2012

Entropy and related principles

Jonathan P. Williams
Eastern Michigan University

Entropy and related principles

Abstract

This thesis is a formal presentation of entropy and related principles as they relate to probability theory. The central focus is on the mathematical definition of entropy, and on methods for obtaining maximum entropy distributions. Transformations which preserve the entropy of a probability density function are also developed. Entropy is a measure of uncertainty, and a maximum entropy distributions has the desirable property of maximizing uncertainty for all unknown information of a given phenomenon.

Degree Type

Open Access Senior Honors Thesis

Department

Economics

First Advisor

Dr. Ovidiu Calin

Second Advisor

Dr. Kemper Moreland

ENTROPY AND RELATED PRINCIPLES

By

Jonathan P. Williams

A Senior Thesis Submitted to the

Eastern Michigan University

Honors College

in Partial Fulfillment of the Requirements for Graduation

with Honors in Economics

Approved at Ypsilanti, Michigan, on this date April 17, 2012

Senior Honors Thesis

Abstract

Author: Jonathan P. Williams

Department: Economics

Area: Economics and Mathematics

Supervising Instructor: Dr. Ovidiu Calin

Honors Adviser: Dr. Kemper Moreland

Title: Entropy and Related Principles

Completion Date: April 30, 2012

This thesis is a formal presentation of entropy and related principles as they relate to probability theory. The central focus is on the mathematical definition of entropy, and on methods for obtaining maximum entropy distributions. Transformations which preserve the entropy of a probability density function are also developed. Entropy is a measure of uncertainty, and a maximum entropy distributions has the desirable property of maximizing uncertainty for all unknown information of a given phenomenon.

Contents

1 Introduction

Entropy, from its probabilistic definition, is a measure of uncertainty for a random variable. Typically, a random variable is described using a distribution with parameters including an expected value and variance. Additional moments are able to convey information such as the kurtosis and skewness of the distribution, but the entropy cannot be found using a moment generating function. In the continuous case, the entropy of a given random variable is the integral of the product of its probability density function and the logarithm of its probability density function, on a specified interval. The interval of integration is of central concern to this analysis. Questions arise regarding which probability distributions have the maximum entropy on varying intervals of integration, and with various moment constraints. The problem becomes one of constrained optimization in which the Lagrange multiplier procedure is an appropriate method to utilize. Using this procedure, various moments can be constrained or left unconstrained to observe how the maximum entropy distribution changes. Additionally, there exist at least one distribution transformation which preserves entropy. Knowing this and other transformations may be very insightful for obtaining unknown maximum entropy distributions.

First and foremost, a discussion of entropy and its definition will follow. Then, the maximum entropy distribution problem will be formally introduced, along with a few well known cases of maximum entropy distributions. Next, cases in which the maximum entropy distributions have not been found are discussed. Finally, a section discussing entropy preserving transformations is presented.

2 Entropy

Definition 2.1 *Let X be a continuous random variable with probability density function, $p(x)$, on the finite interval $[a, b]$. Then the entropy of X, denoted $H(X)$, is defined as*

$$H(X) = \int_a^b p(x) \ln \frac{1}{p(x)} \, dx, \tag{2.1}$$

or

$$H(X) = -\int_a^b p(x) \ln p(x) \, dx. \tag{2.2}$$

Since $p(x) > 0$, $H(X)$ is defined for all $p(x)$. However, $H(X)$ may be negative for certain distributions in which $p(x) > 1$ for some small interval such that $\int_a^b p(x)\, dx = 1$ still holds. The entropy may also be defined for a discrete random variable. For example, in physics the entropy commonly refers to the movement of a particle. Notice that the entropy is defined similar to the Fisher Information aside from differentiation inside the integral with respect to a parameter. However, Fisher Information is most useful for determining a lower bound on the variance of an estimator. While the entropy has no such interpretation it may have a useful application as an estimator. For example, Buchen and Kelly [1] use maximum entropy the determine the distribution of asset prices with prior knowledge of the asset's option prices. The entropy also appears to be similar to the likelihood function, which has a very nice application as an estimator, but is again distinctly different.

The question becomes: Is it desirable to minimize or maximize the entropy of a distribution? Since entropy measures the uncertainty associated with a random variable the maximum entropy distribution will be the least biased distribution. Given all known information about a random variable, such as a certain number of moments and the interval over which the variable is defined, the distribution which makes the least additional assumptions about the variable will be unbiased; this distribution will be maximally uncertain with respect to the unknown information. It follows from this that the maximum entropy distribution will be the least biased distribution approximation of a random variable given the known information. In general, the entropy has an upper bound which can be found using Jensen's inequality.

