Proc. Indian Acad. Sci. (Math. Sci.) Vol. 119, No. 4, September 2009, pp. 531–539.

Entropy maximization

K B ATHREYA

Department of Mathematics and Statistics, Iowa State University, USA
IMI, Department of Mathematics, Indian Institute of Science, Bangalore 560 012, India
E-mail: kba@iastate.edu

MS received 22 October 2008; revised 15 January 2009

Abstract. It is shown that (i) every probability density is the unique maximizer of relative entropy in an appropriate class and (ii) in the class of all pdf f that satisfy $\int f h_i \mathrm{d}\mu = \lambda_i$ for $i = 1, 2, \ldots, \ldots k$ the maximizer of entropy is an f_0 that is proportional to $\exp(\sum c_i h_i)$ for some choice of c_i. An extension of this to a continuum of constraints and many examples are presented.

Keywords. Entropy; relative entropy; entropy maximization.

Let $(\Omega, \mathcal{B}, \mu)$ be a measure space. A $\mathcal{B}$ measurable function f from Ω to $R^+ = [0, \infty)$ is called a probability density function (pdf) if $\int f \mathrm{d}u = 1$. For such an f, let $P_f(A) \equiv \int_A f \mathrm{d}\mu$ for $A \in B$. Then $P_f(\cdot)$ is a probability measure. The *entropy of* P_f *relative to* μ is defined by

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

provided the integral on the right exists

If f_1 and f_2 are two pdfs on $(\Omega, \mathcal{B}, \mu)$ then for all ω (we define $0 \log 0 = 0$),

$$f_1(\omega) \log f_2(\omega) - f_1(\omega) \log f_1 \omega) \le (f_2(\omega) - f_1(\omega)). \tag{2}$$

To see this, note that the function $f(x) = x - 1 - \log x$ has a unique minimum at $x = 1$. This implies that $f(x)$ is positive for all x different from one and at $x = 1$ it is zero.

Now integrating (2) yields

$$\int_\Omega f_1(\omega) \log f_2(\omega) \mathrm{d}\mu - \int f_1(\omega) \log f_1(\omega) \mathrm{d}\mu$$

$$\le \int_\Omega (f_2(\omega) - f_1(\omega)) \mathrm{d}\mu = 0 \tag{3}$$

since

$$\int_\Omega f_1 \mathrm{d}\mu = 1 = \int_\Omega f_2 \mathrm{d}\mu.$$

We note that in view of (2), equality holds in (3) iff equality holds in (2) and that holds iff $f_2(\omega) = f_1(\omega)$ a.e. This simple idea is well-known in the literature and is mentioned in Durret (p. 318 of [1]). We summarize the above discussion as follows.

PROPOSITION 1

Let $(\Omega, \mathcal{B}, \mu)$ *be a measure space. Let* f_1 *and* f_2 *be* B *measurable functions from* Ω *to* $R^+ = [0, \infty)$ *such that* $\int f_1(\omega)\mathrm{d}\mu = 1 = \int f_2(\omega)\mathrm{d}\mu$. *Then*

$$H(f_1, \mu) = -\int f_1(\omega) \log f_1(\omega)\mathrm{d}\mu \leq -\int f_1(\omega) \log f_2(\omega)\mathrm{d}\mu \tag{4}$$

with equality holding iff $f_1(\omega) = f_2(\omega)$ a.e.

Let f_0 be a pdf such that $\lambda = -\int f_0 \log f_0 \mathrm{d}\mu$ exists in R. Let

$$\mathcal{F}_\lambda \equiv \left\{ f \colon f \text{ a pdf and } -\int f \log f_0 \mathrm{d}\mu = \lambda \right\}. \tag{5}$$

From (4) it follows that for $f \in \mathcal{F}_\lambda$,

$$H(f, \mu) = -\int f \log f \mathrm{d}\mu \leq -\int f \log f_0 \mathrm{d}\mu = -\int f_0 \log f_0 \mathrm{d}\mu.$$

Thus we get the following.

