Weighted entropy and its use in industrial management and financial engineering

Yuri Suhov

Math Dept, Penn State (USA)

DPMMS, University of Cambridge (UK)

May 8, 2015, Williamsport

This is an ongoing work, with several participants:

Gerald Frizelle (University of Cambridge, UK),

Mark Kelbert (The Higher School of Economics, Moscow, RF),

Izabella Stuhl (University of Denver, USA),

Salimeh Yasaei Sekeh (Federal University of Sao Carlos, SP, Brazil), and myself.

We all know: the entropy of a probability distribution $\mathbf{p}$ is

$$h(\mathbf{p}) = -\sum p(x_i) \log\, p(x_i). \tag{1}$$

It does not depend on the nature of outcomes x_i,

We all know: the entropy of a probability distribution $\mathbf{p}$ is

$$h(\mathbf{p}) = -\sum p(x_i) \log\, p(x_i). \tag{1}$$

It does not depend on the nature of outcomes x_i, only upon probabilities $p(x_i)$. (One says $h(\mathbf{p})$ is context-free.)

We all know: the entropy of a probability distribution $\mathbf{p}$ is

$$h(\mathbf{p}) = -\sum p(x_i) \log\ p(x_i). \tag{1}$$

It does not depend on the nature of outcomes x_i, only upon probabilities $p(x_i)$. (One says $h(\mathbf{p})$ is context-free.)

Entropy is useful in a number of areas (each time we deal with information/uncertainty/complexity/variability of trajectories).

We all know: the entropy of a probability distribution $\mathbf{p}$ is

$$h(\mathbf{p}) = -\sum p(x_i)\log\, p(x_i). \tag{1}$$

It does not depend on the nature of outcomes x_i, only upon probabilities $p(x_i)$. (One says $h(\mathbf{p})$ is context-free.)

Entropy is useful in a number of areas (each time we deal with information/uncertainty/complexity/variability of trajectories).

However, in many cases the nature of outcomes x_i is important.

So, we may modify the definition making it context-dependent:

We all know: the entropy of a probability distribution $\mathbf{p}$ is

$$h(\mathbf{p}) = -\sum p(x_i) \log\, p(x_i). \tag{1}$$

It does not depend on the nature of outcomes x_i, only upon probabilities $p(x_i)$. (One says $h(\mathbf{p})$ is context-free.)

Entropy is useful in a number of areas (each time we deal with information/uncertainty/complexity/variability of trajectories).

However, in many cases the nature of outcomes x_i is important.

So, we may modify the definition making it context-dependent:

$$h^{\rm w}_{\phi}(\mathbf{p}) = -\sum \phi(x_i) p(x_i) \log\, p(x_i). \tag{2}$$

Here function $x_i \mapsto \phi(x_i)$ represents a *weight* of an outcome x_i,

Here function $x_i \mapsto \phi(x_i)$ represents a *weight* of an outcome x_i, i.e., its ‘value‘ for us.

Here function $x_i \mapsto \phi(x_i)$ represents a *weight* of an outcome x_i,

i.e., its ‘value‘ for us.

I‘ll focus mainly non-negative weight functions $\phi(x) \geq 0$,

Here function $x_i \mapsto \phi(x_i)$ represents a *weight* of an outcome x_i, i.e., its ‘value‘ for us.

I‘ll focus mainly non-negative weight functions $\phi(x) \geq 0$, but my fantasy suggests that negative values may also be of some use,

Here function $x_i \mapsto \phi(x_i)$ represents a *weight* of an outcome x_i,

i.e., its ‘value‘ for us.

I‘ll focus mainly non-negative weight functions $\phi(x) \geq 0$,

but my fantasy suggests that negative values may also be

of some use, both theoretically (analytically) and in applications.

One example is where $\phi(x) = \mathbf{1}(x \in A)$

Here function $x_i \mapsto \phi(x_i)$ represents a *weight* of an outcome x_i, i.e., its ‘value‘ for us.

I‘ll focus mainly non-negative weight functions $\phi(x) \geq 0$, but my fantasy suggests that negative values may also be of some use, both theoretically (analytically) and in applications.

One example is where $\phi(x) = \mathbf{1}(x \in A)$ where A is a particular sub-collection of outcomes.

Here function $x_i \mapsto \phi(x_i)$ represents a *weight* of an outcome x_i,

i.e., its 'value' for us.

