Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Special Issue on Unsolved Problems of Noise in Physics, Biology and Technology

Towards an information-theoretic model of the Allison mixture stochastic process

Lachlan J Gunn[1], **François Chapeau-Blondeau**[2], **Andrew Allison**[1] **and Derek Abbott**[1]

[1] School of Electrical and Electronic Engineering, The University of Adelaide, SA 5005, Australia
[2] Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS), University of Angers, 62 avenue Notre Dame du Lac, 49000 Angers, France
E-mail: lachlan.gunn@adelaide.edu.au, chapeau@univ-angers.fr, andrew.allison@adelaide.edu.au and derek.abbott@adelaide.edu.au

Received 16 September 2015
Accepted for publication 17 November 2015
Published 20 May 2016

Online at stacks.iop.org/JSTAT/2016/054041
doi:10.1088/1742-5468/2016/05/054041

Abstract. The Allison mixture is a random process formed by stochastically switching between two random and uncorrelated input processes. Unintuitively, these samples—independent prior to being drawn—can acquire dependence as a result of the sampling process. It has previously been shown that correlation can occur subject to certain conditions, however in general dependence does not imply correlation. In this paper we provide an initial information-theoretic analysis of the Allison mixture, and derive the autoinformation function of its sampling process as the first step towards a fuller information-theoretic analysis of its output.

J. Stat. Mech. (2016) 054041

1742-5468/16/054041+8$33.00

Contents

I. Introduction 2

II. The Allison mixture 2

III. Autoinformation of the Allison mixture sampling process 5

IV. Open questions 7

Acknowledgments 8

References 8

I. Introduction

The Allison mixture [1, 2], its name coined by Epstein in [3], is a stochastic process formed by random sampling of two parent processes, having applications in such diverse fields as natural language processing [1] and physics [2]. It was originally introduced as a model for word repetition intervals [1], the distance between consecutive uses of a word. Concretely, it models the phenomenon by which word frequency changes throughout a text according to the focus of the narrative. That is to say, if a character or object is present in a scene then it will be mentioned quite often, and thus its name will have a short repetition interval. At other times the name may be only rarely referred to, and thus its repetition interval will be large. This was modelled in [1] by a time-dependent Poisson process, whose rate parameter varies randomly according to a two-state Markov chain.

The Allison mixture is of interest in thermodynamics because of its connection to Brownian ratchets and the Parrondo paradox [4, 5], in which losing games played in combination provide a net gain. Viewing the mixing process from a thermodynamic perspective, one sees that the irreversible nature of the mixing process causes a loss of information and ultimately redundancy in the output process.

This unintuitive feature of the Allison mixture results in the appearance of autocorrelation, despite all of its values being drawn from uncorrelated processes. However, this correlation vanishes [2] if the parent processes are of equal mean, suggesting the use of autoinformation [6, 7] as an alternative to correlation, providing a canonical measure of the strength of the memory of the Allison mixture. We apply this measure to the Allison mixture, producing analytic expressions for the k-step autoinformation of its sampling process.

II. The Allison mixture

The Allison mixture [2] is a process whose samples are drawn from one of two distributions, the choice determined by the state of a Markov chain [8], shown in figure 1. The

doi:10.1088/1742-5468/2016/05/054041

J. Stat. Mech. (2016) 054041

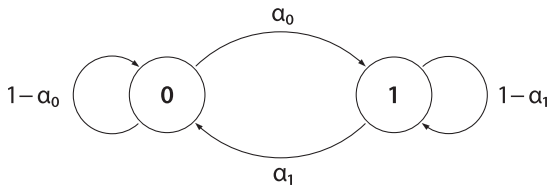

Figure 1. The Markov chain defining the sampling process S_t of the Allison mixture. It is parametrised by the probabilities α_0 and α_1 of leaving states 0 and 1 respectively.

marginal distribution of this process is a mixture of the two source distributions, the mixing constant determined by the stationary distribution of the Markov chain.

Definition II.1 (Allison mixture [2]). An Allison mixture X_t of two processes U_t and V_t is given by

$$X_t = S_t U_t + (1 - S_t) V_t \tag{1}$$

where the sampling process S_t is a Markov chain, shown in figure 1, having states $\{0, 1\}$ and transition probabilities α_0 and α_1 when in states 0 and 1 respectively. The stationary distribution of S_t is given by [2]

$$\begin{aligned} \pi_0 &= \frac{\alpha_1}{\alpha_0 + \alpha_1} \\ \pi_1 &= \frac{\alpha_0}{\alpha_0 + \alpha_1}. \end{aligned} \tag{2}$$

Knowing the means of the input processes U_t and V_t, the single-step autocovariance is given by the following theorem, given by [2] and which we restate without proof:

Theorem II.1. The Allison mixture X_t associated with a fully mixed sampling process S_t as in figure 1 has a lag-one autocovariance of

$$R_{XX}(1) = (E[U] - E[V])^2 \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2} (1 - \alpha_1 - \alpha_2). \tag{3}$$

In order to extend this to the case of arbitrary lags, we require the n-step transition probabilities $\alpha_{0,k}$ and $\alpha_{1,k}$; we use a spectral decomposition of the transition matrix P in order to compute the k-step probability matrix P^k from which we can read the transition probabilities on the minor diagonal.

