NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

———————◆———————

NOTE ON THE LAW OF LARGE NUMBERS AND "FAIR" GAMES

By W. Feller

Cornell University

1. "Fair" games. Let $\{X_k\}$ be a sequence of independent random variables with the same cumulative distribution function $V(x)$. Suppose that the expectation

$$(1) \qquad E(X_k) = \int_{-\infty}^{+\infty} x \, dV(x) = M$$

exists, and put

$$(2) \qquad S_n = X_1 + \cdots + X_n .$$

The weak law of large numbers states[1] that for every $\epsilon > 0$ and $n \to \infty$

$$(3) \qquad \Pr \{| \, S_n - nM \, | < \epsilon n\} \to 1.$$

In the picturesque language of the theory of games this means that, after a large number of trials, the accumulated gain S_n will, with great probability, be of the order of magnitude of nM. This led to the definition that a game is "fair" if the entrance fee for each trial is M. Unfortunately this definition creates the erroneous notion that a "fair" game is necessarily fair. To disprove it we shall (section 3) exhibit an example which will show:

(I)· *A game can be "fair" and nevertheless such that the probability tends to one that, after n trials, the player will have sustained a loss $L_n = nM - S_n$ of the order of magnitude $n(\log n)^{-\eta}$, where $\eta > 0$ is arbitrarily small. In other words, in our example*

$$(4) \qquad \Pr \{nM - S_n > (1 - \epsilon)n(\log n)^{-\eta}\} \to 1.$$

Of course, L_n is necessarily of smaller order of magnitude than n; however, our example can be modified in such a way that the ratio of the loss L_n to the accumulated entrance fees nM decreases as slowly as one pleases.

This shows that a "fair" game can be exceedingly disadvantageous. Conversely, an "unfair" game can very well be advantageous. If a careful driver insures his car, the game is clearly "unfair" according to definition, and yet some

---

[1] Usually (3) is proved only under more restrictive hypotheses. Actually the finiteness of $E(X_k)$ implies even the strong law of large numbers; cf. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Berlin 1933), p. 59.

states impose such games on drivers. Now in this and many other practical cases the game is of such a nature that there is a very small probability p of winning a comparatively great amount A; the "fair" price would be pA. In such cases the law of large numbers would be significant only if n is large compared to $1/p$, whereas actually the maximum number of games to be played is comparatively small. Clearly any theory meets practical requirements only if it makes allowance for the number of trials and makes the "fair" price depend on the number of trials.

2. The Petersburg "paradox." For obvious reasons the classical theory of probability was unable to provide a precise formulation of the law of large numbers and to establish the actual conditions of its validity. Often it has been looked upon as a direct consequence of the definition of probability, and this led to the so-called Petersburg paradox which presents no difficulties to the modern theory. It refers to the case where the expectation (1) is infinite. The usual example exhibits a game in which the possible gains in each trial are distributed according to

$$(5) \qquad\qquad \Pr\{X = 2^k\} = 2^{-k}.$$

Here $M = \infty$. Now the law of large numbers (3) used to be proved (if at all) only assuming the existence of moments of higher order. Nevertheless, the classical theory postulated the validity of (3) even for $M = \infty$, and treating ∞ as a number (with $\infty - \infty = 0$) it argued that ∞ is a "fair" price for the game as defined in (5). Great ingenuity was exercized in order to reconcile this result with commonsense.[2] Actually one can pass from (3) to the limit $M \to \infty$, but the *only* result to be arrived at is trivial and could be anticipated without theory: If the player pays for each trial a *fixed* amount A, he is likely to have a positive gain provided he plays sufficiently long, i.e., provided $n > N(A)$, where $N(A)$ itself *increases with A*.

Instead of a paradox we reach the conclusion that the price should depend on n, that is to say vary as the number of trials increases. For best results this should be the case even if M is finite. It should be noticed that in the Petersburg case (5) a variable price can be determined so that a law of large numbers will hold which is in every respect analogous to (3). In this formula nM is simply the accumulated amount of entrance fees; denoting it by P_n, formula (3) takes on the equivalent form

---

[2] Among the latest textbooks, von Mises (*Wahrscheinlichkeitsrechnung*, Leipzig–Wien 1931, p. 108f.) avoids the difficulty by declaring that (5) can not represent a collectif because of its infinite tail. This viewpoint is legitimate, but makes the law of large numbers inapplicable to practically all useful distributions. Fry (*Probability and its Engineering Uses*, New York, 1928, p. 197) says: "The true explanation of the paradox is . . . based upon the fact that in our every-day experience we have to deal only with individuals who have finite fortunes and who would therefore be incapable of paying back the sums which are required . . .". The problem does not seem to be mentioned in Uspensky's book.