Definition 2.2 *For a given random variable, Y, with probability density function $p(y)$ defined on the finite interval $[a, b]$, its expected value is defined as*

$$E[Y] = \int_a^b y p(y)\, dy \tag{2.3}$$

which is also the first moment about the mean.

Theorem 2.3 (Jensen's inequality) *Let $\varphi : [a, b] \to \mathbb{R}$ be a convex function and let Y be an integrable random variable. Given $\varphi(Y)$ is integrable,*

$$\varphi(E[Y]) \leq E[\varphi(Y)]. \tag{2.4}$$

If $\varphi(Y)$ is concave, then

$$\varphi(E[Y]) \geq E[\varphi(Y)].$$

From Equation 2.1 it follows that $H(X)$ can be written as the expectation

$$H(X) = \int_a^b p(x) \ln \frac{1}{p(x)}\, dx = E\left[\ln \frac{1}{p(x)}\right] \tag{2.5}$$

for the entire domain, $[a, b]$, on which $p(x)$ is defined. Representing $H(X)$ as an expectation is useful for finding the maximum entropy for a given probability density function, $p(x)$.

Theorem 2.4 *Let X be an integrable random variable with probability density function $p(x)$ defined on the finite interval $[a, b]$. Then $H(X) \leq \ln(b - a)$.*

Proof: Let $p(x)$ be the probability density function for the random variable X defined on the finite interval $[a, b]$, and let $\varphi(u) = \ln u$ where $u = \frac{1}{p(x)}$. By definition $p(x) > 0$, so u must be convex, and φ is concave so $-\varphi$ must be convex. Using Jensen's inequality,

$$-\varphi(E[u]) \leq E[-\varphi(u)]$$

$$-\ln E\left[\frac{1}{p(x)}\right] \leq E\left[-\ln \frac{1}{p(x)}\right]$$

then

$$E\left[\ln \frac{1}{p(x)}\right] \leq \ln E\left[\frac{1}{p(x)}\right]. \tag{2.6}$$

Substitute Equation 2.5 into Equation 2.6 to obtain

$$H(X) \leq \ln E\left[\frac{1}{p(x)}\right]. \tag{2.7}$$

Simplify the expression

$$E\left[\frac{1}{p(x)}\right] = \int_a^b \frac{1}{p(x)} p(x)\, dx = \int_a^b 1\, dx = b - a,$$

and then substituting the expression back into Equation 2.7 yields

$$H(X) \leq \ln(b - a). \square$$

3 Maximum Entropy Distributions

In many applications the maximum entropy distribution, given certain constraints and over various intervals, is desirable. To maximize entropy is to maximize uncertainty, and to achieve the least biased probability distribution approximation. In order to determine various maximum entropy distributions the Lagrange Multiplier procedure can be implemented. A few well known maximum entropy distributions are presented in what follows.

Theorem 3.1 *The uniform distribution is the maximum entropy distribution of all distributions on the finite interval* $[a, b]$.

Proof: The maximum entropy distribution can be obtained by maximizing the following formula.

$$p \mapsto -\int_a^b p(x)\ln p(x)\,dx + \lambda_1\left[\int_a^b p(x)\,dx - 1\right]$$

where λ_1 is the Lagrange multiplier. The Lagrangian is

$$\mathcal{L} = -p\ln p + \lambda_1(p-1).$$

Differentiation with respect to p yields

$$\frac{d\mathcal{L}}{dp} = -\ln p - 1 + \lambda_1 \tag{3.8}$$

and

$$\frac{d^2\mathcal{L}}{dp^2} = -\frac{1}{p} < 0.$$

Since the second derivative is less than zero it follows that the first derivative is a maximum when when equated to zero. Accordingly, Equation 3.8 must be equated to zero and solved for p. Simplification yields

$$p = e^{-1+\lambda_1} \tag{3.9}$$

where λ_1 is obtained from the constraint $\int_a^b p(x)\,dx = 1$. Substitute p in Equation 3.9 into the constraint.

$$\int_a^b e^{-1+\lambda_1}\,dx = 1 \implies e^{-1+\lambda_1}(b-a) = 1,$$

or

$$e^{-1+\lambda_1} = \frac{1}{(b-a)} \implies p(x) = \frac{1}{(b-a)},$$

which is the uniform distribution. □

Theorem 3.2 *The exponential distribution, with given mean μ, is the maximum entropy distribution of all distributions on the interval $[0,\infty)$.*

Proof: To find the maximum entropy distribution, again, the following formula must be maximized.

