Ann. Inst. Statist. Math.
37 (1985), Part A, 535-539

A CONDITIONAL LIMIT CONSTRUCTION OF THE NORMAL PROBABILITY DENSITY

B. S. CHOI

(Received Jan. 20, 1984; revised Sept. 26, 1984)

Summary

It is shown that a normal probability density can be characterized as a limit of conditional probability densities of i.i.d. uniform random variables.

1. Introduction

Recently the normal probability density function (p.d.f.) appears in the answer of a conditional limit problem. If we let $X_1, X_2, \cdots$ be i.i.d. random variables with the common p.d.f. $g(x)$ that satisfies certain regularity conditions, then the conditional p.d.f. of X_1 given $\frac{1}{n}\sum_{i=1}^{n} X_i^2=\sigma^2$ tends to a p.d.f. of the form

$$f_\lambda(x)=c(\lambda)\exp(-\lambda x^2)g(x)\,, \tag{1.1}$$

where λ and $c(\lambda)$ are determined by the equations

$$\int f_\lambda(x)dx=1 \qquad \text{and} \qquad \int x^2 f_\lambda(x)dx=\sigma^2\,. \tag{1.2}$$

The limiting p.d.f. $f_\lambda(x)$ is the normalized product of a normal p.d.f. $\sqrt{\lambda/\pi}\exp(-\lambda x^2)$ and the initial p.d.f. $g(x)$. This theorem has been presented by Vincze [13], Bartfai [1], Lanford [8], Tjur [10], Zabell [14], Vasicek [12], and Van Campenhout and Cover [11] under various kinds of regularity conditions. A type of the regularity conditions of the initial p.d.f. $g(x)$, which is named Zabell's conditions, is summarized in Van Campenhout and Cover [11], p. 485. Another possible choice of the conditions is the same as those of the local limit theorem. (See, e.g., Feller [4], pp. 488–491.) The purpose of this paper is to obtain the

Key words: Normal probability density, uniform probability density, conditional probability density, Kullback-Leibler information, maximum entropy.

normal p.d.f. $\phi(x; \mu, \sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$ as an answer of the conditional limit problem. There are two candidates for the initial p.d.f. $g(x)$ that yield a normal p.d.f. as the limiting p.d.f. $f_\lambda(x)$. One is that the initial p.d.f. $g(x)$ itself is a normal p.d.f. Then the limiting p.d.f. $f_\lambda(x)$ is normal by (1.1). However this case doesn't arouse any interest to statisticians. The other is that $g(x)$ is uniform over $(-\infty, \infty)$. Then the limiting function $f_\lambda(x)$ does possibly become a normal p.d.f. Unfortunately the uniform function over $(-\infty, \infty)$ is improper and doesn't satisfy the regularity conditions. In this paper we will overcome these deficiencies and obtain a normal p.d.f. as a limit of conditional p.d.f.'s of uniform random variables.

2. The theorems

Two theorems will be presented to construe a normal p.d.f. as a limit of conditional p.d.f.'s. Even though the two theorems can be combined into one, Theorem 1 is separated from Theorem 2 because of its own interest.

The first theorem will show that the conditional p.d.f. of Y_1 given $\frac{1}{n}\sum_{i=1}^{n} Y_i^2=\sigma^2$, where $Y_1, Y_2, \cdots, Y_n$ are i.i.d. uniform random variables with mean 0 and variance $n\sigma^2/3$, can be interpreted as that of Y_1 given $\frac{1}{n}\sum_{i=1}^{n} Y_i^2=\sigma^2$, where $Y_1, Y_2, \cdots, Y_n$ are i.i.d. normal random variables with mean 0 and variance σ^2.

THEOREM 1. *Let $Y_1, Y_2, \cdots, Y_n$ be a sequence of i.i.d. random variables with uniform probability density*

$$g_n(y)=\frac{1}{2\sigma\sqrt{n}} I_{(-\sigma\sqrt{n}, \sigma\sqrt{n})}(y) , \tag{2.1}$$

where $\sigma>0$. Then the conditional p.d.f. of Y_1 given $\frac{1}{n}\sum_{i=1}^{n} Y_i^2=\sigma^2$ is the same as that of Y_1 given $\frac{1}{n}\sum_{j=1}^{n} Y_j^2=\sigma^2$ where $\{Y_j: 1\leqq j\leqq n\}$ is a sequence of i.i.d. normal random variables with mean 0 and variance σ^2.