Theorem II.2. *The sampling process S_t has k-step transition probabilities*

$$\alpha_{0,k} = \pi_1 [1 - (1 - \alpha_0 - \alpha_1)^k] \tag{4}$$

$$\alpha_{1,k} = \pi_0 [1 - (1 - \alpha_0 - \alpha_1)^k]. \tag{5}$$

Proof. S_t has transition matrix

$$\mathbb{P} = \begin{bmatrix} 1 - \alpha_0 & \alpha_1 \\ \alpha_0 & 1 - \alpha_1 \end{bmatrix} \tag{6}$$

with spectral decomposition

$$\mathbb{P} = \frac{1}{\alpha_0 + \alpha_1}\begin{bmatrix}\alpha_1 & 1\\ \alpha_0 & -1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1-\alpha_0-\alpha_1\end{bmatrix}\begin{bmatrix}1 & 1\\ \alpha_0 & -\alpha_1\end{bmatrix}. \tag{7}$$

We therefore find that

$$\mathbb{P}^k = \frac{1}{\alpha_0 + \alpha_1}\begin{bmatrix}\alpha_1 & 1\\ \alpha_0 & -1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & (1-\alpha_0-\alpha_1)^k\end{bmatrix}\begin{bmatrix}1 & 1\\ \alpha_0 & -\alpha_1\end{bmatrix} \tag{8}$$

$$= \frac{1}{\alpha_0 + \alpha_1}\begin{bmatrix}\alpha_1 + \alpha_0(1-\alpha_0-\alpha_1)^k & \alpha_1(1-(1-\alpha_0-\alpha_1))\\ \alpha_0(1-(1-\alpha_0-\alpha_1)) & \alpha_0 + \alpha_1(1-\alpha_0-\alpha_1)^k\end{bmatrix} \tag{9}$$

$$= \begin{bmatrix}\pi_0 + \pi_1(1-\alpha_0-\alpha_1)^k & \pi_0\left[1-(1-\alpha_0-\alpha_1)^k\right]\\ \pi_1\left[1-(1-\alpha_0-\alpha_1)^k\right] & \pi_1 + \pi_0(1-\alpha_0-\alpha_1)^k\end{bmatrix}, \tag{10}$$

and read off the stated transition probabilities from the minor diagonal. □

Knowing this, we may now calculate the autocorrelation function of the process at arbitrary time-lags.

Theorem II.3. The Allison mixture X_t has k-step autocovariance

$$R_{XX}[k] = R_{XX}[1](1-\alpha_0-\alpha_1)^{k-1}. \tag{11}$$

Proof. We begin by noting that as a decimated Markov chain—that is to say, one where all but every k-th step is discarded—is still a Markov chain, and that therefore a decimated Allison mixture is also an Allison mixture. We may therefore simply calculate arbitrary two-point statistics by simply substituting the k-step transition probabilities from theorem II.2 for α_0 and α_1.

Let $\gamma = (1-\alpha_0-\alpha_1)$. Then, performing the substitution as described,

$$R_{XX}[k] = (E[U]-E[V])^2\frac{\alpha_{0,k}\alpha_{1,k}}{(\alpha_{0,k}+\alpha_{1,k})^2}(1-\alpha_{0,k}-\alpha_{1,k}) \tag{12}$$

$$= (E[U]-E[V])^2\,\pi_0\pi_1\gamma^k \tag{13}$$

$$= (E[U]-E[V])^2\frac{\alpha_0\alpha_1}{(\alpha_0+\alpha_1)^2}(1-\alpha_0-\alpha_1)^k \tag{14}$$

$$= R_{XX}[1](1-\alpha_0-\alpha_1)^{k-1} \tag{15}$$

as originally stated. □

J. Stat. Mech. (2016) 054041

III. Autoinformation of the Allison mixture sampling process

Whereas the autocovariance function considers the covariance between samples from a process, the autoinformation function considers their mutual information. This is advantageous for our purposes, as a lack of correlation does not in general result in a lack of dependence.