COROLLARY 1

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

and f_0 *is the unique maximizer.*

Remark 1. The above corollary says that any probability density f_0 such that $-\int f_0 \log f_0 \mathrm{d}\mu \equiv \lambda$ is defined appears as the unique solution to an entropy maximization problem in an appropriate class of densities. Of course, this has some meaning only if $\mathcal{F}_\lambda$ does not consist of f_0 alone.

A useful reformulation of Corollary 1 is as follows.

COROLLARY 2

Let $h \colon \Omega \to R$ *be* B *measurable. Let* λ *and* c *real be such that*

$$\psi(c) \equiv \int \mathrm{e}^{ch}\mathrm{d}\mu < \infty, \qquad \int |h|\mathrm{e}^{ch}\mathrm{d}\mu < \infty,$$

$$\lambda \int \mathrm{e}^{ch}\mathrm{d}\mu = \int h\mathrm{e}^{ch}\mathrm{d}\mu. \tag{6}$$

Let

$$f_0 = \frac{\mathrm{e}^{ch}}{\psi(c)}. \tag{7}$$

Then, let $\mathcal{F}_\lambda = \{f \colon$ *a pdf and* $\int fh\mathrm{d}\mu = \lambda\}$. *Then* $\sup\{H(f, \mu) \colon f \in \mathcal{F}_\lambda\} = -\int f_0 \log f_0 \mathrm{d}\mu$ *and* f_0 *is the unique maximizer.*

As an application of the above corollary we get the following examples.

Example 1. $\Omega = \{1, 2, \dots, N\}$, $N < \infty$, μ counting measure, $h \equiv 1$, $\lambda = 1$, $\mathcal{F} \equiv \{\{p_i\}_1^N, p_i \geq 0, \sum_1^N p_i = 1\}$.

For any c real (6) holds and (7) becomes

$$f_0(j) = \frac{1}{N}, \quad j = 1, 2, \dots, N, \text{ i.e. } f_0 \text{ is the 'uniform' density.}$$

Example 2. $\Omega = \{1, 2, \dots, N\}$, $N < \infty$, μ counting measure, $h(j) \equiv j$, $1 \leq \lambda \leq N$, $\mathcal{F} \equiv \{\{p_i\}_1^N, p_i \geq 0, \sum_1^N p_i = 1 \sum_1^N jp_j = \lambda\}$. The optimal f_0 is $f_0(j) = p^{j-1}\frac{(p-1)}{(p^N-1)}$ where $p > 0$ is the unique solution of $\sum_1^N (j - \lambda)p^{j-1} = 0$. Since $\varphi(p) = \frac{\sum_1^N jp^{j-1}}{\sum_1^N p^{j-1}}$ is continuous and strictly nondecreasing in $(0, \infty)$ (see Remark 2 below), $\lim_{p\downarrow 0} \varphi(p) = 1$ and $\lim_{p\uparrow\infty 0} \varphi(p) = N$, for each λ in $[1, N]$, there exists a unique p in $(0, \infty)$ such that $\varphi(p) = \lambda$. This f_0 is the conditional geometric (given that '$X \leq N$').

Example 3. $\Omega = \{1, 2, \dots\}$, μ counting measure, $h(j) = j$, $1 \leq \lambda < \infty$, $\mathcal{F}_\lambda = \{\{p_i\}_i^\infty, p_i \geq 0, \sum_1^\infty p_i = 1, \sum_1^\infty jp_j = \lambda\}$. The optimal f_0 is $f_0(j) = (1 - p)p^{j-1}$ where $p = 1 - \frac{1}{\lambda}$. This f_0 is the unconditional geometric.

