Chapter 1

Entropy and Martingale

K. B. Athreya[1] and M. G. Nadkarni[2]

[1] *Department of Mathematics,*
Iowa State University, Ames, Iowa, 50011, USA
and
I. M. I, Department of Mathematics,
Indian Institute of Science, Bangalore, 560012, India
kba@iastate.edu

[2] *Department of Mathematics*
University of Mumbai, Kalina, Mumbai, 400098, India
mgnadkarni@gmail.com

1.1. Introduction

This article discusses the concepts of relative entropy of a probability measure with respect to a dominating measure and that of measure free martingales. There is considerable literature on the concepts of relative entropy and standard martingales, both separately and on connection between the two. This paper draws from results established in [1] (unpublished notes) and [6]. In [1] the concept of relative entropy and its maximization subject to a finite as well as infinite number of linear constraints is discussed. In [6] the notion of measure free martingale of a sequence $\{f_n\}_{n=1}^{\infty}$ of real valued functions with the restriction that each f_n takes only finitely many distinct values is introduced. Here is an outline of the paper.

In section 1.2 the concepts of relative entropy and Gibbs-Boltzmann measures, and a few results on the maximization of relative entropy and the weak convergence of of the Gibbs-Boltzmann measures are presented. We also settle in the negative a problem posed in [6]. In section 1.3 the notion of measure free martingale is generalized from the case of finitely many valued sequence $\{f_n\}_{n=1}^{\infty}$ to the general case where each f_n is allowed to

PERSPECTIVES IN MATHEMATICAL SCIENCES I - Probability and Statistics

http://www.worldscibooks.com/mathematics/7308.html

be a Borel function taking possibly uncountably many values. It is shown that every martingale is a measure free martingale, and conversely that every measure free martingale admits a finitely additive measure on a certain algebra under which it is a martingale. Conditions under which such a measure is countably additive are given. Last section is devoted to an ab initio discussion of the existence of an equivalent martingale measure and the uniqueness of such a measure if they are chosen to maximize certain relative entropies.

1.2. Relative Entropy and Gibbs-Boltzmann Measures

1.2.1. *Entropy Maximization Results*

Let $(\Omega, \mathcal{B}, \mu)$ be a measure space. A $\mathcal{B}$- measurable function $f : \Omega \to [0, \infty)$ is called a probability density function (p.d.f) with respect to μ if $\int_\Omega f(\omega)\mu(d\omega) = 1$. Then $\nu_f(A) = \int_A f(\omega)\mu(d\omega), A \in \mathcal{B}$, is a probability measure dominated by μ. The relative entropy of ν_f with respect to μ is defined as

$$H(f, \mu) = -\int_\Omega f(\omega) \log f(\omega)\mu(d\omega) \tag{1}$$

provided the integral on the right hand side is well defined, although it may possibly be infinite. In particular, if μ is a finite measure, this holds since the function $h(x) = x \log x$ is bounded on $(0, 1)$ and hence $\int_\Omega (-f(\omega) \log f(\omega))^+ \mu(d\omega) < \infty$. This does allow for the possibility that $H(f, \mu)$ could be $-\infty$ when $\mu(\Omega)$ is finite. We will show below that if $\mu(\Omega)$ is finite and positive then $H(f, \mu) \leq \log \mu(\Omega)$ for all p.d.f. f with respect to μ. In particular if $\mu(\Omega) = 1$, $H(f, \mu) \leq 0$.

We recall here for the benefit of the reader that a $\mathcal{B}$ measurable non-negative real valued function f always has a well defined integral with respect to μ. It is denoted by $\int_\Omega f(\omega)\mu(d\omega)$. The integral may be finite or infinite. A real valued $\mathcal{B}$ measurable function f can be written as $f = f_+ - f_-$, where, for each $\omega \in \Omega$,

$$f_+(\omega) = \max\{0, f(\omega)\}, f_-(\omega) = -\min\{0, f(\omega)\}.$$

If at least one of f_+, f_- has a finite integral, then we say that f has a well defined integral with respect to μ and write

$$\int_\Omega f(\omega)\mu(d\omega) = \int_\Omega f_+(\omega)\mu(d\omega) - \int_\Omega f_-(\omega)\mu(d\omega).$$

PERSPECTIVES IN MATHEMATICAL SCIENCES I - Probability and Statistics

Now note the simple fact from calculus. The function $\phi(x) = x-1-\log x$ on $(0, \infty)$ has a unique minimum at $x = 1$ and $\phi(1) = 0$. Thus for all $x > 0$, $\log x \leq x - 1$ with equality holding if and only if $x = 1$. So if f_1 and f_2 are two probability density functions on $(\Omega, \mathcal{B}, \mu)$, then for all ω,

$$f_1(\omega) \log f_2(\omega) - f_1(\omega) \log f_1(\omega) \leq f_2(\omega) - f_1(\omega), \tag{2}$$

with equality holding if and only if $f_1(\omega) = f_2(\omega)$. Assume now that $f_1(\omega) \log f_1(\omega)$, $f_1(\omega) \log f_2(\omega)$ have definite integrals with respect to μ and that one of them is finite. On integrating the two sides of (2) we get

$$\begin{aligned}
&\int_\Omega f_1(\omega) \log f_2(\omega)\mu(d\omega) - \int_\Omega f_1(\omega) \log f_1(\omega)\mu(d\omega) \\
&\leq \int_\Omega f_2(\omega)\mu(d\omega) - \int_\Omega f_1(\omega)\mu(d\omega) \\
&= 1 - 1 = 0.
\end{aligned}$$

The middle inequality becomes an equality if and only if equality holds in (2) a.e. with respect to μ. We have proved:

Proposition 2.1. *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and let f_1, f_2 be two probability density functions on $(\Omega, \mathcal{B}, \mu)$. Assume that the functions $f_1 \log f_1, f_1 \log f_2$ have definite integrals with respect to μ and that one of them is finite. Then*

$$H(f_1, \mu) = -\int_\Omega f_1(\omega) \log f_1(\omega)\mu(d\omega) \leq -\int_\Omega f_1(\omega) \log f_2(\omega)\mu(d\omega), \tag{3}$$

with equality holding if and only if $f_1(\omega) = f_2(\omega)$, a.e. μ.

Note that if $\mu(\Omega)$ is finite and positive and if we set $f_2(\omega) = (\mu(\Omega))^{-1}$, for all ω, then the right hand side of (3) becomes $\log \mu(\Omega)$ and we conclude that relative entropy of $H(f_1, \mu)$ is well defined and at most $\log \mu(\Omega)$.

Let f_0 be a probability density function on $(\Omega, \mathcal{B}, \mu)$ such that $\lambda \equiv H(f_0, \mu)$ is finite and let

$$\mathcal{F}_\lambda = \{f : f \textit{ a p.d.f. wrt } \mu \textit{ and } -\int_\Omega f(\omega) \log f_0(\omega)\mu(d\omega) = \lambda\}. \tag{4}$$

From Proposition 2.1 it follows that for any $f \in \mathcal{F}_\lambda$,

$$H(f,\mu) = -\int_\Omega f(\omega)\log f(\omega)\mu(d\omega) \le -\int_\Omega f(\omega)\log f_0(\omega)\mu(d\omega) = \lambda$$
$$= -\int_\Omega f_0(\omega)\log f_0(\omega)\mu(d\omega),$$

with equality holding if and only if $f = f_0,\ a.e.\ \mu$. We summarize this as:

Theorem 2.1. *Let $f_0, \lambda, \mathcal{F}_\lambda$, be as in (4) above. Then*

$$\sup\{H(f,\mu) : f \in \mathcal{F}_\lambda\} = H(f_0,\mu)$$

and f_0 is the unique maximiser.

Theorem 2.1. says that any probability density function f_0 with respect to μ with finite entropy relative to μ appears as the unique solution to an entropy maximizing problem in an appropriate class of probability density functions. Of course, this assertion has some meaning only if $\mathcal{F}_\lambda$ does not consist of f_0 alone. A useful reformulation of this result is as follows:

Theorem 2.2. *Let $h : \Omega \to \mathbb{R}$ be a $\mathcal{B}$ measurable function. Let c and λ be real numbers such that*

$$\psi(c) = \int_\Omega e^{ch(\omega)}\mu(d\omega) < \infty, \qquad \int_\Omega \mid h(\omega) \mid e^{ch(\omega)}\mu(d\omega) < \infty,\ \textit{and}$$

$$\lambda = \frac{\int_\Omega h(\omega)e^{ch(\omega)}\mu(d\omega)}{\psi(c)}.$$

Let $f_0 = \frac{e^{ch}}{\psi(c)}$ and let

$$\mathcal{F}_\lambda = \{f : f\ \textit{a p.d.f. wrt}\ \mu,\ \text{and}\ \int_\Omega f(\omega)h(\omega)\mu(d\omega) = \lambda\}.$$

Then

$$\sup\{H(f,\mu) : f \in \mathcal{F}_\lambda\} = H(f_0,\mu),$$

and f_0 is the unique maximiser.

Here are some sample examples of the above considerations.

Example 1. Let $\Omega = \{1, 2, \cdots, N\}, N < \infty$, μ counting measure on Ω, $h = 1, \lambda = 1$. Then $\mathcal{F}_\lambda = \{(p_i)_{i=1}^N, p_i \geq 0, \sum_{i=1}^N p_i = 1\}$. Then

$$f_0(j) = \frac{1}{N}, j = 1, 2, \cdots, N,$$

the uniform distribution on $\{1, 2, \cdots, N\}$, maximizes the relative entropy of the class $\mathcal{F}_\lambda$ with respect to μ.

Example 2. Let $\Omega = \mathbb{N}$, the natural numbers $\{1, 2, \cdots\}$, $\mu =$ the counting measure on Ω, $h(j) = j, j \in \mathbb{N}$. Fix $\lambda, 1 \leq \lambda < \infty$ and let

$$\mathcal{F}_\lambda = \{(p_j)_{j=1}^\infty : \forall\ j,\ p_j \geq 0, \sum_{j=1}^\infty p_j = 1, \sum_{j=1}^\infty j p_j = \lambda\}.$$

Then $f_0(j) = (1-p)p^{j-1}, j = 1, 2, \cdots$, where $p = 1 - \frac{1}{\lambda}$, maximizes the relative entropy of the class $\mathcal{F}_\lambda$ with respect to μ.

Example 3. Let $\Omega = \mathbb{R}$, $\mu =$ Lebesgue measure on $\mathbb{R}$, $h(x) = x^2$, $0 < \lambda < \infty$. Set $\mathcal{F}_\lambda = \{f : f \geq 0, \int_{\mathbb{R}} f(x)dx = 1, \int_{\mathbb{R}} x^2 f(x)dx = \lambda\}$. Then

$$f_0(x) = \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{x^2}{2\lambda}},$$

i.e., the Gaussian distribution with mean zero and variance λ, maximizes the relative entropy of the class $\mathcal{F}_\lambda$ with respect to the Lebesgue measure.

