The Student distribution and the principle of maximum entropy

Article *in* Annals of the Institute of Statistical Mathematics · December 1982

DOI: 10.1007/BF02481032

1 author:

Vasile Preda
University of Bucharest

Ann. Inst. Statist. Math.
34 (1982), Part A, 335-338

THE STUDENT DISTRIBUTION AND THE PRINCIPLE OF MAXIMUM ENTROPY

VASILE C. PREDA

(Received Nov. 13, 1980; revised Dec. 11, 1981)

Summary

The Student distribution is obtained by means of the principle of maximum entropy.

1. Introduction

The principle of maximum entropy can be considered as such a criterion to select the best probability distributions compatible to some set of restraints. According to this principle, we choose the probability distribution which maximizes the entropy or the conditional entropy compatible to some set of restraints. This principle was established, according to S. Guiaşu [2], independently, by R. S. Ingarden [3], E. T. Jaynes [5], S. Kullback and R. A. Leibler [7]. Using this principle, E. T. Jaynes [5] obtained the exponential distribution, J. Kampé de Fériet [6] obtained the normal distribution, R. S. Ingarden and A. Kossakowski [4] obtained the Poisson distribution and V. Preda [8] obtained the gamma distribution.

In this paper we will show that the Student distribution can be obtained by means of the principle of maximum entropy.

2. Statement of the problem

Let $(\Omega, \mathcal{K}, P)$ and $(T, \mathcal{T}, \mu)$ be two probability spaces where $T=(-\infty, +\infty)$, $\mathcal{T}$ is the class of Lebesgue measurable sets on T, μ is Lebesgue measure on $\mathcal{T}$. If X is a random variable defined on $(\Omega, \mathcal{K}, P)$ and taking values in $(T, \mathcal{T})$, then the corresponding probability distribution will be: $\nu_X = P\circ X^{-1}$. If $\nu_X \ll \mu$ let $\rho_X(x)=(d\nu_X/d\mu)(x)$ be the probability density function of X and $H(X)=-\mathrm{E}_X \ln \rho_X(X)$ the entropy of the random variable X, where E_X denotes expectation under ν_X.

Let $\mathcal{X}_n$ be the set of all random variables X defined on $(\Omega, \mathcal{K}, P)$ and taking values in $(T, \mathcal{T})$ such that:

$$\nu_X \ll \mu \,, \tag{1}$$

$$\mathrm{E}_X \ln(1+n^{-1}\cdot X^2)=2\cdot a_n \,, \tag{2}$$

where n is a natural number and a_n is a real constant defined by:

$$a_1=\ln 2\,; \quad a_n=\sum_{k=0}^{n-2}(-1)^{n-k}(k+1)^{-1}+(-1)^{n-1}\ln 2\,, \quad n\geqslant 2\,. \tag{3}$$

Now, we consider the following problem:

$$\sup\{H(X)/X\in \mathcal{X}_n\}\,. \tag{4}$$

In this paper we shall prove the theorem:

THEOREM. *The solution of the problem* (4) *is given by the random variable* X *with the Student probability density function with* n *degrees of freedom.*

3. Lemmas

LEMMA 1. *Let be the gamma function*

$$\Gamma(a)=\int_0^{+\infty} x^{a-1}e^{-x}dx$$

for $a>0$. *Then, we have:*

$$\Gamma'(a)/\Gamma(a)=\int_0^1(1-y^{a-1})(1-y)^{-1}dy-C$$

if $a\neq 1$, *and equal to:* $-C$, *if* $a=1$, *where* C *is Euler's constant, and* $\Gamma'(a)=d\Gamma(a)/da$.

PROOF. See H. Bateman and A.Erdelyi [1], Subsection 1.7.