Example 4. $\Omega = \{1, 2, \dots, N\}$, $N \leq \infty$, μ counting measure, $h(j) = j^2$, $1 < \lambda < \infty$, $\mathcal{F}_\lambda = \{\{p_i\}, p_i \geq 0, \sum_1^N p_i = 1, \sum_1^N j^2 p_j = \lambda\}$. The optimal f_0 is the 'discrete folded normal' $f_0(j) = \frac{\mathrm{e}^{-cj^2}}{\sum_1^N \mathrm{e}^{-cj^2}}$ for some $c > 0$ such that

$$\sum_1^N j^2 \mathrm{e}^{-cj^2} = \lambda \sum_1^N \mathrm{e}^{-cj^2}.$$

Since $\varphi(c) = \frac{\sum_1^N j^2\mathrm{e}^{-cj^2}}{\sum_1^N \mathrm{e}^{-cj^2}}$ is continuous and strictly nondecreasing in $(0, \infty)$ (see Remark 2 below), $\lim_{c\downarrow -\infty} \varphi(c) = N^2$ and $\lim_{c\uparrow\infty} \varphi(c) = 1$, for each $1 < \lambda < N^2$ there is a unique c in $(-\infty, \infty)$ such that $\varphi(c) = \lambda$. For $\lambda = 1$ or N^2, $\mathcal{F}_\lambda$ is a singleton.

Example 5. $\Omega = R^+ = [0, \infty)$, μ = Lesbesgue measure, $h(x) \equiv x$, $0 < \lambda < \infty$, $\mathcal{F}_\lambda = \{f = f \geq 0, \int_0^\infty f(x)\mathrm{d}x = 1, \int_0^\infty xf(x)\mathrm{d}x = \lambda\}$. The optimal f_0 is $f_0(x) = \frac{1}{\lambda}\mathrm{e}^{-x/\lambda}$, i.e., the exponential density with mean λ.

Example 6. $\Omega = R$, μ = Lesbesgue measure, $h(x) \equiv x^2$, $0 < \lambda < \infty$, $\mathcal{F}_\lambda = \{f\colon f \geq 0, \int_{-\infty}^\infty f(x)\mathrm{d}x = 1, \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \lambda\}$. The optimal f_0 is $\frac{1}{\sqrt{2\pi\lambda}}\mathrm{e}^{-(x^2/2\lambda)}$, i.e., the normal density with mean 0 and variance λ.

Example 7. $\Omega = R$, μ = Lesbesgue measure, $h(x) = \log(1 + x^2)$, $0 < \lambda < \infty$, $\mathcal{F}_\lambda = \{f\colon f \geq 0, \int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1, \int_{-\infty}^{+\infty} f(x)\log(1+x^2)\mathrm{d}x = \lambda\}$. Let $c > 1/2$ be such that

$$\int \frac{\log(1 + x^2)}{(1 + x^2)^c}\mathrm{d}x = \lambda \int \frac{1}{(1 + x^2)^c}\mathrm{d}x.$$

Then the optimal f_0 is $f_0(x)\alpha\frac{1}{(1+x^2)^c}$ (α means proportional to). If $\lambda = \frac{1}{\pi}\int \frac{\log(1+x^2)}{(1+x^2)}$ $\mathrm{d}(x)$, then f_0 is the Cauchy (0, 1) density.

Since $\varphi(c) = \left(\int \frac{\log(1+x^2)}{(1+x^2)^c}\mathrm{d}(x)\right)\Big/\left(\int \frac{1}{(1+x^2)^c}\mathrm{d}x\right)$ is continuous and strictly decreasing in $\left(\frac{1}{2}, \infty\right)$ (see Remark 2 below), $\lim_{c\downarrow\frac{1}{2}} \varphi(c) = \infty$ and $\lim_{c\uparrow\infty} \varphi(c) = 0$, for each $0 < \lambda < \infty$ there is a unique c in $\left(\frac{1}{2}, \infty\right)$ such that $\varphi(c) = \lambda$.