$$p \mapsto -\int_0^\infty p(x)\ln p(x)\,dx + \lambda_1\left[\int_0^\infty p(x)\,dx - 1\right] + \lambda_2\left[\int_0^\infty xp(x)\,dx - \mu\right]$$

where λ_1 and λ_2 are Lagrange multipliers. The corresponding Lagrangian is

$$\mathcal{L} = -p\ln p + \lambda_1(p-1) + \lambda_2(xp-\mu)$$

Then differentiate with respect to p.

$$\frac{\partial\mathcal{L}}{\partial p} = -\ln p - 1 + \lambda_1 + \lambda_2 x \tag{3.10}$$

and

$$\frac{\partial^2 \mathcal{L}}{\partial p^2} = -\frac{1}{p} < 0.$$

From this, the second order condition is satisfied for the first order condition to be a maximum when set equal to zero. Equation 3.10 must then be set equal to zero and solved for p.

$$\ln p = -1 + \lambda_1 + \lambda_2 x \implies p = e^{-1+\lambda_1+\lambda_2 x} = ce^{\lambda_2 x}$$

where $c = e^{-1+\lambda_1}$. The Lagrange multipliers are obtained using the constraints

$$\int_0^\infty p(x)\,dx = 1, \int_0^\infty xp(x)\,dx = \mu.$$

Substitute $p(x)$, and solve the system of equations for the Lagrange multipliers

$$c\int_0^\infty e^{\lambda_2 x}\,dx = 1, \ c\int_0^\infty xe^{\lambda_2 x}\,dx = \mu$$

or

$$-c\int_{-\infty}^0 e^{\lambda_2 x}\,dx = 1, \ -c\int_{-\infty}^0 xe^{\lambda_2 x}\,dx = \mu.$$

Then

$$\lambda_2 = -c, \ \frac{c}{\lambda_2^2} = \mu$$

With substitution $c = \frac{1}{\mu}$ and $\lambda_2 = -\frac{1}{\mu}$. Therefore,

$$p(x) = \frac{1}{\mu}e^{-\frac{1}{\mu}x},$$

which is the exponential distribution. □

Theorem 3.3 *The normal distribution, with given mean μ and variance σ^2, is the maximum entropy distribution of all distributions on the interval $(-\infty, +\infty)$.*

Proof: Maximize the following formula.

$$p \mapsto -\int_{-\infty}^{+\infty} p(x)\ln p(x)\,dx - \lambda_1\left[\int_{-\infty}^{+\infty} p(x)\,dx - 1\right] - \lambda_2\left[\int_{-\infty}^{+\infty} xp(x)\,dx - \mu\right]$$
$$-\lambda_3\left[\int_{-\infty}^{+\infty}(x-\mu)^2 p(x)\,dx - \sigma^2\right],$$

where λ_1, λ_2, and λ_3 are Lagrange multipliers. The Lagrangian can be written as

$$\mathcal{L} = -p\ln p - \lambda_1(p-1) - \lambda_2(x-\mu)p - \lambda_2\mu - \lambda_2\mu p - \lambda_3[(x-\mu)^2 p - \sigma^2].$$

Then differentiate with respect to p.

$$\frac{\partial \mathcal{L}}{\partial p} = -\ln p - 1 - \lambda_1 - \lambda_2(x-\mu) - \lambda_2\mu - \lambda_3(x-\mu)^2. \tag{3.11}$$

The second order condition, for the first order condition to be maximum, is satisfied as indicated next.

$$\frac{\partial^2 \mathcal{L}}{\partial p^2} = -\frac{1}{p} < 0$$

Next, equate Equation 3.11 to zero to solve for p.

$$\ln p = -1 - \lambda_1 - \lambda_2\mu - \lambda_2(x-\mu) - \lambda_3(x-\mu)^2,$$

or

$$p = e^{-1-\lambda_1-\lambda_2\mu-\lambda_2(x-\mu)-\lambda_3(x-\mu)^2} \implies p = ce^{-\lambda_2(x-\mu)-\lambda_3(x-\mu)^2} \tag{3.12}$$

where $c = e^{-1-\lambda_1-\lambda_2\mu}$. The following constraints are used to obtain the Lagrange multipliers.

$$\int_{-\infty}^{\infty} p(x)\,dx = 1,\ \int_{-\infty}^{\infty} xp(x)\,dx = \mu,\ \int_{-\infty}^{+\infty}(x-\mu)^2 p(x)\,dx = \sigma^2. \tag{3.13}$$

Due to the complexity of the integrals, some formulas will be very useful. They are

$$\int_{-\infty}^{+\infty} e^{-bx^2}\,dx = \frac{\sqrt{\pi}}{\sqrt{b}}, \tag{3.14}$$

$$\int_{-\infty}^{+\infty} x e^{-ax-bx^2}\,dx = \frac{a e^{\frac{a^2}{4b}}\sqrt{\pi}}{2b\sqrt{b}}, \tag{3.15}$$

$$\int_{-\infty}^{+\infty} x^2 e^{-bx^2}\,dx = \frac{\sqrt{\pi}}{2b\sqrt{b}}. \tag{3.16}$$