The autoinformation function is thus an alternative to the autocovariance function as a measure of dependence—though not causality, which is to be the subject of a future paper—and is defined as follows:

Definition III.1 (Autoinformation function [7]). The autoinformation function of a stochastic process S_t is the mutual information [9]

$$I_{\mathrm{SS}}[t, k] = I(S_t, S_{t-k}) \tag{16}$$

$$= H(S_t) + H(S_{t-k}) - H(S_t, S_{t-k}). \tag{17}$$

If S_t is stationary, then we may omit t as a parameter, leaving us with

$$I_{\mathrm{SS}}[k] = I(S_t, S_{t-k}) \tag{18}$$

$$= 2H(S_t) - H(S_t, S_{t-k}). \tag{19}$$

The autoinformation improves on the autocovariance function by providing a condition both sufficient and necessary—whereas a lack of correlation does not necessarily indicate independence, two variables will have zero mutual information if and only if they are statistically independent; this is vital when the processes U_t and V_t of the system being modelled have identical means but differing variances or skew, such as would occur when sampling particle velocities in statistical mechanics.

Substituting the stationary and transition probabilities into the entropy, we find the single-step autoinformation, stated without further detail in the following lemma.

Lemma III.1. *Let S_t be a binary-valued random process with transition probabilities and a stationary distribution equal to that of the Markov chain in definition II.1. Then, in the fully-mixed regime—that is to say, when the state probability distribution is equal to the stationary distribution of the Markov chain—the single-step autoinformation $I_{\mathrm{SS}}[1]$ is given by*

$$I_{\mathrm{SS}}[1] = \frac{\alpha_1(1-\alpha_0)\log_2\frac{1-\alpha_0}{\alpha_1}}{\alpha_0+\alpha_1} + \frac{\alpha_0(1-\alpha_1)\log_2\frac{1-\alpha_1}{\alpha_0}}{\alpha_0+\alpha_1} + \log_2(\alpha_0+\alpha_1), \tag{20}$$

where both α_0 and α_1 are nonzero, zero if exactly one of α_0 and α_1 is equal to zero, and undefined if both are equal to zero.

J. Stat. Mech. (2016) 054041

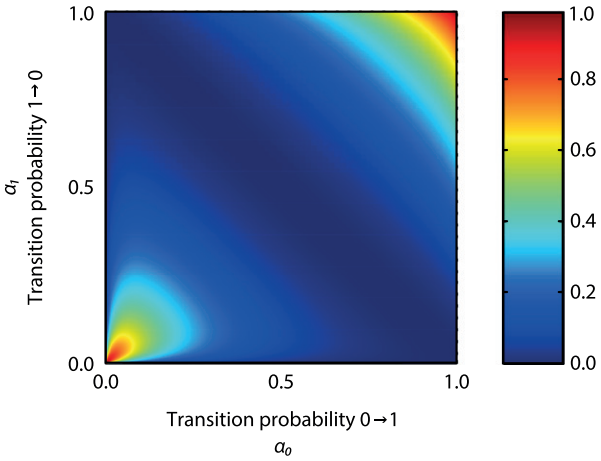

Figure 2. Single-step autoinformation $I_{SS}[1]$ of the Allison mixture sampling process S_t as a function of the transition probabilities α_0 and α_1, calculated according to equation (20). Note the lines of zero autoinformation along $\alpha_0 = 0$, $\alpha_1 = 0$, and $\alpha_0 + \alpha_1 = 1$.

Thus the autoinformation $I_{SS}[1]$ is equal to zero when either $\alpha_0 = 0$, $\alpha_1 = 0$, or $\alpha_0 + \alpha_1 = 1$, and so these three previously-described [2] conditions for decorrelation of the sampling process imply zero mutual information and therefore genuine independence.

Importantly, we have not assumed the Markov property of S_t, instead directly demanding that the formulae for the stationary probabilities hold. This weakening is intended to allow us later to generalise to the Allison mixture proper.

The mutual information $I_{SS}[1]$ as a function of (α_0, α_1) is shown in figure 2. As one would expect, we see a peak near $(\alpha_0, \alpha_1) = (0, 0)$, where consecutive states are highly dependent. Similarly, we see a large autoinformation $I_{SS}[1]$ near (1, 1), where the strong anticorrelation makes consecutive states highly predictable. Between these two extremes lies a valley, its nadir falling along the line $\alpha_0 + \alpha_1 = 1$; along this line, consecutive states of the sampling process are completely independent.

Importantly, these results can be generalised to allow calculation of the autoinformation at arbitrary time-lags, shown in theorem III.1 by substituting the k-step probabilities of the Allison mixture sampling process.