I'll focus mainly non-negative weight functions $\phi(x) \geq 0$,

but my fantasy suggests that negative values may also be

of some use, both theoretically (analytically) and in applications.

One example is where $\phi(x) = \mathbf{1}(x \in A)$ where A is a particular

sub-collection of outcomes. Here

$$h^{\rm w}_{\phi}(\mathbf{p}) = -\sum_{x \in A} \phi(x_i) p(x_i) \log\, p(x_i) \tag{3}$$

i.e., we disregard the information coming from outcomes $x \notin A$.

Quantity $h^{\mathrm{w}}_{\phi}(\mathbf{p})$ in Eqn (2) is called the *weighted entropy* of probability distribution $\mathbf{p}$ with weight function ϕ.

Quantity $h^{\rm w}_{\phi}(\mathbf{p})$ in Eqn (2) is called the *weighted entropy* of probability distribution $\mathbf{p}$ with weight function ϕ.

It was proposed in late 1960's and analyzed in a handful of papers, Quantity $h^{\mathrm{w}}_{\phi}(\mathbf{p})$ in Eqn (2) is called the *weighted entropy* of probability distribution **p** with weight function ϕ.

It was proposed in late 1960's and analyzed in a handful of papers, but largely was left unnoticed. In the form of (3) we used it, on an *ad hoc* basis, in our papers

G. Frizelle and Y. M. Suhov. An entropic measurement of queueing behaviour in a class of manufacturing operations. *Proc. Royal Soc. A*, **457** (2001), 1579–1601

and

G. Frizelle and Y. M. Suhov. The measurement of complexity in production and other commercial systems. *Proc. Royal Soc. A*, **464** (2008), 2649–2668.

We analyzed data from a variety of industrial/commercial systems and numerically calculated the value of the weighted entropy $h^{\mathrm{w}}_{\phi}(\mathbf{p})$, with $\phi = \underline{1}(\cdot \in A)$, for various parts of a production process.

and

G. Frizelle and Y. M. Suhov. The measurement of complexity in production and other commercial systems. *Proc. Royal Soc. A*, **464** (2008), 2649–2668.

We analyzed data from a variety of industrial/commercial systems and numerically calculated the value of the weighted entropy $h^{\rm w}_{\phi}(\mathbf{p})$, with $\phi = \underline{1}(\,\cdot\, \in A)$, for various parts of a production process. We had to make a judicial choice of set A. It helped to determine a **bottleneck** that slowed down the whole system,

and

G. Frizelle and Y. M. Suhov. The measurement of complexity in production and other commercial systems. *Proc. Royal Soc. A*, **464** (2008), 2649–2668.

We analyzed data from a variety of industrial/commercial systems and numerically calculated the value of the weighted entropy $h^{\rm w}_\phi(\mathbf{p})$, with $\phi = \underline{1}(\,\cdot\,\in A)$, for various parts of a production process. We had to make a judicial choice of set A. It helped to determine a **bottleneck** that slowed down the whole system, which was difficult to identify by other methods.

A similar formalism can be proposed for differential entropy

A similar formalism can be proposed for differential entropy when we work with a probability density function (PDF):

$$h^{\rm w}_\phi(f) = -\int \phi(x) f(x) \log\, f(x) {\rm d}x. \tag{4}$$

A similar formalism can be proposed for differential entropy when we work with a probability density function (PDF):

$$h^{\rm w}_{\phi}(f) = -\int \phi(x) f(x) \log\, f(x) {\rm d}x. \tag{4}$$

E.g., take $f(x) = f^{\rm No}_{\mathbf{C}}(x)$ where $f^{\rm No}_{\mathbf{C}}(x)$ is a multi-variate Gaussian/normal PDF with mean 0 and covariance matrix $\mathbf{C}$:

$$f^{\rm No}_{\mathbf{C}}(x) = \frac{\exp\left[-x^{\rm T}\mathbf{C}^{-1}x/2\right]}{\left[(2\pi)^d \det \mathbf{C}\,\right]^{1/2}},\ x \in \mathbb{R}^d.$$

Then the weighted entropy can be expressed as

$$h^{\rm w}_\phi(f^{\rm No}_{\mathbf C}) = \frac{\alpha(\mathbf C)}{2}\log\Big[(2\pi)^d(\det\mathbf C)\Big] + \frac{\log e}{2}{\rm tr}\,\mathbf C^{-1}\mathbf\Phi_{\mathbf C,\phi}. \qquad (5)$$