Remark 2. The claim made about the properties of φ in Examples 2, 4 and 7 is justified as follows. Let h: $\Omega \to R$ be $\mathcal{B}$ measurable and $\psi(c) = \int \mathrm{e}^{ch}\mathrm{d}\mu$ and $I_h = \{c\colon \psi(c) < \infty\}$. It can be shown that I_h is a connected set in R, i.e. an interval [4] that could be empty, a single point, an interval that is half open, fully open, closed, semi-infinite, finite. If I_h has a nonempty interior I_h^0 then in I_h^0, $\psi(\cdot)$ is infinitely differentiable with $\psi'(c) = \int h\mathrm{e}^{ch}\mathrm{d}\mu$, $\psi''(c) = \int h^2\mathrm{e}^{ch}\mathrm{d}\mu$. Further,

$$\psi(c) = \frac{\psi'(c)}{\psi(c)} \text{ satisfies,} \tag{8}$$

$$\psi'(c) = \frac{\psi''(c)}{\psi(c)} - \left(\frac{\psi'(c)}{\psi(c)}\right)^2 = \text{variance of } X_c > 0, \tag{9}$$

where X_c is the random variable $h(\omega)$ with density $g_c = \frac{\mathrm{e}^{ch}}{\psi(c)}$ with respect to μ.

Thus for any $\inf_{I_h^0}\varphi(c) < \lambda < \sup_{I_h^0} \varphi(c)$ there is a unique c such that $\varphi(c) = \lambda$.

Remark 3. Examples 1, 3, 5 and 6 are in Shannon [5] where the method of Lagrange multiplier is used

Corollary 2 can be generalized easily.

COROLLARY 3

Let $h_1, h_2, \ldots, h_k$ be $\mathcal{B}$ measurable functions from Ω to R and $\lambda_1, \lambda_2, \ldots, \lambda_k$, c_1, c_2, c_k be real numbers such that

$$\int \mathrm{e}^{\sum_1^k c_i h_i}\mathrm{d}\mu < \infty, \quad \int \left(\sum_1^k |h_j|\right) \mathrm{e}^{\sum_1^k c_i h_i}\mathrm{d}\mu < \infty \tag{10}$$

and

$$\int h_j \mathrm{e}^{\sum_1^k c_i h_i}\mathrm{d}\mu = \lambda_j \int \mathrm{e}^{\sum_1^k c_i h_i}\mathrm{d}\mu, \quad j = 1, 2, \ldots, k. \tag{11}$$

Let $f_0 \alpha \mathrm{e}^{\sum_1^k c_i h_i}$ *and*

$$\mathcal{F} \equiv \left\{f\colon f \text{ a pdf and } \int f h_j \mathrm{d}\mu = \lambda_j, \quad j = 1, 2, \ldots, k\right\}. \tag{12}$$

Then

$$\sup\left\{-\int f\log f\mathrm{d}\mu, \quad f \in \mathcal{F}\right\} = -\int f_0 \log f_0 \mathrm{d}\mu \tag{13}$$

and f_0 is the unique maximizer.

As an application of the above Corollary we get the following examples.

Example 8. The question whether the Poisson distribution has an entropy maximization characterization is of some interest. This example shows that it does. Let $\Omega = \{0, 1, 2, \dots\}$, μ counting measure, $h_1(j) = j$, $h_2(j) = \log j!$ Let $c_1, c_2, \lambda_1, \lambda_2$ be such that

$$\sum j\mathrm{e}^{c_1 j}(j!)^{c_2} = \lambda_1 \sum \mathrm{e}^{c_1 j}(j!)^{c_2},$$

$$\sum (\log j!)\mathrm{e}^{c_1 j}(j!)^{c_2} = \lambda_2 \sum \mathrm{e}^{c_1 j}(j!)^{c_2}.$$