Using the second constraint in Equation 3.13,

$$\int_{-\infty}^{+\infty} x p(x)\,dx = \mu \implies \int_{-\infty}^{+\infty} (x-\mu)p(x)\,dx = 0.$$

Then

$$c\int_{-\infty}^{+\infty} (x-\mu)e^{-\lambda_2(x-\mu)-\lambda_3(x-\mu)^2}\,dx = 0.$$

Using Equation 3.15

$$\frac{c\lambda_2 e^{\frac{\lambda_2^2}{4b}}\sqrt{\pi}}{2b\sqrt{b}} = 0,$$

and from this it follows that $\lambda_2 = 0$. Next, using Equation 3.14 yields

$$c\int_{-\infty}^{+\infty} e^{-0(x-\mu)-\lambda_3(x-\mu)^2}\,dx = 1 \implies c\int_{-\infty}^{+\infty} e^{-\lambda_3(x-\mu)^2}\,dx = 1,$$

$$\frac{c\sqrt{\pi}}{\sqrt{\lambda_3}} = 1 \implies c = \sqrt{\frac{\lambda_3}{\pi}}. \tag{3.17}$$

Finally, λ_3 can be obtained using the third constraint in Equation 3.12.

$$c\int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\lambda_3(x-\mu)^2}\,dx = \sigma^2,$$

and with Equations 3.16 and 3.17,

$$c\frac{\sqrt{\pi}}{2\lambda_3\sqrt{\lambda_3}} = \sigma^2 \implies \frac{1}{2\sigma^2} = \lambda_3.$$

Then

$$c = \frac{1}{\sigma\sqrt{2\pi}},$$

and therefore, from Equation 3.12

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

which is the normal distribution. $\square$

This maximum entropy distribution result has the following, nice interpretation. The normal distribution dominates the physical and social sciences with applications to numerous random variable phenomena. For all of these phenomena to be normally distributed it would only make sense that this distribution has the maximum entropy with a known mean and variance on the interval of all real numbers.

However, for many applications in finance, the maximum entropy on a closed interval would be much more applicable, given a mean and possibly additional moments. For instance, the distribution of an asset price, such as a stock or a certain derivative, has a finite beginning and ending time, and a sample mean and variance can be easily obtained from a sample of prices. Unfortunately, such a distribution is not easily obtained, and might not even be possible to obtain. Still, something can be said about the existence and uniqueness of such a distribution.

Theorem 3.4 *There exists a unique maximum entropy distribution, with given mean* μ*, on the finite interval* $[a, b]$

Proof: Set up the Lagrangian in the familiar way

$$p \mapsto -\int_a^b p(x) \ln p(x)\, dx + \lambda_1 \left[\int_a^b p(x)\, dx - 1 \right] + \lambda_2 \left[\int_a^b x p(x)\, dx - \mu \right],$$

where λ_1 and λ_2 are the usual Lagrange multipliers. The Lagrangian is

$$\mathcal{L} = -p \ln p + \lambda_1(p-1) + \lambda_2(xp - \mu).$$

Then differentiate with respect to p and set equal to zero. The second order conditions for a maximum are met as follows

$$\frac{\partial \mathcal{L}}{\partial p} = -\ln p - 1 + \lambda_1 + \lambda_2 x,$$

$$\frac{\partial^2 \mathcal{L}}{\partial p^2} = -\frac{1}{p} < 0.$$

Accordingly,

$$\frac{\partial \mathcal{L}}{\partial p} = -\ln p - 1 + \lambda_1 + \lambda_2 x = 0,$$

and

$$\ln p = -1 + \lambda_1 + \lambda_2 x \implies p = e^{-1+\lambda_1+\lambda_2 x} = ce^{\lambda_2 x}.$$

where $c = e^{-1+\lambda_1}$. The Lagrange multipliers are obtained using the constraints

$$\int_a^b p(x)\,dx = 1, \int_a^b xp(x)\,dx = \mu.$$

So

$$c\int_a^b e^{\lambda_2 x}\,dx = 1, \; c\int_a^b xe^{\lambda_2 x}\,dx = \mu. \tag{3.18}$$

Then

$$c\left(\frac{e^{b\lambda_2} - e^{a\lambda_2}}{\lambda_2}\right) = 1, \; c\left[\left(\frac{be^{b\lambda_2} - ae^{a\lambda_2}}{\lambda_2}\right) - \left(\frac{e^{b\lambda_2} - e^{a\lambda_2}}{\lambda_2^2}\right)\right] = \mu.$$