(6) $$\Pr \{ |\, S_n - P_n\, | \, < \, \epsilon P_n \} \to 1.$$

It is this interpretation of (3) that leads to the notion of "fair" games. Now the Petersburg game can also be played in a "fair" way:

(II) *Let the player in the Petersburg game* (5) *at the k-th trial pay the amount* $\log_2 k$. *The accumulated entrance fees up to the n-th trial are* $P_n \sim n \log_2 n$, *and the game is "fair" in the sense that the law of large numbers* (6) *holds. This requirement determines the entrance fees essentially uniquely* (that is to say up to terms of smaller order of magnitude which, by definition, remain undetermined).

3. Proofs. Theorems (I) and (II) follow easily from the following

Lemma: *Let* $a_n \to \infty$ *be a sequence of positive numbers; in order that there exist a sequence* $\{b_n\}$ *such that*

(7) $$\Pr \{ |\, S_n - b_n\, | \, < \, \epsilon a_n \} \to 1$$

it is necessary and sufficient that for every $\delta > 0$ *simultaneously*

(8) $$n \int_{|x| > \delta a_n} dV(x) \to 0, \qquad a_n^{-2} n \int_{|x| < a_n} x^2 \, dV(x) \to 0;$$

in this case (8) *will hold with*

(9) $$b_n = \sum_{k=1}^{n} \int_{|x| < a_k} x \, dV(x)$$

(*and, of course, for any other sequence* $\{b_n^*\}$ *if and only if* $|\, b_n^* - b_n\, | = O(a_n)$). This lemma is a simple consequence of the necessary and sufficient conditions for the generalized law of large numbers[4].

To prove theorem (II) we have to determine a sequence $\{a_n\}$ such that (7) will hold for the distribution function defined in (5) and with $b_n \sim a_n$. A simple computation shows that (8) will hold for any sequence $\{a_n\}$ which increases faster than n. Moreover, the sequence $\{b_n\}$ defined by (9) will be of the same order of magnitude as $\{a_n\}$ if, and only if, $a_n \sim n \log_2 n$. This proves (II).

Now let $\eta > 0$ be arbitrary, and define the distribution function $V(x)$ to have a density

(10) $$V'(x) = \frac{\eta}{x^2 \log^{1+\eta} x} \qquad \text{for} \ \ x > e;$$

at $x = 0$ the function $V(x)$ shall have a jump of magnitude

(11) $$1 - \int_e^\infty \frac{\eta \, dx}{x^2 \log^{1+\eta} x} < 1,$$

while $V(x)$ is constant in the intervals $x < 0$ and $0 < x < e$. For this distribution function we have obviously $M = 1$.

---

[3] $\log_2$ stands for the logarithm to the basis 2.

[4] Cf. FELLER, *Acta Univ. Szeged*, Vol. 8 (1937), pp. 191–201.

Next, let for $n > e$

(12) $$a_n = n \log^{-\eta} n.$$

Then (8) holds and from (9) and (10) we obtain easily for large n

(13) $$b_n = \sum_{k=1}^{n} \{1 - \log^{-\eta} a_k\} < n - (1 - \epsilon)a_n .$$

Substituting into (7) one sees that, again for sufficiently large n,

(14) $$\Pr \{S_n - n + (1 - \epsilon)a_n < \epsilon a_n\} \to 1,$$

or, since $M = 1$,

(15) $$\Pr \{S_n - nM < -(1 - 2\epsilon)a_n\} \to 1.$$

This proves (I).

A NOTE ON RANK, MULTICOLLINEARITY AND MULTIPLE REGRESSION[1]

By Gerhard Tintner

Iowa State College

Let $X_{it}(i = 1, 2 \cdots M)$ be set of M random variables, each being observed at $t = 1, 2 \cdots N$. $X_{it} = M_{it} + y_{it}$. (This is essentially the situation envisaged by Frisch [1]). The systematic part of our variables $M_{it} = EX_{it}$. The y_{it} are normally distributed with means zero. Their variances and covariances are independent of t. The M_{it} and y_{it} are independent of each other. Define $\bar{X}_i = \Sigma_t X_{it}/N$ the arithmetic mean of X_{it} and $x_{it} = X_{it} - \bar{X}_i$ the deviation from the mean. Then $a_{ij} = \Sigma_t x_{it} x_{jt}/(N - 1)$ gives the variances and covariances of the observations. We want to determine the rank of the matrix of the variances and covariances of M_{it}.

Now assume that $||V_{ij}||$ is an estimate of the variance-covariance matrix of the error terms or "disturbances" y_{it}. The elements of this matrix are distributed according to the Wishart distribution and are independent of the M_{it}. They can be estimated as deviations from polynomial trends, as deviations from Fourier series, by the Variate Difference Method, etc. The estimates could also be based upon a priori knowledge if for instance the y_{it} are interpreted as errors of measurement. Assume that the estimate is based upon N' observations.