For convergence we need $c_2 < 0$. In particular, if we take $c_2 = -1$, $\mathrm{e}^{c_1} = \lambda_1$ and $\lambda_2 = \sum_j \frac{\mathrm{e}^{-\lambda_1}\lambda_1^j}{j!} \log j!$, then we find that Poisson λ is the unique maximizer of entropy among all nonnegative integer-valued random variables X such that $EX = \lambda$ and $E(\log X!) = \sum_0^\infty \frac{\mathrm{e}^{-\lambda}\lambda^j}{j!}(\log j!)$. If λ_1 and λ_2 are two positive numbers then the optimal distribution is Poisson-like and is of the form

$$f_0(j) = \frac{\mu^j (j!)^{-c}}{\sum_0^\infty \mu^j (j!)^{-c}},$$

where $0 < \mu, c < \infty$ and satisfy

$$\sum j\mu^j (j!)^{-c} = \lambda_1 \sum_0^\infty \mu^j (j!)^{-c},$$

$$\sum (\log j!)\mu^j (j!)^{-c} = \lambda_2 \sum_0^\infty \mu^j (j!)^{-c}.$$

The function

$$\psi(\mu, c) = \sum_0^\infty \mu^j (j!)^{-c}$$

is well-defined in $(0, \infty) \times (0, \infty)$ and is infinitely differentiable as well. The constraints on μ and c may be rewritten as

$$\frac{\partial \psi}{\partial \mu} = \mu\lambda_1 \psi(\mu, c), \qquad \frac{\partial \psi}{\partial c} = -\lambda_2 \psi(\mu, c). \tag{14}$$

Let $\varphi(\mu, c) = \log \psi(\mu, c)$. Then the map $(\mu, c) \to \left(\frac{1}{\mu}\frac{\partial \varphi}{\partial \mu}, \frac{\partial \varphi}{\partial c}\right)$ from $(0, \infty) \times (0, \infty)$ to $(0, \infty) \times (-\infty, 0)$ can be shown to be one-to-one and onto. Thus for any $\lambda_1 > 0$, $\lambda_2 > 0$ there exist unique $\mu > 0$ and $c > 0$ such that

$$\frac{1}{\mu}\frac{\partial \varphi}{\partial \mu} = \frac{1}{\mu}\frac{1}{\psi(\mu, c)}\frac{\partial \psi}{\partial \mu} = \lambda_1,$$

$$\frac{\partial \varphi}{\partial c} = \frac{1}{\psi}\frac{\partial \psi}{\partial c} = -\lambda_2.$$

Example 9. The exponential family of densities in mathematical statistics literature is of the form

$$f(\theta, \omega) \alpha \mathrm{e}^{\sum_1^k c_i(\theta)h_i(\omega) + c_0 h_0(\omega)}. \tag{15}$$

From Corollary 3 it follows that for each θ, $f(\theta,\omega)$ is the unique maximizer of entropy among all densities f such that

$$\int f(\omega)h_i(\omega)\mathrm{d}\mu = \int f(\theta,\omega)h_i(\omega)\mu(\mathrm{d}\omega)$$

for $i = 0, 1, 2, \ldots, k$.

Given $\lambda_0, \lambda_1, \lambda_2, \lambda_k$ to find a value of θ such that $f(\theta,\omega)$ is the maximizer of entropy subject to $\int f h_i \mathrm{d}\mu = \lambda_i$ for $i = 0, 1, 2, \ldots, k$ is equivalent to first finding $c_0, c_1, c_2, \ldots$ such that if $\psi(c_0, c_1, \ldots, c_k) = \int \mathrm{e}^{\sum_0^k c_i h_i}\mathrm{d}\mu$ and $\phi = \log\psi$, then $\frac{\partial\phi}{\partial c_i} = \lambda_i$, $i = 0, 1, \ldots, k$ and then θ such that $a_i(\theta) = c_i$ for $i = 1, 2, \ldots, k$. Under fairly general assumptions the range of $\left(\frac{\partial\phi}{\partial c_i},\ i = 1, 2, \ldots, k\right)$ is a big enough set so that requiring $(\lambda_0, \lambda_1, \ldots, \lambda_k)$ belongs to that set would not be too stringent.

Corollary 3 can be generalized to an infinite family of functions as follows.