These examples are well known (see [5]) and the usual method is by the use of Lagrange's multipliers. The present method extends to the case of arbitrary number of constraints (see [1], [8]).

Definition 2.1. Let $(\Omega, \mathcal{B}, \mu)$ be a measure space. Let $h : \Omega \to \mathbb{R}$ be $\mathcal{B}$ measurable and let c be a real number. Let $\psi(c) = \int_\Omega e^{ch(\omega)} \mu(d\omega)$ be finite. Let

$$\nu_{(\mu,c,h)}(A) = \frac{\int_A e^{ch(\omega)} \mu(d\omega)}{\psi(c)}, A \in \mathcal{B}.$$

The probability measure $\nu_{(\mu,c,h)}$ is called the *Gibbs-Boltzmann measure* corresponding to (μ, c, h).

Example 4. (Spin system on N states.) Let $\Omega = \{-1, 1\}^N$, N a positive integer, and let $V : \Omega \to \mathbb{R}$, $0 < \beta_T < \infty$ be given. Let μ denote the counting measure on Ω. The measure

$$\nu_{(\mu,\beta_T,V)}(A) = \frac{\sum_{\omega \in A} e^{-\beta_T V(\omega)}}{\sum_{\omega \in \Omega} e^{-\beta_T V(\omega)}}, \qquad A \subset \Omega, \tag{5}$$

is called the *Gibbs distribution* with potential function V and temperature constant β_T for the spin system of N states. The denominator on the right side of (5) is known as the *partition function.*

1.2.2. *Weak Convergence of Gibbs-Boltzmann Distribution*

Let Ω be a Polish space, i.e., a complete separable metric space and let $\mathcal{B}$ denote its Borel σ-algebra. Recall that a sequence $(\mu_n)_{n=1}^{\infty}$ of probability measures on $(\Omega, \mathcal{B})$ converges weakly to a probability measure μ on $(\Omega, \mathcal{B})$, if

$$\int_{\Omega} f(\omega)\mu_n(d\omega) \to \int_{\Omega} f(\omega)\mu(d\omega)$$

for every continuous bounded function $f : \Omega \to \mathbb{R}$. Now let $(\mu_n)_{n=1}^{\infty}$ be a sequence of probability measures on $(\Omega, \mathcal{B})$, $(h_n)_{n=1}^{\infty}$ a sequence of $\mathcal{B}$ measurable functions from Ω to $\mathbb{R}$ and $(c_n)_{n=1}^{\infty}$ a sequence of real numbers. Assume that for each $n \geq 1$, $\int_{\Omega} e^{c_n h_n(\omega)}\mu(d\omega) < \infty$. For each $n \geq 1$, let $\nu_{(\mu_n, c_n, h_n)}$ be the Gibbs-Boltzmann measure corresponding (μ_n, c_n, h_n) as in definition 2.1. An important problem is to find conditions on $(\mu_n, h_n, c_n)_{n=1}^{\infty}$ so that $(\nu_{(\mu_n, c_n, h_n)})_{n=1}^{\infty}$ converges weakly. We address this question in a somewhat special context. We start with some preliminaries.

Let $C \subset \mathbb{R}$ be a compact subset and μ a probability measure on Borel subsets of $\mathbb{R}$ with support contained in C. For $c \in \mathbb{R}$, let

$$\psi(c) = \int_C e^{cx}\mu(dx).$$

Since C is bounded and μ is a probability measure, the function ψ is well defined and infinitely differentiable on $\mathbb{R}$. For any $k \geq 1$,

$$\psi^{(k)}(c) = \int_C e^{cx}x^k\mu(dx).$$

Note that the function $f_c(x) = \frac{e^{cx}}{\psi(c)}$ is a probability density function with respect to μ with mean $\phi(c) = \frac{\psi'(c)}{\psi(c)}$.

Proposition 2.2.

(i) *ϕ is infinitely differentiable and $\phi'(c) > 0$ for all $c \in \mathbb{R}$, provided μ is not supported on a single point. If μ is a Dirac measure, i.e., if μ is supported on a single point, then $\phi'(c) = 0$ for all c.*

PERSPECTIVES IN MATHEMATICAL SCIENCES I - Probability and Statistics

(ii) $\lim_{c\to-\infty}\phi(c) = \inf\{x : \mu(-\infty, x) > 0\} \equiv a$,
(iii) $\lim_{c\to+\infty}\phi(c) = \sup\{x : \mu[x, \infty) > 0\} \equiv b$,
(iv) *for any α, $a < \alpha < b$, there is a unique c such that $\phi(c) = \alpha$.*

Proof: If μ is a Dirac measure then the claims are trivially true, so we assume that μ is not a Dirac measure. Since ψ is infinitely differentiable and $\psi(c) > 0$ for all c, ϕ is also infinitely differentiable. Moreover,

$$\phi'(c) = \frac{(\int_C x^2 e^{cx}\mu(dx))\psi(c) - (\psi'(c))^2}{(\psi(c))^2}$$

can be seen as the variance of a non-constant random variable X_c whose distribution is absolutely continuous with respect to μ with probability density function $f_c(x) = \frac{e^{cx}}{\psi(c)}$. (Note that X_c is non-constant since μ is not concentrated at a single point and f_c is positive on the support of μ.) Thus $\phi'(c) =$ variance of $X_c > 0$, for all c. This proves (i).

Although a direct verification of (ii) is possible we will give a slightly different proof. We will show that as $c \to -\infty$, the random variable X_c converges in distribution to the constant function a so that $\phi(c)$ which is the expected value of X_c converges to a. Note that by definition of a, for all $\epsilon > 0$, $\mu([a, a+\epsilon)) > 0$ while $\mu((-\infty, a)) = 0$. Also $\mu((b, \infty)) = 0$, whence

$$P(X_c > a+\epsilon) = \frac{\int_{(a+\epsilon,b]} e^{cx}\mu(dx)}{\psi(c)} = \frac{\int_{(a+\epsilon,b]} e^{c(x-a)}\mu(dx)}{\int_{[a,b]} e^{c(x-a)}\mu(dx)}.$$

For $c < 0$, and $0 < \epsilon < \frac{b-a}{2}$,

$$\begin{aligned} P(X_c > a+\epsilon) &\le \frac{e^{c\epsilon}\mu((a+\epsilon, b])}{e^{c\frac{\epsilon}{2}}\mu([a, a+\frac{\epsilon}{2}))} \\ &= e^{c\frac{\epsilon}{2}}\frac{\mu((a+\epsilon, b])}{\mu([a, a+\frac{\epsilon}{2}))} \to 0 \quad \text{as} \quad c \to -\infty. \end{aligned}$$

Also, since $\mu((-\infty, a)) = 0$, $\quad P(X_c < a) = 0$. So, $X_c \to a$ in distribution as $c \to -\infty$, whence $\phi(c) \to a$ as $c \to -\infty$. This proves (ii). Proof of (iii) is similar. Finally (iv) follows from the intermediate value theorem since ϕ is strictly increasing and continuous with range (a, b). This completes the proof of Proposition 2.2.

Proposition 2.2 also appears at the beginning of the theory of large deviations (see [10]) thus giving a glimpse of the natural connection between large deviation theory and entropy theory. (See Varadhan's interview, p31, [11].)

The requirement that μ have compact support can be relaxed in the above proposition. The following is a result under a relaxed condition.

Let μ be a measure on $\mathbb{R}$. Let $I = \{c : \int_{\mathbb{R}} e^{cx}\mu(dx) < \infty\}$. It can be shown that I is a connected subset of $\mathbb{R}$ which can be empty, a singleton, or an interval that is half open, open, closed, finite, semi-finite or all of $\mathbb{R}$ (see [2]). Suppose I has a non-empty interior I^0. Then in I^0 the function $\psi(c) = \int_{\mathbb{R}} e^{cx}\mu(dx)$ is infinitely differentiable with $\psi^{(k)}(c) = \int_{\mathbb{R}} e^{cx}x^k\mu(dx)$. Further $\phi(c) = \frac{\psi'(c)}{\psi(c)}$ satisfies

$$\phi'(c) = \frac{\psi''(c) - (\psi'(c))^2}{(\psi(c))^2},$$

which is positive, being equal to the variance of a random variable with probability density function $\frac{e^{cx}}{\psi(c)}$ with respect to μ. Thus, for any α satisfying $\inf_{c\in I^0}\phi(c) < \alpha < \sup_{c\in I^0}\phi(c)$, there is a unique c_0 in I^0 such that $\phi(c_0) = \alpha$.

Let μ be a probability measure on $\mathbb{R}$. Note that a real number λ is the mean $\int_{\mathbb{R}} x\nu(dx)$ of a probability measure ν absolutely continuous with respect to μ if and only if $\mu(\{x : x \leq \lambda\}), \mu(\{x : x \geq \lambda\})$ are both positive.

As a corollary of Proposition 2.2 we have:

Corollary 2.1. *Let the closed support of μ be a compact set C. Let α be such that $\mu\{x : x \leq \alpha\}, \mu\{x : x \geq \alpha\}$ are both positive. Let*

$$\mathcal{F}_\alpha = \{f : f \text{ a pdf}, \quad \int_C xf(x)\mu(dx) = \alpha\}.$$

Then there exists a unique probability density function g with respect to μ such that

$$H(g,\mu) = \max\{H(f,\mu) : f \in \mathcal{F}_\alpha\}.$$

If $\alpha = \inf C$ or if $\alpha = \sup C$, then μ necessarily assigns positive mass to α and $g = \frac{1}{\mu(\{\alpha\})} \times 1_{\{\alpha\}}$. Let $\inf C < \alpha < \sup C$. Then there is a unique c such that with $g = f_c = \frac{e^{cx}}{\int_C e^{cx}\mu(dx)}$ one has $\alpha = \int_C xf_c(x)\mu(dx)$ and

$$H(g,\mu) = H(f_c,\mu) = \max\{H(f,\mu) : f \in \mathcal{F}_\alpha\}.$$

PERSPECTIVES IN MATHEMATICAL SCIENCES I - Probability and Statistics

Keeping in mind the notation of the above corollary, and the fact that α uniquely determines c, we write $\nu_{\alpha,\mu}$ to denote the probability measure $f_c d\mu$, i.e., the measure with probability density function f_c with respect to μ. It is also the Gibbs-Boltzmann measure $\nu_{\mu,c,h}$ with $h(x) \equiv x$. We are now ready to state the result on the weak convergence of Gibbs-Boltzmann measures.

