Ann. Inst. Statist. Math.
38 (1986), Part A, 451-457

THE APPLICATION OF THE PRINCIPLE OF MINIMUM CROSS-ENTROPY TO THE CHARACTERIZATION OF THE EXPONENTIAL-TYPE PROBABILITY DISTRIBUTIONS

MONICA E. BAD DUMITRESCU

(Received Feb. 25, 1985; revised May 15, 1985)

Summary

Systematic and simple characterizations are presented for several familiar distributions in exponential family by means of the principle of minimum cross-entropy (minimum discrimination information). The suitable prior distributions and the appropriate constraints on expected values are given for the underlying distributions.

1. Introduction

The well-known principle of maximum entropy has been used in order to obtain the most of probability distributions on the real line ([3], [4], [8]).

The principle of minimum cross-entropy provides a general method of inference about an unknown probability density, when there exist a prior estimate of the density and new information in the form of constraints on expected values. Introduced by S. Kullback and R. A. Leibler ([5], [6]) under the name of "minimum discrimination information", it has been applied successfully in a remarkable variety of fields, inclusively in statistics. J. E. Shore and R. W. Johnson ([9], [10]) established four consistency axioms for the principle of minimum cross-entropy, so that it is correct in a specified sense.

The approach here applies this principle to the characterization of the exponential-type probability distributions. In Section 2, Kullback's estimation criterion is briefly resumed. Since the exponential-type probability distributions are equivalent and they are absolutely continuous with respect to the Lebesgue measure on the real line, a natural question arises: which prior distribution and new information lead to a specified exponential distribution, by the minimization of the variation of information? In Section 2, we consider that the prior esti-

Key words: Minimum cross-entropy, exponential distribution.

mate of the general probability density of exponential-type is one of "the simplest" members of the exponential family: the standard normal density in the case of a distribution on $(-\infty, \infty)$ and the exponential density with the mean value equal to one in the case of a distribution on $[0, \infty)$. The general forms of the constraints on expected values are established. Section 3 contains the characterizations by the principle of minimum cross-entropy of the normal, the negative exponential, the gamma and the Student distributions.

2. The principle of minimum-cross entropy inside the family of exponential-type probability distributions

Let us consider two probability spaces $(\Omega, \mathcal{K}, \mu_1)$ and $(\Omega, \mathcal{K}, \mu_2)$ such that the probability measures μ_1 and μ_2 are equivalent, i.e., they are absolutely continuous with respect to each other. Let us consider the third probability measure λ defined on $\mathcal{K}$ such that it is equivalent both to μ_1 and μ_2. Let $f_i(w)$ $(i=1, 2)$ be the Radon-Nikodym derivatives of μ_i with respect to λ. The cross-entropy (the discrimination information) is

$$I(1; 2)=\int_\Omega \ln \frac{f_1(w)}{f_2(w)} f_1(w)d\lambda(w) .$$

If we take $\lambda=\mu_2$, $I(1; 2)$ is the variation of information when the initial probability measure λ is replaced by the probability measure μ_1.

In the following, we use Kullback's estimation criterion in the formulation given by S. Guiaşu ([2]):

The estimation criterion

The probability density $f_1^*(w)\geqq 0$, $\int_\Omega f_1^*(w)d\lambda(w)=1$ which minimizes the cross-entropy

$$I(1; 2)=\int_\Omega \ln \frac{f_1(w)}{f_2(w)} f_1(w)d\lambda(w)$$

subject to the constraint

$$\int_\Omega g(w)f_1(w)d\lambda(w)=\theta$$

is given by

$$f_1^*(w)=\frac{1}{\Phi(\hat{\beta})} f_2(w) \exp[-\hat{\beta}g(w)] ,$$

where $\Phi(\beta)=\int_\Omega f_2(w) \exp[-\beta g(w)]d\lambda(w)$ and $\hat{\beta}$ is the unique solution of the equation $d \ln \Phi(\beta)/d\beta = -\theta$, g being a nondegenerate random variable.