COROLLARY 4

Let $(S, \mathcal{S})$ *be a measurable space,*

$$h = S \times \Omega \to R \text{ be } \mathcal{B} \times \mathcal{S} \text{ measurable and}$$

$$\lambda = S \to R \text{ be } \mathcal{S} \text{ measurable}. \tag{16}$$

Let

$$\mathcal{F}_\lambda = \left\{ f\colon f \text{ a pdf such that for } \forall s \text{ in } S \int f(\omega)h(s,\omega)\mathrm{d}\mu = \lambda(s) \right\}.$$

Let ν *be a measure on* $(S, \mathcal{S})$ *and* $c = S \to R$ *be* $\mathcal{S}$ *measurable such that*

$$\int_\Omega \exp\left(\int_S h(s,\omega)c(s)\nu(\mathrm{d}s)\right)\mu(\mathrm{d}\omega) < \infty \tag{17}$$

and

$$\int_\Omega h(s,\omega)\mathrm{e}^{\int_S h(s',\omega)c(s')\nu(\mathrm{d}s')}\mu(\mathrm{d}\omega) = \lambda(s) \text{ for all } s \text{ in } S. \tag{18}$$

Then

$$\sup\left\{-\int_\Omega f\log f\mathrm{d}\mu\colon f \in \mathcal{F}_\lambda\right\} = -\int f_0 \log f_0 \mathrm{d}\mu,$$

where $f_0(\omega)\alpha\exp(\int_S h(s,\omega)c(s)\nu(\mathrm{d}s))$.

Example 10. Let $\Omega = C[0,1]$, $\mathcal{B}$ the Borel σ-algebra generated by the sup norm on Ω, μ be a Gaussian measure with mean function $m(s) \equiv 0$ and covariance $r(s,t)$. Let $\lambda(\cdot)$ be a Borel measurable function on $[0,1] \to R$. Let $\mathcal{F}_\lambda \equiv \{f\colon f$ a pdf on $(\Omega, \mathcal{B}, \mu)$ such that $\int \omega(t)f(\omega)\mu(\mathrm{d}\omega) = \lambda(t)\ \forall 0 \le t \le 1\}$. That is, $\mathcal{F}_\lambda$ is the set of pdf of all those stochastic processes on $[0,1]$ that have continuous trajectories, mean function $\lambda(\cdot)$ and whose probability distribution on Ω is absolutely continuous with respect to μ. Let ν be a Borel measure on $[0,1]$ and $c(\cdot)$ a Borel measurable function. Then

$$f_0(\omega)\alpha\exp\int_0^1 c(s)\omega(s)\nu(\mathrm{d}s)$$

maximizes $-\int_\Omega f\log f\,\mathrm{d}\mu$ over all f in $\mathcal{F}_\lambda$ provided

$$\int_\Omega w(t)\mathrm{e}^{\int_0^1 c(s)\omega(s)\nu(\mathrm{d}s)}\mu(\mathrm{d}\omega)$$
$$= \lambda(t)\int_\Omega \mathrm{e}^{\int_0^1 c(s)\omega(s)\nu(\mathrm{d}s)}\mu(\mathrm{d}\omega) \text{ for all } t \text{ in } [0, 1]. \tag{19}$$

Since μ is a Gaussian measure with mean function 0 and convariance function $r(s, t)$ the joint distribution of $\omega(t)$ and $\int_0^1 c(s)\omega(s)\nu(\mathrm{d}s)$ is bivariate normal with mean 0 and covariance matrix $\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}$ where $\sigma_{11} = r(t, t)$, $\sigma_{12} = \int_0^1 c(s)r(s, t)\nu(\mathrm{d}s)$,

$$\sigma_{22} = \int_0^1\int_0^1 c(s_1)c(s_2)r(s_2, s_2)\nu(\mathrm{d}s_1)\nu(\mathrm{d}s_2).$$

It can be verified by differentiating the joint m.g.f. that if (X, Y) is bivariate normal with mean 0 and covariance matrix $\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}$, then

$$E(X\mathrm{e}^Y) = \mathrm{e}^{\frac{1}{2}\sigma_{22}}\sigma_{12} \quad \text{and} \quad E(\mathrm{e}^Y) = \mathrm{e}^{\frac{1}{2}\sigma_{22}}.$$

