LA-5407-MS

INFORMAL REPORT

c.3

A Method of Sampling Certain Probability Densities Without Inversion of Their Distribution Functions

los alamos
scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87544

UNITED STATES
ATOMIC ENERGY COMMISSION
CONTRACT W-7405-ENG. 36

In the interest of prompt distribution, this LAMS report was not edited by the Technical Information staff.

Printed in the United States of America. Available from
National Technical Information Service
U. S. Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22151
Price: Printed Copy $4.00; Microfiche $0.95

LA-5407-MS
Informal Report
UC-32

ISSUED: September 1973

los alamos scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87544

A Method of Sampling Certain Probability Densities Without Inversion of Their Distribution Functions

by

C. J. Everett
E. D. Cashwell
G. D. Turner

A METHOD OF SAMPLING CERTAIN PROBABILITY DENSITIES WITHOUT INVERSION OF THEIR DISTRIBUTION FUNCTIONS

by

C. J. Everett, E. D. Cashwell, and G. D. Turner

ABSTRACT

A Monte Carlo device is described which bypasses the inversion $x = P^{-1}(r)$ involved in directly sampling the distribution $P(x)$ of a stochastic variable x with given density $p(x)$. The method is practical for all linear and a broad class of quadratic densities.

I. INTRODUCTION

It is a well-known maxim of Monte Carlo practice that one should never compute the square root $x = \sqrt{r}$ of a random number, but rather set x equal to the greater of two such numbers. In general, if $p(x)$ is the density of a stochastic variable x on $[a,b]$, and $P(x) = \int_a^x p(x)dx$ its distribution, the direct way of sampling for x consists in setting a random number $r = P(x)$ and solving for $x = P^{-1}(r)$. This is how the equation $x = \sqrt{r}$ arises from the density $p(x) = 2x$ on $[0,1]$. Since such inversions are usually time consuming if not intractable, it is important to provide simple alternatives when possible. The following is a scheme which generalizes the $\sqrt{r}$ device and applies in particular to the determination $x = \sqrt{1 - (1-\xi^2)r}$ encountered in a previous report[1] on sampling the Klein-Nishina distribution (see Part III below).

II. THE GENERAL METHOD

For a distribution $P(x)$ on $[a,b]$, the function $f(r) = r^{-1}P(x)$, $x = a + (b - a)r$, $0 < r \leq 1$, has the properties

1. $f(0^+) = (b - a)p(a) \geq 0$, $f(1) = 1$
2. $f'(r) = r^{-2}[(x - a)p(x) - P(x)]$
3. $sdr + rds = p(x)dx$, $s = f(r)$, $x = a + (b - a)r$

Hence, if $f(r)$, in particular [by (2)] if $p(x)$, is increasing, then by (1) and (3), the probability $p(x)dx$ of x on $(x, x + dx)$ is the chance of a random point (r',s') of the unit square falling in the lower left region determined by $(r, r + dr)$, $(s, s + ds)$, and the curve $s = f(r)$. But this occurs iff either

(a) r' is on $(r, r + dr)$ and $s' \leq f(r')$, or s' is on $(s, s + ds)$ and $r' \leq f^{-1}(s')$, i.e.,

(b) $f^{-1}(s')$ is on $(r, r + dr)$ and $s' \geq f(r')$.

Thus, x will be obtained with density $p(x)$ if one follows

RULE 1. {Increasing $f(r) = r^{-1}P[a + (b - a)r]$}

I. Generate random numbers r', s'

II. Define $\rho = \begin{cases} r' \text{ if } s' \leq f(r') \\ f^{-1}(s') \text{ if } s' > f(r') \end{cases}$

III. Set $x = a + (b - a)\rho$.

Analogously, the function $g(r) = r^{-1}Q(x)$, $Q(x) = \int_x^b p(x)dx$, $x = b - (b - a)r$, has properties

(1) $g(0^+) = (b - a)p(b) \geq 0$, $g(1) = 1$

(2) $g'(r) = r^{-2}[(b - x)p(x) - Q(x)]$

(3) $sdr + rds = p(x)(-dx) \geq 0,\ s = g(r),$

$x = b - (b - a)r$.

Now, if g(r) is <u>increasing</u>, in particular (by (2)) if p(x) is <u>decreasing</u>, then it is clear that the density p(x) results from

RULE 2. $\{\text{Increasing } g(r) = r^{-1}Q[b - (b - a)r]\}$

I. Generate r', s'

II. Define $\rho = \begin{cases} r' \text{ if } s' \leq g(r') \\ g^{-1}(s') \text{ if } s' > g(r') \end{cases}$

III. Set $x = b - (b - a)\rho$.

III. LINEAR DENSITIES

The method applies to any linear density $p(x) = C^{-1}(c_0 + c_1x) \geq 0$ on $a \leq x \leq b$, where $c_1 \neq 0$, and $C = (b - a)\left[c_0 + \frac{1}{2}c_1(b + a)\right]$, thus bypassing solution of the quadratic equation $r = P(x) = C^{-1}(x - a)\left[c_0 + \frac{1}{2}c_1(x + a)\right]$ for x.

