of Proposition 3 are violated, and the sample median is not a U statistic.

Tukey (1958) suggests that $\text{var}[n^{1/2}T_n]$ can be estimated by the bias-corrected sample variance of the bracketed terms in (4). When T_n is a U statistic, this becomes

$$V_n = (n-1)^{-1}\sum_{i=1}^{n}[nT_n - (n-1)T_{n-1}^{(i)} - T_n]^2$$
$$= (n-1)\sum_{i=1}^{n}(T_n - T_{n-1}^{(i)})^2.$$

By algebraic manipulation, it can be shown that

$$T_n - T_{n-1}^{(i)} = (m/(n-m))[\hat{k}_1(X_i) - T_n],$$

where

$$\hat{k}_1(X_i) = \binom{n-1}{m-1}^{-1}\sum_{S_i(n-1,m-1)} k(X_i, X_{j_1}, X_{j_2}, \ldots, X_{j_{m-1}}).$$

Thus,

$$V_n = [m^2(n-1)/(n-m)^2]\sum_{i=1}^{n}(\hat{k}_1(X_i) - T_n)^2.$$

This result also appears in Sen (1960,1981), where it is derived from a different approach. In particular, Sen proposes the estimator

$$V_s = m^2(n-1)^{-1}\sum_{i=1}^{n}(\hat{k}_1(X_i) - T_n)^2$$

and shows that it is a strongly consistent estimator of $\lim_{n\to\infty}\text{var}(n^{1/2}T_n)$. Thus the jackknife estimator $V_n = (n-1)^2 V_s/(n-m)^2$ is also consistent. For example, when the population is normal with variance σ^2 and T_n is the sample variance, then $EV_n = 2n\sigma^4/(n-2)$ and $EV_s = 2n(n-2)\sigma^4/(n-1)^2$, while $\text{var}(n^{1/2}T_n) = 2n\sigma^4/(n-1)$. Thus, in this example, V_n tends to overestimate the variance, while V_s underestimates it.

Finally, we note that occasionally the jackknife will transform a non$U$ statistic into a U statistic. The kernel can then be found using the previously stated results, and the asymptotic properties of the jackknifed estimator can be deduced. An example is the statistic $n^{-1}\Sigma(X_i - \bar{X})^2$, which becomes $(n-1)^{-1}\Sigma(X_i - \bar{X})^2$ after jackknifing. Some relationships between the jackknife and U statistics are mentioned elsewhere in the literature. Mantel (1967) asserts that T_n^J is composed of U statistics of degrees n and $n-1$ (though technically, the degree must be fixed), and Arvesen (1969) discusses jackknifing a function of a U statistic and suggests that Proposition 3 may be true.

[*Received October 1982. Revised June 1983.*]

REFERENCES

ARVESEN, J.N. (1969), "Jackknifing U-Statistics," *Annals of Mathematical Statistics,* 40, 2076–2100.

HOEFFDING, W. (1948), "A Class of Statistics With Asymptotically Normal Distribution," *Annals of Mathematical Statistics,* 19, 293–325.

MANTEL, N. (1967), "Assumption-Free Estimators Using U Statistics and a Relationship to the Jackknife Method," *Biometrics,* 23, 567–571.

QUENOUILLE, M.H. (1956), "Notes on Bias in Estimation," *Biometrika,* 43, 353–360.

RANDLES, R.H., and WOLFE, D.A. (1979), *Introduction to the Theory of Nonparametric Statistics,* New York: John Wiley.

SEN, P.K. (1960), "On Some Convergence Properties of U-Statistics," *Calcutta Statistical Association Bulletin,* 10, 1–18.

——— (1981), *Sequential Nonparametrics,* New York: John Wiley, 80–83.

TUKEY, J.W. (1958), "Bias and Confidence in Not-Quite Large Samples," *Annals of Mathematical Statistics,* 29, 614.

Teaching Singular Distributions to Undergraduates

L. H. KOOPMANS*

Singular distributions are seldom covered in undergraduate probability courses, although they are of interest in statistics and, as is shown by example, can easily arise through extending mixed discrete and continuous distributions to two or more dimensions. A representation is given that makes the construction of a class of singular distributions in two dimensions simple to carry out. This representation is also used to characterize the types of marginal distributions that members of this class can have.