Now let us consider the family of exponential-type probability distributions on the real line, given by the family of probability densities with respect to the Lebesgue measure $\mathcal{F} = \{f(x, \theta) = \exp[a(\theta)b(x) + c(\theta) + d(x)],\ x \in X,\ \theta \in \Theta\}$, where X is the support of the probability measure with the density $f(x, \theta)$ and Θ is the parameter space,

$$\Theta = \left\{\theta \,\middle|\, \int_X \exp[a(\theta)b(x) + d(x)]dx < \infty\right\} .$$

Since the exponential-type probability distributions are equivalent and they are absolutely continuous with respect to the Lebesgue measure on the real line, we can consider the cross-entropy, or the variation of information inside the exponential family.

In order to obtain a characterization of a member of the family $\mathcal{F}$, we solve the problem of minimization of the variation of information, when a suitable prior probability density $f_2(x) \in \mathcal{F}$ is replaced by the underlying distribution of the family $\mathcal{F}$.

The case $X = [0, \infty)$

Let us consider that "the simplest" member of the family is the exponential density with the mean value equal to one:

$$f_2(x) = \exp(-x) , \qquad x \geqq 0 .$$

Using this distribution as the prior estimate, we get the general probability density of exponential-type on $[0, \infty)$.

The random variable $g(x, \theta) = a(\theta)b(x) + c(\theta) + d(x) + x$ has the expected value with respect to the probability density $f(x, \theta) \in \mathcal{F}$ equal to

$$\mathrm{E}_\theta[g(x, \theta)] = \frac{c(\theta)a'(\theta) - a(\theta)c'(\theta)}{a'(\theta)} + \mathrm{E}_\theta[d(x) + x] ,$$

since $\mathrm{E}_\theta[b(x)] = -c'(\theta)/a'(\theta)$. (Here $a'(\theta)$, $c'(\theta)$ denote the derivatives of $a(\theta)$, $c(\theta)$ with respect to θ).

We denote $M(\theta) = \mathrm{E}_\theta[g(x, \theta)]$ and let us suppose that $M(\theta) < \infty$ for any $\theta \in \Theta$.

$$I(1;2) = \int_0^\infty \ln \frac{f_1(x)}{\exp(-x)} f_1(x)dx$$

$$\begin{aligned}\Phi(\beta; \theta) &= \mathrm{E}_2[\exp(-\beta g(x, \theta))] \\ &= \int_0^\infty \exp\{-\beta[a(\theta)b(x) + c(\theta) + d(x) + x] - x\} dx .\end{aligned}$$

The equation $\partial \ln \Phi(\beta; \theta)/\partial\beta = -M(\theta)$ has the unique solution $\hat{\beta} = -1$ and $\Phi(\hat{\beta}; \theta) = 1$.

Following Kullback's criterion, the solution of the problem $\inf\left\{I(1;2)\,\middle|\,\int_0^\infty f_1(x)dx=1,\ \int_0^\infty g(x,\theta)f_1(x)dx=M(\theta)\right\}$ is $f_1^*(x,\theta)=\exp[a(\theta)b(x)+c(\theta)+d(x)]$.

Hence, the exponential-type probability distributions on $[0,\infty)$ can be obtained by Kullback's criterion, when we use as prior estimation the negative exponential distribution with the mean value equal to one.

The case $X=(-\infty,\infty)$

In this case "the simplest" member of the family $\mathcal{F}$ is the standard normal density

$$f_2(x)=\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\,,\qquad x\in(-\infty,\infty)\,.$$

Using this distribution as prior estimate, we get the general probability density of exponential-type on the real line.

The random variable $h(x,\theta)=a(\theta)b(x)+c(\theta)+d(x)+x^2/2$ has the expected value with respect to the probability density $f(x,\theta)\in\mathcal{F}$ equal to

$$\mathrm{E}_\theta[h(x,\theta)]=\frac{c(\theta)a'(\theta)-a(\theta)c'(\theta)}{a'(\theta)}+\mathrm{E}_\theta[d(x)+x^2/2]\,.$$

We denote $N(\theta)=\mathrm{E}_\theta[h(x,\theta)]$ and let us suppose that $N(\theta)<\infty$, $\theta\in\Theta$.

$$I(1;2)=\int_{-\infty}^{\infty}\ln\frac{f_1(x)}{(1/\sqrt{2\pi})\exp(-x^2/2)}f_1(x)dx$$

$$\begin{aligned}\Phi(\beta;\theta)&=\mathrm{E}_2[\exp(-\beta h(x,\theta))]\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left\{-\beta\left[a(\theta)b(x)+c(\theta)+d(x)+\frac{x^2}{2}\right]-\frac{x^2}{2}\right\}dx\,.\end{aligned}$$

The equation $\partial\ln\Phi(\beta;\theta)/\partial\beta=-N(\theta)$ has the unique solution $\hat\beta=-1$ and $\Phi(\hat\beta;\theta)=1/\sqrt{2\pi}$.

