Maximum Entropy distributions on the circle with given cross-moment.

J.C.S. de Miranda [1]

Introduction

The maximum entropy principle has been used for modeling situations where there is minimum information or where disorder is expected in some sense. It has been applied to the study of physical phenomena. Maximum entropy distributions were used in Ecology to model the distribution of some species in South America.

Consider a continuous real random variable with probability density function f. Maximum entropy distributions are extremal probability distributions that maximize the following functional

$$\mathfrak{S}(f) = -\int_{\mathcal{I}} f(x) \ln f(x) \mathrm{d}x, \tag{1}$$

where $\mathcal{I} \subset \mathbb{R}$ is the domain of f. In case $\mathcal{I}$ is a bounded interval, the maximum entropy distribution is the uniform distribution. If $\mathcal{I} = \mathbb{R}$ and the mean and variance of f are μ and σ^2, then the maximum entropy distribution is the Normal distribution $\mathrm{N}(\mu, \sigma^2)$. In case $\mathcal{I} = \mathbb{R}^+$ and the mean of f is $\frac{1}{\lambda}$, where $\lambda > 0$, then the maximum entropy distribution is the exponential distribution with parameter λ, i.e., $f(x) = \lambda \mathrm{e}^{-\lambda x}$, for $x \geq 0$ and $f(x) = 0$ for $x < 0$.

Maximum entropy distributions are also of interest in higher dimensional spaces. Given a probability distribution, f, with support on $A \subset \mathbb{R}^n$ the entropy of this distribution is given by the following formula:

$$\mathfrak{S}(f) = -\int_A f \ln f \mathrm{d}\ell$$

If no constraints are imposed on the distribution and the measure of A is finite, then, among all distributions, the uniform distribution will be the one that maximizes the entropy. However, if A is a set of infinite measure, and no constraints are imposed on the distribution, the maximum entropy distribution does not exists.

Our interest will be in distributions defined on subsets of $\mathbb{R}^n$ whose measure, $\ell(A)$, is finite and satisfy some moment conditions. We observe that the following problem:

maximize

$$\mathfrak{S}(f) = -\int_A f \ln f \mathrm{d}\ell$$

with mean value constraints

$$\forall i, \ 1 \leq i \leq n, \ \int_A x_i f \mathrm{d}\ell = \mu_i$$

can be hard to solve. This is mostly due to the nature of the set A. Even for "reasonably simple" regions finding analytical solutions may be a challenge. In a previous work of

[1]Institute of Mathematics and Statistics - University of São Paulo. e-mail: *simon@ime.usp.br*

ours, we considered the maximization of the entropy of a probability distribution defined on the elliptical region $\mathcal{E} = \{(x,y) \in \mathbb{R}^2 : \frac{x^2}{a^2} + \frac{y^2}{b^2} < 1\}$, under the constraints

$$\int_{\mathcal{E}} xf\mathrm{d}\ell = \mu_x \quad \text{and} \quad \int_{\mathcal{E}} yf\mathrm{d}\ell = \mu_y.$$

This leads us to find the extremal points of the following functional:

$$J(f) = \int_{\mathcal{E}} -f\ln f + \lambda f + \lambda_x xf + \lambda_y yf\mathrm{d}\ell,$$

which satisfy the following equation:

$$lnf = -1 + \lambda + \lambda_x x + \lambda_y y.$$

and our main result was

Theorem 1: *The maximum entropy distribution on the elliptical region is given by the following formula:*

$$f_{XY}(x,y) = \frac{1}{ab}K(\mu)\exp\left(\lambda(\mu)\left(\frac{\mu_x}{\mu}\frac{x}{a^2} + \frac{\mu_y}{\mu}\frac{y}{b^2}\right)\right)\chi_{\mathcal{E}}(x,y)$$

where $\mu = \sqrt{\frac{\mu_x^2}{a^2} + \frac{\mu_y^2}{b^2}}$ *and* $K(\mu)$ *and* $\lambda(\mu)$ *are the solutions of the following equations:*

$$\frac{I_2(\lambda)}{I_1(\lambda)} = \mu$$

and

$$K = \frac{\lambda}{2\pi I_1(\lambda)},$$

where

$$I_n(x) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(k+n+1)}\left(\frac{x}{2}\right)^{2k+n}$$

is the modified Bessel function of the first kind of order n.

We remark that we used a projective mapping of the disk onto the elliptical region and a symmetry property of the disk to find the solution to the variational problem.