Theorem 2.3. *Let C be a compact subset of $\mathbb{R}$. Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of probability measures such that (i) support of each μ_n is contained in C, (ii) $\mu_n \to \mu$ weakly. Let*

$$a = \inf\{x : \mu((-\infty, x)) > 0\}, b = \sup\{x : \mu((x, \infty)) > 0\}.$$

Let $a < \alpha < b$. Then for all large n, ν_{α,μ_n} is well defined and $\nu_{\alpha,\mu_n} \to \nu_{\alpha,\mu}$ weakly.

Proof: Since $\mu_n \to \mu$ weakly, and $a < \alpha < b$, for n large, $\mu_n(-\infty, \alpha) > 0$, $\mu_n(\alpha, \infty) > 0$. So, by Proposition 2.2 it follows that there is a unique c_n such that with $f_{c_n} = \frac{e^{c_n x}}{\int_C e^{c_n x} \mu_n(dx)}$,

$$\int_C x f_{c_n}(x) \mu_n(dx) = \alpha.$$

Thus ν_{α,μ_n} is well defined for all large n. Next we claim that c_n's are bounded. If c_n's are not bounded, then there is a subsequence of $(c_n)_{n=1}^{\infty}$ which diverges to $-\infty$ or to $+\infty$. Suppose a subsequence of $(c_n)_{n=1}^{\infty}$ diverges to $-\infty$. Note that for all $\epsilon > 0$, and $c_n < 0$,

$$\nu_{\alpha,\mu_n}[a+\epsilon, \infty) = \frac{\int_{[a+\epsilon,\infty)} e^{c_n x} \mu_n(dx)}{\int_{\mathbb{R}} e^{c_n x} \mu_n(dx)} \le \frac{e^{c_n \epsilon} \mu_n([a+\epsilon, \infty)}{e^{c_n \frac{\epsilon}{2}} \mu_n([a, a+\frac{\epsilon}{2}))}. \quad (6)$$

Since $\mu_n \to \mu$ weakly, for each $\epsilon > 0$, $\liminf_{n\to\infty} \mu_n([a, a+\epsilon)) > 0$. Therefore, over the subsequence in question, $\nu_{\alpha,\mu_n}([a+\epsilon, \infty)) \to 0$ by (6), and since $\nu_{\alpha,\mu_n}((-\infty, a)) = 0$ we see that $(\nu_{\alpha,\mu_n})_{n=1}^{\infty}$ converges weakly to Dirac measure at a. Since C is compact, this implies that $\int_C x\nu_{x,\mu_n}(dx) \to a$ as $n \to \infty$, contradicting the fact that $\int_C x\nu_{\alpha,\mu_n}(dx) = \alpha > a$, for all n. Similarly, $(c_n)_{n=1}^{\infty}$ is bounded above. So $(c_n)_{n=1}^{\infty}$ is a bounded sequence, which in fact converges as we see below. Let a subsequence $(c_{n_k})_{k=1}^{\infty}$ converge to a real number c. Then, since $\mu_n \to \mu$ weakly, and since all μ_n have support contained in C, a compact set, we see that

$$\int_C e^{c_{n_k} x} \mu_{n_k}(dx) \to \int_C e^{cx} \mu(dx),$$

and

$$\int_C xe^{c_{n_k}x}\mu_{n_k}(dx) \to \int_C xe^{cx}\mu(dx),$$

as $k \to \infty$, whence, $\alpha = \phi(c_{n_k}) \to \phi(c)$, so that $\phi(c) = \alpha$. Again, by Proposition 2.2., c is uniquely determined by α, so that all subsequential limits of $(c_n)_{n=1}^{\infty}$ are the same. So $c_n \to c$ as $n \to \infty$. Clearly then, since C is compact, $f_{c_n} \to f_c$ uniformly on C, so that $\nu_{\alpha,\mu_n} \to \nu_{\alpha,\mu}$ weakly, thus proving Theorem 2.3.

This theorem allows us to answer a question raised in [6]. Let C be a compact subset of $\mathbb{R}$. For each $\epsilon > 0$ let $C_\epsilon = \{x_{\epsilon,1} < x_{\epsilon,2} \cdots < x_{\epsilon,k_\epsilon}\}$ be a ϵ-net in C, i.e., for all $x \in C$, there is a $x_{\epsilon,j}$ such that $\mid x - x_{\epsilon,j} \mid< \epsilon$. Fix α such that $\inf C < \alpha < \sup C$. Then for small enough ϵ it must hold that $x_{\epsilon,1} < \alpha < x_{\epsilon,k_\epsilon}$. Let μ_ϵ be the uniform distribution on C_ϵ. Let $\nu_{\mu_\epsilon,\alpha}$ be the Gibbs-Boltzmann distribution on C_ϵ corresponding to μ_ϵ and α. The problem raised in [6] was whether $\nu_{\alpha,\mu_\epsilon}$ converges to a unique limit as $\epsilon \to 0$. Theorem 2.3. above answers this in the negative. Take $C = [0,1]$, and let $(x_n)_{n=1}^{\infty}$ be a sequence of points in C which become dense in C and such that if μ_n denotes the uniform distribution on the first n points of the sequence, then the sequence $(\mu_n)_{n=1}^{\infty}$ has no unique weak limit. By Theorem 2.3, the associated Gibbs-Boltzmann distributions $(\nu_{\epsilon,\mu_n})_{n=1}^{\infty}$ will also not have a unique weak limit.

(Here is a way of constructing such a sequence $(x_n)_{n=1}^{\infty}$. Let μ_1 and μ_2 be two different probability measures on $[0,1]$ both equivalent to Lebesgue measure. Let $(X_n)_{n=1}^{\infty}$, $(Y_n)_{n=1}^{\infty}$ be two sequences of points in $[0,1]$ such that the sequence of empirical distributions based on $(X_n)_{n=1}^{\infty}$ converges weakly to μ_1 and that based on $(Y_n)_{n=1}^{\infty}$ converges weakly to μ_2. Let $(Z_i)_{i=1}^{\infty}$ be defined as follows:

$$Z_i = X_i, 1 \le i \le n_1, Z_i = Y_i, n_1 < i \le n_2, Z_i = X_i, n_2 < i \le n_3, \cdots.$$

One can choose $n_1 < n_2 < n_3 < \cdots$ in such a way that the empirical distribution of $(Z_i)_{i=1}^{\infty}$ converges to μ_1 over the sequence $(n_{2k+1})_{k=1}^{\infty}$ and to μ_2 over the sequence $(n_{2k})_{k=1}^{\infty}$. Since μ_1, μ_2 are equivalent to the Lebesgue measure on $[0,1]$, the sequence $(Z_n)_{n=1}^{\infty}$ is dense in $[0,1]$.)

Remarks. See [1] for some further applications of the above discussion. The quantity $H(f,\mu)$ or its negative has been known in statistical literature as Kullback-Leibler information. In financial mathematics, the quantity $-H(f,\mu)$ is called the relative entropy with respect to μ, so then one deals with$f$ minimizing the relative entropy.

1.2.3. *Relative Entropy and Conditioning*

Let $(\Omega, \mathcal{B}, \mu)$ be a probability space, where $(\Omega, \mathcal{B})$ is a standard Borel space (see [7], [9]). For a given countable collection $\{B_n\}_{n=1}^{\infty}$ in $\mathcal{B}$ let $\mathbb{Q}$, $\mathcal{C}$, be the partition and the σ-algebra generated by it. (By partition generated by $\{B_n\}_{n=1}^{\infty}$ we mean the collection:

$$\mathbb{Q} = \{q : q = \cap_{i=1}^{\infty} B_i^{\epsilon_i}\},$$

where $\epsilon_i = 0$ or 1, and $B_i^0 = B_i, B_i^1 = \Omega - B_i$.)

For any subset A of Ω, by saturation of A with respect to $\mathbb{Q}$ we mean the the union of elements of $\mathbb{Q}$ which have non-empty intersection with A.

It is known that $\mathcal{C} = \{C \in \mathcal{B} : C \text{ a union of elements in } \mathbb{Q}\}$. We regard $\mathcal{C}$ also as a σ-algebra on $\mathbb{Q}$. Note that $\mathcal{C}$-measurable functions are the functions which are $\mathcal{B}$ measurable and which are constant on elements of $\mathbb{Q}$. A $\mathcal{C}$-measurable function on Ω is therefore also a $\mathcal{C}$-measurable function on $\mathbb{Q}$, and if f is such a function, we regard it both as a function on Ω and on $\mathbb{Q}$. We write $f(q)$ to denote the constant value of such a function on $q, q \in \mathbb{Q}$. In addition to $\mathcal{C}$, we also need a larger σ-algebra, denoted by $\mathcal{A}$, generated by analytic subsets $\mathbb{Q}$ (see [9]).

Since $(\Omega, \mathcal{B})$ is a standard Borel space for any probability measure μ on $(\Omega, \mathcal{B})$ there exists a regular conditional probability given $\mathcal{C}$ (equivalently disintegration of μ with respect to $\mathbb{Q}$). This means that there is a function $\mu(\cdot, \cdot)$ on $\mathcal{B} \times \mathbb{Q}$, such that

1) $\mu(\cdot, q)$ is a probability measure $\mathcal{B}$,

2) $\mu(q, q) = 1$,

3) for each $A \in \mathcal{B}$, $\mu(A, \cdot)$ is measurable with respect to $\mathcal{A}$ and

$$\mu(A) = \int_{\Omega} \mu(A \cap q(\omega), \omega)\mu(d\omega) = \int_{\mathbb{Q}} \mu(A, q)\mu \mid_{\mathcal{C}} (dq),$$

where $q(\omega)$ is the element of $\mathbb{Q}$ containing ω,

4) if $\mu'(\cdot, \cdot)$ is another such function then $\mu(\cdot, \omega) = \mu'(\cdot, \omega)$ for a.e. ω (with respect to μ).

PERSPECTIVES IN MATHEMATICAL SCIENCES I - Probability and Statistics

Note that sets in $\mathcal{A}$ are measurable with respect to every probability measure on $\mathcal{B}$. Further, we can say that there is a μ-null set N which is a union of elements of $\mathbb{Q}$ and such that (i) $\mathcal{C} \mid_{\Omega - N}$ is a standard Borel structure on $\Omega - N$, (ii) for each $A \in \mathcal{C} \mid_{\Omega-N}$, $\mu(A, \cdot)$ is measurable with respect to this Borel structure.

