Sur les Propriétés Statistiques de l'Entropie de Permutation Multi-échelle et ses Raffinements; applications sur les Signaux Électromyographiques de Surface

Antonio Davalos Trevino

UNIVERSITÉ D'ORLÉANS

ÉCOLE DOCTORALE [MATHÉMATIQUES, INFORMATIQUE, PHYSIQUE THÉORIQUE ET INGÉNIERIE DES SYSTÈMES]

[LABORATOIRE PRISME]

Thèse présentée par :

Antonio DAVALOS-TREVINO

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline/ Spécialité : **Traitement du Signal**

On the Statistical Properties of Multiscale Permutation Entropy and its Refinements, with Applications on Surface Electromyographic Signals

Thèse dirigée par :

Olivier BUTTELLI	Maître de Conférences, Université d'Orléans
Meryem JABLOUN	Maître de Conférences, Université d'Orléans

RAPPORTEURS :

Anne HUMEAU-HEURTIER	Professeur, Université d'Angers
Steeve ZOZOR	Directeur de Recherche, Institut Polytechnique de Grenoble

---

JURY :

Stéphane CORDIER	Professeur, Université d'Orléans, Président du jury
Jean Marc GIRAULT	Professeur, ESEO Grande École d'Ingénieurs Généralistes à Angers
Franck QUAINE	Maître de Conférences, Gipsa-Lab, Grenoble-INP et l'Université de Grenoble-Alpe
Philippe RAVIER	Maître de Conférences, Université d'Orléans

Antonio DÁVALOS

Sur les Propriétés Statistiques de l'Entropie de Permutation Multi-échelle et ses Raffinements; applications sur les Signaux Électromyographiques de Surface

Résumé de l'introduction

L'utilisation de mesures de l'entropie de l'information a permis d'évaluer avec succès la "quantité d'informations" ou la "complexité" d'un système. Parmi les domaines qui peuvent bénéficier de l'application de ces mesures spécifiques, le domaine biomédical mérite d'être mentionné en tant que discipline riche d'enseignements dans laquelle les chercheurs ont accès à des quantités considérables d'informations par l'intermédiaire de leurs examinateurs. En outre, les processus biologiques sont de nature remarquablement complexe, impliquant généralement des étapes complexes, des événements simultanés et des boucles de rétroaction autocorrectrices. Pour illustrer davantage ce point, l'activité bioélectrique est difficile à interpréter, car elle n'est généralement disponible que sous la forme d'un agrégat de multiples signaux élémentaires simultanés qui, même lorsqu'ils sont compris à un niveau isolé, augmentent en complexité dans leurs interactions lorsqu'ils sont observés depuis l'extérieur du corps. C'est précisément pour cette raison que l'application de mesures de l'entropie par des techniques et des processus nouveaux offre une solution prometteuse à ce dilemme, pour autant que nous puissions différencier de manière significative la dynamique des processus en mesurant la complexité.

En outre, la littérature existante associe fortement la réduction de l'entropie à l'avancement des maladies motrices. La diminution de la variabilité suggère une activité motrice stéréotypée, empêchant les ajustements nécessaires pour changer de tâche ou s'adapter à des conditions changeantes. Par conséquent, un ensemble complexe et dynamique d'instructions doit être présent pour accomplir l'une de ces tâches, et une personne souffrant d'une déficience connexe aura du mal ou échouera à les exécuter faute d'une compensation adéquate. Cela s'explique par une réduction de la complexité des instructions que chaque muscle reçoit.

Nous avons décidé d'aborder cette situation en appliquant l'entropie de permutation multi-échelle (EPM). Cette mesure particulière possède un ensemble unique de caractéristiques que nous jugeons appropriées pour la caractérisation de ce problème. Tout d'abord, l'entropie de permutation fonctionne avec des informations ordinales, une qualité qui garantit que la méthode est robuste en ce qui concerne l'amplitude des signaux sources et la distribution de probabilité qu'ils contiennent. Étant donné le large éventail de sorties possibles d'amplitude de signaux provenant de différents sujets - une propriété souhaitable - et le fait que la MPE nécessite un ensemble minimal de paramètres pour fonctionner, il devient facile d'obtenir des résultats

significatifs à partir du système en raison de ses exigences simplifiées. En outre, le calcul du MPE est simple et direct à effectuer, avec une gamme claire de valeurs maximales et minimales. Le prétraitement multi-échelle permet également au chercheur d'explorer les signaux à différentes échelles de temps, permettant ainsi l'identification d'informations cachées à longue portée qui ne sont pas nécessairement disponibles par l'acquisition directe de signaux bruts.

Bien que les mesures du MPE et de ses variantes aient déjà été appliquées avec succès auparavant, la discussion de cette méthode dans la littérature est généralement traitée d'un point de vue algorithmique. Étant donné que les mesures de l'entropie - en particulier l'EMT - sont définies en fonction de la distribution de probabilité de la source, il est logique de compléter ces techniques en y incluant un point de vue et une approche statistiques.

Il existe des travaux antérieurs sur la caractérisation de ce problème, avec un accent particulier sur la dimension d'encastrement et la longueur du signal, mais l'interaction de la statistique avec la procédure de granulation grossière n'a pas été abordée auparavant dans la littérature. Il convient également de mentionner que certaines propriétés statistiques - en particulier le biais - ont déjà été signalées dans la littérature, mais qu'elles ne sont généralement mentionnées qu'après coup. En outre, on connaît peu la variance du PPE, et il s'agit d'une mesure qui ne se conforme pas nécessairement à l'hypothèse de normalité largement appliquée. Compte tenu des raisons susmentionnées, il est possible de mal interpréter le comportement du PPE observé comme une propriété émergente du phénomène en question, alors que la statistique elle-même peut être la source de ces effets indésirables.

Notre premier objectif dans le présent travail est de développer la théorie qui sous-tend la mesure de l'EMT, ainsi que de comprendre son comportement par rapport aux modèles de signaux et aux processus stochastiques connus. Avec ces nouvelles connaissances, nous serions en mesure d'énoncer et de développer notre second objectif : proposer une nouvelle technique de mesure de la MPE comme amélioration par rapport à d'autres méthodes bien établies, y compris le calcul original et ses raffinements. Enfin, nous appliquerons ces méthodes à des signaux électromyographiques (EMG) réels, dans le but d'obtenir une interprétation actualisée de la mesure de l'EMG sur des systèmes biologiques réels, et d'améliorer la précision des techniques ordinales existantes à des fins de classification.

Dans le chapitre 1, nous commençons par présenter au lecteur le paysage global des techniques d'entropie disponibles, en les classant de manière à ce que le lecteur puisse différencier la nature des options possibles qui s'offrent aux chercheurs. Dans le chapitre 2, nous nous concentrons sur les propriétés statistiques de l'entropie de permutation multi-échelle, en fournissant une forme fermée pour l'approximation de la statistique de l'EMT - en prenant en considération sa distribution et l'effet de la longueur du signal - et des expressions générales pour ses deux premiers moments. Au chapitre 3, nous appliquons l'estimation de l'EMT à des modèles de

signaux et des processus stochastiques largement répandus afin d'évaluer plus avant les propriétés de l'entropie dans des conditions connues, ce qui donne une valeur théorique de l'EMT attendue pour les processus stochastiques gaussiens. Dans le chapitre 4, nous explorons les propriétés des affinements de l'EMT afin de mieux comprendre les raisons de leur précision accrue, ce qui a permis de produire une nouvelle méthode d'EMT pour réduire davantage l'incertitude de l'estimation de l'entropie. Enfin, dans le chapitre 5, nous abordons le problème complexe des signaux électromyographiques de surface réels (sEMG) en appliquant les méthodes discutées jusqu'à présent dans divers dispositifs expérimentaux, et nous fournissons une interprétation actualisée des mesures de l'entropie dans ce contexte, tant d'un point de vue statistique que biologique.

Antonio DÁVALOS

Sur les Propriétés Statistiques de l'Entropie de Permutation Multi-échelle et ses Raffinements; applications sur les Signaux Électromyographiques de Surface

Résumé

Introduction

L'utilisation de mesures de l'entropie de l'information a permis d'évaluer avec succès la "quantité d'informations" ou la "complexité" d'un système. Parmi les domaines qui peuvent bénéficier de l'application de ces mesures spécifiques, le domaine biomédical mérite d'être mentionné en tant que discipline riche en idées dans laquelle les chercheurs ont accès à des quantités considérables d'informations par l'intermédiaire de leurs examinateurs. En outre, les processus biologiques sont de nature remarquablement complexe, impliquant généralement des étapes complexes, des événements simultanés et des boucles de rétroaction autocorrectrices. Pour illustrer davantage ce point, l'activité bioélectrique est difficile à interpréter, car elle n'est généralement disponible que sous la forme d'un agrégat de multiples signaux élémentaires simultanés qui, même lorsqu'ils sont compris à un niveau isolé, augmentent en complexité dans leurs interactions lorsqu'ils sont observés depuis l'extérieur du corps. C'est précisément pour cette raison que l'application de mesures de l'entropie par des techniques et des processus nouveaux offre une solution prometteuse à ce dilemme, pour autant que nous puissions différencier de manière significative la dynamique des processus en mesurant la complexité.

En outre, la littérature existante associe fortement la réduction de l'entropie à l'avancement des maladies motrices. L'idée est la suivante : la diminution de la variabilité suggère une activité motrice stéréotypée, entravant les ajustements nécessaires pour changer de tâche ou s'adapter à des conditions changeantes ; se lever ou marcher, par exemple, sont des actions qui nécessitent une compensation constante et fluctuante de la gravité, du vent et d'autres forces externes pour maintenir l'équilibre. Par conséquent, un ensemble complexe et dynamique d'instructions doit être présent pour accomplir l'une de ces tâches, et une personne souffrant d'une déficience connexe aura du mal ou échouera à les accomplir faute d'une compensation adéquate. Cela s'explique par une réduction de la complexité des instructions que chaque muscle reçoit.

Nous avons décidé d'aborder cette situation en appliquant l'entropie de permutation multi-échelle (EPM). Cette mesure particulière possède un ensemble unique de caractéristiques que nous jugeons appropriées pour la caractérisation de ce problème. Tout d'abord, l'entropie de permutation fonctionne avec des informations ordinales, une qualité qui garantit que la méthode

est robuste en ce qui concerne l'amplitude des signaux sources et la distribution de probabilité qu'ils contiennent. Étant donné le large éventail de sorties possibles d'amplitude de signaux provenant de différents sujets - une propriété souhaitable - et le fait que la MPE nécessite un ensemble minimal de paramètres pour fonctionner, il devient facile d'obtenir des résultats significatifs à partir du système en raison de ses exigences simplifiées. En outre, le calcul du MPE est simple et direct à effectuer, avec une gamme claire de valeurs maximales et minimales. Le prétraitement multi-échelle permet également au chercheur d'explorer les signaux à différentes échelles de temps, permettant ainsi l'identification d'informations cachées à longue portée qui ne sont pas nécessairement disponibles par l'acquisition directe de signaux bruts.

Bien que les mesures du MPE et de ses variantes aient déjà été appliquées avec succès auparavant, la discussion de cette méthode dans la littérature est généralement traitée d'un point de vue algorithmique. Étant donné que les mesures de l'entropie - en particulier l'EMT - sont définies en fonction de la distribution de probabilité de la source, il est logique de compléter ces techniques en y incluant un point de vue et une approche statistiques ; par exemple, nous n'obtiendrions qu'une approximation de la distribution réelle du phénomène en raison des informations limitées présentes sur les sources de signaux.

Il existe des travaux antérieurs sur la caractérisation de ce problème, avec un accent particulier sur la dimension d'intégration et la longueur du signal, mais l'interaction de la statistique avec la procédure de grossièreté n'a pas été abordée auparavant dans la littérature. Il convient également de mentionner que certaines propriétés statistiques - en particulier le biais - ont déjà été signalées dans la littérature, mais qu'elles ne sont généralement mentionnées qu'après coup. En outre, on connaît peu la variance du PPE, et il s'agit d'une mesure qui ne se conforme pas nécessairement à l'hypothèse de normalité largement appliquée. Compte tenu des raisons susmentionnées, il est possible de mal interpréter le comportement du PPE observé comme une propriété émergente du phénomène en question, alors que la statistique elle-même peut être la source de ces effets indésirables.

Notre premier objectif dans le présent travail est de développer la théorie qui sous-tend la mesure de l'EMT, ainsi que de comprendre son comportement par rapport aux modèles de signaux et aux processus stochastiques connus. Avec ces nouvelles connaissances, nous serions en mesure d'énoncer et de développer notre second objectif : proposer une nouvelle technique de mesure de la MPE comme amélioration par rapport à d'autres méthodes bien établies, y compris le calcul original et ses raffinements. Enfin, nous appliquerons ces méthodes à des signaux électromyographiques (EMG) réels, dans le but d'obtenir une interprétation actualisée de la mesure de l'EMG sur des systèmes biologiques réels, et d'améliorer la précision des techniques ordinales existantes à des fins de classification.

Chapitre 1. Entropie de l'information - Concepts et définitions

Avec son article fondateur, C.E. Shannon a établi les bases de la théorie de l'information. L'auteur a défini le concept de mesure de l'information, y compris l'entropie de l'information.

En termes simples, l'entropie est une mesure de l'imprévisibilité. Étant donné une chaîne de symboles, un niveau d'entropie élevé implique qu'un symbole donné ne peut pas être facilement prédit en regardant les symboles qui le précèdent dans la chaîne ; inversement, une faible valeur d'entropie implique que chaque symbole peut être déduit de son histoire. Une chaîne de symboles totalement imprévisible produira l'entropie maximale possible par le système, tandis que l'entropie minimale se produira lorsqu'un seul symbole se répète sans faute ; dans un sens plus général, cependant, il est habituel qu'il y ait une certaine probabilité pour que chaque symbole apparaisse dans une chaîne donnée. On peut donc en déduire intuitivement que cette mesure nécessite une certaine connaissance de la distribution de probabilité et, par conséquent, que l'entropie doit être fonction de la fonction de probabilité associée à la chaîne.

Bien que l'entropie mesure une propriété précise d'un ensemble aléatoire basé sur des données brutes, une interprétation correcte dépend fortement du contexte, car il manque une interprétation universelle de ces mesures. Alors que l'entropie est définie dans le contexte de la thermodynamique - la discipline dont elle s'inspire - comme une mesure de l'irréversibilité de tout processus thermodynamique (selon la définition de Rudolph Clausius), l'entropie est appelée contenu, diversité ou complexité dans le contexte de l'information. Néanmoins, l'utilisation de l'entropie de l'information peut toujours être justifiée d'un point de vue pragmatique, car elle a été utilisée avec succès dans une grande variété d'applications, telles que la compression de données et les applications dans les domaines de la finance et de la biomédecine. Même lorsque la véritable nature et la signification de l'entropie de l'information font l'objet d'un débat, cette mesure a ses utilisations et ses mises en œuvre.

Depuis sa formulation originale en 1948, un grand nombre de variantes de l'entropie ont été proposées. La quantité d'options proposées à ce jour est impressionnante : pour chaque petit changement dans la méthode, la source de données ou l'application, il existe une mesure de l'entropie avec un nom propre et ses propres considérations. De plus, chaque variante présente ses propres avantages et inconvénients, ce qui prouve qu'il n'existe pas de solution unique à ce jour.

Avant de pouvoir fouiller directement dans les rouages de cette méthode, nous devons d'abord présenter un paysage approprié pour l'utilisation de l'entropie de l'information, y compris ses variantes et ses applications courantes.

Nous présenterons tout au long de ce chapitre les variantes d'entropie les plus courantes à chaque niveau d'analyse. Nous explorerons ensuite les méthodes avec une approche "inside-out", en commençant par les variations de la définition de l'entropie de base. Ensuite, nous

suivrons les variantes possibles dans la formulation de l'ensemble des événements relatifs au phénomène. La dernière étape consistera à examiner les variantes de prétraitement possibles sur le signal source - où les données brutes sont préparées pour l'analyse de l'entropie - et à discuter brièvement des raffinements apportés et du raisonnement qui sous-tend leur sélection. Bien que nous fournissions un panorama général de toutes les variantes d'entropie, la liste présentée et la discussion qui l'accompagne sont loin d'être exhaustives.

Pour les besoins de ce projet, nous nous concentrerons sur l'entropie de permutation multi-échelle, car ses propriétés particulières répondent aux besoins et aux exigences de l'analyse des signaux bioélectriques.

Chapitre 2. Entropie de permutation multi-échelle - Statistiques théoriques

À partir de toutes les mesures d'entropie possibles, nous avons plaidé en faveur de l'entropie de permutation multi-échelle (MPE) comme extension de l'entropie de permutation (PE) classique. L'EPM peut être conceptualisée comme suit :

- Équation : L'entropie de Shannon.
- Partition d'événement : motifs ordinaux.
- Preprocessing : procédure de grossièreté.

Le travail avec des modèles ordinaux produit une mesure de l'entropie qui est invariante par rapport à l'amplitude du signal, et en particulier à la présence de valeurs aberrantes. Bien que l'amplitude contienne des informations pertinentes, nous nous intéressons davantage à la forme fonctionnelle du signal - un scénario digne d'intérêt serait celui où la variabilité biologique entre les sujets rendrait les comparaisons difficiles. En outre, le MPE présente également l'avantage de disposer d'un ensemble d'événements définis avant l'introduction de l'ensemble de données à analyser, ce qui évite de travailler avec des partitions d'événements approximatives.

La simplicité de l'analyse MPE permet également de réduire au minimum le besoin de calibrage des paramètres, puisqu'aucune valeur de tolérance n'est nécessaire. En outre, d'un point de vue informatique, le comptage de modèles ordinaux est un processus rapide, un facteur qui peut potentiellement conduire à des applications en temps réel. Le processus de mise à l'échelle multiple, ajouté à l'EPP originale, permet aux chercheurs d'explorer le contenu d'informations dans des échelles de temps qui ne sont pas directement mesurées avec le signal brut, ce qui élargit la portée de l'analyse originale. Néanmoins, la méthode n'est pas sans inconvénients : le comptage de modèles ordinaux nécessite l'utilisation d'un ensemble de données suffisamment important pour être fiable, une question qui devient cruciale lorsque nous explorons l'approche multi-échelle dans une situation où la longueur du signal est réduite avec l'augmentation des

échelles. Toute considération initiale concernant l'analyse MPE doit tenir compte de la longueur du signal.

Il est important de se rappeler que les formulations de l'EMT sont généralement présentées sous forme algorithmique et que la distribution ordinale du modèle est mesurée directement à partir du signal source, car cela implique que toute information recueillie de cette manière n'est qu'une estimation de la véritable distribution du modèle, et donc que la mesure de l'EMT elle-même est également une estimation. Par conséquent, la théorie statistique qui sous-tend l'EPP n'est pas très étudiée dans la littérature, à quelques exceptions près. Cela justifie l'approche consistant à considérer le PPE comme une statistique.

Cela dit, dans ce chapitre, nous allons approfondir les propriétés statistiques du PPE et améliorer le cadre théorique existant en développant les résultats que nous avons présentés pour la première fois dans des articles précédents. Nous commençons par une définition formelle du PPE et du processus de formation des gros grains. Nous commentons également les considérations concernant le signal source en tant que processus aléatoire, puis nous développons un modèle statistique de la mesure du MPE au moyen de l'expansion polynomiale de la série de Taylor. Cela nous permettra d'avoir une expression approximative de la valeur, du biais et de la variance attendus de l'EPP. Nous fournirons également l'expression de la limite inférieure de Cramér-Rao afin d'évaluer l'efficacité de l'estimateur. Ensuite, nous testons nos résultats théoriques par rapport à un modèle de substitution avec un ensemble de paramètres facilement modifiables. Enfin, nous discuterons et commenterons les idées préliminaires que nous avons obtenues grâce au développement de la théorie statistique du PPE.

Nous avons d'abord constaté que la valeur attendue de l'EMT est un estimateur biaisé. De plus, le biais dépend uniquement des paramètres de l'analyse de l'EPP, en particulier la dimension, l'échelle et la longueur du signal. Cela implique que l'EPP présentera le même biais par rapport à l'échelle de temps, quelle que soit la distribution de probabilité du signal.

Deuxièmement, nous avons constaté que la variance de l'EPP augmente presque linéairement avec l'augmentation de l'échelle de temps pour presque toute distribution de modèle. L'exception apparaît lorsque la MPE est proche d'une valeur maximale (probabilités uniformes). Dans ce scénario, la variance augmente de façon quadratique par rapport à l'échelle de temps. Notre formulation ressemble beaucoup à la limite inférieure de Cramér-Rao pour la statistique de l'EPP, ce qui signifie qu'elle est presque aussi efficace. Nous devons également ajouter que la variance présente une valeur maximale pour des valeurs spécifiques de l'EPP et des distributions de probabilité de modèle, car cela informe les autres chercheurs sur une région de l'EPP où nous avons une incertitude maximale.

Enfin, nous avons pu suggérer un critère plus précis pour la longueur du signal que l'expression habituelle dans la littérature. En définissant un biais maximum autorisé, nous avons pu spécifier une longueur minimum (et une échelle de temps maximum) basée uniquement sur la dimension du modèle pour l'analyse.

Chapitre 3. MPE sur les modèles de signaux communs

Nous avons étudié les rouages de l'entropie de permutation multi-échelle d'un point de vue empirique et, surtout, statistique. C'est grâce à la série de Taylor que nous avons pu développer une approximation du PPE qui nous permet de calculer sa valeur et sa variance attendues. Nous savons à ce stade que l'EPP est un estimateur biaisé dont la valeur linéaire ne dépend que des paramètres de l'analyse et non de la distribution des modèles. Nous avons également constaté que la variance est proche de la limite inférieure de Cram\'er-Rao, et donc, approximativement efficace. Nous avons également été en mesure de proposer un critère de longueur plus précis pour un signal suffisamment long pour l'analyse MPE. Enfin, nous avons établi une plage de MPE normalisée où la variance est maximale pour une longueur de signal donnée.

Bien que ces résultats soient valables pour des signaux arbitraires, nous ne pouvons pas ignorer le fait que la procédure à gros grains a un effet notable sur la distribution des modèles trouvée à chaque échelle. Il est maintenant temps d'aborder directement cette relation.

Nous pouvons appliquer l'analyse MPE à n'importe quel signal discret sans avoir besoin de connaître au préalable sa dynamique sous-jacente. En fait, le calcul empirique du PPE peut nous donner quelques indications à cet égard. D'autre part, si nous connaissons la nature du processus, nous pouvons calculer un PPE théorique basé sur le modèle du signal. Nous saurons si le modèle du signal que nous proposons peut être approprié pour expliquer toutes les informations pertinentes du phénomène lorsque le PPE théorique correspond au PPE empirique.

C'est pourquoi, dans ce chapitre, nous étudions les valeurs du PPE pour certains modèles de signaux bien connus, en abordant d'abord le PPE attendu des signaux déterministes communs. Comme nous nous attendons à ce que l'EPP soit robuste aux perturbations, nous analyserons également l'effet du bruit, dans le but d'évaluer quand le bruit domine sur le signal et de tester les limites de robustesse des méthodes. Bien que notre exploration de l'EMT envisage des processus aléatoires, les signaux gaussiens sont particulièrement pertinents dans le cadre du présent travail ; plus précisément, nous caractériserons à la fois le bruit blanc gaussien (wGn) et le bruit fractionnaire gaussien (fGn). En outre, nous explorerons les modèles autorégressifs (AR) et à moyenne mobile (MA) du premier ordre. Comme tous les signaux gaussiens présentent des symétries particulières, nous proposerons une formulation explicite et générale de l'EPM théorique pour ce type de processus aléatoires, et nous conclurons en testant notre technique

proposée sur des modèles plus élaborés, tels que le modèle général autorégressif et à moyenne mobile (ARMA).

Ces résultats permettront aux chercheurs de mieux évaluer les résultats attendus du PPE à partir de modèles bien établis. Nous fournirons le MPE de référence approprié pour comparer le contenu informatif entre les ensembles de données réels et les modèles utilisés pour les décrire.

En traitant les signaux déterministes, nous avons constaté que la distribution de probabilité du modèle est fixée par la région où la pente de la courbe est positive ou négative. Le taux d'échantillonnage joue également un rôle important, et augmente presque invariablement la valeur de l'EPP. Avec un taux d'échantillonnage élevé, l'EPP converge lentement avec l'EPP d'une courbe continue théorique. L'ajout de bruit blanc a également été pris en compte. Nous avons constaté que l'effet du bruit sur l'EPP dépend fortement de la relation entre la pente de la courbe et l'amplitude du bruit. Si la pente est suffisamment élevée, le bruit n'a pas d'effet visible sur l'EMT. En revanche, dans les régions à faible pente, le bruit domine. Néanmoins, si le taux d'échantillonnage d'un signal bruyant est trop élevé, nous obtiendrons des valeurs de l'EMT caractéristiques du bruit blanc, quelle que soit la pente.

Nous avons également étudié le MPE attendu pour les processus gaussiens corrélés couramment utilisés. Au moyen de symétries de motifs pour ces signaux, nous sommes en mesure de calculer l'EPP en fonction de la fonction d'autocorrélation du signal. Étant donné que la procédure à gros grains est une combinaison linéaire des éléments du signal, le signal à gros grains qui en résulte est également gaussien. Par conséquent, il suffit de connaître la fonction d'autocorrélation à gros grains pour obtenir l'EPP pour n'importe quelle échelle.

En explorant le bruit blanc gaussien et le bruit fractionnaire gaussien, nous concluons que l'EMT est invariante à l'échelle temporelle pour ces processus. Les processus autorégressifs et de moyenne mobile de premier ordre ont une expression élaborée, mais finalement fermée, pour la distribution de probabilité du modèle, qui ne dépend que des paramètres des modèles et de l'échelle de temps.

Dans ce chapitre, nous avons également proposé une expression générale pour l'autocorrélation à gros grains pour un signal arbitraire, au moyen de formes matricielles quadratiques. Cela nous permet de calculer l'EPP théorique d'un signal sans connaître explicitement la fonction d'autocorrélation à gros grains. Nous avons présenté quelques exemples de modèles ARMA avec un nombre arbitraire de paramètres, et avons testé les résultats théoriques par rapport aux simulations, avec des résultats satisfaisants.

Cette analyse prétend être une exploration approfondie des multiples facteurs qui influencent la MPE d'un signal arbitraire. La pente du signal, le taux d'échantillonnage et les fonctions d'autocorrélation s'avèrent primordiales dans la valeur attendue de l'EMP. La recherche des

interactions entre ces facteurs et l'étude des dimensions arbitraires feront l'objet de travaux futurs.

Chapitre 4. Affinements des MPE composites

L'EMT et d'autres méthodes d'entropie ont été soumises à différents perfectionnements afin d'augmenter la précision de l'estimation de l'entropie résultante. La MPE composite (cMPE) et la MPE composite affinée (rcMPE) visent toutes deux à mesurer le plus grand nombre de modèles possibles dans le signal original sans modifier l'idée sous-jacente de l'approche MPE. Il a été prouvé expérimentalement que ces deux méthodes donnent de meilleurs résultats en réduisant la variance de l'estimateur MPE.

C'est pourquoi, dans ce chapitre, nous allons développer davantage la théorie statistique du MPE en incluant les algorithmes cMPE et rcMPE. Nous présenterons et discuterons les améliorations qu'ils offrent par rapport à l'approche classique du PPE, ainsi que leurs inconvénients et leurs éventuelles lacunes. En outre, nous présentons une alternative à l'approche composite classique de la méthode des grains grossiers - connue sous le nom de "downsampling" - qui améliore encore ces méthodes raffinées, en évitant les informations redondantes provenant des corrélations croisées d'artefacts inhérentes à la procédure des grains grossiers. Enfin, comparer expérimentalement toutes les méthodes discutées précédemment pour évaluer leur précision et recommander l'algorithme le plus approprié pour la mesure de l'entropie ordinale dans les séries temporelles.

Nous avons constaté que les techniques de sous-échantillonnage composite réduisent considérablement la variance des approches cMPE et rcMPE que l'on trouve dans la littérature. Plus précisément, la rcMPE sous-échantillonnée, en plus de présenter la variance minimale, a également montré une valeur attendue qui est restée invariante par rapport à l'échelle de temps. Cela sera particulièrement utile, car nous ne serons pas confrontés à des dégradations notables lorsque nous explorerons de grandes échelles de temps. C'est pourquoi, pour des raisons pratiques, nous recommandons l'utilisation de cette méthode.

Étant donné que la corrélation croisée des artefacts des techniques composites à gros grains est complètement évitée par l'utilisation du processus de sous-échantillonnage composite, nous n'avons pas caractérisé ce phénomène explicitement à des fins pratiques - c'est toutefois un problème mathématique intéressant qui mérite des recherches plus approfondies.

Chapitre 5. Applications des signaux bioélectriques

Nous disposons maintenant des outils nécessaires pour appliquer ces méthodes à des ensembles de données réels. Nous allons donc maintenant présenter les applications biomédicales du MPE.

Les mesures de modèles ordinaux ont été utiles ces dernières années pour mesurer la complexité des systèmes biologiques, en particulier ceux liés à l'activité électrique. Ces types de signaux sont caractérisés par une dynamique complexe, même au repos. Certains cas notables impliquent une activité cérébrale spontanée présentant un comportement complexe et non aléatoire, et même l'activité pathologique des crises d'épilepsie est caractérisée par une séquence ordonnée d'événements. Alors que la plupart des méthodes nécessitent des hypothèses supplémentaires concernant les caractéristiques déterministes et aléatoires du signal, les mesures de l'EP ont l'avantage supplémentaire d'être exemptes de modèle et robustes. Enfin, comme nous l'avons déjà mentionné, les techniques de PE sont rapides à exécuter, une caractéristique intéressante pour les applications en temps réel sans autre prétraitement, comme les contextes impliquant des signaux bioélectriques.

Le placement d'électrodes à la surface de la peau permet de mesurer les champs électriques générés par le cœur (électrocardiogramme, ECG), le cerveau (électroencéphalogramme, EEG) ou le système neuromusculaire (électromyographie, EMG). Dans ce dernier cas, l'électromyographie a été largement utilisée pour acquérir des connaissances fondamentales sur le processus de recrutement de l'unité fonctionnelle musculaire - appelée unité motrice - depuis les premiers travaux de Piper utilisant les signaux EMG.

Les études susmentionnées ont montré jusqu'à présent que les méthodes EMG sont bien adaptées à l'analyse des comportements qui impliquent une contraction musculaire. Au-delà des aspects fondamentaux, leurs domaines d'application sont divers : c'est le cas du sport et de l'ergonomie, qu'il s'agisse de réaliser des exercices isométriques ou dynamiques ; dans des applications cliniques et technologiques, telles que la rééducation, le biofeedback et les interfaces myoélectriques pour le contrôle de dispositifs prothétiques ou l'interaction informatique. En outre, cette technique apporte également un éclairage sur l'apprentissage moteur et les troubles neurologiques.

Ce chapitre se penchera sur les applications de l'EP sur les signaux EMG, en particulier les EMG de surface (sEMG).

Comme l'acquisition de données sEMG à partir de la surface de la peau est nettement moins invasive que les techniques traditionnelles de détection à l'aiguille et au fil, la première peut être mise en œuvre dans un ensemble de conditions beaucoup plus diverses et flexibles, tandis que la seconde conserve des applications cliniques limitées.

Tout au long de ce chapitre, nous avons passé en revue les propriétés et les mécanismes physiologiques des signaux sEMG, tant du point de vue de la biomédecine que de la théorie de

l'information. Nous avons conçu une expérience concernant les contractions isométriques qui envisageait différentes conditions de force et de fatigue, avec l'intention d'évaluer la validité et la performance de différentes mesures de l'entropie de permutation (MPE, rcMPE et rcDPE). Par la suite, alors que nous cherchions les paramètres optimaux pour maximiser la différence d'entropie entre les conditions de contraction musculaire, nous avons décidé de tester ces calculs sous l'effet de l'échelle de temps et des variables dimensionnelles d'intégration. Enfin, nous avons effectué une batterie de tests statistiques afin de déterminer la signification statistique des résultats obtenus.

Avant tout, nous avons déterminé que la méthode rcDPE surpasse les autres méthodes dans la différenciation des niveaux de fatigue dans les contractions isométriques. Nous avons trouvé les différences d'entropie maximales à une échelle de temps correspondant à une fréquence de 1 000 Hz, et nous avons conclu que la dimension d'encastrement est un facteur important, car une augmentation de la valeur de la dimension rend la différence d'entropie plus évidente.

Ces méthodes d'entropie ont également permis de détecter les différences entre les niveaux de force. Comme l'échelle de temps optimale pour la différenciation des MPE était la première échelle de temps, les méthodes MPE donnent exactement les mêmes résultats - le rcDPE offrant de petits avantages lorsqu'il est utilisé dans cette échelle. Néanmoins, la dimension d'encastrement s'est révélée importante en montrant une capacité de différenciation accrue à des valeurs plus élevées, bien que moins prononcée que dans le cas de l'ensemble de données sur la fatigue.

Toutes les contractions observées présentent une augmentation rapide de l'entropie à des échelles de temps basses et se stabilisent ensuite. Comme nous savons, grâce aux chapitres précédents, que cette ligne d'entropie horizontale ne peut pas être distinguée du bruit non corrélé, il est sous-entendu que les échelles de temps élevées ne fournissent aucune information concernant l'activité dynamique des sEMG.

D'un point de vue biomédical, les résultats de ce chapitre sont en accord avec la littérature actuelle. La réduction observée de l'EPM est liée à la réduction de la complexité du signal sEMG, qui sont des effets bien établis de la fatigue. En ce qui concerne les niveaux de production de force, la différence de MPE suggère un modèle d'activité différent à différents pourcentages de contraction volontaire maximale. Il est intéressant de noter que la MPE, par définition, ne tient pas compte des informations contenues dans l'amplitude du signal sEMG. Par conséquent, toute différenciation des signaux à différents niveaux de force doit provenir d'un schéma différent de recrutement et de cadence de tir des unités motrices.

Conclusions

Dans le présent travail, nous avons exploré les propriétés cachées et sous-jacentes de l'entropie de permutation multi-échelle (EPM). Notre premier objectif était de définir et d'enrichir le corpus de connaissances théoriques qui sous-tend l'algorithme de l'EMP, car cela nous permettrait d'évaluer correctement ses propriétés statistiques, ses avantages et ses limites. Notre deuxième objectif était de proposer une nouvelle méthode d'EMT qui tire parti de ces résultats théoriques. Notre troisième et dernier objectif était d'appliquer ces connaissances à un problème biomédical complexe, tel que l'activité électrique des muscles, afin de différencier la fatigue et la production de force en tant qu'états de performance.

Pour mieux nous positionner dans le contexte de la théorie de l'information, nous avons proposé au chapitre 1 un critère général de classification des mesures d'entropie les plus couramment utilisées en ce qui concerne sa formulation de base, la définition de l'ensemble des événements et les techniques de prétraitement utilisées avant le calcul de l'entropie. Bien que de nombreuses extensions, améliorations et généralisations différentes aient été proposées depuis les travaux initiaux de Shannon, il n'existe actuellement aucune méthode universelle optimale pour le calcul de l'entropie, car les particularités du phénomène en question doivent être prises en compte ; même la simple définition et la nature de l'entropie ne peuvent être interprétées que dans le contexte de l'expérience spécifique (complexité, quantité d'informations, etc.). Par conséquent, notre objectif est de fournir une vue d'ensemble des variantes d'entropie que nous pouvons mettre en œuvre et explorer, du point de vue mathématique et statistique.

Nous avons présenté notre principal développement théorique du PPE au chapitre 2. Au moyen d'une expansion polynomiale, nous avons pu trouver une approximation analytique pour la mesure de l'EMT, ce qui nous a permis de trouver une expression fermée pour ses deux premiers moments. Nous avons trouvé le biais de MPE, dont l'approximation est indépendante de la distribution du modèle, en ne prenant en compte que la dimension d'encastrement, la longueur du signal et l'échelle. Nous avons également caractérisé la variance de MPE, qui est étroitement liée à la mesure de MPE elle-même. Ici, nous avons trouvé que notre approximation de la variance de l'EPP ressemble beaucoup à la limite inférieure de Cramér-Rao pour la variance minimale. Même en tant que statistique biaisée, l'estimation est presque efficace. Forts de ces nouvelles connaissances, nous avons proposé un critère de longueur minimale plus précis pour l'EPP. Nous avons également indiqué les valeurs de l'EPP avec une incertitude maximale le long de la plage d'entropie normalisée.

Nous avons exploré les résultats attendus pour différents modèles de signaux au chapitre 3, dans le but d'étudier l'effet des différentes propriétés des signaux sur les résultats globaux du MPE. Nous avons constaté que l'entropie des signaux déterministes est affectée par la pente du signal, le taux d'échantillonnage et l'amplitude du bruit - bien que la méthode soit assez robuste au bruit lorsque le signal a une pente prononcée. Par la suite, nous avons exploré l'EMT des processus

stochastiques, en particulier le bruit gaussien fractionnaire (qui est fractal) et les modèles ARMA, où nous avons constaté qu'il est possible d'estimer un résultat EMT théorique à partir des paramètres des processus. D'une manière générale, cela implique que les paramètres qui définissent un processus aléatoire contiennent toutes les informations du processus lui-même, et qu'il est possible de tester les signaux réels par rapport aux modèles proposés en comparant les mesures d'entropie.

À ce stade, notre base mathématique était suffisante pour aborder les propriétés statistiques des méthodes MPE plus raffinées. Nous avons exploré les propriétés du MPE composite bien établi (cMPE) et du MPE composite raffiné (rcMPE). Bien que l'amélioration de l'estimation du MPE - en particulier la rcMPE, où le biais et la variance sont tous deux réduits - soit bien établie, nous avons constaté que la méthode des grains grossiers composites introduit une corrélation croisée entre les signaux grossiers possibles. Bien que la variance globale soit réduite par rapport à l'algorithme MPE original, cet effet ajoute une source artificielle d'incertitude. Nous avons proposé ici une procédure de sous-échantillonnage composite comme substitut au gros grain classique utilisé pour les techniques d'entropie multi-échelles. Cette approche a entièrement évité le problème de la corrélation croisée des artefacts, ce qui a permis d'augmenter la précision par rapport aux méthodes utilisées dans la littérature existante. En particulier, l'entropie composite de permutation de sous-échantillonnage (rcDPE), en plus d'avoir la plus petite variance parmi les méthodes discutées ici, a également l'avantage d'un biais constant pour tout paramètre de sous-échantillonnage à une échelle particulière. Contrairement aux deux autres méthodes, cela permet à la méthode d'utiliser des valeurs plus élevées à la fois en termes d'échelle et de dimension, et c'est donc la technique que nous recommandons pour une approche d'entropie ordinale.

Enfin, au chapitre 4, nous avons pu tester ces outils et méthodes sur des signaux réels. Les ensembles de données que nous avons choisis sont des signaux électromyographiques de surface (sEMG), qui sont pratiques à mettre en œuvre car leurs méthodes sont de nature non invasive. Néanmoins, certains des défis que présente cette technique sont les sources de bruit et la superposition de signaux multiples, qui dépend elle-même de facteurs tels que la géométrie, la conductivité et une myriade de considérations biologiques. Nous avons découvert que les méthodes d'EPM - le rcDPE affichant les meilleurs résultats parmi elles - sont capables de discriminer de manière cohérente entre différents états de fatigue musculaire, en particulier pour les dimensions à forte encombrement. Malgré le fait que les méthodes MPE n'étaient pas aussi cohérentes lorsque nous avons essayé de trouver des différences entre les différentes sorties de force, nous avons quand même pu différencier les pourcentages de contraction volontaire maximale (CMV) avec des résultats statistiquement significatifs. Comme les méthodes ordinales excluent normalement l'amplitude, cette divergence implique qu'il existe encore une

dynamique de contraction musculaire non découverte lorsque différentes puissances de force sont appliquées.

D'autre part, le choix des bons paramètres est important pour cette classification, et une sélection adéquate des valeurs tant pour la dimension que pour l'échelle n'est pas nécessairement évidente a priori : en général, les dimensions les plus élevées (dans une fourchette raisonnable) permettent une meilleure différenciation entre les états d'activité, même si les rendements diminuent ; inversement, il n'existe pas d'échelle de temps universellement définie à choisir pour une analyse adéquate, et elles doivent être évaluées au cas par cas. En ce qui concerne les implications biomédicales de ces résultats, nous avons constaté une réduction significative de l'entropie lorsque les muscles se fatiguent. Une explication possible est l'allongement du potentiel d'action - un produit des changements électrophysiologiques - dû à la contraction continue. Nous avons également constaté une diminution significative de l'entropie en présence de contractions avec une force de sortie élevée, qui peut être expliquée par des chevauchements avec les potentiels d'action des unités motrices, ainsi que par la synchronisation observée de la vitesse de tir des unités motrices.

Il y a largement place pour des recherches plus approfondies sur le front théorique de ce sujet, telles que l'utilisation de nouvelles définitions de l'entropie de base dans le contexte ordinal, l'exploration de nouvelles partitions d'événements, ou même l'essai de modèles stochastiques plus généraux. Plus important encore, il serait possible de revisiter certains des ensembles de données biomédicales bien établis et d'obtenir une interprétation plus approfondie des résultats, en particulier lors de l'exploration du comportement entropique à des échelles élevées. En raison de la précision accrue des méthodes proposées ici, elles peuvent être appliquées à la recherche de comportements dynamiques précédemment cachés. Nous espérons que ce projet de recherche contribuera à la fois à élargir le corpus de connaissances mathématiques existant et à améliorer encore l'utilisation des techniques d'entropie au service des sciences médicales.

En outre, la recherche sur les méthodes d'entropie ordinale est loin d'être terminée. D'un point de vue théorique, nous pouvons explorer les propriétés statistiques de l'EMT en utilisant des formulations de base qui diffèrent de la définition originale de Shannon. Bien entendu, nous devrions également intégrer la dynamique des techniques "conscientes de l'amplitude" à la théorie générale du comportement statistique de l'EMT. Dans le domaine des modèles de signaux du MPE, la différence entre la complexité et le caractère aléatoire n'est pas encore complètement établie, et nous pensons que l'étude de processus statistiques plus élaborés et de signaux déterministes chaotiques pourrait apporter un éclairage supplémentaire sur ce sujet.

En ce qui concerne les méthodes composites, la compréhension de l'interaction entre les signaux grossiers reste incomplète en raison d'un manque de caractérisation correcte de l'effet de corrélation croisée des artefacts. Théoriquement, il est fondamental de disposer d'une meilleure

proposition concernant la distribution de probabilité des modèles ordinaux, qui est assez similaire - mais pas strictement identique - à une distribution multinomiale.

Enfin, l'étude des méthodes MPE pour la caractérisation des signaux bioélectriques est encore un domaine d'étude fertile. Nous espérons que les méthodes d'entropie que nous proposons - en particulier le rcDPE - permettront de mieux différencier les signaux sEMG dans diverses conditions. Ces méthodes pourraient même être appliquées à des conditions réelles en raison de leur temps de traitement court et de leur précision accrue, car ces scénarios n'offrent généralement pas le luxe de bonnes conditions de mesure, grâce à des facteurs tels que les sources de bruit externes ou les salves d'activité de longue durée. En outre, l'effet du biais et de la variance de l'EMT deviendra crucial dans les études ultérieures impliquant des signaux courts et des conditions plus dynamiques, car ces effets affecteront plus directement l'EMT résultante. En outre, l'étude des simulations de signaux sEMG peut éclairer davantage la dynamique sous-jacente de ces méthodes d'entropie, en particulier pour les contractions de différents niveaux de force, qui nécessitent une exploration plus approfondie.

Le présent travail montre la grande complexité des techniques d'entropie ordinale et leurs applications potentielles ultérieures sur les systèmes biologiques. Même si le corps théorique reste encore incomplet, les améliorations possibles des résultats permettent aux chercheurs de faire des calculs et des prévisions plus fins et plus précis concernant des questions de santé, telles que les processus moteurs. Cela dit, nous espérons que ce projet de recherche apportera plus de clarté sur les méthodes susmentionnées, et qu'il ouvrira la voie à d'autres recherches et mises en œuvre technologiques.

Centre Val de Loire UNIVERSITÉ Spirit of the Loire Valley

UNIVERSITÉ D'ORLÉANS

UNIVERSITE D'ORLEANS

ÉCOLE DOCTORALE [MATHÉMATIQUES, INFORMATIQUE, PHYSIQUE THÉORIQUE ET INGÉNIERIE DES SYSTÈMES]

[LABORATOIRE PRISME]

Thèse présentée par :

Antonio DAVALOS-TREVINO

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline/ Spécialité : **Traitement du Signal**

On the Statistical Properties of Multiscale Permutation Entropy and its Refinements, with Applications on Surface Electromyographic Signals

Thèse dirigée par :

Olivier BUTTELLI	Maître de Conférences, Université d'Orléans
Meryem JABLOUN	Maître de Conférences, Université d'Orléans

RAPPORTEURS :

Anne HUMEAU-HEURTIER	Professeur, Université d'Angers
Steeve ZOZOR	Directeur de Recherche, Institut Polytechnique de Grenoble

---

JURY :

Stéphane CORDIER	Professeur, Université d'Orléans, Président du jury
Jean Marc GIRAULT	Professeur, ESEO Grande École d'Ingénieurs Généralistes à Angers
Franck QUAINE	Maître de Conférences, Gipsa-Lab, Grenoble-INP et l'Université de Grenoble-Alpe
Philippe RAVIER	Maître de Conférences, Université d'Orléans

Acknowledgements

I want to thank Olivier Buttelli for helping me bring my abstract work back to Earth through his guidance, advice and pragmatic perspective. I would also like to thank Meryem Jabloun, whose dynamic discussions and sharp suggestions ignited many of the ideas contained here, and Philippe Ravier, whose methodical and systematic observations aided me in condensing the topics at hand.

I would like to extend my gratitude to the jury members: Anne Humeau-Heutier, Steeve Zozor, Stéphane Cordier, Jean-Marc Girault, and Franck Quaine for their examination. I would also like to acknowledge and thank the Consejo Nacional de Ciencia y Tecnología (CONACYT), for they provided the necessary funding for this research project.

I would also like to thank Juan Mattei, whose editing advice was invaluable.

Special thanks my parents, Antonio and Dora, as well as Lydia, Alejandro, and Pablo for their unconditional support.

Last but not least, I want to thank Edna, for she inspires me every day to be the best version of myself.

On the Statistical Properties of Multiscale Permutation Entropy and its Refinements, with Applications on Surface Electromyographic Signals

Antonio Dávalos-Treviño

Abstract

Permutation entropy (PE) and multiscale permutation entropy (MPE) are extensively used to measure regularity in the analysis of time series, particularly in the context of biomedical signals. As accuracy is crucial for researchers to obtain optimal interpretations, it becomes increasingly important to take into account the statistical properties of MPE.

Therefore, in the present work we begin by expanding on the statistical theory behind MPE, with an emphasis on the characterization of its first two moments in the context of multiscaling. Secondly, we explore the composite versions of MPE in order to understand the underlying properties behind their improved performance; we also created an entropy benchmark through the calculation of MPE expected values for widely used Gaussian stochastic processes, since that gives us a reference point to use with real biomedical signals. Finally, we differentiate between muscle activity dynamics in isometric contractions through the application of the classical and composite MPE methods on surface electromyographic (sEMG) data.

As a result of our project, we found MPE to be a biased statistic that decreases with respect to the multiscaling factor, regardless of the signal's probability distribution. We also noticed that the variance of the MPE statistic is highly dependent on the value of MPE itself, and almost equal to its Cramér-Rao lower bound —in other words, confirming it is an efficient estimator. Despite showing improved results, we realized that the composite versions also modify the MPE estimation due to the measuring of redundant information. In light of our findings, we decided to replace the multiscaling coarse-graining procedure with one of our own, with the intention of improving our estimations.

Since our team observed the MPE statistic to be completely characterized by the model parameters when applied to correlated Gaussian models, we developed a general formulation for expected MPE with low-embedding dimensions. When applied to real sEMG signals, we were able to distinguish between fatigue and non-fatigue states with all methods, especially for high-embedding dimensions. Moreover, we found that our proposed MPE method makes an even clearer difference between the two aforementioned activity states.

Declaration

I, Antonio Dávalos, hereby declare that this thesis is my own original work and that all sources have been accurately reported and acknowledged, and that this document has not been previously, in its entirety or in part submitted at any university in order to obtain academic qualifications. I also hereby declare no conficts of interest.

29/02/2020

Orléans

Antonio Dávalos

Contents

List of Figures

Introduction

Nadie puede dudar de que las cosas recaen
un señor se enferma y de golpe un miércoles recae...

- Julio Cortázar, *Me caigo y me levanto*

The use of information entropy measurements has made it possible to successfully assess the "amount of information" or "complexity" in a system. Among the areas that can benefit from applying these specific measurements, the biomedical field is worth mentioning as an insight-rich discipline in which researchers have access to considerable amounts of information through their examinees. Moreover, biological processes are remarkably complex in nature, usually involving elaborate steps, simultaneous events, and self-correcting feedback loops. To further illustrate the point, bioelectrical activity is difficult to interpret, as it is usually available only as an aggregate of multiple simultaneous elementary signals that, even when understood in an isolated level, grow in complexity in their interactions when observed from outside the body. It is for this very reason that the application of entropy measurements through novel techniques and processes offers a promising solution to this dilemma, as long as we can significantly differentiate between process dynamics by measuring complexity.

Additionally, existing literature strongly associates the reduction of entropy with the advancement of motor diseases. The idea is as follows: the decline in variability suggests a stereotypical motor activity, impeding the adjustments needed to switch tasks or adapt to changing conditions; standing up or walking, for example, are actions that require a constant, fluctuating compensation of gravity, wind, and other external forces to maintain balance. Therefore, a complex and dynamic set of instructions must be present to accomplish any of these tasks, and an individual with a related impairment will either struggle or fail to perform them due to a lack of proper compensation. This is explained as a reduction of the complexity of the instructions each muscle receives.

We have decided to approach this situation by applying multiscale permutation entropy (MPE). This particular measurement has a unique set of features that we deem suitable for the characterization of this problem. First of all, permutation entropy works with ordinal information, a quality that ensures that the method is robust with respect to the amplitude of the source signals and the probability distribution contained within. Given the wide range of possible signal amplitude outputs from different subjects —a desirable property— and the fact that MPE requires a minimum set of parameters to operate, it becomes easy to obtain meaningful results out of the system due to its simplified requirements. Furthermore,

the MPE computation is simple and straightforward to perform, with a clear range of maximum and minimum values. The multiscaling preprocessing also allows the researcher to explore signals at different time scales, thus allowing the identification of hidden long-range information that is not necessarily available from direct raw signal acquisition.