In this work we determine the maximum entropy distribution on the circle, $D = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$, i.e. we determine the probability density function of the bivariate random vector (X,Y) whose image is contained in D, under the following moment and cross-moment restrictions: $\mathbb{E}X = 0$, $\mathbb{E}Y = 0$ and $\mathbb{E}XY = m$, for an arbitrary real constant m. Using techniques from the calculus of variations, we show that $f_{X,Y}(x,y) = c\exp(\lambda xy)\chi_D(x,y)$ where c and λ are such that $\int_D f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = 1$ and $\int_D xyf_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = m$. This distribution can be used in the simulation of complex valued time series in situations where we know that noise has zero mean and is bounded, and we have no additional information on it but the second order, $(1,1)$, cross-moment of its real and imaginary parts.

The entropy maximization viewed as an isoperimetric problem

In order to find the maximum entropy distribution on the circle that satisfies the zero mean constraints and the cross-moment constraint, we have to solve the following variational problem:

Maximize

$$-\int_D f \ln f \mathrm{d}\ell$$

subjected to

$$f \geq 0 \text{ and } \int_D f \mathrm{d}\ell = 1,$$

$$\int_D x f(x,y) \mathrm{d}x \mathrm{d}y = \int_D y f(x,y) \mathrm{d}x \mathrm{d}y = 0,$$

and

$$\int_D xy f(x,y) \mathrm{d}x \mathrm{d}y = m.$$

This leads us to find the extremal of the following functional:

$$J(f) = \int_D -f \ln f + \lambda_0 f + \lambda_x x f + \lambda_y y f + \lambda_{xy} xy \mathrm{d}\ell,$$

which satisfy the following equation:

$$\ln f = -1 + \lambda_0 + \lambda_x x + \lambda_y y + \lambda_{xy} xy.$$

Using symmetry and the zero mean constraint we are able to prove that $\lambda_x = 0 = \lambda_y$ and f is written as

$$f(x,y) = c\mathrm{e}^{\lambda xy}$$

for some real constants, $c > 0$ and λ.

Now, using $\int_D f \mathrm{d}\ell = 1$ and $\int_D xy f(x,y) \mathrm{d}x \mathrm{d}y = m$ we arrive at the following system of two equations on the unknowns c and λ:

$$c\Psi(\lambda) = 1$$

and

$$c\Psi'(\lambda) = m.$$

Here

$$\Psi(\lambda) = \pi \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k+1)2^{4k}(k!)^2}.$$

Finally, λ and c are determined using $\frac{\Psi'(\lambda)}{\Psi(\lambda)} = m$, and $c = \frac{1}{\Psi(\lambda)}$.

Main result

The main result of this work is the following:

Theorem 2: *The maximum entropy distribution on the circular region, whose mean is zero and whose second order* $(1,1)$ *moment is* m, *is given by the following formula:*

$$f_{XY}(x,y) = c\mathrm{e}^{\lambda xy}\chi_D(x,y)$$

where c *and* λ *are determined using the following equations:*

$$\frac{\Psi'(\lambda)}{\Psi(\lambda)} = m$$

and

$$c = \frac{1}{\Psi(\lambda)},$$

where

$$\Psi(\lambda) = \pi \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda}{4}\right)^{2k}}{(2k+1)(k!)^2}.$$

Final Remarks

We observe that the function Ψ that appears in Theorem 2 is related to the modified Bessel function of first kind an order 0. As a matter of fact, for $\lambda \neq 0$, the following formula relates Ψ and I_0 :

$$\Psi(\lambda) = \frac{2\pi}{\lambda}\int_0^{\frac{\lambda}{2}} I_0(x)\mathrm{d}x.$$

Note that in case $m = 0$ we have $\Psi'(\lambda) = 0$ which implies $\lambda = 0$ and $c = \frac{1}{\Psi(0)} = \frac{1}{\pi}$, and the uniform distribution is the solution to the entropy maximization problem. As we previously mentioned, finding these distributions for arbitrary m is of importance in generating complex valued bounded noise for use in complex valued time series simulations.

We observe that the problem of finding the maximum entropy distribution on the circle with given correlation $corr(X,Y) = \rho$, has a higher level of difficulty.

References

BOROVKOV, A. *Statistique Mathématique.* Éditions Mir, Moscou, 1987.

BOWMAN, F. *Introduction to Bessel Functions.* Dover Publications Inc., New York, 1958.

ELSGOLTS, L. *Differential equations and the calculus of variations.* Mir Publishers, Moscow, 1980.

PHILIPS S. J. ; ANDERSON R. P. ; SCHAPIRE R. E. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* , v.190, p.231-259, 2006.