Here $\alpha(\mathbf C)=\alpha_\phi(\mathbf C)>0$ and a positive-definite matrix $\mathbf\Phi_{\mathbf C}$ $=\mathbf\Phi_{\mathbf C,\phi}$ are given by

$$\alpha_\phi(\mathbf C)=\int\limits_{\mathbb R^d}\phi(x)f^{\rm No}_{\mathbf C}(x){\rm d}x,\;\; \mathbf\Phi_{\mathbf C,\phi}=\int\limits_{\mathbb R^d}x\,x^{\rm T}\phi(x)f^{\rm No}_{\mathbf C}(x){\rm d}x. \qquad (6)$$

For $\phi(x)\equiv 1$, we get back to the standard differential entropy:

Then the weighted entropy can be expressed as

$$h^{\mathrm{w}}_{\phi}(f^{\mathrm{No}}_{\mathbf{C}}) = \frac{\alpha(\mathbf{C})}{2}\log\Big[(2\pi)^d(\det\mathbf{C})\Big] + \frac{\log e}{2}\mathrm{tr}\,\mathbf{C}^{-1}\mathbf{\Phi}_{\mathbf{C},\phi}. \qquad (5)$$

Here $\alpha(\mathbf{C}) = \alpha_{\phi}(\mathbf{C}) > 0$ and a positive-definite matrix $\mathbf{\Phi}_{\mathbf{C}}$ $=\mathbf{\Phi}_{\mathbf{C},\phi}$ are given by

$$\alpha_{\phi}(\mathbf{C}) = \int\limits_{\mathbb{R}^d}\phi(x)f^{\mathrm{No}}_{\mathbf{C}}(x)\mathrm{d}x, \quad \mathbf{\Phi}_{\mathbf{C},\phi} = \int\limits_{\mathbb{R}^d} x\,x^{\mathrm{T}}\phi(x)f^{\mathrm{No}}_{\mathbf{C}}(x)\mathrm{d}x. \qquad (6)$$

For $\phi(x)\equiv 1$, we get back to the standard differential entropy:

$$h(f^{\mathrm{No}}_{\mathbf{C}}) = \frac{1}{2}\log\Big[(2\pi)^d(\det\mathbf{C})\Big] + \frac{d\log e}{2}. \qquad (7)$$

It is well-known that $\mathbf{C} \mapsto \delta(\mathbf{C}) := \log \det \mathbf{C}$ is a concave function of a (strictly) positive-definite matrix $\mathbf{C}$:

It is well-known that $\mathbf{C} \mapsto \delta(\mathbf{C}) := \log \det \mathbf{C}$ is a concave function of a (strictly) positive-definite matrix $\mathbf{C}$:

$$\begin{aligned}&\delta(\lambda_1\mathbf{C}_1 + \lambda_2\mathbf{C}_2) \geq \lambda_1\delta(\mathbf{C}_1) + \lambda_2\delta(\mathbf{C}_2);\\ &\quad\text{equivalently,}\quad h(f^{\rm No}_{\lambda_1\mathbf{C}_1+\lambda_2\mathbf{C}_2}) \geq \lambda_1 h(f^{\rm No}_{\mathbf{C}_1}) + \lambda_2 h(f^{\rm No}_{\mathbf{C}_2}).\end{aligned} \tag{8}$$

It is well-known that $\mathbf{C} \mapsto \delta(\mathbf{C}) := \log \det \mathbf{C}$ is a concave function of a (strictly) positive-definite matrix $\mathbf{C}$:

$$\begin{aligned} &\delta(\lambda_1 \mathbf{C}_1 + \lambda_2 \mathbf{C}_2) \geq \lambda_1 \delta(\mathbf{C}_1) + \lambda_2 \delta(\mathbf{C}_2); \\ &\qquad \text{equivalently,} \quad h(f^{\rm No}_{\lambda_1 \mathbf{C}_1 + \lambda_2 \mathbf{C}_2}) \geq \lambda_1 h(f^{\rm No}_{\mathbf{C}_1}) + \lambda_2 h(f^{\rm No}_{\mathbf{C}_2}). \end{aligned} \tag{8}$$

This is an example from a (famous) series of Ky Fan inequalities. In the context of entropy, (8) is a simple corollary of the fact that $h(f) := -\int f(x) \log f(x) \mathrm{d}x$ is maximised at $f = f^{\rm No}_{\mathbf{C}}$.