Although MPE measurements and its variants have been successfully applied before, the discussion of this method in literature is usually treated from an algorithmic point of view. Since entropy measurements —especially MPE— are defined as a function of the source's probability distribution, it makes sense to complement these techniques by including a statistical point of view and approach; for instance, we would only obtain an approximation of the real distribution of the phenomenon as a consequence of the limited information present on signal sources.

Previous work exists on the characterization on this problem [1], with special emphasis on the embedding dimension and signal length, yet the interaction of the statistic with the coarse-graining procedure has not been previously addressed in literature. It is also worth mentioning that some statistical properties —particularly bias— have been reported in literature before, but they are usually only mentioned as an afterthought. Also, there is little knowledge on the MPE variance, and this is a measurement that does not necessarily conform to the widely applied assumption of normality. Considering the aforementioned reasons, it is possible to misinterpret observed MPE behavior as an emerging property of the phenomenon in question, when the statistic itself can be the source of these unwanted effects.

Our first objective in the present work is to further develop the theory behind MPE measurement, as well as to understand its behavior over known signal models and stochastic processes. With this knowledge, we would be able to state and develop our second goal: to propose a new MPE technique as an improvement over other well-established methods, including the original computation and its refinements. Lastly, we will apply these methods to real electromyographic (EMG) signals, with the intention of obtaining an updated interpretation of the MPE measurement over real biological systems, and to improve the precision of existing ordinal techniques for the purpose of classification.

In the interest of clarity, as we present our findings throughout this research project, we will also discuss results previously published by our team and expand on said findings: a study pertaining to the statistical properties of MPE measurement in [2] and [3], and findings concerning the interaction of MPE with autoregressive and moving average processes [4].

In Chapter 1 we begin by introducing readers to the overall landscape of available entropy techniques, classifying them in a way that readers can differentiate between the nature of the possible options available to researchers. In Chapter 2 we focus on the statistical properties of multiscale permutation entropy, providing a closed form for the approximation of the MPE statistic —taking in consideration its distribution and the effect of the signal's length— and general expressions for its first two moments. In Chapter 3 we apply the MPE estimation over widely common signal models and stochastic processes to further evaluate entropy properties under known conditions, resulting in a theoretical MPE expected value for Gaussian stochastic processes. In Chapter 4 we explore the properties of MPE refinements in order

to obtain a deeper understanding of the reasons behind their improved precision, subsequently producing an MPE method to further reduce the entropy estimation uncertainty. Finally, in Chapter 5 we tackle the complex problem of real surface electromyographic (sEMG) signals by applying the methods discussed so far in a variety of experimental setups, and provide an updated interpretation of entropy measurements in this context, both from a statistical and biological perspective.

Chapter 1

Information Entropy - Concepts and Definitions

Teóricamente a nada o a nadie se le ocurriría recaer
pero lo mismo está sujeto
sobre todo porque recae sin conciencia
recae como si nunca antes.

- Julio Cortázar, *Me caigo y me levanto*

1.1 Introduction

With his seminal paper [5], C.E. Shannon established the basis of information theory. The author defined the concept of information measurements, including information entropy.

In layman's terms, entropy is a measure of unpredictability [5]. Given a string of symbols, a high entropy level implies that any given symbol cannot be easily predicted by looking at its preceding symbols in the string; conversely, a low entropy value implies that each symbol can be deduced from its history. A completely unpredictable string of symbols will yield the maximum entropy possible by the system, whereas the minimum entropy will occur when only one symbol repeats itself without fail; in a more general sense, however, it is usual that there is a certain probability for each symbol to appear in any given string. Hence, it can be inferred intuitively that this measurement requires some knowledge of the probability distribution and, consequently, that the entropy must be a function of the probability function associated with the string.

Although entropy measures a precise property of a random set based on raw data, a proper interpretation depends heavily on context, since there is a lack of a universal interpretation of said measurements. While entropy is defined in the context of thermodynamics —the discipline that entropy took its inspiration from— as a measure of the irreversibility of any thermodynamic process (by the definition of Rudolph Clausius [6]), entropy is referred to as content, diversity, or complexity in the context of information. Nonetheless, the use of information entropy can still be

justified from a pragmatic perspective, as it has been used successfully in a wide variety of applications, such as data compression and applications in both finances and biomedicine [7], [8], [9]. Even when the true nature and meaning of information entropy is up to debate, this measurement has its uses and implementations.

Since its original formulation in 1948, a large number of entropy variants have been proposed [10]. The sheer amount of options to date are daunting: for each small change in the in method, data source, or application, there exists an entropy measurement with a proper name and its own considerations. Not only that, but each variant comes with its own advantages and disadvantages, further establishing that there is not a one-size-fits-all solution as of today.

In the following chapters, we will focus our efforts in the proper use and characterization of multiscale permutation entropy, including its statistical properties, expected results from known time series models, and real-life applications on electromyographic signals. Before we can delve directly in the inner workings of this method, we need first to present a proper landscape for the use of information entropy, including its variants and common applications.

In Figure 1.1, we present a general map of the practical entropy analysis on time series. Any step in this process can have its own set of variants, either addressing different problems or trying to capture different phenomena within the same dataset. We will outline throughout this chapter the most common entropy variants at each level of analysis. We will then explore the methods with an "inside-out" approach, beginning with the variations of the core entropy definition. Afterwards we will follow the possible variants in the formulation of the event set pertaining to the phenomenon. The last step will consist in reviewing the possible preprocessing variants on the source signal —where the raw data is prepared for entropy analysis— and briefly discussing the refinements made and the reasoning behind their selection. Although we provide a general landscape of all the entropy variants, the presented list and accompanying discussion are far from exhaustive.

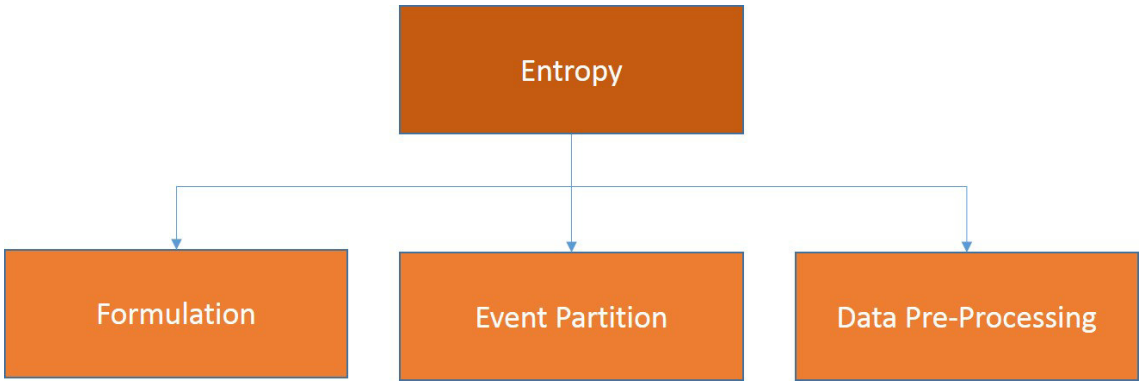


Figure 1.1: Entropy analysis stage components. We can conceptualize the components of any entropy measure in the following three consecutive steps: we must select the proper entropy formulation to use, define the partition that properly describes the system we are to measure, and decide which kind of pre-processing (if any) will be performed on the experimental data

This chapter ends with the discussion of the multiscale permutation Entropy (MPE),

since its general characteristics and properties are particularly suited to our intended application within the analysis of electromyographic signals.

1.2 Entropy Formulations

To properly define and measure the "quantity of information", Shannon outlined the minimum requirements for it to work. If we have a set of n events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$, where $\sum_{i=1}^{n} p_i = 1$, the measure of uncertainty $H(p_1, p_2, \ldots, p_n)$ must have the following properties [5],

1. H should be continuous on p_i.
2. If all the probabilities are the same $\left(p_i = \frac{1}{n}\right)$, then H should be a monotonically increasing function of n.
3. If a choice is broken into two successive choices, the original H should be the weighted sum of the individual values of H associated with each step.(Recursion Property)

In this section we will review the most common entropy formulation from the perspective of information theory. We will first discuss the original Shannon's entropy, as well as some of its alternatives, which also satisfy the properties above.

1.2.1 Classical Shannon's Entropy

After proposing the properties of the measure of uncertainty, Shannon defined the following equation,

$$H = -K \sum_{i=1}^{n} p_i \log p_i, \tag{1.1}$$

where K is any real positive constant. This equation is strikingly similar to the definition of the Boltzmann-Gibbs entropy in a thermodynamic system [6],

$$H = -k_b \sum_{i=1}^{n} p_i \log p_i, \tag{1.2}$$

where k_b is the Boltzmann constant, and each p_i represents the probability of a microstate in the context of statistical mechanics —and later, in quantum physics. Therefore, the name "entropy" for Shannon's equation seems fit for the task. Moreover, by removing the equation from the physical phenomenon, it is possible to apply this measurement to any system with appropriate partitions that obey the fundamental axioms of probability. Instead of particle microstates, we can talk about alphabets, patterns, or symbols, as long as they form an appropriate event set.

Some interesting properties arise from Shannons's entropy definition (1.1),

1. $H = 0$ if and only if $p_i = 1$ for any particular i, and zero otherwise. Any other case yields to a positive value for H.
2. For any given n, H is maximum when all $p_i = \frac{1}{n}$. This is the case of maximum uncertainty, and corresponds with the discrete uniform distribution.
3. For any two joint events r and s, the joint entropy obeys the inequality,

$$H(r, s) \leq H(r) + H(s), \tag{1.3}$$

   when equality is achieved only when the events r and s are independent.
4. Any change of probabilities that makes the probability distribution approach the uniform case increases the value of H.
5. Given two joint events r and s (not necessarily independent) for any particular value of r, there is a conditional probability $p_r(s)$, given by

$$p_r(s) = \frac{p(r, s)}{\sum_s p(r, s)}.$$

   Shannon defined the conditional entropy of s, labeled $H_r(s)$ [11], as the weighted average of the entropy s for each value of r.

$$H_r(s) = -\sum_{r,s} p(r, s) \log p_r(s)$$

   Therefore,

$$H(r, s) = H(r) + H_r(s).$$

Equation (1.1) is the most straightforward equation that meets the properties described above. We must note that this is a deterministic equation of the probability distribution of the phenomenon to describe. This implies that, if we work in a practical, data-driven application, we can only approximate the mass probability function. Therefore, the event probabilities p_i need to be estimated, and thus, H would become a statistic. This will be extensively discussed in Chapter 2.

1.2.2 Tsallis Entropy

If the Shannon's recursivity property is not strictly enforced, the logarithmic function in (1.1) is not the only equation to satisfy the properties outlined for $H(p_1, \ldots, p_n)$. In 1967, Havrda and Charvát proposed a new equation for entropy called Havrda–Charvát structural α-entropy [12]. Later in 1988, Constantino Tsallis proposed a measure now known as Tsallis entropy within the context of thermodynamics. Both formulations are functionally the same.

For a complete discrete set of probabilities $\{p_i\}$, and a real parameter q, the Tsallis entropy is defined as [13]

$$H_q = \frac{K}{q-1}\left(1 - \sum_i p_i^q\right). \tag{1.4}$$

In the case where q approaches the value of 1, the Tsallis entropy converges with Shannon's entropy

$$lim_{q\to 1} H_q = -K \sum_{i=1}^{n} p_i \log p_i = H.$$

By contrast, the main difference in the Tsallis entropy comes from the fact that it is not completely additive. For any two joint independent events r and s, such that

$$P(r,s) = P(r)P(s), \tag{1.5}$$

the Tsallis entropy satisfies the following relationship:

$$H_q(r,s) = H_q(r) + S_q(s) + (1-q)H_q(r)S_q(s).$$

Hence, the parameter q, in this case, produces a deviation from the traditional constraints of entropy additivity, which is particularly well-suited for physical systems with long-range interactions, long-term memory, or fractal properties [14]. Tsallis Entropy is specially useful in the description of particle velocity which present power-law distributions [15] —conditions that are widely present in the field of plasma astrophysics [16]. Additionally, Tsallis entropy has been used in other fields, most notably image processing for segmentation[17] [18], where the images present similar properties as described above [14].

1.2.3 Rényi's Entropy

In a similar fashion, Alfred Rényi proposed a generalized entropy measurement [19]. For a real non-negative value $\alpha \geq 0$ and $\alpha \neq 1$, Rényi's entropy is defined as

$$H_\alpha = \frac{1}{\alpha - 1} \log \left(\sum_{i=1}^{n} p_i^\alpha \right), \tag{1.6}$$

where each p_i, as usual, represents the probabilities of all the possible events in the event set, and n is the number of such events.

Rényi's entropy is heavily modulated by the parameter α, since it modifies the influence of the events in the final entropy measurement. For values of α close to zero, all events tend to have the same weight, regardless of their probability of occurrence. For high values of α (close to infinity) only the most probable events have an influence in the final entropy value.

There are some interesting special cases for the Rényi's Entropy. In the case that $\alpha = 0$, we obtain the max-entropy —or Hartley entropy— as long as the probabilities are not zero,

$$H_{\alpha=0} = \log(n), \tag{1.7}$$

which is the logarithm of the cardinality of the event set (the number of the events in the set). Conversely, when the value of α approaches infinity, we have the min-entropy, defined as [19]

$$lim_{\alpha\to\infty} H_\alpha = min(-\log(p_i)) = -max(\log(p_i)) = -\log(max(p_i)), \tag{1.8}$$

which takes only the most probable event in account.

In the limit case when α approaches 1, we have,

$$lim_{\alpha\to 1} H_\alpha = -\sum_{i=1}^{n} p_i \log p_i, \tag{1.9}$$

which is, once again, Shannon's entropy.

Rényi's entropy generalization is particularly well-suited to analyze phenomena with probability distributions which are notoriously different from a Gaussian behavior [20]. This adds flexibility in the application of spectral estimations, pattern recognition, and source separation [20]. Some applications can be found in biomedical engineering [21], such as the measurement of Gaussianity of heart rate signals.

1.2.4 Entropy Formulations Remarks

Tsallis and Rényi's entropies are some of the most widely used generalizations of the classical Shannon entropy. Other generalized entropies are also available [22], which present further ways to add flexibility by introducing weights to the probability distributions of the phenomena they describe. Once again, the context and applications will define the suitability of each entropy measurement.

1.3 Event Partitions for Entropy

All entropy measurements, as defined here so far, act over a discrete probability function. It is possible to extend these procedures for a continuous probability density function (pdf). In practice, however, we are interested in particular events. Therefore, even continuous distributions are partitioned in such a manner that they reflect the events we are interested in measuring, and the definition of the sample space Ω is utterly important.

Shannon's work [5] portrays the original source as a string of symbols. These can be directly in binary code format —the reason behind the use of $\log_2$— or any other set of symbols, like the Latin alphabet. By having a string of symbols, it is possible to interpret the entropy of the signal as a measure of the quantity of information [5].

It is obvious that for a binary code, $\Omega = \{0, 1\}$, there are only two possible events: a random string of bits will produce either maximum entropy (each bit is completely unpredictable, given the knowledge of previous bits) or an entropy value

of zero if we receive a string consisting of the same value (the knowledge of the first bit is enough to predict the whole string). For the Latin alphabet, we have 26 different events for each letter if we exclude spaces, special symbols, uppercase, accents, or punctuation marks. Although the logic behind the entropy measurement does not change, the sample space is completely different. For example, we obtain a binary string if we translate the Latin alphabet to binary code by using the ASCII convention, but the events will be defined as $\Omega = \{'a', 'b', \ldots, 'z'\} = \{0000000001100001, 0000000001100010, \ldots, 0000000001111010\}$, using 16 bits for encoding. In literature, the event set partition Ω is enough to rename entropy measurements altogether. Since there is a potentially infinite number of sample space definitions, context and application, once again, will suggest the most useful approach.

Even though information entropy seems to be so far constrained to the use of symbols, this limitation can be easily bypassed in the case of a time series composed of measurements with continuous distributions. If we carefully select the partitions for the sample space, a continuous distribution becomes discrete as the events are clearly —and mutually exclusive— defined. This opens the entropy measurement to any time series.

The most common approach for defining the event set in this scenario is to define the sample set Ω by using patterns within the raw signal. By comparing the occurrence of a certain pattern among the rest of the signal, it is possible to have an estimator of the occurrence of that pattern. This, in turn, estimates its probability, which allows entropy analysis to occur. We will briefly discuss some of the most commonly used techniques in the context of signal processing.

1.3.1 Approximate Entropy

S. M. Pincus first developed and proposed [23] approximate entropy (ApEn) specifically for the analysis of medical data. ApEn compares, for a particular cardinal pattern, the similarity with all other patterns of the same length contained in the signal. Typically, ApEn is presented in algorithmical form. The steps for calculating ApEn are as follows:

1. We obtain a time series $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]$, with a uniform sampling rate. N refers to the signal's length; i.e. the number of samples.
2. We fix the parameter m for the size of the pattern, consisting of the number of successive data points used for analysis.
3. We form a sequence of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N-m+1}$, where each vector is defined as $\boldsymbol{x}_i = [x_i, \ldots, x_{i+m-1}] \in \mathbb{R}^m$.
4. For each vector i, we count the number of vectors j whose distance is equal or less than a fixed tolerance $r \in \mathbb{R}^+$,

$$C_i^m = \frac{\#(d[\boldsymbol{x}_i, \boldsymbol{x}_j] \leq r)}{N - m + 1}, \tag{1.10}$$

where the symbol # denotes cardinality, and the distance $d[\boldsymbol{x}_i, \boldsymbol{x}_j]$ is defined

as,

$$d[\boldsymbol{x}_i, \boldsymbol{x}_j] = \max_a |x_i(a) - x_j(a)|; \tag{1.11}$$

that is, the position a within the vector that has the maximum difference among their scalar elements. This distance measure is known as the Chebyschev's Distance [24].

5. We build up the following measurement,

$$\Phi^m(r) = \frac{\sum_{i=1}^{N-m+1} \log(C_i^m)}{N-m+1}, \tag{1.12}$$

which is the average of the logarithm of all calculated C_i^m.

6. Finally, the ApEn measurement is:

$$ApEn = \Phi^m(r) - \Phi^{m+1}(r). \tag{1.13}$$

As we can see, the event partition for ApEn does not include all the possible patterns in the signal, since that would require a large amount of data to make a proper estimation [23]. The method herein uses the present time series, and assumes all possible patterns included within —this is the "approximate" part. Although not an exact solution, this allows researchers to construct an appropriate partition to use as an entropy measurement.

This method is reported to have satisfactory results for a short signal length, and it is fast to compute [23]. Nonetheless, ApEn is still dependent on the signal length and the choice of a proper tolerance parameter, since the value of r is not a trivial choice. In practice, the value r is chosen as a proportion of the signal's standard deviation.

We notice here that ApEn includes vector self-comparisons. Without this feature, the individual C_i^m could be zero, making it impossible to compute ApEn. However, proceeding in such a manner makes ApEn report a higher regularity than it should, which is only a problem when working with short time series.

1.3.2 Sample Entropy

Sample entropy (SampEn) is a refinement over ApEn proposed by Richman and Moorman [25]. In principle, the process is identical, yet with some key modifications:

1. We follow steps 1 to 3 from ApEn.
2. We define C_A and C_B in a way that

$$C_A = \#(d[\boldsymbol{x}_i, \boldsymbol{x}_j] \leq r) \text{ for vector size } m+1$$
$$C_B = \#(d[\boldsymbol{x}_i, \boldsymbol{x}_j] \leq r) \text{ for vector size } m$$

for all $i \neq j$ (that is, without vector self-comparisons).

3. We calculate SampEn as:

$$SampEn = -\log \frac{C_A}{C_B}. \tag{1.14}$$

In the case of SampEn, the vector comparisons are done in a way that no vector is compared with itself, which is indeed done in ApEn. Since count variables A and B have a small chance of being undefined, as they are the sum of all possible vector comparisons, the regular overestimation of ApEn is not present here. Given that the method is similar to ApEn, the event set partition approximation is almost equal. The authors found SampEn results that agree better than ApEn when tested over random numbers with known probability distributions [25]. In the particular case of short time series, SampEn also shows a reduced bias in respect to ApEn [25]. These properties make SampEn a more widely used entropy measurement, specially as the base formulation for more elaborate algorithms [26].

1.3.3 Permutation Entropy

Permutation entropy (PE), proposed by Bandt and Pompe [27], uses the pattern approach with the ordinal information. Instead of relying on approximate distance measurements, PE only takes the order of the data points inside the pattern segments. This has the advantage of offering a full partition, which consists of all the possible order permutations for the segment size at the cost of losing information concerning the signal amplitude. This issue is addressed in [28] by weighting the contributions to the probability distribution based on the amplitude of the patterns.

Since the event set is automatically defined by the pattern size d —the *embedding dimension*— we can estimate the mass probability distribution by counting the number of patterns of each type within the signal,

$$\widehat{p}_i = \frac{\#\{t|\ t < N - d,\ (x_t, \ldots, x_{t+d-1}) \text{ has type } i\}}{N - d + 1}, \tag{1.15}$$

for each of the $i = 1, \ldots, d!$ possible patterns. With this distribution, PE is obtained by calculating the classical Shannon's entropy. The PE algorithm is also fast to compute for a small dimension, and it is invariant to nonlinear monotonous transformations [27].

Nonetheless, by working on a complete event set, PE takes the opposite approach of ApEn, where the set itself is approximated. This implies that he signal length must be large to have a proper estimation, yet it is recommended in practice to work with the restriction $N \gg d!$ to have an adequate probability distribution estimation [29]. Some applications include the stock market [30], where PE was successfully used as a measure of market efficiency; in mechanical engineering, PE has been used on motor bearing fault diagnosis [31], where the method proposed was not only capable of detecting anomalies, but also in discriminating between fault types and fault severity. In [32], where the authors implemented multivariate approach to PE on electroencephalograms from patients with Alzheimer's disease, PE was successful in detecting the "slowing" effect related to the disease, as well as in detecting anomalies in synchronicity between channels. Further examples of PE applications in market analysis and medical research are presented in [33].

1.3.4 Fuzzy Entropy

At this level of analysis, we have so far worked with clearly delimited, mutually exclusive partitions. This is the place where set theory and probability theory intersect (pun intended). A novel approach to analyze the problem of quantifying information is to use fuzzy logic as an alternative to classic probability definitions of event sets and partitions. For the sake of brevity, we will only discuss fuzzy entropy briefly, outlining its basic tenets as a variation to the event set definitions used so far.

In the context of fuzzy logic, the boundaries between partitions are not strongly defined, and the membership to a certain event is replaced by a membership function, usually a ramp or a sigmoid [34]. Any particular event can be a member of different partitions, with a particular weight defined by its membership.

De Luca and Termini [35] defined a set of axioms for fuzzy entropy, which are analogous to the properties outlined by Shannon. Being A and B partitions defined in a fuzzy set, x_i representing any particular event, and both $m_A(x_i)$ and $m_B(x_i)$ being corresponding membership functions of A and B, respectively, then H_F is an entropy measure if it satisfies the following conditions:

1. $H_F(A) = 0$ if and only if A is not Fuzzy (partitions are defined and mutually exclusive).
2. $H_F(A) = 1$ if and only if $m_A(x_i) = 0.5$ for all i (all events are "halfway" members of A).
3. $H_F(A) \leq H_F(B)$ if A is less fuzzy than B; i.e. $m_A(x) \leq m_B(x)$ if $m_B(x) \leq 0.5$, and $m_A(x) \geq m_B(x)$ if $m_B(x) \geq 0.5$.
4. $H_F(A) = H_F(A^c)$ (the entropy of A is equal to the entropy of the complement of A).

Given these conditions, de Luca and Termini defined the fuzzy analog to information entropy as follows [35],

$$H_{F,k}(A) = D_F(A) + D_F(A^c), \tag{1.16}$$

where

$$D_F(A) = -k \sum_{i=1}^{n} m_A(x_i) \log m_A(x_i). \tag{1.17}$$

A more generalized version is proposed by Kosko [36], based on the relative distances between the fuzzy set and its non-fuzzy counterparts,

$$R_p(A) = \frac{l^p(A, \bar{A})}{l^p(A, \underline{A})}, \tag{1.18}$$

where the distance l^p is defined as

$$l^p(A,B) = \left(\sum_i |m_A(x_i) - m_B(x_i)|^p \right)^{1/p}, \tag{1.19}$$

where $\bar{A}$ is the closest non-fuzzy set to A, and $\underline{A}$ is the farthest one.

In this context, entropy refers to the "vagueness" of the membership of any event to subset A. Therefore, the interpretation of fuzzy entropy is not grounded on probability, as Shannon entropy is. Nonetheless, fuzzy entropy has been used as a suitable alternative to SampEn in the analysis of surface electromyographic signal dynamics [37] [38]. In general, fuzzy entropy provides a better relative consistency than methods such as SampEn and ApEn, as well as improving on its statistical stability [38].

1.4 Signal Pre-Processing: Multiscaling

Now we have explored and visited some of the possible variants for entropy analysis at the level of the core equation and the event set partition. These calculations can be done on the raw signal directly, but it is beneficial to explore the information content at different time scales —a particularly desirable approach when we are looking for information contained inside longer trends. Costa et al. [26] introduced the concept of multiscale entropy (MSEn) to take longer correlations and trends within the signal. The process consists on applying a coarse-graining procedure to the original signal, and implementing an entropy measurement afterwards. This process has several variants, as pointed out by Humeau-Heurtier [10], and some of them will be mentioned in this section. The coarse-graining procedure is completely decoupled from the entropy equations, so it can be applied in conjunction with any of the previous techniques as long as the proper event set is defined.

1.4.1 Multiscale Entropy

Given an arbitrary time series $\boldsymbol{x} = [x_1, \ldots, x_N]'$ (where the symbol $'$ denotes transposition), the MSEn [26] consists of two main steps;

1. The coarse-graining procedure. For a particular time scale m, we build a coarse signal $\boldsymbol{x}^{(m)} = [x_1^{(m)}, \ldots, x_{N/m}^{(m)}]'$ by dividing the original signal in nonoverlapping segments of m data points in length. The coarse-graining procedure consists on taking the mean value of the segment,

$$x_j^{(m)} = \frac{1}{m} \sum_{i=m(j-1)+1}^{jm} x_i, \tag{1.20}$$

   where $j = 1, \ldots, m \in \mathbb{N}$.

2. Compute SampEn (1.14) in this new coarse signal.

A schematic representation is provided in Figure 1.2.

With this procedure, it is possible to find long-range measures of regularity hidden within the signal. For the purposes of classification and diagnosis —the original context of [26]— we are not bound to the raw data. When we compare MSEn from different sources, we can choose the scale with the most pronounced differences, and thus, improve the classification process.

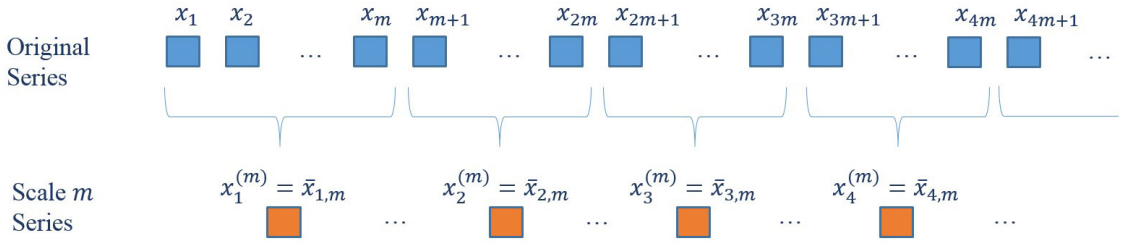


Figure 1.2: The coarse-graining procedure takes the average of all the data points within non-overlapping segments of size m. This diagram is based on the one presented in [39].

There are, of course, some drawbacks to this method, which we will take into account. The coarse-graining procedure can be further described as a process with two steps:

1. A moving average filtering with window size m.
2. A downsampling of the averaged signal by a factor m.

This implies that the coarse-grained signal's length is reduced by a factor of $1/m$. This length reduction yields an increasingly imprecise sampling of the possible events in the set, and thus, compromises the reliability of the entropy measurement.

We should also observe the frequency properties of the coarse-graining procedure, which is equivalent to a finite-impulse low-pass filter that cannot prevent aliasing when the downsampling procedure is applied [40].

It is also necessary to revisit the SampEn tolerance parameter r. As previously mentioned, r is usually chosen as a percentage of the signal's standard deviation [25]. The coarse-graining procedure yields signals with a reduced variance, and therefore can lead to increased coincidences between patterns. Therefore, r must be revisited at each scale to adequately maintain proportion.

1.4.2 Refined Multiscale Entropy

Valencia et al. [40] proposed a series of refinements to address these drawbacks on the original MSEn. The modifications can be summarized as follows:

To avoid the aliasing problem in the frequency domain present by using the coarse-graining procedure as a filter, the authors proposed the use of a Butterworth low-pass

filter instead of a moving average filter,

$$|H(e^{2\pi if})|^2 = \frac{1}{1 + (f/f_c)^{2\eta}}$$

where η is the filter order, f_c is the cutoff frequency, and $i = \sqrt{-1}$. This reduces the aliasing when the filtered time series is subsequently downsampled.

This measurement is known as refined multiscale entropy (rMSE). By testing the method in white Gaussian noise and $1/f$ noise, they found significant differences: rMSE keeps the entropy flat on white Gaussian noise and shows an increase in $1/f$ noise, while MSEn presents a monotonic decrease for both signals along the time scale. These effects are most notorious on short time series, where fast oscillations are dominant.

1.4.3 Composite and Refined Composite Multiscale Entropy

Both the composite multiscale entropy (cMSE) [41] and refined composite multiscale entropy (rcMSE) [42] aim to increase the number of coarse-grained segments obtained from a single raw signal. Given an original signal $\boldsymbol{x} = [x_1, \ldots, x_N]'$ and a time scale m, we can get a different partition from $\boldsymbol{x}$ by starting the partition process at different points; for the original MSEn, it is assumed that the segments starts at x_1. If we set a starting point at $k = 1, 2, \ldots, m$, we can build up to m different coarse-grained signals for that time scale. Thus, the composite coarse-crained procedure can be written as

$$x_{k,j}^{(m)} = \frac{1}{m} \sum_{i=m(j-1)+k}^{jm+(k-1)} x_i \tag{1.21}$$

where $x_{k,j}^{(m)}$ are the elements of the segment vector $\boldsymbol{x}_k^{(m)}$. The cMSE consists of the average SampEn measured for all possible coarse-grained signals $\boldsymbol{x}_k^{(m)}$ for time scale m:

$$cMSE(\boldsymbol{x}, m, r) = \frac{1}{m} \sum_{1}^{m} SampEn(\boldsymbol{x}_k^{(m)}, r). \tag{1.22}$$

This measurement takes advantage of the reduced variance of a mean value to increase the precision of the entropy measurement. Wu *et al.* [41] show the improved precision of cMSE over MSEn on white noise —$1/f$ noise— and real fault-bearing vibration signals.

1.4.4 Refined Composite Multiscale Entropy

The rcMSE [42] uses a similar approach and utilizes the same composite coarse-graining procedure in equation (1.21). Nonetheless, the procedure first makes the

pattern count along all the coarse signals for scale m, and then performs a single SampEn calculation. Hence,

$$rcMSE = -\log \frac{\sum_{k=1}^{m} C_{A,k}}{\sum_{k=1}^{m} C_{B,k}}, \tag{1.23}$$

where

$$C_{A,k} = \#(d[\boldsymbol{x}_k^{(m)}(i), \boldsymbol{x}_k^{(m)}(j)] \leq r) \text{ for vector size } m+1 \tag{1.24}$$

$$C_{B,k} = \#(d[\boldsymbol{x}_k^{(m)}(i), \boldsymbol{x}_k^{(m)}(j)] \leq r) \text{ for vector size } m. \tag{1.25}$$

This calculation further improves the precision of the cMSE by relying on the increased number of sample pair comparisons, instead of the average of SampEn over composite signals. This allows rcMSE to maintain accurate entropy estimations on shorter time series [42].

1.4.5 Modified Multiscale Entropy

In order to reduce the effect of the coarse-graining procedure on the signal length, Wu et al. [43] propose a modification to the original MSEn algorithm named, appropriately, modified multiscale entropy (mMSEn) . The authors applied a moving average filtering without the subsequent downsampling, effectively taking all the possible segments within the signal with overlap.

This approach presents the advantage of increasing the number of pattern comparisons, and thus, reducing the variance of the probability estimation. This is particularly well-suited for short time series, where there are less samples to work with. Nonetheless, the comparison of segments with common data points can lead to unexpected pattern matches, a phenomenon that will be further explored in Chapter 4.

1.4.6 Generalized Multiscale Entropy

In order to explore other properties of the signal, such as the dynamics of the variance, Costa *et al.* [44] propose the use of the generalized MSEn by using higher moments in the coarse-graining procedure, other than the average:

1. The signal is divided into nonoverlapping segments, as it is done with the classical MSEn.
2. Instead of obtaining the mean value of each of the elements within the segment's higher-order moments —such as the variance of each segment—are used. We build the coarse-grained signal from these calculations.
3. We calculate the SampEn for each generalized coarse-grained signal.

This type of coarse-graining leads to the analysis of a completely different aspect of the signal. Generalized MSEn follows the signal's volatility changes when applying

the second moment; for the third moment, we have a measure of the variation in symmetry.

There are multiple versions of multiscale entropy analysis, each built with a different type of filtering, procedure or decomposition —giving place to an endless amount of possibilities. Once again, the proper choice of multiscaling procedure depends heavily on the properties of interest in the time series, as well as the suitability of the properties of each method. The list here is by no means exhaustive.

1.5 A Case for Permutation Entropy

With all the possible options, variations and refinements available for the measurement of information entropy, the question of the appropriate method to use remains open. This is not an easy task, in and by itself. As we have hinted through this chapter, the answer lies in the context, the particular application, and the interactions with the specific properties of the entropy variant.

We must state here some of the peculiarities concerning the analysis of surface electromyographic (sEMG) signals. This analysis consists of the superposition of multiple signal sources from motor neurons, which greatly vary in amplitude and frequency, the latter two being influenced by multiple factors. The result is an sEMG signal with a complex behavior, and particularly prone to respond to artifacts. Therefore, it is imperative to decompose the effects of each of these factors to properly establish their effect on the measurement, especially when the aim is to diagnose a particular anomaly in motor control.

Here, the properties of permutation entropy seem notably appealing, since the analysis of ordinal patterns takes away the variations regarding the amplitude of the signal. This implies that PE analysis is naturally invariant to the force output of the measured sEMG signal. Also, the ordinal analysis is particularly robust in the presence of noise: when noise is sufficiently small in comparison to the signal, these variations do not affect the order of the data points, which lead to the same entropy measurement. While this has its limits, it is convenient to have a method that is inherently resistant to the effects of noise, since this quality remains even if the noise is not white or normally distributed.

We will also take advantage of the multiscale coarse-graining procedure in conjunction with PE, a measurement known as multiscale permutation entropy (MPE) [45]. This variant allows us to explore the long-trend components within the signal, and look for regularity in muscle activity that can otherwise remain obscured by the fine resolution of the sampling frequency used for data acquisition. Since the firing frequency of the motor units has a biological upper limit, the exploration and emphasis on lower frequencies (and thus, longer trends) is justified.

Since one of the main goals of this project is to properly characterize the statistical properties of multiscale permutation entropy, it is necessary to approach this method with the least possible number of modifications. Hence, we are using the classical Shannon's entropy as proposed by Bandt and Pompe for PE. Other entropy formulations, like Tsallis or Rényi, contain an additional level of complexity,

worthy of a vast exploration work on their own. At this moment, even when MPE is widely studied for its biomedical applications, its statistical properties are not completely explored. Therefore, before we take into consideration further entropy equations —or alternative multiscaling techniques and refinements— we need the proper statistical characterization for the original MPE.

1.6 Closing Remarks

Thus far we have reviewed and summarized a wide array of entropy measurements, starting with the original information theory formulation by Shannon [5]. We have explained the variations in entropy formulations, event partition definitions, and signal multiscaling. We covered the general spirit behind these entropy measurement proposals, as well as the specific motivations behind them. Despite not being fully comprehensive, this chapter provides a "big picture" regarding information entropy.

For the purposes of this project, we will focus on multiscale permutation entropy, as explained above, because its particular properties suit the needs and requirements of bioelectrical signal analysis. The particular mathematical and statistical properties will be explored and discussed in Chapters 2 and 4, while the MPE applicability will be further explored in Chapter 5.

Chapter Summary

- Information entropy is a measure of unpredictability within a particular system. It is also interpreted as complexity, diversity or amount of information, depending on the context.
- Entropy analysis is structured as:
  - Equation: the actual computation of the entropy measurement.
  - Event partition: the possible events within the system, as defined by researchers. This implies corresponding probabilities assigned to each event.
  - Data Preprocessing: in most cases, entropy analysis and the event partition are not applied directly to the system data, but a transformation of it. This is specially relevant when working on time series at different scales.
- Shannon's entropy is the classical formulation. A couple of example generalizations include the measurements postulated by Tsallis and Renyi, where different weights are assigned to different events.
- Since entropy works with probabilities, it is equally important to define event partitions. Since we will work within the context of time series, some of the most prominent examples include
  - Approximate/sample Entropy: approximate partitions taken directly from the raw signal, which define the event set by similarity between signal

segments.

  - Permutation entropy: measures the rank patterns within segments in a signal. Since all possible rank permutations are defined by the embedded dimension, the event set is completely characterized.
  - Fuzzy entropy. Based on membership functions with imperfect classifiers, the entropy here measures "fuzziness" of the event set instead of the probabilities of delimited, mutually exclusive events.

- In this context, signal preprocessing deals mainly with the filtering of the original signal before defining the adequate event partitions for entropy analysis.
  - Multiscale entropy applies a Coarse-Graining procedure: a moving average filter with downsampling, such that there is no overlap between data points. This procedure captures the information contained inside long-range trends. Several refinements of multiscaling exist, mainly to address the problem of signal length reduction.
  - Generalized multiscale entropy computes the multiscale information not only from the average filter, but also taking advantage of higher moments.

- For the purposes of this work, we will use Multiscale Permutation Entropy, a procedure which is particularly suitable for sEMG analysis:
  - The MPE procedure is fast to compute and robust over outliers.
  - MPE is invariant to signal amplitude, which eliminates the problem of variability in signal strength proper of subject biological variability. This implies MPE is also invariant to force output.
  - Although there is sufficient development in literature regarding PE, the multiscale variant is not completely understood from a statistical point of view, which limits the interpretation of the results. We will develop the necessary theory in the next chapter.

Chapter 2

Multiscale Permutation Entropy - Theoretical Statistics

Pero el problema, para nosotros los que pensamos nuestra vida,
es confuso y casi infinito...

- Julio Cortázar, *Me caigo y me levanto*

2.1 Introduction

So far, we have explored the diverse entropy analysis options we have available when trying to measure the amount of information in a system. From all these possibilities, we made a case in favor of multiscale permutation entropy (MPE) [45] as an extension from the classical Permutation Entropy (PE) [27]. MPE can be conceptualized as follows:

- Equation: Shannon's entropy.
- Event partition: ordinal patterns.
- Preprocessing: coarse-graining procedure.

In Section 1.5 we briefly stated the reasons for this particular choice in the context of biomedical signals. Working with ordinal patterns produces an entropy measurement which is invariant respect to the signal's amplitude, and particularly to the presence of outliers [27]. Although there is relevant information contained within the amplitude, we are more concerned with the functional shape of the signal —a noteworthy scenario would include when the biological variability between subjects makes comparisons difficult. Additionally, MPE also presents the advantage of having a defined event set prior to the introduction of the data set to analyze, avoiding the necessity of working with approximate event partitions.

The simplicity of MPE analysis also keeps the need of parameter calibration to a bare minimum, since no tolerance value is needed. Also, computationally speaking, the count of ordinal patterns is a fast process, a factor that can potentially lead to real-time applications. The multiscaling process, added to the original PE, allows

researchers to explore the information content in time scales not directly measured with the raw signal, which expands the scope of the original analysis. Nonetheless, the method is not without drawbacks: the ordinal pattern count requires the use of a sufficiently large dataset to be reliable [46], a matter that becomes crucial when we explore the multiscale approach in a situation where the signal length is reduced with increasing scales. Any initial considerations regarding the MPE analysis must take the signal length in account.

It is important to remember that the MPE formulations are typically presented in algorithmic form and the ordinal pattern distribution is measured directly from the source signal, since this implies that any information collected in this manner is only an estimation of the true pattern distribution, and hence, the MPE measurement itself is also an estimation. Therefore, the statistical theory behind MPE is not extensively explored in literature, with some notable exceptions [1]. This justifies the approach of considering MPE as a statistic.

Having said this, in this Chapter we will provide a deeper development of the statistical properties of MPE as well as improving the existing theoretical framework by expanding on the results we first presented in [2] and [3]. We will begin with a formal definition of MPE and the coarse-graining process, which were already briefly explained in Sections 1.3.3 and 1.4.1. We will also comment on the considerations regarding the source signal as a random process, followed by us developing a statistical model of the MPE measurement by means of the Taylor series polynomial expansion. This will allow us to have an approximate expression of the expected value, bias, and variance of MPE. We will also provide the expression for the Cramér Rao lower bound to assess the efficiency of the estimator. Next, we will test our theoretical results against a surrogate model with an easily-modifiable parameter set. Lastly, we will discuss and comment on the preliminary insights we obtained from the development of the MPE statistical theory. This will allow us to better understand the behavior of MPE, which in turn will improve the interpretation of the results obtained when we apply this analysis to real data.

2.2 Multiscale Permutation Entropy Background

2.2.1 Permutation Entropy

For a signal $\boldsymbol{x} = [x_1, x_2, ..., x_N]'$ with N elements, we define ordinal patterns of embedded dimension d as any possible ordinal permutation between adjacent d points of the signal. For example, for $d = 2$, only two possible ordinal patterns exist: $x_t < x_{t+1}$ and $x_t > x_{t+1}$; if dimension $d = 3$, we could obtain pattern $x_t < x_{t+1} < x_{t+2}$, one of the six possible patterns. In general, for embedded dimension d, there are $d!$ possible patterns.

For the aforementioned signal, we will assume no particular structure or statistical properties and establish that said signal must be uniformly sampled as the only restriction. Since this assumption implies that no further information is known, we can only estimate the probability of each pattern by measuring the pattern

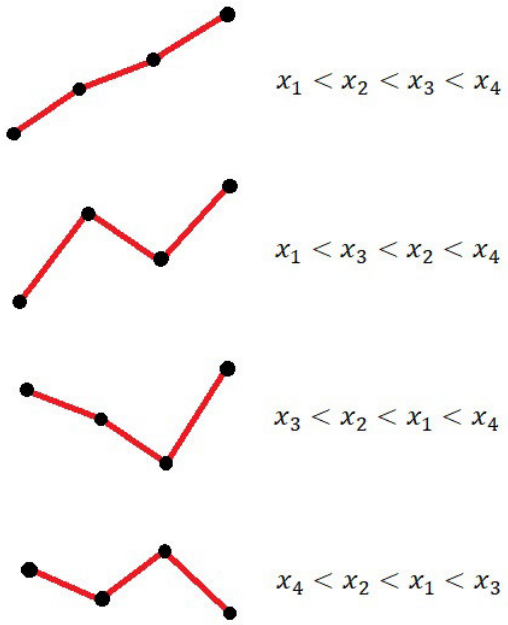


Figure 2.1: Ordinal pattern examples. The figures represent discrete data points from a uniformly sampled signal. There are 24 possible patterns for $d = 4$.

counts inside the signal; in other words, we measure the cardinality of the pattern $i = 1, \ldots, d!$ [27],

$$y_i = \#\{n|\ n < N - (d-1), (x_n, \ldots, x_{n+d-1}) \text{ has pattern } i\}, \tag{2.1}$$

where y_i is the number of patterns of type i in the signal $\boldsymbol{x}$. Some examples of possible ordinal patterns are shown in Figure 2.1. We can now use y_i to build an estimate of the pattern probability $\widehat{p}_i$:

$$\widehat{p}_i = \frac{\#\{n|\ n < N - (d-1),\ (x_t, \ldots, x_{t+d-1}) \text{ has pattern } i\}}{N - d + 1}. \tag{2.2}$$

We use the symbol $\widehat{p}_i$ here to denote that this is an estimation of the pattern probability, as opposed to p_i, which represents the true value. For a given dimension d, the estimate probabilities for all possible patterns i form a mass probability distribution function (pmf). Therefore, it is possible to obtain the permutation entropy measurement using Shannon's definition (1.1),

$$\widehat{H} = -\sum_{i=1}^{d!} \widehat{p}_i \ln \widehat{p}_i, \tag{2.3}$$

which was already presented in Section 1.2.1. Here, the value $\widehat{H}$ is an estimation,

and thus, a statistic. We can also write the normalized version of (2.3) as follows,

$$\widehat{\mathcal{H}} = \frac{\widehat{H}}{\ln(d!)} = \frac{-1}{\ln(d!)} \sum_{i=1}^{d!} \widehat{p}_i \ln \widehat{p}_i \tag{2.4}$$

which guarantees an entropy value between zero and one.

From the algorithmic point of view, PE is easy to implement and fast to compute given a signal $\boldsymbol{x}$ and a dimension d. Since the nature of this procedure is ordinal, PE is invariant to nonlinear monotonous transformations [27], which is in turn a desired property when we expect to work with signals containing different amplitudes, noise, or outliers. On the other hand, this robustness works against us if we intent to extract information from the signal's amplitude.

Another obvious constraint is the length of the signal. For very short signals, the pattern counts in (2.2) are not sufficient to provide a precise estimation of the pattern distribution. There are several proposed guidelines for the minimum length required, with the condition $N \gg d!$ [27] being the most notable one —this, however, offers no practical guidelines to the proper size of N. Other length constrain formulations are $N \geq 5d!$ [47] and $N > (d+1)!$ [48], which were chosen empirically. Later in this chapter we will propose a more specific length criterion, based on theoretical guidelines.

We should note, for the sake of completeness, that the PE definition usually includes a downsampling factor τ [49] [30], since its use is beneficial to avoid oversampling scenarios. For the purposes of this chapter, we will assume that the signal is uniformly sampled and that $\tau = 1$, since the coarse-graining procedure fulfills a similar function in the latter case. However, we will revisit τ in Chapter 4.

2.2.2 Multiscale Coarse-Graining Procedure

Using the original signal $\boldsymbol{x}$, we can construct coarse-grained signals for a fixed time scale m, following a similar route to the one taken by the MSE method explained in Section 1.4.1. We partition the data points in consecutive, nonoverlapping segments of size m, computing the average of each segment afterwards and constructing $\boldsymbol{x}^{(m)} = [x_1^{(m)}, \ldots, x_{N/m}^{(m)}]'$, where each element is,

$$x_j^{(m)} = \frac{1}{m} \sum_{i=m(j-1)+1}^{jm} x_i, \tag{2.5}$$

where $j \in \mathbb{N}$ and $m \in \mathbb{N}$. The MPE measurement consists on calculating the permutation entropy on each $\boldsymbol{x}^{(m)}$ for different time scales.

However, MPE also has the disadvantage of being sensitive to signal length: as N decreases, the estimation of MPE will be less reliable. This effect becomes more pronounced with increasing scales, where the size of the coarse-grained signal decreases by a factor of $1/m$. The general condition $N/m \gg d!$ must be satisfied.

The motivation behind the coarse-graining procedure is to capture long-term information, usually lost in the ordinal comparisons between adjacent data points. The assessment of this information can be useful in detecting trends or recurring patterns that are not usually evident in raw data.

2.3 Multiscale Permutation Entropy Statistics

2.3.1 Previous Considerations

Before exploring the statistical properties of MPE, we will first present the following assumptions regarding the pattern counts:

1. For any particular set of consecutive data points in $\boldsymbol{x}^{(m)}$, the occurrence of any of the possible patterns can be modeled as a random variable in and into itself. This random variable is defined as an indicator function, with a Bernoulli distribution with multiple outcomes

$$I_i(x_t^{(m)}) = \begin{cases} 1, & [x_t^{(m)}, \ldots, x_{t+d-1}^{(m)}] \text{ has type } i \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

2. All the elements $I_i(x_t^{(m)})$ are independent, and identically distributed (iid).

These conditions guarantee that the sum of all the indicator functions leads to a pattern count with a binomial random variable. While the first assumption holds true —by definition— for any signal, the second one requires the signal to be completely uncorrelated, which is not true in the general case. Therefore, when the indicator functions are independent, but not identically distributed, the most appropriate model is a Poisson binomial distribution [50]; if they are not independent, the distribution has no closed form, and approximations are needed [51]. As a first approach to the statistical problem, we will assume that the pattern counts satisfy both assumptions. This approach is justified, to an extent, by the fact that the practical application of the MPE algorithm does not take in account the evolution of ordinal patterns in time. Since the pattern probability distribution is built around the pattern counts of the raw signal as an aggregate, the pattern probability is not computed as a function of time.

At this point, it is useful to define the appropriate values for the time scale m. First and foremost, m is a positive integer, so it cannot take the value of zero or any negative number. Secondly, the maximum theoretical value is $m = N$, the length of the original time scale itself, where the resulting coarse-grained signal consists of a single data point —a nonfeasible trait for MPE analysis. If we define m/N as a normalized scale, then $0 < m/N \leq 1$; consequently, we will consider m/N to be very close to zero for practical reasons, as this will help explore the MPE statistic in a standardized manner.

Before proceeding any further, we will also introduce a vectorial form of Shannon's entropy. Since the mathematical expressions that follow require a considerable use

of linear algebra, it is convenient to rewrite equation (2.3) as,

$$H = -\boldsymbol{l}\,'\boldsymbol{p}, \tag{2.7}$$

where,

$$\boldsymbol{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_{d!} \end{bmatrix}, \qquad \boldsymbol{l} = \begin{bmatrix} \ln p_1 \\ \vdots \\ \ln p_{d!} \end{bmatrix}. \tag{2.8}$$

As long as $\boldsymbol{l}'$ ($\boldsymbol{l}$ transposed) is an horizontal vector and $\boldsymbol{p}$ is a vertical vector, the scalar product in equation (2.7) is identical to the sum expressed in (2.3). Similarly we will take advantage of the scalar product and matrix quadratic forms whenever possible to simplify the long summations that could naturally arise.