Applying this to (19) with $X = w(t)$ and $Y = \int_0^1 c(s)w(s)\nu(\mathrm{d}s)$ we get

$$\int_0^1 c(s)r(s, t)\nu(\mathrm{d}s) = \lambda(t), \quad 0 \le t \le 1.$$

Thus, if $c(\cdot)$ and $\nu(\cdot)$ satisfy the above equation and

$$\int_0^1\int_0^1 |c(s_1)c(s_2)r(s_1, s_2)|\nu(\mathrm{d}s_1)\nu(\mathrm{d}s_1) < \infty,$$

then

$$\sup\left\{-\int f\log f\,\mathrm{d}\mu\colon f\in\mathcal{F}\right\}_\lambda = -\int f_0\log f_0\mathrm{d}\mu$$

and f_0 is the unique maximizer. Notice that

$$f_0(\omega) = \frac{\mathrm{e}^{\int_0^1 c(s)w(s)\nu(\mathrm{d}s)}}{\mathrm{e}^{\frac{\sigma_{22}}{2}}}. \tag{20}$$

The joint m.g.f. of $(\omega(t_1), \omega(t_2), \ldots, \omega(t_k))$ under $P_{f_0}(A) \equiv \int_A f_0\mathrm{d}\mu$ is

$$E_{P_{f_0}}\left(\mathrm{e}^{\sum_1^k \theta_i\omega(t_1)}\right) = \int_\Omega \mathrm{e}^{\sum_1^k \theta_i\omega(t_i)}\frac{\mathrm{e}\int_0^1 c(s)\omega(s)\nu(\mathrm{d}s)}{\mathrm{e}^{\frac{\sigma_{22}}{2}}}\mu(\mathrm{d}\omega). \tag{21}$$

But $\sum_1^k \theta_i\omega(t_i) + \int_0^1 c(s)\omega(s)\nu(\mathrm{d}s)$ is a Gaussian random variable under μ with mean 0 and variance

$$\sigma^2 = \sum_{i,j}\theta_i\theta_j r(t_i, t_j) + \sigma_{22} + 2\sum_1^k\theta_i\int_0^1 c(s)r(s, t_i)\nu(\mathrm{d}s))$$
$$= \sum_{i,j}\theta_i\theta_j r(t_i, t_j) + \sigma_{22} + 2\sum_1^k\theta_i\lambda(t_i).$$

The right-hand side of (20) becomes

$$\exp\left(\frac{1}{2}\left(\sum_{i,j}\theta_i\theta_j r(t_i,t_j)\right)+\sum_1^k \theta_i\lambda(t_i)\right). \tag{22}$$

That is, P_{f_0} is Gaussian with mean $\lambda(\cdot)$ and covariance $r(\cdot,\cdot)$, same as μ. Thus, among all stochastic processes on Ω that are absolutely continuous with respect to μ and whose mean function is specified to be $\lambda(\cdot)$ the one that maximizes the relative entropy is a Gaussian process with mean $\lambda(\cdot)$ and same covariance kernel as that of μ. This suggests that the density $f_0(\cdot)$ in (20) should be independent of $c(\cdot)$ and $\nu(\cdot)$ so long as (18) holds. This is indeed so. Let (c_1, ν_1) and (c_2, ν_2) be two solutions to (18). Let f_1 and f_2 be the corresponding densities. We claim $f_1 = f_2$. a.e. μ. That is,

$$\frac{\mathrm{e}^{\int_0^1 c_1(s)\omega(s)\nu_1(\mathrm{d}s)}}{\int_\Omega \mathrm{e}^{\int_0^1 c_1(s)\omega(s)\nu_1(\mathrm{d}s)}\mu(\mathrm{d}\omega)}=\frac{\mathrm{e}^{\int_0^1 c_2(s)\omega(s)\nu_2(\mathrm{d}s)}}{\int_\Omega \mathrm{e}^{\int_0^1 c_2(s)\omega(s)\nu_2(\mathrm{d}s)}\mu(\mathrm{d}\omega)}.$$