It is well-known that $\mathbf{C} \mapsto \delta(\mathbf{C}) := \log \det \mathbf{C}$ is a concave function of a (strictly) positive-definite matrix $\mathbf{C}$:

$$\begin{aligned}&\delta(\lambda_1\mathbf{C}_1 + \lambda_2\mathbf{C}_2) \geq \lambda_1\delta(\mathbf{C}_1) + \lambda_2\delta(\mathbf{C}_2);\\ &\qquad\text{equivalently,}\quad h(f^{\mathrm{No}}_{\lambda_1\mathbf{C}_1+\lambda_2\mathbf{C}_2}) \geq \lambda_1 h(f^{\mathrm{No}}_{\mathbf{C}_1}) + \lambda_2 h(f^{\mathrm{No}}_{\mathbf{C}_2}).\end{aligned} \tag{8}$$

This is an example from a (famous) series of Ky Fan inequalities. In the context of entropy, (8) is a simple corollary of the fact that $h(f) := -\int f(x)\log f(x)\mathrm{d}x$ is maximised at $f = f^{\mathrm{No}}_{\mathbf{C}}$. But a similar property can be verified for the weighted entropy.

It is well-known that $\mathbf{C} \mapsto \delta(\mathbf{C}) := \log \det \mathbf{C}$ is a concave function of a (strictly) positive-definite matrix $\mathbf{C}$:

$$\delta(\lambda_1\mathbf{C}_1 + \lambda_2\mathbf{C}_2) \geq \lambda_1\delta(\mathbf{C}_1) + \lambda_2\delta(\mathbf{C}_2); \\ \text{equivalently,} \quad h(f^{\mathrm{No}}_{\lambda_1\mathbf{C}_1+\lambda_2\mathbf{C}_2}) \geq \lambda_1 h(f^{\mathrm{No}}_{\mathbf{C}_1}) + \lambda_2 h(f^{\mathrm{No}}_{\mathbf{C}_2}). \tag{8}$$

This is an example from a (famous) series of Ky Fan inequalities. In the context of entropy, (8) is a simple corollary of the fact that $h(f) := -\int f(x)\log f(x)\mathrm{d}x$ is maximised at $f = f^{\mathrm{No}}_{\mathbf{C}}$. But a similar property can be verified for the weighted entropy. This immediately leads to a **new** inequality for positive-definite matrices: under some assumptions about weight function ϕ,

$$h^{\rm w}_\phi(f^{\rm No}_{\lambda_1\mathbf{C}_1+\lambda_2\mathbf{C}_2}) \geq \lambda_1 h^{\rm w}_\phi(f^{\rm No}_{\mathbf{C}_1}) + \lambda_2 h^{\rm w}_\phi(f^{\rm No}_{\mathbf{C}_2}). \tag{9}$$

Inequality (9) becomes more explicit when the weight function is $\phi(x) = \exp\left(x^{\rm T} t\right)$ (parametrized by $t \in \mathbb{R}^d$). Namely:

$$h^{\rm w}_\phi(f^{\rm No}_{\mathbf{C}}) = h(f^{\rm No}_{\mathbf{C}}) \exp\left(\frac{1}{2} t^{\rm T}\mathbf{C}t\right). \tag{10}$$

Introduce the set $\mathbb{S} = \mathbb{S}(\mathbf{C}_1, \mathbf{C}_2; \lambda_1, \lambda_2) \subset \mathbb{R}^d$:

$$\mathbb{S} = \left\{ t \in \mathbb{R}^d : \ F^{(1)}(t) \geq 0, \ \text{ and } \ F^{(2)}(t) \leq 0 \right\}. \tag{11}$$

Here

$$F^{(1)}(t) = \sum_{i=1,2} \lambda_i \exp\left(\frac{1}{2} t^{\mathrm{T}} \mathbf{C}_\alpha t\right) - \exp\left(\frac{1}{2} t^{\mathrm{T}} \mathbf{C} t\right)$$

and

$$F^{(2)}(t) = \left[\sum_{i=1,2} \lambda_i \exp\left(\frac{1}{2} t^{\mathrm{T}} \mathbf{C}_i t\right) - \exp\left(\frac{1}{2} t^{\mathrm{T}} \mathbf{C} t\right)\right]$$
$$\times \log\left[(2\pi)^d (\det \mathbf{C})\right]$$
$$+ \sum_{i=1,2} \lambda_i \exp\left(\frac{1}{2} t^{\mathrm{T}} \mathbf{C}_i t\right) \mathrm{tr}\left[\mathbf{C}^{-1} \mathbf{C}_i\right] - d \exp\left(\frac{1}{2} t^{\mathrm{T}} \mathbf{C} t\right).$$