2.3.2 MPE Taylor Series Approximation

It is necessary for us to build an explicit model before properly exploring the statistical properties of H. Even though building the distribution function of H can be a daunting challenge, it is not strictly necessary to know this function to extract practical information. We have particular interest in knowing the expected value and variance of H, and as we explain below, this can be accomplished by using equation (2.7) along with a pattern count with a binomial distribution.

In this section we will explicitly formulate a statistical model to estimate H by means of Taylor series expansions. For any coarse grained signal $\boldsymbol{x}^{(m)}$ of length $n_m = N/m - d + 1$ at time scale $m \in \mathbb{N}^+$ and dimension $d \in \mathbb{N}^+$, we can define the random vector $\boldsymbol{Y}$ of size $d!$ as the pattern count vector, and $\widehat{\boldsymbol{p}}$ as the pattern pmf in vectorial form:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{d!} \end{bmatrix} = \begin{bmatrix} n_m p_1 + \Delta Y_1 \\ \vdots \\ n_m p_{d!} + \Delta Y_{d!} \end{bmatrix} = n_m \boldsymbol{p} + \Delta \boldsymbol{Y}, \qquad \boldsymbol{Y} \sim Mu(n_m, \boldsymbol{p}) \tag{2.9}$$

$$\widehat{\boldsymbol{p}} = \tfrac{1}{n_m}\boldsymbol{Y} = \boldsymbol{p} + \tfrac{1}{n_m}\Delta \boldsymbol{Y}. \tag{2.10}$$

When we measure the random variables $Y_1, \ldots, Y_{d!}$, we will obtain the individual pattern counts $y_1, \ldots, y_{d!}$ from equation (2.1).

The random variable $\widehat{\boldsymbol{p}}$ works as an estimator of $\boldsymbol{p}$ —the true pattern pmf vector. $\Delta\boldsymbol{Y}$ is the random part of (2.9), which is a multinomial random variable. $\Delta\boldsymbol{Y}$ has zero mean and the same probability distribution as $\boldsymbol{Y}$.

The next step is to consider the following modifications for equation (2.3): first, the size of the coarse-grained signal will approximately be N/m instead of N. The

true multinomial parameter is $n_m = N/m - d + 1$, but since $N/m \gg d$, we can use the approximation $n_m \approx N/m$; additionally, we should also note that the ordinal pattern count $\boldsymbol{Y}$ can depend on m. For fractional Gaussian noise [2], the pattern probabilities remain constant (See Section 3.3.2). This will, in general, not be the case for other signals. For the purposes of this section, we will assume that the pattern probabilities remain constant along m, and analyze the interaction between these parameters at a later point.

Using the vectorial form of Shannon's entropy (2.7) and the vectorial pmf $\widehat{\boldsymbol{p}}$ in (2.10), we can write the MPE estimator as

$$\hat{H} = H(\widehat{\boldsymbol{p}}) = -\sum_{i=1}^{d!} \widehat{p}_i \ln(\widehat{p}_i) = -\widehat{\boldsymbol{l}}'\widehat{\boldsymbol{p}}, \tag{2.11}$$

which is the vectorial form for the MPE estimator in (2.3). Although H is a function of the vector $\widehat{\boldsymbol{p}}$, the result is a scalar value. This allows us to use the multivariable version of the Taylor series. In this case, we will use the following quadratic approximation around the point $\boldsymbol{p}$ [52],

$$H(\widehat{\boldsymbol{p}}) \approx H(\boldsymbol{p}) + \nabla H(\boldsymbol{p})'(\widehat{\boldsymbol{p}} - \boldsymbol{p}) + \tfrac{1}{2!}(\widehat{\boldsymbol{p}} - \boldsymbol{p})'\nabla^2 H(\boldsymbol{p})(\widehat{\boldsymbol{p}} - \boldsymbol{p}), \tag{2.12}$$

where $\nabla H(\boldsymbol{p})$ is the gradient of $H(\widehat{\boldsymbol{p}})$ at point $\boldsymbol{p}$, and $\nabla^2 H(\boldsymbol{p})$ is the Hessian matrix.

Next, using equation (2.11), we will obtain the explicit expression for the gradient and the hessian of $H(\widehat{\boldsymbol{p}})$, as follows,

$$\nabla H(\boldsymbol{p})|_{\boldsymbol{p}} = \begin{bmatrix} \frac{\partial H(\boldsymbol{p})}{\partial p_1} \\ \vdots \\ \frac{\partial H(\boldsymbol{p})}{\partial p_{d!}} \end{bmatrix}_{\boldsymbol{p}} = \begin{bmatrix} -(1+\ln p_1) \\ \vdots \\ -(1+\ln p_{d!}) \end{bmatrix} = -(\boldsymbol{1} + \boldsymbol{l}) \tag{2.13}$$

$$\nabla^2 H(\boldsymbol{p})|_{\boldsymbol{p}} = \begin{bmatrix} \frac{\partial^2 H(\boldsymbol{p})}{\partial p_1^2} & \cdots & \frac{\partial^2 H(\boldsymbol{p})}{\partial p_1 \partial p_{d!}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 H(\boldsymbol{p})}{\partial p_{d!} \partial p_1} & \cdots & \frac{\partial^2 H(\boldsymbol{p})}{\partial p_{d!}^2} \end{bmatrix}_{\boldsymbol{p}} = -\begin{bmatrix} p_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_{d!}^{-1} \end{bmatrix} = -diag(\boldsymbol{p}^{\circ -1}), \tag{2.14}$$

where $\boldsymbol{1}$ is a vector of ones, and $\boldsymbol{p}^{\circ -1}$ is the Hadamard power (element-wise) of $\boldsymbol{p}$ [53]. $diag(\boldsymbol{p}^{\circ -1})$ is a diagonal matrix with all the elements of $\boldsymbol{p}^{\circ -1}$. With these expressions, we can write equation (2.12) more explicitly:

$$H(\widehat{\boldsymbol{p}}) \approx H(\boldsymbol{p}) - (\boldsymbol{1} + \boldsymbol{l})'(\widehat{\boldsymbol{p}} - \boldsymbol{p}) - \tfrac{1}{2}(\widehat{\boldsymbol{p}} - \boldsymbol{p})' diag(\boldsymbol{p}^{\circ -1})(\widehat{\boldsymbol{p}} - \boldsymbol{p}). \tag{2.15}$$

We note from (2.9) that $\widehat{\boldsymbol{p}} - \boldsymbol{p} = \frac{1}{n_m}\Delta\boldsymbol{Y}$. We will use this fact to write (2.15) in terms of $\Delta\boldsymbol{Y}$:

$$H(\widehat{\boldsymbol{p}}) \approx H(\boldsymbol{p}) - \tfrac{1}{n_m}(\boldsymbol{1} + \boldsymbol{l})'\Delta\boldsymbol{Y} - \tfrac{1}{2}\left(\tfrac{1}{n_m}\right)^2 \Delta\boldsymbol{Y}' diag(\boldsymbol{p}^{\circ -1})\Delta\boldsymbol{Y}. \tag{2.16}$$

To simplify the calculations of the moments of $H(\widehat{\boldsymbol{p}})$, it is desirable to rewrite the last term in (2.16) using the following rearrangement,

$$\begin{aligned}
\Delta\boldsymbol{Y}' diag(\boldsymbol{p}^{\circ -1})\Delta\boldsymbol{Y} &= \Delta\boldsymbol{Y}' \begin{bmatrix} p_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_{d!}^{-1} \end{bmatrix} \Delta\boldsymbol{Y} \\
&= \begin{bmatrix} p_1^{-1}\Delta Y_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_{d!}^{-1}\Delta Y_{d!} \end{bmatrix} \Delta\boldsymbol{Y} \\
&= \left(\boldsymbol{p}^{\circ -1}\right)' \begin{bmatrix} \Delta Y_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Delta Y_{d!} \end{bmatrix} \Delta\boldsymbol{Y} \\
&= \left(\boldsymbol{p}^{\circ -1}\right)' \Delta\boldsymbol{Y}^{\circ 2}, \qquad (2.17)
\end{aligned}$$

where $\Delta\boldsymbol{Y}^{\circ 2}$ is the Hadamard square of $\Delta\boldsymbol{Y}$. By replacing (2.17) with (2.16), we finally arrive to our MPE statistic approximation:

$$H(\widehat{\boldsymbol{p}}) \approx H(\boldsymbol{p}) - \frac{1}{n_m}\left(\mathbf{1} + \boldsymbol{l}\right)' \boldsymbol{\Delta Y} - \frac{1}{2}\left(\frac{1}{n_m}\right)^2 \left(\boldsymbol{p}^{\circ -1}\right)' \boldsymbol{\Delta Y}^{\circ 2}. \qquad (2.18)$$

At this point, the only term that is not explicitly shown in (2.18) is m. If we use the approximation $n_m \approx N/m$, we get,

$$\boxed{H(\widehat{\boldsymbol{p}}) \approx H(\boldsymbol{p}) - \frac{m}{N}\left(\mathbf{1} + \boldsymbol{l}\right)' \boldsymbol{\Delta Y} - \frac{1}{2}\left(\frac{m}{N}\right)^2 \left(\boldsymbol{p}^{\circ -1}\right)' \boldsymbol{\Delta Y}^{\circ 2}}, \qquad (2.19)$$

which is a polynomial with respect to m.

This approximation for the MPE statistic has several advantages. The dependence of $\Delta\boldsymbol{Y}$ (the error of the ordinal pattern count), a variable with a binomial distribution, is evident even if the distribution of $H(\widehat{\boldsymbol{p}})$ is not. Moreover, the deterministic and random parts of $H(\widehat{\boldsymbol{p}})$ are clearly shown here. Also, the role of the time scale m is polynomial, which simplifies future calculations of moments. Although $\boldsymbol{p}$ can be a function of m in general, this expression is compact and relatively easy to handle. In practice, we expect $\boldsymbol{p}$ to be different at each time scale m, so we cannot immediately assume that $\boldsymbol{p}$ and m are independent. For the purposes of exploring the properties of equation (2.19), we will assume m is a fixed parameter. The relationship between $\boldsymbol{p}$ and m will be directly addressed on Chapter 3, section 3.3.

2.3.3 MPE Expected Value and Bias

Since we have a polynomial form of the MPE statistic (2.19) it is possible to obtain the first moment directly. From the definition of Y in (2.1), we know that $E[\Delta Y_i] = 0$

and $\text{var}(\Delta Y_i) = E[\Delta Y_i^2] = n_m p_i(1-p_i)$. It follows that

$$E[H(\widehat{\boldsymbol{p}})] \approx H(\boldsymbol{p}) - \frac{1}{2}\left(\frac{m}{N}\right)\left(\boldsymbol{p}^{\circ -1}\right)'(\boldsymbol{p} - \boldsymbol{p}^{\circ 2}). \tag{2.20}$$

We can rewrite the expressions as,

$$\begin{aligned}\left(\boldsymbol{p}^{\circ -1}\right)'\boldsymbol{p} &= \sum_{i=1}^{d!}\frac{p_i}{p_i} = \sum_{i=1}^{d!} 1 = d! \\ \left(\boldsymbol{p}^{\circ -1}\right)'\boldsymbol{p}^{\circ 2} &= \sum_{i=1}^{d!}\frac{p_i^2}{p_i} = \sum_{i=1}^{d!} p_i = 1,\end{aligned} \tag{2.21}$$

so we can express (2.20) as

$$k. \tag{2.22}$$

It is interesting that the expected value of the MPE statistic approximation is biased, but said bias is not dependent on the pattern probabilities. The only major variable is m, and this means that the expected value bias decreases linearly with scale, and that it can be corrected regardless of the pattern distribution. The expression of the bias of the expected value is,

$$B[H(\widehat{\boldsymbol{p}})] \approx -\frac{1}{2}(d!-1)\left(\frac{m}{N}\right). \tag{2.23}$$

The MPE bias is rarely taken into account in the interpretation of MPE of real-life applications. Without the knowledge that MPE is a biased estimator, the gradual decrease in entropy with respect to m can be mistaken for a real effect from the phenomenon. This is our first contribution to the MPE theory, which we presented in [2].

2.3.4 MPE Variance

The calculation of the variance of the MPE estimator is, not surprisingly, more complex to compute than the expected value. If we compute the variance of equation (2.19), we get:

$$\begin{aligned} var\left(H(\widehat{\boldsymbol{p}})\right) &= E[H^2(\widehat{\boldsymbol{p}})] - E^2[H(\widehat{\boldsymbol{p}})] \\ &\approx H(\boldsymbol{p})^2 \; - \; (\tfrac{m}{N})^2(\boldsymbol{p}^{\circ -1})'E\left[\Delta\boldsymbol{Y}^{\circ 2}\right]H(\boldsymbol{p}) \\ &\quad + (\tfrac{m}{N})^2(\mathbf{1}+\boldsymbol{l})'E\left[\Delta\boldsymbol{Y}\Delta\boldsymbol{Y}'\right](\mathbf{1}+\boldsymbol{l}) \\ &\quad + (\tfrac{m}{N})^3(\mathbf{1}+\boldsymbol{l})'E\left[\Delta\boldsymbol{Y}(\Delta\boldsymbol{Y}^{\circ 2})'\right](\boldsymbol{p}^{\circ -1}) \\ &\quad + \tfrac{1}{4}(\tfrac{m}{N})^4(\boldsymbol{p}^{\circ -1})'E\left[\Delta\boldsymbol{Y}^{\circ 2}(\Delta\boldsymbol{Y}^{\circ 2})'\right](\boldsymbol{p}^{\circ -1}) \\ &\quad + (\tfrac{m}{N})(d!-1)H(\boldsymbol{p}) \; - \; \tfrac{1}{4}(\tfrac{m}{N})^2(d!-1)^2 \; - \; H(\boldsymbol{p})^2. \end{aligned} \tag{2.24}$$

We will proceed to simplify this equation by cancelling the terms $H(\boldsymbol{p})^2$. We also note that the expression

$$\begin{aligned}\left(\boldsymbol{p}^{\circ -1}\right)' E\left[\Delta \boldsymbol{Y}^{\circ 2}\right] &= (\tfrac{N}{m})(\boldsymbol{p}^{\circ -1})'(\boldsymbol{p} - \boldsymbol{p}^{\circ 2}) \\ &= (\tfrac{N}{m})(d! - 1)\end{aligned} \tag{2.25}$$

effectively cancels $(\frac{m}{N})(d! - 1)H(\boldsymbol{p})$. We can now rewrite equation (2.24) as

$$\begin{aligned}var\left(H(\widehat{\boldsymbol{p}})\right) \approx\ & (\tfrac{m}{N})^2 \left[(\mathbf{1} + \boldsymbol{l})' E\left[\Delta \boldsymbol{Y} \Delta \boldsymbol{Y}'\right](\mathbf{1} + \boldsymbol{l}) \ -\ \tfrac{1}{4}(d! - 1)^2\right] \\ & + (\tfrac{m}{N})^3 (\mathbf{1} + \boldsymbol{l})' E\left[\Delta \boldsymbol{Y} (\Delta \boldsymbol{Y}^{\circ 2})'\right](\boldsymbol{p}^{\circ -1}) \\ & + \tfrac{1}{4}(\tfrac{m}{N})^4 (\boldsymbol{p}^{\circ -1})' E\left[\Delta \boldsymbol{Y}^{\circ 2} (\Delta \boldsymbol{Y}^{\circ 2})'\right](\boldsymbol{p}^{\circ -1}).\end{aligned} \tag{2.26}$$

We know that $E\left[\Delta \boldsymbol{Y} \Delta \boldsymbol{Y}'\right]$ is the covariance matrix of $\Delta \boldsymbol{Y}$, matrix $E\left[\Delta \boldsymbol{Y} (\Delta \boldsymbol{Y}^{\circ 2})'\right]$ is the coskewness matrix, and $E\left[\Delta \boldsymbol{Y}^{\circ 2} (\Delta \boldsymbol{Y}^{\circ 2})'\right]$ is the cokurtosis. If we obtain these matrices explicitly, we obtain:

$$E\left[\Delta \boldsymbol{Y} \Delta \boldsymbol{Y}'\right] = \tfrac{N}{m}(diag(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}') = \tfrac{N}{m}\boldsymbol{\Sigma}_p \tag{2.27}$$

$$\begin{aligned}E\left[\Delta \boldsymbol{Y} (\Delta \boldsymbol{Y}^{\circ 2})'\right] &= 2\tfrac{N}{m}\left(\boldsymbol{p}^{\circ 2}\boldsymbol{p}' - diag(\boldsymbol{p}^{\circ 2})\right) + \tfrac{N}{m}(diag(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}') \\ &= \tfrac{N}{m}\boldsymbol{\Sigma}_p\left(\boldsymbol{I} - 2diag(\boldsymbol{p})\right)\end{aligned} \tag{2.28}$$

$$\begin{aligned}E\left[\Delta \boldsymbol{Y}^{\circ 2} (\Delta \boldsymbol{Y}^{\circ 2})'\right] =\ & 3\tfrac{N}{m}(\tfrac{N}{m} - 2)\boldsymbol{p}^{\circ 2}(\boldsymbol{p}^{\circ 2})' \ -\ \tfrac{N}{m}(\tfrac{N}{m} - 2)(\boldsymbol{p}^{\circ 2}\boldsymbol{p}' + \boldsymbol{p}(\boldsymbol{p}^{\circ 2})') \\ & + (\tfrac{N}{m})^2 \boldsymbol{p}\boldsymbol{p}' \ -\ 4\tfrac{N}{m}(\tfrac{N}{m} - 2)diag(\boldsymbol{p}^{\circ 3}) \ +\ 2\tfrac{N}{m}(\tfrac{N}{m} - 3)diag(\boldsymbol{p}^{\circ 2}) \\ & + \tfrac{N}{m}(diag(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}').\end{aligned} \tag{2.29}$$

(For the calculation of covariance, coskewness and cukortisis, see Appendix A).

After taking out the term $H(\boldsymbol{p})$ from equation (2.24), we substitute the expressions for the covariance, coskewness, cokurtosis, and the expected value of $\Delta \boldsymbol{Y}^{\circ 2}$ (equations (2.27), (2.28), (2.29), and (2.25), respectively). After some simplifications, we get:

$$\begin{aligned}var\left(H(\widehat{\boldsymbol{p}})\right) \ \approx\ & (\tfrac{m}{N})\boldsymbol{l}'\boldsymbol{\Sigma}_p\boldsymbol{l} \ +\ (\tfrac{m}{N})^2\left(\mathbf{1}'\boldsymbol{l} + d!H(\boldsymbol{p}) + \tfrac{1}{2}(d! - 1)\right) \\ & +\ \tfrac{1}{4}(\tfrac{m}{N})^3\left(\mathbf{1}'\boldsymbol{p}^{\circ -1} - (d!^2 + 2d! - 2)\right).\end{aligned} \tag{2.30}$$

This expression is a cubic polynomial equation with respect to the normalized time scale m/N. Recalling the domain limitations in Section 2.3.1, m/N will tend to have values very close to zero, implying that the high degree terms have a tendency to vanish, regardless of the values of $\boldsymbol{p}$. Furthermore, since the original Taylor series approximation is quadratic, the cubic term of $var\left(H(\widehat{\boldsymbol{p}})\right)$ (and higher order elements) would be incomplete. For these reasons, it is justifiable to further simplify equation (2.30) to at least a quadratic function in respect to m.

$$var\left(H(\widehat{\boldsymbol{p}})\right) \;\approx\; (\tfrac{m}{N})\boldsymbol{l}'\boldsymbol{\Sigma}_p\boldsymbol{l} \;+\; (\tfrac{m}{N})^2\left(\mathbf{1}'\boldsymbol{l} + d!H(\boldsymbol{p}) + \tfrac{1}{2}(d!-1)\right). \tag{2.31}$$

Written in scalar form, (2.31) read as follows:

$$var\left(H(\widehat{\boldsymbol{p}})\right) \;\approx\; (\tfrac{m}{N})\left(\sum_{i=1}^{d!} p_i \ln^2 p_i - H^2(\boldsymbol{p})\right) \;+\; (\tfrac{m}{N})^2\left(\sum_{i=1}^{d!} \ln p_i + d!H(\boldsymbol{p}) + \tfrac{1}{2}(d!-1)\right). \tag{2.32}$$

This is the result we presented on [3], where not only did we obtain this expression, but we also tested the accuracy of the approximations from (2.32), using surrogate signals with constant pattern distribution for $d = 2$. We will comment on these results later in Section 2.5.

2.3.5 MPE Cramér-Rao Lower Bound

To assess and evaluate our MPE estimator, we need it to compare the variance we obtained from equation (2.31). If we want to evaluate $H(\widehat{\boldsymbol{p}})$ as an estimator of $H(\boldsymbol{p})$, we can obtain the Cramér-Rao lower bound [54] as follows:

$$\text{var}(H(\widehat{\boldsymbol{p}})) \geq \frac{\left[1 - \frac{dB(H(\widehat{\boldsymbol{p}}))}{dH(\boldsymbol{p})}\right]^2}{I(H(\widehat{\boldsymbol{p}}))} = CRLB(H(\boldsymbol{p})), \tag{2.33}$$

where $B(H(\hat{\boldsymbol{p}}))$ is the MPE bias from equation (2.23) and $I(H(\widehat{\boldsymbol{p}})$ is the Fisher's information, which is defined as:

$$I(H(\boldsymbol{p})) = -E\left[\frac{\partial^2 \ln(f_{H(\widehat{\boldsymbol{p}})}(H(\boldsymbol{p}); n_m, \boldsymbol{p}))}{\partial H^2(\boldsymbol{p})}\right]. \tag{2.34}$$

(Note that by having a bias that is constant with respect to $\boldsymbol{p}$, its derivative is zero).

We need the distribution function for $H(\widehat{\boldsymbol{p}})$ —a function we do not explicitly know— to get Fisher's information. Moreover, $H(\boldsymbol{p})$ is not a given parameter, but a measure dependent of $\boldsymbol{p}$.

However, there is a way around these limitations. First, although we do not directly know the distribution of H, we are certain that the distribution of its parameter estimator is multinomial. From equation (2.10), we know that $\hat{\boldsymbol{p}} = \frac{\boldsymbol{Y}}{n_m}$ is an unbiased estimator for $\boldsymbol{p}$. We know that the explicit pmf of $\boldsymbol{Y}$ is:

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; n_m, \boldsymbol{p}) = n_m! \prod_{i=1}^{d!} \frac{p_i^{y_i}}{y_i!}. \tag{2.35}$$

We can calculate the Cramér-Rao bound for $\boldsymbol{p}$ by stating the multivariate form of its definition:

$$\boldsymbol{I}(\boldsymbol{p})^{-1} = CRLB(\boldsymbol{p}), \tag{2.36}$$

where the elements of $\boldsymbol{I}(\boldsymbol{p})$ are

$$I_{j,k}(\boldsymbol{p}) = -E\left[\frac{\partial}{\partial p_j}\ln(f_{\boldsymbol{Y}}(\boldsymbol{y}; n_m, \boldsymbol{p}))\frac{\partial}{\partial p_k}\ln(f_{\boldsymbol{Y}}(\boldsymbol{y}; n_m, \boldsymbol{p}))\right]. \tag{2.37}$$

Before going any further, we must note that $\boldsymbol{I}(\boldsymbol{p})$ is a square matrix of size $d!$. It is convenient to remember that vector $\boldsymbol{p}$ is constrained by

$$\sum_{i=1}^{d!} p_i = 1, \tag{2.38}$$

since the parameters p_i of $\boldsymbol{p}$ represent all the probabilities of the event set. Therefore, if we know the values of $p_1, \ldots, p_{d!-1}$, we know the last probability $p_{d!}$ as a consequence is

$$p_{d!} = 1 - \sum_{i=1}^{d!-1} p_i. \tag{2.39}$$

The particular choice of $p_{d!}$ is arbitrary. Since we lose no information, we can take $p_{d!}$ out from our calculations and subsequently recover this value from equation (2.39). This fact allows to define an auxiliary parameter vector $\boldsymbol{p}_*$ as

$$\boldsymbol{p}_* = [p_1, \ldots, p_{d!-1}]'. \tag{2.40}$$

Because $\boldsymbol{p}_*$ does not lose information, then

$$CRLB(\boldsymbol{p}) = CRLB(\boldsymbol{p}_*) = \boldsymbol{I}(\boldsymbol{p_*})^{-1}, \tag{2.41}$$

where $\boldsymbol{I}(\boldsymbol{p_*})$ is a square matrix of size $d! - 1$. Its elements are obtained as per equation (2.37).

To obtain the Cramér-Rao bound of $H(\boldsymbol{p})$, we use the relation between $CRLB(\boldsymbol{p}_*)$ and $CRLB(H(\boldsymbol{p}))$ from [55],

$$CRLB(H(\boldsymbol{p})) = \left(\frac{\partial H(\boldsymbol{p})}{\partial \boldsymbol{p}_*}\right)' CRLB(\boldsymbol{p}_*)\left(\frac{\partial H(\boldsymbol{p})}{\partial \boldsymbol{p}_*}\right), \tag{2.42}$$

where

$$\frac{\partial H(\boldsymbol{p})}{\partial \boldsymbol{p}_*} = \left[\frac{\partial H(\boldsymbol{p})}{\partial p_1}, \ldots, \frac{\partial H(\boldsymbol{p})}{\partial p_{d!-1}}\right]'. \tag{2.43}$$

If we take advantage of equation (2.39), the values of the elements in equation (2.43) are

$$\begin{aligned}
H(\boldsymbol{p}) &= -\sum_{i=1}^{d!} p_i \ln p_i = -\sum_{i=1}^{d!-1} p_i \ln p_i - p_{d!} \ln p_{d!} \\
\frac{\partial H}{\partial p_j} &= -1 - \ln p_j - \ln p_{d!} \frac{\partial p_{d!}}{\partial p_j} - p_{d!} \frac{\partial \ln p_{d!}}{\partial p_j} \\
&= -1 - \ln p_j + \ln p_{d!} + p_{d!} \frac{1}{p_{d!}} \\
&= \ln p_{d!} - \ln p_j \\
\frac{\partial H}{\partial \boldsymbol{p_*}} &= \ln p_{d!}\, \mathbf{1} - \boldsymbol{l}_*,
\end{aligned} \tag{2.44}$$

where $\mathbf{1}$ is a column vector of ones with size $d! - 1$, and

$$\boldsymbol{l}_* = [\ln(p_1), \ldots, \ln(p_{d!-1})]'. \tag{2.45}$$

We now need to obtain the explicit expression for $CRLB(\boldsymbol{p}_*)$. To obtain the elements of the Fisher matrix from equation (2.37), we need to obtain the natural logarithm of the multinomial distribution (2.35) and its derivatives:

$$\begin{aligned}
\ln(f_{\boldsymbol{Y}}(\boldsymbol{y}; n_m, \boldsymbol{p})) &= \ln(n_m!) + \sum_{i=1}^{d!-1} y_i \ln(p_i) - \sum_{i=1}^{d!-1} \ln(y_i!) + y_{d!} \ln(p_{d!}) - \ln(y_{d!}!) \\
\frac{\partial \ln(f_{\boldsymbol{Y}})}{\partial p_j} &= y_j p_j^{-1} - y_{d!} p_{d!}^{-1} \\
\frac{\partial^2 \ln(f_{\boldsymbol{Y}})}{\partial p_j^2} &= -y_j p_j^{-2} - y_{d!} p_{d!}^{-2} \\
\frac{\partial^2 \ln(f_{\boldsymbol{Y}})}{\partial p_j \partial p_k} &= -y_{d!} p_{d!}^{-2} \\
-E\left[\frac{\partial^2 \ln(f_{\boldsymbol{Y}})}{\partial p_j^2}\right] &= n_m p_j^{-1} + n_m p_{d!}^{-1} \\
-E\left[\frac{\partial^2 \ln(f_{\boldsymbol{Y}})}{\partial p_j \partial p_k}\right] &= n_m p_{d!}^{-1} \\
\therefore \boldsymbol{I}(\boldsymbol{p_*}) &= n_m \left(diag(\boldsymbol{p_*}^{\circ -1}) + p_{d!}^{-1} \mathbf{1} \cdot \mathbf{1}'\right),
\end{aligned} \tag{2.46}$$

where $\mathbf{1} \cdot \mathbf{1}'$ is a square matrix of ones, and $n_m \approx N/m$. The derivatives here retain the probability $p_{d!}$, since it remains a function of all the other elements of $\boldsymbol{p}_*$, as stated in equation (2.39).

The next step is to find the inverse of $\boldsymbol{I(p_*)}$. Here, it is useful to use the following lemma [56]: if $\boldsymbol{A}$ and $\boldsymbol{A}+\boldsymbol{B}$ are nonsingular matrices, and $\boldsymbol{B}$ has rank 1, then

$$(\boldsymbol{A}+\boldsymbol{B})^{-1} = A^{-1} - \frac{1}{1+tr(\boldsymbol{B}\boldsymbol{A}^{-1})}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}. \tag{2.47}$$

From equation (2.46), we have expression $\boldsymbol{p}_*$ as the sum of two matrices. The diagonal of $\boldsymbol{p}_*$ is nonsingular, by definition, and $\boldsymbol{I(p_*)}$ is nonsingular too. Since $\mathbf{1}\cdot\mathbf{1}'$ is rank 1, the lemma (2.47) applies.

Therefore,

$$\begin{aligned} CRLB(\boldsymbol{p}_*) = \boldsymbol{I(p_*)}^{-1} &= \tfrac{1}{n_m}(diag(\boldsymbol{p}_*^{\circ -1}) + p_{d!}^{-1}\mathbf{1}\cdot\mathbf{1}')^{-1} \\ &= \tfrac{1}{n_m}\left(diag(\boldsymbol{p}_*) - \frac{p_{d!}^{-1}}{1+p_{d!}^{-1}(1-p_{d!})}\boldsymbol{p}_*\boldsymbol{p}_*'\right) \\ \boldsymbol{I(p_*)}^{-1} &= \tfrac{1}{n_m}\left(diag(\boldsymbol{p}_*) - \boldsymbol{p}_*\boldsymbol{p}_*'\right). \end{aligned} \tag{2.48}$$

It is worth mentioning that he $CRLB(\boldsymbol{p}_*)$ looks surprisingly similar to the covariance matrix in equation (2.27).

Finally, if we introduce equations (2.44) and (2.48) into equation (2.42), we obtain

$$\begin{aligned} CRLB(H(\boldsymbol{p}_*)) &= \tfrac{1}{n_m}\left(\ln(p_{d!})\,\mathbf{1} - \boldsymbol{l}_*\right)'\left(diag(\boldsymbol{p}_*) - \boldsymbol{p}_*\boldsymbol{p}_*'\right)\left(\ln(p_{d!})\,\mathbf{1} - \boldsymbol{l}_*\right) \\ &= \tfrac{1}{n_m}(\ln^2(p_{d!})\,\mathbf{1}'diag(\boldsymbol{p}_*)\mathbf{1} - \ln^2(p_{d!})\,\mathbf{1}'\boldsymbol{p}_*\boldsymbol{p}_*'\mathbf{1} + \boldsymbol{l}_*'diag(\boldsymbol{p}_*)\boldsymbol{l}_* - \boldsymbol{l}_*'\boldsymbol{p}_*\boldsymbol{p}_*'\boldsymbol{l}_* \\ &\quad - 2\ln(p_{d!})\,\mathbf{1}^T diag(\boldsymbol{p}_*)\boldsymbol{l}_* + 2\ln(p_{d!})\,\mathbf{1}'\boldsymbol{p}_*\boldsymbol{p}_*'\boldsymbol{l}_*). \end{aligned} \tag{2.49}$$

By noting the following relations,

$$
\begin{aligned}
\mathbf{1}' diag(\boldsymbol{p}_*)\mathbf{1} &= \sum_{i=1}^{d!-1} p_i = 1 - p_{d!} \\
\mathbf{1}'\boldsymbol{p}_* &= \sum_{i=1}^{d!-1} p_i = 1 - p_{d!} \\
\mathbf{1}'\boldsymbol{p}_*\boldsymbol{p}_*'\mathbf{1} &= (\mathbf{1}'\boldsymbol{p}_*)^2 = (1 - p_{d!})^2 \\
\boldsymbol{l}_*' diag(\boldsymbol{p}_*)\boldsymbol{l}_* &= \sum_{i=1}^{d!-1} p_i \ln^2 p_i \\
\boldsymbol{l}_*'\boldsymbol{p}_* &= \sum_{i=1}^{D-1} p_i \ln p_i = -H - p_{d!} \ln p_{d!} \\
\boldsymbol{l}_*'\boldsymbol{p}_*\boldsymbol{p}_*'\boldsymbol{l}_* &= (\boldsymbol{l}_*'\boldsymbol{p}_*)^2 = H^2 + 2Hp_{d!} \ln p_{d!} + p_{d!}^2 \ln^2 p_{d!} \\
\mathbf{1}' diag(\boldsymbol{p}_*)\boldsymbol{l}_* &= \boldsymbol{p}_*'\boldsymbol{l}_* = -H - p_{d!} \ln p_{d!} \\
\mathbf{1}'\boldsymbol{p}_*\boldsymbol{p}_*'\boldsymbol{l}_* &= -(1 - p_{d!})(H + p_{d!} \ln p_{d!}) \\
&= -H - p_{d!} \ln p_{d!} + p_{d!} H + p_{d!}^2 \ln p_{d!}.
\end{aligned} \tag{2.50}
$$

we can simplify and rewrite equation (2.49) as

$$
CRLB(H(\boldsymbol{p})) = \tfrac{1}{n_m} \left(\sum_{i=1}^{d!} p_i \ln^2(p_i) - H^2 \right) = \tfrac{m}{N} \boldsymbol{l}' \boldsymbol{\Sigma}_p \boldsymbol{l}. \tag{2.51}
$$

By referring back to (2.31), we note that the $CRLB(H(\boldsymbol{p}))$ is *exactly* the first term in our MPE's model variance (2.32). As long as we stay in the low end of the time scale, we can be sure that the MPE statistic will be approximately efficient (i.e. m/N is close to zero), regardless of the pattern probability distribution. This is one more reason to try to stay in the lower end of the time scales when doing a multiscale analysis. Here, the upper practical constraint of m will only depend on the original signal's data length N.

2.4 Simulations and Results

So far we have found some relevant information regarding the statistical properties of MPE. first, the MPE statistic (2.3) is not unbiased, and this bias is completely independent of the pattern distribution (2.23), since it is only affected by the embedding dimension d. Secondly, we found the MPE variance (2.32) to closely resemble the Cramér Rao lower bound (2.51), suggesting that the MPE estimator, although not unbiased, is close to the minimum variance for the MPE.

In this section, we will test these properties of MPE. In order to simplify the visualization, we will restrict ourselves to the dimension $d = 2$, where the pattern distribution is binomial and only has one parameter: $p = P(x_t < x_{t+1})$. We will

propose and design a surrogate model with an explicit p as a parameter, for the purpose of controlling the pattern distribution, and thus generate the appropriate random signals for testing. We will then compute MPE using (2.3), and compare both the mean results and variance with our corresponding predictions from equations (2.22) and (2.31).

2.4.1 Surrogate Model

To test the results here obtained (2.22), (2.31), we need to design a proper signal model with the following goals in mind: the model must preserve the pattern probabilities across all of the signal, and the function must have the pattern probability as an explicit parameter —which, in turn, is easily modifiable. The following equation,

$$x_t = x_{t-1} + \epsilon_t - \delta(p), \tag{2.52}$$

where t is a discrete time step that satisfies these criteria for dimension $d = 2$. Here, ϵ_t is a Gaussian noise term with variance $\sigma^2 = 1$ without loss of generality [46]. The function $\delta(p)$ represents the trend function, which is built up from the Gaussian cumulative distribution function (cdf) as follows:

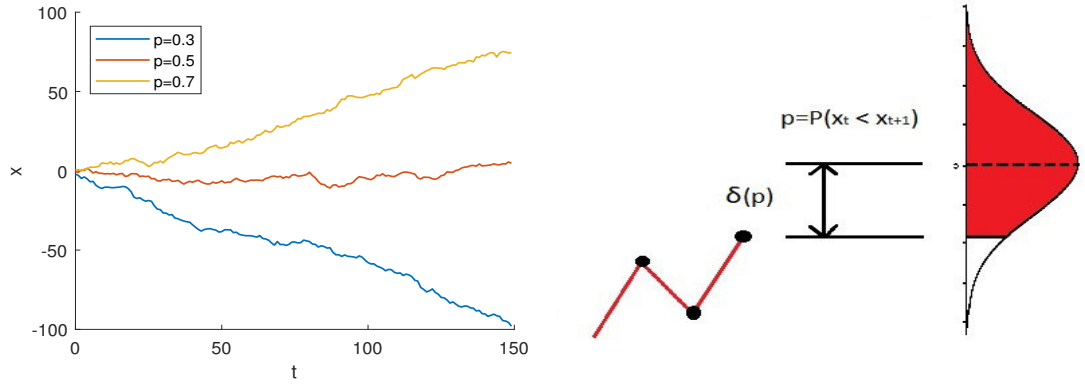
$$p = P(x_t < x_{t+1}) = \tfrac{1}{2}\left(1 - erf\left(\frac{\delta(p)}{\sqrt{2}}\right)\right) \tag{2.53}$$

$$\delta(p) = \sqrt{2}\ erf^{-1}(1 - 2p). \tag{2.54}$$

By using the expression (2.54) in our surrogate model (2.52), we can control the pattern probabilities present in the generated signal just by modifying the value of p; this relationship is illustrated in Figure 2.2. Consequently, we can compare our resulting mean MPE values with the surface generated by equation (2.22) for $d = 2$ (higher dimensions lead to hypersurfaces), as well as comparing the resulting MPE variance from our surrogate model simulations with the surface from equation (2.31). We present both surfaces explicitly on Figure 2.3.

The surrogate model (2.52) was implemented in Matlab, generating 500 signals per each of the 99 different values of $p = 0.01, 0.02, \ldots, 0.99$ we utilized. Additionally, the signal length was also set to $N = 1000$, and we used the time scales $m = 1, \ldots, 50$; for each value of p and m, we obtained the MPE along all the corresponding signals. Finally, we obtained the mean MPE and variance for each case. The results are shown in Figure 2.4.

By directly applying the coarse-graining procedure (2.5) to the surrogate signals in (2.52), the probability p will be modified at each scale. Therefore, instead of applying (2.5) directly, we generated a new set of 500 signals using the original surrogate model (2.52) with length N/m at increasing values of m. This will retain the effect of decreasing signal length, without modifying the parameter p. For the purposes of this test, p and m should be completely independent.



(a) (b)

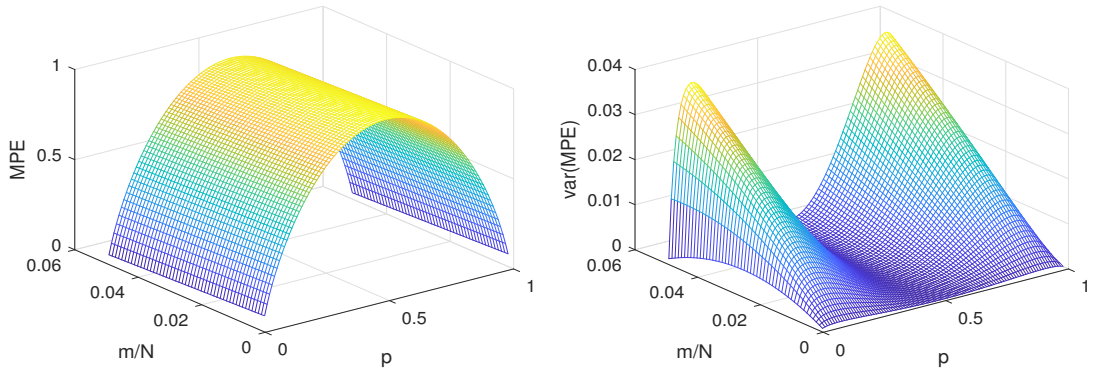
Figure 2.2: Test surrogate model from equation (2.52) for dimension $d = 2$. (a) Model's sample paths for different values of $p = P(x_t < x_{t+1})$. (b) The shift term $\delta(p)$ is modified in accordance with the Gaussian cumulative distribution function, in a way that the variation for the next point in the process has probability p.

2.4.2 Results

Figure 2.4a displays the value of the expected normalized MPE value versus the pattern probability p, with each curve representing a different value of m. As we can see, the general shape of the curve is preserved across time scales, with a small shift downward for each increasing scale. In fact, as we can see from Figure 2.4c, the decrease along m is almost linear, which agrees with the predicted MPE bias (2.23). Moreover, all the lines present the same downward slope, regardless of pattern probability.

Figure 2.4b shows the shape of the MPE variance along p. We can see that the symmetry around $p = 0.5$, which corresponds to the maximum entropy. We note, interestingly, that the curves present clearly defined maxima near the extremes of the distribution, and they preserved along all time scales. We can also note, unsurprisingly, that the variance greatly increases with the time scale m; as shown in Figure 2.4d, where we can see that, for almost all fixed pattern probabilities, the variance increase linearly. This is both present in the theoretical lines and the simulation results, further supporting the MPE variance formulation in (2.31). Lastly, we notice that the variance from simulations is consistently over the predicted values from equations (2.22) with regards to the mean MPE and (2.31) the MPE variance.

We observe a notable exception in the behavior of the MPE variance curve with respect to the time scale. In the case of a uniform pattern distribution and maximum entropy (where all $p_i = \frac{1}{d!}$, $\forall i = 1, \ldots, d!$), the linear term of (2.31) vanishes, leaving only a quadratic curve; for any distribution which deviates from uniformity, the linear term dominates. This effect is clearly shown in Figure 2.5.



(a) (b)

Figure 2.3: Three-dimensional theoretical normalized MPE (2.4) for $d = 2$. (a) Mean MPE value (2.22) in respect to the pattern probability p and normalized time scale m/N. (b) MPE variance (2.31) in respect to p and m/N.

2.5 Discussion

We need to compare our results with previous literature regarding the MPE moments. In the specific case pertaining white noise (i.e. uniform pattern distribution), Little and Kane [1] developed the expected value of classical PE that is subject to finite-length constraints. They found the normalized MPE expected value and variance to be

$$E[\widehat{\mathcal{H}}] \approx 1 - \frac{d! - 1}{2N \ln d!} \tag{2.55}$$

$$\mathrm{var}(\widehat{\mathcal{H}}) \approx \frac{d! - 1}{2N^2 (\ln d!)^2}. \tag{2.56}$$

Our results from equations (2.22) and (2.31) prove to be generalizations of the findings in [1]. First, our MPE statistic (2.19) is valid for any arbitrary pattern distribution, implying that the statistic is robust with respect to the underlying dynamics of any real time series. Secondly, our results extend the finite-length constraints by including the multiscaling component; since the moments are length-dependent, we cannot ignore the decreasing size of coarse-grained signals when applying MPE. Therefore, if we compute the MPE moments (2.22) (2.31) by using $m = 1$ and $p_i = \frac{1}{d!}$, $\forall i = 1, \ldots, d!$, we replicate Little's results [1] (note that the results in equation (2.55) are normalized by $1/ln(d!)$, which is obtained by using the normalized MPE definition of (2.4) instead of the one found in (2.3)).

Now, the MPE variance presents some interesting properties worth discussing. As we can see in Figure 2.4b, the variance is particularly sensitive to the pattern distribution. The points of minimum variance correspond to the maximum MPE at $p = 0.5$, and the minimum MPE at points $p = 0$ and $p = 1$; on the other hand, it is interesting to know the distribution which yields the maximum variance (and thus, uncertainty) of the MPE statistic. Since the first term in (2.31) dominates the overall MPE variance curve, as well as corresponding to the CRLB, we will proceed

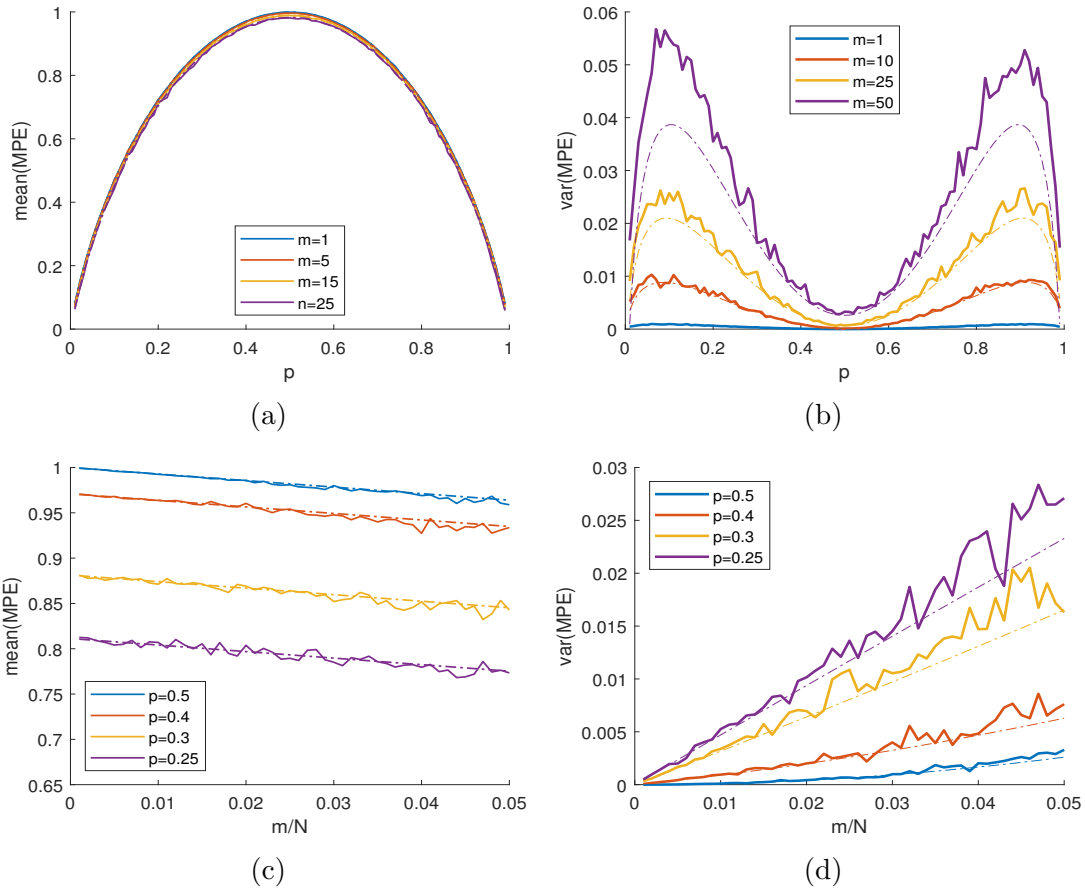
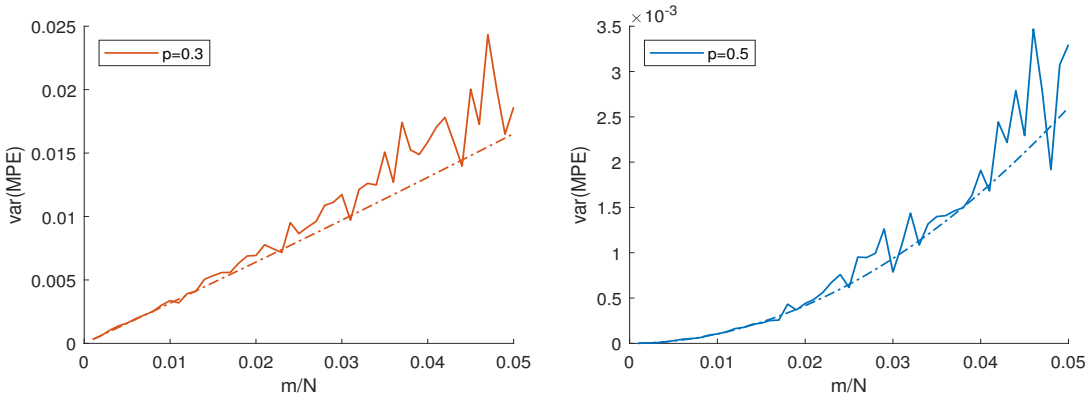


Figure 2.4: Normalized MPE (2.4) for $d = 2$. (a) Mean MPE (2.22) with respect to pattern probability p, which shows a clear maximum at $p = 0.5$ (the point of equiprobable patterns). (b) MPE variance (2.31) in respect to p. We observe minimum points at $p = 0$, $p = 0.5$, and $p = 1$, as well as maximum points at $p = 0.083$ and $p = 0.917$. (c) Mean MPE (2.22) in respect to the normalized time scale m/N. We observe here the linear bias from (2.23), which has the same slope regardless of p. (d) MPE variance (2.31) in respect to m/N. We observe a linear increase, showing that the first element of (2.31) is dominant.



(a) (b)

Figure 2.5: MPE variance (2.31) for d=2 with respect to normalized time scale m/N. (a) Pattern probability $p = 0.3$. We observe an almost linear increase with scale, where the first term of (2.31) is dominant. (b) Pattern probability $p = 0.5$, which corresponds to uniform pattern distribution. Here, the linear term in (2.31) vanishes, leaving only a quadratic increase with scale.

to find the local maxima and minima. For $d = 2$, we can write (2.51) as

$$\mathrm{var}(\hat{H}) \approx \left(\tfrac{m}{N}\right) \boldsymbol{l}^T \boldsymbol{\Sigma}_p \boldsymbol{l}|_{d=2} \;=\; \left(\tfrac{m}{N}\right) p(1-p) \ln^2 \left(\frac{p}{1-p}\right). \tag{2.57}$$

The minimum points become obvious if we look for the zeros of (2.57): if $p = 0.5$, $p = 0$, or $p = 1$, then (2.57) vanishes. The maximum variance, on the other hand, does not correspond to a particularly value of interest in the MPE curve. To find the maximum points, we need to take the derivative of equation (2.57):

$$\left(\tfrac{m}{N}\right) \tfrac{d}{dp} \left(\boldsymbol{l}^T \boldsymbol{\Sigma}_p \boldsymbol{l}|_{d=2}\right) = \left(\tfrac{m}{N}\right) \ln \left(\frac{p}{1-p}\right) \left((1-2p) \ln \left(\frac{p}{1-p}\right) + 2\right) = 0. \tag{2.58}$$

When equation (2.58) is equal to zero, we find the extreme points of the variance MPE curve. It is obvious, once again, that $p = 0.5$ corresponds to a minimum value. The maximum points are found by solving the transcendental function

$$\ln \left(\frac{p}{1-p}\right) = \frac{2}{2p-1}. \tag{2.59}$$

The maximum entropy variance for $d = 2$ is found at points $p = 0.083$ and $p = 0.917$, both equidistant from $p = 0.5$. This implies that we need to be cautious with the MPE measurement we obtain from the signal, as the MPE variance changes nonlinearly with respect to MPE itself.

