A Note on Bivariate Distributions That Are Conditionally Normal

ANDREW GELMAN and XIAO–LI MENG*

It is possible for a nonnormal bivariate distribution to have conditional distribution functions that are normal in both directions. This article presents several examples, with graphs, including a counterintuitive bimodal joint density. The graphs simultaneously display the joint density and the conditional density functions, which appear as Gaussian curves in the three-dimensional plots.

KEY WORDS: Bimodality; Bivariate normal distribution; Conditional probability.

1. INTRODUCTION

It is well known that the pair of marginal distributions does not uniquely determine a bivariate distribution; for example, a bivariate distribution with normal marginals need not be jointly normal (Feller 1966, p. 69). In contrast, the *conditional* distribution functions uniquely determine a joint density function (Arnold and Press 1989). A natural question then arises: Must a bivariate distribution with normal conditional distributions be jointly normal? The answer is no; in fact, the joint distribution thus specified must fall in a parametric exponential family that we show includes such oddities as bimodal densities and a distribution with constant conditional means but nonconstant conditional variances. This article presents a simple expression for the distributional result derived in Castillo and Galambos (1987); we then graph some examples of bivariate density functions.

In general, a multivariate distribution on the variables $(x_1, \ldots, x_n)$ may be characterized by its joint distribution or the conditional distributions of $(x_i \mid x_j, \text{ all } j \neq i)$ for all i. For many models, one can specify the set of conditional distributions but cannot directly identify the joint distribution; Brook (1964) and Besag (1974) connected these two specifications for nearest-neighbor and Gibbs distributions and showed that the set of conditional distributions for all x_i determines the joint distribution. In addition, the set of conditional distributions is constrained by the requirement that they be consistent; that is, a single joint distribution should exist that reduces to each conditional distribution. Even in the bivariate case, interesting complications arise, as in the example of this article.

Dawid (1979) and others stressed the importance of identifying models by their conditional distributions; our work may be of practical importance because we expand the class of multivariate distributions that can be simply specified by conditionals. The supply of tractable joint distributions is limited, and it may be useful, for example, to model a bimodal joint density using only conditional normal densities (see Fig. 3).

2. PARAMETRIC FAMILY

Let x_1 and x_2 be two jointly distributed random variables, for which x_1 is normally distributed given x_2 and vice versa. Then their joint distribution, after location and scale transformations in each variable, can be written as

$$f(x_1, x_2) \propto \exp(-\tfrac{1}{2}[Ax_1^2x_2^2 + x_1^2 + x_2^2 - 2Bx_1x_2 - 2C_1x_1 - 2C_2x_2]), \quad (1)$$

whence the conditional distributions are

$$x_1 \mid x_2 \sim N\left(\frac{Bx_2 + C_1}{Ax_2^2 + 1}, \frac{1}{Ax_2^2 + 1}\right),$$

$$x_2 \mid x_1 \sim N\left(\frac{Bx_1 + C_2}{Ax_1^2 + 1}, \frac{1}{Ax_1^2 + 1}\right).$$

The only restrictions for (1) to be a probability density function are that $A \geq 0$, and if $A = 0$, then $|B| < 1$. One can see the conditional variances are constant iff $A = 0$, in which case the conditional mean functions are linear and the joint distribution is Gaussian.

This result can be extended to the general multivariate problem of variables $x_1, \ldots, x_n$ whose conditional distributions $(x_i \mid x_j, \text{ all } j \neq i)$ are Gaussian for all i. The resulting joint density must be of the form

$$f(x_1, \ldots, x_n) \propto \exp\left(-\frac{1}{2}\sum A_k x_1^{k_1} \ldots x_n^{k_n}\right).$$

The summation is taken over all 3^n values of the exponents defined by each k_i attaining the values 0, 1, or 2. The coefficients A_k are allowed to take on any real values for which the joint density function has a finite integral.

3. EXAMPLES

We illustrate the diversity of this distributional family with graphs of three bivariate densities that clearly differ from joint normality. Consider for simplicity the symmetric subfamily in which $A = 1$, $B = 0$, $C_1 = C_2 = C$, with conditional distributions

$$x_1 \mid x_2 \sim N\left(\frac{C}{1 + x_2^2}, \frac{1}{1 + x_2^2}\right),$$

*Andrew Gelman is Assistant Professor, Department of Statistics, University of California, Berkeley, CA 94720. Xiao-Li Meng is Assistant Professor, Department of Statistics, University of Chicago, Chicago, IL 60637. The authors thank Chuanhai Liu for posing this problem, Donald Rubin for comments, and the National Science Foundation for financial support.