Theorem 1. *Given positive definite matrices* $\mathbf{C}_1$, $\mathbf{C}_2$ *and* $\lambda_1, \lambda_2 \in [0,1]$ *with* $\lambda_1 + \lambda_2 = 1$, *set* $\mathbf{C} = \lambda_1\mathbf{C}_1 + \lambda_2\mathbf{C}_2$. *Assume that* $t \in \mathbb{S}$. *Then*

$$h(f^{\rm No}_{\mathbf{C}})\exp\left(\frac{1}{2}t^{\rm T}\mathbf{C}t\right) \\ -\lambda_1 h(f^{\rm No}_{\mathbf{C}_1})\exp\left(\frac{1}{2}t^{\rm T}\mathbf{C}_1 t\right) - \lambda_2 h(f^{\rm No}_{\mathbf{C}_1})\exp\left(\frac{1}{2}t^{\rm T}\mathbf{C}_2 t\right) \geq 0; \tag{12}$$

equality iff $\lambda_1\lambda_2 = 0$ *or* $\mathbf{C}_1 = \mathbf{C}_2$.

This is a previously unknown series of bounds of a Ky Fan type.

Theorem 1. *Given positive definite matrices* $\mathbf{C}_1$, $\mathbf{C}_2$ *and* $\lambda_1, \lambda_2 \in [0,1]$ *with* $\lambda_1 + \lambda_2 = 1$, *set* $\mathbf{C} = \lambda_1 \mathbf{C}_1 + \lambda_2 \mathbf{C}_2$. *Assume that* $t \in \mathbb{S}$. *Then*

$$h(f^{\rm No}_{\mathbf{C}}) \exp\left(\frac{1}{2} t^{\rm T} \mathbf{C} t\right) \\ -\lambda_1 h(f^{\rm No}_{\mathbf{C}_1}) \exp\left(\frac{1}{2} t^{\rm T} \mathbf{C}_1 t\right) - \lambda_2 h(f^{\rm No}_{\mathbf{C}_1}) \exp\left(\frac{1}{2} t^{\rm T} \mathbf{C}_2 t\right) \geq 0; \tag{12}$$

equality iff $\lambda_1 \lambda_2 = 0$ *or* $\mathbf{C}_1 = \mathbf{C}_2$.

This is a previously unknown series of bounds of a Ky Fan type.

For $t = 0$ it becomes the standard Ky Fan bound (8).

We recently produced a host of inequalities involving determinants of positive-definite matrices; they all are proven with the help of weighted entropies. See

We recently produced a host of inequalities involving determinants of positive-definite matrices; they all are proven with the help of weighted entropies. See

Y. Suhov, S.Yasaei Sekeh. Simple inequalities for weighted entropies. arXiv:1409.4102v3, 2014

Y. Suhov, S.Yasaei Sekeh, I.Stuhl. Weighted Gaussian entropy and determinant inequalities. arXiv:1505.01437v1, 2015

My last example will be from financial engineering.

More precisely, I will use a background of betting.

-12-

You are betting on results ϵ_n of subsequent random trials, $n = 1, 2, \ldots$.

You are betting on results ϵ_n of subsequent random trials, $n = 1, 2, \ldots$. Each ϵ_n produces a value x_n, say, $x_n \in \mathbb{R}^d$. We suppose that a random string $\underline{\epsilon}_1^n = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$ has a joint PDF $f_n(\underline{x}_1^n)$. A conditional PDF $\mathtt{f}_n(x_n|\underline{x}_1^{n-1})$ will be also used, with

$$\mathtt{f}_n(x_n|\underline{x}_1^{n-1}) f_{n-1}(\underline{x}_1^{n-1}) = f_n(\underline{x}_1^n), \quad \int_{\mathbb{R}^d} \mathtt{f}_n(x_n|\underline{x}_1^{n-1}) \mathrm{d}x_n = 1.$$

Let us agree that if you stake \$ C_n on game n you win \$ $C_n g_n(x_n)$ if the result is $x_n \in \mathbb{R}^d$. (So, you make a profit when $C_n g_n(x_n) > 0$ and incur a loss when $C_n g_n(x_n) < 0$.) Here g_n are given real-valued functions $x_n \in \mathbb{R}^d \mapsto g_n(x_n) \in \mathbb{R}$. We say that g_n are return functions.