PROOF. Let $Y_1, Y_2, \cdots, Y_n$ be i.i.d. random variables symmetric about zero with p.d.f. $f(y)$. If we define $Z_1, Z_2, \cdots, Z_n$ by

$$Z_i=Y_i , \qquad i=1, 2, \cdots, n-1 ,$$

and

$$Z_n=\frac{1}{n}\sum_{i=1}^{n} Y_i^2 ,$$

then the joint p.d.f. of $(Z_1, Z_2, \cdots, Z_n)$ is

$$h(z_1,\cdots, z_n)=f(z_1)\cdots f(z_{n-1})f((nz_n-z_1^2-\cdots-z_{n-1}^2)^{1/2}) \cdot n(nz_n-z_1^2-\cdots z_{n-1}^2)^{-1/2} . \tag{2.2}$$

If $z_n=\sigma^2$, then $|z_i|\leqq\sigma\sqrt{n}$, $i=1, 2,\cdots, n-1$. Thus the p.d.f. in (2.2) becomes

$$h(z_1,\cdots, z_{n-1}, \sigma^2) =\begin{cases} n(2\sigma\sqrt{n})^{-n}(n\sigma^2-z_1^2-\cdots-z_{n-1}^2)^{-1/2} , & \text{if } f(y)=g_n(y) \\ n(2\pi\sigma^2)^{-n/2}(n\sigma^2-z_1^2-\cdots-z_{n-1}^2)^{-1/2}e^{-n/2} , & \text{if } f(y)=\phi(y; 0, \sigma^2) . \end{cases}$$

The conditional p.d.f. of Z_1 given $Z_n=\sigma^2$ is, in both cases,

$$\begin{aligned} &\int\cdots\int h(z_1,\cdots, z_{n-1}, \sigma^2)dz_2\cdots dz_{n-1}\Big/\int\cdots\int h(z_1,\cdots, z_{n-1}, \sigma^2)dz_1\cdots dz_{n-1} \\ &=\int\cdots\int (n\sigma^2-z_1^2-\cdots-z_{n-1}^2)^{-1/2}dz_2\cdots dz_{n-1}\Big/ \\ &\quad\int\cdots\int (n\sigma^2-z_1^2-\cdots-z_{n-1}^2)^{-1/2}dz_1\cdots dz_{n-1} \\ &=\left(1-\frac{z_1^2}{n\sigma^2}\right)^{-(n-2)/2}(n\sigma^2-z_1^2)^{-1/2}\pi^{-1/2}\Gamma\left(\frac{n}{2}\right)\Big/\Gamma\left(\frac{n-1}{2}\right) . \end{aligned} \tag{2.3}$$

The last equality holds by

$$\int_{x_1^2+\cdots+x_p^2\leq 1}\cdots\int (1-x_1^2-\cdots-x_p^2)^{-1/2}dx_1\cdots dx_p=\pi^{(p+1)/2}\Big/\Gamma\left(\frac{p+1}{2}\right) ,$$

(see Equation 4.632 of Gradshteyn and Ryzhik [5]). Q.E.D.

Theorem 1 can also be proved by Chernoff's tilting technique [2], which is useful to generalize Theorem 1 to other univariate cases (Van Campenhout and Cover [11]) and multivariate cases (Choi [3]).

It can be easily shown that the conditional p.d.f. of Y_1 given $\frac{1}{n}\sum_{i=1}^{n} Y_i^2=\sigma^2$, where $Y_1, Y_2,\cdots, Y_n$ are i.i.d. normal random variables with mean 0 and variance σ^2, tends to the nonconditional p.d.f. $\phi(y_1; 0, \sigma^2)$. Thus, Theorem 1 implies that the normal p.d.f. can be constructed as a limit of conditional p.d.f. of i.i.d. uniform random variables. However, if $n\to\infty$ then the initial p.d.f. $g_n(y)$ in (2.1) tends to an improper density. To avoid the improperness of the uniform density over $(-\infty, \infty)$, we'll define a double array of uniform random variables $\{X_{jn}: 1\leqq j\leqq n,\ n=1, 2,\cdots\}$.

THEOREM 2. *For each* n $(\geqq 1)$ *let* $X_{1n}, X_{2n}, \cdots, X_{nn}$ *be a sequence of i.i.d. uniform random variables with p.d.f.* $g_n(x)$. *Then the conditional p.d.f. of* X_{1n} *given* $\frac{1}{n}\sum_{i=1}^{n} X_{in}^2=\sigma^2$ *tends to the normal p.d.f. with mean* 0 *and variance* σ^2 *as* $n\to\infty$.

PROOF. Theorem 1 has shown that the conditional p.d.f. of $X_{1n}=x$ given $\frac{1}{n}\sum_{i=1}^{n} X_{in}=\sigma^2$ is Equation 2.3. Applying Stirling's formula to the equation gives us that the conditional p.d.f. tends to the normal p.d.f. with mean 0 and variance σ^2 as $n\to\infty$. Q.E.D.