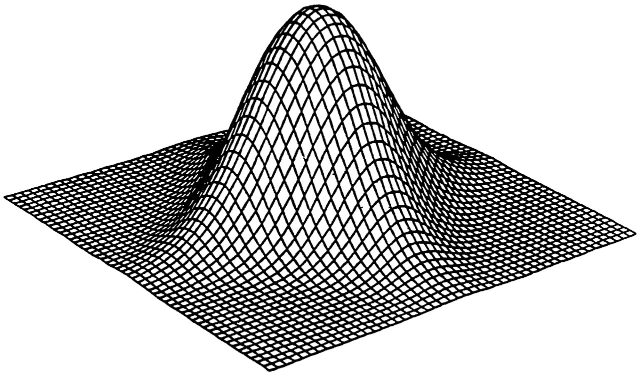

Figure 1. $f(x_1, x_2) \propto \exp(-\frac{1}{2}[x_1^2x_2^2 + x_2^2])$.

and similarly for $x_2 \mid x_1$. Figures 1–3 illustrate the corresponding joint densities for the values $C = 0, 1, 4$. Note that the grid lines in the graphs, which are just unnormalized conditional density functions, are clearly Gaussian. Figure 1 shows a joint density with zero conditional means that differs from a Gaussian by having nonconstant conditional variances. The distribution shown in Figure 2 is amusing in that $(x_1 \mid x_2) \sim N(1/(x_2^2 + 1), 1/(x_2^2 + 1))$ and vice versa, so the conditional mean equals

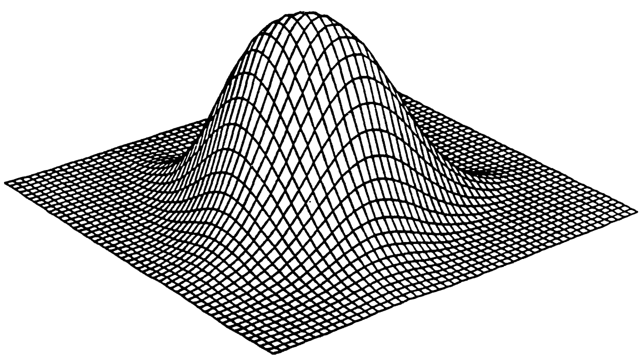

Figure 2. $f(x_1, x_2) \propto \exp(-\frac{1}{2}[x_1^2x_2^2 + x_1^2 + x_2^2 - 2x_1 - 2x_2])$.

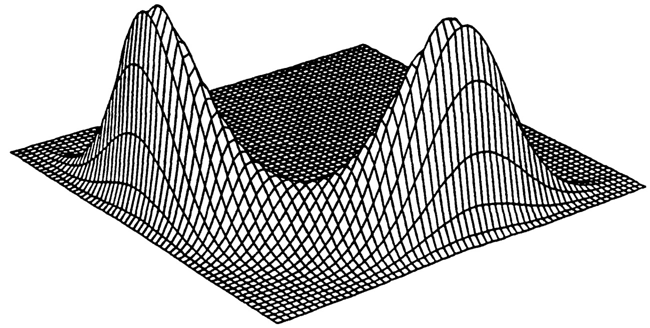

Figure 3. $f(x_1, x_2) \propto \exp(-\frac{1}{2}[x_1^2x_2^2 + x_1^2 + x_2^2 - 8x_1 - 8x_2])$.

the conditional variance at all points. Figure 3 presents a counterintuitive example of a bimodal joint density with bimodal marginals but Gaussian conditional densities. It is easily shown that, within this subfamily, the joint density is bimodal iff $C > 2$.

[*Received May 1990.*]

REFERENCES

Arnold, B. C., and Press, S. J. (1989), "Compatible Conditional Distributions," *Journal of the American Statistical Association,* 84, 152–156.

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society,* Ser. B, 36, 192–225.

Brook, D. (1964), "On the Distinction Between the Conditional Probability and the Joint Probability Approaches in the Specification of Nearest-Neighbour Systems," *Biometrika,* 51, 481–483.

Castillo, E., and Galambos, J. (1987), "Bivariate Distributions With Normal Conditionals," in *Proceedings of the International Association of Science and Technology for Development* (International Symposium on Simulation Modelling and Development, Cairo, Egypt), Anaheim, CA: Acta Press, pp. 59–62.

Dawid, A. P. (1979), "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society,* Ser. B, 41, 1–31.

Feller, W. (1966), *An Introduction to Probability Theory and Its Applications* (Vol. 2), New York: John Wiley.