Let $Z_0 > 0$ be your initial capital. More generally, given $n \geq 1$, denote by $Z_{n-1} > 0$ your fortune after $n-1$ trials and impose the restriction that variable $C_n = C_n(Z_0; \underline{\epsilon}_1^{n-1})$ depends on Z_0 and $\underline{\epsilon}_1^{n-1}$ but not on ϵ_n.

(One says that C_n is a previsible strategy.)

Then $Z_{n-1} = Z_{n-1}(Z_0, \underline{\epsilon}_1^{n-1})$. It also makes sense to require that $C_n \geq 0$. One also may demand that $-C_n g_n(x_n) \leq Z_{n-1}$.

(In applications, this is required to guarantee the deposit.)

We have the recursion

$$Z_n = Z_{n-1} + C_n g_n(\epsilon_n) = Z_{n-1}\left(1 + \frac{C_n g_n(\epsilon_n)}{Z_{n-1}}\right) \qquad (13)$$

and wish to maximize $\mathbb{E}S_N$ where

$$S_N := \sum_{j=1}^{N} \phi_j(\epsilon_j; \underline{\epsilon}_1^{j-1}) \log \frac{Z_j}{Z_{j-1}}. \qquad (14)$$

Here the weight function (WF) $\underline{x}_1^j \mapsto \phi_j(x_j; \underline{x}_1^{j-1})$ depends on $\underline{x}_j$ and the vector $\underline{x}_1^{j-1} = \begin{pmatrix} x_1 \\ \vdots \\ x_{j-1} \end{pmatrix}$. Quantity $\phi_j(x_j; \underline{x}_1^{j-1})$ represents a 'sentimental' value of outcome x_n (given that it succeeds a sequence $\underline{x}_1^{j-1}$) taken into account when one calculates S_N.

Value $\mathbb{E}S_N$ is the weighted expected interest rate after N rounds of investment.

When $\phi \equiv 1$, the sum (14) becomes telescopic and equal to $\log \dfrac{Z_N}{Z_0}$, the standard interest rate. Recursion (13) suggests a martingale-based approach.

We also consider a sequence of positive functions $b_n(x_n)$, $x_n \in \mathbb{R}_n$, figuring in Eqns (15) – (16) below. More precisely, we will use the following conditions.

$$\int_{\mathbb{R}^d} \phi_n(x_n;\underline{\epsilon}_1^{n-1}) b_n(x_n) g_n(x_n) \mathrm{d}x_n = 0, \tag{15}$$

and

$$\begin{aligned}\int_{\mathbb{R}^d} \phi_n(x_n;\underline{\epsilon}_1^{n-1}) b_n(x_n) \mathrm{d}x_n \\ \leq \int_{\mathbb{R}_n} \phi_n(x_n;\underline{\epsilon}_1^{n-1}) \mathtt{f}_n(x_n|\underline{\epsilon}_1^{n-1}) \mathrm{d}\mu_n(x_n),\end{aligned} \tag{16}$$

Next, define a RV $\alpha_n = \alpha_n(\underline{\epsilon}_1^{n-1})$ by

$$\alpha_n = \int_{\mathbb{R}_n} \phi_n(x_n;\underline{\epsilon}_1^{n-1}) \mathtt{f}_n(x_n|\underline{\epsilon}_1^{n-1}) \log \frac{\mathtt{f}_n(x_n|\underline{\epsilon}_1^{n-1})}{b_n(x_n)} \mathrm{d}x_n. \tag{17}$$

Theorem 2. *Given* $1 < N \leq \infty$, *assume conditions* (15), (16) *for* $1 \leq n < N$. *Then:*