KEY WORDS: Singular distributions; Mixtures of distributions.

*L. H. Koopmans is Professor, Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131.

1. INTRODUCTION

While preparing an exam for an undergraduate course in probability some years ago, I came upon the following simple cumulative distribution function (cdf)

$$F(x,y) = (x+y)/2$$
$$0 \le x \le 1,\ 0 \le y \le 1. \qquad (1)$$

My students were equipped with the usual knowledge about analyzing cdf's: The probability function of a discrete distribution, $p(x,y)$, can be computed by a standard method at the jump points of $F(x,y)$, while the probability density function $f(x,y)$ is computed by taking the mixed second-order partial derivative of $F(x,y)$. I had even discussed mixed distributions for which both components can be present (i.e., nonzero) at the same time.

Now, the problem with the cdf of (1) is that both $p(x,y)$ and $f(x,y)$ are easily seen to be identically 0, even though the cdf obviously represents a completely proper probability distribution. At this point in an exam, it is not unreasonable to expect the normal undergraduate to panic. While I didn't have the courage to include the analysis of this cdf in the exam, it did become the basis of some subsequent lectures (and a short note (Koopmans 1969)).

The student of probability who is familiar with the representation of probability distributions for random variables via the Lebesgue decomposition theorem (see, e.g., McCord and Maroney 1964, p. 71) will recall that there are *three* pure distributional types, not just the two mentioned previously. It follows, by process of elimination, that the given cdf must be of the third type; that is, it is an example of a pure singular distribution.

The usual method of teaching the probability distributions of random variables guarantees that singular distributions will not ordinarily come up. We begin with physically interesting one-dimensional models that are all of pure type, either discrete or continuous. Then by a process of extension, multidimensional distributions are constructed that are also of pure discrete or continuous type. The most popular extension uses statistical independence, which guarantees that if the marginals are all discrete or all continuous, then the joint distributions will be of the same type. Thus, singular distributions tend not to arise in the usual course and are simply not mentioned—with, perhaps, the one exception of the singular normal distribution.

This discussion suggests that for singular distributions to come up in a probability course one would have to begin with one-dimensional singular distributions. This is by no means the case. In fact, it will be seen in Example 1 that the unwary instructor who has the courage to present mixed discrete and continuous one-dimensional distributions, but is not willing to tackle singular distributions, is only one step from pedagogical disaster in an extension to two dimensions.

Why is it desirable to treat singular distributions in an undergraduate course? For the instructor with a mathematical soul there is the satisfaction of being able to tell students that they have now covered every possible form of probability distribution for random variables. Every such distribution must be a mixture of the three basic types. A second reason is that once the basic tools of discrete and continuous marginal and conditional distributions for two random variables have been introduced, essentially nothing more is needed to discuss an extensive class of singular distributions in two dimensions. This will be seen when the representation for singular distributions is given in Section 3.

Part of the bargain in using this representation is that we can easily characterize the types of marginals that singular distributions can have. This will bring home the fact that you can't always tell the type of a joint distribution from its marginals.

Finally, whereas singular distributions are difficult (if not impossible) to visualize in one dimension, being distributions concentrated on uncountable zero-dimensional sets (so to speak), in two dimensions they are, or can be constructed to be, concentrated on rather familiar one-dimensional figures such as lines and circles. The representation in Section 3 will enable students to construct singular distributions on one-dimensional figures of their choice. All of this helps to bring home what multidimensional probability distributions can be like and provides an opportunity for some additional practice with marginal and conditional distributions. Some examples are now given to illustrate these points.

2. A SIMPLE EXAMPLE IN WHICH A SINGULAR DISTRIBUTION ARISES

I should reiterate that the representation given in Section 3 is not for all singular distributions, but is only for those whose marginals are mixtures of discrete and continuous distributions of the form

$$F(x) = F_c(x) + F_d(x), \tag{2}$$

where $F_c(x)$ is the (absolutely) continuous component and $F_d(x)$ the discrete component. I prefer to use this form for mixtures because it is symbolically clean and convenient. However, it does have the pedagogical disadvantage that the components are generally improper distributions; that is, they have total probabilities less than 1. For those who prefer the use of proper distributions with the total probabilities appearing as component coefficients, it is a simple matter to make this change.