Under μ, $\int_0^1 c(s)\omega(s)\nu(\mathrm{d}s)$ is univariate normal with mean 0 and variance

$$\int_0^1\int_0^1 c(s_1)c(s_2)r(s_1,s_2)\nu(\mathrm{d}s)\nu(\mathrm{d}s)=\int_0^1 c(s)\lambda(s)\nu(\mathrm{d}s)$$

if (c, ν) satisfy (18). Now, if $Y_1 = \int_0^1 c_1(s)\omega(s)\nu_1(\mathrm{d}s)$ and $Y_2 = \int_0^1 c_2(s)\omega(s)\nu_2(\mathrm{d}s)$ then $EY_1 = EY_2 = 0$ and since (c_1, ν_1), (c_2, ν_2) satisfy (18) we get

$$\begin{aligned}\mathrm{Cov}(Y_1,Y_2)&=\int_0^1 c_1(s)\lambda(s)\nu_1(\mathrm{d}s)=\int_0^1 c_2(s)\lambda(s)\nu_2(\mathrm{d}s)\\&=\int_0^1 c_2(s)\lambda(s)\nu_1(\mathrm{d}s)=\int_0^1 c_2(s)\lambda(s)\nu_2(\mathrm{d}s),\\ V(Y_1)&=\int_0^1 c_1(s)\lambda(s)\nu_1(\mathrm{d}s),\\ V(Y_2)&=\int_0^1 c_2(s)\lambda(s)\nu_2(\mathrm{d}s).\end{aligned}$$

Thus $(Y_1 - Y_2)^2 = 0$ implying $Y_1 = Y_2$ a.e. μ and hence $f_1 = f_2$ a.e. μ.

The result that the measure maximizing relative entropy with respect to a given Gaussian measure with a given covariance kernel and subject to a given mean function $\lambda(\cdot)$ is a Gaussian with mean $\lambda(\cdot)$ and covariance $r(\cdot,\cdot)$ is a direct generalization of the corresponding univariate result that says of all pdf f on R subject to $\frac{1}{\sqrt{2\pi}}\int xf(x)\mathrm{e}^{-\frac{x^2}{2}\mathrm{d}x} = \mu$ the one that maximizes $-\frac{1}{\sqrt{2\pi}}\int f(x)\log f(x)\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x$ is $f(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2}}$. Although the generalization that is stated above is to the case of Gaussian measure on $C[0, 1]$ the result and the argument hold much more generally. If $\Omega = C[0, 1]$ and μ is the standard Wiener measure then by Girsanov's theorem [2] the process $\omega(t) + \int_0^t \alpha(s,\omega)\mathrm{d}\omega(s)$ where $\alpha(\cdot)$ is a nonanticipating functional induces a probability measure that is absolutely continuous with respect to μ and has a pdf of the form

$$\exp\left(\int_0^1 \alpha(s,\omega)\mathrm{d}\omega(s) - \frac{1}{2}\int_0^1 \alpha^2(s,\omega)\mathrm{d}s\right),$$

where the first integral is an Ito integral and the second a Lebesgue integral. Our result says that among these the one that maximizes the relative entropy subject to a mean function $\lambda(\cdot)$ restriction is a process where the Ito integral can be expressed as $\int_0^1 c(s)\omega(s)\mathrm{d}s$ i.e. of the type that Weiner defined for nonrandom integrands [3].

References

[1] Durrent R, Probability: Theory and Examples (CA: Wadsworth and Brooks – Cole, Pacific Grove) (1991)
[2] Karatzas I and Shreve S, Brownian Motion and Stochastic Calculus, second edition (GTM, Springer-Verlag) (1991)
[3] McKean H P, Stochastic Integrals (New York: Academic Press) (1969)
[4] Rudin W, Real and Complex Analysis, third edition (New York: McGraw Hill) (1984)
[5] Shannon C, A Mathematical Theory of Communication, *Bell Systems Technical J.* **23** (1948) 379–423, 623–659