Dividing these two equations yields

$$\frac{be^{b\lambda_2} - ae^{a\lambda_2}}{e^{b\lambda_2} - e^{a\lambda_2}} - \frac{1}{\lambda_2} = \mu \implies \frac{be^{b\lambda_2} - ae^{a\lambda_2}}{e^{b\lambda_2} - e^{a\lambda_2}} = \mu + \frac{1}{\lambda_2}.$$

For simplification, divide the numerator and denominator on the left side of the equation by $e^{b\lambda_2}$.

$$\frac{b - ae^{(a-b)\lambda_2}}{1 - e^{(a-b)\lambda_2}} = \mu + \frac{1}{\lambda_2}.$$

Next, add and subtract an a from the numerator of the left side of the equation to obtain

$$\frac{b - a + a(1 - e^{(a-b)\lambda_2})}{1 - e^{(a-b)\lambda_2}} = \mu + \frac{1}{\lambda_2} \implies \frac{b-a}{1 - e^{(a-b)\lambda_2}} + a = \mu + \frac{1}{\lambda_2}. \tag{3.19}$$

The solution for λ_2 in Equation 3.19 is nontrivial, so it is unclear what the maximum entropy distribution is. However, the existence and uniqueness of the solution can still be proven. Consider the system of equations

$$F_1(\lambda_1, \lambda_2) = \int_a^b e^{\lambda_2 x} e^{\lambda_1 - 1}\, dx,$$

and

$$F_2(\lambda_1, \lambda_2) = \int_a^b x e^{\lambda_2 x} e^{\lambda_1 - 1}\, dx.$$

These are the constraints from Equation 3.18, with $c = e^{\lambda_1 - 1}$, such that

$$F_1(\lambda_1, \lambda_2) = 1,$$

and

$$F_2(\lambda_1, \lambda_2) = \mu.$$

By the inverse function theorem, if

$$\Delta = \begin{vmatrix} \frac{\partial F_1}{\partial \lambda_1} & \frac{\partial F_1}{\partial \lambda_2} \\ \frac{\partial F_2}{\partial \lambda_1} & \frac{\partial F_2}{\partial \lambda_2} \end{vmatrix} \neq 0,$$

then $\exists!$ G_1 and G_2 such that $\lambda_1 = G_1(\mu)$ and $\lambda_2 = G_2(\mu)$. The components of Δ are obtained next.

$$\frac{\partial F_1}{\partial \lambda_1} = \int_a^b e^{\lambda_2 x} e^{\lambda_1 - 1}\, dx \implies \int_a^b p(x)\, dx = 1,$$
$$\frac{\partial F_1}{\partial \lambda_2} = \int_a^b x e^{\lambda_2 x} e^{\lambda_1 - 1}\, dx \implies \int_a^b x p(x)\, dx = \mu,$$
$$\frac{\partial F_2}{\partial \lambda_1} = \int_a^b x e^{\lambda_2 x} e^{\lambda_1 - 1}\, dx \implies \int_a^b x p(x)\, dx = \mu,$$
$$\frac{\partial F_2}{\partial \lambda_2} = \int_a^b x^2 e^{\lambda_2 x} e^{\lambda_1 - 1}\, dx \implies \int_a^b x^2 p(x)\, dx = m_2(x).$$

Hence

$$\Delta = \begin{vmatrix} 1 & \mu \\ \mu & m_2(x) \end{vmatrix} = m_2(x) - \mu^2 = Var(x) \geq 0.$$

This leaves two possible cases: Case 1. $Var(x) > 0$ or Case 2. $Var(x) = 0$.

Case 1. $Var(x) > 0$. If the $Var(x) > 0$, then $\Delta \neq 0$, and $\exists!$ λ_2.

Case 2. $Var(x) = 0$. $Var(x) = 0$ for $X_0 = x_0$ (some constant) $\in [a, b]$. Then $p(x) = \delta(x - x_0)$. By the definition of *entropy*,

$$H(X) = -\int_a^b \delta(x - x_0) \ln \delta(x - x_0)\, dx.$$

Let ϵ be an interval around x_0, and let $\delta(x - x_0) = \varphi_\epsilon(x)$. So

$$H_\epsilon(X) = -\int_a^b \varphi_\epsilon(x) \ln \varphi_\epsilon(x)\, dx.$$

As $\epsilon \longrightarrow x_0$, $\varphi_\epsilon(x) \longrightarrow +\infty$ since $\int_\epsilon p(x)\, dx = 1$. Let $\varphi_\epsilon(x) = \frac{1}{\epsilon}$. Define $a = x_0 - \epsilon/2$ and $b = x_0 + \epsilon/2$ so that entire interval is $x_0 + \epsilon/2 - (x_0 - \epsilon/2) = \epsilon$, as defined previously. Then

$$H_\epsilon(X) = -\int_{x_0 - \epsilon/2}^{x_0 + \epsilon/2} \frac{1}{\epsilon} \ln \frac{1}{\epsilon}\, dx.$$

