# A NOTE ON THE STOCHASTIC INDEPENDENCE OF FUNCTIONS OF ORDER STATISTICS

By Gerald S. Rogers

*University of Arizona*

The theorem presented below appears to be useful in determining whether certain functions of order statistics are stochastically independent (briefly, independent). The following result has appeared in the literature in various forms, eg. [1]; the statement here is the particular form used in the proof of the second part of the theorem.

LEMMA: *Let s be a complete sufficient statistic for a family of probability density functions indexed by a parameter $\theta$. Let t be any other statistic, not a function of s alone. Then s and t are independent if and only if the distribution of t does not depend on $\theta$.*

THEOREM: *Let x be a real random variable with distribution function* $F(x) = \int_{-\infty}^{x} f(X)\, dX$, *where* $f(x)$ *is a non-degenerate probability density function* (pdf). *Let* $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ *be the order statistics based on a random sample of size* $n \geqq 2$ *from this x distribution. Let* $z = z(x_1, \cdots, x_j)$ *be a statistic based on the first* $j < n$ *order statistics only. Then the following two statements are equivalent*:

(a) $z$ *is independent of* $x_k$ *for some* $k \geqq j$;

(b) $z$ *is independent of the set* $\{x_k : j \leqq k \leqq n\}$.

PROOF: Notation—let $g(A)[g(A|C)]$ denote the ordinary [conditional] pdf of $A$ [given $C$]. To show that (a) implies (b), first suppose that in (a), $k = j$. It follows directly from the definition of conditional pdf's that

$$g(x_1, \cdots, x_j | x_j) = g(x_1, \cdots, x_j | x_j, \cdots, x_n),$$

and hence that

$$g(z|x_j) = g(z|x_j, \cdots, x_n).$$

Under the hypothesis (a), $g(z|x_j) = g(z)$, and therefore, $g(z|x_j, \cdots, x_n) = g(z)$.[1] Thus, $z$ is independent of the set in (b).

Now suppose that in (a), $k > j$. Then, (as is readily shown by direct computation), in the conditional pdf $g(x_1, \cdots, x_{k-1}|x_k)$, $x_k$ may be considered as a "parameter" for which the conditional random variable $x_{k-1}$ given $x_k$, written $(x_{k-1}|x_k)$, is a "complete sufficient statistic." Under the hypothesis (a), $g(z|x_k) = g(z)$, so that the distribution of $z$ given $x_k$ actually does not depend upon the "parameter" $x_k$. Therefore, by the lemma, $(z|x_k)$ and $(x_{k-1}|x_k)$ are independent. In terms of the pdf's,

$$g(z, x_{k-1}|x_k) = g(z|x_k)g(x_{k-1}|x_k).$$

Since $g(z|x_k) = g(z)$, $g(z, x_{k-1}, x_k) = g(z)g(x_{k-1}, x_k)$, and hence, $g(z, x_{k-1}) = g(z)g(x_{k-1})$.

Received July 15, 1958; revised July 6, 1959.

[1] One referee pointed out that this is well known in the theory of Markov Chains.



Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org