$(Q_2, Q_3, \cdots, Q_m)$, $(Q_3, Q_4, \cdots, Q_m)$, etc., we finally obtain

$$\Pr\{T_{j1} \leqq \beta_1, T_{j2} \leqq \beta_2, \cdots, T_{jm} \leqq \beta_m\} = \prod_{k=1}^{m} \Pr\{T_{jk} \leqq \beta_k\}$$

as was to be shown.

Finally, it should be mentioned that if the waiting times are defined so as *not* to include the service times, that is, as the quantities $T_{jk} - s_{jk}$, the question of mutual independence of these quantities for $k = 1, 2, \cdots, m$ is apparently an open problem.

REFERENCES

[1] Reich, Edgar (1957). Waiting times when queues are in tandem. *Ann. Math. Statist.* **28** 768–773.

[2] Saaty, Thomas L. (1961). *Elements of Queueing Theory*. McGraw-Hill, New York.

[3] Takács, Lajos (1962). *Introduction to the Theory of Queues*. Oxford Univ. Press, New York.

A NOTE ON THE RE-USE OF SAMPLES[1]

By David R. Brillinger[2]

London School of Economics and Political Science

There are situations in statistical estimation in which the basic underlying distribution is invariant under some family of transformations. In this note a theorem similar to the Blackwell-Rao Theorem is proved demonstrating that this additional structure can sometimes be exploited to improve an estimator.

Theorem. *Consider a random variable x, sample space X, σ-algebra $\mathfrak{X}$, probability measure $P(\ )$. Suppose that G is a set of measure-preserving transformations for the measure P,* i.e. *$P(gA) = P(A)$ for all A in $\mathfrak{X}$, g in G. Let $\mu(\ )$ be a measure of total mass* 1, *defined on a σ-algebra $\mathcal{G}$ of subsets of G. Let $\phi(x)$ be an estimator such that $\phi(gx)$ is $\mathcal{G} \times \mathfrak{X}$ measurable.*

(i) If $\phi(x)$ is an unbiased estimator of θ then,

$$\gamma(x) = \int_G \phi(gx)\, d\mu(g)$$

is also an unbiased estimator of θ.

(ii) If $\phi(x)$ takes values in a k-dimensional space and has an associated real-valued, convex, bounded from below loss function $W[\phi(x)]$, such that $W[\phi(gx)]$ is $\mathcal{G} \times \mathfrak{X}$ measurable then, $R_\phi \geqq R_\gamma$ where R is the associated risk function, and in particular the ellipsoid of concentration of γ is everywhere

Received May 15, 1962.

[1] Part of this research was carried out while the author had the support of a Research Training Fellowship of the Social Science Research Council.

[2] Presently at Bell Telephone Laboratories, Murray Hill, N. J. and Princeton University.

contained in the ellipsoid of concentration of ϕ. (ϕ is not necessarily an unbiased estimator in this part of the theorem.)

PROOF.

(i) This part of the theorem follows trivially from Fubini's Theorem.

(ii) It must be shown that $\int_X W[\phi(x)]\, dP(x) \geqq \int_X W[\gamma(x)]\, dP(x)$.

Take the following randomized procedure: choose a g according to μ, then use the estimator $\phi(gx)$. This procedure has risk $\int\int W[\phi(gx)]\, dP(x)\, d\mu(g)$ or R_ϕ using Fubini's Theorem once again. The result of the theorem now follows from Lemma 3.1 in [1].

If W is strictly convex one easily sees that there will be equality of risks if and only if,

$$\phi(x) = \int_G \phi(gx)\, d\mu(g).$$

(If the measure μ is a Haar measure on a locally compact group this implies that ϕ is invariant under G.)

EXAMPLES. In applications the measure μ on G will most likely be a discrete measure or a Haar measure.

The theorems concerning the unbiasedness and smaller variance properties of Hoeffding's U statistics may be derived from the theorem proved above. In this case the probability distribution is invariant under the symmetric group and the (Haar) measure attached to each group element is $1/\binom{n}{m}$ for an estimator of degree m. Note that in fact one need not have independent observations from the same distribution, but need only have a realisation of an n-dimensional symmetric distribution.

The broadest class of techniques to which it appears useful to try to apply the theorem are those of Monte Carlo. Here the theorem states in effect that under certain conditions it is possible to use the same randomly generated sample more than once. This is important if samples from the distribution under consideration are expensive to come by relative to the calculations being carried out on the samples. For example, suppose one wishes to derive the average of some statistic from samples of size n from an $N(0, 1)$ distribution by means of Monte Carlo. The method suggested by the above theorem is as follows: draw a sample of size n from $N(0, 1)$. Apply k orthogonal transformations to this sample obtaining k further "samples." Calculate the statistic for each of the $k + 1$ samples and average the $k + 1$ values so obtained. If the statistic is not invariant under orthogonal transformations the result will have a smaller variance than that that would have been obtained by using the original sample alone. If this method were taken to the extreme, one sample would be drawn and all orthogonal transformations would be averaged out by means of say the Haar measure on the orthogonal group. If the average did not depend on the original sample it would have zero variance.

In [2] it is suggested that when one is estimating $E[\phi(x)]$, x uniform, instead of using simply $\phi(x)$, x a single observation from the uniform distribution, one should use,

$$\{\phi(x) + \phi(1 - x) + \phi(y)\}/3$$

where

$$y = x \qquad 0 \leqq x \leqq \tfrac{1}{2}$$

$$= \tfrac{3}{2} - x \qquad \tfrac{1}{2} < x \leqq 1$$

for example. The reduced variance property of this estimate is a result of the above theorem. G consists of the identity transformation, the transformation $x \rightarrow 1 - x$, and the transformation $x \rightarrow y$. Each of these transformations then has weight $\frac{1}{3}$.

REFERENCES

[1] Hodges, J. L., Jr. and Lehmann, E. L. (1950). Some problems in minimax point estimation. *Ann. Math. Statist.* **21** 182–197.
[2] Hammersley, J. M. and Morton, K. W. (1956). A new Monte Carlo technique, antithetic variates. *Proc. Cambridge Philos. Soc.* **52** 449–475.

ON STOCHASTIC APPROXIMATIONS

By Shmuel Friedman

The Hebrew University, Jerusalem

0. Summary. The procedure of stochastic approximations suggested by Robbins-Monro [1], for reaching a zero point x_0 of a regression function, was shown by Dvoretzky [4], to be a convergent w. p. 1. and in mean square under certain conditions. In this paper we deal with two problems of modifying the process to acquire convergence under weaker conditions.

1. Introduction. Let $H(y/x)$ be a family of distribution functions, which correspond to the parameter x.

Let us write: $m(x) = \int y\, dH(y/x)$; $\sigma^2(x) = \int (y - m(x))^2\, dH(y/x)$.

Let $\{a_n\}$ be a sequence of positive members, such that, $\sum a_n = \infty$, $\sum a_n^2 < \infty$.

Let x_1 be an arbitrary number. The Robbins-Monro process is defined recursively for all n by $x_{n+1} = x_n - a_n y_n$, where y_n is a chance variable with distribution function $H(y/x_n)$. The conditions for its convergence were shown to be:

(1) $$|m(x)| \leqq L\,|x| + K.$$

(2) $$\sigma^2(x) \leqq \sigma^2 < \infty.$$

(3) If $x < x_0$, then $m(x) < 0$, while if $x > x_0$, then $m(x) > 0$.

Received August 28, 1961; revised January 1, 1962.