High Speed Sampling

1. Introduction. Within the last ten years several electronic digital computing devices have been constructed or have reached advanced stages of development, e.g., the ENIAC, EDVAC, UNIVAC, SSEC, SEAC, SWAC, and a device at the Institute for Advanced Study. All these machines will perform additions, subtractions, multiplications, and divisions at high speeds. At least one of the devices can add two numbers in about one millionth of a second.

It would be desirable to introduce these high speed operations into "synthetic" sampling in statistics. At present such sampling may involve very time-consuming procedures. For example, suppose one wishes to estimate the distribution of a function (say, the standard deviation) of a random sample of size 50 from a population having the frequency function $f(x)$. ($f(x)$ may be troublesome to work with analytically; e.g., it might be an integral that is difficult to evaluate (see G in section 4).) An exact or even approximate mathematical formula for the sampling distribution may be impossible to calculate. However, the estimate can be obtained by drawing a large number of samples of size 50, computing the standard deviation of each sample, and making a frequency distribution of the standard deviations. In such work a machine for carrying out the sampling at high speed would be useful in conjunction with a high speed computing machine.

A high speed sampling machine (HSSM) might be such that directly from any given distribution function, $F(x)$, a large number of random samples of a given size would be drawn. In this case standard high speed computation could then be carried out on each of the samples. Research is needed in regard to specifying physically a distribution function, $F(x)$, and in regard to sampling at high speed from a physically specified $F(x)$. Such phenomena as radioactivity, thermionic flow, and distributions of particle velocities may provide means of solving these problems. Unless the sampling machine's fundamental selection would be directly from $F(x)$, operations with the selected values would be necessary to effect random selections from $F(x)$.

A second possibility is that from a standard distribution function (say, the standard uniform distribution), a HSSM would draw a large number of samples of given size and by means of a high speed computing unit would transform each into a sample from any preassigned distribution function.

A third possibility is that a HSSM (by means of special high speed acceptance-rejection procedures regarding randomly selected values from a standard distribution function) would effectively draw random samples directly from a preassigned distribution function. Professor JOHN VON NEUMANN pointed out and discussed this possibility at a conference August 30, 1948. The nature of the acceptance-rejection procedures is discussed in section 4 of this paper.

At present it seems that a HSSM of the second or of the third type would be easier to develop than one of the first type.

The purposes of this paper are: (1) to give brief descriptions of certain procedures for selecting numbers at random and for applying acceptance-rejection criteria to the numbers to obtain random sampling distributions;
to give arguments for the validity of the interpretations of the procedures. The notation and other preliminary matters are given in section 2. The procedures, interpretations, and derivations are given in sections 3 and 4.


A chance quantity, \( Y \), having only two possible values (say 0 and 1) will be called a "random binary digit" when the following conditions are fulfilled precisely or to a close approximation:

\[
\Pr(Y = 0) = \frac{1}{2}; \quad \Pr(Y = 1) = \frac{1}{2}.
\]

Let \( Y_1, Y_2, \ldots, Y_c \) be \( c \) independent random binary digits. Let

\[
Z = 2^{-1}Y_1 + 2^{-2}Y_2 + \cdots + 2^{-c}Y_c + 2^{-c+1};
\]

the set of possible values of \( Z \) consists of: \( 2^{-c+1}, 3 \cdot 2^{-c+1}, 5 \cdot 2^{-c+1}, \ldots, (2^{c+1} - 1) \cdot 2^{-c} \). From (2.1) and (2.2) it follows that

\[
\Pr(Z = (j + \frac{1}{2})2^{-c}) = 2^{-c}, \quad (j = 0, 1, \ldots, 2^c - 1).
\]

As \( c \) becomes larger, the probability distribution of \( Z \) is specified more and more closely by the standard uniform density function

\[
f(z) = 1, \quad (0 \leq z \leq 1).
\]

In many situations \( Z \) could be considered for practical purposes as having the density function given in (2.4) if the value of \( c \) was not less than 10, say. The definition of \( Z \) in (2.2) indicates explicitly how random binary digits can be used to select at random a chance quantity having approximately the standard uniform density function.