All examples, except for the one dealing with the singular normal distribution, are for distributions with positive probability only on the unit square

$$\{(x,y)\colon 0 \le x \le 1,\ 0 \le y \le 1\}.$$

Thus, the marginal distributions are concentrated on the unit interval from 0 to 1. For this reason, the total probabilities contained in the continuous and discrete components of (2) are $F_c(1)$ and $F_d(1)$, respectively. To save space, distribution specifics are given only for arguments in the unit square; the standard extension to the entire x–y plane is implicitly assumed. The fact that $F(x,y) = 0$ when either x or y is less than 0 is important in some of the examples, such as the next one.

Example 1. Two identical electronic devices that generate uniform random numbers between 0 and 1 are available to be issued to a statistician who is carrying out a sampling survey in the field. An equipment manager selects one of them at random for issue. Unfortunately, one of the machines is broken and always gives the value 0. A device is (independently) issued on two successive days and is used by the statistician to obtain one number each day. Let X represent the number obtained on the first day and Y the number obtained on the second. What is the joint distribution of X and Y?

We assume the statistician is not aware that one of the devices is broken and wouldn't be bothered by seeing a 0 on day one. In this case, it is reasonable to take X and

Y to be independent and identically distributed. With probability $\frac{1}{2}$ the statistician will use a device for which the distribution of values is uniform and with probability $\frac{1}{2}$ the distribution will have probability 1 of giving 0. Thus, the cdf is of the form (2) with $F_c(x) = x/2$ and $F_d(x) = I_0(x)/2$, where the symbol $I_r(x)$ represents the function that is 0 for $x < r$ and 1 for $x \geq r$.

The density for this distribution is $f(x) = \frac{1}{2}$ and the probability function $p(x)$ equals $\frac{1}{2}$ for $x = 0$ and is 0 elsewhere.

Now, by using the independence of the variables and doing some rearranging of terms, the joint cdf of X and Y can be written in the form

$$\begin{aligned} F(x,y) &= F(x)F(y) \\ &= F_c(x)F_c(y) + F_d(x)F_d(y) \\ &\quad + [F_d(x)F_c(y) + F_c(x)F_d(y)]. \end{aligned} \tag{3}$$

The Lebesgue decomposition theorem for two-dimensional cdf's can be written as

$$F(x,y) = F_c(x,y) + F_d(x,y) + F_s(x,y). \tag{4}$$

It is easy to verify that the three terms of the expansion (4) are, term for term, the corresponding three components in (3). In particular, because $F_d(x) = \frac{1}{2}$ for all $x \geq 0$, the singular term reduces to $F_s(x,y) = (x+y)/4$.

Note that $F_s(1,1) = \frac{1}{2}$. It follows that half of the total joint probability is tied up in the singular component, which, if normalized to have total probability 1, would be exactly the singular distribution (1).

The other half of the probability is evenly split between a uniform distribution on the unit square and a point mass at $(0,0)$. Physically, the continuous term corresponds to the statistician getting the good device on both days. The discrete term corresponds to using the bad device both days. Thus, the singular term applies to those situations in which the good device was issued on one day and the bad one on the other. This suggests that the singular component consists of a uniform distribution on the vertical line for which $x = 0$ and another on the horizontal line for which $y = 0$. We will confirm this conjecture after the representation of the singular component is given.

3. REPRESENTATION OF $F_s(x,y)$

Let $p(x)$ and $f(x)$ denote the probability (mass) and probability density function of X, respectively, and let X_d denote the set of x's for which $p(x) > 0$. The cdf of the conditional distribution of Y given X is denoted by $F(y|x)$. This distribution is also a mixture of discrete and continuous distributions $F_d(y|x)$ and $F_c(y|x)$. The associated conditional probability function and probability density are denoted by $p(y|x)$ and $f(y|x)$. The set of y values, $Y_d(x)$, for which $p(y|x) > 0$ are especially important in what follows.