The function $\mu(\cdot,\cdot)$ is called conditional probability distribution of μ with respect to $\mathbb{Q}$, or, the disintegration of μ with respect to the partition $\mathbb{Q}$. If f is a $\mathcal{B}$ measurable function with finite integral with respect to μ, then the function $h(\omega) = \int_q f(y)\mu(dy, q), \omega \in q$ is called the conditional expectation of f with respect to $\mathbb{Q}$ (or with respect to $\mathcal{C}$) and denoted by $E_\mu(f \mid \mathbb{Q})$ or $E_\mu(f \mid \mathcal{C})$. If $\mathbb{Q}$ is the partition induced by a measurable function g, $E_\mu(f \mid \mathbb{Q})$ is called the conditional expectation of f given g and written $E_\mu(f \mid g)$. (See [9], p. 209)

The measure μ is completely determined by $\mu(\cdot,\cdot)$ together with the restriction of μ to $\mathcal{C}$, denoted by $\mu \mid_{\mathcal{C}}$. We note also that if ν' is any probability measure on $\mathcal{C}$ then $\nu \equiv \int_{\mathbb{Q}} \mu(\cdot, q)\nu'(dq)$ is a measure on $\mathcal{B}$ having the same conditional distribution (or disintegration) with respect to $\mathbb{Q}$ as that of μ.

Let μ and ν be two probability measures on $\mathcal{B}$, with ν absolutely continuous with respect to μ. Let $\mu(\cdot, q), \nu(\cdot, q), q \in \mathbb{Q}$ be the disintegration of μ and ν with respect to the partition $\mathbb{Q}$. Then, for a.e. ω with respect to μ,

$$\frac{d\nu}{d\mu}(\omega) = \frac{d\nu(\cdot, q)}{d\mu(\cdot, q)} \cdot \frac{d\nu \mid_{\mathcal{C}}}{d\mu \mid_{\mathcal{C}}}(\omega), \text{if } \omega \in q.$$

A calculation using this identity shows that

$$H(\frac{d\nu}{d\mu}, \mu) = \int_{\Omega} H(\frac{d\nu(\cdot, q)}{d\mu(\cdot, q)}, \mu(\cdot, q))\nu \mid_{\mathcal{C}} (dq) + H(\frac{d\nu \mid_{\mathcal{C}}}{d\mu \mid_{\mathcal{C}}}, \mu \mid_{\mathcal{C}}). \tag{7}$$

Assume now that f is a real valued function having finite expectation with respect to μ, and let g be a real valued function on $\mathbb{Q}$ for which there is a probability measure ν, absolutely continuous with respect to μ, such that for all $q \in \mathbb{Q}$,

$$\int_q f(\omega)\nu(d\omega, q) = g(q). \tag{8}$$

(Note that g is necessarily $\mathcal{C}$ measurable.).

Theorem 2.4. *If ν_0 is a probability measure which maximizes $H(\frac{d\nu}{d\mu}, \mu)$ among all probability measure ν, absolutely continuous with respect to μ and satisfying (8), then for a.e. q (with respect to $\mu \mid_{\mathcal{C}}$), $\nu_0(\cdot, q)$ maximizes the relative entropy $H(\frac{d\lambda}{d\mu(\cdot,q)}, \mu(\cdot, q))$ as λ ranges over all probability measures on q absolutely continuous with respect to $\mu(\cdot, q)$ and satisfying*

$$\int_q f(\omega)\lambda(d\omega) = g(q), q \in \mathbb{Q}.$$

Proof: Assume in order to arrive at a contradiction, that the theorem is false. Then there is a set $E \subset \mathbb{Q}$ of positive $\mu \mid_{\mathcal{C}}$ measure and a transition probability $\lambda(\cdot, \cdot)$ on $\mathcal{B} \times E$ such that for each $q \in E$,

(i) $\lambda(q, q) = 1$,

(ii) $\lambda(\cdot, q)$ is absolutely continuous with respect to $\mu(\cdot, q)$,

(iii) $H(\frac{d\nu_0(\cdot,q)}{d\mu(\cdot,q)}, \mu(\cdot, q)) < H(\frac{d\lambda(\cdot,q)}{d\mu(\cdot,q)}, \mu(\cdot, q))$, and

(iv) $\int_q f(\omega)\lambda(d\omega, q) = g(q)$.

The existence of such an E and the transition probability $\lambda(\cdot, \cdot)$ is easy to see if the partition $\mathbb{Q}$ is finite or countable. In the general case the proof relies on some non-trivial measure theory. Define a new transition probability on $\mathcal{B} \times \mathbb{Q}$ as follows: For all $A \in \mathcal{B}$,

$$T(A, q) = \nu_0(A, q) \text{ if } \ q \in \mathbb{Q} - E, T(A, q) = \lambda(A, q) \text{ if } \ q \in E.$$

The measure T defined on $\mathcal{B}$ by

$$T(A) = \int_A T(A, q)\nu_0 \mid_{\mathcal{C}} (dq),$$

is absolutely continuous with respect to μ, $T(\cdot, \cdot)$ is its disintegration with respect to $\mathbb{Q}$, $T \mid_{\mathcal{C}} = \nu_0 \mid_{\mathcal{C}}$ and for each $q \in \mathbb{Q}$

$$\int_q f(\omega)T(d\omega, q) = g(q).$$

Finally, by formula (7), and in view of the values of T on E,

$$H(\frac{dT}{d\mu}, \mu) = \int_{\mathbb{Q}} H(\frac{dT(\cdot, q)}{d\mu(\cdot, q)}, \mu(\cdot, q))\nu_0 \mid_{\mathcal{C}} (dq) + H(\frac{d\nu_0 \mid_{\mathcal{C}}}{d\mu \mid_{\mathcal{C}}}, \mu \mid_{\mathcal{C}})$$

PERSPECTIVES IN MATHEMATICAL SCIENCES I - Probability and Statistics

is strictly bigger than $H(\frac{d\nu_0}{d\mu}, \mu)$ which is equal to

$$H(\frac{d\nu_0}{d\mu}, \mu) = \int_{\mathbb{Q}} H(\frac{d\nu_0(\cdot, q)}{d\mu(\cdot, q)}, \mu(\cdot, q))\nu_0 \mid_{\mathcal{C}} (dq) + H(\frac{d\nu_0 \mid_{\mathcal{C}}}{d\mu \mid_{\mathcal{C}}}, \mu \mid_{\mathcal{C}}).$$

This contradicts the maximality of $H(\frac{d\nu_0}{d\mu}, \mu)$, and proves the theorem.

Note that the measures $\nu_0(\cdot, q), q \in \mathbb{Q}$, remain unchanged even if ν_0 maximizes $H(\frac{d\nu}{d\mu}, \mu)$ under the additional constraint that $\int_\Omega f(\omega)\nu(d\omega) = \alpha$ for some fixed α. However, $\nu_0 \mid_{\mathcal{C}}$ need not maximize $H(\frac{dm}{d\mu|_{\mathcal{C}}}, \mu \mid_{\mathcal{C}})$ among all probability measures m on $\mathcal{C}$ satisfying $\int_{\mathbb{Q}} g(q)m(dq) = \alpha$.

1.3. Measure Free Martingales, Weak Martingales, Martingales

1.3.1. *Finite Range Case*

In this section we will discuss the notion of measure free martingales, and and its relation to the usual martingale. In [6] the simpler case of measure free martingale, where functions assume only finitely many values, was introduced and we recall it below.

Let Ω be a non-empty set. Let $(f_n)_{n=1}^\infty$ be a sequence of real valued functions such that each f_n has finite range, say $(x_{n1}, x_{n2}, \cdots, x_{nk_n})$, and these values are assumed on subsets $\Omega_{n1}, \Omega_{n2}, \cdots, \Omega_{nk_n}$. These sets form a partition of Ω which we denote by $\mathbb{P}_n$. We denote by $\mathbb{Q}_n$ the partition generated by $\mathbb{P}_1, \mathbb{P}_2, \cdots, \mathbb{P}_n$ and the algebra generated by $\mathbb{Q}_n$ is denoted by $\mathcal{A}_n$. Let $\mathcal{A}_\infty$ denote the algebra $\cup_{n=1}^\infty \mathcal{A}_n$.

Define $\mathcal{A}_n$ measurable functions m_n, M_n as follows: for $Q \in \mathbb{Q}_n$ and $\omega \in Q$,

$$m_n(\omega) = \min_{x \in Q} f_{n+1}(x),$$

$$M_n(\omega) = \max_{x \in Q} f_{n+1}(x).$$

Definition 3.1. The sequence $(f_n, \mathcal{A}_n)_{n=1}^\infty$ is said to be a measure free martingale or probability free martingale if

$$m_n(\omega) \le f_n(\omega) \le M_n(\omega), \forall \omega \in \Omega,\ n \ge 1.$$

Clearly, for each $Q \in \mathbb{Q}_n$, the function f_n is constant on Q. We denote this constant by $f_n(Q)$. With this notation, it is easy to see that

$(f_n, \mathcal{A}_n)_{n=1}^{\infty}$ is a measure free martingale or probability free martingale if and only if for each n and for each $Q \in \mathbb{Q}_n$, $f_n(Q)$ lies between the minimum and the maximum values of $f_{n+1}(Q')$ as Q' runs over $Q \cap \mathbb{Q}_{n+1}$. It is easy to see that if there is a probability measure on $\mathcal{A}_\infty$ with respect to which $(f_n, \mathcal{A}_n)_{n=1}^{\infty}$ is a martingale then $(f_n, \mathcal{A}_n)_{n=1}^{\infty}$ is also a measure free martingale. Indeed let P be such a measure. Then, for any Q in $\mathbb{Q}_n$, $f_n(Q)$, is equal to

$$\frac{1}{P(Q)} \sum_{\{Q' \in \mathbb{Q}_{n+1}, Q' \subseteq Q\}} f_{n+1}(Q') P(Q'),$$

so that $f_n(Q)$ lies between the minimum and the maximum values $f_{n+1}(Q')$, $Q' \in Q \cap \mathbb{Q}_{n+1}$. In [6], the following converse is proved.

Theorem 3.1. *Given a measure free martingale $(f_n, \mathcal{A}_n)_{n=1}^{\infty}$, there exists for each $n \geq 0$, a measure P_n on $\mathcal{A}_n$ such that*

$$P_{n+1} |_{\mathcal{A}_n} = P_n, \quad E_{n+1}(f_{n+1} \mid \mathcal{A}_n) = f_n,$$

where E_{n+1} denotes the conditional expectation with respect to the probability measure P_{n+1}. There is a finitely additive probability measure P on the algebra $\mathcal{A}_\infty$ such that, for each n, $P |_{\mathcal{A}_n} = P_n$. Moreover such a P is unique if certain naturally occurring entropies are maximized.

1.3.2. *The General Case*

In the rest of this section we will dispense with the requirement that the functions f_n assume only finitely many values.