The second interesting property of the MPE variance is its relationship with the time scale. As we can see from figure 2.4d, the variance increases almost linearly with m. This is true for almost any pattern probability p, even the most unbalanced values. Nonetheless, when MPE is close to its maximum value, the first (linear) term in (2.31) is almost zero, so we need to take into account the second (quadratic) term

of the MPE variance (2.31). This is actually displayed in Figure 2.5, where this particular plot increases in a parabolic curve, both in theory and in simulations.

When solving (2.59) for p, we found that the points $p = 0.083$ and $p = 0.917$ correspond to the maximum variance for any given m/N, and that they both lead to a normalized MPE value of $\mathcal{H} = H/\ln d! = 0.413$. Therefore, the entropy value around this point produces the maximum variance possible for the MPE statistic. We can clearly see these maximum points on Figure 2.4b.

Further elaborating on this result, it is in our interest to know the distribution of $\boldsymbol{p}$ that produces the maximum uncertainty $H(\hat{\boldsymbol{p}})$ for all dimensions d, since it would provide us with information about the worst-case scenario regarding the precision of the MPE statistic. For this purpose, we will define the following maximization problem: we want to maximize the first term of the MPE variance,

$$\boldsymbol{l}^T \boldsymbol{\Sigma}_p \boldsymbol{l} \;=\; \sum_{j=1}^{d!} p_j (\ln p_j)^2 - \left(\sum_{j=1}^{d!} p_j \ln p_j \right)^2, \tag{2.60}$$

subject to restriction

$$\sum_{j=1}^{d!} p_j = 1. \tag{2.61}$$

For this very purpose, we capitalize on the advantages offered by the Lagrangian method of multipliers. First, we define the function as

$$\mathcal{L}(\lambda, \boldsymbol{p}) = \sum_{j=1}^{d!} p_j (\ln p_j)^2 - \left(\sum_{j=1}^{d!} p_j \ln p_j \right)^2 - \lambda \left(\sum_{j=1}^{d!} p_j - 1 \right). \tag{2.62}$$

By taking the partial derivative of $\mathcal{L}$ with respect to $p_i, \forall\ i = 1, \ldots, d!$, we obtain the following equation system:

$$\frac{\partial \mathcal{L}}{\partial p_i} = \ln^2 p_i + 2 \ln p_i - 2(\ln p_i + 1) \left(\sum_{j=1}^{d!} p_j \ln p_j \right) = \lambda, \qquad \forall\ i. \tag{2.63}$$

Since all $\frac{\partial \mathcal{L}}{\partial p_i} = \lambda$, we can take any probability p_i as a reference point, similarly to what we did in Section 2.3.5. We will again use the last variable $p_{d!}$ without loss of generality. By setting $\frac{\partial \mathcal{L}}{\partial p_i} = \frac{\partial \mathcal{L}}{\partial p_{d!}}$, we obtain the following equation:

$$(\ln p_i - \ln p_{d!}) \left[(\ln p_i + \ln p_{d!}) + 2 - 2 \sum_{j=1}^{d!} p_j \ln p_j \right] = 0. \tag{2.64}$$

This expression has two multiplied terms. Even if one or both terms turn to be zero, the equation (2.64) still holds true. If the first term is zero, then we have

$$\begin{aligned} \ln p_i - \ln p_{d!} &= 0 \\ p_i &= p_{d!}, \qquad \forall i. \end{aligned} \tag{2.65}$$

This condition can only be met if all the probabilities are the same. Therefore, this condition corresponds to the uniform distribution of $\boldsymbol{p}$ for any dimension d. We know this distribution produces the maximum MPE, and corresponds to a minimum variance (2.61).

Now, if the second term in (2.64) is equal to zero, this implies that $p_i \neq p_{d!}$, $\forall i$. We can write this second term as

$$\ln p_i + \ln p_{d!} + 2 - 2\sum_{j=1}^{d!} p_j \ln p_j = 0, \qquad \forall\, i \neq d!. \tag{2.66}$$

We can rewrite equation (2.66) as

$$\ln p_i + \ln p_{d!} = -2 + 2\sum_{j=1}^{d!} p_j \ln p_j, \qquad \forall\, i \neq d!. \tag{2.67}$$

We note that the right side of the equation is constant for all i. Once again, this implies that

$$\begin{aligned} \ln p_i + \ln p_{d!} &= \ln p_j + \ln p_{d!} \\ p_i &= p_j, \qquad \forall\, i, j \neq d!. \end{aligned} \tag{2.68}$$

Once again, all probabilities p_i are identical, given $i \neq d!$. By using the constraint in equation (2.61), we see that

$$p_{d!} = 1 - (d! - 1)p_i. \tag{2.69}$$

Also, from equation (2.67), we can see that the summation term is exactly the definition of entropy in (2.3). Therefore,

$$\ln p_i + \ln p_{d!} + 2 + 2H(\boldsymbol{p}) = 0, \qquad \forall\, i \neq d!. \tag{2.70}$$

Given the restrictions found in (2.68) and (2.69), we can obtain the entropy expression for this case:

$$\begin{aligned} H(\boldsymbol{p}) &= -\sum_{j=1}^{d!} p_j \ln p_j \\ &= -\sum_{j=1}^{d!-1} p_j \ln p_j - p_{d!} \ln p_{d!} \\ &= -(d!-1)p_i \ln p_i - p_{d!} \ln p_{d!}. \end{aligned} \tag{2.71}$$

We then substitute (2.71) into (2.70),

$$\begin{aligned} \ln p_i + \ln p_{d!} + 2 - 2(d!-1)p_i \ln p_i - 2p_{d!} \ln p_{d!}) = 0, \\ (1 - 2(d!-1)p_i) \ln p_i + (1 - 2p_{d!}) \ln p_{d!} + 2 = 0. \end{aligned} \tag{2.72}$$

Finally, by introducing once again our expression for $p_{d!}$ (2.69), we obtain the following:

$$\begin{aligned} (1 - 2(d!-1)p_i) \ln p_i + (1 - 2(d!-1)p_i) \ln 1 - (d!-1)p_i + 2 = 0 \\ \ln p_i + \ln 1 - (d!-1)p_i = \frac{-2}{1 - 2(d!-1)p_i} \\ \ln\left(\frac{p_i}{1-(d!-1)p_i}\right) = \frac{2}{2(d!-1)p_i - 1}. \end{aligned} \tag{2.73}$$

This is a transcendental equation for p_i. With this information at hand, we know that the critical points that yield maximum variance for an arbitrary dimension d must have a pattern probability distribution that satisfies (2.59):

$$\begin{aligned} &\ln\left(\frac{p_i}{1-(d!-1)p_i}\right) = \frac{2}{2(d!-1)p_i - 1}, \qquad \forall\, i \neq d! \\ &p_{d!} = 1 - (d!-1)p_i. \end{aligned} \tag{2.74}$$

As long as we have this structure in the pattern distribution, we will have a maximum variance for the system. For $d = 2$, this equation is reduced to equation (2.59).

By computing the results in (2.74) for dimensions $d = 3, \ldots, 7$, we found the particular distributions which maximize the MPE variance (2.31) for each case. These are the worst-case scenarios regarding the precision of our MPE statistic. The results are shown in Table 2.1.

Although the precise value of the normalized MPE (2.4) differs across dimensions, it is clear that there is a region around $\mathcal{H} \approx 0.450$ where our normalized MPE statistic will have maximum variance. However, we will not explore the uniqueness of this result in the present work, since any arbitrary signal that presents a distribution close to (2.74) requires particular consideration.

d	p_i	$p_{d!}$	$\mathcal{H}$
3	0.036	0.082	0.425
4	0.0113	0.7394	0.425
5	0.0027	0.6811	0.438
6	0.0005	0.6405	0.450
7	0.00008	0.61200	0.466

Table 2.1: Probability distributions that yield maximum variance for normalized MPE $\mathcal{H}$, at dimensions $d = 3, \ldots, 7$. The probabilities are subject to the restrictions $p_i = p_j$ for $i \neq j$, and $p_i \neq p_{d!}$.

2.6 MPE Length Criterion

By knowing the moments of the MPE statistic (2.22) (2.31), we are now able to propose some improvements on this technique's implementation on real signals. We will revisit the problem of the length constraint $N/m \gg d!$, which is not sufficiently clear as a criteria for a long signal. We present a reformulation of this constraint as follows: we know that a signal consisting of uncorrelated noise will lead to a uniform pattern probability distribution, and thus, to the maximum possible entropy value for the system. We also know that uncorrelated noise will retain maximum MPE, regardless of scale [2]. Therefore, with these assumptions in mind, the decrease of MPE with scale comes exclusively from the bias in equation (2.23). We then define a maximum deviation tolerance α that measures the percentage of the MPE decline from the maximum possible MPE. We propose a length criterion where the maximum bias is less than the value of α, such that

$$\begin{aligned} |B[\mathcal{H}(\widehat{\boldsymbol{p}})]| &< \alpha \\ \frac{1}{2}\frac{d!-1}{\ln d!}\left(\frac{m}{N}\right) &< \alpha \\ \frac{N}{m} &> \frac{1}{2\alpha}\frac{d!-1}{\ln d!}. \end{aligned} \tag{2.75}$$

Here we use the normalized MPE estimator $\mathcal{H}(\widehat{\boldsymbol{p}})$ (2.4), so that we can interpret α as a percentage —which should be quantitatively small. Although obtaining the value of d for this equation is not trivial, abiding by the suggestion of Bandt and Pompe [27] gives us the advantage of only working with dimensions $d = 3, \ldots, 7$. Therefore, we provide a table comparison for these values in Table 2.2.

Although this is in no way a true improvement on the length constraint itself, it allows researchers to have a more precise gauge over the limits of their study. This criterion takes away the ambiguity of the valid parameter selection.

The surrogate model (2.52) was implemented in Matlab, generating 500 signals per each of the 99 different values of $p = 0.01, 0.02, \ldots, 0.99$ we utilized. Additionally, the signal length was also set to $N = 1000$, and we used the time scales $m = 1, \ldots, 50$; f. For each value of p and m, we obtained the MPE along all the

d	$\frac{d!-1}{\ln d!}$	min $\frac{N}{m}$
3	2.79	333
4	7.23	643
5	24.86	1,667
6	109.28	5,724
7	591.86	25,164

Table 2.2: Critical points for N/m for different values of the embedded dimension d. Minimum signal length N/m at $\alpha = 0.05$

corresponding signals. Finally, we obtained the mean MPE and variance for each case. The results are shown in Figure 2.4.

2.7 Closing Remarks

In this chapter, we have developed the multiscale permutation entropy statistic, presenting its expected value, bias, and variance by means of a Taylor series approximation. We also tested our theoretical results against previous literature [1], as well as with simulations from a suitable surrogate model. In both cases, the results match our predictions, further supporting our initial formulations.

We first found that the MPE expected value is a biased estimator. Moreover, the bias is solely dependent on the parameters of the MPE analysis, particularly dimension, scale, and signal length. This implies that the MPE will present the same bias with respect to time scale, regardless of the pattern probability distribution of the signal.

Secondly, we found the MPE variance to increase almost linearly with increasing time scale for almost any pattern distribution. The exception emerges when the MPE is close to a maximum value (uniform probabilities). In this scenario, the variance increases quadratically in respect to the time scale. Our formulation closely resembles the Cramér-Rao lower bound for the MPE statistic, which means it is almost as efficient. We must also add that the variance presents a maximum value for specific MPE values and pattern probability distributions, as this informs other researchers about an MPE region where we have maximum uncertainty.

Finally, we were able to suggest a more precise criterion for signal length than $N/m \gg d!$—which is usually found in literature. By defining a maximum allowed bias, we were able to specify a minimum length (and maximum time scale) based solely on the pattern dimension for the analysis.

We are aware of the possible refinements available for MPE, which are reported to increase the precision of the measurement [57]. This chapter's main purpose is to develop the theory on classical MPE; a necessary approach to understand the increasingly refined methods that will be discussed in Chapter 4. Furthermore, the relationship between pattern distribution and time scale were left out of the analysis intentionally, since it is necessary to first establish the most simplified model possible. The evolution of the pattern distribution with respect to scale will be discussed in Chapter 3 for well-known stochastic processes.

Chapter Summary

- Permutation entropy (PE) measures the information content within the probability distribution of the ordinal patterns in a signal for a given embedding dimension.
- Multiscale permutation entropy (MPE) performs the PE calculation over a coarse-grained signal that aims to capture the ordinal information content over longer trends.
- The pattern probability distribution is assumed to follow a multinomial distribution. This is in general, not true, since the patterns present along time are generally neither independent nor identically distributed. Nonetheless, the multinomial approach is justified, since PE and MPE estimate the probabilities from pattern counts.
- Since the pattern counts estimate the true pattern probability distribution, the MPE must be regarded as a statistic. We propose a Taylor series expansion on MPE in order to obtain its moments.
- The MPE expected value is a biased statistic, which decreases its value when the time scale increases. Moreover, the bias' linear approximation is independent of the estimated pattern probability distribution, and hence, the same could be stated for any given time series.
- The MPE variance, on the other hand, does not depend on the pattern distribution. The variance tends to be small when the MPE itself is close to its maximum value or really close to zero. Regardless, the variance presents maximum points for specific pattern probability distributions.
- The Taylor approximation to the MPE variance is close to the Cramér-Rao lower bound for the MPE estimator, with the first term being exactly the CRLB. The variance deviates from the CRLB only when the MPE is close to its maximum.
- We tested the accuracy of our MPE model against surrogate signals, which further valides our results, specially in the case of the downward MPE bias and the almost linearly increasing variance with respect to the scale.
- To address the ambiguity of the length constraint of MPE, we proposed a precise criterion based on a maximum bias tolerance. This provides a more explicit rule for parameter selection than the often cited $N/m \gg d!$.
- With the MPE variance, we were also able to specify a particular MPE region where the variance will be maximum, regardless of scale and almost independently of the embedding dimension of choice.

Chapter 3

MPE on Common Signal Models

Un jazmín, para dar un ejemplo perfumado.
A esa blancura, ¿de dónde le viene su penosa
amistad con el amarillo?

- Julio Cortázar, *Me caigo y me levanto*

3.1 Introduction

So far, we have delved inside the workings of multiscale permutation entropy from both the empirical and —most importantly— the statistical point of view. It is through the Taylor series that we have been able to develop an MPE approximation which allows us to compute its expected value and variance. We know by this point that MPE is a biased estimator with a linear value that only depends on the parameters of the analysis and not the pattern distribution. We also found the variance to be close to the Cramér-Rao lower bound, and thus, approximately efficient. We were also able to propose more precise length criterion for a sufficiently long signal for MPE analysis. Finally, we established a normalized MPE range where the variance is maximum for a given signal length.

Although these results hold true for arbitrary signals, we cannot ignore the fact that the coarse-graining procedure has a noticeable effect on the pattern distribution found at each scale. In Chapter 2 we intentionally left this fact out of the analysis, so that we could isolate the effect of the signal length over MPE statistic. Now it is time we address this relationship directly.

We can apply the MPE analysis on any discrete signal without the need of prior knowledge of its underlying dynamics. In fact, the empirical computation of MPE can give us some insight in this regard. On the other hand, if we know the nature of the process, we can compute a theoretical MPE based on the signal's model. We will know if our proposed signal model can be appropriate for explaining all the relevant information of the phenomenon when the theoretical MPE matches the empirical MPE.

Therefore, in this chapter we will study the MPE values for some well-known sig-

nal models, addressing first the expected MPE from common deterministic signals. Since we expect the MPE to be robust to perturbations [27] [45] [58], we will also analyze the effect of noise, with the intention of assessing when the noise dominates over the signal and testing the robustness limits of the methods. Although our exploration of MPE contemplates random processes, Gaussian signals are particularly relevant in the present work; more specifically, we will characterize —as we already did in [2]— both white Gaussian noise (wGn) and fractional Gaussian noise (fGn). Additionally, we will explore —as done by our team in [4]— first order autoregressive (AR) and moving average (MA) models. Since all Gaussian signals present particular symmetries [58], we will propose an explicit, general formulation of the theoretical MPE for this type of random processes, and conclude by testing our proposed technique on more elaborated models, such as the general Autoregressive and Moving average model (ARMA). These results will allow researchers to better gauge the expected MPE results from well-established models. We will provide the appropriate benchmark MPE to compare the information content between real datasets and the models used to describe them.

3.2 MPE on Models with Deterministic Signals

In this section we will briefly discuss the MPE results we should expect from deterministic signals. For known functions of time, we would expect to observe a low value of MPE, since the signal is easily predictable if we know its analytic function. We will also discuss the effects of the sampling rate, since we will work with discrete signals in practice. Finally, we will evaluate the introduction of random noise and gauge its overall effect on our entropy measurements. In this section, without loss of generality, we will limit our analysis to the time scale $m = 1$.

3.2.1 Deterministic Signals

First and foremost, we need to approach the simplest signal models available: deterministic curves. By introducing a signal shape with a known function and absence of randomness, it is almost trivial to obtain the pattern probabilities, regardless of the embedded dimension d used for the analysis. Nevertheless, we can address here some of the most basic concepts regarding the MPE implementation.

A continuous signal can be regarded as a series of points with an infinitesimal distance between them. If we were able to measure such a system, we will only find two possible patterns: either an always increasing (pattern 1) or always decreasing (pattern $d!$). For simplicity, we will call them *monotonic patterns*. These cases correspond with the regions where the slope of the curve is positive and negative, respectively. The only special case here comes from the local maxima and minima of the curve, where the slope is zero —we exclude the case where the curve is horizontal, since patterns with data points of equal value must be properly classified by the researcher. In the case of reaching its limits, all the nonmonotonic pattern probabilities $p_2, \ldots, p_{d!-1}$ are zero almost surely. Thus, we would only measure the monotonic pattern probabilities by measuring the proportion of time where the curve

has the corresponding slope. More formally, given an embedding dimension d and a continuous function $x = g(t)$ within a bounded time t, such that $t_{min} \leq t \leq t_{max}$, we can obtain the probability of the pattern p_1 (increasing) and pattern $p_{d!}$ (decreasing) by measuring the proportion of time where the derivative of $g(t)$ is positive, such that

$$p_1 = \frac{\left(\sum_{k \in K} \left(g^{-1}(x_{max,k}) - g^{-1}(x_{min,k})\right) | \frac{dx}{dt} > 0\right)}{t_{max} - t_{min}} \tag{3.1}$$

$$p_{d!} = 1 - p_1, \tag{3.2}$$

where $k \in K$ is the number of local minima within $t_{max} - t_{min}$, $g^{-1}(x)$ is the inverse function of $x = g(t)$, $x_{min,k}$ and $x_{max,k}$ are the local minima and maxima, and $g^{-1}(x_{min,k}) < g^{-1}(x_{max,k})$ for all k. As we can see, the probability $p_{d!}$ is just the complement of p_1. All other probabilities $p_2, \ldots, p_{d!-1} = 0$.

Therefore, we can write the normalized PE of such a system as

$$lim_{p_2,\ldots,p_{d!-1} \to 0} \mathcal{H} = \frac{-1}{\ln(d!)} \left(p_1 \ln(p_1) + p_{d!} \ln(p_{d!})\right]). \tag{3.3}$$

We use the limit since, strictly speaking, $\ln(0)$ is not defined. Figure 3.1 shows an example of such a calculation with a deterministic cubic polynomial equation.

There is an additional effect we will discuss, given that we are working with discrete signals. In the regions close to the maximum and minimum points of $x = g(t)$, we would expect non-monotonic patterns to appear. Intuitively, we can see that sampled signals from $f(t)$ would not exactly adhere to equation (3.1). For a sufficiently large sampling rate, the nonmonotonic patterns probabilities would tend toward zero, so (3.1) would still be a good approximation. On the contrary, when the pattern size is comparable to the number of data points between local maxima

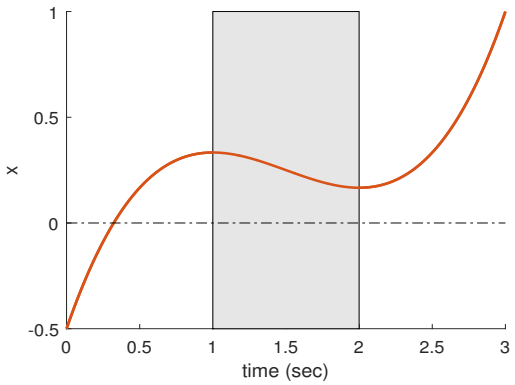


Figure 3.1: Sampled cubic polynomial $x = \frac{1}{3}t^3 - (\frac{2}{3})t^2 + 2t - \frac{1}{2}$ for $t = [0, 3]$ seconds. The regions $t = [0, 1]$ and $t = [2, 3]$ sec have a positive slope; therefore $p_1 = 2/3$ and $p_{d!} = 1/3$. It follows from equation (3.3) that the normalized PE is $\mathcal{H} = 0.3552$ for $d = 3$.

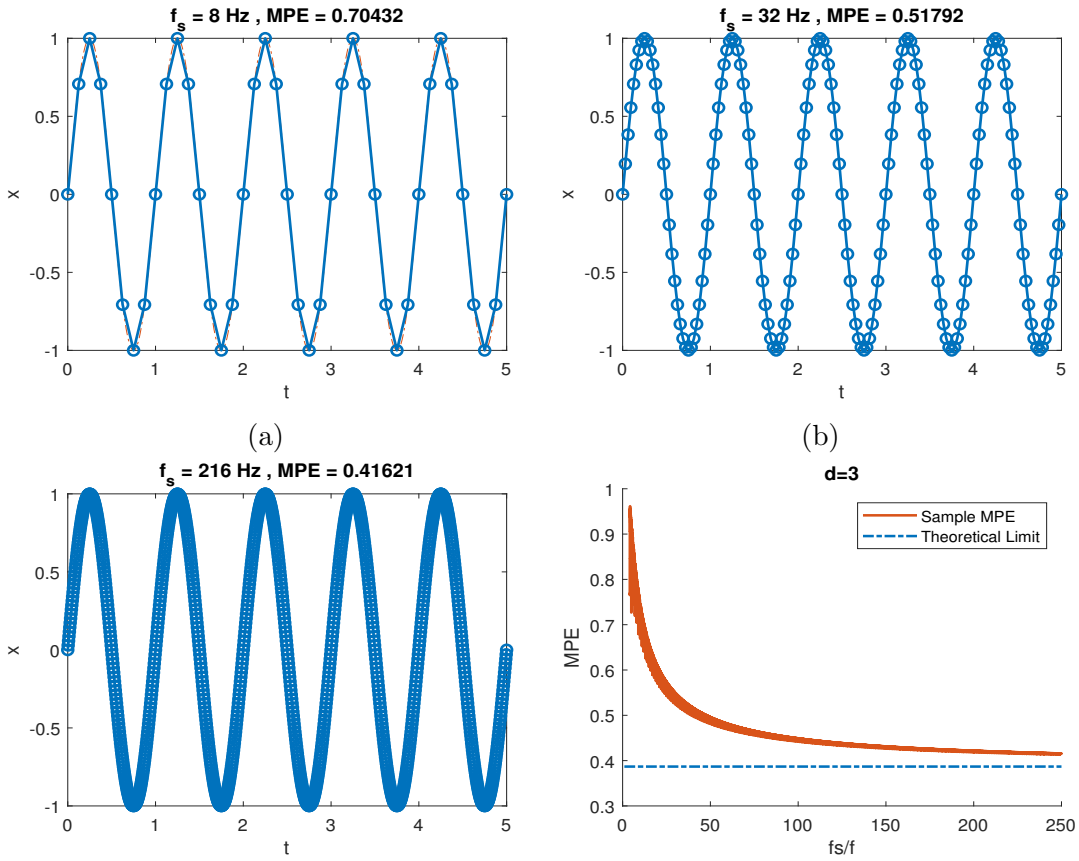


Figure 3.2: Sine wave $x = \sin(2\pi f t)$ with wave frequency $f = 1$, from $0 \leq t \leq 5$ seconds. Here we show the sampled signals with sampling frequency (a) $f_s = 8$ Hz, (b) $f_s = 32$ Hz, and (c) $f_s = 216$ Hz, with their corresponding values of normalized PE at dimension $d = 3$. (d) shows the PE of the sine wave x at different sampling frequencies f_s. The measured PE converges with the theoretical normalized PE (3.3) for the continuous sine wave ($\mathcal{H} = 0.387$).

and minima, the nonmonotonic pattern probabilities would be comparable to the monotonic ones. In figure 3.2, we can observe the effect of the sampling frequency f_s over the measured MPE ($m = 1$ and $d = 3$) of a sine wave. As we increase f_s, the number of patterns that overlap with a local maximum or minimum decreases considerably. Therefore, as we increase the sampling rate, the overall MPE slowly converges with the theoretical value predicted by (3.3).

From these examples we can offer the following observations. First, even if f_s satisfies the Nyquist-Shannon theorem $f_s > 2f_{max}$, this criterion is not enough to guarantee an accurate MPE estimation from a continuous deterministic signal. It is necessary that $f_s \gg f_{max}$ to avoid any significant deviations from the theoretical PE value in (3.3). Second, contrary to the bias effect discussed in Section (2.3.3), a small f_s produces an *overestimation* of MPE. The source of this effect comes from the number of patterns that overlap with a maxima or minima. Therefore, this is a geometric effect, rather than a statistical one.

So far, we can observe two important factors that have an effect on the MPE of deterministic signals: the sampling rate and the presence of maximum and minimum points. Since we have not explored neither non-differentiable functions nor chaotic systems, we cannot claim this list of factors is exhaustive. Nonetheless, here we present the basic considerations regarding the properties of deterministic functions for MPE calculations.

3.2.2 Deterministic Signals with Noise

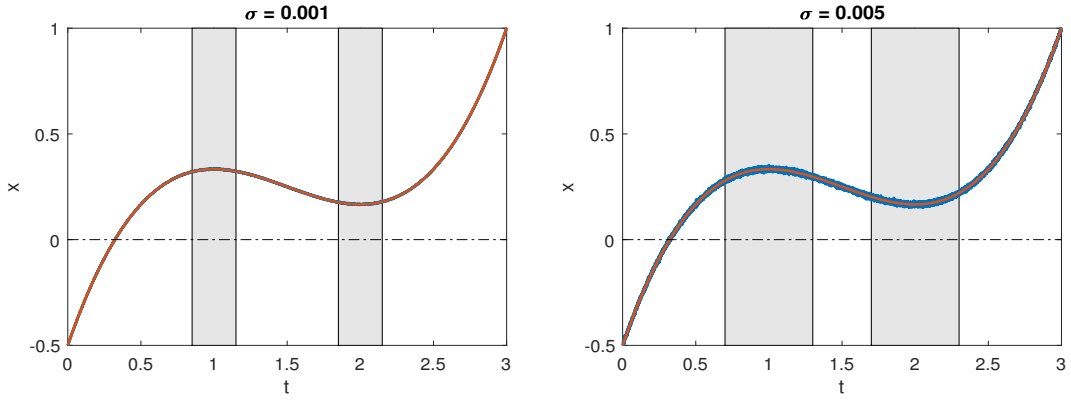
It is a well-known fact that white noise yield the maximum permutation entropy value. This corresponds to a uniformly distributed pattern mass probability function, where each pattern has the exact same chance of appearing in the signal. This presents itself in stark contrast to a deterministic signal, which will present a low entropy value if we measure it with a high enough sampling rate. As a consequence, it is natural to ask what would we expect from the MPE measurement of deterministic signal in the presence of noise. This is particularly relevant, since we expect to have at least a small amount of noise from a real signal measurement, and the MPE method is regarded as a robust method.

Intuitively, there will be no difference between noisy and clean signals from the perspective of entropy if the amount of noise is small in comparison to the amplitude of the patterns. Nonetheless, if the added noise has a sufficiently large variance, it will override the deterministic pattern, as described in Figure 3.3.

Therefore, in the case of deterministic signals with added noise, the magnitude of the slope becomes important —in contrast to the direction of the slope, as we saw in Section 3.2.1). In the case of a completely horizontal line, the addition of noise, regardless of its amplitude, will inevitably shift the MPE from zero to its maximum value. In contrast, curves with a pronounced slope should preserve their ordinal patterns, even in the presence of noise. Therefore, for a signal with clear local maxima and minima, we should see an increased number of nonmonotonic patterns as a product of noise, since the regions near these points have a slope close to zero. An example of one of such cases is shown in Figure 3.4.

In order to gauge the effect of amplitude over MPE of deterministic signals with white noise, we will test the case of a parabolic curve $x = t^2$ for $0 \leq t \leq 15$ seconds. The parabola has a slope that increases linearly with time for this region. This is particularly well-suited for our experiment. We thereby add white noise to the parabola, testing for different standard deviation values σ. We will calculate local

Figure 3.3: The presence of noise does not affect the signal patterns, as long as the variation is small compared to the curve's slope.



(a) (b)

Figure 3.4: Sampled cubic polynomial $x = \frac{1}{3}t^3 - (\frac{2}{3})t^2 + 2t - \frac{1}{2}$ for $t = [0, 3]$ seconds, with added white Gaussian noise with standard deviation of (a) $\sigma = 0.001$ and (b) $\sigma = 0.005$. In the regions near the local maximum and minimum, the white Gaussian noise, rather than the polynomial, determines the ordinal patterns present.

MPE for a sliding window of $\Delta t = 0.05$ seconds. The results are shown in Figures 3.5a and 3.5b.

Additionally, we want to explore the phenomenon from the opposite perspective, by having a straight line $x = At$ with fixed slope A, and white noise with a standard deviation σ that increases with time. By following the same procedure, we obtained the results shown in Figures 3.5c and 3.5d.

As we could expect, as we increase the slope of the signal, MPE is reduced from a region where the noise dominates (maximum entropy) to a region where the deterministic line has most effect (minimum entropy). Figure 3.5b confirms this trend. The opposite effect occurs when we increase the standard deviation of the noise, as it moves from minimum to maximum entropy. Surprisingly, the shift from one regime to the other is not sharp. Instead, we observe a transition curve which depends on the exact slope and standard deviation values. Here, we will not attempt to characterize this behavior formally. It suffices to say that there is a strong interaction between the geometry of the signal and the intensity of the noise, regarding the overall resulting MPE measurement.

Now that we know the overall MPE effect of the interaction between noise amplitude and slope, we should reintroduce the sampling rate. Geometrically, if the sampled data points for a noisy signal are close together in time, the vertical distance between values decreases, even for a pronounced slope. Therefore, for a constant noise standard deviation, we should expect an increase of MPE values when the sampling frequency increases.

To test this effect over the MPE of noisy signals, we revisit the example of the sinusoidal wave function $x = sin(2\pi ft)$ from $0 \leq t \leq 5$ seconds, with increasing sampling frequency f_s. We perform the MPE calculation for $m = 1$ at different values of SNR. The results are shown in Figure 3.6.

At first glance, we can see in Figure 3.6a that the SNR has a profound impact on the

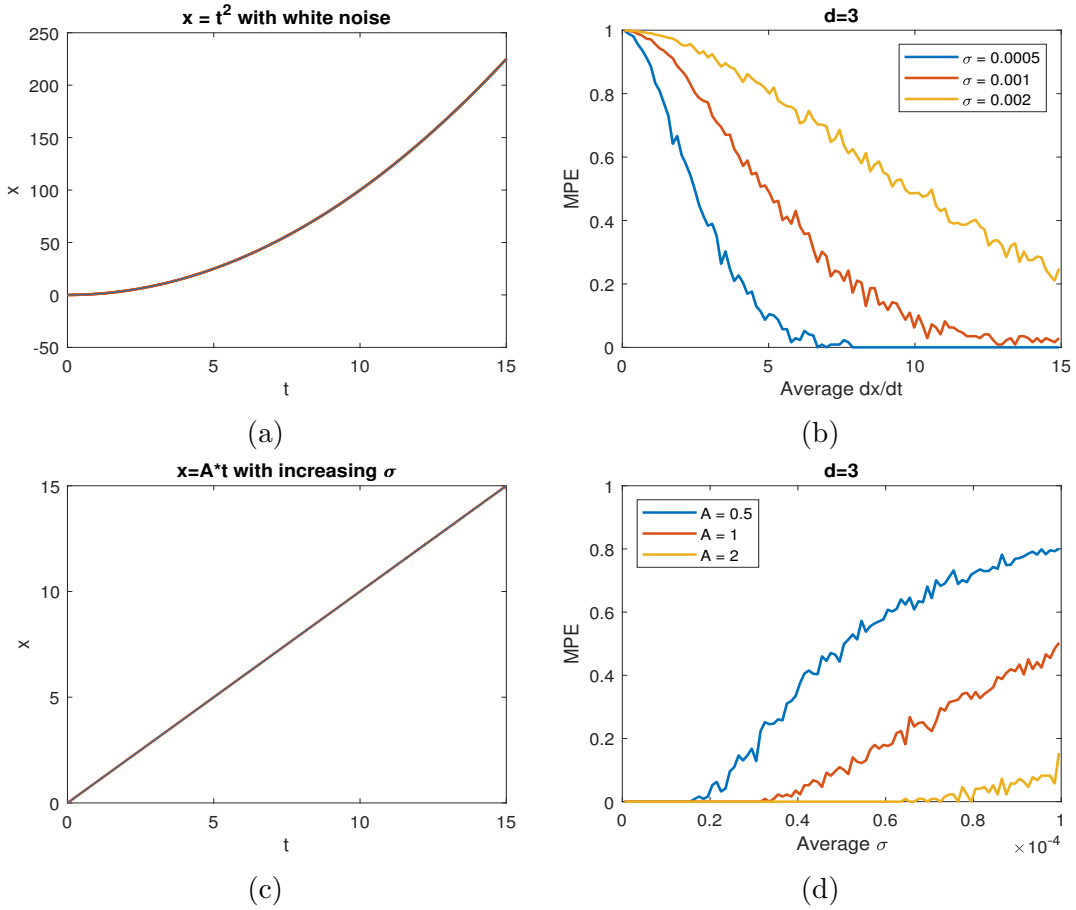
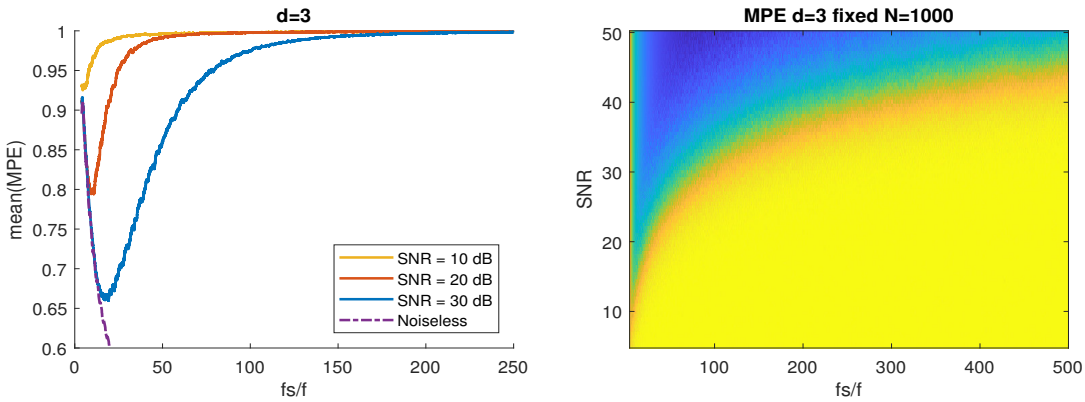


Figure 3.5: (a) Parabolic curve $x = t^2$ for $0 \leq t \leq 15$ seconds, with added white noise at $\sigma = \{0.005, 0.001, 0.002\}$. (b) MPE at $d = 3$ within respect to the linearly increasing slope of the parabolic curve at different values of σ. (c) Straight line $x = At$ with added white noise at increasing $\sigma = [1e-9, 1e-6]$. (d) MPE at $d = 3$ within respect to a linearly increasing σ at different values for the slope A. The MPE values in (b) and (d) come from a local sliding window of $\Delta t = 0.05$ sec. The sampling rate for this measurements is $f_s = 6670Hz$.

overall results, compared to the noiseless MPE vs. f_s curve in Figure 3.2d. Instead of reducing MPE asymptotically to the theoretical continuous entropy, the value of noisy sine waves presents a minimum, and then increases with higher f_s. For a sufficiently large sampling frequency, the noise effect dominates over the signal, regardless of SNR. In Figure 3.6b, we present an MPE surface, respect to both f_s and SNR. We can see a clear frontier between the regions where the sine wave dominates (in blue), and the region where the noise effect prevails (in yellow). As we increase the SNR, the noise effect requires higher sampling frequencies to appear, and the frontier within regions is not linear.

This entropy increase with sampling rate can be explained by the vertical distance between data points: the closer the data points are in time, the closer they are in vertical distance, even for a steep slope. Therefore, we should also be wary of oversampling the signal, since we are increasing the effect of noise, as exemplified in



(a) (b)

Figure 3.6: Sine wave function $x = sin(2\pi ft)$ from $0 \leq t \leq 5$ seconds, with increasing sampling frequency f_s, in the presence of white noise at different signal-to-noise ratio (SNR). (a) Mean MPE vs. f_s at SNR = 10 dB, 20 dB, and 30 dB. The MPE follows the MPE of the noiseless sine wave for low f_s, and approaches maximum entropy at high sampling rates. (b) MPE surface representation, with f_s and SNR as independent variables. Low entropy values are shown in blue, and high entropy in yellow. We observe a clear frontier between regions where noise dominates (yellow), or the underlying deterministic signal is more important (blue).

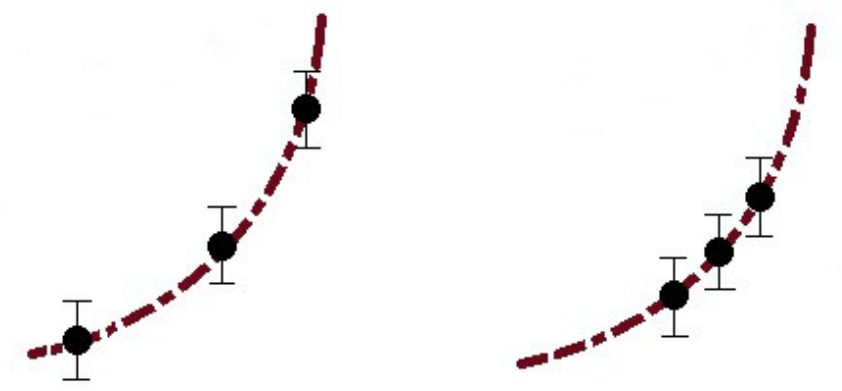


Figure 3.7: For a fixed signal-to-noise ratio (SNR), an increased sampling rate f_s implies the data points are closer together, both in time and amplitude. Therefore, when f_s is high, the pattern noise dominates over the deterministic signal, and thus, the ordinal pattern is modified.

Figure 3.7. Fortunately, the coarse-graining procedure, characteristic of the MPE, acts as a downsampling parameter, which can correct this effect. Therefore, a careful exploration of the MPE signals with respect to noise is always recommended before further analysis.

3.3 MPE on Models with Random Gaussian Signals

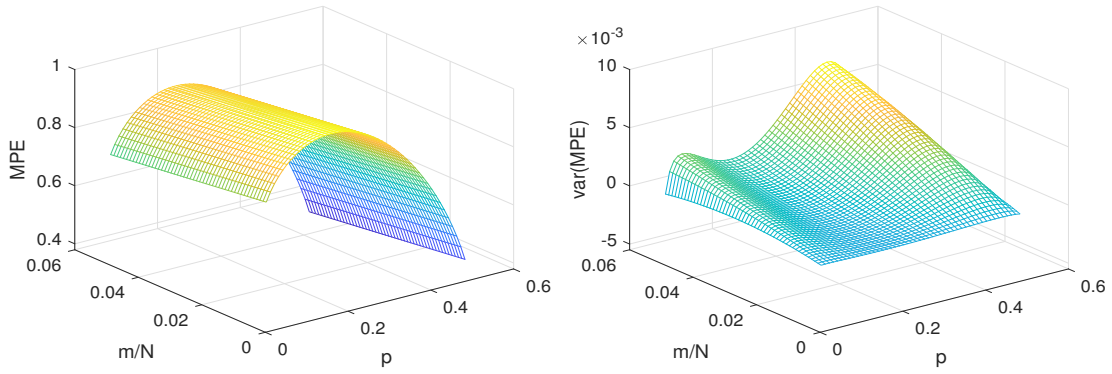
In contrast to the previous section, here we will focus on characterizing the MPE of random stochastic processes, focusing on Gaussian processes. Here, the information content from the signal does not come directly from the geometry of the model, but from its autocorrelation function [58]. White Gaussian noise (wGn), as we discussed before, produces a uniform ordinal pattern distribution. Since wGn is uncorrelated, we can make no inferences regarding possible future patterns, given the information we have. This is not the case when we have an autocorrelation other than zero.

The models in this section share the following characteristics. First, it is evident that each data point in the series has at least one Gaussian error term. Second, the signals are stationary —this property is not strictly necessary, as we can see from section 3.2.2. Third, the Gaussian random variable has a constant variance, and hence, presents homoscedasticity. Models with nonconstant variance will be studied in future work. Lastly, and most importantly, we will not restrict ourselves to white Gaussian noise (wGn). As we will see, the autocorrelation in this process presents some information content, manifest in the MPE.

The coarse-graining procedure presents an additional challenge for the calculation of Gaussian ordinal patterns. In the general case, there is no guarantee that the pattern probability distribution of these signals will remain the same across time scales. Therefore, as part of our analysis of Gaussian models, we will also provide the evolution of the pattern probabilities as a function of time scale. This step is essential to obtain the MPE as a function of the process parameters.

3.3.1 Gaussian Ordinal Pattern Distributions

Bandt and Shiha [58] first observed the pattern distributions of Gaussian noise by taking advantage of the pattern symmetries present in these models. For embedded dimension $d = 2$, the two patterns have the exact same probability $p_1 = p_2 = 1/2$, with no effect stemming from the autocorrelation in the noise. This is true because of the stationary constraint outlined before, and the fact that, for a Gaussian random variable, the median is equal to the mean. By the definition of the median, the probability of having an increasing pattern is $P(X_t > X_{t+1}) = 1/2$. Therefore, unless the signal is non-stationary, the signal will be balanced [58]. Since the coarse-graining procedure is an averaging transformation, the resulting coarse signal is also Gaussian and stationary. Although this is a general result, it is also not useful in the characterization of the Gaussian process.



(a) (b)

Figure 3.8: Three-dimensional surface for (a) MPE, and (b) var(MPE) for Gaussian models and dimension $d = 3$. This representation is possible since the Gaussian pattern symmetries (3.4) allow the pattern pdf to be dependent on only one variable p_1.

For $d = 3$, the patterns present an interesting structure. By exploiting the general symmetry properties of patterns of dimension three, applied to Gaussian distributions, the following relationships arise between pattern probabilities [58],

$$\begin{aligned} p_1 &= p_6 \\ p_2 &= p_3 = p_4 = p_5 = \tfrac{1}{4}(1 - 2p_1), \end{aligned} \tag{3.4}$$

being p_1 the monotonically increasing pattern, and the remaining p_i follow the lexicographic order of the permutations [59]. By using Plackett's lemma [58], they found an explicit relationship between the autocorrelation function of the process ($\rho(\lambda)$), and the increasing pattern probability,

$$p_1 = \frac{1}{\pi} \arcsin\left(\frac{1}{2}\sqrt{\frac{1-\rho(2)}{1-\rho(1)}}\right), \tag{3.5}$$

where $\rho(1)$ is the autocorrelation between adjacent points, and $\rho(2)$ is the autocorrelation between data points two steps apart.

As we can clearly see, all of the pattern's distribution is completely characterized by computing the first pattern probability. This reduces the problem from five dimensions (3!-1 degrees of freedom) to just one. We also note that this relationship will hold for any stationary Gaussian process. Any coarse-grained signal whose source time series is a Gaussian process, will itself be a Gaussian process, and will still obey equation (3.5). From this point, we rewrite the original PE definition (2.3) using the pattern symmetries in (3.4):

$$H = -2p_1 \ln(p_1) - (1 - 2p_1) \ln\left(\tfrac{1}{4}(1 - 2p_1)\right). \tag{3.6}$$

The problem now lies in obtaining the autocorrelation functions for the coarse-grained versions of these models. By introducing the time scale to the autocorrelation function, we can easily obtain the pattern probability $p_1^{(m)}$ (the pattern probability at scale m) and the theoretical MPE value for the model.

Before exploring more specific Gaussian processes, we should address the embedded dimension. For the purposes of this section, we will limit our analysis to the case of $d = 3$. Although there are explicit symmetries obtained for $d = 4$, some of the resulting pattern probabilities lie in the complex plane, which complicates its interpretation. For $d \geq 5$, the pattern probabilities have no closed form [58].

3.3.2 White Gaussian Noise and Fractional Gaussian Noise

As described by Manderlbrot [60], the fGn models a surprisingly large amount of natural phenomena across time, from hydrology to stock markets. Interestingly, fGn does include white Gaussian noise as a special case. The fGn fractal properties are of special interest under the MPE approach.

In this section we will lay down the necessary theory about fGn, with special emphasis on its autocorrelation function, as we have previously published in [2]. This will allow us to build a model for coarse-grained fGn signals (cgfGn). By obtaining the autocorrelation function of cgfGn, we will completely characterize the pattern distribution function, and thus, the theoretical MPE with respect to the time scale m.

It is not necessary to state the explicit form of the fGn signals, but we will need to relate it to the fractional Brownian motion (fBm). For $n \in \mathbb{N}$, we will write fBm signal as $X_B(n)$ and fGn as $X_G(n)$, corresponding to time t_n. Since fGn and fBm are continuous, we need to work with the discrete sampled version.

These models are dependent of the Hurst exponent $0 < h < 1$, which is used to model long-term autocorrelations that are proportional to t^h [60] (the Hurst exponent is usually found in literature as uppercase H. Here, we use lowercase h to avoid using H, which we use for MPE). For $h > 0.5$, each new data point is positively correlated to all previous ones. For $h < 0.5$, each new point is inversely correlated to all its history. In the case of $h = 0.5$, the model is identical to uncorrelated Gaussian noise. For this reason, we will consider wGn as a special case of fGn. Therefore, all the results for fGn also apply to the classical, uncorrelated wGn.

Regarding some properties of fBm, we can state that [61],

	Fractional Brownian motion	
1	Model	$X_B(n) = \sum_{i=0}^{n} X_G(i)$
2	Expected value	$E[X_B(n)] = 0$
3	Variance	$var(X_B(n)) = E[X_B^2(n)] = \sigma^2 n^{2h}$
4	Covariance	$cov(X_B(n), X_B(n+\lambda)) = \frac{\sigma^2}{2}\left(n^{2h} + \lvert n+\lambda\rvert^{2h} - (\lambda)^{2h}\right)$
5	Correlation	$\rho_B(n,\lambda) = \frac{1}{2n^{2h}}\left(n^{2h} + \lvert n+\lambda\rvert^{2h} - (\lambda)^{2h}\right)$

(3.7)

where λ is the index shift between data points. The value of $X_B(n)$ is the sum of all the prior individual, nonindependent Gaussian steps. The autocorrelation between different points in the fBm depends on the distance between data points, as well as the absolute position in time. Similarly, we can write the properties of fGn as follows,

	Fractional Gaussian noise	
1	Model	$X_G(n) = X_B(n) - X_B(n-1)$
2	Expected value	$E[X_G(n)] = 0$
3	Variance	$var(X_G(n)) = E[X_G^2(n)] = \sigma^2$
4	Covariance	$cov(X_G(n), X_G(n+\lambda)) = \frac{\sigma^2}{2}\left(\lvert\lambda+1\rvert^{2h} + \lvert\lambda-1\rvert^{2h} - 2\lvert\lambda\rvert^{2h}\right)$
5	Correlation	$\rho_G(\lambda) = \frac{1}{2}\left(\lvert\lambda+1\rvert^{2h} + \lvert\lambda-1\rvert^{2h} - 2\lambda^{2h}\right)$

(3.8)

Here, the fGn is defined with respect to fBm. Each fGn instance is the increment between the fBm at the same time t_n, compared to the realization at t_{n-1}. In this case, the autocorrelation is not dependent directly on the position in time. Instead, it is solely dependent on the relative distance between data points.