Write $F(x,y) = \int_{-\infty}^{x} F(y|x')F(dx')$. Now, expand both the marginal distribution and the conditional distribution in this expression into the sums of their discrete and continuous components. Collect the resulting four terms into three terms exactly as was done in (3). Then, make the same identification of these terms with the components of the Lebesgue decomposition (4). The discrete and continuous components have familiar forms:

$$F_c(x,y) = \int_{-\infty}^{x} F_c(y|x')f(x')dx',$$

and

$$F_d(x,y) = \sum_{\substack{x' \leq x \\ x' \in X_d}} F_d(y|x')p(x').$$

However, the *singular component representation* we are after comes from the third term:

$$F_s(x,y) = \sum_{\substack{x' \leq x \\ x' \in X_d}} F_c(y|x')p(x') + \int_{-\infty}^{x} F_d(y|x')f(x')dx'. \tag{5}$$

For all three terms,

$$F_c(y|x') = \int_{-\infty}^{y} f(y'|x')dy',$$

and

$$F_d(y|x') = \sum_{\substack{y' \leq y \\ y' \in Y_d(x')}} p(y'|x'). \tag{6}$$

The important thing to note is that, in the singular component (5), the *continuous* conditional distribution is involved with the *discrete* marginal, while the *discrete* conditional distribution is integrated with respect to the *continuous* marginal density.

As an illustration of the use of the representation (5) we now verify that the singular distribution (1) consists of uniform distributions on the lines $x = 0$ and $y = 0$ as conjectured in Example 1.

Example 2. Set $y = 1$ in Expression 1. The resulting function, $F(x) = (1+x)/2$ is the X-marginal cdf of the desired distribution. Note that it puts probability $\frac{1}{2}$ at $x = 0$ and a uniform density on the rest of the line. That is, $X_d = \{0\}$, $p(0) = \frac{1}{2}$, and $f(x) = \frac{1}{2}$. For the singular distribution representation (5), the continuous conditional density is defined only on X_d. We hypothesize that the mass at the point 0 represents the projection of the uniform probability along the line $x = 0$ onto the x axis. So we try the conditional density $f(y|0) = 1$. The uniform distribution on the x axis is guaranteed by the marginal density $f(x) = \frac{1}{2}$, and we keep it there for the joint distribution by defining the conditional probability function to be $p(0|x) = 1$ for all x.

It follows that $Y_d(x) = \{0\}$; that is, this set is independent of x. This is seen later to be the key property distinguishing singular distributions for independent variables from those for dependent variables.

Now, apply Expressions (6) to obtain $F_c(y|0) = y$ and

$F_d(y|x) = I_0(y)$. Then, an application of (5) gives

$$F(x,y) = \tfrac{1}{2}y + \int_0^x 1 \cdot \tfrac{1}{2} dx'$$
$$= (y + x)/2$$

as was to be shown.

In the following example, X and Y are dependent variables.

Example 3. Uniform Distribution on the Line $y = x$. This is a fairly well-known distribution with cdf $F(x, y) = \text{minimum}(x, y)$. The sensible construction takes the uniform density $f(x) = 1$ for X and the conditional probability function that puts probability 1 at $y = x$ for each x: $Y_d(x) = \{x\}$ and $p(x|x) = 1$. Since X has no discrete component, $f(y|x)$ need not be defined.

From (6) we get $F_d(y|x) = I_x(y)$. Now if $y < x$ then $I_{x'}(y) = 1$ for x' between 0 and y and is 0 beyond y. Thus,

$$F(x,y) = \int_0^x I_{x'}(y) dx' = y.$$

However, if $y > x$, since the integrand is 1 from 0 to y, this integral yields $F(x, y) = x$. Thus, $F(x, y) = \text{minimum}(x, y)$ as stated.

Example 3 shows that you can't always tell the distribution type of a two- (or higher-) dimensional distribution from its marginals. Both marginals are uniform distributions, thus they are pure continuous, whereas the joint distribution is pure singular.

Example 4. A Distribution on a Circle. This example is primarily an exercise in integration that puts all of the probability of a singular distribution on the circle $(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 = (\frac{1}{2})^2$.