After integrating,

$$H_\epsilon(X) = -\epsilon\frac{1}{\epsilon}\ln\frac{1}{\epsilon} = -\ln\frac{1}{\epsilon}.$$

So

$$H_\epsilon(X) = \ln\epsilon \longrightarrow -\infty \; as \; \epsilon \searrow 0,$$

or

$$H(\delta(x-x_0)) = \lim_{\epsilon\to 0} H_\epsilon = -\infty,$$

is the minimum entropy. This result is trivial because if $Var(x) = 0$ then $p(x)$ is no longer a probability density function, and there will be no uncertainty. □

It is unfortunate that the maximum entropy distribution on a finite interval, with a known mean, is not apparent, but it is very interesting to be able to show that it exists and is unique. It has been shown previously that the uniform distribution is the maximum entropy distribution on a finite interval, with no fixed moments. And it was just shown that when the mean is fixed on the same interval there also exists a maximum entropy distribution, but it is unknown. So what happens when the first and second moments are fixed on a finite interval? Does it become apparent what the maximum entropy distribution is? Or does there even exist a unique maximum entropy distribution?

As it turns out, when the first two moments are fixed on a finite interval the maximum entropy distribution is not apparent, and its existence and uniqueness are unclear, as well. More formally, the maximum entropy distribution, with given mean μ and second moment m_2, on the finite interval $[a, b]$ is unknown, and may not exits and/or is not unique.

Maximize the formula

$$p \mapsto -\int_a^b p(x)\ln p(x)\,dx + \lambda_1\left[\int_a^b p(x)\,dx - 1\right] + \lambda_2\left[\int_a^b xp(x)\,dx - \mu\right] + \lambda_3\left[\int_a^b x^2p(x)\,dx - m_2\right]$$

where λ_1, λ_2, and λ_3 are the Lagrange multipliers. The Lagrangian is

$$\mathcal{L} = -p\ln p + \lambda_1(p-1) + \lambda_2(xp-\mu) + \lambda_3(x^2p(x) - m_2).$$

Then differentiate with respect to p and set equal to zero. The second order conditions for a maximum are met as follows

$$\frac{\partial \mathcal{L}}{\partial p} = -\ln p - 1 + \lambda_1 + \lambda_2 x + \lambda_3 x^2$$

$$\frac{\partial^2 \mathcal{L}}{\partial p^2} = -\frac{1}{p} < 0.$$

Accordingly,

$$\frac{\partial \mathcal{L}}{\partial p} = -\ln p - 1 + \lambda_1 + \lambda_2 x + \lambda_3 x^2 = 0,$$

and

$$\ln p = -1 + \lambda_1 + \lambda_2 x + \lambda_3 x^2 \implies p = e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}.$$

The Lagrange multipliers are obtained using the constraints

$$\int_a^b p(x)\,dx = 1, \int_a^b xp(x)\,dx = \mu, \int_a^b x^2p(x)\,dx = m_2.$$

So

$$\int_a^b e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\,dx = 1, \int_a^b xe^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\,dx = \mu, \int_a^b x^2e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\,dx = m_2.$$

However, solving this system of equations for λ_2 and λ_3 is a nontrivial task, as it was for the previous theorem with only the mean constrained. Prior to determining parameters of the distribution, it would be helpful to show whether the solution exists and/or is unique. Consider the following system of equations.

$$F_1(\lambda_1, \lambda_2, \lambda_3,) = \int_a^b e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\, dx,$$

$$F_2(\lambda_1, \lambda_2, \lambda_3,) = \int_a^b x e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\, dx,$$

and

$$F_3(\lambda_1, \lambda_2, \lambda_3,) = \int_a^b x^2 e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\, dx.$$

From the constraints imposed, the system of equations is such that

$$F_1(\lambda_1, \lambda_2, \lambda_3,) = 1,$$

$$F_2(\lambda_1, \lambda_2, \lambda_3,) = \mu,$$

and

$$F_3(\lambda_1, \lambda_2, \lambda_3,) = m_2.$$

By the inverse function theorem, if

$$\Delta = \begin{vmatrix} \frac{\partial F_1}{\partial \lambda_1} & \frac{\partial F_1}{\partial \lambda_2} & \frac{\partial F_1}{\partial \lambda_3} \\ \frac{\partial F_2}{\partial \lambda_1} & \frac{\partial F_2}{\partial \lambda_2} & \frac{\partial F_2}{\partial \lambda_3} \\ \frac{\partial F_3}{\partial \lambda_1} & \frac{\partial F_3}{\partial \lambda_2} & \frac{\partial F_3}{\partial \lambda_3} \end{vmatrix} \neq 0$$

then $\exists!$ G_1, G_2, and G_3 such that $\lambda_1 = G_1(\mu, m_2)$, $\lambda_2 = G_2(\mu, m_2)$, and $\lambda_3 = G_3(\mu, m_2)$. The partial differentials are solved next, and then substituted into Δ.