Following Kullback's criterion, the solution of the problem $\inf\left\{I(1;2)\,\middle|\,\int_{-\infty}^\infty f_1(x)dx=1,\ \int_{-\infty}^\infty h(x,\theta)f_1(x)dx=N(\theta)\right\}$ is $f_1^*(x,\theta)=\exp[a(\theta)b(x)+c(\theta)+d(x)]$.

Hence, the exponential-type probability distributions on the real line can be obtained by Kullback's criterion, when we use as prior estimate the standard normal distribution.

3. Applications

In this section we establish the prior distributions which are "the closest" (in the sense of the minimization of the variation of information) to several familiar distributions in exponential family. Thus we obtain new characterizations of the normal, the negative exponential, the gamma and the Student distributions.

1) *The negative exponential distribution with the mean value* λ $(\lambda>0)$

Let us consider the prior density $f_2(x)=\exp(-x)$, $x\geqq 0$. The negative exponential distribution with the mean value λ is the solution of the problem

$$\inf\left\{I(1;2)\,\middle|\,\int_0^\infty f_1(x)dx=1,\ \int_0^\infty xf_1(x)dx=\lambda\right\}.$$

In fact, $\Phi(\beta)=\mathrm{E}_2[\exp(-\beta x)]=1/(\beta+1)$ and the equation $d\ln\Phi(\beta)/d\beta=-\lambda$ has the unique solution $\hat{\beta}=1/\lambda-1$. Then $\Phi(\hat{\beta})=\lambda$ and

$$f_1^*(x,\lambda)=\frac{1}{\lambda}\exp\left(-\frac{1}{\lambda}x\right),\qquad x\geqq 0.$$

2) *The gamma distribution with parameters* a, b $(a\in N,\ a>1,\ b>0)$

We consider again the prior density $f_2(x)=\exp(-x)$, $x\geqq 0$. The gamma distribution is the solution of the problem

$$\inf\left\{I(1;2)\,\middle|\,\int_0^\infty f_1(x)dx=1,\ \int_0^\infty[(a-1)\ln x-(b-1)x]f_1(x)dx=M(a,b)\right\}$$

where $M(a,b)=(a-1)\left[\sum_{k=1}^{a-1}\frac{1}{k}-\ln b-C\right]-\frac{a(b-1)}{b}$ and C is Euler's constant. In fact,

$$\begin{aligned}\Phi(\beta;a,b)&=\mathrm{E}_2\{\exp[-\beta((a-1)\ln x-(b-1)x)]\}\\&=\int_0^\infty\exp[-\beta((a-1)\ln x-(b-1)x)-x]dx.\end{aligned}$$

The equation $\partial\ln\Phi(\beta;a,b)/\partial\beta=-M(a,b)$ has the unique solution $\hat{\beta}=-1$. Then $\Phi(\hat{\beta};a,b)=\Gamma(a)/b^a$ and

$$f_1^*(x;a,b)=\frac{b^a}{\Gamma(a)}x^{a-1}\exp(-bx),\qquad x\geqq 0.$$

We notice that the case $a=1$ reduces to the previous one.

3) *The normal distribution* $N(m,1)$, $(m\in R)$

Let us consider that the prior probability density $f_2(x)$ is the standard normal density. The normal distribution $N(m,1)$ is the solution of the problem

$$\inf\left\{I(1;2)\,\middle|\,\int_{-\infty}^\infty f_1(x)dx=1,\ \int_{-\infty}^\infty xf_1(x)dx=m\right\}.$$