Using these properties with the definition of the coarse-graining procedure (2.5), we can express the properties of a cgfGn by introducing the relationship between fBm and fGn (3.8), as follows,

$$\begin{aligned} X_G^{(m)}(n) &= \frac{1}{m}\sum_{j=m(n-1)+1}^{mn} X_G(j) \\ &= \tfrac{1}{m}\left[X_B(mn) - X_B(m(n-1))\right]. \end{aligned} \tag{3.9}$$

By writing the cgfGn in terms of fBm, we simplify the expression enough to obtain the moments manually.

$$E[X_G^{(m)}(n)] = 0 \tag{3.10}$$

$$\begin{aligned} var(X_G^{(m)}(n)) &= E[{X_G^{(m)}}^2(n)] \\ &= \tfrac{1}{m^2} E\left[(X_B(mn) - X_B(m(n-1)))^2\right] \\ &= \sigma^2 m^{2(h-1)}. \end{aligned} \tag{3.11}$$

We will take advantage of the variance and covariance of fBm in equation (3.7), where $\lambda = m$. More importantly, we need to obtain the autocovariance and autocorrelation of (3.9). This can be done by using the same reasoning as with the cgfGn variance, as follows,

$$\begin{aligned} cov(X_G^{(m)}(n), X_G^{(m)}(n+\lambda)) &= E[X_G^{(m)}(n) X_G^{(m)}(n+\lambda)] \\ &= \tfrac{1}{m^2} E\big[(X_B(mn) - X_B(m(n-1))) \\ &\qquad (X_B(m(n+\lambda)) - X_B(m(n+\lambda-1))\big] \\ &= \tfrac{\sigma^2}{2} m^{2(h-1)} \left(|\lambda+1|^{2h} + |\lambda-1|^{2h} - 2\lambda^{2h}\right) \end{aligned} \tag{3.12}$$

$$\begin{aligned} \rho_{\tilde{G}}^{(m)}(\lambda) &= \tfrac{1}{2} \left(|\lambda+1|^{2h} + |\lambda-1|^{2h} - 2\lambda^{2h}\right) \\ &= \rho_G(\lambda). \end{aligned} \tag{3.13}$$

By dividing (3.12) by (3.11), we get the autocorrelation of the coarse-grained signal, which is *exactly* the same as the autocorrelation of the original fGn signal in (3.8). This is consistent with the self-similarity property of the fractional Gaussian noise signals [46],

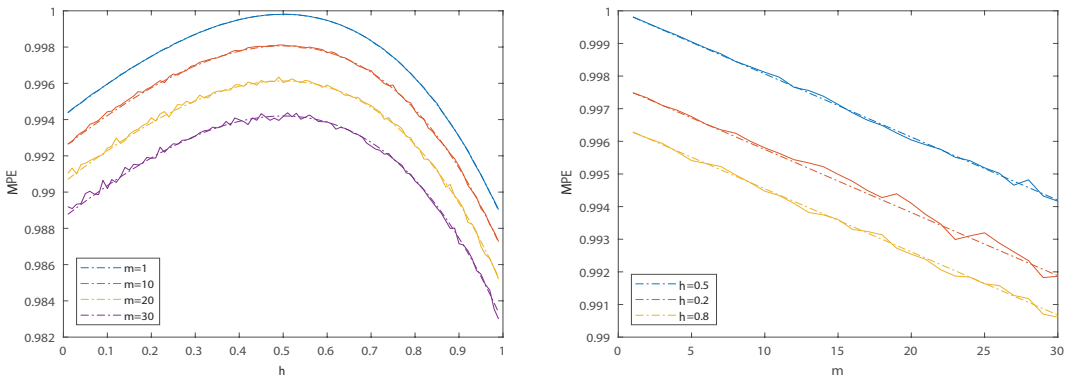
$$X_G(n) \overset{d}{=} c^h X_G(c^{-1}n), \tag{3.14}$$

which are equal in distribution. Therefore, by revisiting the pattern probabilities for Gaussian signals in equation (3.5), the MPE of coarse-grained fGn remains constant for all time scales and for any given Hurst exponent h. As a consequence of this result, the MPE of white Gaussian noise is invariant to the time scale coarse-graining transformation. When we compute the pattern probabilities from (3.5), we obtain,

$$p_{1,fGn}^{(m)} = \frac{1}{\pi} \arcsin \frac{1}{4} \sqrt{\frac{1 + 2^{2h+1} - 3^{2h}}{1 - 2^{2(h-1)}}}, \tag{3.15}$$

which we use to obtain the MPE of fGn by means of equation (3.6).

In Figure 3.9 we observe the mean MPE from 1500 fGn signal simulations (length $N = 5000$). Contrary to the result found in (3.13), we observe a linear downward trend for the MPE with respect to time scale m. This is easily explained by the MPE bias from equation (2.23). We included the bias in the theoretical prediction (dotted lines) in Fig. 3.9b. Since the simulated results only on a downward trend, it implies that the MPE for fGn is not affected by the scale, other than the linear bias effect. Therefore, we conclude that the fGn, by virtue of the self-similarity property, contains the same MPE regardless of time scale.



(a) (b)

Figure 3.9: (a) Average MPE of fGn with respect to the Hurst exponent, for different time scales m. The curves get downshifted with increasing m. (b) MPE of fGn respect to m, for Hurst exponents $h = \{0.2, 0.5, 0.8\}$. The dotted lines represent the theoretical predictions, while the solid lines measure the mean MPE from 1500 signals of length $N = 5000$.

3.3.3 First-Order AR Models

Autorregressive (AR) [62] Gaussian processes are ubiquitous tools in signal processing, used to represent complicated time series phenomena. The AR process takes into account the influence of the parameter $\tilde{p}$ (not to be confused with p, which refers to probability) past data points in a new iteration of the process, plus a new random Gaussian innovation. The general AR ($\tilde{p}$) process is described as,

$$X_{AR(\tilde{p})}(n) = c + \varepsilon_n + \sum_{i=1}^{\tilde{p}} \phi_i X_{AR(\tilde{p})}(n-i), \tag{3.16}$$

for time t_n, where the ϵ_n terms are assumed to be Gaussian, independent and identically distributed (iid), with mean zero and constant σ^2 variance. The term $\tilde{p}$ denotes the number of elements, or lags, taken in account for the AR model.

The first degree AR ($\tilde{p} = 1$) is expressed as,

$$X_{AR(1)}(n) = c + \varepsilon_n + \phi X_{AR(1)}(n-1), \tag{3.17}$$

where

$$E[X_{AR(1)}(n)] = c/(1-\phi) \tag{3.18}$$
$$var(X_{AR(1)}(n)) = \sigma^2/(1-\phi^2) \tag{3.19}$$
$$\rho_{AR(1)}(\lambda) = \phi^{|\lambda|}. \tag{3.20}$$

At this point, we will limit the analysis to the first-order AR processes, so that we can ensure the coarse-grained pattern probabilities have a closed form. More general cases will be explored in Section 3.3.5.

As we have previously done in [4], we will apply the coarse-grained procedure to the AR(1) processes. We here set the constant $c = 0$, without loss of generality. If we apply the coarse-graining procedure expression in (2.5) into first-order AR process (3.17), we get,

$$X^{(m)}_{AR(1)}(n) = \frac{\phi}{m}\Big(\frac{1-\phi^m}{1-\phi}\Big)X_{AR(1)}(m(n-1)) + \frac{1}{m}\sum_{j=1}^{m}\Big(\frac{1-\phi^j}{1-\phi}\Big)\varepsilon_{mn+1-j}. \tag{3.21}$$

The variance of (3.21) is

$$var(X^{(m)}_{AR(1)}(n)) = \frac{\sigma^2}{m^2(1-\phi^2)}\left[m\left(\frac{1+\phi}{1-\phi}\right) - \frac{2\phi}{1-\phi}\left(\frac{1-\phi^m}{1-\phi}\right)\right]. \tag{3.22}$$

The autocovariance function γ_λ for (3.21) is given by

$$\gamma_\lambda = cov\big(X^{(m)}_{AR(1)}(n), X^{(m)}_{AR(1)}(n+\lambda)\big) = \\ \frac{\phi^{m\lambda+2}}{m^2}\Big(\frac{1-\phi^m}{1-\phi}\Big)^2 var(X_{AR(1)}(m(n-1))) \; + \frac{\sigma^2}{m^2}\phi^{m(\lambda+1)}\frac{1-\phi^m}{1-\phi}\sum_{j=1}^{m}\left(\frac{1-\phi^j}{1-\phi}\right)\phi^j. \tag{3.23}$$

and the autocorrelation is

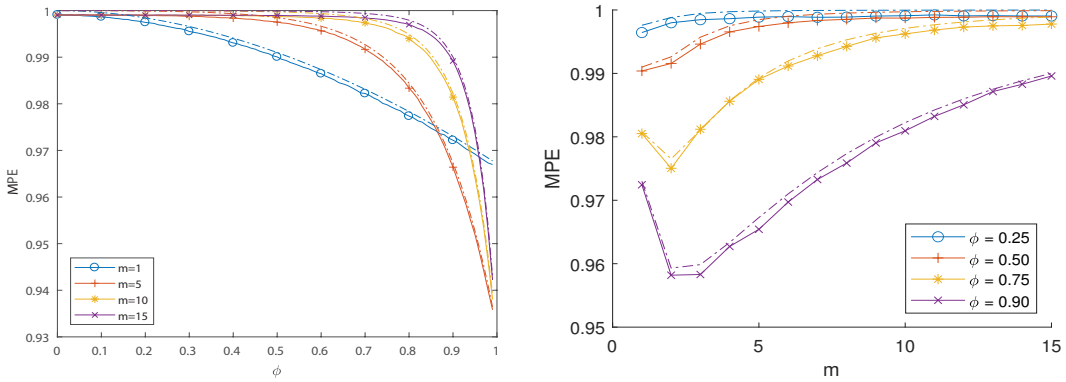
$$\rho^{(m)}_{AR(1)}(\lambda) = \phi^{m(\lambda-1)+1}\left[\frac{(1-\phi^m)^2}{m(1-\phi^2) - 2\phi(1-\phi^m)}\right]. \tag{3.24}$$

for $|\lambda| > 0$, and $\rho^{(m)}_{AR(1)}(0) = 1$.

From this point, it is straightforward to obtain the probability of obtaining the increasing pattern probability for embedded dimension $d = 3$ by using this autocorrelation function (3.24) to obtain the first pattern probability (3.5), and the PE equation in (3.6). We note here that (3.21) is *not* itself an AR(1) process. Nonetheless, it is still a stationary Gaussian process, so the assumptions required for (3.5) still apply. Therefore, the increasing ordinal pattern for coarse-grained AR(1) is

$$p^{(m)}_{1,AR(1)} = \frac{1}{\pi}arcsin\left(\frac{1}{2}\sqrt{\frac{m(1-\phi^2) - \phi(2-\phi^m)(1-\phi^{2m})}{m(1-\phi^2) - \phi(1-\phi^m)(3-\phi^m)}}\right). \tag{3.25}$$

Even when the equation (3.25) is cumbersome, we have a closed expression for the pattern distribution in the AR(1) model. Moreover, the distribution is only a function of the model's parameters and the time scale. This means the MPE of this model is completely characterized, and any deviation from these results come from effects outside the statistical properties of the signal. In figure 3.10 we can observe the comparison between our MPE models and actual MPE measurements from simulated AR(1) processes.



(a) (b)

Figure 3.10: (a) MPE curves for AR(1) with respect to their corresponding model parameter ϕ. Different curves correspond to different time scales m, as shown directly in the plots. (b) MPE curves with respect to m, with $\phi = \{0.25, 0.50, 0.75, 0.90\}$. Dotted lines represent theoretical MPE values, while solid lines show the resulting mean MPE from 1500 signals of $N = 1000$.

As we can observe from Figure 3.10, increasing time scales tend to increase the MPE, except for ϕ values close to 1. For high ϕ, the signal presents a minimum entropy at a scale different than $m = 1$. Consequently, we would want to find the critical value of ϕ above which this effect begins to occur. Therefore, in the limit case, where we set the probabilities $p^{m=1}_{1,AR(1)} = p^{m=2}_{1,AR(1)}$ to be equal, we get

$$\begin{aligned} \frac{1-\rho^{(m=1)}_{AR(1)}(2)}{1-\rho^{(m=1)}_{AR(1)}(1)} &= \frac{1-\rho^{(m=2)}_{AR(1)}(2)}{1-\rho^{(m=2)}_{AR(1)}(1)} \\ \phi(\phi-1)(\phi^2+\phi-1) &= 0 \\ \phi = -1/2 + \sqrt{5}/2 &\approx 0.618. \end{aligned} \tag{3.26}$$

Thus, as we previously published in [4], this result states that coarse-grained AR(1) models a ϕ parameter greater than the Golden Ratio, presents more regularity and structure on longer time scales than the original signal.

3.3.4 First-Order MA Models

As in section 3.3.3, the moving average process [62] is also one of the most referenced techniques in the modelization of random time series. Contrary to the AR model, the

MA process computes each data point as the weigthed sum of $\tilde{q}$ previous innovations, in addition to the present one. The general MA($\tilde{q}$) process is described as,

$$X_{MA(\tilde{q})}(n) = c + \varepsilon_n + \sum_{j=1}^{\tilde{q}} \theta_j \varepsilon_{n-j}. \tag{3.27}$$

for time t_n, where the ϵ terms are assumed to be Gaussian, independent and identically distributed (iid), with mean zero and constant σ^2 variance. The term $\tilde{q}$, is the number of lags taken in account for the MA model. The first degree MA ($\tilde{p} = 0$, $\tilde{q} = 1$), which is the simplest case, is explicitly written as

$$X_{MA(1)}(n) = c + \varepsilon_n + \theta \varepsilon_{n-1}, \tag{3.28}$$

with moments

$$E[X_{MA(1)}(n)] = c \tag{3.29}$$
$$var(X_{MA(1)}(n)) = \sigma^2(1 + \theta^2). \tag{3.30}$$

The normalized autocorrelation is given by

$$\rho_{MA(1)}(\lambda) = \begin{cases} 1, & \text{if } \lambda = 0 \\ \theta/(1+\theta^2), & \text{if } |\lambda| = 1 \\ 0, & \text{otherwise,} \end{cases} \tag{3.31}$$

where λ is the time shift between data points $X_{MA(1)}(n)$ and $X_{MA(1)}(n \pm \lambda)$.

As previously done in [4], we will apply the coarse-grained procedure to $MA(1)$ processes. We will, once again, take the coarse-grained definition from equation (2.5), as in section 3.3.3. We will once again set $c = 0$ without loss of generality. For any time scale m, the coarse-grained MA(1) (cgMA(1)) process is,

$$X^{(m)}_{MA(1)}(n) = \frac{\theta}{m}\varepsilon_{m(n-1)} + \frac{1}{m}\varepsilon_{mn} + \frac{1+\theta}{m} \sum_{j=m(n-1)+1}^{mn-1} \varepsilon_j, \tag{3.32}$$

being $n \in \mathbb{N}$ the index variable of the new coarse-grained signal, and m the scale. From this expression, we will derive the autocovariance function,

$$Cov\big(X^{(m)}_{MA(1)}(n), X^{(m)}_{MA(1)}(n+\lambda)\big) = \begin{cases} \frac{\sigma^2}{m}\Big(1 + \theta^2 + 2\Big(\frac{m-1}{m}\Big)\theta\Big), & \text{if } \lambda = 0 \\ \frac{\theta}{m}\sigma^2, & \text{if } |\lambda| = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.33}$$

For $\lambda = 0$, we have the variance of the coarse-grained MA(1) model. If we divide (3.33) by its own variance, we obtain the autocorrelation function,

$$\rho^{(m)}_{MA(1)}(\lambda) = \begin{cases} 1, & \text{if } \lambda = 0 \\ \frac{\theta}{m(1+\theta^2)+2(m-1)\theta}, & \text{if } |\lambda| = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.34}$$

Once again, it is straightforward to calculate the probability of obtaining the increasing pattern probability for embedded dimension $d = 3$, by using this autocorrelation function (3.34) to obtain the first pattern probability (3.5), and the PE equation in (3.6). The new cgMA(1) process is still Gaussian, even when the original properties of MA(1) do not apply. Therefore, the use of (3.5) is still valid. The increasing ordinal pattern for coarse-grained MA(1) is

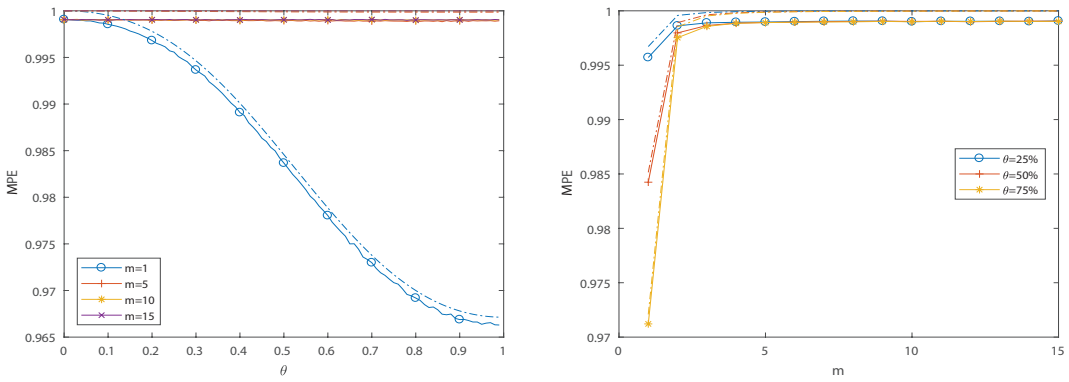
$$p^{(m)}_{1,MA(1)} = \frac{1}{\pi} arcsin\left(\frac{1}{2}\sqrt{\frac{m(1+\theta^2)+(2m-2)\theta}{m(1+\theta^2)+(2m-3)\theta}}\right). \tag{3.35}$$

Here, in contrast to the case of fractional Gaussian noise, the pattern probabilities do not stay constant across time scale. Nonetheless, we have an explicit equation that governs the evolution of the pattern probability distribution, based solely on the MA(1) parameter θ and time scale m. It is cumbersome to express the MPE of MA(1) using this expression, but the MPE value can be easily computed by using Equation (3.6). We must note here that the expression inside the *arcsin* function is almost equal to 0.5, which yields a pattern probability of 1/6, even for small m. This implies that, other than the case where $m = 1$, the coarse-grained MA(1) process will be virtually indistinguishable from noise. Figure 3.11 shows the MPE for the coarse-grained MA(1) process for different values of θ and time scale m.

For the MA(1) process, we can appreciate that only the first time scale $m = 1$ presents a noticeable deviation from the maximum entropy. This implies that the coarse-graining procedure, in fact, nullifies the autocorrelation effect on the original signal. For scales greater than $m = 1$, the process is indistinguishable from noise, regardless of the model parameter θ. This is not surprising, since distant points in a MA(1) process are not correlated. The coarse-graining procedure reflects this.

3.3.5 General Formulation for Correlated Gaussian Models

When we consider the general formulation of the Gaussian models, we observe the coarse-grained signals are still Gaussian, albeit with different autocorrelation functions. This property still allows the use of the symmetries in (3.4) for all time scales to obtain the MPE based solely on the autocorrelation function at dimension $d = 3$. If we solely rely on these assumptions about a model, we can formulate a general expression to properly describe the signal autocorrelation for all m.



(a) (b)

Figure 3.11: (a) MPE curves for MA(1) with respect to their corresponding model parameter θ. Different curves correspond to different time scales m, as shown directly in the plots. (b) MPE curves with respect to m, with $\theta = \{0.25, 0.5, 0.75\}$. Dotted lines represent theoretical MPE values, while solid lines show the resulting mean MPE from 1500 signals of $N = 1000$.

Therefore, in this section we will obtain the autocorrelation function for the general coarse-grained Gaussian stationary signal. First we need to recall the coarse-grained procedure definition from Equation (2.5), now for a random process,

$$X^{(m)}(n) = \frac{1}{m} \sum_{j=m(n-1)+1}^{mn} X(j),$$

where n is the time index of the original signal and m is the time scale. We will rewrite this expression in vectorial form as a dot product, as follows,

$$X^{(m)}(n) = \frac{1}{m}\mathbf{1}' \begin{bmatrix} X(m(n-1)+1) \\ X(m(n-1)+2) \\ \vdots \\ X(mn) \end{bmatrix} = \frac{1}{m}\mathbf{1}'\boldsymbol{X}^{(m)}(n^{(m)}). \tag{3.36}$$

Without any knowledge of the signal, we cannot know its expected value *a priori*. Nonetheless, we can at least state the form the general expression for the covariance.

$$
\begin{aligned}
&cov(X^{(m)}(n), X^{(m)}(n+\lambda)) \\
&\qquad = E\left[X^{(m)}(n)X^{(m)}(n+\lambda)\right] \\
&\qquad\quad - E\left[X^{(m)}(n+\lambda)\right] E\left[X^{(m)}(n+\lambda)\right] \\
&\qquad = \frac{1}{m^2} E\left[\mathbf{1}'\boldsymbol{X}^{(m)}(n)\boldsymbol{X}^{(m)}(n+\lambda)'\mathbf{1}\right] \\
&\qquad\quad - \frac{1}{m^2} E\left[\mathbf{1}'\boldsymbol{X}^{(m)}(n+\lambda)\right] E\left[\boldsymbol{X}^{(m)}(n+\lambda)'\mathbf{1}\right] \\
&\qquad = \frac{1}{m^2}\mathbf{1}'\left(E\left[\boldsymbol{X}^{(m)}(n)\boldsymbol{X}^{(m)}(n+\lambda)'\right]\right. \\
&\qquad\quad \left. - E\left[\boldsymbol{X}^{(m)}(n+\lambda)\right] E\left[\boldsymbol{X}^{(m)}(n+\lambda)'\right]\right)\mathbf{1} \\
&\qquad = \frac{1}{m^2}\mathbf{1}'\,\boldsymbol{K}^{(m)}\left(\boldsymbol{X}^{(m)}(n), \boldsymbol{X}^{(m)}(n+\lambda)\right)\,\mathbf{1}.
\end{aligned}
\tag{3.37}
$$

In this expression, $\boldsymbol{K}$ is the mxm covariance matrix:

$$
\begin{aligned}
&\boldsymbol{K}^{(m)}\left(\boldsymbol{X}^{(m)}(n), \boldsymbol{X}^{(m)}(n+\lambda)\right) = \\
&\begin{bmatrix}
cov(X(m(n-1)+1), X(m(n+\lambda-1)+1)) & \dots & cov(X(m(n-1)+1), X(m(n+\lambda))) \\
cov(X(m(n-1)+2), X(m(n+\lambda-1)+1)) & \dots & cov(X(m(n-1)+2), X(m(n+\lambda))) \\
\vdots & \dots & \vdots \\
cov(X(mn), X(m(n+\lambda-1)+1)) & \dots & cov(X(mn), X(m(n+\lambda)))
\end{bmatrix}
\end{aligned}
\tag{3.38}
$$

It is evident from equation (3.37) that the sum all the elements of $\boldsymbol{K}^{(m)}$ will lead to the $cov(X^{(m)}(n), X^{(m)}(n+\lambda))$ we are looking for. This sum is written in a compact form by means of a matrix quadratic form. Regarding the inner structure of $\boldsymbol{K}$, we observe the distance between data points change in a predictable way. Each time we move one diagonal below, the distance between data points compared by the covariance is reduced by one. The opposite is true if we move above the main diagonal of $\boldsymbol{K}$. If the signal model has constant variance σ_x^2, we can extract this common element from the covariance matrix. This will help in the display of the diagonal properties:

$$
\begin{aligned}
&\boldsymbol{K}^{(m)}\left(\boldsymbol{X}^{(m)}(n), \boldsymbol{X}^{(m)}(n+\lambda)\right) = \sigma_x^2 \boldsymbol{R}^{(m)}(m\lambda) \\
&= \sigma_x^2 \begin{bmatrix}
\rho(m\lambda) & \rho(m\lambda+1) & \dots & \rho(m\lambda+m-1) \\
\rho(m\lambda-1) & \rho(m\lambda) & \dots & \rho(m\lambda+m-2) \\
\vdots & \vdots & \ddots & \vdots \\
\rho(m\lambda-m+1) & \rho(m\lambda-m+2) & \dots & \rho(m\lambda))
\end{bmatrix}
\end{aligned}
\tag{3.39}
$$

Here, we define the autocorrelation matrix $\boldsymbol{R}(m\lambda)$ as a Toeplitz matrix, with the same geometric properties as $\boldsymbol{K}$. The time shift in the autocorrelation function decreases one point as we go down in the diagonals, and increases as we go up. At this point, we must address a special case where this behavior is not true. If we

compute the autocorrelation matrix between a vector segment $\boldsymbol{X}^{(m)}(n)$ and itself, we obtain,

$$\boldsymbol{R}^{(m)}(0) = \begin{bmatrix} \rho(0) & \rho(1) & \dots & \rho(m-1) \\ \rho(1) & \rho(0) & \dots & \rho(m-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(m-1) & \rho(m-2) & \dots & \rho(0)) \end{bmatrix}, \tag{3.40}$$

since $\rho(-\lambda) = \rho(\lambda)$. $\boldsymbol{R}(0)$ is the only instance where the autocorrelation matrix is symmetric, and the strictly increasing distance with the matrix diagonals do not apply.

From Equation (3.37) we can now compute the autocorrelation function of the coarse-grained signal as follows,

$$\rho^{(m)}(\lambda) = \frac{cov(X^{(m)}(n), X^{(m)}(n+\lambda))}{var(X^{(m)}(n))} = \frac{\mathbf{1}'\boldsymbol{K}^{(m)}(\boldsymbol{X}^{(m)}(n), \boldsymbol{X}^{(m)}(n+\lambda))\mathbf{1}}{\mathbf{1}'\boldsymbol{K}^{(m)}(\boldsymbol{X}^{(m)}(n), \boldsymbol{X}^{(m)}(n))\mathbf{1}}. \tag{3.41}$$

If we have a signal with homoscedasticity, the equation further simplifies to

$$\rho^{(m)}(\lambda) = \frac{\mathbf{1}'\boldsymbol{R}^{(m)}(m\lambda)\mathbf{1}}{\mathbf{1}'\boldsymbol{R}^{(m)}(0)\mathbf{1}}. \tag{3.42}$$

These equations can, in practice, be difficult to compute in an explicit, scalar form. Nonetheless, (3.41) can be used with any evenly sampled signal. If, furthermore, the signal is itself stationary and homoscedastic, we can use (3.42) to obtain its coarse-grained autocorrelation function. If we go even further, if we only know the autocorrelation values for $\lambda = 1, 2, \dots, m-1$, we can obtain the coarse-grained autocorrelation value without even knowing the underlying functions. It is sufficient for the signal to be Gaussian to use the results in (3.41) and (3.42) to obtain the pattern probability distribution (3.5), and thus, the MPE (3.6).

3.3.6 ARMA Models Revisited

At this point, we can return to our results in section 3.3.3 and 3.3.4, to use the MA(1) and AR(1) models as an example of an explicit computation for (3.42). In the case of MA(1), the coarse-grained autocorrelation is,

$$
\begin{aligned}
\rho^{(m)}_{MA(1)}(\lambda) &= \frac{\mathbf{1}'\boldsymbol{R}^{(m)}(m\lambda)\mathbf{1}}{\mathbf{1}'\boldsymbol{R}^{(m)}(0)\mathbf{1}}, \qquad \text{for } |\lambda| = 1 \\
&= \frac{\frac{\theta}{1+\theta^2}}{m + 2(m-1)\frac{\theta}{1+\theta^2}}, \qquad \text{for } |\lambda| = 1 \\
&= \frac{\theta}{m(1+\theta^2) + 2(m-1)\theta}, \qquad \text{for } |\lambda| = 1 \\
\therefore \rho^{(m)}_{MA(1)}(\lambda) &= \begin{cases} 1, & \text{if } \lambda = 0 \\ \frac{\theta}{m(1+\theta^2)+2(m-1)\theta}, & \text{if } |\lambda| = 1 \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{3.43}
$$

which is the same result as obtained in (3.34). For the AR(1) process, we have,

$$
\begin{aligned}
\rho^{(m)}_{AR(1)}(\lambda) &= \frac{\mathbf{1}'\boldsymbol{R}^{(m)}(m\lambda)\mathbf{1}}{\mathbf{1}'\boldsymbol{R}^{(m)}(0)\mathbf{1}} \\
&= \frac{\phi^{m(\lambda-1)}\left((1+\phi^m)\frac{\phi}{(1-\phi)^2}\left(1+(m-1)\phi^m - m\phi^{m-1}\right) + m\phi^m\left(\frac{1-\phi^m}{1-\phi}\right)\right)}{2m\left(\frac{1-\phi^m}{1-\phi}\right) - m - 2\frac{\phi}{(1-\phi)^2}\left(1+(m-1)\phi^m - m\phi^{m-1}\right)} \\
&= \phi^{m(\lambda-1)+1}\left[\frac{(1-\phi^m)^2}{m(1-\phi^2) - 2\phi(1-\phi^m))}\right],
\end{aligned} \tag{3.44}
$$

which is, again, equal to equation (3.24).

Now, we can use equation (3.42) to obtain the general coarse-grained autocorrelation for an arbitrary $ARMA(\tilde{p}, \tilde{q})$. Since the explicit derivation would be long and cumbersome, it would suffice to have a general form for the autocorrelation function of the original ARMA process, to obtain the coarse-grained version.

If we already know the ARMA parameters, it is enough to obtain the autocorrelation for $X_{ARMA(\tilde{p},\tilde{q})}(n)$. This can be accomplished solving the generalized Yule-Walker (YW) equations for the autocovariance function γ. If $\tilde{p} > \tilde{q}$,

$$
\begin{aligned}
\gamma_0 &= \phi_1\gamma_1 + \cdots + \phi_{\tilde{p}}\phi_{\tilde{p}} + \sigma^2 + \theta_1 E[X_n\epsilon_{n-1}] + \cdots + \theta_{\tilde{q}} E[X_n\epsilon_{n-\tilde{q}}] \\
\gamma_1 &= \phi_1\gamma_0 + \cdots + \phi_{\tilde{p}}\phi_{\tilde{p}-1} + 0 + \theta_1\sigma^2 + \cdots + \theta_{\tilde{q}} E[X_n\epsilon_{n-\tilde{q}+1}] \\
&\vdots \\
\gamma_{\tilde{q}} &= \phi_1\gamma_{\tilde{q}-1} + \cdots + \phi_{\tilde{p}}\phi_{\tilde{p}-\tilde{q}} + 0 + 0 + \cdots + \theta_{\tilde{q}}\sigma^2 \\
&\vdots \\
\gamma_{\tilde{p}} &= \phi_1\gamma_{\tilde{p}-1} + \cdots + \phi_{\tilde{p}}\phi_0 \\
&\vdots \\
\gamma_\lambda &= \phi_1\gamma_{\lambda-1} + \cdots + \phi_{\tilde{p}}\phi_{\lambda-\tilde{p}}.
\end{aligned} \tag{3.45}
$$

If $\tilde{p} < \tilde{q}$, the YW equations change sightly,

$$\begin{aligned}
\gamma_0 &= \phi_1\gamma_1 + \cdots + \phi_{\tilde{p}}\phi_{\tilde{p}} + \sigma^2 + \theta_1 E[X_n\epsilon_{n-1}] + \cdots + \theta_{\tilde{q}} E[X_n\epsilon_{n-\tilde{q}}] \\
\gamma_1 &= \phi_1\gamma_0 + \cdots + \phi_{\tilde{p}}\phi_{\tilde{p}-1} + 0 + \theta_1\sigma^2 + \cdots + \theta_{\tilde{q}} E[X_n\epsilon_{n-\tilde{q}+1}] \\
&\vdots \\
\gamma_{\tilde{p}} &= \phi_1\gamma_{\tilde{p}-1} + \cdots + \phi_{\tilde{p}}\phi_0 + 0 + 0 + \cdots + \theta_{\tilde{q}} E[X_n\epsilon_{n-\tilde{q}+\tilde{p}}] \\
&\vdots \\
\gamma_{\tilde{q}} &= \phi_1\gamma_{\tilde{q}-1} + \cdots + \phi_{\tilde{p}}\phi_{\tilde{q}-\tilde{p}} + 0 + 0 + \cdots + \theta_{\tilde{q}}\sigma^2 \\
&\vdots \\
\gamma_\lambda &= \phi_1\gamma_{\lambda-1} + \cdots + \phi_{\tilde{p}}\phi_{\lambda-\tilde{p}}.
\end{aligned} \tag{3.46}$$

All the autocovariance terms γ_λ must be divided by γ_0 to obtain the first $max(\tilde{p}, \tilde{q})$ autocorrelations values. Any autocorrelation $\lambda > max(\tilde{p}, \tilde{q})$ follows the autorregressive recursive relation, and can be computed *a posteriori*.

With the YW equation solution for $\rho(\lambda) = \gamma_\lambda/\gamma_0$, we can use equation (3.42) utilizing the pattern probability in equation (3.5) (for $d = 3$) to obtain the MPE in (3.6). We can observe some examples in Fig. 3.12 with the average MPE of 100 signals at $N = 5000$ with bias correction. The models shown correspond to:

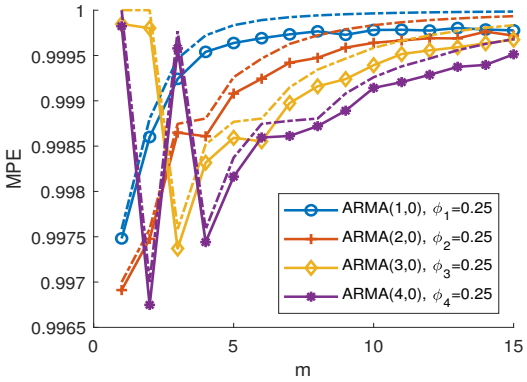
1. AR(p) with a single parameter $\phi_p = 0.25$, with increasing order p, with all lower order parameters set to zero. Fig. 3.12a.
2. ARMA(1,q), with fixed AR parameter $\phi_1 = 0.5$, and adding a new MA term $\theta_1 = \cdots = \theta_q = 0.1$ with increasing order. Fig. 3.12b.
3. MA(q) with a single parameter $\theta_q = 0.25$ with increasing q, also with lower order parameters equal to zero. Fig. 3.12c.
4. ARMA(p,1), with fixed MA parameter $\theta_1 = 0.5$, and adding a new AR term $\phi_1 = \cdots = \phi_p = 0.1$, also with increasing order. Fig. 3.12d.

Once again, albeit the behavior of each model presents more complications than the AR(1) and MA(1) models, it is evident that the simulations closely follow the MPE predictions. This fact further proves the utility of equation (3.42) to obtain the MPE of elaborate Gaussian models.

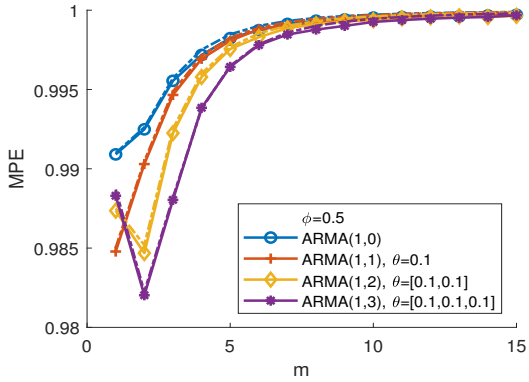
3.4 Closing Remarks

In this chapter we have explored the different factors that have an effect on the final MPE value measured from a signal. For this purpose, we studied the MPE results from different types of signals and models.

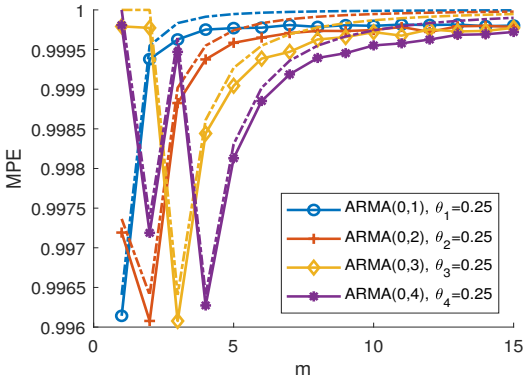
By dealing with deterministic signals, we found the pattern probability distribution to be set by the region where the curve's slope is positive or negative. The sampling rate also plays an important role, and almost invariably increases the MPE value.



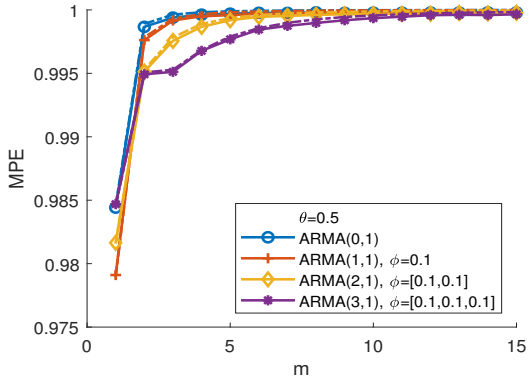
(a) AR(p) with increasing model order



(b) ARMA(1,q) with increasing θ parameters



(c) MA(q) with increasing model order



(d) ARMA(p,1) with increasing ϕ parameters

Figure 3.12: MPE curves vs. time scale for $ARMA(\tilde{p}, \tilde{q})$ models. Figures (a) and (c) correspond to AR and MA models increasing order, respectively, with only a single parameter (of the highest order). Figure (b) and (d) are models with one fixed AR or MA parameter. The particular values used are shown in their respective plots. Dotted lines represent the theoretical results, and the solid lines show the results from simulations.

With a high sampling rate, the MPE slowly converges with the MPE of a theoretical continuous curve. The addition of white noise was also considered. We found noise's effect on MPE depends heavily on the relationship between the curve's slope and the noise amplitude. If the slope is sufficiently high, the noise has no visible effect on MPE. For regions with low slopes, in contrast, the noise dominates. Nonetheless, if the sampling rate of a noisy signal is too high, we will obtain MPE values characteristic of white noise, regardless of the slope.

We also studied the expected MPE for commonly-used correlated Gaussian processes. By means of pattern symmetries for these signals, we are able to compute the MPE as a function of the signal's autocorrelation function. Since the coarse-graining procedure is a linear combination of the signal's elements, the resulting coarse-grained signal is also Gaussian. Therefore, it is sufficient to know the coarse-grained autocorrelation function to obtain the MPE for any scale.

By exploring white Gaussian noise and fractional Gaussian noise, we conclude that the MPE is invariant to time scale for these processes. First order Autoregressive and Moving Average processes have an elaborate, but ultimately closed, expression for the pattern probability distribution, which depends only on the models' parameters and the time scale.

In this chapter we also proposed a general expression for the coarse-grained autocorrelation for an arbitrary signal, by means of matrix quadratic forms. This allows us to compute the theoretical MPE of a signal without knowing the coarse-grained autocorrelation function explicitly. We presented some examples of ARMA models with an arbitrary number of parameters, and tested the theoretical results against simulations, with satisfactory results.

This analysis pretends to be an in-depth exploration on the multiple factors that influence the MPE of an arbitrary signal. The signal's slope, sampling rate, and autocorrelation functions prove to be paramount in the expected MPE value. The research of the interactions between these factors, and the study of arbitrary dimensions, will be subject of future work.

Chapter Summary

- The MPE of deterministic, noiseless continuous signals depends solely on the proportion of time the curve has positive or negative slope.
- For sampled deterministic signals, the sampling rate also has the effect of increasing the theoretical MPE. When the sampling frequency increases, the measured MPE slowly converges with the theoretical continuous case.
- The addition of uncorrelated noise to deterministic signals adds a new factor to the MPE measurement. MPE sensitivity to noise depends heavily on the relationship between the noise's amplitude and the curve's slope. Near the maximum and minimum points, random patterns appear, and the pattern distribution is uniform. In zones with high enough slope, the presence of noise presents no modification to the MPE.
- A high sampling rate enhances the effect of noise over deterministic signals. If the sampling rate is high enough, the noise dominates, regardless of the slope.
- Correlated Gaussian processes have a particular structure which can be exploited to obtain the theoretical expected MPE, reducing the degrees of freedom of the system. We focus our study to dimension $d = 3$, since $d = 2$ yields trivial results, $d = 4$ yields to complex probability distributions, and $d \geq 5$ has no closed form.
- For Gaussian processes, pattern probability distributions (and MPE) depend explicitly on the signal's autocorrelation function. Therefore, if we obtain the autocorrelation function for coarse-grained Gaussian signals, we can obtain a closed expression for their MPE as a function of the models' parameters and time scale.
- White Gaussian noise and fractional Gaussian noise, having the property of

self-similarity, are invariant to the coarse-graining process and time scale. The signal only presents a downward trend, explained by the bias in Chapter 2.

- We obtained the expressions for the coarse-grained first-order autoregressive and moving average processes, validated by simulations.
- A general coarse-grained autocorrelation function is presented by means of matrix quadratic forms. This formulation allows us to compute coarse-grained autocorrelations of elaborate Gaussian models without a closed form. Several examples are presented for autoregressive and moving average models of arbitrary order.

Chapter 4

Composite MPE Refinements

Y si sospechamos lo recayente de nuestro estado,
¿cómo nos rehabilitaremos?

- Julio Cortázar, *Me caigo y me levanto*

4.1 Introduction

MPE and other entropy methods have been subjected to a variety of different refinements in order to increase the precision of the resulting entropy estimation. Some examples have been briefly mentioned by us in Chapter 1, with the composite approach to coarse-graining — which was discussed in general terms in Chapter 1, section 1.4.3— being a natural progression of the original MPE algorithm. Both composite MPE (cMPE) and refined composite MPE (rcMPE) [57] aim to measure the maximum number of possible patterns within the original signal without modifying the underlying idea of the MPE approach. Both of these methods have been experimentally proven to yield better results by reducing the variance of the MPE estimator.

Therefore, in this Chapter we will further expand on the MPE statistical theory by including the cMPE and rcMPE algorithms. We will outline and discuss the improvements they offer over the classical MPE approach, as well as their drawbacks and possible shortcomings. Moreover, we present an alternative to the classic composite coarse-graining approach —which is known as "downsampling"— that further improves these refined methods. Finally, we compare all the previously discussed methods experimentally to evaluate their precision and recommend the most appropriate algorithm for the measurement of ordinal entropy in time series.

4.2 Composite Coarse-Graining Techniques

4.2.1 Composite Coarse-Graining Procedure

Composite coarse-graining stems from the notion that, for a given time series $\boldsymbol{x}$ and a set time scale m, we can build an m different coarse signals if we change the starting element for the coarse-graining procedure. Up to this moment, classical MPE assumes that the starting point is equal to the first element of the time series (x_1). At most, we can have an m number of difference signals for the same time scale that are similar to one another yet contain slightly different information. We apply the general procedure to build all the possible coarse signals for m,

$$x_{k,j}^{(m)} = \frac{1}{m} \sum_{i=m(j-1)+k}^{jm+(k-1)} x_i, \tag{4.1}$$

with $k = 1, \ldots, m$ for the starting element. Applying the procedure in (4.1) gives us coarse signals $\boldsymbol{x}_1^{(m)}, \ldots, \boldsymbol{x}_m^{(m)}$ for any given m.

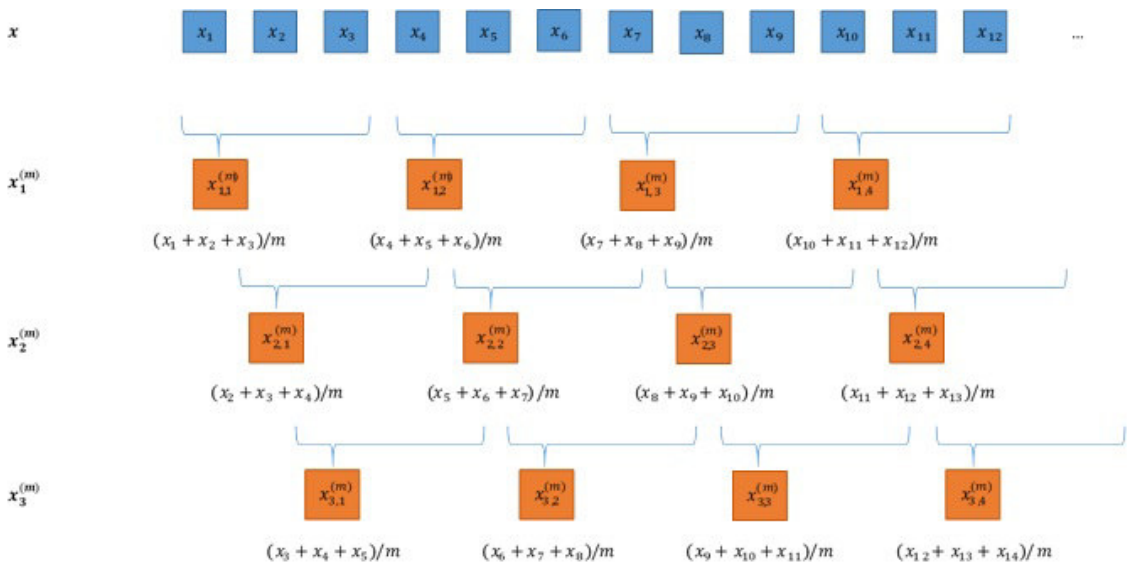
This refinement allows us to access previously unaccounted ordinal patterns in the series, thus increasing their number for the purposes of building the empirical pattern probability distribution. Therefore, utilizing composite coarse signals allows us to partially overcome the length constraints imposed by the multiscaling process..

We can make some comments regarding procedure (4.1). Since we are working with ordinal patterns, there is no need to perform the averaging by $1/m$ for each coarse signal element; this implies that, if $x_{k=1,j}^{(m)} < x_{k=1,j+1}^{(m)}$, then $mx_{k=1,j}^{(m)} < mx_{k=1,j+1}^{(m)}$. Thus, in stark contrast to the cardinal entropy techniques, we do not need to take the average of each segment to preserve the patterns. The sum of the elements is enough.

But more importantly, we will discuss now one of the main shortcomings present in this approach that is not mentioned in existing literature. If we compare the same elements from different coarse signals at a given m and revisit the definition in (4.1), we observe that the elements share information with another and that segments from different coarse signals overlap. For example, a closer look at Fig. **??** reveals that the first elements of the coarse signals $\boldsymbol{x}_1^{(m)}$ and $\boldsymbol{x}_2^{(m)}$ are

$$\begin{aligned} x_{k=1,1}^{(m=3)} &= \tfrac{1}{m}(x_1 + x_2 + x_3) \\ x_{k=2,1}^{(m=3)} &= \tfrac{1}{m}(x_2 + x_3 + x_4). \end{aligned}$$

Since elements x_2 and x_3 appear in both signals, all the information that we could measure from coarse signals $\boldsymbol{x}_1^{(m)}$ and $\boldsymbol{x}_2^{(m)}$ will have some level of redundancy. This will become more evident as we increase the value of m, and consequently, the number of shared elements increases.



This redundancy is bound to create cross-correlation between coarse signals, even if the original signal is uncorrelated noise. This effect will influence the overall MPE estimators that rely on composite signals, possibly resulting in an increased variance due to this redundant information. For the remainder of this chapter, we will refer to this effect as an artifact cross-correlation: the presence of correlation between coarse signals originating from shared elements and not from the inherent dynamics of the signals in question.

In the following subsections we will outline the most common composite methods: composite MPE (cMPE) [39] and the refined composite MPE (rcMPE) [57]. The characterization of the explicit statistical properties of these methods, including their moments, is beyond the scope of this work. Nonetheless, we will provide some guidelines regarding their performance over classical MPE.

4.2.2 Composite MPE

Following the composite coarse-graining procedure in equation (4.1) [39] makes it possible to achieve better precision by averaging the MPE result for all the composite signals with the same time scale. Even though this approach was originally named "improved multiscale permutation entropy" [39], we will refer to this procedure as composite MPE due to its shared similarities with composite multiscale entropy —as proposed by Wu et al. [43] using SampEn— in its mathematical approach.

Given the original time series $\boldsymbol{x}$ and embedding dimension d, we compute the classical MPE on each of the possible composite coarse signals $\boldsymbol{x}_1^{(m)}, \ldots, \boldsymbol{x}_m^{(m)}$ for each time scale m to obtain the cMPE. The cMPE value is the average of each of the resulting MPE measurements of all m coarse signals:

$$H_c(\widehat{\boldsymbol{p}}^{(m)}) = \frac{1}{m}\sum_{k=1}^{m} H_k(\widehat{\boldsymbol{p}}^{(m)}), \tag{4.2}$$

where $H_k(\widehat{\boldsymbol{p}}^{(m)})$ are the k possible entropy values for scale m. The approach of this method relies on reducing the variance by taking the average of multiple MPE measurements.

If we suppose all $\widehat{H}(\boldsymbol{x}_k^{(m)}, d)$ are independent for $k = 1, \ldots, \tau$, we expect to obtain the traditional moments for the mean of $\widehat{H}$, namely

$$E[H_c(\widehat{\boldsymbol{p}}^{(m)})] = \frac{1}{m}\sum_{k=1}^{m} E[H(\widehat{\boldsymbol{p}}_k^{(m)})] \approx H(\boldsymbol{p}^{(m)}) - \frac{1}{2}(d! - 1)\left(\frac{m}{N}\right). \tag{4.3}$$

We assume all the values of $H(\widehat{\boldsymbol{p}}_k^{(m)})$ to have a positive correlation, since the coarse signals share almost the same data points

$$cov(H(\widehat{\boldsymbol{p}}_{k_1}^{(m)}), H(\widehat{\boldsymbol{p}}_{k_2}^{(m)})) > 0, \qquad \forall k. \tag{4.4}$$

Therefore, the general expression of the cMPE variance can be written as

$$\begin{aligned} var\left(H_c(\widehat{\boldsymbol{p}}^{(m)})\right) &= \frac{1}{m^2} var(\sum_{k=1}^{m} H(\widehat{\boldsymbol{p}}_k^{(m)})) \\ &= \frac{1}{m^2}\sum_{k=1}^{m} var(H(\widehat{\boldsymbol{p}}_k^{(m)})) + \frac{1}{m^2}\sum\sum_{i\neq j}^{m} cov(H(\widehat{\boldsymbol{p}}_{k_1}^{(m)}), H(\widehat{\boldsymbol{p}}_{k_2}^{(m)})) \\ &= \frac{1}{m} var(H(\widehat{\boldsymbol{p}}_k^{(m)})) + \frac{1}{m^2}\sum\sum_{i\neq j}^{m} cov(H(\widehat{\boldsymbol{p}}_{k_1}^{(m)}), H(\widehat{\boldsymbol{p}}_{k_2}^{(m)})), \qquad \forall k \\ &\geq (\tfrac{1}{N})\boldsymbol{l}^{(m)\prime}\boldsymbol{\Sigma}_p^{(m)}\boldsymbol{l}^{(m)} \;+\; (\tfrac{1}{N})(\tfrac{m}{N})\left(\mathbf{1}'\boldsymbol{l}^{(m)} + d!H(\boldsymbol{p}^{(m)}) + \tfrac{1}{2}(d! - 1)\right), \end{aligned} \tag{4.5}$$

where $k_1 = 1, \ldots, m$ and $k_2 = 1, \ldots, m$. The measure of equality should be reached when the $H_k(\widehat{\boldsymbol{p}}^{(m)})$ are not correlated.

As we can see from equation (4.3), it should not come as a surprise that the expected value does not change with respect to classical MPE. Nonetheless, the variance in equation (4.5) is indeed reduced by a factor of $1/m$. This has a visible effect on the polynomial approximation, reducing the degree by one. Now the first element is constant with respect to m, and the second term is linear. Equation (4.5) provides a benchmark for the minimum variance the cMPE can obtain in the presence of uncorrelated coarse signals.