Let $(\Omega, \mathcal{B})$ be a standard Borel space, and let $(f_n)_{n=1}^{\infty}$ be a sequence of real valued Borel functions on Ω. For each n, let $\mathbb{P}_n = \{f_n^{-1}(\{\omega\}) : \omega \in \mathbb{R}\}$ denote the partition of Ω generated by f_n, and let $\mathbb{Q}_n$ denote the partition generated $f_1, f_2, \cdots, f_n$, i.e., $\mathbb{Q}_n$ is the superposition of $\mathbb{P}_1, \mathbb{P}_2, \cdots, \mathbb{P}_n$. Let $\mathcal{B}_n$ be the σ-algebra generated by $f_1, f_2, \cdots, f_n$. Since $(\Omega, \mathcal{B})$ is a standard Borel space, for each n, $\mathcal{B}_n$ is the collection of sets in $\mathcal{B}$ which can be written as a union of elements in $\mathbb{Q}_n$. For $q \in \mathbb{Q}_n$, f_n is constant on q and we denote this value by $f_n(q)$. The algebra $\cup_{n=1}^{\infty} \mathcal{B}_n$ will be denoted by $\mathcal{B}_\infty$ and we will assume that it generates $\mathcal{B}$. Note that we have changed the notation slightly. In section 1.3.1 above we denoted an element in $\mathbb{Q}_n$ by Q, while from now we will use the lower case q.

Definition 3.2. Let $(f_n)_{n=1}^{\infty}$ be a sequence $\mathcal{B}$ measurable real valued functions on Ω.

(i) Say that sequence $(f_n)_{n=1}^{\infty}$ is a measure free martingale if for each n and for each $q \in \mathbb{Q}_n$, $f_n(q)$ is in the convex hull of the values assumed by f_{n+1} on q (note that f_{n+1} need not be constant on $q \in \mathbb{Q}_n$.)

(ii) Let ν_∞ be a finitely additive probability measure on $\mathcal{B}_\infty$ such that for each $n < \infty$, its restriction to $\mathcal{B}_n$ is countably additive. The sequence $(f_n)_{n=1}^{\infty}$ is called a weak martingale with respect to ν_∞ if for each n, $E_{\nu_{n+1}}(f_{n+1} \mid f_n) = f_n \;\; a.e.\; \nu_{n+1}$.

(iii) A weak martingale $(f_n)_{n=1}^{\infty}$ with respect to ν_∞ is a martingale if ν_∞ is countably additive, in which case the countably additive extension of ν_∞ to $\mathcal{B}$ is denoted by ν, and we call $(f_n)_{n=1}^{\infty}$ a martingale with respect to ν.

Clearly every martingale is a weak martingale, and if $(f_n)_{n=1}^{\infty}$ is a weak martingale with respect to ν_∞, then for each n we can modify f_n on a ν_n null set so that the resulting new sequence of functions is a measure free martingale. Indeed if $\nu_n(\cdot,\cdot)$ is the conditional probability distribution of ν_{n+1} given $\mathbb{Q}_n$, then

$$f_n(q) = \int_q f_{n+1}(\omega)\nu_n(d\omega, q), \;\; \nu_n \; a.\; e.\; q \in \mathbb{Q}_n. \tag{9}$$

At those q's where the equality in (9) holds $f_n(q)$ lies between the infimum and the supremum of the values assumed by f_{n+1} on q. On the other q's we modify f_{n+1} by simply setting $f_{n+1}(\omega) = f_n(q), \omega \in q$. The modified sequence $(f_n)_{n=1}^{\infty}$ is the required measure free martingale. In the converse direction we show that every measure free martingale admits a finitely additive measure on $\mathcal{B}_\infty$ under which it is a weak martingale.

Proposition 3.1. *Let $(\Omega, \mathcal{B})$ be a standard Borel space and let $\mathbb{Q}$, $\mathcal{C}$ be the partition and the σ-algebra generated by a countable collection of sets in $\mathcal{B}$. Let $\mathcal{A}$ be the σ-algebra generated by analytic subsets of Ω which are unions of elements of $\mathbb{Q}$. Let f and g be respectively $\mathcal{B}$ and $\mathcal{C}$ measurable real valued functions on Ω such that for each $q \in \mathbb{Q}$, $g(q)$ is in the convex hull of the values assumed by f on q. Then there exists a transition probability $\nu(\cdot,\cdot)$ on $\mathcal{B} \times \mathbb{Q}$ such that for each $A \in \mathcal{B}$, the function $\nu(A,\cdot)$ is $\mathcal{A}$ measurable, while $\nu(\cdot, q)$ is a probability measure on $\mathcal{B}$ supported on at most two points of q satisfying*

$$g(q) = \int_q f(\omega)\nu(d\omega, q), \forall q \in \mathbb{Q}.$$

Proof: The sets

$$S_1 = \{\omega \in \Omega : f(\omega) \leq g(\omega)\}, S_2 = \{\omega \in \Omega : g(\omega) \leq f(\omega)\}$$

are in $\mathcal{B}$. For each $q \in \mathbb{Q}$, since $g(q)$ is in the convex hull of the values assumed by f on q, both S_1 and S_2 have non-empty intersection with q. By von-Neumann selection theorem (see [9], p 199) there exist coanalytic sets $C_1 \subset S_1, C_2 \subset S_2$ which intersect each $q \in \mathbb{Q}$ in exactly one point. For each $q \in \mathbb{Q}$, let

$$\omega_1(q) = S_1 \cap q, \omega_2(q) = S_2 \cap q.$$

Then

$$f(\omega_1(q)) \leq g(q) \leq f(\omega_2(q)),$$

so that the middle real number $g(q)$ is a unique convex combination of $f(\omega_1(q)), f(\omega_2(q))$. If $f(\omega_1(q)) = f(\omega_2(q)) = g(q)$ write $p_1(q) = 1, p_2(q) = 0$, otherwise write

$$p_1(q) = \frac{f(\omega_2(q)) - g(q)}{f(\omega_2(q)) - f(\omega_1(q))}, p_2(q) = \frac{g(q) - f(\omega_1(q))}{f(\omega_2(q)) - f(\omega_1(q))}.$$

Then

$$p_1(q)f(\omega_1(q)) + p_2(q)f(\omega_2(q)) = g(q).$$

For each $q \in \mathbb{Q}$, let $\nu(\cdot, q)$ be the probability measure on q with masses $p_1(q), p_2(q)$ at $\omega_1(q), \omega_2(q)$ respectively. The sets C_1, C_2 are co-analytic and functions $f\mid_{C_1}, f\mid_{C_2}$ are $\mathcal{B}\mid_{C_1}, \mathcal{B}\mid_{C_2}$ measurable respectively, whence $p_1(\cdot), p_2(\cdot)$ are $\mathcal{A}$ measurable. For any $A \in \mathcal{B}$, and $q \in \mathbb{Q}$,

$$\nu(A, q) = p_1(q)1_A(\omega_1(q)) + p_2(q)1_A(\omega_2(q)),$$

whence, for each $A \in \mathcal{B}$, $\nu(A, \cdot)$ is $\mathcal{A}$ measurable. The proposition is proved.

Theorem 3.2. *A measure free martingale admits a finitely additive measure under which it is weak martingale.*

Proof: Let $(f_n)_{n=1}^{\infty}$ be a measure free martingale. Let $\mathbb{Q}_1, \mathbb{Q}_2$, $\mathcal{B}_1, \mathcal{B}_2$ be the partition and the σ-algebra generated by f_1, f_2 respectively. Let ν_1 be a probability measure on $\mathcal{B}_1$. Since, for each $q \in \mathbb{Q}_1$, $f_1(q)$ is in the convex hull of the values assumed by f_2 on q, by the Proposition 3.1, there is a transition probability $\nu_1(\cdot,\cdot)$ on $\mathcal{B}\times\mathbb{Q}_1$ such that for each $q\in\mathbb{Q}_1$, $\nu_1(q,q)=1$, and

$$\int_q f_2(\omega)\nu_1(d\omega,q)=f_1(q).$$

For any $A\in\mathcal{B}_2$, define

$$\nu_2(A)=\int_{\mathbb{Q}_1}\nu_1(A,q)\nu_1(dq).$$

Then ν_2 is a countably additive measure on $\mathcal{B}_2$,

$$\nu_2\mid_{\mathcal{B}_1}=\nu_1,\quad E_{\nu_2}(f_2\mid f_1)=f_1.$$

Having defined ν_2, it is now clear how to construct $\nu_3,\nu_4,\cdots,\nu_n,\cdots$ such that for each n,

$$\nu_{n+1}\mid_{\mathcal{B}_n}=\nu_n$$

and

$$E_{\nu_{n+1}}(f_{n+1}\mid f_n)=f_n.$$

The finitely additive measure ν_∞ defined on $\mathcal{B}_\infty$ by

$$\nu_\infty(A)=\nu_n(A), A\in\mathcal{B}_n$$

satisfies, for each n,

$$\nu_\infty\mid_{\mathcal{B}_n}=\nu_n$$

and $(f_n)_{n=1}^\infty$ is thus a a weak martingale with respect to ν_∞. This proves the theorem.

It is natural to ask the question as to when is ν_∞ countably additive. There is an answer to this. The refining system of partitions $(\mathbb{Q}_n)_{n=1}^\infty$ as well as the associated σ-algebras $(\mathcal{B}_n)_{n=1}^\infty$ is called a filtration. It is said to be complete if for every decreasing sequence $(q_n)_{n=1}^\infty$, of non-empty elements, $q_n\in\mathbb{Q}_n$, their intersection $\cap_{n=1}^\infty q_n$ is non-empty. We have:

Theorem 3.3. *If the filtration $(\mathbb{Q}_n)_{n=1}^\infty$ associated with the measure free martingale is complete, then ν_∞ is countably additive.*

This is indeed a consequence of the Kolmogorov consistency theorem formulated in terms of filtrations which is as follows. Let the filtration $(\mathbb{Q}_n)_{n=1}^{\infty}$ arise from a sequence of Borel functions $(f_n)_{n=1}^{\infty}$, not necessarily a measure free martingale. For $1 \leq i \leq j$, we define the natural projection map $\Pi_{ij} : \mathbb{Q}_j \to \mathbb{Q}_i$ by

$$\Pi_{ij}(q) = r, \quad \text{if } q \subset r, q \in \mathbb{Q}_j, r \in \mathbb{Q}_i.$$

For each i let $\Pi_i : \Omega \to \mathbb{Q}_i$ be defined by

$$\Pi_i(\omega) = q \in \mathbb{Q}_i, \text{if } \omega \in q.$$

If $q \in \mathbb{Q}_n$ and $i \leq n$, then f_i is constant on q, and we write $f_i(q)$ to denote this constant value. For each n let τ_n be the smallest topology on $\mathbb{Q}_n$ which makes the map $q \to (f_1(q), f_2(q), \cdots, f_n(q)), q \in \mathbb{Q}_n$, continuous. We note that maps $\Pi_{ij}, i \leq j$ are continuous. The topology τ_n generates the σ-algebra $\mathcal{B}_n$. Any probability measure μ on $\mathcal{B}_n$ is compact approximable with respect to this topology, i.e., given $B \in \mathcal{B}_n$ and $\epsilon > 0$, there is a set $C \subset B$, C compact w.r.t. τ_n, such that $\mu(B - C) < \epsilon$. For each n, let P_n be a countably additive probability measure on $\mathcal{B}_n$. Assume that $P_{n+1} |_{\mathcal{B}_n}= P_n$. Define P_∞ on $\cup_{n=1}^{\infty}\mathcal{B}_n$ by $P(A) = P_n(A)$, if $A \in \mathcal{B}_n$. P is obviously finitely additive.