Theorem III.1. *The k-step autoinformation of a fully mixed two-state Markov chain with exit probabilities α_0 and α_1, as in figure 1, is given by lemma III.1 under the substitution*

$$\alpha_0 \longrightarrow \pi_1 [1 - (1 - \alpha_0 - \alpha_1)^k] \tag{21}$$

$$\alpha_1 \longrightarrow \pi_0 [1 - (1 - \alpha_0 - \alpha_1)^k] \,. \tag{22}$$

Proof. By substituting the k-step transition probabilities, calculated in equations (4) and (5), in place of the single-step probabilities α_0 and α_1, equation (20) yields the k-step autoinformation $I_{SS}[k]$ rather than the single-step autoinformation $I_{SS}[1]$. □

J. Stat. Mech. (2016) 054041

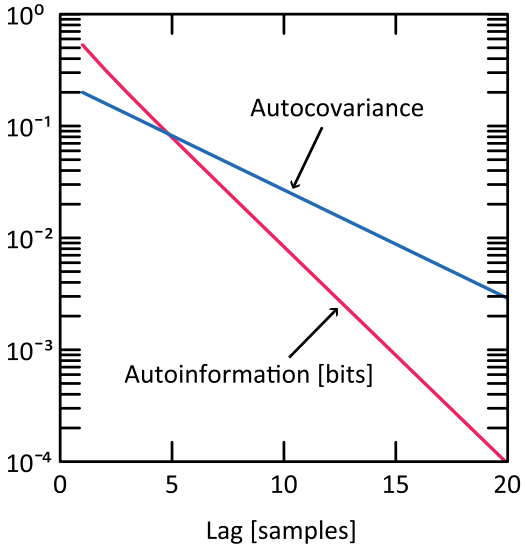

Figure 3. The exponentially-decaying autoinformation $I_{\mathrm{SS}}[k]$ and autocovariance $R_{\mathrm{SS}}[k]$ of an Allison mixture sampling process with $\alpha_0 = 0.1, \alpha_1 = 0.1$. The slope of the autoinformation line is approximately double that of the autocorrelation line; the results of [7] hint that this may be exactly so asymptotically.

We show the autoinformation $I_{\mathrm{SS}}[k]$ in figure 3 as a function of the lag k; it can be seen to decay at a roughly exponential rate.

IV. Open questions

The theorems that we have presented allow computation of the autoinformation function of the Allison mixture sampling process S_t, and can be readily extended to binary-valued Allison mixtures, that is to say those for which X_t takes only two values; the input processes U_t and V_t might each take a single distinct value, or perhaps a common pair of values. However, many physical systems are described by continuous-valued processes, and their autoinformation cannot be calculated by lemma III.1 due to the infinite number of values that they may take. It remains to be seen whether the autoinformation $I_{\mathrm{XX}}[k]$ of the Allison mixture X_t can be computed by transformation of the sampling process autoinformation $I_{\mathrm{SS}}[k]$ in a similar fashion to that of the autocovariance function [2], potentially yielding a more practically-manipulated alternative to the rather cumbersome formulae that can be derived by manual calculation of mixture transition probabilities to be substituted into equation (20). There exists also the possibility that simpifying approximations will be possible in the large-k regime to allow further comparison of its properties with those of the autocorrelation $R_{\mathrm{XX}}[k]$.

Furthermore, the information-theoretic approach that we have presented provides the starting point for an investigation of the transfer entropy [10] between the sampling process and the Allison mixture; previous works on transfer entropy have focussed on complex systems, leaving room for the analysis of simpler and analytically tractable models in order to better probe its properties.

J. Stat. Mech. (2016) 054041

Acknowledgment

L J Gunn is currently the recipient of an Endeavour Research Fellowship from the Australian Government, undertaking research at the University of Angers, France.

References

[1] Allison A G, Pearce C E M and Abbott D 2007 Finding keywords amongst noise: automatic text classification without parsing *Proc. SPIE* **6601** 660113
[2] Gunn L J, Allison A and Abbott D 2014 Allison mixtures: where random digits obey thermodynamic principles *Int. J. Mod. Phys.* **33** 1460360
[3] Epstein R A 2009 *The Theory of Gambling and Statistical Logic* 2nd edn (New York: Academic)
[4] Harmer G P and Abbott D 1999 Game theory: losing strategies can win by Parrondo's paradox *Nature* **402** 864
[5] Abbott D 2010 Asymmetry and disorder: a decade of Parrondo's paradox *Fluctuation Noise Lett.* **9** 129–56
[6] Darbellay G A and Wuertz D 2000 The entropy as a tool for analysing statistical dependences in financial time series *Physica* A **287** 429–39
[7] Chapeau-Blondeau F 2007 Autocorrelation versus entropy-based autoinformation for measuring dependence in random signal *Physica* A **380** 1–18
[8] Goodman R 2006 *Introduction to Stochastic Models* (New York: Dover)
[9] Cover T M and Thomas J A 2006 *Elements of Information Theory* 2nd edn (New York: Wiley)
[10] Schreiber T 2000 Measuring information transfer *Phys. Rev. Lett.* **85** 461–4