It is worth mentioning that Theorem 2 can also be proved by the local limit theorem.

3. Some comments

If $f_n(x)$ is defined by

$$f_n(x)=c(n)\exp\{-\lambda(n)x^2\}I_{(-\sigma\sqrt{n},\sigma\sqrt{n})}(x)\ , \tag{3.1}$$

where $c(n)$ and $\lambda(n)$ are determined by the constraints

$$\int f_n(x)dx=1 \qquad \text{and} \qquad \int x^2 f_n(x)dx=\sigma^2\ , \tag{3.2}$$

the p.d.f. $f_n(x)$ minimizes the Kullback-Leibler [7] mean information for discrimination between $f(x)$ and $g_n(x)$, i.e., $D(f;g_n)=\int f(x)\ln\{f(x)/g_n(x)\}dx$, among the p.d.f.'s satisfying the constraints

$$\int f(x)dx=1 \qquad \text{and} \qquad \int x^2 f(x)dx=\sigma^2\ . \tag{3.3}$$

(Sanov [9].) Therefore the p.d.f. $f_n(x)$ is said to be closest to the initial p.d.f. $g_n(x)$, in the Kullback-Leibler sense, among the p.d.f.'s satisfying the constraints (3.3). It can be easily calculated from the equations in (3.2) that $\lim_{n\to\infty}\lambda(n)=(2\sigma^2)^{-1}$ and $\lim_{n\to\infty}c(n)=(2\pi\sigma^2)^{-1/2}$, i.e., $\lim_{n\to\infty}f_n(x)=\phi(x;0,\sigma^2)$. Thus we may roughly say that the normal p.d.f. $\phi(x;0,\sigma^2)$ is the closest density in the Kullback-Leibler sense to the limit of p.d.f.'s of the uniform random variables $\{X_{1n}|n=1,2,\cdots\}$ among the p.d.f.'s satisfying the constraints (3.3).

A uniform p.d.f. is the maximum entropy p.d.f. without any restrictions, and the normal p.d.f. $\phi(x;0,\sigma^2)$ is the maximum entropy p.d.f. subject to the constraints $E(X)=0$ and $E(X^2)=\sigma^2$, which corresponds to $\frac{1}{n}\sum_{i=1}^{n} X_i=0$ and $\frac{1}{n}\sum_{i=1}^{n} X_i^2=\sigma^2$, respectively. (Kagan et al. [6], p. 410.) Thus Theorem 2 looks like a natural result.

Acknowledgements

The author is grateful to a referee whose comments led to a shortening of the proof of Theorem 1.

YONSEI UNIVERSITY

REFERENCES

[1] Bartfai, P. (1972). On a conditional limit theorem, *Progress in Statistics: Colloquia Mathematica Societatis János Bolyai*, Vol. **9-I**, 85-91.

[2] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of hypothesis based on the sum of observations, *Ann. Math. Statist.*, **23**, 493-507.

[3] Choi, B. S. (1983). A conditional limit characterization of the maximum entropy spectral density in time series analysis, Ph. D. Dissertation, Stanford University, Stanford, California.

[4] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. **2**, John Wiley and Sons, New York.

[5] Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Tables of Integrals, Series and Products*, Academic Press, New York.

[6] Kagan, A. M., Linnik, Y. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*, John Wiley and Sons, New York.

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

[8] Lanford, D. E. (1973). Entropy and equilibrium states in classical statistical mechanics, *Statistical Mechanics and Mathematical Problems, Lecture Notes in Physics*, No. **20**, 1-113, Springer-Verlag, Berlin.

[9] Sanov, I. N. (1957). On the probability of large deviations of random variables, *IMS and AMS Selected Translations in Mathematical Statistics and Probability*, Vol. **1**, (1961), 213-244.

[10] Tjur, T. (1974). *Conditional Probability Distributions*, Institute of Mathematical Statistics, University of Copenhagen.

[11] Van Campenhout, J. M. and Cover, T. M. (1981). Maximum entropy and conditional probability, *IEEE Trans. on Information Theory*, Vol. **IT-27**, 483-489.

[12] Vasicek, O. A. (1980). A conditional law of large numbers, *Ann. Prob.*, **8**, 142-147.

[13] Vincze, I. (1972). On the maximum probability principle in statistical physics, *Progress in Statistics: Colloquia Mathematica Societatis Janos Bolyai*, **9-II**, 869-893.

[14] Zabell, S. L. (1974). A limit theorem for conditional expectations with applications to probability and statistical mechanics, Ph. D. Dissertation, Harvard University, Cambridge, Mass.