Case 1. If $c_1 > 0$, then for $x = a + (b - a)r$ one finds

$$f(r) = r^{-1}P(x) = \left[c_0 + c_1a + \frac{1}{2}c_1(b - a)r\right] \div \left[c_0 + \frac{1}{2}c_1(b + a)\right] ,$$

increasing for $0 \leq r \leq 1$, and RULE 1 defines

$$x = a + \max\left[(b - a)r',\ \left(b + a + 2c_0c_1^{-1}\right)s' - 2\left(a + c_0c_1^{-1}\right)\right] .$$

In particular, for ξ fixed, $0 \leq \xi < 1$ and $p(x) = 2x/(1 - \xi^2)$ on $[\xi,1]$, this reads

$$x = \xi + \max[(1 - \xi)r',\ (1 + \xi)s' - 2\xi] .$$

For $\xi > 0$, the latter provides an alternative to the choice $x = \sqrt{\xi^2 + (1 - \xi^2)r}$, while for $\xi = 0$, it becomes $x = \max(r',s')$ in lieu of $x = \sqrt{r}$, the example cited at the outset.

Case 2. If $c_1 < 0$, then for $x = b - (b - a)r$, we have

$$g(r) = r^{-1}Q(x) = \left[c_0 + c_1b - \frac{1}{2}c_1(b - a)r\right] \div \left[c_0 + \frac{1}{2}c_1(b + a)\right] ,$$

increasing on [0,1], and RULE 2 sets

$$x = b - \max\left[(b - a)r',\ -\left(b + a + 2c_0c_1^{-1}\right)s' + 2\left(b + c_0c_1^{-1}\right)\right] .$$

IV. QUADRATIC DENSITIES

For a quadratic density $p(x) = C^{-1}\ p_1(x)$, $p_1(x) = c_0 + c_1x + c_2x^2$ on [a,b], with $c_2 \neq 0$,

$$C = (b - a)\left[c_0 + \frac{1}{2}c_1(b + a) + \frac{1}{3}c_2\left(b^2 + ba + a^2\right)\right] ,$$

one obtains

$$f(r) = r^{-1}\ P(x) = (b - a)\left[p(a) + \frac{1}{2}p'(a)\lambda + \frac{1}{6}p''(a)\lambda^2\right],\quad x = a + \lambda,\quad p''(a) = 2C^{-1}c_2,$$

$$\lambda = (b - a)r ,$$

whence

$$f'(r) = (b - a)^2\left[\frac{1}{2}p'(a) + \frac{1}{3}p''(a)\lambda\right],$$

$$f'(0) = \frac{1}{2}(b - a)^2\ p'(a) .$$

Similarly,

$$g(r) = r^{-1}Q(x) = (b - a)\left[p(b) - \frac{1}{2}p'(b)\lambda + \frac{1}{6}p''(b)\lambda^2\right],\ x = b - \lambda,\ p''(b) = 2C^{-1}c_2,$$

$$\lambda = (b - a)r ,$$

with

$$g'(r) = (b - a)^2\left[-\frac{1}{2}p'(b) + \frac{1}{3}p''(b)\lambda\right],$$

$$g'(0) = -\frac{1}{2}(b - a)p'(b) .$$

Now for such a p(x) with $c_2 > 0$, it is evident that, since our method requires either $f'(0) \geq 0$ or

$g'(0) \geq 0$, we must have $p'(a) \geq 0$ or $p'(b) \leq 0$, and therefore $p(x)$ must be monotone on the whole range $[a,b]$. (Graphically, $y = p(x)$ is a parabola opening up.) The method of course applies to such densities, and we omit the obvious details.

More interesting is the fact that quadratic densities with $c_2 < 0$ (parabolas opening down), which are not necessarily monotone, are covered by the rules, provided the interval $[a,b]$ (lying between the zeros of $p(x)$) is sufficiently restricted to render $f(r)$ or $g(r)$ increasing on $[0,1]$. By the above remarks, it is clear that we are limited to the two cases:

Case 1. $f'(0) > 0$, $f'(1) \geq 0$, equivalently, $a < -\frac{1}{2} c_1 c_2^{-1}$ and $b \leq -\frac{1}{2}\left(a + \frac{3}{2} c_1 c_2^{-1}\right)$, with RULE 1 applicable.

Case 2. $g'(0) > 0$, $g'(1) \geq 0$, equivalently, $b > -\frac{1}{2} c_1 c_2^{-1}$ and $a \geq -\frac{1}{2}\left(b + \frac{3}{2} c_1 c_2^{-1}\right)$. Here RULE 2 applies. Obviously no $p(x)$ falls under both cases.

For quadratic $p(x)$, the method, when applicable, avoids solution of the cubic equation

$$r = P(x) = \sum_0^2 \frac{p^{(\nu)}(a)}{(\nu + 1)!} (x - a)^{\nu+1} ,$$

by means of a single square root. Even the latter might be avoided by further application of the rules to a linear density, but this we do not discuss, save to remark that one is led in this way to the well-known alternative $x = \max(r', s', t')$ for $x = r^{1/3}$ in the case of $p(x) = 3x^2$ on $[0,1]$.