We have the Kolmogorov consistency theorem in our setting, arrived at after a discussion with B. V. Rao, and as pointed out by Rajeeva Karandikar, it is proved also in [7].

Theorem 3.4. *If the the filtration $(\mathbb{Q}_n)_{n=1}^{\infty}$ is complete, then P_∞ is countably additive.*

Proof: For simplicity, write P for P_∞. If P is not countably additive, then there exists a decreasing sequence $(A_n)_{n=1}^{\infty}$ in $\cup_{n=1}^{\infty}\mathcal{B}_n$, with $\cap_{n=1}^{\infty}A_n = \emptyset$, such that for all n, $P(A_n) > a > 0$ for some positive real a. Without loss of generality we can assume that for each n, A_n is in $\mathcal{B}_n$. We can choose a set $C_1 \subset A_1$, $C_1 \in \mathcal{B}_1$, C_1 compact with respect to the topology τ_1, such that

$$P(A_1 - C_1) = P_1(A_1 - C_1) < \frac{a}{4}.$$

Note that

$$A_2 \cap C_1 \in \mathcal{B}_2, P(A_2 \cap (A_1 - C_1) \leq P(A_1 - C_1) < \frac{a}{4},$$

whence

$$P(A_2 \cap C_1) > \frac{3}{4}a.$$

We next choose $C_2 \subset A_2 \cap C_1$, $C_2 \in \mathcal{B}_2$, C_2 compact in the topology τ_2, such that $P(A_2 \cap C_1 - C_2) < \frac{a}{4^2}$. Then $P(A_2 - C_2) \le \frac{a}{4} + \frac{a}{4^2}$. Note that

$$A_3 \cap C_2 \in \mathcal{B}_3, \quad P(A_3 \cap C_2) > a - (\frac{a}{4} + \frac{a}{4^2}).$$

Proceeding thus we get a decreasing sequence $(C_n)_{n=1}^{\infty}$ such that for all n, $C_n \in \mathcal{B}_n$, $C_n \subset A_n$, C_n compact in the topology on τ_n, and

$$P(A_n - C_n) < \frac{a}{4} + \frac{a}{4^2} + \cdots + \frac{a}{4^n}.$$

Clearly each C_n is non-empty. For each n choose an element q_n in C_n. Since C_n is compact the sequence $(\Pi_{nj} q_j)_{j=n}^{\infty}$ has a subsequence converging to a point in C_n. By Cantor's diagonal procedure it is possible to choose the sequence $(q_n)_{n=1}^{\infty}$ in such a way that for each i, $(\Pi_{ij} q_j)_{j=i}^{\infty}$ is convergent in the topology τ_i to an element p_i in $\mathbb{Q}_i$. By continuity of the map Π_{ij} we have $\Pi_{ij} p_j = p_i$ if $i \le j$, i.e., if $i \le j$ then $p_j \subset p_i$. By completeness of the filtration we conclude that $\cap_{i=1}^{\infty} p_i \neq \emptyset$. But

$$\cap_{i=1}^{\infty} p_i \subset \cap_{i=1}^{\infty} C_i \subset \cap_{i=1}^{\infty} A_i = \emptyset.$$

The contradiction proves the theorem.

Remark. (i) The requirement that the filtration be complete has been crucial in the above discussion. Here is an example due to S. M. Srivastava of a filtration on the real line which is not complete, but the quotient topologies are locally compact second countable. Let $\Omega = \mathbb{R}$, and let, for $n \ge 1$, $\mathbb{Q}_n = \{\{r\}, r < n, [n, \infty)\}$, i.e., the nth partition $\mathbb{Q}_n$ consists of all singletons less than n together with the interval $[n, \infty)$). Clearly $(\mathbb{Q}_n)_{n=1}^{\infty}$ is a filtration. For each n, the quotient topology on $\mathbb{Q}_n$ is isomorphic to the usual topology on $\mathbb{R}$. The set $C_n = [n, \infty)$ is compact in the quotient topology, $C_{n+1} \subset C_n$, but $\cap_{n=1}^{\infty} C_n$ is empty, so the filtration is not complete.
(ii) S. Bochner ([3]) has formulated and proved the Kolmogorov consistency theorem for projective families. One can derive the above version by proper identification of our sets and maps as a projective system, once topologies τ_n are described.
(iii) The totality of finitely additive measure on $\mathcal{B}_\infty$ which render the measure free martingale $f_n, n = 1, 2, 3, \cdots$ into a weak martingale is a convex

set whose extreme points are precisely those ν_∞ for which ν_1 is a point mass, and for which, for each n, the disintegration $\nu_n(\cdot,\cdot)$ of ν_{n+1} with respect to $\mathbb{Q}_n$ has the property that for each $q \in \mathbb{Q}_n$, $\nu(\cdot, q)$ is supported on at most two points.
(iv) We note that measure free martingales have some nice properties not shared by the usual martingales. If $f_n, n = 1, 2, 3, \cdots$ is a measure free martingale, then $[f_n], n = 1, 2, 3, \cdots$, where $[x]$ means the integral part of x, and $min\{f_n, K\}, k = 1, 2, 3, \cdots$, where K is fixed real number are also measure free martingale. In other words, measure free martingales are closed under discretization and truncation.

1.4. Equivalent Martingale Measures

Let $(\Omega, \mathcal{B})$ be a standard Borel space and let μ be a probability measure on $\mathcal{B}$. Let $(f_n)_{n=1}^{\infty}$ be a sequence of Borel measurable real valued functions on Ω, not necessarily a measure free martingale. In this section we discuss conditions, necessary as well as sufficient, for there to exist a measure ν, having the same null sets as μ, and which renders the sequence $(f_n)_{n=1}^{\infty}$ a martingale. Clearly, if such ν exists then we can modify $(f_n)_{n=1}^{\infty}$ on a ν-null set, which is therefore also μ-null, so that the new sequence of functions is a measure free martingale. Thus a necessary condition for the existence of a ν, equivalent to μ, under which $(f_n)_{n=1}^{\infty}$ is a martingale is that $(f_n)_{n=1}^{\infty}$ admit a modification on a μ-null set so that the resulting sequence is a measure free martingale.

Again assume that such a ν exists. For each n, let μ_n, ν_n respectively be the restriction of μ, ν to $\mathcal{B}_n$. Let $\mu_n(\cdot,\cdot), \nu_n(\cdot,\cdot)$ denote the disintegration of μ_{n+1}, ν_{n+1} with respect to the partition $\mathbb{Q}_n$. Since μ_{n+1} and ν_{n+1} have the same null sets, for μ_n almost every $q \in \mathbb{Q}_n$, $\mu_n(\cdot, q)$ and $\nu_n(\cdot, q)$ have the same null sets, and since

$$\int_q f_{n+1}(\omega)\nu_n(d\omega, q) = f_n(q),$$

it follows that for μ_n a.e. q

$$\mu_n(\{\omega \in q : f_{n+1}(\omega) \leq f_n(q)\}, q) > 0, \mu_n(\{\omega \in q : f_{n+1}(\omega) \geq f_n(q)\}, q) > 0. \tag{11}$$

Thus, if a ν equivalent to μ under which $(f_n)_{n=1}^{\infty}$ is a martingale exists, then for each n, for μ_n a.e. $q \in \mathbb{Q}_n$, (11) holds.

Fix n, fix a $q \in \mathbb{Q}_n$, and write $m = \mu_n(\cdot, q)$. Assume that

$$\int_q | f_{n+1}(\omega) | m(d\omega) < \infty,$$

and that

$$m(\{\omega \in q : f_{n+1}(\omega) < f_n(q)\}) > 0, m(\{\omega \in q : f_{n+1}(\omega) > f_n(q)\}) > 0.$$

Write

$$E = \{\omega \in q : f_{n+1}(\omega) < f_n(q)\}, F = \{\omega \in q : f_{n+1}(\omega) > f_n(q)\},$$

$$m(E) = \alpha, m(F) = \beta,$$

$$c = \frac{\int_E f_{n+1}(\omega)m(d\omega)}{\alpha}, d = \frac{\int_F f_{n+1}(\omega)m(d\omega)}{\beta}.$$

Note that $c < f_n(q) < d$ and with $a = \frac{d - f_n(q)}{d-c}, b = \frac{f_n(q) - c}{d-c}$, we have

$$a + b = 1, ac + bd = f_n(q).$$

Define $\nu'_n(\cdot, \cdot)$ as follows:

$$\frac{d\nu'_n(\cdot, q)}{dm} = (\frac{a}{\alpha} 1_E + \frac{b}{\beta} 1_F).$$

Note here that q ranges over $\mathbb{Q}_n$, so that m will vary with it. Further, $a, \alpha, b, \beta, m = \mu_n(\cdot, q)$ are measurable functions of q, so that $\nu'_n(\cdot, \cdot)$ is a transition probability. If $m(\{\omega \in q : f_{n+1}(\omega) = f_n(q)\}) = 0$, then $\nu'_n(\cdot, q)$ and m are equivalent, and

$$\int_q f_{n+1}(\omega)\nu'_n(d\omega, q) = ac + bd = f_n(q).$$

In any case, if we set $s = \{\omega \in q : f_{n+1}(\omega) = f_n(q)\}$, then $s \in \mathbb{Q}_{n+1}$, so that we can speak of Dirac measure $\delta_{\{s\}}$ at s, and consider the measure $\nu_n(\cdot, q)$ defined by:

$$\nu_n(\cdot, q) = (1 - m(s))\nu'_n(\cdot, q) + \delta_{\{s\}} m(s)$$

The measure $\nu_n(\cdot, q)$ and m have the same null sets, $\nu_n(\cdot, q)$ is measurable in q, and

$$\int_q f_{n+1}(\omega)\nu_n(d\omega, q) = f_n(q)$$

We define inductively,

$$\nu_1 = \mu_1, \nu_2 = \int_\Omega \nu_1(\cdot, q)\nu_1(dq), \cdots, \nu_{n+1}(\cdot) = \int_\Omega \nu_n(\cdot, q)\nu_n(dq).$$

Then for each n, ν_n is a probability measure on $\mathcal{B}_n$, equivalent to μ_n,

$$\nu_{n+1} \mid_{\mathcal{B}_n} = \nu_n, \quad E_n(f_{n+1} \mid \mathcal{B}_n) = f_n,$$

where E_n is the conditional expectation operator in $(\Omega, \mathcal{B}_{n+1}, \nu_{n+1})$.