4.2.3 Refined Composite MPE

Originally proposed by Humeau-Heutier et al. [57], rcMPE approaches composite coarse signals through a different mechanism. Instead of using the average MPE, this method counts all the ordinal patterns contained in composite coarse signals

for a given scale m. This results in a single pattern probability estimation, which is used thereafter to obtain the entropy measurement. Computing rcMPE requires us to first take the average estimation for each pattern probability,

$$\widehat{\bar{\boldsymbol{p}}}^{(m)} = \begin{bmatrix} \widehat{\bar{p}}_1^{(m)} \\ \widehat{\bar{p}}_2^{(m)} \\ \vdots \\ \widehat{\bar{p}}_{d!}^{(m)} \end{bmatrix} = \frac{1}{m} \begin{bmatrix} \sum_{k=1}^{m} \widehat{p}_{k,1}^{(m)} \\ \sum_{k=1}^{m} \widehat{p}_{k,2}^{(m)} \\ \vdots \\ \sum_{k=1}^{m} \widehat{p}_{k,d!}^{(m)} \end{bmatrix}, \tag{4.6}$$

where the pattern probability estimator $\widehat{p}_{k,i}^{(m)}$ is obtained using equation (2.2) for each composite coarse signal $k = 1, \ldots, m$. At this point, it is enough to make a single MPE computation over this pattern probability,

$$H_{rc}(\widehat{\boldsymbol{p}}^{(m)}) = -\sum_{i=1}^{d!} \widehat{\bar{p}}_i^{(m)} \ln \widehat{\bar{p}}_i^{(m)}, \tag{4.7}$$

following the same procedure as the original MPE definition (2.4), using $\widehat{\bar{\boldsymbol{p}}}^{(m)}$ instead of $\widehat{\boldsymbol{p}}^{(m)}$.

The explicit representation of the first two moments $H_c(\widehat{\boldsymbol{p}}^{(m)})$ require further explanation. First, we modify the original multinomial pattern count expression from equations 2.9 and 2.10 (from Chapter 2) as follows,

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{d!} \end{bmatrix} = \begin{bmatrix} n_m p_1 + \Delta Y_1 \\ \vdots \\ n_m p_{d!} + \Delta Y_{d!} \end{bmatrix} = n_m \boldsymbol{p} + \Delta \boldsymbol{Y}, \qquad \sim Mu(n_m, \boldsymbol{p})$$

$$\widehat{\boldsymbol{p}} = \tfrac{1}{n_m} \boldsymbol{Y} = \boldsymbol{p} + \tfrac{1}{n_m} \Delta \boldsymbol{Y},$$

where the random variable Y represents the counts for each possible ordinal pattern and $n_m = N/m$. Similarly, we can define the estimated probability vectors $\widehat{\boldsymbol{p}}_k$ for $k = 1, \ldots, \tau$ as follows,

$$\boldsymbol{Y}_k = \begin{bmatrix} Y_{k,1} \\ Y_{k,2} \\ \vdots \\ Y_{k,d!} \end{bmatrix} = \begin{bmatrix} n_m p_1 + \Delta Y_{k,1} \\ n_m p_2 + \Delta Y_{k,2} \\ \vdots \\ n_m p_{d!} + \Delta Y_{k,d!} \end{bmatrix} = n_m \boldsymbol{p} + \Delta \boldsymbol{Y}_k, \qquad \sim Mu(n_m, \boldsymbol{p})$$

$$\widehat{\boldsymbol{p}_k} = \tfrac{1}{n_m} \boldsymbol{Y}_k = \boldsymbol{p} + \tfrac{1}{n_m} \Delta \boldsymbol{Y}_k. \tag{4.8}$$

Before making the entropy computation, we obtain the average of all the pattern counts along the composite signals k:

$$\overline{\boldsymbol{Y}} = \sum_{k=1}^{m} \boldsymbol{Y}_k = \frac{1}{m}\sum_{k=1}^{m} \tfrac{N}{m}\boldsymbol{p} + \frac{1}{m}\sum_{k=1}^{m} \Delta\boldsymbol{Y}_k$$
$$\overline{\boldsymbol{Y}} = \frac{N}{m}\boldsymbol{p} + \frac{1}{m}\sum_{k=1}^{m} \Delta\boldsymbol{Y}_k$$
$$m\overline{\boldsymbol{Y}} = N\boldsymbol{p} + \sum_{k=1}^{m} \Delta\boldsymbol{Y}_k. \tag{4.9}$$

We note that $m\overline{\boldsymbol{Y}}$ is the vector containing the sum of the patterns contained in all composite signals. We proceed to define that

$$\boldsymbol{Z} = m\overline{\boldsymbol{Y}} \tag{4.10}$$
$$\Delta\boldsymbol{Z} = \sum_{k=1}^{m} \Delta\boldsymbol{Y}_k, \tag{4.11}$$

where vector $\Delta\boldsymbol{Z}$ contains the sum of all the errors for each pattern. The new variable $\boldsymbol{Z}$ is defined as

$$\boldsymbol{Z} = N\boldsymbol{p} + \Delta\boldsymbol{Z} \tag{4.12}$$
$$\widehat{\boldsymbol{p}} = \boldsymbol{p} + \tfrac{1}{N}\Delta\boldsymbol{Z}. \tag{4.13}$$

Rewriting the refined composite technique in such a way is revealing: equation (4.13) is identical to the probability estimation in (4.6) and this formulation shows explicitly that the estimation is now independent of the time scale value in m. However, the existing artifact cross-correlation effect indicates that all the $\Delta\boldsymbol{Y}_k$ in (4.11) are not uncorrelated; therefore, $\boldsymbol{Z} \sim Mu(N, \boldsymbol{p})$ is not satisfied in a general sense.

If we use (4.12) in our classical MPE Taylor series approximation (2.19) from section 2.3.2, we obtain

$$H(\widehat{\boldsymbol{p}}) \approx H(\boldsymbol{p}) - \frac{1}{N}\left(\mathbf{1} + \boldsymbol{l}\right)' \boldsymbol{\Delta Z} - \frac{1}{2}\left(\frac{1}{N}\right)^2 \left(\boldsymbol{p}^{\circ -1}\right)' \boldsymbol{\Delta Z}^{\circ 2}. \tag{4.14}$$

Obtaining the expected value of (4.14) is enough to also obtain the mean rcMPE. Since $\boldsymbol{Z}$ is not strictly multinomial, the moments of $\boldsymbol{\Delta Z}$ do not correspond to the results in Appendix A. However, we assume all $\Delta\boldsymbol{Y}_k$ have a positive correlation, since we expect all $\Delta\boldsymbol{Y}_k$ measurements to have similar results. Therefore,

$$C_\gamma = cov(\Delta\boldsymbol{Y}_{k_1}, \Delta\boldsymbol{Y}_{k_2}) > 0,$$

so

$$\begin{aligned} E[\boldsymbol{\Delta Z}^{\circ 2}] &= N(\boldsymbol{p} - \boldsymbol{p}^{\circ 2}) + C_\gamma \\ &\geq N(\boldsymbol{p} - \boldsymbol{p}^{\circ 2}). \end{aligned} \tag{4.15}$$

where C_γ represents added value due to artifact cross-correlations. Therefore, we can clearly write the expected value as,

$$E[H_{rc}(\widehat{\bar{\boldsymbol{p}}}^{(m)})] \leq H(\boldsymbol{p}^{(m)}) - \frac{1}{2}(d! - 1)\left(\frac{1}{N}\right), \tag{4.16}$$

and its variance as,

$$var(H_{rc}(\widehat{\boldsymbol{p}}^{(m)})) \geq (\tfrac{1}{N})\boldsymbol{l}^{(m)\prime}\boldsymbol{\Sigma}_p^{(m)}\boldsymbol{l}^{(m)} \; + \; (\tfrac{1}{N})^2\left(\mathbf{1}'\boldsymbol{l}^{(m)} + d!H(\boldsymbol{p}^{(m)}) + \tfrac{1}{2}(d! - 1)\right). \tag{4.17}$$

Despite the fact that there is still bias in the rcMPE expected value and its variance, having eliminated the m dependence indicates that we now have a constant bias which relies solely on the signal's original length and the embedded dimension. Thanks to this refinement, it is possible for us now to explore higher m values without worrying about loss of precision due to signal length reduction.

The artifact cross-correlation effect does not allow the rcMPE moments (4.16) and (4.17) to achieve equality, since the definition of the composite coarse-graining procedure (4.1) itself imposes some level information redundancy. Therefore, we should look for alternatives to the classical coarse-graining to further improve the precision of rcMPE (and cMPE).

4.3 Composite Downsampling Techniques

4.3.1 Composite Downsampling Procedure

Instead of fully characterizing this artifact cross-correlation effect, a complex and time-consuming mathematical endeavor, we will present an alternative that is exempt of this redundancy from the beginning: composite downsampling. Downsampling in the context of PE is not a new concept. Most modern literature [49] define the pattern probability estimation for PE as

$$\widehat{p}_i = \frac{\#\{n|\ n < N/\tau - (d-1),\ [x_n, \ldots, x_{n+d-1}]\ has\ pattern\ i\}}{N/\tau - (d-1)}. \tag{4.18}$$

The downsampling parameter $\tau \in \mathbb{N}^+$ is reintroduced in this definition of the pattern probability estimator. From the cardinal point of view, the coarse-graining procedure represents a better smoothing filter than a simple downsampling procedure,

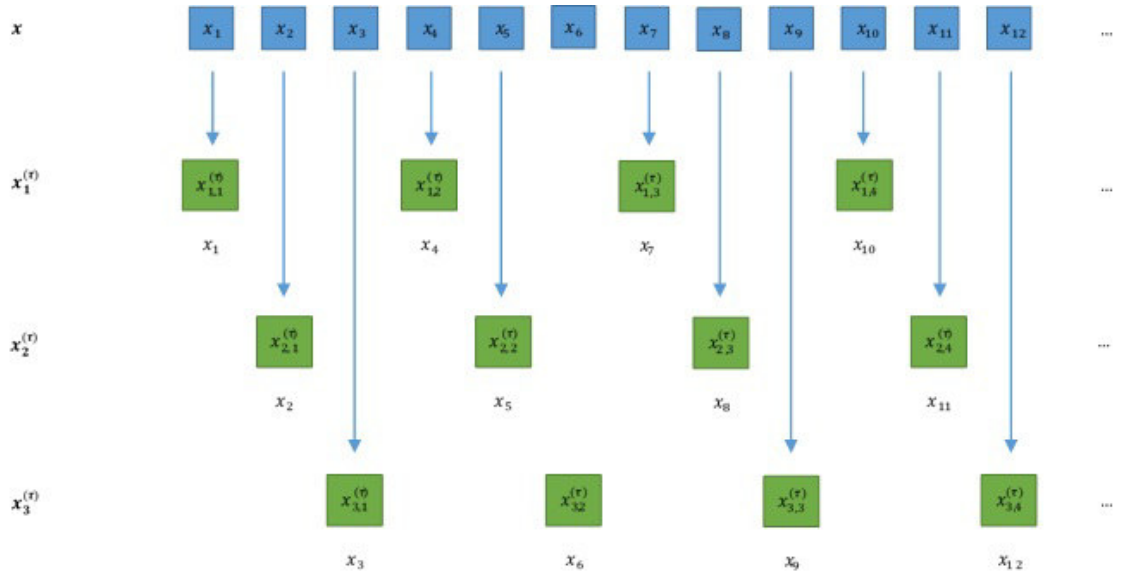


Figure 4.1: Schematic representation of the composite downsampling procedure at $\tau = 3$: we downsample the original signals by taking data points that are τ spaces apart; if we shift the initial position, we can build τ signals. The present downsampling signals share no mutual data points between them.

since the former incorporates more signal information and the latter implies a loss of resolution as a trade-off. Nonetheless, both procedures behave similarly for the purpose of ordinal patterns.

If we introduce the composite approach to the classical downsampling procedure, we can define composite downsampled signals as follows for $k = 1, \ldots, \tau$:

$$x_{k,j}^{(\tau)} = x_{k+\tau(j-1)}. \tag{4.19}$$

Changing the starting element k allows us to obtain a τ number of downsampled signals from the original signal $\boldsymbol{x}$. This implies no information loss, since all the elements in $\boldsymbol{x}$ are still present in the composite signals $\boldsymbol{x}_k^{(\tau)}$ (see Fig. 4.1). Additionally, since we can also appreciate that the resulting signals have no elements in common, we know that the artifact cross-relation effect will not be present. This is justified if we regard the process (4.19) as a systematic sampling, where each downsampled signal is a sample of the "population" signal $\boldsymbol{x}$, with the constraint that no samples share mutual elements.

It is necessary for us to first revisit some concepts from Chapter 2 before going into detail about the effects of composite approaches on MPE. For a signal $\boldsymbol{x}$ of length N, the MPE estimator is expected to have the following moments (see Sections 2.3.3 and 2.3.4):

$$E[H(\widehat{\boldsymbol{p}}^{(m)})] \approx H(\boldsymbol{p}^{(m)}) - \frac{1}{2}(d!-1)\left(\frac{m}{N}\right)$$
$$var\left(H(\widehat{\boldsymbol{p}}^{(m)})\right) \approx (\tfrac{m}{N})\boldsymbol{l}'^{(m)}\boldsymbol{\Sigma}_p^{(m)}\boldsymbol{l}^{(m)} + (\tfrac{m}{N})^2\left(\mathbf{1}'\boldsymbol{l}^{(m)} + d!H(\boldsymbol{p}^{(m)}) + \tfrac{1}{2}(d!-1)\right).$$

Both the expected value and the variance depend on time scale m, both implicitly (by means of $\widehat{\boldsymbol{p}}^{(m)}$)) and explicitly. It is also worth mentioning that these moments are heavily dependent on signal length N and the embedding dimension d.

Since the downsampling procedure also reduces the signal length in a similar fashion, it stands to reason that applying a classical downsampling procedure with any given τ value will present the moments as follows:

$$E[H(\widehat{\boldsymbol{p}}^{(\tau)})] \approx H(\boldsymbol{p}^{(\tau)}) - \frac{1}{2}(d!-1)\left(\frac{\tau}{N}\right) \tag{4.20}$$
$$var\left(H(\widehat{\boldsymbol{p}}^{(\tau)})\right) \approx (\tfrac{\tau}{N})\boldsymbol{l}'^{(\tau)}\boldsymbol{\Sigma}_p^{(\tau)}\boldsymbol{l}^{(\tau)} + (\tfrac{\tau}{N})^2\left(\mathbf{1}'\boldsymbol{l}^{(\tau)} + d!H(\boldsymbol{p}^{(\tau)}) + \tfrac{1}{2}(d!-1)\right). \tag{4.21}$$

Given that the only explicit change is switching from τ instead of m, we can capitalize on the MPE theory and apply it to the downsampling case.

Certainly, we cannot expect to obtain the same pattern probabilities from these two procedures due to the fact that, in general, $\boldsymbol{p}^{(\tau)} \neq \boldsymbol{p}^{(m)}$. Still, we expect to get the exact same pattern distribution (i.e. uniform) for both procedures in some specific circumstances, such as in the presence of uncorrelated noise.

Having said that, we now present the new entropy measurements: composite downsampling permutation entropy (cDPE) and refined composite downsampling permutation entropy (rcDPE).

4.3.2 Composite Downsampling Permutation Entropy

Similarly to the case of cMPE, utilizing the composite downsampling procedure (4.19) allows us to define cDPE as:

$$H_c(\widehat{\boldsymbol{p}}^{(\tau)}) = \frac{1}{\tau}\sum_{k=1}^{\tau} H_k(\widehat{\boldsymbol{p}}^{(\tau)}). \tag{4.22}$$

It is by means of the downsampling procedure (4.19) that we can deduce that all $\widehat{H}(\boldsymbol{x}_k^{(\tau)}, d)$ are independent for $k = 1, \ldots, \tau$. Therefore, we find the traditional moments for the mean of $\widehat{H}$, namely

$$E[H_c(\widehat{\boldsymbol{p}}^{(\tau)})] = \frac{1}{\tau}\sum_{k=1}^{\tau} E[H(\widehat{\boldsymbol{p}}_k^{(\tau)})] \approx H(\boldsymbol{p}^{(\tau)}) - \frac{1}{2}(d!-1)\left(\frac{\tau}{N}\right), \tag{4.23}$$

and

$$
\begin{aligned}
var(H_c\left(\widehat{\boldsymbol{p}}^{(\tau)}\right)) &= \frac{1}{\tau^2} var(\sum_{k=1}^{\tau} H(\widehat{\boldsymbol{p}}_k^{(\tau)})) \\
&= \frac{1}{\tau^2} \sum_{k=1}^{\tau} var(H(\widehat{\boldsymbol{p}}_k^{(\tau)})) \\
&= \frac{1}{\tau} var(H(\widehat{\boldsymbol{p}}_k^{(\tau)})), \qquad \forall k \\
&\approx (\tfrac{1}{N}) \boldsymbol{l}^{(\tau)'} \boldsymbol{\Sigma}_p^{(\tau)} \boldsymbol{l}^{(\tau)} \; + \; (\tfrac{1}{N})(\tfrac{\tau}{N}) \left(\mathbf{1}' \boldsymbol{l}^{(\tau)} + d! H(\boldsymbol{p}^{(\tau)}) + \tfrac{1}{2}(d! - 1) \right).
\end{aligned}
\tag{4.24}
$$

Once again, the cDPE expected value (4.23) does not change with respect to classical MPE, and we still have a downward bias whose only dependencies are the embedding dimension (d), signal length (N), and the downsampling parameter (τ). Conversely, the cDPE variance (4.24) is reduced by a factor of $1/\tau$ in this case, and we expect it to be approximately close to (4.24) due to the lack of an artifact cross-correlation by virtue of the definition shown in (4.19). This implies an improvement over the cMPE variance (4.5), where the cross-correlations will invariable reduce the estimator's precision.

4.3.3 Refined Composite Downsampling PE

By following the same reasoning as with rcMPE, rcDPE takes the average pattern probability distribution from all the downsampled signals for the parameter τ. As with (4.6), we proceed to define the probability vector,

$$
\widehat{\widehat{\boldsymbol{p}}}^{(\tau)} = \begin{bmatrix} \widehat{\widehat{p}}_1^{(\tau)} \\ \widehat{\widehat{p}}_2^{(\tau)} \\ \vdots \\ \widehat{\widehat{p}}_{d!}^{(\tau)} \end{bmatrix} = \tfrac{1}{\tau} \begin{bmatrix} \sum_{k=1}^{\tau} \widehat{p}_{k,1}^{(\tau)} \\ \sum_{k=1}^{\tau} \widehat{p}_{k,2}^{(\tau)} \\ \vdots \\ \sum_{k=1}^{\tau} \widehat{p}_{k,d!}^{(\tau)} \end{bmatrix}. \tag{4.25}
$$

where the pattern probability estimator $\widehat{p}_{k,i}^{(\tau)}$ is obtained using equation (2.2) for each composite coarse signal $k = 1, \ldots, m$. At this point, it suffices to apply a single MPE computation to produce this pattern probability.

For the composite downsampling procedure, we present rcDPE by using the exact same approach as before,

$$
H_{drc}(\widehat{\widehat{\boldsymbol{p}}}^{(\tau)}) = -\sum_{i=1}^{d!} \widehat{\widehat{p}}_i^{(\tau)} \ln \widehat{\widehat{p}}_i^{(\tau)}, \tag{4.26}
$$

and following the same procedure as the original MPE definition (2.4), using $\widehat{\widehat{\boldsymbol{p}}}^{(\tau)}$ instead of $\widehat{\widehat{\boldsymbol{p}}}^{(m)}$.

Once again we can enunciate the explicit moments of rcDPE, for the downsampled signals display no artifact cross-correlation. By following the procedure outlined for rcMPE in section 4.2.3, we obtain the rcDPE expected value,

$$E[H_{drc}(\widehat{\widehat{\boldsymbol{p}}}^{(\tau)})] \approx H(\boldsymbol{p}^{(\tau)}) - \frac{1}{2}(d!-1)\left(\frac{1}{N}\right), \tag{4.27}$$

and the rcDPE variance,

$$var(H_{drc}(\widehat{\widehat{\boldsymbol{p}}}^{(\tau)})) \approx (\tfrac{1}{N})\boldsymbol{l}^{(\tau)\prime}\boldsymbol{\Sigma}_p^{(\tau)}\boldsymbol{l}^{(\tau)} \;+\; (\tfrac{1}{N})^2\left(\mathbf{1}'\boldsymbol{l}^{(\tau)} + d!H(\boldsymbol{p}^{(\tau)}) + \tfrac{1}{2}(d!-1)\right). \tag{4.28}$$

In contrast to rcMPE, we do not expect in this case for the expected value (4.27) and its variance (4.28) to raise above the aforementioned mathematical expressions, since the artifact cross-correlation is not present. We still have the advantage of taking the explicit τ parameter dependence out of the equation, which suggests a stable behavior across τ.

4.4 Results and Discussion

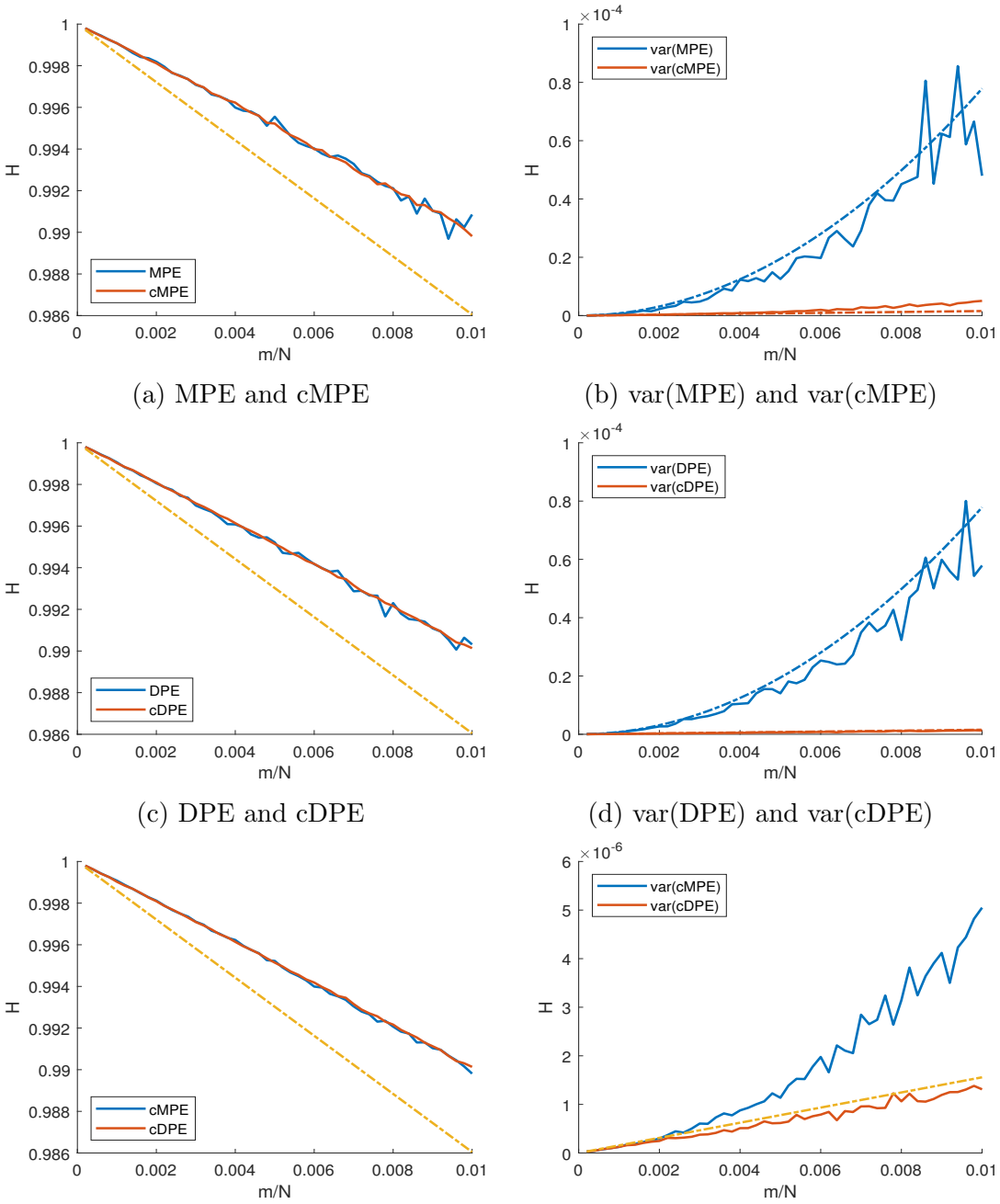
4.4.1 Results

Composite Methods

We know from [39] that cMPE is indeed more accurate than MPE. Since the composite coarse-grained signals are bound to present artifact autocorrelation, we expect the variance to be significantly more than the one predicted in (4.5).

Therefore, we will proceed to test these results against simulations in order to assess the precision of both cMPE and cDPE. For this purpose, we will apply both procedures on simulated white Gaussian noise, with 500 signals of length $N = 1000$ for $d = 3$ as our sample. We will also set $max(m/N) = max(\tau/N) = 0.01$, which is well within the length criterion defined in Chapter 2.5.

As we can see from Figure 4.2, cMPE and cDPE closely follow the path of classical MPE, with a reduced variance. Nonetheless, when we compare both composite entropy methods (Figure 4.2d), we can see that the cDPE variance follows the theoretical curve (dotted) while the cMPE variance consistently shows higher values. While cDPE does follow equation (4.24) and validates our assumptions, it follows that cMPE does not adhere to equation (4.5), since it requires the absence of artifact cross-correlations between coarse-grained signals. In other words, if we look for an entropy ordinal statistic with reduced variance, cDPE outperforms cMPE, specially when utilizing large time scales.



(a) MPE and cMPE

(b) var(MPE) and var(cMPE)

(c) DPE and cDPE

(d) var(DPE) and var(cDPE)

(e) cMPE and cDPE

(f) var(cMPE) and var(cDPE)

Figure 4.2: Composite vs. classical MPE measurements for white Gaussian noise, with dimension $d = 3$ and normalized time scale m/N. (a) Comparison of MPE and cMPE. (b) var(MPE) and var(cMPE). (c) DPE and cDPE. (b) var(DPE) and var(cDPE). (e) Comparison between composite methods: cMPE and cDPE. (b) var(cMPE) and var(cDPE). Solid lines are the product of 500 iterations of wGn signals with $N = 1000$ and $d = 3$. Dotted lines are the predicted values from equation (2.23), (2.31), and (4.24).

It is worth mentioning that, although cDPE follows the theoretical variance, the expected value for all PE follow a different path than the previously predicted linear bias. This effect did not manifest in previous chapters due to our experimental

setup, and deserves further comments. We will revisit this discrepancy in Section 4.4.2.

Refined Composite Methods

By following the same experimental setup, we now compare the differences in performance between the rcMPE and rcDPE algorithms. As we can see in Figure 4.3, the refined composite approach outperforms their composite entropy counterparts by reducing the estimator's variance even further. When we compare both composite procedures, we can observe again that rcDPE follows the theoretical variance (Fig. 4.3f) while rcMPE is considerably higher. This is explained again by the presence of artifact cross-correlation between coarse-grained signals. In terms of precision, the rcDPE is the best approach discussed in this work so far.

Regarding the expected value, we observe that rcDPE presents no scale-dependent bias, as predicted from equation (4.28). On the other hand, rcMPE does indeed show a decrease in time scale, albeit not as pronounced as in the case of cMPE (Fig. 4.3a). This effect, once again, can be attributed to the artifact cross-correlations from composite coarse-grained signals.

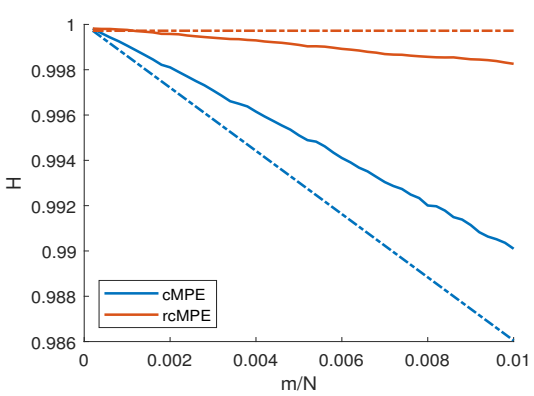
Even when the rcDPE bias is scale-independent, we can observe that the results from simulations are slightly higher than the predicted rcDPE line. This is the same effect as the discrepancy found between the theoretical linear bias and the simulation results present in MPE, DPE, cMPE and cDPE. Thus, this divergence is not a product of the different composite techniques and must be analyzed independently (see Section 4.4.2).

4.4.2 Discussion

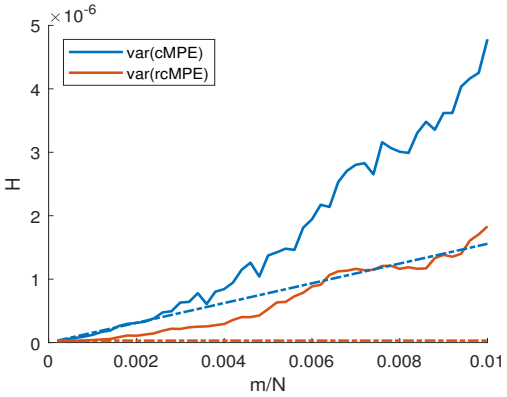
Expected Value Divergence

The first topic to discuss is the pervasive discrepancy between the MPE expected value and the simulation results, since these discrepancies are present regardless of the PE technique used. When we observe a difference between a composite coarse-graining and a composite downsampling method, we can be sure that said difference comes from the presence or absence of artifact cross-correlations, respectively. If the deviations appear on all instances, this implies that such artifact cross-correlations are not the main source. As we see in Figure 4.4a, this effect is not only present across all pattern dimensions d, but it is also more noticeable when d is high.

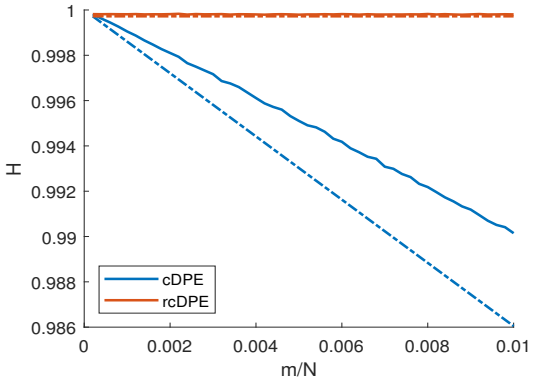
If cross-correlations are not the source of these differences, the next step is to revisit our assumptions from Chapter 2.3.1. Our assumptions therein mention that every ordinal pattern stemming from an uncorrelated random process is itself uncorrelated. If we simulate pattern counts directly from a multinomial variable instead of using an uncorrelated noise signal —under the same parameters ($N = 1000$, $d = 3$, and 500 signals)— we indeed recover predicted the linear bias that was mentioned in Chapter 2, as shown in Figure 4.4b. This fact implies that the pattern distribution is not constant across scales, as previously expected from white Gaussian noise in Section 3.3.2.



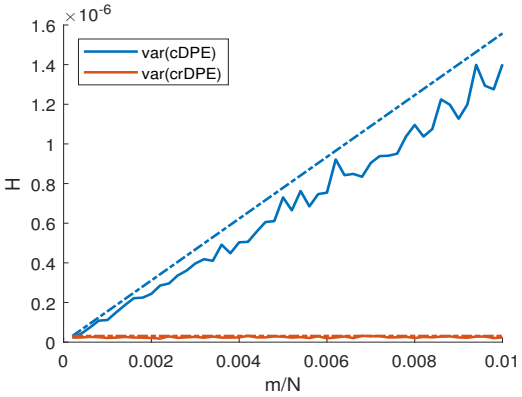
(a) cMPE and rcMPE



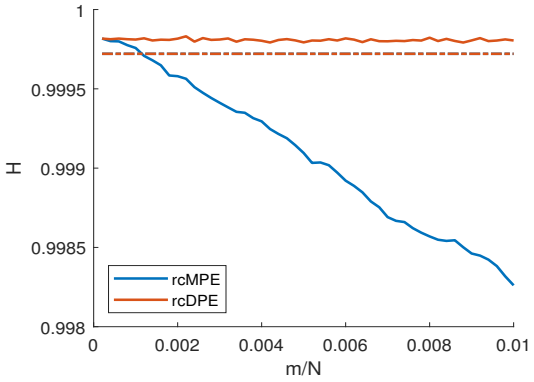
(b) var(cMPE) and var(rcMPE)



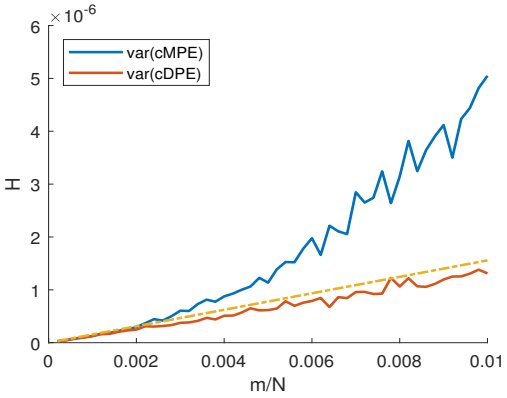
(c) cDPE and rcDPE



(d) var(cDPE) and var(rcDPE)

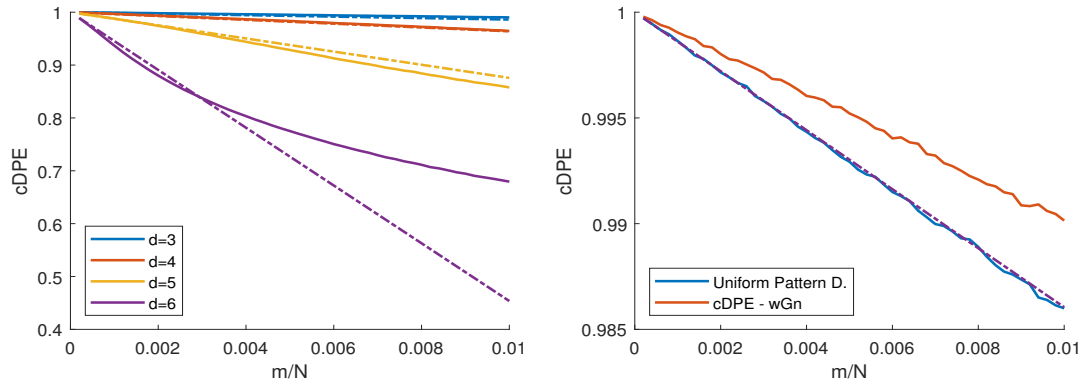


(e) rcMPE vs. rcDPE



(f) var(rcMPE) vs. var(rcDPE)

Figure 4.3: Composite vs. refined composite MPE measurements for white Gaussian noise, with dimension $d = 3$ and normalized time scale m/N. (a) Comparison of cMPE and rcMPE. (b) var(cMPE) and var(rcMPE). (c) cDPE and rcDPE. (b) var(cDPE) and var(rcDPE). (e) Comparison between composite methods: rcMPE and rcDPE. (b) var(rcMPE) and var(rcDPE). Solid lines are the product of 500 iterations of wGn signals with $N = 1000$ and $d = 3$. Dotted lines are the predicted values from equation (4.27), (4.5), and (4.28).



(a) (b)

Figure 4.4: (a) Entropy discrepancy between expected value and simulation results for cDPE for different d values. (b) Entropy discrepancy vs. simulated pattern counts from multinomial (equiprobable) distribution.

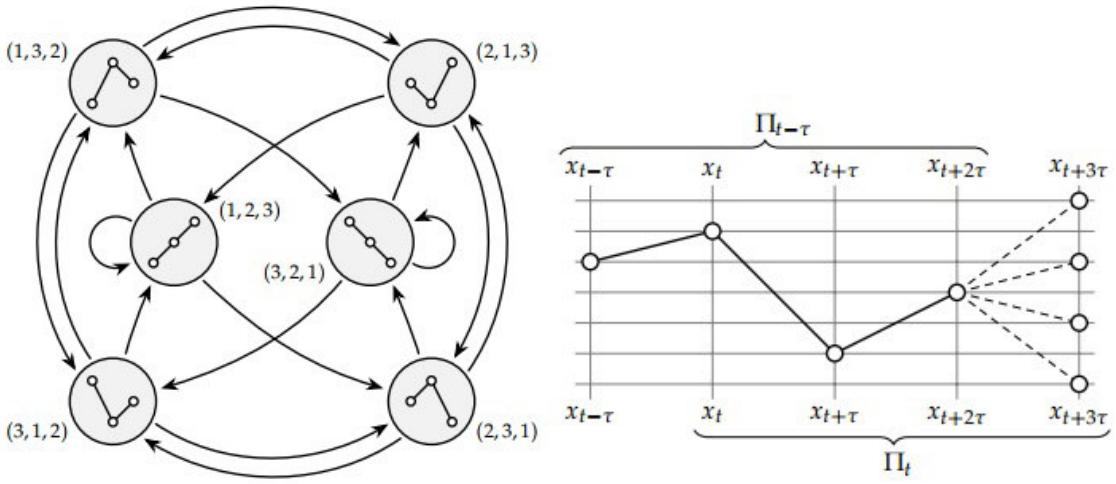
This apparent contradiction can be explained by further reexamining our assumptions from the pattern counts for PE —even before taking into account multiscaling or other refinements. Strictly speaking, when we assume a multinomial distribution, we accept that the patterns inside the signal are uncorrelated and identically distributed. However, we know this is not the case for a general signal, which changes its properties over time; according to [59] and [63], a more appropriate description of an ordinal process would be a first-order Markov chain. Even when we obtain our patterns from uncorrelated noise, the patterns themselves are not independent, since they share most of their elements. Considering these circumstances, we expect the pattern distribution, under these circumstances, to not remain invariant across scales, therefore producing a deviation from the expected linear downward trend. The characterization of this phenomenon is beyond the scope of the present work, but its exploration is nonetheless worthy of attention.

Composite Techniques and Precision

Regarding the precision of the different MPE refinements presented in this chapter, we observe rcDPE to present both the smallest variance across time scales and a desired scale-independent bias. The refined composite approach, in conjunction with a composite downsampling process instead of traditional coarse-graining, renders the problem of the artifact cross-correlations obsolete.

If we revisit the length constraints discussed in Section 2.6, since the time scale is no longer a problem, the only limiting factor is the signal's original length N. Table 4.1 shows the minimum length required for rcDPE analysis at different embedded dimensions, for a precision of $\alpha = 0.05$. For practical purposes, the maximum τ for rcDPE is such that $\frac{N}{\tau} > d$, in order to ensure that composite downsampled signals contain at least one pattern of dimension d; a signal this short is still not advisable.

Lastly, we should note that the experimental rcDPE line is horizontal with respect to time scale, hence proving that the rcDPE result is independent of the downsam-



(a) (b)

Figure 4.5: (a) Markov chain state diagram for an ordinal process of $d = 3$, where not all states are accessible in one step. (b) Two successive ordinal patterns for $d = 4$. Given an ordinal pattern, only d possible patterns from the state-space of size $d!$ are accessible. Figures from [63].

d	$\frac{\ln d!}{d!-1}$	min N
3	2.79	333
4	7.23	643
5	24.86	1,667
6	109.28	5,724
7	591.86	25,164

Table 4.1: Minimum length N for rcDPE at embedded dimension d and $\alpha = 0.05$. With these conditions in mind, coarse signal length is not dependent on τ (or m).

pling parameter τ. Albeit not truly an unbiased estimator, the bias present in our calculations is guaranteed to be stable. Furthermore, we can approximate this bias by knowing signal length (N) and dimension (d) values, regardless of the pattern distribution present.

4.5 Closing Remarks

We have explored and expanded on throughout this chapter the theory behind the composite refinements over classical MPE analysis. We also presented composite downsampling as an alternative to the composite coarse-graining procedure, with the intention of shedding the artifact cross-correlations between composite coarse signals at the same time scale.

We found that composite downsampling techniques vastly reduce the variance of the

cMPE and the rcMPE approaches found in literature. Specifically, downsampled rcMPE, on top of presenting the minimum variance, also showed an expected value that remained invariant with respect to the time scale. This will be particularly useful, since we will not face noticeable degradations when exploring large time scales. Therefore, for practical purposes, we recommend the use of this method. We also proposed an updated version of the length constraint criterion presented in Chapter 2 that is independent of scale.

Since the artifact cross-correlation from composite coarse-graining techniques is completely avoided by the use of the composite downsampling process, we did not characterize this phenomenon explicitly for practical purposes —this is, however, an interesting mathematical problem deserving further research.

During the exploration of MPE refinements, we unexpectedly came across a deviation from the MPE expected values and the actual MPE measurements over uncorrelated white Gaussian noise. Since this deviation is present on all techniques tested here, we conclude the artifact cross-correlations is not the source of this effect. Instead, we reevaluated our assumptions over a scale-independent pattern probability distribution from a white Gaussian noise process. Additionally, further literature research [59] [63] revealed that the ordinal pattern count follows a first-order Markov process. Although this effect merits further research, rcDPE makes a sufficient bias correction to justify continuing the use of the multinomial approach.

Chapter Summary

- The composite coarse-graining procedure consists on utilizing the classical coarse-graining algorithm at different starting points, with the intention of increasing the number of coarse signals for a given time scale.
- Although the composite coarse-graining procedure increases the precision of MPE estimation, it also introduces artifact cross-correlations between composite signals, which consequently increases the expected variance.
- To avoid artifact cross-correlations, we present a composite downsampling process by using the classical downsampling procedure in conjunction with the composite techniques.
- Composite MPE produces no deviation from the original MPE expected value. Nonetheless, the variance is reduced with respect to MPE.
- Composite downsampling PE (cDPE) still follows the predicted, biased expected value, but improves the variance beyond the cMPE by virtue of avoiding artifact cross-correlation effects.
- Refined composite MPE, on top of outperforming cMPE, also mitigates the scale-dependent bias.
- Refined composite downsampling PE (rcDPE), a method devised in this research project, presents the minimum variance value over all other refined MPE variants, again by virtue of completely avoiding artifact cross-correlations between composite signals.

- So far, rcDPE is the only ordinal entropy measurement whose expected value is explicitly independent from time scale for uncorrelated noise.
- An unexpected result arises when comparing white Gaussian noise entropy measurements, which diverge from the theoretical expected value. This effect challenges our assumptions of scale-independent pattern probability distributions for uncorrelated noise. Nonetheless, rcDPE is sufficiently stable across scales to mitigate these scale-dependent probabilities.

Chapter 5

Bioelectrical Signal Applications

Hagamos una cosa:
usted se rehabilita y yo la observo

- Julio Cortázar, *Me caigo y me levanto*

5.1 Introduction

We have so far developed and expanded on the statistical properties of multiscale permutation entropy, including the functional forms for its first two moments (Chapter 2). These new theoretical improvements were tested by us on common signal models and stochastic processes (Chapter 3), and our latest contribution has consisted in further developing MPE by exploring the method's refinements, including the proposal of an alternative to the widely used coarse-graining procedure (Chapter 4). The progress has done more than just leading to a greater, in-depth understanding of MPE algorithm and methods, since it also has contributed to the correction of the MPE bias while increasing the statistical precision of the method. As hinted in Chapter 1, we now have the necessary tools to apply these methods on real data sets. Therefore, we will now proceed to introduce the biomedical applications of MPE.

Ordinal pattern metrics have been useful in recent years to measure complexity in biological systems, particularly the ones related to electrical activity [33]. These types of signals are characterized by complex dynamics, even when on resting conditions [64]. Some noteworthy cases involve spontaneous brain activity presenting complex, non-random behavior [65] [66], and even pathological activity from epileptic seizures are characterized by an ordered sequence of events [67]. While most methods require further assumptions regarding the signal's deterministic and random features, PE measurements have the added advantage of being model-free and robust [33]. Finally, as we have previously discussed, PE techniques are fast to execute, an attractive feature for real-time applications without further pre-processing, such as contexts involving bioelectrical signals.

Placing electrodes in the skin's surface allows us to measure electric fields generated by the heart (electrocardiogram, ECG), the brain (electroencephalogram, EEG) or

the neuromuscular system (electromyography, EMG). In the latter case, electromyography has been widely used to gain fundamental knowledge on the recruitment process of the muscular functional unit —known as motor unit [68]—since Piper's first work using EMG signals [69].

The aforementioned studies have shown so far that EMG methods are well-suited to analyze behavior that involve muscle contraction. Beyond fundamental aspects, their application domains are diverse: there is the case of sports and ergonomics, whether performing isometric or dynamic exercises [70] [71]; in clinical and technological applications, such as rehabilitation, biofeedback, and myoelectric interfaces for the control of prosthetic devices or computer interaction [72]. Additionally, this technique also sheds light on motor learning [73] and neurological disorder [74].

This chapter will delve into the PE applications over EMG signals, particularly surface EMGs (sEMG). Since acquiring sEMG data from the surface of the skin is significantly less invasive than the traditional needle and wire detection techniques, the former can be implemented in a much more diverse and flexible set of conditions, while the latter maintains limited clinical applications.

That being said, we will begin this chapter by presenting the necessary biological background involving the neuromuscular dynamics of muscle contraction and the general characteristics of EMG methods. We will then move on to explain the factors affecting the EMG signal shape and defining the biological complexity within it from an information entropy perspective. Lastly, we will apply the MPE techniques on a series of isometric muscle contractions datasets for the purpose of characterizing the information content both on fatigue conditions and contractions at different force level outputs.

5.2 Motivation

Before going further, we must first express the motivations behind the proposed experiments in the chapter. So far, our approach to PE techniques has been purely theoretical, even when testing our methods and models. This is necessary and desirable from the academic perspective. Nonetheless, if the intention of this work is to eventually contribute to biomedical applications, we must address the practical, data driven viewpoint. For this reason, it is necessary to test the performance of the MPE methods on real datasets with complex biological dynamics. In particular, we choose to work with isometric force contractions, since these are the best documented experiments– and easiest to reproduce– in literature [75] [76]. Likewise, the signal length of the datasets was chosen to be large in order to minimize bias and variance from the different MPE estimators. This will properly measure the methods' interaction and performance with the biological information, with the least amount of interference from the statistical properties previously discussed. Further testing of the MPE moments on biological datasets with more challenging conditions will be addressed in future work, after these methods' feasibility is confidently established.

5.3 EMG Signals and Biological Complexity

5.3.1 Physiology

If we consider all the possible demands that the human body can be subjected to,, the nervous and motor systems must be able to regulate the muscle force outputs for both powerful and precise movements, as well as to maintain balance, posture, locomotion, and even gestures. This process implies a vastly complex and adaptable set of instructions and processes, for voluntary movement and reflex reactions alike. When we see a muscle as an actuator, the most fundamental component is the motor unit (MU) —the end-effector of the motor control. The MU consists of a single alpha-motoneuron —with the body of the cell located in the spinal cord— and the individual muscle fibers (MF) it connects through its long axon. The alpha-motoneuron integrates all the input from the higher-level central control system (the brain), the peripheral reflex system, and the activity coming from other muscles by means of afferent feedback [77]. MU fibers are activated/inhibited through this process, leading to muscle contraction or a lack of thereof (see Figure 5.1).

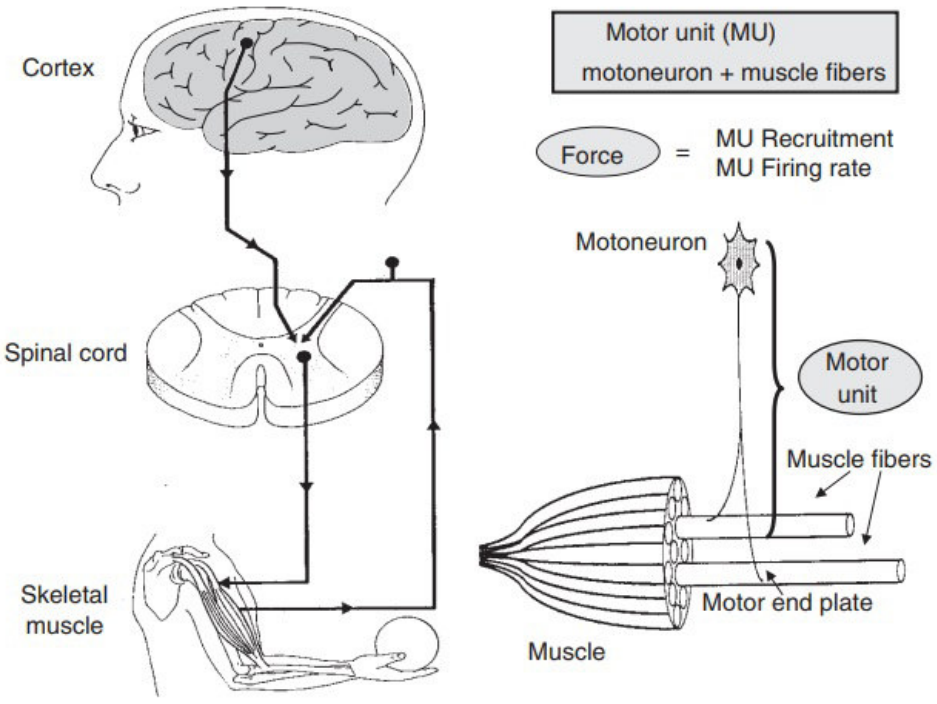


Figure 5.1: Command of the voluntary contraction arising from the cerebral cortex, the latter reaching the motoneuron at the spinal level through projections. The motoneuron axon leaves the spinal cord in order to link with muscle fibres. However, one motoneuron receives several inputs (activation and/or inhibition), with some arriving directly and others via interneurons located at the spinal level. Some of the information it receives stems from proprioceptive feedback. Figure from [78] [79].

MU activation is supported through membrane voltage polarity reversal due to an

increase of the sodium and potassium conductance across the cell membrane. This transient membrane voltage variation is labelled action potential (AP) and grows for the duration —typically between 2 and 5 ms— of a few milliseconds, after which the membrane returns to its state of restom voltage; step by step, this phenomenon propagates along the cell membrane from the neuromuscular junction to the tendon's ending. At the MF, there are regularly spaced invaginations along the membrane (tubular system) and radially oriented inside the cell. This allows to provide the AP to the middle of the cell, where the electromechanical coupling is being done. Indeed, at the tubular system, the AP leads to the release of the calcium (Ca+) accumulated inside the storage tank toward the intracellular surroundings. Hence, Ca+ concentration increases close to the contractile protein's structure (myofibril), which allows it to uncover the binding site between the myosin filaments and the actin filaments —the functional protein for MF contraction. This provides the acto-myosin bridges association in order to produce the filament sliding (Huxley theory [80]) and therefore the MF contraction.