A chance quantity, \( S \), having only ten possible values 0, 1, \ldots, 9 will be called a random decimal digit when the following conditions are fulfilled precisely or very closely:

\[
\Pr(S = j) = \frac{1}{10}, \quad (j = 0, 1, \ldots, 9).
\]

The chance quantity

\[
W = S_110^{n-1} + S_210^{n-2} + \cdots + S_n,
\]

where \( S_1, S_2, \ldots, S_n \) are mutually independent random decimal digits, will be called a random \( n \)-digit decimal integer. A random \( n \)-digit binary integer could be defined similarly.

Throughout this report a chance quantity will be represented by a capital letter and the variable of the chance quantity's distribution function will be represented by the corresponding small letter.

Let \( V \) be a chance quantity whose cumulative distribution function, \( F(v) \), is continuous and monotone increasing. By definition

\[
\Pr(V \leq v) = F(v).
\]

It can be shown easily that the distribution function, say \( Q(x) \), of \( X = F(V) \) is

\[
Q(x) = \Pr(X \leq x) = x, \quad (0 \leq x \leq 1);
\]
thus the density function, say \( q(x) \), of \( X \) is

\[
q(x) = 1, \quad (0 \leq x \leq 1).
\]

It is interesting also to consider a certain transformation from \( X \) to \( V \). Owing to its special properties, the function \( F(v) \) has an inverse, say \( H(x) \), which is such that \( F(H(x)) = x \). If \( X \) has the standard uniform distribution, then \( V = H(X) \) has the distribution \( F(v) \). It follows that if \( V \)'s distribution is \( F(v) \) and if \( x_0 \) is a randomly selected value of \( X \), then \( v_0 = H(x_0) \) is a randomly selected value of \( V \). A machine that could at high speed obtain independently randomly selected values \( x_1, x_2, \ldots, x_b \) of \( X \), then perform the transformations \( v_i = H(x_i) \) (\( i = 1, \ldots, b \)), and then carry out standard, high speed computations with the \( v \)'s would be a high speed sampling machine (see the fourth paragraph of section 1).

3. Selecting Values of Random Numbers.¹

A. Use of Radioactivity.

**Procedure.** Use a radioactive substance having a high average emission rate, \( m \), over a unit time interval. Note whether the number, \( N \), of emissions for a given unit time interval is odd or even. Let a quantity \( Y \) take the value 1 or 0 according as \( N \) is odd or even.

**Interpretation.** \( Y \) is a random binary digit when \( m \) is large.

**Argument.** Assume that \( N \) has a Poisson distribution,

\[
p(n) = e^{-m}m^n/n!, \quad (n = 0, 1, \ldots),
\]

where \( m \) is the average emission rate. The quantity \( Y = \frac{1}{2}[1 - (-1)^N] \). From (3.1) it follows that

\[
\begin{align*}
\Pr(Y = 0) &= e^{-m} \sum_{a=0}^{\infty} m^{2a}/(2a)! = e^{-m}[e^m + e^{-m}]/2 = \frac{1}{2}[1 + e^{-2m}] \\
\Pr(Y = 1) &= \frac{1}{2}[1 - e^{-2m}].
\end{align*}
\]

The convergence of \( \Pr(Y = 0) \) and \( \Pr(Y = 1) \) to the limit \( \frac{1}{2} \) as \( m \) becomes large is unusually fast.

B. Use of Several Coin Tosses.

**Procedure.** Perform \( K \) coin tosses. Note whether \( N' \), the number of Heads obtained, is odd or even. Let \( Y \) be 1 or 0 according as \( N' \) is odd or even.

**Interpretation.** \( Y \) is a random binary digit when \( K \) is large, if no coin is badly biased.

**Argument.** Let \( p_i = \frac{1}{2} + \alpha_i \), say, be the probability of Head on the \( i \)-th toss (\( i = 1, \ldots, K \)). Assume that the tosses are mutually independent. Let \( Y = \frac{1}{2}[1 - (-1)^{N'}] \). It can be shown that

\[
\begin{align*}
\Pr(Y = 0) &= \frac{1}{2} + (-2)^{K-1} \prod_{i=1}^{K} \alpha_i, \\
\Pr(Y = 1) &= \frac{1}{2} - (-2)^{K-1} \prod_{i=1}^{K} \alpha_i.
\end{align*}
\]
The assertions in (3.3) are obviously true for $K = 1$. By mathematical induction on $K$ the assertions can be shown to be true for all values of $K$. Assume that for each $i$, $|\alpha_i| < .2$, say; then for large $K$, $Y$ is approximately a random binary digit since $|2^{K-1} \prod_{i=1}^{K} \alpha_i| < (.4)^K$.

