1970

A note on a compound distribution: the demand for hospital beds

John Heineke
Santa Clara University, jheineke@scu.edu

A NOTE ON A COMPOUND DISTRIBUTION: THE DEMAND FOR HOSPITAL BEDS

J. M. Heineke

Dr. J. M. Heineke is Assistant Professor Economics at the University of Santa Clara. He received a Ph. D. in Economics from the University of Iowa in 1968. While he was a graduate student he published two articles in *The American Economist.* Subsequent to joining the University of Santa Clara, an article, "The Demand for Refined Lead" appeared in the *Review of Economics and Statistics.*

I

There has been growing interest in the past few years in the economics of medical services. Due to the nature of the medical "industry" a number of problems heretofore considered to be of minor importance have assumed major proportions. It is the purpose of this paper to present a model for predicting the demand for hospital beds, which is free from a rather serious specification error. This error arises from a failure to recognize the nature of the underlying stochastic mechanism and is by no means unique to the demand for hospital beds. Indeed, the demand for hospital beds is merely the vehicle used in explicating, what seems to be, an often overlooked source of error.

Consider the random variable $X(t)$ - the total number of bed days demanded in a time interval of length t. If one knew or could estimate the distribution of $X(t)$, say $f[X(t)]$, one could make statements concerning the likelihood that total bed day demand would be within any set of reals. It is not the purpose of this paper to discuss the various methods available for estimating $f[X(t)]$, but instead to derive, under a quite plausible set of assumptions, the distribution of $X(t)$. It will be shown, for example, that if one were to estimate $\text{Var}[X(t)]$ on the basis of past realizations of $X(t)$ the result could be a grossly inaccurate assessment of the amount of uncertainty in the $X(t)$ process.

II

Let us examine more closely the nature of the events which give rise to the random variable $X(t)$. The basic event underlying $X(t)$ is the *number* of demands for hospital beds in a period of length t. Since demands arise from a vast number of causes it seems quite reasonable to assume that the number of demands in any two periods are independent. Since the length of the period, t, is arbitrary it may be adjusted to insure that the probability of more than one demand in a period is negligible. Therefore the *number*, N, of demands during a time period of length t has a Poisson distribution with parameter rt. (r is a factor of proportionality.)

But we still haven't completely isolated the stochastic mechanism which gives rise to $X(t)$. *Each* demand for a hospital bed is accompanied by a second random variable, Y - the length of time in days the bed will be occupied. Let $g(y)$ be the probability distribution of Y. Then

$$(1) \qquad X(t) = \sum_{i=1}^{N} Y_i,$$

where Y_i is the number of days the i^{th} demander will occupy a bed, and

$$(2) \qquad f[X(t)] = P[X(t) = b] = \sum_{n=0}^{\infty} P(N = n) \cdot P\left[\sum_{i=1}^{n} Y_i = b\right], \; ^{(1)}$$

using the fundamental formula for conditional probability. Since there is no reason to suppose that the number of days the i^{th} demander occupies a bed in any way affects the length of demander j's period in a bed, $X(t)$ is the sum of N independent random variables with common probability distribution $g(y)$.

Therefore for any fixed n, the distribution of $\sum_{1}^{n} Y_i$ [2] is given by the n fold convolution of $g(y)$ with itself. [3]

Call this distribution $g^n(x)$. [4] The distribution of $X(t)$ is then

$$(2') \qquad f[X(t)] = \sum_{n=0}^{\infty} \frac{e^{-rt}(rt)^n}{n!} g^n(x), \; x = 0, 1, 2, \ldots \text{ zero elsewhere.}$$

(1) Capital letters denote random variables and small letters specific realizations.

(2) Note that this random variable is not $X(t)$ since the variable $\sum_{1}^{n} Y_i$ is conditioned on $N = n$.

(3) See Feller [2].

(4) $g^n(x) = P[\sum_{1}^{n} Y_i = x]$.

Distributions of this type are characteristic of a large group of phenomena, e. g., Feller [2] lets N be the number of lighting strikes in a certain space interval and Y be the damage caused by an individual strike, while Birch and Heineke [1] take N as the number of reserve losses in a period of length t and Y as the size of an individual loss, making X(t) the loss of reserves in a period of length t.

According to equation (1) X(t) is the sum of a random number of independent random variables. In this case it is well known that

(3) $E[X(t)] = E(N) \cdot E(Y)$ and

(4) $Var[X(t)] = E(N) \cdot Var(Y) + Var(N) \cdot E^2(Y).$[5]

If one did not recognize the "compound" nature of the variable X(t) and elected to estimate E[X(t)] and Var[X(t)] from a set of observations on X(t) extremely misleading results could follow. Consider the following example:[6]

Period	N	Y	X(t)
		7	
1	4	1	20
		3	
		9	
		13	20
2	2	7	
3	1	20	20
		3	
4	3	10	20
		7	
5	2	12	20
		8	
		2	
		1	
6	5	6	20
		4	
		7	
Totals	17	120	120
Means	2.83	7.06	20
Variance	2.17	23.93	0

From equation (4), $Var[X(t)] = 175.87$

If one calculated the variance of X(t) directly from the six values in the table without recognizing the compound nature of X(t), one would conclude Var [X(t)] = 0 when in fact Var [X(t)] = 175.87. That is, the "direct" calculation drastically understates the amount of uncertainty in the process. Even worse, another set of data could lead to the opposite result, viz. that the amount of uncertainty is heavily overstated when the compound nature of X(t) is overlooked.

(5) See Parzen [4], pp. 55 - 56.

(6) This example is from Birch and Heineke [1], p. 25.

To use equation (2′) as a model for predicting the demand for hospital beds, two approaches are possible. First, one could estimate the distribution of Y, the number of days an individual demander will occupy a bed, and then compute $g^n(x)$ from this distribution. The second alternative is of considerable practical importance. Note that the time period, t, in the distribution of N can be adjusted to make the parameter rt equal to t. Let q_1 and q_2 be the first two moments of g(y), then by equations (3) and (4) $E[X(t)] = q_1 t$ and $Var[X(t)] = q_2 t$, and the variable $[X(t) - q_1 t] / (q_2 t)^{1/2}$ is asymptotically normal. The importance of this result lies in the fact that g(y) can be observed and q_1 and q_2 can be estimated. Note that this appeal to the Central Limit Theorem and the asymtotic normality of the above variable is justified from the nature of the basic event and *is not the same* as assuming X(t) is a normal random variable.

REFERENCES

[1] E. M. Birch and J. M. Heineke, "Stochastic Reserve Losses", *The American Economist*, Vol. XI, no. 1, 1967.

[2] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I, J. Wiley and Sons, 1966.

[3] P. Kahn, "An Introduction to Collective Risk Theory and Its Application to Stop-Loss Reinsurance", in *Society of Actuaries Transactions*, Vol. XIV, Part 1, 1962.

[4] E. Parzen, *Stochastic Processes*, Holden and Day, 1962.