However, a unique sequence of bridges does not have the ability to develop mechanical force. This is achieved by a close succession of bridge formations and dissociations, produced by a train of AP. There are two ways in which the neuromuscular control system drives MU to adjust force development: increasing or

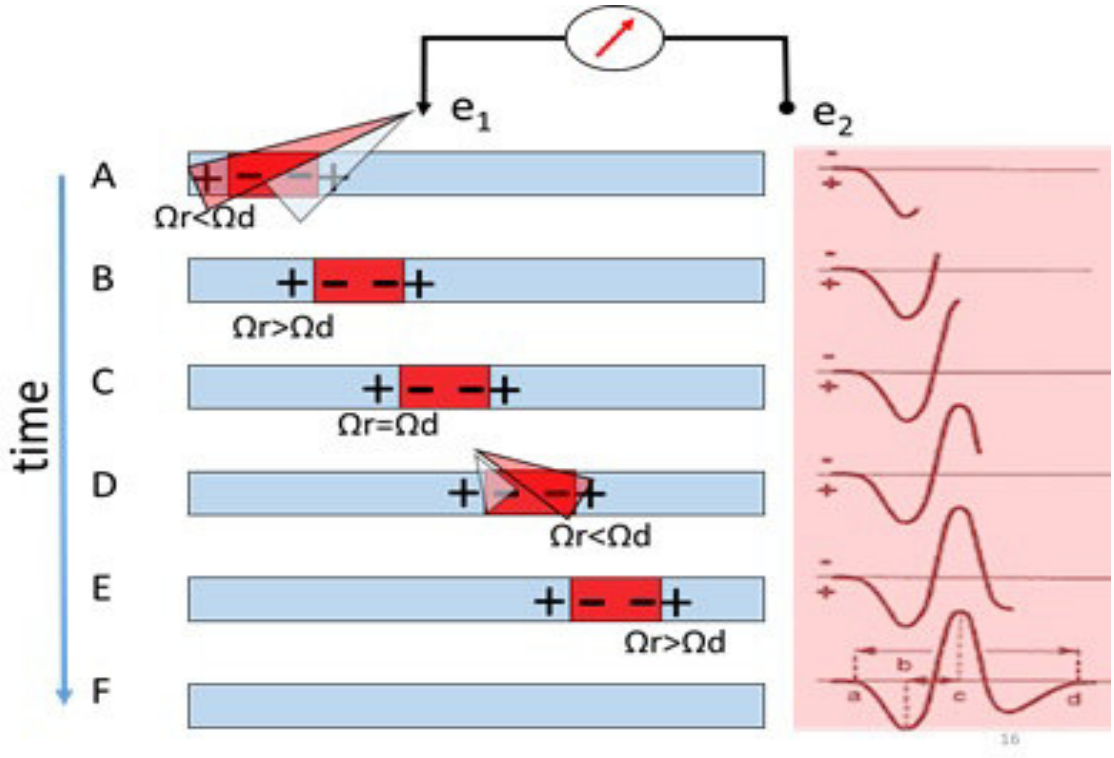


Figure 5.2: Schematic representation of the AP propagating along the muscle fibre, and considered in terms of a leading and trailing dipole pair (extracellular sign depicted) for which the record voltage depends on the angle of electrode (e) view (left panel). The solid angles (Ω) are modified regarding the AP location during its movement and explain the voltage shape detected by the sensor (right panel). Image from [81].

decreasing these interspike AP intervals —defined as the MU firing frequency (rate coding or temporal recruitment) [82]— and the modulation of the number of active MU (spatial recruitment).

AP is a phenomenon localized in a restricted area along the MF that, for any given moment, is surrounded by two areas in which their voltage membrane is at resting state and thus in reverse polarity. This phenomenon can be modelled as a tripole (+ - - +), with two parts have to be taken into account: one for the rising edge —the leading dipole pairs (- +)— and one for the tail —the trailing dipole pair (+ -)— [78]. As this phenomenon propagates along the muscle fibre,the contribution of each part on the final voltage result —which can be observed with electrodes— changes in accordance to the angle of view thus changes the measured potential (Figure 5.2) [81]. Therefore, the conduction velocity of this phenomenon will be affected and, consequently, have a defining effect on the frequency content in the signal.

5.3.2 Motor Units and EMG

Since any given motoneuron activates all of the muscle fibres within a MU, the "remote sensor" will not detect the activity of one muscle fiber, but the voltage field of several muscle fibers. Therefore, one MU produces an individual signal —known as motor unit action potential (MUAP)— that is equal to the sum of all voltage fields stemming from every muscle fiber pertaining to a specific MU. Furthermore, the geometric characteristics of the muscle fibers (such as size, number and location) will determine the shape of the MUAP (Figure 5.3). Lastly, the resulting sEMG signal will emerge by taking into account the interference from a set of generated MUAP.

There are other factors besides the intrinsic characteristics of the MU that influence the surface EMG measuring —or any EMG signal. Indeed, the MUAP propagation medium acts as a low-pass filter, which in turn is composed of several types of tissue that include fat, skin, and the interface between the skin and the sensor. With that in mind, we can see that, the bigger the distance between the MU and the sensor, the greater the decrease in the MUAP amplitude and frequency content [78]. Moreover, there are different MU types available inside a single muscle: MU with greater fiber diameter (large vs. small MU) present an increased AP conduction velocity, which in turn increases the frequency content of the MUAP wave; MU of a larger size contain more muscle fibers, which in turn produce a sEMG amplitude increase [83]. Further complications arise if we opt for sEMG data acquisition through classical surface sensors, since that would not allow us to isolate activity of an individual MU, limiting our detection capacity to only the overall activity of a particular target muscle. The most common spatial recruitment patterns for the latter follow a recruitment order in accordance to the size of all the MU involved [84] (i.e. the smaller ones activate first and the larger ones follow afterwards) to increase the force output; nonetheless, this canonical recruitment order is affected in the presence of several factors, such as contraction speed, fatigue, or stretch reflex. All these parameters show how difficult it is to apprehend and interpret sEMG signals.

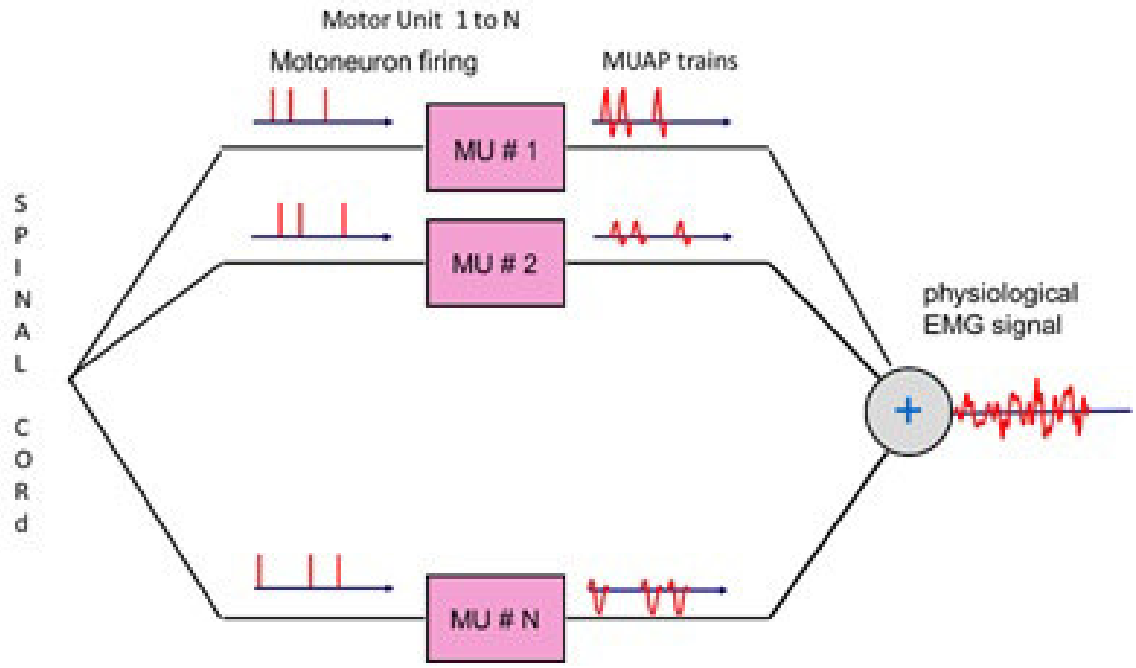


Figure 5.3: The individual motor units (MU) receive activation/inhibition information from the spinal cord. Each activated MU produces a motor unit action potential (MUAP) with a specific firing rate and conduction velocity. The overall sEMG consists of the aggregate interference of all generated MUAP, which in turn suffer nonlinear transformations due to medium propagation. sEMG signal model from [83].

5.3.3 EMG and Complexity

Adaptability and survival are usually seen together in the biological sciences, as it has become a hallmark of living organisms that are fit to prolong their survival. Among the myriad of known adaptability strategies, it is common for them to express multiscale, nonlinear variability, since said feature allows the subject to quickly adapt to any situation; unsurprisingly, it is also the reason why numerous functional or structural studies in the living system report such complex behavior and structures [84]. With that in mind, it becomes necessary to develop analysis techniques that contemplate these types of behaviours, and sEMG is not the exception to the rule. Complex by nature of its own genesis, they consist of a greater number of MU [85], and are composed of various MUAP shapes with a nonlinear mixture. Additionally, sEMG expresses different behaviors, including stochastic and deterministic components that confer a notable level of complexity to its signals [86]. At the same time, just as there is no universal definition for biological complexity, there is also a lack of a single formulation capable of characterizing all the dynamical behaviours in a biological system to date [86]. Despite that fact, we have previously discussed in Chapter 1 that entropy measurements can be regarded as a measure of information complexity, therefore making it possible for us to use it in order to obtain some insight regarding EMG signals.

Permutation entropy techniques were selected due to their two previously mentioned

properties: invariance to amplitude and robustness with respect to noise. In particular, we expect to isolate the effect of pattern shape from the force output (increased amplitude) by using the ordinal pattern approach —an approach that complements other EMG analysis techniques, such as fractal analysis [87]. Therefore, it stands to reason that PE and MPE measurements have become widespread tools in the study of bioelectrical signals [45] [33] [75] [76] [88].

By taking into account our mathematical contributions —from Chapter 2 to 4— in the present work, we can further contribute to the interpretation of PE and MPE results. Our first advancement is acknowledging the bias effect, which improves our interpretations whenever we observe a monotonic (or even linear) MPE decrease with respect to a time scale. We also studied the properties of the MPE variance and presented improvements on MPE techniques by removing variance artifact sources within the estimator itself. Lastly, our new permutation entropy proposed methods allow us to correctly assign observed variations to biological factors and possibly artifacts from measuring protocols [89], rather than labeling them as estimation errors.

5.4 MPE on Real sEMG Signals

We will work directly with real EMG signals in this section as a means to illustrate the applicability of the MPE technique and outline both its practical limits and shortcomings. For the purpose of this research project, we will focus our efforts in the analysis of isometric contractions.

It is usual to observe a decrease of the EMG frequency parameters —the mean or median frequency computed on the power spectral density— during sustained isometric exercise at challenging intensity levels (above 20% of the maximal voluntary force) [90] [91] along with an increase of the EMG amplitude, often quantified by means of the root Mean square (RMS), during the first stage of exercise. While the first measurement provides information about some AP conduction propagation disturbances along the muscle fibres, the second one suggests compensatory strategies provided by the neuromuscular system in order to sustain the requested task [91], such as increasing the firing rate. However, these parameters are also sensitive to other factors (i.e. force variation); to illustrate further, an increase in force involves MU recruitments (spatially and temporally) [92] and, consequently, an increase in EMG amplitude [93]. Such modifications will lead to changes in the nature of the mixture of active MU and in their activities, which results in the modification of the shapes of the MUAP and, eventually, the shape of the sEMG signal.

In view of this, we propose to explore the consequences of this physiological upheaval as a function of the force factor and as a function of fatigue through entropic indicators of sEMG.

5.4.1 Methods

Experimental Setup

Data was collected during a previous study conducted by the signal team of the PRISME laboratory, in which ten healthy subjects (three women and seven men, ages 24 ± 1.5 year, all right-handed) participated in this study. They were fully informed about the experimental procedures and every subject gave their signed consent.

An isometric ergometer was specifically designed for this experiment in order to secure the subject's body on a chair, focusing on the trunk and joints involved in the isometric flexion of the right elbow (shoulder and elbow); in regards to the arm, it was positioned as to rest horizontally —perpendicular in relation to the the rest of body, with the elbow joint angle being immobilized at a 100º of extension— and the hand was oriented midways between a supination and pronation position. Subjects had to pull on a rigid wire, which in turn was connected to a strain gauge, to activate the measuring of the isometric force output by means of a wrist-cuff attached close to the styloid process. Moreover, a visual feedback displayed the supplied force in comparison to the requested level.

The sEMG signal of the *biceps brachii* was recorded by means of electrodes located on the muscle belly, halfway between the motor innervation point and the tendon, and was also boosted by a bipolar isolated amplifier. Force and sEMG signals acquisitions were synchronized by an analog-to-digital card (PCI 6023E, National Instrument, USA) at a 10 kHz sampling frequency.

The protocol was built around maximal isometric elbow flexion contractions as its cornerstone activity: subjects would sustain each contraction for three seconds each, and have a three-minute rest between each contraction; additionally, it was determined that the trial would revolve around the subjects' maximum voluntary contraction (MVC). The pre-exercise test had subjects perform at least three contractions as a warm-up, while the main phase of the trial had subjects perform the same kind of contractions at 20%, 40%, 60% and 80% MVC in a randomized order. After a final three-minute rest, subjects were asked to perform a sustained 70% MVC contraction until exhaustion, when they could not support the required force level.

Data Setup and Entropy Techniques

MPE has been so far explored as a function of the *fatigue* factor by carrying out calculations on datasets obtained during the sustained force until exhaustion. Each sample obtained during this work were split in four non-overlapping segments of equal length, labeled as windows (W_1, W_2, W_3, W_4), in chronological order ($0\% - 25\%$, $25\% - 50\%$, $50\% - 75\%$ and $75\% - 100\%$ time to exhaustion, respectively). MPE has also been explored as a function of the *force level* factor by carrying out calculations on datasets obtained from the four short contraction exercises at 20%, 40%, 60%, and 80% MVC.

MPE, rcMPE and rcDPE calculations were applied to each segment at different

values of scale (m), from 1 to 100 in 1-step increments and at dimension (d) values of 3, 4, and 5. Since the average signal from the trials is $\overline{N} = 274{,}000$ data points, the maximum dimension of analysis is set at $d = 5$ and the length criterion adheres to the aforementioned discussion on classic MPE (Chapter 2.6). For practical purposes, the down sampling parameter for rcDPE was set to $\tau = m$.

Statistical Tests

The entropy averages were calculated on eleven values for each of the conditions present in this study: MPE methods, scale, dimension, fatigue step, and force level. The study of the differences between methods and their parameters (scale and dimension) has been carried out on gross results from segment W_1 to W_4 (fatigue condition) and from segment 20% to 80% (force level condition), as well as the difference between these segment results. For the *fatigue* (delta step) and the *force level* (delta force) conditions, five and six combinations have been taken into account, respectively: $[W_1 - W_2;\ W_1 - W_3;\ W_1 - W_4;\ W_2 - W_3\ ;\ W_3 - W_4]$, and $[20\% - 40\%;\ 20\% - 60\%;\ 20\% - 80\%;\ 40\% - 60\%;\ 40\% - 80\%;\ 60\% - 80\%]$.

Statistical analysis was performed by means of the Statistica 7.1 Software (Stat Soft. Inc.). Given that samples were drawn by normality and equal variances in all groups (evaluated by the Lilliefors and the Levene tests [94], respectively) in the case of the fatigue study, statistics were conducted by means of three-way repeated measures analysis of variance (ANOVA). The repeated measure is the step fatigue (or the delta step fatigue) and corresponds to the first factor. The other two factors were the *method* (MPE, rcMPE, rcDPE) and the *dimension* (3, 4, 5). When a significant difference was observed, the Bonferroni comparison procedure was used in order to isolate the differing groups.

Since both the assumptions of normality and homogeneity of variances failed in the first part of the force study, the Kruskal-Wallis multicomparison test was performed to compare the methods' ability to discriminate between the different levels of force (20%-40%; 20%-60%; 20%-80%; 40%-60%; 40%-80%; 60%-80%) and to evaluate the effect of the *dimension* factor (3, 4, and 5). In the second part of this study, from the previous results, only one method, scale and dimension have been selected. Under these conditions, the assumption on conducting parametric models has been satisfied, hence repeated measures ANOVA were carried out to investigate the entropy evolution together with force level. Following this analysis, the Bonferroni comparison procedure was used to isolate the force level groups (20%, 40%, 60%, and 80% of MVC) that differ. All significance thresholds were fixed at $\alpha < 0.05$.

5.4.2 Results

Comparison between methods and their settings

Fatigue

As a first approach, four mean MPE curves were plotted as a function of time scale m at dimension $d = 3$ (Figure 5.4, left-hand column) for all four window

segments (W_1 to W_4) while contemplating the three methods (MPE, rcMPE, rcDPE ; 3 subplots). Two phases can be observed in these curve kinematics: first, a rapid increase is achieved on shorter time scales (from $m = 1$ to $m \approx 60$), followed by a steady state ($m > 60$) for the three algorithms. The four curves belonging to the four segments differ below this value $m = 60$ and are superimposed above it. Regarding the methods, the curve profiles are similar, although both of the two refined composite methods show smoother lines in comparison to MPE, matching our previous results from Chapter 4. Presentation of MPE as the difference between segment results (ΔMPE) allows us to refine the scale area that serves as the best differentiator between segments(Figure 5.4, right-hand column). Indeed, the curves present a greater difference in the 8 to 15 m value range, and since this area remains unchanged regardless of the method applied, we can deduce that it corresponds to the optimal scale range of analysis for this particular dataset. This range prevails even if a higher embedding dimension is selected (Figure 5.5). Therefore, for the subsequent steps in our analysis, we kept the $m = 10$ value fixed.

The *method* factor has a significant effect (Figure 5.6a) on the ΔMPE value ($F(2, 486) = 4.091$, $p = 0.0077$), with rcDPE producing higher differences than the other two methods (MPE and rcMPE, $p = 0.017$ and $p = 0.025$, respectively). *Dimension* also has a significant effect (Figure 5.6b) on the ΔMPE value ($F(2, 486) = 94.837$, $p < 0.001$), with *dimension* $d = 5$ producing higher differences than $d = 4$ ($p < 0.001$) or $d = 3$ ($P < 0.001$). No evidence of interaction between factors has been reported.

Force Levels

Similarly to the isometric sustained contraction dataset, the four MPE curves —corresponding to the four force levels (from 20% to 80% MVC) with respect to time scale m— follow a rapid entropy increase (Figure 5.7, left-hand column). The curves achieve a steady state around $m \approx 60$, up to the specified maximum ($m = 100$). As previously seen, the difference is that the two refined composite methods show smoother lines when compared to regular MPE. Differences between force level (Figure 5.7, right-hand column) are more relevant around $m = 1$, and drop off rapidly with increasing scale. Hence, the smallest time scale ($m = 1$) is the most suitable setting to discriminate MPE differences between force levels for the isometric condition of this study. This behavior is preserved for dimensions $d = 4$ and $d = 5$, as shown in Figure 5.8.

No significant statistical effect stemming from the method of choice was observed despite an observed trend—Kruskal-Wallis test $H(2, N = 8100) = 4.045$, $p = 0, 13]$. The P value between methods is indicative of the aforementioned trend, with a higher ΔMPE result for rcDPE (sum of rank 4121) when compared with the classical MPE result (sum of rank 3997). On the other hand, dimension —Kruskal-Wallis test $H(2, N = 8100) = 85.82$, $p < 0, 001$— has a significant effect on the ΔMPE value, with dimension $d = 5$ (sum of rank 4249, $p < 0, 001$) producing higher differences than both $d = 4$ (sum of rank 4191, $p < 0.001$) and $d = 3$ (sum of rank 3712, $p < 0.001$). Significantly different force level comparisons are shown in Figure 5.9.

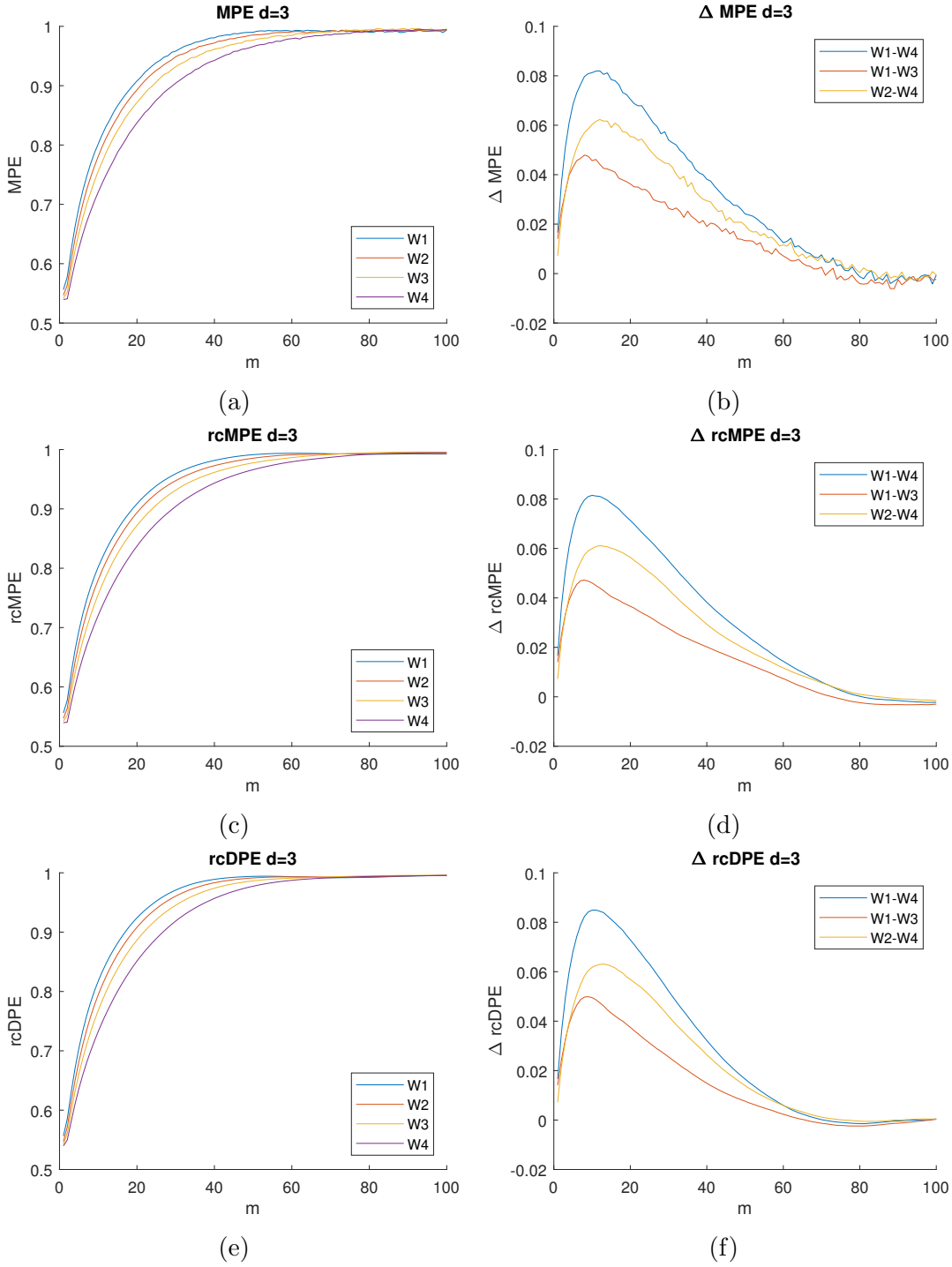


Figure 5.4: Left column: mean entropy values as a function of scale (m) for the fatigue steps (from W_1 to W_4). Right column: variation in entropy mean values as a function of scale (m) on the pairwise differences between steps of fatigue (only $W_1 - W_3$, $W_2 - W_4$, and $W_1 - W_4$ are shown). From top to bottom: MPE (a, b), rcMPE (c, d) and rcDPE (e, f). All values at dimension $d = 3$.

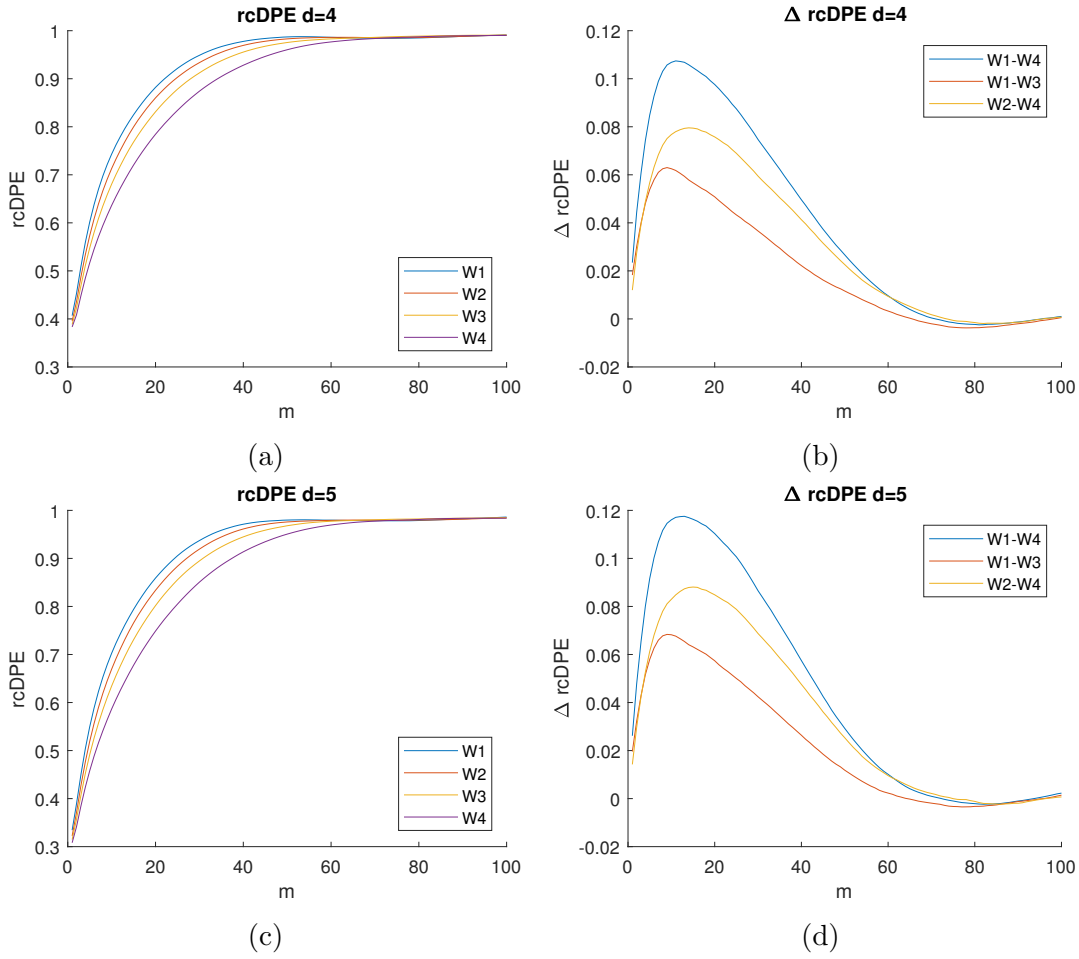
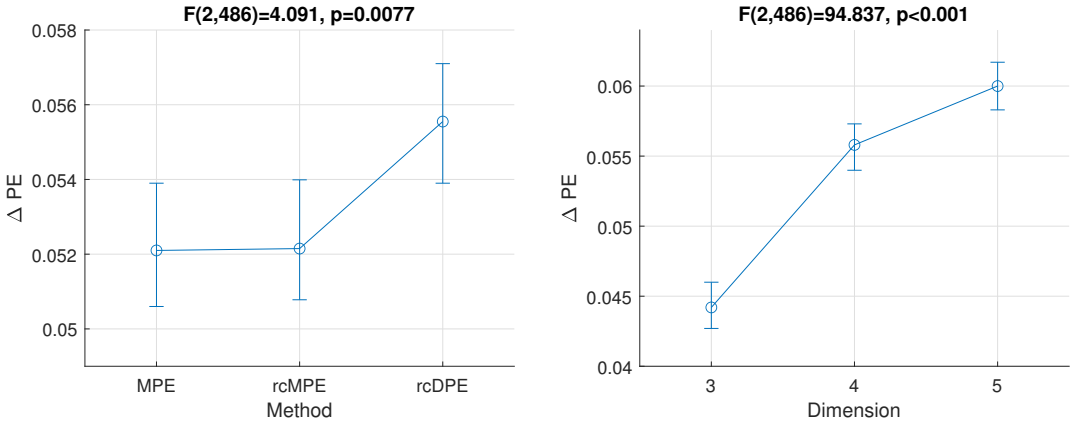


Figure 5.5: Left column: mean entropy values as a function of scale (m) for the fatigue steps (from W_1 to W_4). Right column: variation in entropy mean values as a function of scale (m) on the pairwise differences between steps of fatigue ($W_1 - W_3$, $W_2 - W_4$, and $W_1 - W_4$). Dimension values, from top to bottom: $d = 4$ (a, b) and $d = 5$ (c, d).



(a) (b)

Figure 5.6: Effective hypothesis decomposition for fatigue isometric contraction entropy for the following factors: (a) *methods*, (b) *dimension*. The vertical lines denote 95% confidence intervals for the ANOVA F distribution.

MPE Application

The physiological MPE application has been managed using the rcDPE developed in the present study by setting it at the dimension $d = 4$ and scale $m = 10$ for the sustained exercise conditions, and at the dimension $d = 4$ and scale $m = 1$ for the isometric force level.

Repeated ANOVA measurements concerning the fatigue and force factor revealed statistically significant effects ($[F(3, 27) = 72.3,\ p < 0.0001]$ and $[F(3, 27) = 96.298$, $p < 0.0001]$, respectively). Additionally, there was a significant rcDPE decrease for both factors: in the case of fatigue, this was associated with time progression to exhaustion (Figure 5.10a) and occurred throughout all stages ($p < 0.0001$); in the case of force level, this was associated with force level increase (Figure 5.10b) and occurred throughout all force levels ($p < 0.0001$)

5.4.3 Discussion

Optimal Parameter Settings

We will begin discussing the optimal parameters for this analysis by taking the observed results as our reference. As we saw from the fatigue signals, the time scale with the most pronounced difference between segments is around $m = 10$ (or $\tau = 10$). Since the original sampling frequency used for the experiment is $f_s = 10$ kHz, the effective sampling rate over the coarse/downsampled signals is $f_s = 1$ kHz, which only captures information the range 0 - 500 Hz as a consequence of Shannon's theorem. According to [95], this is the frequency range where sEMG information is produced by the MU contractions. Although more evidence is needed in this regard, we expect that the most adequate sampling rate for PE calculations will be close to 1,000 Hz. Conversely, the % MVC categories present the most pronounced differ-

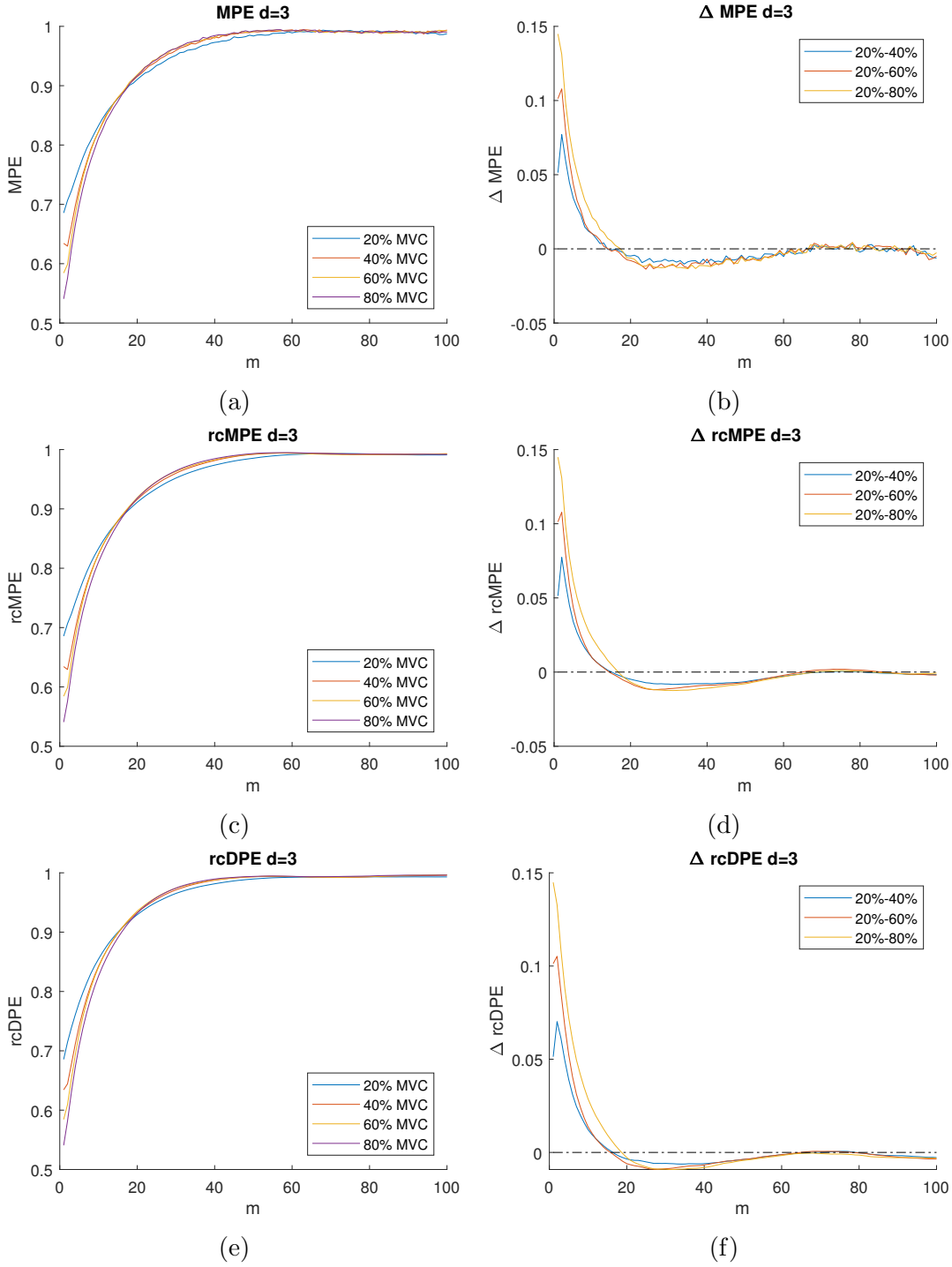


Figure 5.7: Left column: mean entropy values as a function of scale (m) for the force level (from 20% to 80% MVC). Right column: variation in entropy mean values as a function of scale (m) on the differences between force levels (from $20\% - 40\%$ to $20\% - 80\%$). From top to bottom: MPE (a, b), rcMPE (c, d) and rcDPE (e, f). All values at dimension $d = 3$.

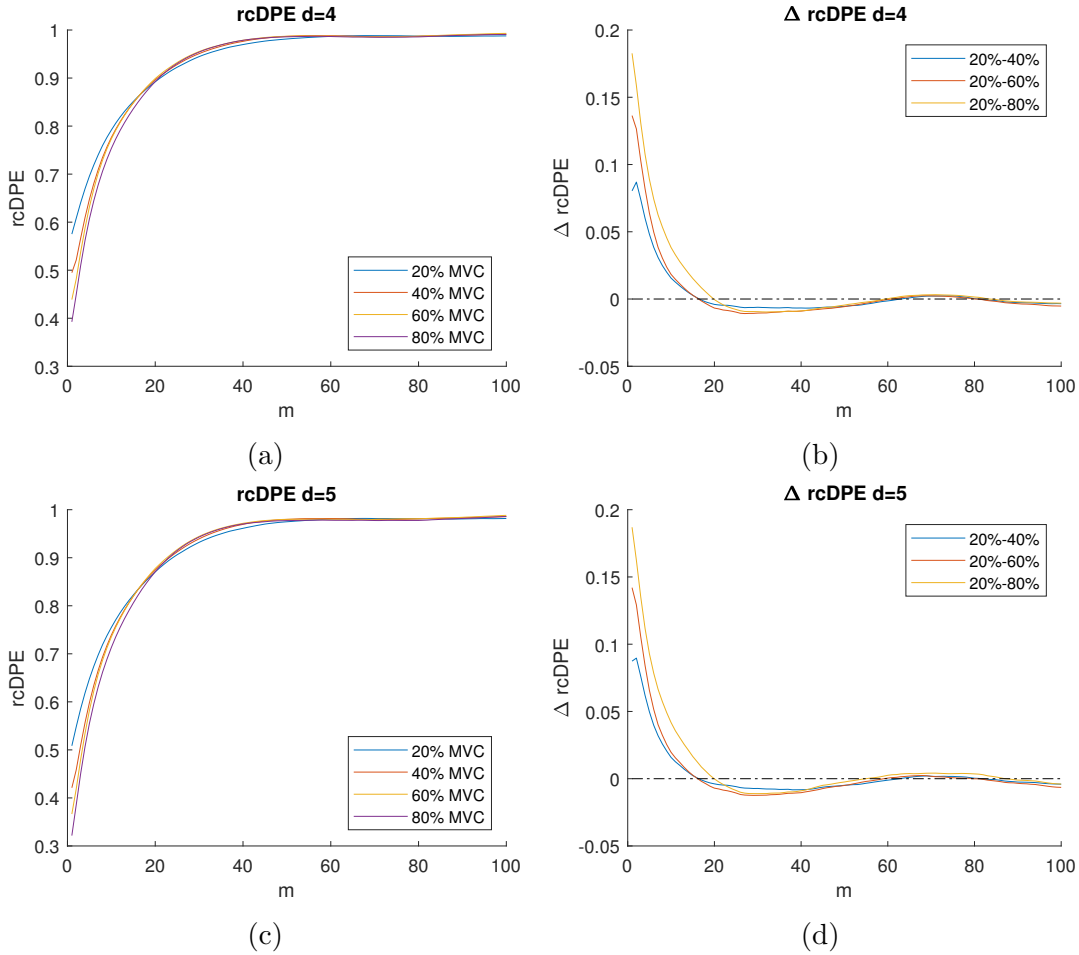


Figure 5.8: Left column: mean entropy values as a function of the scale (m) for the force level (from 20% to 80% MVC). Right column: variation in entropy mean values as a function of scale (m) on the differences between force levels (from $20\% - 40\%$ to $20\% - 80\%$). Dimension values, from top to bottom: $d = 4$ (a, b) and $d = 5$ (c, d).

ences at low time scales. In this particular dataset, a scale around $m = 1$ is the most appropriate, for this setting seems to be task dependent. Therefore, a preliminary visual exploration is recommended to select the adequate scale. We should also note that, for this case, $m = 20$ —which corresponds to a 500 Hz sampling frequency— should be avoided, since any difference between % MVC disappears at this scale.

We found the embedding dimension to be particularly important in the differentiation between entropy values. Here, the maximum dimension used ($d = 5$) yields the highest differences between segments. As we mentioned in Section 5.4.1, we did not use higher dimensions in order to maintain the MPE bias below acceptable levels, due to the signal length constraints. This restriction does not apply to rcMPE and rcDPE, where we can use higher dimensions (See Section 4.4.2). In the case of rcDPE, signal length allows the use of higher d values without significant increases in bias. Further testing needs to be done regarding the performance of rcDPE at higher dimensions, with a special emphasis on the increasing computational costs involved.

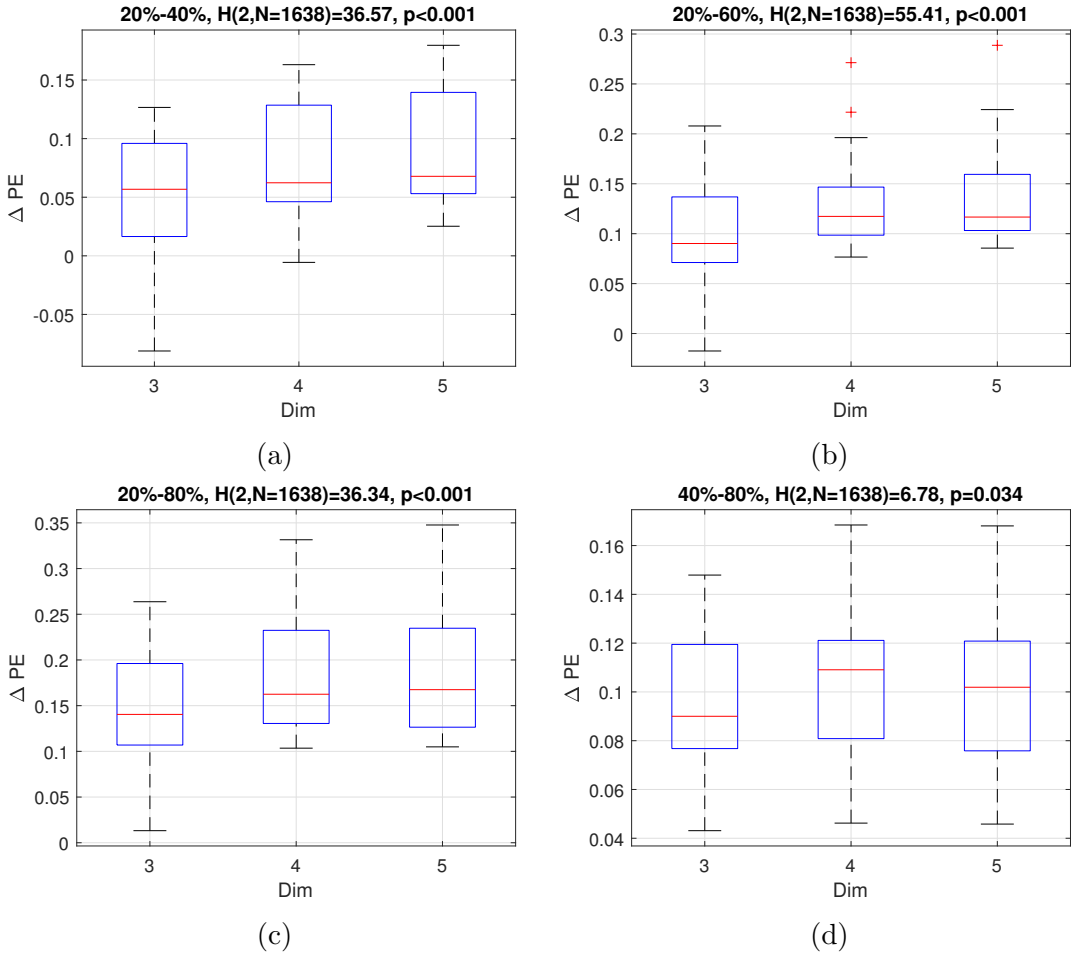


Figure 5.9: Range boxplots for MPE differences with increasing dimension $d = [3, 4, 5]$. (a) $20\% - 40\%$, (b) $20\% - 60\%$, (c) $20\% - 80\%$, (d) $40\% - 80\%$. Only statistically significant results are shown (Kruskal-Wallis test with $\alpha = 0.05$).

Therefore, as a general rule of thumb when working with sEMG signals that include fatigue, occur during isometric contractions and are a product of a high level of force (above 20% MVC), we recommend the use of a time scale that adjusts the coarse (or downsampled) signals close to 1,000 Hz, with at least an embedding dimension close to $d = 5$. Regarding the time scale, we should adjust our parameter (m or τ) in order to capture the maximum difference between groups. These parameters will not be standard, and depend on the particular characteristics of the dataset. The differentiation between % MVC offers a better performance when using a scale that fits a high frequency (10 kHz) and dimension $d = 5$, albeit the advantages of using this dimension instead of $d = 4$ are negligible.

Entropy Technique Performance

Regarding the different available entropy methods utilized in the present work, we found that the rcDPE method makes the differences between fatigue activity windows at isometric contraction stand out better than methods using either rcMPE or classical MPE. On the other hand, methods are no longer relevant in order to

distinguish between force levels, since all MPE algorithms are mathematically identical when we set the dimension value as $m = 1$. It should be noted, however, that the entropy curves were more stable in both of the refined composite methods when compared them to classical MPE.

Even though all methods have been proven to be valid due to the statistically significant results obtained, it should be noted that choosing a time scale that is either low or too high would be affect the experiment negatively, albeit for very different reasons: while choosing a time scale that is too small would make the differences between methods almost indistinguishable, a time scale that is too high would render all of the methods useless, since they cannot distinguish complexity from noise, even when the length constraints are satisfied.

If we compare these results with the information from previous chapters, we note that the effect of the MPE bias is not evident, even in the case of classical MPE. Since this experiment collected long length signals (the 70% MVC signals averaged a signal length of $\overline{N} = 274,000$, while the other % MVC signals had a $N = 30,000$ value each), the bias effect remains small even at high time scales. Also, a high signal length yields a small MPE variance, which contributes little to the overall observed variability between subjects.

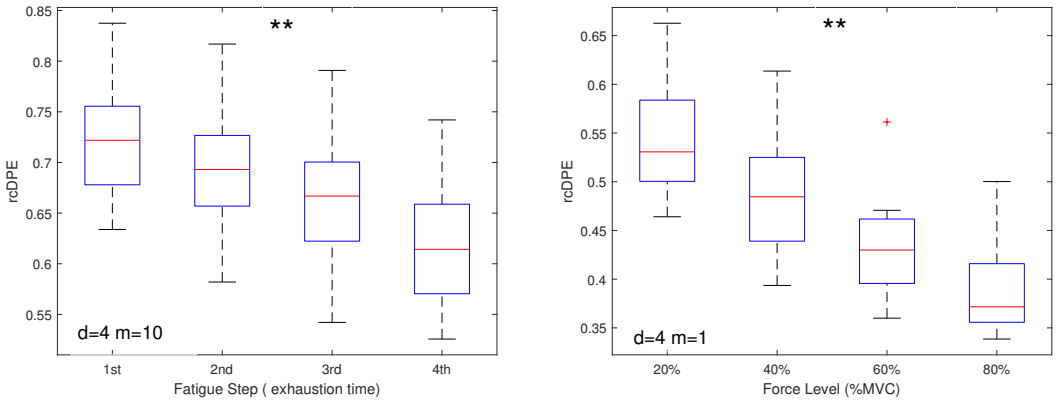
By comparing the MPE curves, we observe results which are close to the ones reported by Cashaback [86] for multiscale sample entropy over continuous 70% MVF activity. This result is also consistent with the one found by Navaneethakrishna [76] when using MPE directly: for short-time scales, the curves present a sharp increase in MPE, while MPE remains almost stationary in long-term scales.

According to [86], short-term scales have a significantly different multiscale sample entropy at different stages of fatigue while the long term scales report no statistical significance, although they present a noticeable difference. In the context of MPE, high-time scale entropy stability suggests that there is no difference between sEMG and uncorrelated noise [76]; this is supported by previous work present in Chapters 2 and 4. While the exact scale between the two regions is not available before the MPE calculation, MPE stabilizes around scale $m \approx 60$ for the fatigue and % MVC sEMG data sets.

In regards to the% MVC signals case, the fact that the MPE methods were able to distinguish between force outputs is worthy of notice. By the definition of these ordinal patterns, no information concerning sEMG amplitude was used, therefore suggesting that the differences in MPE do not come from a measurement of force output, but from the information contained within the sEMG ordinal patterns. This indicates the presence of different dynamics within the signals while using different force outputs.

Physiological Findings

The changes in entropy values as a function of fatigue are in line with existing literature: the study done by Cashaback et al. [86] has a similar experimental setup (muscle model, isometric contraction, different levels of strength and fatigue test) and it also reported that entropy decreased in the presence of fatigue. However,



(a) (b)

Figure 5.10: Range boxplots for rcDPE measurements from isometric contractions with dimension $d = 4$. (a) Four activity windows from sustained 70% MVC at scale $m = 10$, (b) % MVC force levels at scale $m = 1$. From the repeated ANOVA test measurements, the pairwise comparisons are statistically significant ($p < 0.001$). statistically significant pairwise comparisons ($p < 0.001$) are shown with a **.

study contrasts with the present work, as a decrease in entropy is observed as the force level increases in the latter, while an increase and then a decrease in this sEMG indicator is reported in the former under the same circumstances. However, this discrepancy can be explained by different force levels, since they ranged from 40%, 70% and 100% MVC in the former whereas the latter utilized a 20% to 80% with 20% increments. Moreover, our research shows more pronounced kinematics in the entropy indicator.

Observed entropy depletion can be explained through the relationship between force intensity and the increase in probability of temporal overlap between MUAP [96], with firing rate and the number of recruited MU showing a similar behavior in the presence of force output. Although the main recruitment strategy used by the motor command to increase the force level up to 80% MVC is the spatial mode (the number of active MU, in the case of the biceps brachii muscle), rate coding is also expressed [97]. Moreover, MU firing rate synchronization may also play a part in this phenomenon due to the fact that it takes place at higher levels of force (from 80% MVC) [85]. All these modifications can reduce variations in the sEMG path as reported in the study by Gabriel et al. [97] and consequently reduce the entropy signal . Indeed, the present work reports a continuous decrease in the average number of peaks per spike when force intensity increases (from 40% to 100% MVC). Here, "spike" is defined as the consecutive succession of a peak and a valley (positive and negative, respectively) that crosses the isoelectric line, while "peak" corresponds to a fluctuationdeflection that does not cross that line.

Some of the factors that contribute to entropy decrease in the force intensity condition may also explain the effect of fatigue in entropy decrease, for it has been reported that MU firing rate synchronization increases continuously until exhaustion during sustained effort [98]. Another factor which is fully expressed during

peripheral fatigue is the reduction of muscle fibre conduction velocity (MFCV) [91], with part of this reduction being attributed to an increase of AP duration due to disruptive electrophysiological phenomena [99]. Furthermore, this increased duration of the AP phenomena should widen the MUAP waveform [100].

Lastly, the biological interpretation of the time scale deserves further comments. We know the information content inside sEMG lies in the range of 20 Hz – 500 Hz [95], the most appropriate sampling frequency for measurements is $f_s = 1,000$ Hz (due to the Shannon's Sampling theorem), as shown by the Fatigue signal dataset. By having the original signal sampled at $f_s = 1$ kHz with scale $m = 10$ (or $\tau = 10$), we capture only the desired range where the information is meaningful. For the refined composite methods, it is better to use this setup, since we are measuring multiple coarse/downsampled signals which improve the final entropy precision. Furthermore, the use of rcMPE or rcDPE avoids the overemphasis of the noise due to high sampling frequency; an effect already discussed in section 3.2.2. In this context, we still recommend the use of rcDPE over rcMPE, since the latter does not perform the filtering effect discussed in section 1.4.1.