C. Use of Two Coin Tosses.

PROCEDURE. Perform two coin tosses, $C_1$, $C_2$. Note results if and only if $C_1$ is Head and $C_2$ is Tail or $C_1$ is Tail and $C_2$ is Head; when this condition is fulfilled, note whether $C_1$ is Head. If $C_1$ is Head and $C_2$ is Tail, select the value 0; if the reverse holds, select the value 1.

INTERPRETATION. $Y$ is precisely a random binary digit.

ARGUMENT. Assume that the probability, $p$, of Head is the same for both tosses ($0 < p < 1$). Assume that the two tosses are independent; then

$$\Pr(Y = 0) = \Pr(Y = 1) = \frac{[p(1-p)]/[p(1-p) + (1-p)p]} = \frac{1}{2}.$$ 

D. “Middle of the Square” Method for Generating “Pseudo-Random” Numbers.

PROCEDURE. Form a sequence $x_0$, $x_1$, \ldots, $x_{N-1}$ of integers, where $x_j$ ($j = 0, 1, \ldots, N - 1$) is an $n$-digit decimal integer, $x_{j+1}$ consists of the successive $n$ digits in the middle of the $(2n)$-digit number $x_j^2$, and $x_0$ is a number whose successive digits are, say, $0, 1, 2, \ldots, (n - 1)$.

INTERPRETATION. Under certain conditions on $N$ and $n$ (discussed below) $x_0, x_1, \ldots, x_{N-1}$ can be considered as independently selected values of a random $n$-digit decimal integer (see (2.6)).

ARGUMENT. The argument given here will be mainly empirical, although an a priori argument, based on certain results in ergodic theory, could be used. It should be noted that $x_1, \ldots, x_{N-1}$ are completely determined by $x_0$ and $x_0$ is not necessarily chosen at random; nevertheless, experience (discussed below) indicates that the interpretation above is reasonable. Accordingly, $x_0, x_1, \ldots, x_{N-1}$ are not appropriately termed values of “pseudo-random” numbers.

For fixed $n$ the sequence $x_0, x_1, \ldots, x_{N-1}$ becomes periodic as $N \to \infty$; and there are reasons for expecting trouble for $N \approx \sqrt{2 \cdot 10^{n/2}}$. For $n = 7, 8, 10$ trouble actually arose for smaller values of $N$ (roughly, $N = \sqrt{2 \cdot 10^{n/2-1}}$). The experience for $n = 8, N \approx 700$ and for $n = 10, N \approx 3,500$ has been satisfactory; it is based on various distribution and correlation tests. (For example, for the case $n = 10$ a $\chi^2$-test was applied to the $10 \times 10$ frequency table of occurrence of the value $k$ in the $q$-th digit of $x_i$ for each $x_i$. Also, the correlation of the $k$-th digit of $x_i$ with the $q$-th digit of $x_{i+p}$ was evaluated for $k, q = 0, \ldots, 9$ and $p = 0, \ldots, 5$; the distribution of the 545 correlation coefficients was examined.) The experience for $n = 7$ was unsatisfactory.


PROCEDURE. Form a sequence $x_0, x_1, \ldots$, as follows. Select at random any non-zero 8-digit (decimal) integer as $x_0$. To compute $x_1$ multiply $x_0$ by 23. Next remove the 9-th and 10-th digits of this product and subtract this two-digit number from the remaining 8-digit number. This difference
is \( x_1 \). The number \( x_2 \) is produced in the same way from \( x_1 \) that \( x_1 \) was produced from \( x_0 \).