$$\frac{\partial F_1}{\partial \lambda_1} = \int_a^b e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\, dx \implies \int_a^b p(x)\, dx = 1,$$

$$\frac{\partial F_1}{\partial \lambda_2} = \int_a^b x e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\, dx \implies \int_a^b x p(x)\, dx = \mu,$$

$$\frac{\partial F_1}{\partial \lambda_3} = \int_a^b x^2 e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\,dx \implies \int_a^b x^2 p(x)\,dx = m_2,$$

$$\frac{\partial F_2}{\partial \lambda_1} = \int_a^b x e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\,dx \implies \int_a^b x p(x)\,dx = \mu,$$

$$\frac{\partial F_2}{\partial \lambda_2} = \int_a^b x^2 e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\,dx \implies \int_a^b x^2 p(x)\,dx = m_2,$$

$$\frac{\partial F_2}{\partial \lambda_3} = \int_a^b x^3 e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\,dx \implies \int_a^b x^3 p(x)\,dx = m_3,$$

$$\frac{\partial F_3}{\partial \lambda_1} = \int_a^b x^2 e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\,dx \implies \int_a^b x^2 p(x)\,dx = m_2,$$

$$\frac{\partial F_3}{\partial \lambda_2} = \int_a^b x^3 e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\,dx \implies \int_a^b x^3 p(x)\,dx = m_3,$$

and

$$\frac{\partial F_3}{\partial \lambda_3} = \int_a^b x^4 e^{-1+\lambda_1+\lambda_2 x+\lambda_3 x^2}\,dx \implies \int_a^b x^4 p(x)\,dx = m_4,$$

where $\int_a^b x^k p(x)\,dx = m_k$, $k \geq 2$, with the special cases $m_0 = 1$ and $m_1 = \mu$. Then

$$\Delta = \begin{vmatrix} 1 & \mu & m_2 \\ \mu & m_2 & m_3 \\ m_2 & m_3 & m_4 \end{vmatrix} = \begin{vmatrix} m_2 & m_3 \\ m_3 & m_4 \end{vmatrix} - \mu \begin{vmatrix} \mu & m_3 \\ m_2 & m_4 \end{vmatrix} + m_2 \begin{vmatrix} \mu & m_2 \\ m_2 & m_3 \end{vmatrix}.$$

Further simplification yields

$$\Delta = (m_2 - \mu^2) m_4 + 2\mu m_2 m_3 - m_3^2 - m_2^3.$$

From the inverse function theorem,

$$\Delta \neq 0 \implies \sigma^2 m_4 + 2\mu m_2 m_3 \neq m_3^2 + m_2^3 \tag{3.20}$$

must be true. However, it is not clear whether Equation 3.20 holds. Therefore, it is unclear whether there exists a unique maximum entropy distribution, or a maximum entropy distribution at all.

In summary, it seems to be the case that on a finite interval the maximum entropy distribution becomes increasingly ambiguous as more moments are fixed. With no moments fixed the uniform distribution is the maximum entropy distribution. With one moment fixed the distribution is unclear, but there does exists a unique maximum entropy distribution. Finally, with two moments fixed it is unclear whether a maximum entropy solution even exists. The problem of determining the maximum entropy distribution becomes increasingly more tedious and difficult with additional moments fixed. For this reason, and due to little success in fixing two moments, fixing three or more moments has not be analyzed.

The next section focuses on preserving the entropy of a probability density function.

4 Transformations that Preserve Entropy

Another interesting question related to the entropy of a given probability density function, $p(x)$, is how it changes as $p(x)$ is transformed. It may be a very useful property if it is found that entropy is preserved over various transformations of the underlying probability density function. For instance, assume $\hat{p}(x)$ is $p(x)$ after a certain transformation which is known to preserve entropy. Also assume that $p(x)$ has an entropy given by $H(X)$. Then it follows that $\hat{p}(x)$ has the same entropy, $H(X)$, since the transformation is known to preserve the entropy. Further, in the case that $p(x)$ happens to be a maximum entropy distribution, $\hat{p}(x)$ will also be a maximum entropy distribution. The major implication is that these transformations may be a way to find additional maximum entropy distributions using already known maximum entropy distributions. The crux of the problem, then, is to determine if such transformations exist, and if so, what they are. One such transformation is given next.