Let the X marginal be the uniform distribution and let $Y_d(x)$ consist of the two points $\frac{1}{2} \pm (\frac{1}{4} - (x - \frac{1}{2})^2)^{1/2}$. Put discrete conditional probability $\frac{1}{2}$ on each point. The joint cdf is a bit messy, but it is reasonably straightforward to evaluate the Y marginal cdf $F(y) = F(1, y)$ to be

$$F(y) = \sqrt{\tfrac{1}{4} - (y - \tfrac{1}{2})^2} \qquad y < \tfrac{1}{2}$$
$$= 1 - \sqrt{\tfrac{1}{4} - (y - \tfrac{1}{2})^2} \qquad y \geq \tfrac{1}{2}.$$

Example 5. Verify that the cdf of the singular distribution that puts a uniform distribution on the line $x + y = 1$ is

$$F(x, y) = (x + y - 1)I_1(x + y).$$

Example 6. The Singular Standard Normal Distribution. Let $\phi(x)$ denote the standard normal density. Put conditional probability 1 on the regression line $y = x$ if the correlation coefficient $\rho = 1$ and on the line $y = -x$ if $\rho = -1$. If Φ denotes the standard normal cdf, verify that

$$F(x, y) = \Phi(\text{minimum}(x, y)) \text{ if } \rho = 1$$

and

$$F(x, y) = [\Phi(x) - \Phi(-y)]I_0(x + y) \text{ if } \rho = -1.$$

Show that the Y marginal is $\phi(y)$ in both cases. Note the similarity between this example and Examples 2 and 4.

4. CHARACTERIZATION OF THE POSSIBLE MARGINAL TYPES

A straightforward application of the criterion for uniform convergence for infinite series, given for example in Titchmarsch (1939, p. 5), shows that the first term of (5) is always (absolutely) continuous. Thus, if the X marginal is pure discrete, the second term in (5) is 0 and the Y marginal is necessarily continuous. Thus, even when X and Y are dependent random variables, the discrete-discrete case for the two marginals can only arise from the discrete term in the Lebesgue decomposition (4).

If the X marginal is of mixed type, then mixed, discrete, and continuous marginals are all possible for Y. Example 1 is an example of the mixed-mixed case. Examples 2, 3, and 4 could have been applied to the continuous component of a mixed marginal distribution for X, leading to a pure continuous marginal for Y.

Note that the continuity of the second term in (5) depends on the fact that the set $Y_d(x)$ of jump points of $p(y|x)$ can vary with x. Suppose that X is pure continuous or is of mixed type with pure discrete conditional distribution for x in X_d, then, only the second term is present in (5). By allowing $Y_d(x)$ to (a) vary on a set of X probability 1, (b) vary on a set with X probability between 0 and 1 and be constant but nonempty for a second set of positive X probability, (c) be constant on a set of X probability 1, examples can be produced for which the Y marginal is (a) continuous, (b) mixed, and (c) discrete, respectively.

In summary, if X and Y are dependent and have a joint singular distribution represented by (5), then the possible marginal-type combinations are (i) discrete-continuous, (ii) mixed-discrete, (iii) mixed-mixed, (iv) mixed-continuous, and (v) continuous-continuous.

If X and Y are independent, then $Y_d(x)$ can't vary with x, and in fact (5) reduces to the last term in (3). It follows that the marginals must be either both mixed, or the marginal pairs must be mixed-continuous, mixed-discrete, or of opposite pure types; the continuous-continuous case is now excluded. This explains why singular distributions with continuous marginals, like the singular normal, do not arise in the usual construction of joint distributions by independence.

[*Received January 1983. Revised April 1983.*]

REFERENCES

KOOPMANS, L.H. (1969), "Some Simple Singular and Mixed Probability Distributions," *The American Mathematical Monthly*, 76, 297–300.

McCORD, JAMES R., III, and MARONEY, RICHARD M., JR. (1964), *Introduction to Probability Theory*, New York: Macmillan.

TITCHMARSH, E.C. (1939), *The Theory of Functions* (2nd ed.), Oxford, England: Oxford Press.