**INTERPRETATION.** The pseudo-random numbers \( x_j \) are being produced according to the general law

\[
x_j = x_0 k^j \pmod{M},
\]

where in the above \( k = 23 \) and \( M = 10^8 + 1 \). These special values are particularly well adapted to a 10-digit decimal machine. Although in all cases the \( x_j \) are periodic, the periods are extremely large; in the above examples the \( x \)'s are periodic with period 5882352. The method is due to Lehmer.3

**ARGUMENT.** The argument here is again empirical. The standard tests applied to such numbers gave satisfactory results. The method avoids the difficulty of the "middle of the square" method in which the sequence may unexpectedly terminate in a sequence of zero terms. It is designed primarily for use with parallel machines like the ENIAC and the SSEC.

4. **Acceptance-Rejection Methods of Obtaining Empirical Random Sampling Distributions.**

In this section it is assumed that if by some procedure a machine can, at high speed, randomly select a value of a chance quantity, then the machine can repeat the selection process a large number of times at high speed, where the selections are mutually independent. In view of this assumption only the random selection of a single value of a chance quantity need be considered.

**A. Random Selection of a Value of a Purely Discrete Chance Quantity.**

Let \( W \) be a purely discrete chance quantity whose possible values are, say, \( w_0, w_1, w_2, \ldots, w_c \), and let

\[
(4.1) \quad \Pr(W = w_j) = p_j, \quad (j = 0, \ldots, c),
\]

where \( \sum_0^c p_j = 1 \). Divide the interval \((0 \leq x \leq 1)\) into \( c + 1 \) mutually exclusive and exhaustive intervals \( I_j (j = 0, 1, \ldots, c) \), where the length of \( I_j \) equals \( p_j \). Let \( x_0 \) be a randomly selected value of a chance quantity having the standard uniform distribution; associate with \( x_0 \) the index, \( j \), of the interval \( I_j \) in which \( x_0 \) lies; then, the value \( w_j \) can be considered as a randomly selected value of \( W \).

**B. Random Selection of a Value of \( Y = \cos Z \), where \( Z \) has the Uniform Density Function \( f(z) = (2\pi)^{-1} (0 \leq z \leq 2\pi) \).**

**Procedure.** Let \( X \) be a chance quantity having the density function \( f(x) = \frac{1}{2}, (-1 \leq x \leq 1) \). Select two values of \( X \) independently at random. If the sum of squares of the values exceeds 1, reject both values and repeat the selection process until a pair of values is obtained that satisfies the condition. Let \( x_1 \) and \( x_2 \) be such a pair. Form the ratio \( x_1/(x_1^2 + x_2^2)^{1/2} \).

**INTERPRETATION.** \( Y = x_1/(x_1^2 + x_2^2)^{1/2} \) is a randomly selected value of \( Y \).

**ARGUMENT.** The conditional probability element of a pair of acceptable values is \[ \left[ \frac{(1/4)/(\pi/4)}{x_1^2 + x_2^2} \right] dx_1 dx_2 = \frac{1}{\pi} dx_1 dx_2. \] Let \( x_1 = \rho \cos z, x_2 = \rho \sin z \); thus \( \rho = (x_1^2 + x_2^2)^{1/2}, \cos z = (x_1)/(x_1^2 + x_2^2)^{1/2} \). The joint probability ele-
ment of the two new chance quantities is \((1/\pi)\rho d\rho dz\); and the probability element of \(Z\) is
\[
(4.2) \quad f(z)dz = \pi^{-1} \int_0^1 \rho d\rho dz = (2\pi)^{-1}dz, \quad (0 \leq z \leq 2\pi).
\]

C. Random Selection of a Value of \(U = -\ln X\), where \(X\) has the Standard Uniform Distribution.

**Procedure.** Let \(x_0\) and \(x_i\) be two values of \(X\) selected independently at random. If and only if \(x_1 < x_0\), select independently a third value, \(x_2\), of \(X\) at random; if and only if \(x_1 + x_2 < x_0\), select a fourth value, \(x_3\), of \(X\) at random; etc. Discontinue the procedure upon first obtaining a value, \(x_n\), of \(X\) such that
\[
(4.3) \quad x_1 + x_2 + \cdots + x_n \geq x_0.
\]
If \(n\) is an even number, reject all the values obtained and repeat the procedure until a sequence \(x_0, x_1, \ldots, x_n\) satisfying the relation in (4.3) is obtained, where \(n\) is an odd number. Retain \(x_0\); and note the number, \(t'\), of sequences formed to obtain the first acceptable sequence. Let \(t = t' - 1\).