If the filtration $(\mathbb{Q}_n)_{n=1}^\infty$ is complete, the naturally defined measure ν_∞ on $\mathcal{B}_\infty$ has a countably additive extension, say ν, to all of $\mathcal{B}$. However, in general, ν need not be equivalent to μ (see example 4.1. below). If there are positive constants A and B such that for all n, $A < \frac{d\nu_n}{\mu_n} < B$, then clearly the ν_∞ defined on $\mathcal{B}_\infty$ has an extension to $\mathcal{B}$ which has the same null sets as μ. We have proved:

Theorem 4.1.

(a) *Let $(f_n)_{n=1}^\infty$ be a sequence of Borel measurable functions on $(\Omega, \mathcal{B}, \mu)$ and let $\mathbb{Q}_n, \mathcal{B}_n$, $\mu_n, \mu_n(\cdot, \cdot)$ be as above. If there is a probability measure ν equivalent to μ with respect to which $(f_n)_{n=1}^\infty$ is a martingale, then $(f_n)_{n=1}^\infty$ can be modified on a μ-null set so that the resulting sequence of functions is a measure free martingale. Further, for each n, for almost every $q \in \mathbb{Q}_n$, the sets $\{\omega \in q : f_{n+1}(\omega) \leq f_n(q)\}, \{\omega \in q : f_n(q) \geq f_{n+1}(\omega)\}$ have positive $\mu_n(\cdot, q)$- measure.*

(b) *If for every n, and for almost every $q \in \mathbb{Q}_n$, the sets $\{\omega \in q : f_{n+1}(\omega) \leq f_n(q)\}, \{\omega \in q : f_n(q) \leq f_{n+1}(\omega)\}$ have positive $\mu_n(\cdot, q)$-measure, then for each n we have a ν_n equivalent to μ_n such that $\nu_{n+1} \mid_{\mathcal{B}_n} = \nu_n$, $E_{n+1}(f_{n+1} \mid \mathcal{B}_n) = f_n$, where E_{n+1} stands for the conditional expectation operator on $(\Omega, \mathcal{B}_{n+1}, \nu_{n+1})$ with respect to $\mathcal{B}_n$.*

(c) *Finally, if there are positive constants A and B such that for all n, $A < \frac{d\nu_n}{d\mu_n} < B$, then the naturally defined ν_∞ on $\mathcal{B}_\infty$ has an extension ν to $\mathcal{B}$ which has the same null sets as μ and $(f_n)_{n=1}^\infty$ is a martingale with respect to ν.*

PERSPECTIVES IN MATHEMATICAL SCIENCES I - Probability and Statistics

The equivalent martingale measure ν in the above theorem is obtained by a rather naive modification of μ. Indeed, $\mu_n(\cdot, q)$ is only rescaled over the sets $\{\omega \in q : f_{n+1}(\omega) < f_n(q)\}, \{\omega \in q : f_n(q) > f_{n+1}(\omega)\}$, so that if $\mu_n(\cdot, q)$ is not 'well distributed ' on these sets, then this persists with $\nu_n(\cdot, q)$. This circumstance can be changed if we assume that for each n and for each $q \in \mathbb{Q}_n$, f_{n+1} is bounded on q, in addition to the requirement that sets $\{\omega \in q : f_{n+1}(\omega) \leq f_n(q)\}, \{\omega \in q : f_n(q) \geq f_{n+1}(\omega)\}$ have positive $\mu_n(\cdot, q)$ measure. We know from entropy considerations of section 1.1 that there exists a unique $c_n = c_n(q)$ such that

$$\frac{\int_q f_{n+1}(\omega) e^{c_n(q) f_{n+1}(\omega)} \mu_n(d\omega, q)}{\int_q e^{c_n(q) f_{n+1}(\omega)} \mu_n(d\omega, q)} = f_n(q).$$

The function $q \to c_n(q)$ is $\mathcal{B}_n$ measurable. We set, for each n and for each $q \in \mathbb{Q}_n$,

$$d\nu_n(\cdot, q) = \frac{e^{c_n(q) f_{n+1}(\omega)}}{\int_q e^{c_n(q) f_{n+1}(\omega)} \mu_n(d\omega, q)} \cdot d\mu_n(\cdot, q),$$

$$\nu_1 = \mu_1, \nu_{n+1} = \int_\Omega \nu_n(\cdot, q) \nu_n(dq).$$

We change notation and write $\nu_n = B_n$. The natural finitely additive measure B_∞ on $\mathcal{B}_\infty$ renders $(f_n)_{n=1}^\infty$ into a weak martingale. If we assume that there are positive constants A, C such that for all n, $A < \frac{dB_n}{d\mu_n} < C$, then B_∞ extends to a countably additive measure B on $\mathcal{B}$, B equivalent to μ, and, $(f_n)_{n=1}^\infty$ is a martingale with respect to B.

We may summarise this as

Theorem 4.2. *If for each n, for μ_n a.e. $q \in \mathbb{Q}_n$,*

(i) *f_{n+1} is bounded on q,*

(ii) *$\mu(\{\omega \in q : f_{n+1}(\omega) \leq f_n(q)\}, q) > 0$, $\mu(\{\omega \in q : f_{n+1}(\omega) \geq f_n(q)\}, q) > 0$,*

then there exists a unique finitely additive measure B_∞ on $\mathcal{B}_\infty$ such that for each n,

(a) *the restriction B_n of B_∞ to $\mathcal{B}_n$ is countably additive and equivalent to μ_n,*

(b) *B_{n+1} maximizes the relative entropy $H(\frac{d\lambda_{n+1}}{d\mu_{n+1}}, \mu_{n+1})$ among all finitely additive probability measures λ on $\mathcal{B}_\infty$ which render $(f_n)_{n=1}^\infty$ into a weak martingale and such that $\lambda_1 = \mu_1$, λ_n equivalent to μ_n for all n,*

(c) *if there exist constants* A, B *such that for each* n, $A < \frac{dB_n}{d\mu_n} < B$, *then* B_∞ *extends to a countably additive measure measure* B *on* $\mathcal{B}$, B *and* ν *are equivalent, and* $(f_n)_{n=1}^{\infty}$ *is a martingale under* B.

Remarks. 1) The measure B, however, need not maximize the relative entropy $H(\frac{d\lambda}{d\mu}, \mu)$ among all measures λ on $\mathcal{B}$ equivalent to μ and under which $(f_n)_{n=1}^{\infty}$ is a martingale.
2) One may call B a Boltzmann measure equivalent to μ and the associated sequence $(f_n)_{n=1}^{\infty}$ a Boltzmann martingale.

We now give a more general condition for the existence of a martingale measure equivalent to a given μ than the one given in Theorem 4.1. (c). Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions on $(\Omega, \mathcal{B}, \mu)$ for which there exists a measure ν_∞ on $\mathcal{B}_\infty$ such that (i) $(f_n)_{n=1}^{\infty}$ is a weak martingale with respect to ν_∞, (ii) for each n, μ_n and ν_n are equivalent, (iii) $\mu_1 = \nu_1$.

Now the Radon-Nikodym derivative $\frac{d\nu_n}{d\mu_n}$ is computed as follows:

$$\frac{d\nu_{n+1}}{d\mu_{n+1}}(\omega) = \frac{d\nu_n(\cdot, q_n)}{d\mu_n(\cdot, q_n)}(\omega)\frac{d\nu_n}{d\mu_n}(\omega), \omega \in q_n \in \mathbb{Q}_n.$$

So, on iteration, we have:

$$\frac{d\nu_{n+1}}{\mu_{n+1}}(\omega) = (\Pi_{i=1}^{n}\frac{d\nu_i(\cdot, q_i)}{d\mu_i(\cdot, q_i)}(\omega)) \times \frac{d\nu_1}{d\mu_1}, \omega \in q_i \in \mathbb{Q}_i, i = 1, 2, \cdots, n.$$

Since we have chosen $\nu_1 = \mu_1$, $\frac{d\nu_1}{d\mu_1}(\omega) = 1$ for all ω, so that

$$\frac{d\nu_{n+1}}{\mu_{n+1}}(\omega) = \Pi_{i=1}^{n}\frac{d\nu_i(\cdot, q_i)}{d\mu_i(\cdot, q_i)}(\omega), \omega \in q_i \in \mathbb{Q}_i, i = 1, 2, \cdots, n.$$

Further,

$$E_\mu(\frac{d\nu_{n+1}}{d\mu_{n+1}} \mid \mathcal{B}_n)(\omega) = \frac{d\nu_n}{d\mu_n}(\omega)$$

where E_μ denotes the conditional expectation operator with respect to μ.

Indeed for any set $A \in \mathcal{B}_n$,

$$\begin{aligned}\int_A \frac{d\nu_{n+1}}{d\mu_{n+1}}(\omega)\mu(d\omega) &= \int_A \frac{d\nu_n}{d\mu_n}(\omega)\frac{d\nu_n(\omega,q)}{d\mu_n(\omega,q)}d\mu \\ &= \int_A \frac{d\nu_n}{d\mu_n}(\omega)\frac{d\nu_n(\omega,q)}{d\mu_n(\omega,q)}d\mu_{n+1} \\ &= \int_A \frac{d\nu_n}{d\mu_n}(\omega)(\int_q \frac{d\nu_n(\omega,q)}{d\mu_n(\omega,q)}\mu_n(d\omega,q))\mu_n(d\omega) \\ &= \int_A \frac{d\nu_n}{d\mu_n}(\omega)\mu_n(d\omega) \\ &= \int_A \frac{d\nu_n}{d\mu_n}(\omega)\mu(d\omega).\end{aligned}$$

The sequence of functions $(g_n = \frac{d\nu_n}{d\mu_n})_{n=1}^{\infty}$ is therefore a martingale of non-negative functions on the probability space $(\Omega, \mathcal{B}, \mu)$, so converges μ a.e. to a function g. If $\int_\Omega g(\omega)\mu(d\omega) = 1$, or if the sequence $(g_n)_{n=1}^{\infty}$ is uniformly integrable, then, by martingale convergence theorem (see [4], p. 319), for each n, $g_n = E_\mu(g \mid \mathcal{B}_n)$, equivalently, for each n, $gd\mu \mid_{\mathcal{B}_n} = d\nu_n$ and so ν_∞ is countably additive and extends to a measure ν on $\mathcal{B}$, with $\frac{d\nu}{d\mu} = g$. Further $(f_n)_{n=1}^{\infty}$ is a martingale with respect to it. If, in addition, $g > 0$ a.e. μ, then ν is equivalent to μ and $(f_n)_{n=1}^{\infty}$ is a martingale with respect to it.

We have proved:

Theorem 4.3.

(a) *The sequence* $(g_n = \frac{d\nu_n}{d\mu_n})_{n=1}^{\infty}$ *of Radon-Nikodym derivatives is a martingale of non-negative functions and converges* μ *a.e. to a function* g.