Theorem 4.1 (linear transformation) *The probability density functions $p(x)$, $\hat{p}(x)$: $[a,b] \to \mathbb{R}$, with $\hat{p}(x) = p(a+b-x)$ have the same entropy, $H(p) = H(\hat{p})$.*

Proof: Let $\hat{p}(x) = p(a+b-x)$. By the definition of cntropy,

$$H(\hat{p}) = -\int_a^b \hat{p}(x) \ln \hat{p}(x)\, dx = -\int_a^b p(a+b-x) \ln p(a+b-x)\, dx.$$

Let $u = a+b-x$. Then $du = -dx$. With substitution,

$$H(\hat{p}) = \int_b^a p(u) \ln p(u)\, du = -\int_a^b p(u) \ln p(u)\, du.$$

Therefore,

$$H(\hat{p}) = H(p). \square$$

Another interesting result is obtained from the previous transformation.

Theorem 4.2 *If the means $\hat{\mu}+\mu = a+b$ for two probability density functions defined on the interval $[a,b]$, then the distributions have the same entropy.*

Proof: Define

$$\hat{\mu} = \int_a^b x\hat{p}(x)\, dx = \int_a^b x p(a+b-x)\, dx$$

where $\hat{p}(x) = p(a+b-x)$, which is a transformation that is known to preserve entropy. Let $u = a + b - x$. Then $du = -dx$.

$$\hat{\mu} = \int_b^a (a + b - u)p(u)\,(-du) = \int_a^b (a + b - u)p(u)\,du,$$

and

$$\hat{\mu} = a + b - \int_a^b up(u)\,du = a + b - \mu.$$

Therefore,

$$\hat{\mu} + \mu = a + b. \square$$

Unfortunately, there has been little success in discovering additional transformations which preserve the entropy of a distribution. It could be the case that only a linear transformation will preserve entropy, but that cannot be said for certain at this time. Possibly, there is a family of transformations with the desired property that has not been discovered yet.

5 Conclusion

Entropy is a very interesting concept which is very applicable for approximating certain phenomena with probability distributions. The maximum entropy distribution will always be the least biased distribution because it makes the fewest assumptions of all approximation methods. This statement is true because the maximum entropy distribution maximizes uncertainty for all unknown information; any approximation method which does not have this property is imposing additional, uncertain assumptions regarding the underlying phenomena. A few cases of maximum entropy distributions have been presented. To summarize:

The uniform distribution is the maximum entropy distribution of all distributions on the finite interval $[a, b]$.

The exponential distribution, with mean μ, is the maximum entropy distribution of all distributions on the interval $[0, \infty)$.

The normal distribution, with mean μ and variance σ^2, is the maximum entropy distribution of all distributions on the interval $(-\infty, +\infty)$.

For many applications in finance, the maximum entropy on a closed interval would be much more applicable, given a mean and possibly additional moments. For instance, the distribution of an asset price, such as a stock or a certain derivative, has a finite beginning and ending time, and a sample mean and variance can be easily obtained from a sample of prices. It seems to be the case, on a finite interval, that the maximum entropy distribution becomes increasingly ambiguous as more moments are fixed. With no moments fixed the uniform distribution is the maximum entropy distribution. With one moment fixed the distribution is unclear, but there does exists a unique maximum entropy distribution. Finally, with two moments fixed it is unclear whether a maximum entropy distribution exists.

A final, question related to the entropy of a given probability density function, $p(x)$, is how it changes as $p(x)$ is transformed. A major implication is that such a transformation may be a way to find additional maximum entropy distributions using already known maximum entropy distributions. This method could possibly be a sidestep around determining these distributions using the Lagrange multiplier approach which appears to have many limitations for this problem, as discussed previously for a finite interval.

References

[1] Buchen, Peter W. & Kelly, Michael. 1996. The Maximum Entropy Distribution of an Asset Inferred from Option Prices. *The Journal of Financial and Quantitative Analysis.* 31, No. 1:(March) 143-159.

[2] DeGroot, Morris H. & Schervish, Mark J. 2002. *Probability and Statistics.* Edition 3. Addison-Wesley.

[3] Kleeman, Richard. Information Thoery and Predictability: Lecture 6: Maximum Entropy Techniques. http://www.math.nyu.edu/faculty/kleeman/infolect6.pdf (accessed February, 2012).

[4] Mead, Lawrence R. & Papanicolaou, N. 1984. Maximum entropy in the problem of moments. *J. Math. Phys.* 25, No. 8:(August) 2404-2417.