**Interpretation.** \(t + x_0\) is a randomly selected value of \(U = -\ln x\).

**Argument.** Let \(r_j\) by the probability that the sum of the elements of a sample of \(j\) values of \(X\) will be not less than a given value \(x_0(0 < x_0 < 1)\). It can be shown easily that
\[
(4.4) \quad r_j = 1 - (x_0^{j-1})/j!, \quad (j = 1, 2, \cdots).
\]
For any given attempt let \(N\) be the number of values of \(X\) selected, excluding \(x_0\), to satisfy (4.3). We have that \(Pr(N = n) = r_n - r_{n-1}\), \((n = 1, \cdots; r_0 = 0)\). The conditional probability that \(N\) is odd, given \(x_0\), is
\[
(4.5) \quad Pr(N odd|x_0) = (r_1 - r_0) + (r_3 - r_2) + \cdots = e^{-x_0}.
\]

Let \(T'\) be the number of sequences formed to obtain the first acceptable sequence. Let \(X_0\) be the initial chance quantity in the acceptable sequence; \(X_0\) takes the value \(x_0\) in (4.3) when (4.3) is an acceptable sequence. Let \(T = T' - 1\). The probability that a given attempt will be unsuccessful is \(Pr(N even) = 1 - \int_0^1 e^{-x_0}dx_0 = e^{-1}\); thus the joint probability element of \(T\) and \(X_0\) is
\[
(4.6) \quad e^{-t} \cdot e^{-x_0}dx_0.
\]
This element is "mixed" since \(T\) is discrete and \(X_0\) is continuous. (4.6) implies that the probability element of \(U = T + X_0\) is
\[
(4.7) \quad e^{-u}du,
\]
which is the same as the probability element of \(U = -\ln X\). (It can be shown that the condition \(x_0 \geq x_1 \geq \cdots \geq x_n\) can be used in place of (4.3).)

D. Selection at Random of a Value of \(X\), where the Density Function of \(X\) is \(f(x), (0 \leq x \leq a)\), and there is a Maximum Value of \(f(x)\).

**Procedure.** Let \(b\) be the maximum of \(f(x)\). Let \(g(x) = f(x)/b\). Let \(U, V\) be independent chance quantities having the standard uniform distribution. Let \(Z = aU\). At random select values \(v\) and \(z\) of \(V\) and \(Z\), respectively.
Retain \( z \) if and only if \( v < g(z) \). Continue the procedure of selecting values of \( V \) and \( Z \) until an acceptable \( z \) is obtained.

**INTERPRETATION.** \( z \) is a randomly selected value of \( X \).

**ARGUMENT.** The joint probability element of \( Z \) and \( V \) before application of the acceptance-rejection procedure is \( a^{-1} dz \). Clearly

\[
Pr(V < g(Z)) = a^{-1} \int_0^a dz \int_0^{g(z)} dv = (ab)^{-1};
\]

therefore the joint probability element of \( Z \) and \( V \) given that \( V < g(Z) \) is

\[
f(z, v | v < g(z)) = bdz dv,
\]

and so

\[
f(z | v < g(z)) = b \left( \int_0^{g(z)} dv \right) dz = bg(z)dz = f(z)dz.
\]

**E. Random Selection of a Value of \( W = e^{-x} \), \( (0 \leq x \leq k) \), where \( X \) has a Uniform Density Function \( f(x) = 1/k \).**

**PROCEDURE.** Let \( u, v \) be randomly selected values of \( U, V \) which are independent chance quantities each having the standard uniform distribution. Retain \( u \) if and only if \( uv < e^{-k} \) and \( u > e^{-k} \). Continue until an acceptable \( u \) is obtained.

