The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 377-378

SUBSTITUTION IN CONDITIONAL EXPECTATION

BY R. R. BAHADUR[1] AND P. J. BICKEL[2]

University of Chicago and University of California, Berkeley

Let X be a sample space of points x and P a probability measure on a σ-algebra $\mathcal{A}$ of sets of X. Let Y be a space of points y, $\mathcal{B}$ a σ-algebra of sets of Y, and $T:X \to Y$ a function such that B in $\mathcal{B}$ implies $T^{-1}B$ in $\mathcal{A}$. Let $f(x, y)$ be a real valued $\mathcal{A} \times \mathcal{B}$-measurable function on $X \times Y$, and consider the conditional expectation of $f(x, T(x))$ given $T(x) = y$. It is natural to presume that this equals the conditional expectation, with y held fixed, of $f(x, y)$ given $T(x) = y$. This note points out that the presumption is essentially correct. In the final paragraph of the note we show, as an application, that a regular conditional probability measure automatically assigns probability one to the set specified by the condition.

Since in the general case a conditional expectation given $T(x) = y$ is not quite uniquely determined as a function on Y, we must first restate the present issue more precisely, as follows: Suppose for simplicity that f is non-negative. Let $g(y, \eta) \geqq 0$ be a function on $Y \times Y$ such that (i) for each fixed η in Y, $g(y, \eta)$ is $\mathcal{B}$-measurable in y and serves as the conditional expectation of $f(x, \eta)$ given $T(x) = y$, i.e., $\int_{T^{-1}B} g(T(x), \eta)\, dP = \int_{T^{-1}B} f(x, \eta)\, dP$ for all B in $\mathcal{B}$. Then is it true that (ii) $g(y, y)$ is $\mathcal{B}$-measurable in y and serves as the conditional expectation of $f(x, T(x))$ given $T(x) = y$? In general, for an unfortunate choice of g, (i) does not imply (ii). Suppose, for example, that each one-point subset $\{\eta\}$ of Y is $\mathcal{B}$-measurable and of induced measure zero, that $f \equiv 0$, and that $g(y, \eta)$ is the indicator of the set $\{(y, \eta): y = \eta\}$, i.e., $g = 1$ if $y = \eta$ and $g = 0$ otherwise. Then (i) holds but (ii) does not. It is true, however, that there always exists a suitable g, i.e., a g which satisfies both (i) and (ii). The existence of a suitable g (i.e., the essential validity of substitution) is well known in certain special cases, e.g., $f(x, y) \equiv f_1(x) \cdot f_2(y)$, or $f(x, T(x)) \equiv \varphi(U(x), T(x))$ where U and T are independently distributed.

To establish the existence of a suitable g in the general case, let $\mathcal{F}$ be the class of all non-negative $\mathcal{A} \times \mathcal{B}$-measurable functions f on $X \times Y$, and let $\mathcal{F}_0$ be the class of all f in $\mathcal{F}$ for which a suitable g exists. According to one of the special cases mentioned above, $\mathcal{F}_0$ includes all indicator functions of sets $A \times B$ with A in $\mathcal{A}$ and B in $\mathcal{B}$; for such an $f = I_{A \times B}(x, y)$, $g(y, \eta) = h(y) \cdot I_B(\eta)$ satisfies (i) and (ii), where $h \geqq 0$ is any $\mathcal{B}$-measurable function which serves as the con-

Received 21 August 1967.

[1] Research supported in part by Research Grant No. NSF GP 6562 from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation, and in part by the Statistics Branch, Office of the Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Research supported by Research Grant No. NSF GP 5059 from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation.

ditional probability of A given $T(x) = y$. It now follows by approximation that $\mathfrak{F}_0 = \mathfrak{F}$. The approximation required here is parallel to the approximation in proofs of Fubini's theorem.

The argument outlined above yields only the existence of a suitable g for the given f. Constructive definitions of a suitable g can be given in certain cases. Suppose, for example, that x admits a regular conditional probability measure given y, i.e., there exists a function $Q(A, y)$ on $\mathcal{A} \times Y$ such that $Q(\cdot, \eta)$ is a probability measure on $\mathcal{A}$ for each η, and such that, for each A, $Q(A, y)$ is $\mathcal{B}$-measurable in y and serves as the conditional probability of A given $T(x) = y$. (According to Doob, such a Q always exists if x represents a finite or denumerably infinite set of real valued random variables.) In this case $g(y, \eta) = \int_X f(x, \eta)$ $Q(dx, y)$ satisfies (i) and (ii). The proof consists in verifying first that the formula produces a suitable g for $f(x, y) \equiv I_A(x) \cdot I_B(y)$ and then showing by approximation that the formula is successful for any non-negative $\mathcal{A} \times \mathcal{B}$-measurable f.

By way of an application of substitution, suppose that $C = \{(x, y): T(x) = y\}$ is $\mathcal{A} \times \mathcal{B}$-measurable. (This holds, in particular, if y represents a finite or denumerably infinite set of real valued random variables). Then $T^{-1}\{\eta\}$ is $\mathcal{A}$-measurable for each η in Y, and a regular conditional probability measure Q necessarily satisfies the natural condition $Q(T^{-1}\{y\}, y) = 1$ for almost all y in Y. To see this, let f be the indicator of the set C. Then, by the preceding paragraph, $Q(T^{-1}\{\eta\}, y)$ is a suitable g for this f. Hence $Q(T^{-1}\{y\}, y)$ serves as the conditional expectation of $I_C(x, T(x)) \equiv 1$ given $T(x) = y$; hence $Q(T^{-1}\{y\}, y) = 1$ for almost all y. Similar conclusions have been established in special settings by other methods in [1], [3], [4]. It is shown in [2] that in general there does not exist a Q such that $Q(T^{-1}\{y\}, y) = 1$ for *all* y.

REFERENCES

[1] Blackwell, D. (1955). On a class of probability spaces. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **2** 1–6. Univ. of California Press.

[2] Blackwell, D. and Ryll-Nardzewski, C. (1963). Non-existence of everywhere proper conditional distributions. *Ann. Math. Statist.* **34** 223–225.

[3] Parthasarathy, K. R., (1967). *Probability Measures on Metric Spaces*, Academic Press, New York, 146–150.

[4] Varadarajan, V. S. (1962). Lecture notes on "Special topics in probability", Courant Institute of Mathematical Sciences, New York Univ. (Unpublished).