(b) *If* $\int_\Omega g(\omega)\mu(d\omega) = 1$ *or if the* g_n*'s are uniformly integrable with respect to* μ, *then* $(f_n)_{n=1}^{\infty}$ *is a martingale with respect to* ν *given by* $d\nu = gd\mu$. *If in addition* $g > 0$ *a.e.* μ, ν *is equivalent to* μ.

(c) *If* $\sum_{n=1}^{\infty} \mid 1 - \frac{d\nu_n(\cdot,q)}{d\mu_n(\cdot,q)} \mid < \infty$ a.s. μ *then* $g > 0$ *a.e.* μ.

Write

$$H(\frac{d\nu_{n+1}}{d\mu_{n+1}}, \mu_{n+1}) \mid \mathbb{Q}_n) = \int_{\mathbb{Q}_n} H(\frac{d\nu_n(\cdot,q)}{d\mu_n(\cdot,q)}, \mu_n(\cdot,q))\nu_n(dq).$$

We know from formula (7) of section 1.2.3 that

$$H(\frac{d\nu_{n+1}}{d\mu_{n+1}}, \mu_{n+1}) = H(\frac{d\nu_{n+1}}{d\mu_{n+1}}, \mu_{n+1}) \mid \mathbb{Q}_n) + H(\frac{d\nu_n}{d\mu_n}, \mu_n).$$

Iterating we get

$$H(\frac{d\nu_{n+1}}{d\mu_{n+1}}, \mu_{n+1}) = \sum_{i=1}^{n+1} H(\frac{d\nu_i}{d\mu_i}, \mu_i) \mid \mathbb{Q}_{i-1}).$$

(Here, when $i = 1$, $\mathbb{Q}_{i-1} = \mathbb{Q}_0$ which we take to be the trivial partition $\{\emptyset, \Omega\}$.)

Since $(g_n = \frac{d\nu_n}{d\mu_n})_{n=1}^{\infty}$ is a martingale with respect to μ, the sequence $(g_n \log^+ g_n)_{n=1}^{\infty}$ is a submartingale provided, for each n, $E_\mu(g_n \log^+ g_n)$ is finite ([4], p. 296). We assume that this is the case. Then

$$E_\mu(g_n \log^+ g_n) \leq E_\mu(g_{n+1} \log^+ g_{n+1}), n = 1, 2, \cdots,$$

so that $\lim_{n\to\infty} E_\mu(g_n \log^+ g_n)$ exists, which may be finite or infinite. We assume that this limit is finite, say c. Since $(g_n \log^+ g_n)_{n=1}^{\infty}$ is a submartingale of non-negative functions it has a limit which is indeed $g \log^+ g$, since $(g_n)_{n=1}^{\infty}$ has limit g, μ a.e. Moreover by Fatou's lemma, $E_\mu(g \log^+ g) \leq c$. Assume that the sequence $(g_n)_{n=1}^{\infty}$ is uniformly integrable so that this sequence together with g forms a martingale. From martingale theory ([4], p. 296) the sequence $(g_n \log^+ g_n)_{n=1}^{\infty}$ together with the function $g \log^+ g$ is a submartingale of non-negative functions, so, again from martingale theory ([4], p. 325) we conclude that

$$\lim_{n\to\infty} E_\mu(g_n \log g_n^+) = E_\mu(g \log^+ g).$$

Since $g_n \log^- g_n$ remains bounded independent of n (which is the case at ω where $g_n(\omega) \leq 1$), we also have

$$\lim_{n\to\infty} \int_\Omega g_n(\omega) \log^- g_n(\omega) \mu_n(d\omega) = \int_\Omega g \log^-(\omega) \mu(d\omega).$$

Thus we have

$$\lim_{n\to\infty} H(\frac{d\nu_n}{d\mu_n}, \mu_n) = H(\frac{d\nu}{d\mu}, \mu).$$

We have proved:

Theorem 4.4.

(a) *Assume that the martingale* $(g_n = \frac{d\nu_n}{d\mu_n})_{n=1}^{\infty}$ *is uniformly integrable and that* $\int_\Omega g_n \log^+ g_n \mu_n(d\omega) \leq c$, *for some real number* c. *Then* $\lim_{n\to\infty} H(\frac{d\nu_n}{d\mu_n}, \mu_n)$ *exists and we have:*

$$\lim_{n\to\infty} H(\frac{d\nu_n}{d\mu_n}, \mu_n) = \sum_{i=1}^{\infty} H(\frac{d\nu_i}{d\mu_i}, \mu_i) \mid \mathbb{Q}_{i-1}) = H(\frac{d\nu}{d\mu}, \mu).$$

(b) *In addition to the hypothesis and notations of Theorem 4.2, assume that the martingale* $(\frac{dB_n}{d\mu_n})_{n=1}^{\infty}$ *is uniformly integrable and that the sequence of relative entropies* $(H(\frac{dB_n}{d\mu_n}, \mu_n))_{n=1}^{\infty}$ *remains bounded. Then* $\lim_{n\to\infty} H(\frac{dB_n}{d\mu_n}, \mu_n)$ *exists and we have:*

$$\lim_{n\to\infty} H(\frac{dB_n}{d\mu_n}, \mu_n) = \sum_{n=1}^{\infty} H(\frac{dB_n}{d\mu_n}, \mu_n) \mid \mathbb{Q}_{n-1}) = H(\frac{dB}{d\mu}, \mu).$$

Among all ν *absolutely continuous with respect to* μ *under which* $(f_n)_{n=1}^{\infty}$ *is a martingale and* $\nu_1 = \mu_1$, *B is the unique one which maximizes, for each* n, *the relative entropy entropy* $H(\frac{d\nu_n}{d\mu_n}, \mu_n)$.

Example 4.1. Consider $\mathbb{R}^2$ together with the measure $\mu = \sigma \times \sigma$ where σ is the normal distribution with mean zero and variance one. Let $\mathbb{Q}_1$ be the partition $\{x\} \times \mathbb{R}, x \in \mathbb{R}$. Let $f_i, i = 1, 2$ be the co-ordinate maps. The partition of $\mathbb{R}^2$ given by f_1 is $\mathbb{Q}_1$. The distribution ν on $\mathbb{R}^2$ equivalent to μ, satisfying $E_\nu(f_2 \mid f_1) = E_\nu(f_2 \mid \mathbb{Q}_1) = f_1$, and maximizing the relative entropy with respect to μ is the bivariate distribution of $(f_1, f_1 + f_2)$, where f_1, f_2 are independent with normal distribution of mean zero and variance one.

More generally, let $\mathbb{R}^n$ be given the measure $\mu = \sigma^n$, the n-fold product of σ. Let $f_1, f_2, \cdots, f_n$ be the co-ordinate random variables. Then

$$\mathbb{Q}_i = \{\{\omega_1, \omega_2, \cdots, \omega_i\} \times \mathbb{R}^{n-i} : (\omega_1.\omega_2, \cdots, \omega_i) \in \mathbb{R}^i\}$$

is the partition of $\mathbb{R}^n$ given by the functions $f_1, f_2, \cdots, f_i$. Let ν_n be the measure induced on $\mathbb{R}^n$ by the vector random variable $(f_1, f_1 + f_2, \cdots, f_1 + f_2 + \cdots + f_n)$ where $(f_1, f_2, \cdots .f_n)$ has distribution $\mu = \sigma^n$. Then, among all probability measures λ on $\mathbb{R}^n$ equivalent to μ and satisfying $E_\lambda(f_{i+1} \mid f_i) = f_i, 1 \le i \le n-1$, ν_n is the unique one which simultaneously maximizes the relative entropies

$$-\int_{\mathbb{R}^i} \frac{d\lambda_i}{d\mu_i}(\omega) \log \frac{d\lambda_i}{d\mu_i}(\omega)\mu_i(d\omega),$$

$1 \le i \le n$, where μ_i, λ_i are respectively the measures μ and λ restricted to the the σ-algebra $\mathcal{B}_i$ generated $f_1, f_2, \cdots, f_i$.

Finally, let (i) $\Omega = \mathbb{R}^{\mathbb{N}}$, where $\mathbb{N}$ is the set of natural numbers, (ii) $\mu =$ countable product of σ with itself, (iii) for each n, $f_n =$ projection on the

nth co-ordinate space. The partition $\mathbb{Q}_n$ of Ω generated by $f_1, f_2, \cdots, f_n$ is the collection

$$\{\{(\omega_1, \omega_2, \cdots, \omega_n)\} \times \mathbb{R}^{\{n+1,n+2,\cdots\}} : (\omega_1, \omega_2, \cdots, \omega_n) \in \mathbb{R}^n.\}$$

For each n, let ν_n denote the measure on $\mathcal{B}_n$, induced by the martingale $f_1, f_1 + f_2, \cdots, \sum_{i=1}^n f_i)$, where $f_1, f_2, \cdots, f_n$ are independent random variable, each with distribution σ. Let ν_∞ be the measure on the algebra $\cup_{n=1}^{\infty} \mathcal{B}_n$ whose restriction to each $\mathcal{B}_n$ is ν_n. Then among all probability measures λ_∞ on $\mathcal{B}_\infty$ which satisfies (a) for each n, μ_n and λ_n are equivalent, (b) for each n, $E_{\lambda_n}(f_{n+1} \mid f_n) = f_n$, (c) $\lambda_1 = \sigma$, the measure ν_∞ is the unique one which maximizes simultaneously the relative entropies $H(\frac{d\lambda_n}{d\mu_n}, \mu_n)$, $n = 1, 2, \cdots$. The extension ν of ν_∞ to the Borel σ-algebra of $\mathbb{R}^{\mathbb{N}}$ is, however, singular to μ.

References

[1] Athreya, K. B. (1994). *Entropy Maximization*, Preprint Series No 1231, April 1994, Institute for Mathematics and Applications, University of Minnesota.

[2] Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*, Springer.

[3] Bochner, S. (1960). *Harmonic Analysis and The Theory of Probability*, Dover.

[4] Doob, J. L. (1955). *Stochastic Processes*, John Wiley.

[5] Durret, S. (1991). *Probability: Theory and Examples*, Wadsworth and Brooks-Cole, Pacific Grove, CA.

[6] Karandikar, R. L. and Nadkarni, M. G. (2005). Measure Free Martingales. *Proc. Indian Acad. Sci (Math. Sci)*. **15** 111–116.

[7] Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press.

[8] Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, Second Edition. John Wiley.

[9] Srivastava, S. M. (1998). *A course on Borel Sets*. Graduate Texts in Mathematics, 180, Springer-Verlag, New York.

[10] Stroock, D. (1993). *Probability Theory, An Analytic View*. Cambridge University Press.

[11] Bhatia, R. (2008). A Conversation with S. R. S. Varadhan, Varadhan's Interview by Rajendra Bhatia. *Mathematical Intellingencer*. **30** 24–42.