**INTERPRETATION.** \( u \) is a randomly selected value of \( W \).

**ARGUMENT.** The density function of \( W \) is

\[
h(w) = \begin{cases} 
1/(kw), & (e^{-k} \leq w \leq 1), \\
0, & (0 \leq w < e^{-k}).
\end{cases}
\]

The argument in D above may be applied, where \( a = 1, b = e^k/k \). The inequality condition in D, that \( v \) be less than \( g(z) \), becomes here that \( v \) be less than \( e^{-k}/u \) when \( u \geq e^{-k} \) and that \( v \) be less than 0 when \( u < e^{-k} \). The latter requirement means rejection of any \( u < e^{-k} \); thus the application of D leads to the procedure of accepting \( u \) if and only if \( u \geq e^{-k} \) and \( uv < e^{-k} \).

**F. Random Selection of a Value of \( X \), where the Density Function of \( X \) is \( f(x)h(x) \), \( (0 \leq x \leq a) \), and \( f, h \) have Maxima.**

**PROCEDURE.** Let \( b \) and \( c \) be the maxima of \( f(x) \) and \( h(x) \), respectively; let \( g(x) = f(x)/b, p(x) = h(x)/c \). Let \( U, V, W \) be mutually independent chance quantities having the standard uniform distribution. Let \( Z = aU \). At random select values of \( Z, V, W \). Retain \( z \) if and only if \( v < g(z) \) and \( w < p(z) \). Continue the procedure of selecting values of \( Z, V, W \) until an acceptable \( z \) is obtained.

**INTERPRETATION.** \( z \) is a randomly selected value of \( X \).

**ARGUMENT.** Assume that both \( f(x) \) and \( h(x) \) are non-negative. By an argument very similar to that used in D above it can be shown that the conditional probability element of \( Z \) given \( V < g(Z) \) and \( W < p(Z) \) is

\[
(bc)g(z)p(z)dz = f(z)h(z)dz.
\]
G. Random Selection of a Value of $X$, where the Density Function of $X$ is $f(x) = \int_0^a g(x, u)du$, $(0 \leq u \leq a)$, $(0 \leq x \leq b)$, and $g$ has a Maximum.

PROCEDURE. Let $c$ be the maximum of $g(x, u)$, $(0 \leq u \leq a; 0 \leq x \leq b)$. Let $p(x, u) = g(x, u)/c$. Let $T$, $V$, $W$ be independent chance quantities having the standard uniform distribution. Let $S = aT$ and $Y = bV$. At random select values of $S$, $Y$, and $W$. Retain $y$ if and only if $w < p(y, s)$. Continue the procedure of selecting values of $S$, $Y$, $W$ until an acceptable $y$ is obtained.

INTERPRETATION. $y$ is a randomly selected value of $X$.

ARGUMENT. Assume that $g(x, u) > 0$. Before application of the acceptance-rejection procedure the probability element of $S$, $Y$, $W$ is

\[ (ab)^{-1}dsdydw. \]

From (4.13) it follows that

\[ \Pr(W < p(Y, S)) = (ab)^{-1} \int_0^b \int_0^a p(y, s)dsdy \]

\[ = (abc)^{-1} \int_0^b \int_0^a g(y, s)dsdy \]

\[ = (abc)^{-1} \int_0^b f(y)dy = (abc)^{-1}; \]

therefore, the conditional probability element of $S$ and $Y$ given $W < p(Y, S)$ is

\[ f(s, y|w < p(y, s))dsdy = c\rho(y, s)dsdy, \]

and

\[ f(y|w < p(y, s))dy = \left(\int_0^a c\rho(y, s)ds\right)dy = \left(\int_0^a g(y, s)ds\right)dy = f(y)dy. \]

D. F. Votaw, Jr.

J. A. Rafferty

School of Aviation Medicine

Randolph Field, Texas

1 This paper was presented under the title, “High speed selection of values of chance quantities,” at a meeting of the American Statistical Association Dec. 28, 1949 and is a revision of reports submitted to the USAF School of Aviation Medicine, Aug. and Sept., 1948. The authors wish to acknowledge the valuable help of Professor J. Von Neumann.

2 For simplicity of discussion the only chance quantities considered will be one-dimensional.

3 In connection with this topic we may cite the following references.


4 In empirical studies conducted at Los Alamos.