LEMMA 2. *For* $\alpha\in(-\infty,+\infty)$ *we define the function:*

$$J(\alpha)=\int_0^{\pi/2}(\cos t)^{\alpha-1}dt\,.$$

Then:

a) $J(\alpha)<+\infty$ *if and only if* $\alpha>0$; *in this case we have*

$$J(\alpha)=2^{-1}\sqrt{\pi}\,\Gamma(\alpha/2)/\Gamma((\alpha+1)/2)\,.$$

b) *For* $\alpha>0$, *we have:*

$$\begin{aligned} J'(\alpha)&=dJ(\alpha)/d\alpha\\ &=4^{-1}\sqrt{\pi}\,[\Gamma((\alpha+1)/2)\Gamma'(\alpha/2)-\Gamma(\alpha/2)\Gamma'((\alpha+1)/2)]\\ &\qquad\times[\Gamma((\alpha+1)/2)]^{-2}\,.\end{aligned}$$

PROOF. It is immediate if we use H. Bateman and A. Erdelyi [1], Subsection 1.5.1.

LEMMA 3. *If* $n \geqq 2$ *is a natural number, then*

$$\int_0^1 z^{n-1}(z+1)^{-1}dz = \sum_{k=0}^{n-2}(-1)^{n-k}(k+1)^{-1} + (-1)^{n-1}\ln 2$$

and for $n=1$,

$$\int_0^1 z^{n-1}(z+1)^{-1}dz = \ln 2 .$$

PROOF. It is immediate.

LEMMA 4. *Let* a_n *be defined by* (3). *Then, the equation:*

$$h(\alpha) = J'(\alpha)/J(\alpha) + a_n = 0$$

has a unique solution, $\alpha = n$.

PROOF. Because

$$h'(\alpha) = (J''(\alpha)J(\alpha) - (J'(\alpha))^2)/(J(\alpha))^2$$

by using the Schwartz inequality we have $h'(\alpha)>0$ for any $\alpha>0$ and h is an increasing function. Hence the equation $h(\alpha)=0$ has at most a solution. But according to Lemma 2, Lemma 1, Lemma 3 and relation (3) we have $h(n)=0$.

4. Proof of Theorem

Using Lagrange's multipliers method, we get

$$\begin{aligned} H(X) - u - 2a_n v &= \mathrm{E}_X \ln e^{-u}(1+n^{-1}X^2)^{-v}(\rho_X(X))^{-1} \\ &\leqq \mathrm{E}_X\, e^{-u}(1+n^{-1}X^2)^{-v}(\rho_X(X))^{-1} - 1 , \end{aligned}$$

with equality if and only if:

$$\rho_X(X) = e^{-u}(1+n^{-1}X^2)^{-v} , \qquad X \in \mathcal{X}_n . \tag{5}$$

Because $\rho_X(x)$ is a probability density function, by using a change of variable, we obtain

$$2\sqrt{n}\, J(2v-1) = e^u , \tag{6}$$

where $v>1/2$ (Lemma 2).

From (2) and (5), using a change of variable, too, it results:

$$-4\sqrt{n}\, J'(2v-1) = 2a_n e^u . \tag{7}$$

Now, from (6) and (7) we have

$$J'(\alpha)/J(\alpha)+a_n=0 \tag{8}$$

where $\alpha=2v-1$. But, according to Lemma 4 the equation (8) has a unique solution, $\alpha=n$. Hence

$$v=(n+1)/2 \tag{9}$$

and by (6) we obtain

$$e^u=\sqrt{n\pi}\,\Gamma(n/2)/\Gamma((n+1)/2)\,. \tag{10}$$

From (5), (9) and (10) it results

$$\rho_X(x)=(n\pi)^{-1/2}(1+n^{-1}x^2)^{-(n+1)/2}\Gamma((n+1)/2)/\Gamma(n/2)$$

which gives us the Student distribution with n degrees of freedom and the theorem is proved.

Acknowledgement

The author would like to thank the referee for his valuable comments on the paper.

UNIVERSITY OF BUCHAREST

REFERENCES

[1] Bateman, H. and Erdelyi, A. (1953). Higher Transcedental Functions, McGraw-Hill, New York.

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

[3] Ingarden, R. S. (1963). Information theory and variational principles in statistical theories, *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys.*, **11**, 541-547.

[4] Ingarden, R. S. and Kossakowski, A. (1970). Poisson probability distribution and information thermodynamics, *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys.*, **19**, 83-86.

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

[6] 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.

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

[8] Preda, V. (1980). Sur la distribution gamma et la principle du maximum information, submitted for publication to the *C. R. Acad. Sci. Paris.*