(A) *For all previsible* C_n *such that* $1 + \dfrac{C_n g(\epsilon_n)}{Z_{n-1}} > 0$, *sequence* $S_n - \mathrm{A}_n$ *is a supermartingale, where* $\mathrm{A}_n := \sum_{j=1}^{n} \alpha_j$. *Consequently,*

$$\mathbb{E}\, S_n \leq \sum_{j=1}^{n} \mathbb{E}\, \alpha_j.$$

(B) *Sequence* $S_n - \mathrm{A}_n$, $1 \le n < N$, *is a martingale for some previsible* C_n *satisfying* $0 \le C_n \le Z_{n-1}$ *and* $1 + \dfrac{C_n g_n(\epsilon_n)}{Z_{n-1}} > 0$ *iff the following holds. There exists a function*

$$\underline{x}_1^{n-1} \mapsto D_{n-1}(\underline{x}_1^{n-1}) \in [0,1] \text{ with } 1 + D_n(\underline{\epsilon}_1^{n-1}) g_n(\epsilon_n) > 0$$

such that

$$\mathtt{f}_n(x_n|\underline{x}_1^{n-1}) = g_n(x_n) b_n(x_n) D_{n-1}(\underline{x}_1^{n-1}) + b_n(x_n). \tag{18}$$

In this case

$$C_n(\underline{\epsilon}_1^{n-1}) = D_{n-1}(\underline{\epsilon}_1^{n-1}) Z_{n-1}(\underline{\epsilon}_1^{n-1}). \tag{20}$$

E.g., assume that trials ϵ_n are IID, and each trial produces one of $m > 1$ outcomes $E_1,\ldots, E_m \in \mathbb{R}$ with probabilities $p_1,\ldots, p_m > 0$. We also set the return function $g_n(E_i) = E_i$ and use uniform probabilities to emulate functions b_n: $b_n(E_i) = \dfrac{1}{m}$.
Here if you stake \$ C_n on game n you win \$ C_nE_i if the result is E_i. As above, let $Z_{n-1} > 0$ be the fortune after $n-1$ trials ($Z_0 > 0$ is the initial capital). As before, let $\mathfrak{F}_n = \sigma(Z_0)$ and $\mathfrak{F}_n = \sigma(Z_0, \underline{\epsilon}_1^n)$, $n \geq 1$, and consider a sequence of RVs C_n where C_n is $\mathfrak{F}_{n-1}$-measurable (a previsible strategy).
Recursion (13) becomes

$$Z_n = Z_{n-1} + C_n \epsilon_n = Z_{n-1}\left(1 + \frac{\epsilon_n C_n}{Z_{n-1}}\right).$$

We wish to maximize, in C_n, the weighted expected interest rate $\mathbb{E}S_N$ where

$$S_n := \sum_{j=1}^{n} \phi(\epsilon_j) \log \frac{Z_j}{Z_{j-1}}.$$

Here $E \mapsto \phi(E) \geq 0$ is a weight function (for simplicity depending only upon a one-time outcome).

Theorem 2 then takes the following form:

Theorem 3. *Suppose that*

$$\sum_i \phi(E_i)E_i = 0 \ \ \text{and} \ \ \frac{1}{m}\sum_i \phi(E_i) \leq \sum_i \phi(E_i)p_i.$$

Set:

$$\alpha = \sum_i \phi(E_i)p_i \log(p_i m).$$

Then

(A) *For all previsible* C_n *with* $1 + \dfrac{\epsilon_n C_n}{Z_{n-1}} > 0$*, sequence* $S_n - \alpha n$ *is a supermartingale; consequently,* $\mathbb{E}\, S_n \leq n\alpha$*.*

(B) $S_n - \alpha n$ *is a martingale for a previsible* C_n *with* $0 \le C_n \le Z_{n-1}$ *and* $1 + \dfrac{\epsilon_n C_n}{Z_{n-1}} > 0$ *iff* $D := \dfrac{mp_i - 1}{E_i}$ *is a non-negative number between* 0 *and* 1 *which does not depend upon outcome* E_i, *and*

$$C_n = DZ_n.$$

In case $m = 2$, the above martingale strategy exists only if $E_1 = -E_2$ and $\phi(E_1) = \phi(E_2)$ (no weight preference). Assume for definiteness that $E_1 > 0$ and $p_1 \geq 1/2$. Then $D = \frac{2p_1 - 1}{E_1} = \frac{1 - 2p_1}{E_2}$, and the martingale strategy is

$$C_n = \frac{Z_{n-1}}{E_1}(2p_1 - 1).$$

It means that you repeatedly bet the proportion $\frac{2p_1 - 1}{E_1}$ of your current capital on outcome E_1. See

Y. Suhov, I. Stuhl, M. Kelbert. Weight functions and log-optimal investment portfolios. arXiv:1505.01417, 2015