The % MVC dataset is harder to interpret. Here, the higher differences were obtained at $m = 1$, suggesting high frequency information content is important for classification. Moreover, for $m = 20$, the information comes from the effective range 20 Hz – 83 Hz, which lead to no difference between force levels whatsoever. This implies that no relevant information content is present at low frequencies for ordinal entropy techniques. Further experiments are needed to properly understand and interpret this result.

5.5 Closing Remarks

Throughout this chapter we have reviewed the properties and physiological mechanisms behind sEMG signals, from the perspective of both biomedicine and information theory. We designed an experiment concerning isometric contractions that contemplated different force output and fatigue conditions, with the intention of gauging the validity and performance of different permutation entropy measurements (MPE, rcMPE, and rcDPE). Thereafter, as we looked for the optimal parameters to maximize the entropy difference between muscle contraction conditions, we decided to test these calculations under the effect of the time scale and embedding dimension variables. Finally, we performed a battery of statistical tests in order to determine the statistical significance of the results obtained.

First and foremost, we determined that the rcDPE method outperforms the other methods in the differentiation of fatigue levels in isometric contractions. We found the maximum entropy differences at a time scale corresponding to a frequency of 1,000 Hz, and we concluded that the embedding dimension is an important factor, since an increase in the dimension value makes the entropy difference more evident.

These entropy methods were also able to detect differences between force levels. Since the optimal time scale for MPE differentiation was the first time scale, the MPE methods yield the exact same results —with rcDPE offering small benefits when used in this scale. Nonetheless, the embedding dimension still proved to be

important by showing an increased differentiation capacity at higher values, albeit less pronounced than in the case of the fatigue dataset.

All observed contractions have a rapid entropy increase at low time scales and stabilize afterwards. Since we know from previous chapters that this horizontal entropy line cannot be distinguished from uncorrelated noise, it is implied that high-time scales provide no information regarding sEMG dynamic activity.

From the biomedical point of view, the results in this chapter agree with current literature. The observed reduction of MPE is linked to the reduction of the sEMG signal's complexity, which are well-established effects of fatigue. Regarding the force output levels, the difference in MPE suggests a different activity pattern at different percentages of maximum voluntary contraction. It is interesting to note that MPE, by definition, disregards the information contained in the amplitude of the sEMG signal. Therefore, any differentiation from the signals at different force levels must originate in a different pattern of motor unit recruitment and firing rate.

Chapter Summary

- Surface electromyographic (sEMG) signals are measurements of the electrical activity stemming from muscle motor units (MU), which are in turn responsible for muscle contraction. Skin electrodes record the aggregate activity of all the MU involved in the contraction.
- The final shape of the sEMG signal is the result of several intrinsic and extrinsic factors, such as MU size, conduction velocity, firing rate, MU recruitment strategy, sensor distance, intermediate tissue filter, as well as the subject's age, fatigue level, and possible motor pathologies.
- sEMG signals, being complex by their own genesis, can be characterized by the amount of information contained within them. Particularly, ordinal entropy methods provide a good measure of the aggregate sEMG patterns.
- It has been observed that a compromised biological system presents a systematic reduction of entropy, and thus, can help in the detection and differentiation of different activity states.
- In our experiment, ten subjects performed a series of isometric contractions of their right *biceps brachii* at different levels of maximum voluntary contraction (MVC), followed by a sustained force effort until exhaustion. We took several permutation entropy measurements (MPE, rcMPE, and rcDPE) at different dimensions ($d = 3, 4$,and 5) in order to observe the difference between force levels and fatigue states. We also explored the signals at different time scales ($m = 1, \ldots, 100$) in order to gauge the appropriate value which maximizes the differences in each case.
- When analyzing sustained isometric contractions, we found that the rcDPE method outperforms the other methods with statistical significance. We also found the highest dimension used ($d = 5$) to be the best parameter to differentiate between fatigue states.

- When analyzing different % MVC signals, we found that rcDPE has a slight comparative advantage in the differentiation between force levels. We found again $d = 5$ to be the best dimension value, albeit its improvement being small when compared to $d = 4$, suggesting diminishing returns for higher dimensions.
- The time scale range that maximizes differences for sustained contraction was found to be between $m = [8, \ldots, 15]$, regardless of method. The highest differences between % MVC levels were found in the first time scales.
- When we applied the rcDPE method with $d = 4$, we found significant results between the different entropy values at different sections of the fatigue (at $m = 10$) and force levels ($m = 1$) datasets.
- These different rcDPE measurements are consistent with existing literature, where a reduction of entropy is observed in the presence of fatigue.
- The reduction of rcDPE at high % MVC can be explained by a synchronization of the MU firing rate, as well as an overlap between MUAP.

Conclusions

. . . y será tan hermoso decir . . .
ahora nos vamos al centro y nos compramos un helado
el mío todo de frutilla
y el de usted con chocolate y un bizcochito.

- Julio Cortázar, *Me caigo y me levanto*

In the present work we have explored the hidden, underlying properties of multiscale permutation entropy (MPE). Our first goal was to set and further enrich the theoretical body of knowledge behind the MPE algorithm, since that would allow us to have a proper assessment of its statistical properties, advantages, and limitations. Our second goal was to propose a new MPE method that takes advantage of said theoretical findings. Our third and last goal was to apply this knowledge in a complex biomedical problem, such as electrical muscle activity, in order to differentiate between fatigue and force output as states of performance.

To better position ourselves in the context of information theory, we proposed in Chapter 1 a general criterion for classifying the most commonly used entropy measurements with respect to its core formulation, the definition of the event set, and the preprocessing techniques used prior to entropy computation. Although many different expansions, improvements and generalizations have been proposed since Shannon's original work, there is no current universal best method for entropy computation, since the peculiarities of the phenomenon in question must be taken into account; even the mere definition and nature of entropy can only be interpreted within the context of the specific experiment (complexity, amount of information, etc.). Therefore, our aim is to provide a broad view of the entropy variants we can implement and explore, from the mathematical and statistical point of view.

We introduced our main theoretical development of MPE in Chapter 2. By mean of a polynomial expansion, we were able to find an analytical approximation for the MPE measurement, which allowed us to find a closed expression for its first two moments. We found the bias of MPE, whose approximation is independent of the pattern distribution, by only taking into account the embedding dimension, signal length, and scale. We also characterized the MPE variance, which is tightly linked to the MPE measure itself. Here, we found our MPE variance approximation to closely resemble the Cramér-Rao lower bound for minimum variance. Even as a biased statistic, the estimation is almost efficient. Armed with this newfound knowledge, we proposed a more precise minimum length criterion for MPE. We also pointed at the MPE values with maximum uncertainty along the normalized entropy range.

We explored the expected results for different signal models in Chapter 3, with the aim of exploring the effect of different signal properties on the overall MPE results. We found the entropy of deterministic signals to be affected by the slope of the signal, the sampling rate, and the amplitude of the noise —albeit the method is quite robust to noise when the signal has a pronounced slope. Subsequently, we explored the MPE of stochastic processes, particularly fractional Gaussian noise (which is fractal) and ARMA models, where we found that it is possible to estimate a theoretical MPE result from the processes' parameters. In a general sense, this implies that the parameters that define a random process contain all the information from the process itself, and it is possible to test real signals vs. proposed models by comparing entropy measurements.

At this point, our mathematical base was sufficient to tackle the statistical properties of more refined MPE methods. We explored the properties of the well-established composite MPE (cMPE) and refined composite MPE (rcMPE). While the improvement of the MPE estimation —particularly rcMPE, where both the bias and the variance are reduced— is well-established, we found that composite coarse-graining introduces an artifact cross-correlation between the possible coarse signals. Although the overall variance is reduced with respect to the original MPE algorithm, this effect adds an artificial source of uncertainty. Here, we proposed a composite downsampling procedure as a substitute to the classical coarse-graining used for multiscale entropy techniques. This approach avoided the problem of artifact cross-correlation entirely, yielding an increase in precision over the methods in existing literature. In particular, refined composite downsampling permutation Entropy (rcDPE), on top of having the smallest variance among the methods discussed herein, also has the added benefit of a constant bias for any particular scale/downsampling parameter. In contrast to the other two methods, this allows the method to utilize higher values in both scale and dimension, and therefore, it is the technique we recommend for an ordinal entropy approach.

Finally, in Chapter 5 we were in the position to test these tools and methods on real signals. Our chosen datasets consisted of surface electromyographic (sEMG) signals, which are convenient to implement due to their methods being noninvasive in nature. Nonetheless, some of the challenges present in this technique are noise sources and the superposition of multiple signals, which turns depend on factors such as geometry, conductivity, and a myriad of biological considerations. We found that MPE methods —with rcDPE showing the better results among them— are able to consistently discriminate between different states of muscle fatigue, specially for high-embedding dimensions. Despite the fact that the MPE methods were not as consistent when attempting to find differences between various force outputs, we were still able to differentiate the maximum voluntary contraction (MVC) percentages with statistically significative results. Since ordinal methods normally exclude amplitude, this divergence implies that there are still undiscovered muscle contraction dynamics when different force outputs are applied. On the other hand, choosing the right parameters is important for this classification, and an adequate value selection for both dimension and scale is not necessarily obvious *a priori*: generally speaking, higher-embedding dimensions (within a reasonable range) yield a better differentiation between activity states, albeit with diminishing returns; conversely, there is no universally defined time scale to choose for a proper analysis, and they

must be evaluated on a case-by-case basis. Regarding the biomedical implications of these results, we found a significant entropy reduction when muscles become fatigued. One possible explanation points to the action potential elongation —a product of electrophysiological changes— due to continuous contraction. We also found a significant entropy decrease in the presence of contractions with a high-force output, which can be explained by overlaps with motor unit action potentials, as well as observed motor unit fire rate synchronization.

There is ample room for further research on the theoretical front of this topic, such as utilizing new core entropy definitions in the ordinal context, exploring further event partitions, or even testing more general stochastic models. Most importantly, it would be possible to revisit some of the well-established biomedical datasets and obtain a more in-depth interpretation of the results, particularly when exploring entropy behavior at high scales. On account of the improved precision of the methods herein proposed, they can be applied to search for previously hidden dynamic behavior. We hope this research project will contribute both to broaden the existing mathematical body of knowledge and to further improve the use of entropy techniques at the service of the medical sciences.

Additionally, the research on ordinal entropy methods is far from complete. From the theoretical perspective, we can explore the statistical properties of MPE using core formulations that differ from Shannon's original definition. Of course, we should also incorporate the dynamics of "amplitude-aware" techniques to the general MPE statistical behavior theory. In the domain of MPE signal models, the difference between complexity and randomness is still not completely settled, and we believe that further light could be shed on this topic by studying more elaborate statistical processes and chaotic deterministic signals.

Regarding composite methods, understanding the interaction between coarse signals remains incomplete due to a lack of the proper characterization of the artifact cross-correlation effect. Theoretically speaking, a better proposition concerning the probability distribution of ordinal patterns is fundamental, which is quite similar —yet not strictly the same— to a multinomial distribution.

Finally, the study of MPE methods for the characterization of bioelectrical signals is still a fertile area for study. We expect our proposed entropy methods —particularly rcDPE— to better differentiate between sEMG signals in a variety of conditions. These methods could even be brought to real-life conditions due to their short processing time and the improved precision, since these scenarios usually lack the luxury of good measuring conditions, thanks to factors such as external noise sources or long duration activity bursts. Furthermore, the effect of the MPE bias and variance will become crucial in later studies involving short signals and more dynamic conditions, as these effects will affect the resulting MPE more directly. Additionaly, the study of sEMG signal simulations can further shed light in the underlying dynamics of these entropy methods, particularly for different force level contractions, which require more in-depth exploration.

The present work shows the rich complexity behind ordinal entropy techniques and their subsequent, potential applications on biological systems. Even when the theoretical body still remains incomplete, the possible refinements in the results allow researchers to make finer, more accurate calculations and predictions concerning

health issues, such as motor processes. Having said that, we hope that this research project offers more clarity regarding the aforementioned methods, as well as lighting the way for further research and technology implementations.

Appendix A

Covariance, Coskewness, and Cokurtosis Matrices

In this section we will briefly derive the expressions for the covariance, coskewness, and cokurtosis matrices for a multinomial distribution.

First, we recall from equation (2.9) and (2.10) the structure of the pattern count distribution from Chapter 2,

$$\begin{aligned} \boldsymbol{Y} &= n\boldsymbol{p} + \Delta\boldsymbol{Y} && \sim Mu\left(n, p_1, \ldots, p_{d!}\right) \\ \hat{\boldsymbol{p}} &= \tfrac{1}{n}\boldsymbol{Y} = \boldsymbol{p} + \tfrac{1}{n}\Delta\boldsymbol{Y}, \end{aligned}$$

using the same definitions as in Equations (2.1) and (2.2). As before, $\boldsymbol{Y}$ is the random variable which represents the pattern count in the signal while $\hat{\boldsymbol{p}}$ is the estimator of the pattern probabilities. For an embedded dimension d, there are $d!$ possible patterns.

We should also define m as the time scale for MPE analysis, and let n be the greatest integer number below $N/m + d - 1$, which will represent the length of the coarse-grained signal at scale m.

Additionally, we notice that elements Y_i from vector $\boldsymbol{Y}$ are composed of two parts: a deterministic np_i constituent and a random ΔY_i constituent.

We note that the elements Y_i from the vector $\boldsymbol{Y}$ are composed of a deterministic part np_i, and a random part ΔY_i. It should be evident that $E[\Delta Y_i] = 0$ and ΔY_i is identically distributed to Y.

We start by obtaining the expected values of the vector multiplication $\Delta\boldsymbol{Y}\Delta\boldsymbol{Y}'$. We know that

$$E\left[\Delta Y_i^2\right] = np_i(1 - p_i) \tag{A.1}$$
$$E\left[\Delta Y_i \Delta Y_j\right] = -np_i p_j, \tag{A.2}$$

for $i = 1, \ldots, d!$ and $j = 1, \ldots, d!$. Thus, if we gather all possible combinations of i and j in the covariance matrix, we get

$$\begin{aligned} E\left[\Delta \boldsymbol{Y} \Delta \boldsymbol{Y}'\right] &= n(diag(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^T) \\ &= n\boldsymbol{\Sigma}_p. \end{aligned} \tag{A.3}$$

Similarly, the skewness and coskewness can be expressed as,

$$E\left[\Delta Y_i^3\right] = 2np_i^3 - 3np_i^2 + np_i \tag{A.4}$$

$$E\left[\Delta Y_i \Delta Y_j^2\right] = 2np_i p_j^2 - np_i p_j, \tag{A.5}$$

which yields to the coskewness matrix

$$\begin{aligned} E\left[\Delta \boldsymbol{Y}(\Delta \boldsymbol{Y}^{\circ 2})'\right] &= n(diag(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}') + 2n\left(\boldsymbol{p}(\boldsymbol{p}^{\circ 2})' - diag(\boldsymbol{p}^{\circ 2})\right) \\ &= n(diag(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}') - 2n\left(diag(\boldsymbol{p}^{\circ 2}) - \boldsymbol{p}\boldsymbol{p}' diag(\boldsymbol{p})\right) \\ &= n\boldsymbol{\Sigma}_p\left(\boldsymbol{I} - 2\, diag(\boldsymbol{p})\right), \end{aligned} \tag{A.6}$$

where we use again the vector definitions in Equation (2.27).

Lastly, we follow the same procedure to obtain the cokurtosis matrix, by first obtaining the values

$$E\left[\Delta Y_i^4\right] = 3n(n-2)p_i^4 - 6n(n-2)p_i^3 + n(3n-7)p_i^2 + np_i \tag{A.7}$$

$$E\left[\Delta Y_i^2 \Delta Y_j^2\right] = 3n(n-2)p_i^2 p_j^2 - n(n-2)(p_i^2 p_j + p_i p_j^2) + n(n-1)p_i p_j, \tag{A.8}$$

which combines in the matrix as follows

$$\begin{aligned} E[\Delta \boldsymbol{Y}^{\circ 2}(\Delta \boldsymbol{Y}^{\circ 2})'] =& 3n(n-2)\boldsymbol{p}^{\circ 2}(\boldsymbol{p}^{\circ 2})' - n(n-2)(\boldsymbol{p}^{\circ 2}\boldsymbol{p}' + \boldsymbol{p}(\boldsymbol{p}^{\circ 2})') + n^2\boldsymbol{p}\boldsymbol{p}' \\ & - 4n(n-2)diag(\boldsymbol{p}^{\circ 3}) + 2n(n-3)diag(\boldsymbol{p}^{\circ 2}) + n(diag(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}') \end{aligned} \tag{A.9}$$

$$\begin{aligned} =& -3n(n-2)\, diag(\boldsymbol{p})\boldsymbol{\Sigma}_p\, diag(\boldsymbol{p}) \\ & + n(n-2)\left(diag(\boldsymbol{p})\, \boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_p\, diag(\boldsymbol{p})\right) - n(n-1)\boldsymbol{\Sigma}_p \\ & + n^2\, diag(\boldsymbol{p})\left(\boldsymbol{I} - diag^2(\boldsymbol{p})\right) - 2n\, diag^2(\boldsymbol{p})\left(\boldsymbol{I} - diag(\boldsymbol{p})\right). \end{aligned} \tag{A.10}$$

By taking advantage of this expressions, we are able to calculate the MPE variance in Chapter 2.

Appendix B

Math Glossary

This is a list of the most relevant symbols used through this thesis.

$\boldsymbol{x}$ The vector containing all the data points in an arbitrary time series. This is the starting point of the MPE analysis.

x_t Refers to the element t of $\boldsymbol{x}$, from $t = 1, \ldots, T$.

t The index T will be used to refer to any particular data point in $\boldsymbol{x}$.

N Signal length of $\boldsymbol{x}$.

d Embedded dimension (integer value) that corresponds to the number of consecutive data points that will form a pattern. In most cases, the number of ordinal patterns is equal to the factorial of the embedded dimension, $d!$.

τ Downsampling parameter for signal $\boldsymbol{x}$. For $\tau = 1$, we keep the original signal.

y_i Pattern count $i = 1, \ldots, d!$. The number of times pattern i appearing in the signal $\boldsymbol{x}$.

$\boldsymbol{y}$ Vector of size $d!$ containing all the pattern counts y_i.

p_i Pattern probability $i = 1, \ldots, d!$. The probability of pattern i appearing in the signal $\boldsymbol{x}$. As a reference point, the pattern $i = 1$ will always refer to the increasing pattern (i.e. $x_t < \cdots < x_{t+d-1}$), and the pattern $i = d!$ will refer to the decreasing pattern.

$\boldsymbol{p}$ Vector of size $d!$ containing all the pattern probabilities p_i.

$\boldsymbol{l}$ Vector of size $d!$ containing all the natural logarithms of the pattern probabilities $\ln p_i$.

H This is the general symbol used to denote multiscale permutation entropy. This includes the original formulation of permutation entropy, which corresponds to the first time scale.

H_q Tsallis entropy

H_α Rényi's entropy

$\mathcal{H}$ Normalized multiscale permutation entropy. We obtain this measure by dividing H by $\ln d!$.

m Time scale of MPE. By following the coarse-graining procedure, m represents the size of the non-overlapping segments inside the signal $\boldsymbol{x}$.

$\boldsymbol{x}_{(k)}^{(m)}$ The coarse-grained version of $\boldsymbol{x}$, for time scale m. Note that there are $k = 1, \ldots, m$ possible coarse-grained signals, each beginning with the k^{th} element of $\boldsymbol{x}$. When the starting point is not specified, we can write $\boldsymbol{x}^{(m)}$, and assume $k = 1$.

n_m Signal length of $\boldsymbol{x}^{(m,k)}$.

$p_i^{(m,k)}$ Pattern probability $i = 1, \ldots, d!$ in the coarse grained signal $\boldsymbol{x}^{(m,k)}$. Likewise, when the starting point is not specified for the coarse-graining procedure, we can write $p_i^{(m)}$ and assume $k = 1$.

$\boldsymbol{X}$ A vector which represents a sequential random process with a given model. Not to be confused with $\boldsymbol{x}$, which corresponds to a given time series with unknown *a priori* properties.

Y_i Random variable representing the pattern count $i = 1, \ldots, d!$. Unless otherwise specified, Y_i is assumed to be a binomial random variable.

$\hat{p}_i$ Random variable representing the pattern probability $i = 1, \ldots, d!$. Unless otherwise specified, $\hat{p}_i$ is assumed to be a binomial random variable, identically distributed to Y_i.

$\boldsymbol{Y}$ Vector of size $d!$ containing all the pattern counts Y_i. Unless otherwise specified, $\boldsymbol{Y}$ is assumed to be a multinomial random variable.

$\hat{\boldsymbol{p}}$ Vector of size $d!$ containing all of the pattern probabilities $\hat{p}_i$. Unless otherwise specified, $\hat{\boldsymbol{p}}$ is assumed to be a multinomial random variable, identically distributed to $\boldsymbol{Y}$.

$\hat{\boldsymbol{p}}^{(m)}$ When the vector $\hat{\boldsymbol{p}}$ is not constant respect to m.

$\hat{\boldsymbol{l}}$ Vector of size $d!$ containing all the natural logarithms of the pattern probabilities $\ln p_i$.

$\hat{H}$ Multiscale permutation entropy estimator. This is a scalar function of $\hat{\boldsymbol{p}}$.

h Hurst parameter for fractional Gaussian motion and fractional Gaussian noise.

$\boldsymbol{K}$ Covariance matrix.

σ^2 Variance.

ρ Normalized autocorrelation function.

λ Refers to the autocorrelation lag in the autocorrelation function $\rho(\lambda)$.

$\boldsymbol{R}$ Normalized autocorrelation matrix.

ϵ_t Gaussian error innovation with zero mean and variance σ.

$\tilde{p}$ Order of the autoregressive model.

$\tilde{q}$ Order of the moving average model.

ϕ_i Any parameters from an autoregressive model AR($\tilde{p}$), for index $i = 1, \ldots, \tilde{p}$.

$\boldsymbol{\phi}$ Vector of autoregressive model parameters for AR($\tilde{p}$).

θ_j Any parameters from an moving average model MA($\tilde{q}$), for index $j = 1, \ldots, \tilde{q}$.

$\boldsymbol{\theta}$ Vector of moving average model parameters for MA($\tilde{q}$).

H_c Composite MPE.

H_rc Refined composite MPE.

H_cd Composite downsampling PE.

H_rcd
Refined composite downsampling PE.

Appendix C

Acronym Glossary

This is a list of the most relevant acronyms used throughout this thesis.

pdf Probability density function.

pmf Probability mass function.

ApEn
Approximate entropy.

SampEn
Sample entropy.

MSEn
Multiscale entropy.

rMSE
Refined multiscale entropy.

cMSE
Composite multiscale entropy.

rcMSE
Refined composite multiscale entropy.

MPE
Multiscale permutation entropy.

cMPE
Composite multiscale permutation entropy.

rcMPE
Refined composite multiscale permutation entropy.

DPE
Downsampling multiscale permutation entropy.

cDPE
Composite downsampling multiscale permutation entropy.

rcDPE
Refined composite downsampling multiscale permutation entropy.

wGn
White Gaussian noise.

fbm Fractional Brownian motion.

fGn Fractional Gaussian noise.

cgfGn
Coarse-grained fractional Gaussian noise.

AR Autoregressive process.

MA Moving average process.

cgAR
Coarse-grained autoregressive process.

cgMA
Coarse-grained moving average process.

ARMA
Autoregressive and moving average process.

cgARMA
Coarse-Grained autoregressive and moving average process.

ECG
Electrocardiogram.

EEG
Electroencephalogram.

EMG
Electromyogram Electromyography.

sEMG
Surface electromyography.

MU Motor unit.

MF Muscle fibers.

MUAP
Motor unit action potential.

MVC
Maximum voluntary contraction.

Bibliography

[1] D. J. Little and D. M. Kane, "Permutation entropy of finite-length white-noise time series," *Physical Review. E*, vol. 94, no. 2, p. 022 118, Aug. 2016, ISSN: 2470-0053.

[2] A. Dávalos, M. Jabloun, P. Ravier, and O. Buttelli, "Theoretical study of multiscale permutation entropy on Finite-Length fractional gaussian noise," *26th European Signal Processing Conference (EUSIPCO)*, pp. 1092–1096, Sep. 2018.

[3] ——, "On the statistical properties of multiscale permutation entropy: Characterization of the estimator's variance," *Entropy*, vol. 21, no. 5, p. 450, May 2019.

[4] A. Dávalos, M. Jabloun, P. Ravier, and O. Buttelli, "Multiscale permutation entropy: Statistical characterization on autoregressive and moving average processes," in *2019 27th European Signal Processing Conference (EUSIPCO)*, Sep. 2019, pp. 1–5.

[5] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1, 1948, ISSN: 0005-8580.

[6] R. Clausius, *The Mechanical Theory of Heat: With Its Applications to the Steam-engine and to the Physical Properties of Bodies*, en. J. Van Voorst, 1867.

[7] M. Costa, A. Goldberg, and C. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E 71, 021906*, 2005.

[8] Y. Yin and P. Shang, "Weighted multiscale permutation entropy of financial time series," *Nonlinear Dynamics*, vol. 78, no. 4, pp. 2921–2939, Dec. 1, 2014, ISSN: 0924-090X, 1573-269X.

[9] L. Zunino, A. F. Bariviera, M. B. Guercio, L. B. Martinez, and O. A. Rosso, "Monitoring the informational efficiency of european corporate bond markets with dynamical permutation min-entropy," *Physica A: Statistical Mechanics and its Applications*, vol. 456, pp. 1–9, Aug. 15, 2016, ISSN: 0378-4371.

[10] A. Humeau-Heurtier, "The multiscale entropy algorithm and its variants: A review," *Entropy*, vol. 17, no. 5, pp. 3110–3123, May 12, 2015.

[11] D. MacKay, *Information Theory, Pattern Recognition and Neural Networks: The Book*. Cambridge University Press, 2005.

[12] H. Havrda and F. Chavrat, "Quantification method of classification processes," *Kybernetika*, vol. 3, no. 1, May 1967.

[13] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," en, *Journal of Statistical Physics*, vol. 52, no. 1, pp. 479–487, Jul. 1988, ISSN: 1572-9613.

[14] M. P. Leubner, "A nonextensive entropy approach to kappa-distributions," *Astrophysics and Space Science*, vol. 282, no. 3, pp. 573–579, Nov. 1, 2002, ISSN: 1572-946X.

[15] G. Livadiotis, "Introduction to special section on origins and properties of kappa distributions: Statistical background and properties of kappa distributions in space plasmas," *Journal of Geophysical Research (Space Physics)*, vol. 120, pp. 1607–1619, Mar. 1, 2015, ISSN: 0148-0227.

[16] A. R. Plastino and A. Plastino, "Stellar polytropes and tsallis' entropy," *Physics Letters A*, vol. 174, no. 5, pp. 384–386, Mar. 22, 1993, ISSN: 0375-9601.

[17] M. Portes de Albuquerque, I. A. Esquef, A. R. Gesualdi Mello, and M. Portes de Albuquerque, "Image thresholding using Tsallis entropy," *Pattern Recognition Letters*, vol. 25, no. 9, pp. 1059–1065, Jul. 2, 2004, ISSN: 0167-8655.

[18] N. Cvejic, C. Canagarajah, and D. Bull, "Image fusion metric based on mutual information and tsallis entropy," *Electronics Letters*, vol. 42, no. 11, pp. 626–627, May 2006, ISSN: 0013-5194.

[19] A. Renyi, "On the measures of entropy and information," *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability*, pp. 547–561, 1961.

[20] E. Beadle, J. Schroeder, B. Moran, and S. Suvorova, "An overview of renyi entropy and some potential applications," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, ISSN: 1058-6393, Oct. 2008, pp. 1698–1704.

[21] D. Lake, "Renyi entropy measures of heart rate gaussianity," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 1, pp. 21–27, Jan. 2006, ISSN: 0018-9294, 1558-2531.

[22] J. M. Amigó, S. G. Balogh, and S. Hernández, "A brief review of generalized entropies," *Entropy*, vol. 20, no. 11, p. 813, Nov. 2018.

[23] S. M. Pincus, "Approximate entropy as a measure of system complexity.," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, Mar. 15, 1991, ISSN: 0027-8424, 1091-6490.

[24] C. Cantrell, *Modern Mathematical Methods for Physicists and Engineers*. Cambridge University Press, 2000, ISBN: 0521598273.

[25] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, H2039–H2049, Jun. 1, 2000, ISSN: 0363-6135.

[26] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, p. 068 102, Jul. 19, 2002.

[27] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, p. 174 102, Apr. 11, 2002.

[28] H. Azami and J. Escudero, "Amplitude-aware Permutation entropy: Illustration in Spike Detection and Signal Segmentation," *Computer Methods and Programs in Biomedicine*, vol. 128, Feb. 2016.

[29] A. M. Kowalski, M. T. Martín, A. Plastino, and O. A. Rosso, "Bandt–pompe approach to the classical-quantum transition," *Physica D: Nonlinear Phenomena*, vol. 233, no. 1, pp. 21–31, Sep. 1, 2007, ISSN: 0167-2789.

[30] L. Zunino, M. Zanin, B. M. Tabak, D. G. Pérez, and O. A. Rosso, "Forbidden patterns, permutation entropy and stock market inefficiency," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 14, pp. 2854–2864, Jul. 15, 2009, ISSN: 0378-4371.

[31] X. Zhang, Y. Liang, J. Zhou, and Y. zang, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, Jun. 1, 2015, ISSN: 0263-2241.

[32] F. C. Morabito, D. Labate, F. La Foresta, A. Bramanti, G. Morabito, and I. Palamara, "Multivariate multi-scale permutation entropy for complexity analysis of alzheimer's disease EEG," *Entropy*, vol. 14, no. 7, pp. 1186–1202, Jul. 4, 2012.

[33] M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, "Permutation entropy and its main biomedical and econophysics applications: A review," *Entropy*, vol. 14, no. 8, pp. 1553–1577, Aug. 23, 2012.

[34] I. Ahmad, *Introduction to Applied Fuzzy Electronics*. Englewood Cliffs, NJ, USA: Prentice Hall, 1997.

[35] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory," en, *Information and Control*, vol. 20, no. 4, pp. 301–312, May 1972, ISSN: 0019-9958.

[36] B. Kosko, "Fuzzy entropy and conditioning," *Information Sciences*, vol. 40, no. 2, pp. 165–174, Dec. 1986, ISSN: 0020-0255.

[37] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 2, pp. 266–272, Jun. 2007, ISSN: 1534-4320.

[38] H. Azami, P. Li, S. E. Arnold, J. Escudero, and A. Humeau-Heurtier, "Fuzzy entropy metrics for the analysis of biomedical signals: Assessment and comparison," *IEEE Access*, vol. 7, pp. 104 833–104 847, 2019, ISSN: 2169-3536.

[39] H. Azami and J. Escudero, "Improved multiscale permutation entropy for biomedical signal analysis: Interpretation and application to electroencephalogram recordings," *Biomedical Signal Processing and Control*, vol. 23, pp. 28–41, 2016.

[40] J. F. Valencia, A. Porta, M. Vallverdu, F. Claria, R. Baranowski, E. Orlowska-Baranowska, and P. Caminal, "Refined Multiscale entropy: Application to 24-h Holter Recordings of Heart Period Variability in Healthy and Aortic Stenosis Subjects," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 9, pp. 2202–2213, Sep. 2009, ISSN: 0018-9294, 1558-2531.

[41] S.-D. Wu, C.-W. Wu, S.-G. Lin, C.-C. Wang, and K.-Y. Lee, "Time series analysis using composite multiscale entropy," *Entropy*, vol. 15, no. 3, pp. 1069–1084, Mar. 18, 2013.

[42] S.-D. Wu, C.-W. Wu, S.-G. Lin, K.-Y. Lee, and C.-K. Peng, "Analysis of complex time series using refined composite multiscale entropy," *Physics Letters A*, vol. 378, no. 20, pp. 1369–1374, Apr. 4, 2014, ISSN: 0375-9601.

[43] S.-D. Wu, C.-W. Wu, K.-Y. Lee, and S.-G. Lin, "Modified multiscale entropy for short-term time series analysis," en, *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 23, pp. 5865–5873, Dec. 2013, ISSN: 0378-4371.

[44] M. D. Costa and A. L. Goldberger, "Generalized multiscale entropy analysis: Application to quantifying the complex volatility of human heartbeat time series," *Entropy*, vol. 17, no. 3, pp. 1197–1203, Mar. 12, 2015.

[45] W. Aziz and M. Arif, "Multiscale permutation entropy of physiological time series," in *2005 Pakistan Section Multitopic Conference*, Dec. 2005, pp. 1–6.

[46] L. Zunino, D. G. Pérez, M. T. Martín, M. Garavaglia, A. Plastino, and O. A. Rosso, "Permutation entropy of fractional brownian motion and fractional gaussian noise," *Physics Letters A*, vol. 372, no. 27, pp. 4768–4774, Jun. 30, 2008, ISSN: 0375-9601.

[47] M. Matilla-García, "A non-parametric test for independence based on symbolic dynamics," *Journal of Economic Dynamics and Control*, vol. 31, no. 12, pp. 3889–3903, Dec. 1, 2007, ISSN: 0165-1889.

[48] J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán, "True and false forbidden patterns in deterministic and random dynamics," *EPL (Europhysics Letters)*, vol. 79, no. 5, p. 50 001, 2007, ISSN: 0295-5075.

[49] L. Zunino, D. G. Pérez, A. Kowalski, M. T. Martín, M. Garavaglia, A. Plastino, and O. A. Rosso, "Fractional brownian motion, fractional gaussian noise, and tsallis permutation entropy," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 24, pp. 6057–6068, Oct. 15, 2008, ISSN: 0378-4371.

[50] H. Wang, "On the number of successes in independent trials," *Statistica Sinica*, no. 3, pp. 71–84, 1993.

[51] K. Teerapabolarn, "An improvement of poisson approximation for sums of dependent bernoulli random variables," *Communications in Statistics - Theory and Methods*, vol. 43, no. 8, pp. 1758–1777, Apr. 18, 2014, ISSN: 0361-0926.

[52] L. Hörmander, *Linear Partial Differential Operators*, ser. Grundlehren der mathematischen Wissenschaften. Berlin Heidelberg: Springer-Verlag, 1963, ISBN: 978-3-642-46177-4.

[53] R. Horn and C. Johnson, *Matrix Analysis, 2nd Ed.* Cambridge University Press, 2013.

[54] R. C. Geary, "Review of mathematical methods of statistics," *The Economic Journal*, vol. 57, no. 226, in collab. with H. Cramér, pp. 200–202, 1947, ISSN: 0013-0133.

[55] B. Friedlander and J. M. Francos, "Estimation of amplitude and phase parameters of multicomponent signals," *IEEE Transactions on Signal Processing*, vol. 43, no. 4, pp. 917–926, Apr. 1995.

[56] K. S. Miller, "On the inverse of the sum of matrices," *Mathematics Magazine*, vol. 54, no. 2, pp. 67–72, 1981, ISSN: 0025-570X.

[57] A. Humeau-Heurtier, C.-W. Wu, and S.-D. Wu, "Refined composite multiscale permutation entropy to overcome multiscale permutation entropy length dependence - IEEE journals & magazine," *IEEE Signal Processing Letters*, vol. 22, pp. 2364–2367, 12 2015.

[58] C. Bandt and B. Shiha, "Order patterns in time series," *Journal of Time Series Analysis - Wiley Online Library*, 2007.

[59] K. Keller, M. Sinn, and J. Emonds, "Time series from the ordinal viewpoint," *Stochastics and Dynamics*, vol. 07, no. 2, pp. 247–272, Jun. 1, 2007, ISSN: 0219-4937.

[60] B. Mandelbrot and J. Van Ness, "Fractional brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, Oct. 1, 1968, ISSN: 0036-1445.

[61] D. Delignières, "Correlation properties of (discrete) fractional gaussian noise and fractional brownian motion," Mathematical Problems in Engineering, *Mathematical Problems in Engineering - Hindawi*, vol. 2015, 2015.

[62] P. Brockwell and R. Davis, *Time Series Analysis, Forecasting and Control. Revised Edition*. Holden-Day, Jan. 1, 1976, 575 pp.

[63] S. Berger, A. Kravtsiv, G. Schneider, and D. Jordan, "Teaching ordinal patterns to a computer: Efficient encoding algorithms based on the lehmer code," *Entropy*, vol. 21, no. 10, p. 1023, Oct. 2019.

[64] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley, "Fractal dynamics in physiology: Alterations with disease and aging," en, *Proceedings of the National Academy of Sciences*, vol. 99, no. suppl 1, pp. 2466–2472, Feb. 2002, ISSN: 0027-8424, 1091-6490.

[65] J. M. Beggs and D. Plenz, "Neuronal Avalanches in Neocortical Circuits," en, *Journal of Neuroscience*, vol. 23, no. 35, pp. 11 167–11 177, Dec. 2003, ISSN: 0270-6474, 1529-2401.

[66] G. Dragoi and S. Tonegawa, "Preplay of future place cell sequences by hippocampal cellular assemblies," en, *Nature*, vol. 469, no. 7330, pp. 397–401, Jan. 2011, ISSN: 1476-4687.

[67] K. Schindler, H. Gast, L. Stieglitz, A. Stibal, M. Hauf, R. Wiest, L. Mariani, and C. Rummel, "Forbidden ordinal patterns of periictal intracranial EEG indicate deterministic dynamics in human epileptic seizures," en, *Epilepsia*, vol. 52, no. 10, pp. 1771–1780, 2011, ISSN: 1528-1167.

[68] E. G. T. Liddell and C. S. Sherrington, "Recruitment and some other features of reflex inhibition," *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, vol. 97, no. 686, pp. 488–518, Apr. 1, 1925.

[69] H. Piper, *Elektrophysiologie menschlicher Muskeln*. Berlin: J. Springer, 1912.

[70] S. Bercier, R. Halin, P. Ravier, J.-F. Kahn, J.-C. Jouanin, A.-M. Lecoq, and O. Buttelli, "The vastus lateralis neuromuscular activity during all-out cycling exercise," *Journal of Electromyography and Kinesiology*, vol. 19, no. 5, pp. 922–930, Oct. 1, 2009, ISSN: 1050-6411.

[71] P. Madeleine, A. Samani, A. T. Binderup, and A. K. Stensdotter, "Changes in the spatio-temporal organization of the trapezius muscle activity in response to eccentric contractions," *Scandinavian Journal of Medicine & Science in Sports*, vol. 21, no. 2, pp. 277–286, 2011, ISSN: 1600-0838.

[72] M. Hakonen, H. Piitulainen, and A. Visala, "Current state of digital signal processing in myoelectric interfaces and related applications," *Biomedical Signal Processing and Control*, vol. 18, pp. 334–359, Apr. 1, 2015, ISSN: 1746-8094.

[73] M. Dimitriou, "Enhanced muscle afferent signals during motor learning in humans," *Current Biology*, vol. 26, no. 8, pp. 1062–1068, Apr. 25, 2016, ISSN: 0960-9822.

[74] R. Parry, O. Buttelli, J. Riff, J. Roussel, N. Sellam, M. L. Welter, and E. Lalo, "Rethinking gait and motor activity in daily life: A neuroergonomic perspective of Parkinson's disease," fr, *Le travail humain*, vol. Vol. 80, no. 1, pp. 23–50, Apr. 2017, ISSN: 0041-1868.

[75] A. L. Goldberger, C.-K. Peng, and L. A. Lipsitz, "What is physiologic complexity and how does it change with aging and disease?" *Neurobiology of Aging*, vol. 23, no. 1, pp. 23–26, Jan. 1, 2002, ISSN: 0197-4580, 1558-1497.

[76] M. Navaneethakrishna and S. Ramakrishnan, "Multiscale feature based analysis of surface EMG signals under fatigue and non-fatigue conditions," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, ISSN: 1558-4615, Aug. 2014, pp. 4627–4630.

[77] T. W. Boonstra, L. Faes, J. N. Kerkman, and D. Marinazzo, "Information decomposition of multichannel EMG to map functional interactions in the distributed motor system," *NeuroImage*, vol. 202, p. 116 093, Nov. 15, 2019, ISSN: 1053-8119.

[78] R. Merletti and P. Parker, *Electromyography: Physiology, Engineering, and Non-Invasive Applications*. IEEE Press Series in Biomedical Engineering, 2004.

[79] D. G. Sale, "Neural adaptation to strength training," in *Strength and Power in Sport*, John Wiley & Sons, Ltd, 2008, pp. 281–314, ISBN: 978-0-470-75721-5.

[80] A. F. Huxley and R. Niedergerke, "Structural changes in muscle during contraction: Interference microscopy of living muscle fibres," *Nature*, vol. 173, no. 4412, pp. 971–973, May 1954, ISSN: 1476-4687.

[81] J. Kimura, *Electrodiagnosis in Diseases of Nerve and Muscle: Principles and Practice*. OUP USA, Oct. 2013, 1177 pp., ISBN: 978-0-19-973868-7.

[82] R. M. Enoka and J. Duchateau, "Rate coding and the control of muscle force," *Cold Spring Harbor Perspectives in Medicine*, vol. 7, no. 10, a029702, Jan. 10, 2017, ISSN: , 2157-1422.

[83] J. Basmajian and de Luca C.J., *Muscles Alive: their functions revealed by electromyography, 5th Edition*. Williams and Wilkins, 2017.

[84] A. L. Goldberger, "Non-linear dynamics for clinicians: Chaos theory, fractals, and complexity at the bedside," *The Lancet - Elsevier*, vol. 347, no. 9011, pp. 1312–1314, May 11, 1996, ISSN: 0140-6736, 1474-547X.

[85] H. Miyano and T. Sadoyama, "Theoretical analysis of surface EMG in voluntary isometric contraction," *European Journal of Applied Physiology and Occupational Physiology*, vol. 40, no. 3, pp. 155–164, Sep. 1, 1979, ISSN: 1439-6327.

[86] J. G. A. Cashaback, T. Cluff, and J. R. Potvin, "Muscle fatigue and contraction intensity modulates the complexity of surface electromyography," *Journal of Electromyography and Kinesiology*, vol. 23, no. 1, pp. 78–83, Feb. 1, 2013, ISSN: 1050-6411.

[87] P. Ravier, O. Buttelli, R. Jennane, and P. Couratier, "An EMG fractal indicator having different sensitivities to changes in force and muscle fatigue during voluntary static muscle contractions," *Journal of Electromyography and Kinesiology*, vol. 15, no. 2, pp. 210–221, Apr. 1, 2005, ISSN: 1050-6411.

[88] Y. Liu, Y. Lin, J. Wang, and P. Shang, "Refined generalized multiscale entropy analysis for physiological signals," *Physica A: Statistical Mechanics and its Applications*, vol. 490, pp. 975–985, Jan. 15, 2018, ISSN: 0378-4371.

[89] H. J. Hermens, B. Freriks, R. Merletti, D. Stegeman, J. Blok, G. Rau, C. Disselhorst-Klug, and G. Hägg, "European recommendations for surface ElectroMyoGraphy," p. 4,

[90] C. De Luca, "Myoelectrical manifestations of localized muscular fatigue in humans.," *Critical Reviews in Biomedical Engineering*, vol. 11, no. 4, pp. 251–279, Jan. 1984, ISSN: 0278-940X, 1943-619X.

[91] R. Merletti and L. R. Lo Conte, "Surface EMG signal processing during isometric contractions," en, *Journal of Electromyography and Kinesiology*, vol. 7, no. 4, pp. 241–250, Dec. 1997, ISSN: 1050-6411.

[92] C. G. Kukulka and H. P. Clamann, "Comparison of the recruitment and discharge properties of motor units in human brachial biceps and adductor pollicis during isometric contractions," *Brain Research*, vol. 219, no. 1, pp. 45–55, Aug. 1981, ISSN: 0006-8993.

[93] J. Hogrel, "Use of surface EMG for studying motor unit recruitment during isometric linear force ramp," en, *Journal of Electromyography and Kinesiology*, vol. 13, no. 5, pp. 417–423, Oct. 2003, ISSN: 1050-6411.

[94] M. Brown and A. Forsythe, "Robust tests for equality of variances," *Journal of the American Statistical Association - Taylor and Francis Group*, vol. 69, pp. 364–367, 1974.

[95] C. J. De Luca, "The use of surface electromyography in biomechanics," *Journal of Applied Biomechanics*, vol. 13, no. 2, pp. 135–163, May 1, 1997, ISSN: 1065-8483.

[96] A. J. Fuglevand, D. A. Winter, and A. E. Patla, "Models of recruitment and rate coding organization in motor-unit pools," *Journal of Neurophysiology*, vol. 70, no. 6, pp. 2470–2488, Dec. 1, 1993, ISSN: 0022-3077.

[97] D. A. Gabriel, S. M. Lester, S. A. Lenhardt, and E. D. J. Cambridge, "Analysis of surface EMG spike shape across different levels of isometric force," *Journal of Neuroscience Methods*, vol. 159, no. 1, pp. 146–152, Jan. 15, 2007, ISSN: 0165-0270.

[98] A. Holtermann, C. Grönlund, J. S. Karlsson, and K. Roeleveld, "Motor unit synchronization during fatigue: Described with a novel sEMG method based on large motor unit samples," *Journal of Electromyography and Kinesiology*, vol. 19, no. 2, pp. 232–241, Apr. 1, 2009, ISSN: 1050-6411.

[99] A. Luttmann, *Physiological basis and concepts of electromyography*. Taylor and Francis Group, Nov. 13, 2017.

[100] A. Bingham, S. P. Arjunan, B. Jelfs, and D. K. Kumar, "Normalised mutual information of high-density surface electromyography during muscle fatigue," *Entropy*, vol. 19, no. 12, p. 697, Dec. 2017.

Antonio DÁVALOS

Sur les Propriétés Statistiques de l'Entropie de Permutation Multi-échelle et ses Raffinements; applications sur les Signaux Électromyographiques de Surface

L'entropie de permutation (PE) et l'entropie de permutation multi-échelle (MPE) sont largement utilisées pour mesurer la régularité dans l'analyse des séries temporelles, en particulier dans le contexte des signaux biomédicaux. Comme la précision est cruciale pour les chercheurs afin d'obtenir des interprétations optimales, il devient de plus en plus important de prendre en compte les propriétés statistiques de l'EMP.

C'est pourquoi, dans le présent travail, nous commençons par développer la théorie statistique qui sous-tend l'EMT, en mettant l'accent sur la caractérisation de ses deux premiers moments dans le contexte de la multi-échelle. Ensuite, nous explorons les versions composites de MPE afin de comprendre les propriétés sous-jacentes à l'amélioration de leurs performances ; nous avons également créé un point de référence d'entropie par le calcul des valeurs attendues de MPE pour les processus stochastiques gaussiens largement utilisés, puisque cela nous donne un point de référence à utiliser avec de vrais signaux biomédicaux. Enfin, nous différencions la dynamique de l'activité musculaire dans les contractions isométriques par l'application des méthodes MPE classique et composite sur des données électromyographiques de surface (sEMG).

À la suite de notre projet, nous avons constaté que l'EPM est une statistique biaisée qui diminue par rapport au facteur multidimensionnel, quelle que soit la distribution de probabilité du signal. Nous avons également remarqué que la variance de la statistique MPE dépend fortement de la valeur de la MPE elle-même, et est presque égale à sa limite inférieure Cram'er-Rao - en d'autres termes, confirmant qu'il s'agit d'un estimateur efficace. Malgré l'amélioration des résultats, nous avons réalisé que les versions composites modifient également l'estimation de l'EPP en raison de la mesure d'informations redondantes. À la lumière de nos conclusions, nous avons décidé de remplacer la procédure de grossier calibrage à plusieurs échelles par une de nos propres procédures, dans l'intention d'améliorer nos estimations.

Comme notre équipe a observé que la statistique de l'EMT était entièrement caractérisée par les paramètres du modèle lorsqu'elle était appliquée à des modèles gaussiens corrélés, nous avons développé une formulation générale de l'EMT attendue avec des dimensions à faible encombrement. Lorsqu'elle est appliquée à des signaux sEMG réels, nous avons été en mesure de distinguer les états de fatigue et de non-fatigue avec toutes les méthodes, en particulier pour les dimensions à haute imbrication. De plus, nous avons constaté que la méthode MPE que nous proposons fait une différence encore plus nette entre les deux états d'activité susmentionnés.

Mots clés : Entropie de Permutation Multi-échelle,Électromyographie,Traitement de Signaux,Statistique,

On the Statistical Properties of Multiscale Permutation Entropy and its Refinements, with Applications on Surface Electromyographic Signals

Permutation entropy (PE) and multiscale permutation entropy (MPE) are extensively used to measure regularity in the analysis of time series, particularly in the context of biomedical signals. As accuracy is crucial for researchers to obtain optimal interpretations, it becomes increasingly important to take into account the statistical properties of MPE.

Therefore, in the present work we begin by expanding on the statistical theory behind MPE, with an emphasis on the characterization of its first two moments in the context of multiscaling. Secondly, we explore the composite versions of MPE in order to understand the underlying properties behind their improved performance; we also created an entropy benchmark through the calculation of MPE expected values for widely used Gaussian stochastic processes, since that gives us a reference point to use with real biomedical signals. Finally, we differentiate between muscle activity dynamics in isometric contractions through the application of the classical and composite MPE methods on surface electromyographic (sEMG) data.

As a result of our project, we found MPE to be a biased statistic that decreases with respect to the multiscaling factor, regardless of the signal's probability distribution. We also noticed that the variance of the MPE statistic is highly dependent on the value of MPE itself, and almost equal to its Cram´er-Rao lower bound -in other words, confirming it is an efficient estimator. Despite showing improved results, we realized that the composite

versions also modify the MPE estimation due to the measuring of redundant information. In light of our findings, we decided to replace the multiscaling coarse-graining procedure with one of our own, with the intention of improving our estimations.

Since our team observed the MPE statistic to be completely characterized by the model parameters when applied to correlated Gaussian models, we developed a general formulation for expected MPE with low-embedding dimensions. When applied to real sEMG signals, we were able to distinguish between fatigue and non-fatigue states with all methods, especially for high-embedding dimensions. Moreover, we found that our proposed MPE method makes an even clearer difference between the two aforementioned activity states.

Keywords : Multiscale Permutation Entropy, Electromyography, Signal Processing, Statistics.

Laboratoire PRISME
Polytech site Galilée
12 Rue de Blois
45100, Orléans