The method, for the quadratic densities covered, is summarized below.

$$\text{Define } \alpha = 3\left(a + \frac{c_1}{2c_2}\right), \ \beta = 3\left(b + \frac{c_1}{2c_2}\right)$$

$$\lambda(s) = \frac{1}{2}\left\{-\alpha + \operatorname{sgn} c_2 \sqrt{\alpha^2 + 12c_2^{-1}\left[\frac{C}{b - a} s - p_1(a)\right]}\right\}$$

$$\mu(s) = \frac{1}{2}\left\{\beta + \operatorname{sgn} c_2 \sqrt{\beta^2 + 12c_2^{-1}\left[\frac{C}{b - a} s - p_1(b)\right]}\right\}$$

(a) If $c_2 > 0$, $p_1'(a) \geq 0$, or if $c_2 < 0$, $a < -c_1/2c_2$, $b \leq -\frac{1}{2}\left(a + \frac{3c_1}{2c_2}\right)$

$$\text{set} \qquad x = \begin{cases} a + (b - a)r' \ ; & s' \leq f(r') \\ a + \lambda(s') \ ; & s' > f(r') \end{cases}$$

(b) If $c_2 > 0$, $p_1'(b) \leq 0$, or if $c_2 < 0$, $b > -c_1/2c_2$, $a \geq -\frac{1}{2}\left(a + \frac{3c_1}{2c_2}\right)$

$$\text{set} \qquad x = \begin{cases} b - (b - a)r' \ ; & s' \leq g(r') \\ b - \mu(s') \ ; & s' > g(r') \end{cases}$$

V. NOTE ON STATISTICS

For a general density $p(x)$ on $[a,b]$, the probability of x falling on a particular subinterval $[c,d]$ is $p = \int_c^d p(x)dx$. If, in an experiment of any kind, the event of assigning x to $[c,d]$ has probability p of success, and hence probability $q = 1 - p$ of failure; and if M successes are observed in a large number N of such experiments, then the central limit theorem asserts the approximate relation

$$P\left\{\left|\frac{M}{N} - p\right| < \varepsilon\right\} \cong \frac{2}{\sqrt{2\pi}} \int_0^t e^{-u^2/2} du; \ t = \varepsilon \sqrt{N/pq} ,$$

the difference depending only on N, p, and q.

It follows that the direct method $x = P^{-1}(x)$, and the method of choosing x by the RULES, both involving experiments assigning x to $[c,d]$ with probability p, are of identical statistical reliability. This is reflected in the following part.

VI. TWO EXAMPLES

Example 1. The density

$$p(x) = 2x/(1 - \xi^2) = 8x/3 \text{ on } \xi = \frac{1}{2} \leq x \leq 1$$

was sampled $N = 10{,}000$ times by each of the two methods

$$x = \sqrt{1 - \frac{3}{4} r} \quad , \quad \text{and} \quad x = \frac{1}{2} + \max\left(\frac{1}{2} r', \frac{3}{2} s' - 1\right),$$

the values of x obtained being classified in 10 equal subintervals of $\left[\frac{1}{2},1\right]$. The resulting M_i/N with the exact probabilities p_i are tabulated as follows.

i	1	2	3	4	5	6	7	8	9	10
ROOT	0.0695	0.0797	0.0816	0.0884	0.0983	0.0962	0.1088	0.1230	0.1265	0.1280
RULE	0.0688	0.0767	0.0788	0.0898	0.0971	0.1023	0.1152	0.1209	0.1212	0.1292
p_i	0.0700	0.0767	0.0833	0.0900	0.0967	0.1033	0.1100	0.1167	0.1233	0.1300

Example 2. The non-monotone density $p(x) = \frac{3}{164}\left(15 - 2x - x^2\right)$ on [-2,2] was sampled 10,000 times using RULE 2. The value assigned to x by a trial involving r', s' was $x = 2 - 4\rho$, where $\rho = r'$ if $41s' < 21 + r'(36 - 16r')$, and $\rho = \frac{1}{8}\left(9 - \sqrt{165 - 164s'}\right)$ otherwise. The result of classifying the x obtained in 10 equal subintervals of [-2,2] is shown below, with corresponding exact probabilities p_i.

i	1	2	3	4	5	6	7	8	9	10
RULE	0.1139	0.1123	0.1221	0.1156	0.1143	0.1080	0.1002	0.0852	0.0730	0.0554
p_i	0.1123	0.1158	0.1170	0.1158	0.1123	0.1064	0.0982	0.0877	0.0748	0.0596

REFERENCE

1. C. J. Everett, E. D. Cashwell, G. D. Turner, "A New Method of Sampling the Klein-Nishina Probability Distribution for All Incident Photon Energies Above 1 keV," Los Alamos Scientific Laboratory report LA-4663 (May 1971).

ALT:283(80)