In fact, $\Phi(\beta)=\mathrm{E}_2[\exp(-\beta x)]=\exp(\beta^2/2)$ and the equation $d\ln\Phi(\beta)/d\beta=-m$ has the unique solution $\hat{\beta}=-m$. Then $\Phi(\hat{\beta})=\exp(m^2/2)$ and

$$f_1^*(x, m)=\frac{1}{\sqrt{2\pi}}\exp[-(x-m)^2/2], \qquad x\in R.$$

4) *The normal distribution* $N(0, \sigma^2)$, $(\sigma^2>0)$

We consider again as prior estimate $f_2(x)$ the standard normal density. The normal distribution $N(0, \sigma^2)$ is the solution of the problem

$$\inf\left\{I(1;2)\,\middle|\int_{-\infty}^{\infty}f_1(x)dx=1,\ \int_{-\infty}^{\infty}x^2f_1(x)dx=\sigma^2\right\}.$$

$$\Phi(\beta)=\mathrm{E}_2[\exp(-\beta x^2)]=1/\sqrt{1+2\beta}, \qquad \beta>-1/2.$$

The equation $d\ln\Phi(\beta)/d\beta=-\sigma^2$ has the unique solution $\hat{\beta}=1/2\sigma^2-1/2$ and $\Phi(\hat{\beta})=\sqrt{\sigma^2}$. Then

$$f_1^*(x, \sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-x^2/2\sigma^2), \qquad x\in R.$$

5) *The Student distribution* $t(n)$, $(n\geqq 2)$

Let us consider as prior probability density the standard normal density. The Student distribution is the solution of the problem

$$\inf\left\{I(1;2)\,\middle|\int_{-\infty}^{\infty}f_1(x)dx=1,\ \int_{-\infty}^{\infty}\left(\frac{x^2}{2}-\frac{n+1}{2}\ln\left(1+\frac{x^2}{n}\right)\right)f_1(x)dx=N(n)\right\}$$

where

$$N(n)=n\Gamma\left(\frac{n}{2}-1\right)\Big/4\Gamma\left(\frac{n}{2}\right)-(n+1)\left(\sum_{k=0}^{n-2}\frac{(-1)^{n-k}}{k+1}+(-1)^{n-1}\ln 2\right).$$

In fact,

$$\Phi(\beta;n)=\mathrm{E}_2\left\{\exp\left[-\beta\left(\frac{x^2}{2}-\frac{n+1}{2}\ln\left(1+\frac{x^2}{n}\right)\right)\right]\right\}$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left[-\beta\left(\frac{x^2}{2}-\frac{n+1}{2}\ln\left(1+\frac{x^2}{n}\right)\right)-\frac{x^2}{2}\right]dx.$$

The equation $\partial\ln\Phi(\beta;n)/\partial\beta=-N(n)$ has the unique solution $\hat{\beta}=-1$ and

$$\Phi(\hat{\beta};n)=\frac{1}{\sqrt{2\pi}}\frac{\Gamma(n/2)\sqrt{n\pi}}{\Gamma((n+1)/2)}.$$

Then

$$f_1^*(x, n)=\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}, \qquad x\in(-\infty,\infty).$$

Acknowledgment

The author wishes to thank the referee for his valuable suggestions for improving the presentation.

UNIVERSITY OF BUCHAREST

REFERENCES

[1] Bad Dumitrescu, M. (1984). The minimum cross-entropy estimation of a parameter, *Bull. Math. Soc. Sci. Math Roum.*, 28, **4**, 291-297.

[2] Guiaşu, S. (1977). *Information Theory with Applications*, McGraw-Hill, New York.

[3] Jaynes, E. T. (1957). Information theory and statistical mechanics, *Phis. Rev.*, **106**, 620-630, **108**, 171-182.

[4] Kampé de Fériet, J. (1963). Théorie de l'Information. Principle du Maximum de l'Entropie et ses Applications à la Statistique et à la Mécanique, Publications du Laboratoire de Calcul de la Faculté de Sciences de l'Université de Lille, Lille.

[5] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, *Ann. Math. Statist.*, **22**, 79-86.

[6] Kullback, S. (1959). *Information Theory and Statisics*, Wiley, New York.

[7] Kullback, S. and Khairat, M. A. (1966). A note on minimum discrimination information, *Ann. Math. Statist.*, **37**, 279-280.

[8] Preda, V. (1982). The Student distribution and the principle of maximum entropy, *Ann. Inst. Statist. Math.*, **34**, 335-338.

[9] Shore, J. E. and Johnson, R. W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Trans. Inf. Theory*, IT, **26**, no. 1, 26-37.

[10] Shore, J. E. and Johnson, R. W. (1981). Properties of cross-entropy minimization, *IEEE Trans. Inf. Theory*, IT, **27**, no. 4, 472-482.