RANDOM ALMS

By Paul R. Halmos

Syracuse University

1. Statement of the problem. Consider the problem of distributing one pound of gold dust at random among a countably infinite set of beggars. Let the beggars be enumerated and let the procedure for distribution be as follows: the first beggar is given a random portion of the gold; the second beggar gets a random portion of the remainder; $\cdots$ and so on ad infinitum. In this description the phrase "random portion" occurs an infinite number of times: it seems reasonable to require that it have the same interpretation each time. To be precise: let x_j $(j = 0, 1, 2, \cdots)$ be the amount received by the jth beggar. Let the distribution of x_0 be given by a density function $p(\lambda)$:

$$p(\lambda) \geqq 0, \qquad 0 \leqq \lambda \leqq 1; \tag{1}$$

$$\int_0^1 p(\lambda)\, d\lambda = 1; \tag{2}$$

$$P(a < x_0 < b) = \int_a^b p(\lambda)\, d\lambda, \qquad 0 \leqq a < b \leqq 1. \tag{3}$$

After the first beggar has received his alms and the amount of gold dust left is μ, (i.e. $x_0 = 1 - \mu$), the value of x_1 will be between 0 and μ. The uniformity requirement mentioned above means that the proportion of μ that the second beggar is to receive is again determined by the probability density p: in other words the conditional probability that x_1 be between $\lambda\mu$ and $(\lambda + d\lambda)\mu$, given that $x_0 = 1 - \mu$, is $p(\lambda)\, d\lambda$. In symbols:

$$P(a\mu < x_1 < b\mu \mid x_0 = 1 - \mu) = \int_a^b p(\lambda)\, d\lambda. \tag{4}$$

Writing $\alpha = a\mu$, $\beta = b\mu$, (4) becomes

$$P(\alpha < x_1 < \beta \mid x_0 = 1 - \mu) = \int_\alpha^\beta \frac{1}{\mu}\, p\left(\frac{\lambda}{\mu}\right) d\lambda. \tag{5}$$

More generally I shall assume that the conditional probability distribution of x_n, assuming that after the preceding donations there is left an amount μ, is given in the interval $(0, \mu)$ by $\frac{1}{\mu}\, p\left(\frac{\lambda}{\mu}\right)$. In symbols:

$$P(a < x_n < b \mid \sum_{j<n} x_j = 1 - \mu) = \int_a^b \frac{1}{\mu}\, p\left(\frac{\lambda}{\mu}\right) d\lambda, \qquad 0 \leqq a < b \leqq \mu. \tag{6}$$

This assumption completely determines (in terms of p) the joint distribution of the whole infinite sequence $\{x_0, x_1, x_2, \cdots\}$. Several interesting special questions may be asked about this distribution. For example: What are the expectation, dispersion, and higher moments of the x_n ? What, similarly, are the moments of the partial sum $S_n = \sum_{j \leqq n} x_j$? More generally what are the exact distributions of x_n and of S_n ? Will the process described really distribute all the gold, or is there a positive probability that some is left even after every beggar had his turn? What is the rate of convergence of the series $\sum_{n \geqq 0} x_n$? It is the purpose of this paper to answer these and a few related questions.

2. Calculation of distributions. The $n + 1$ dimensional probability density of the distribution of $(x_0, x_1, \cdots, x_n)$ is given by[1]

$$\prod_{i \leqq n} \frac{1}{1 - \sum_{j<i} \lambda_j} p\left(\frac{\lambda_i}{1 - \sum_{j<i} \lambda_j}\right) \tag{7}$$

in the region defined by $\lambda_j \geqq 0$, $\lambda_0 + \cdots + \lambda_n \leqq 1$. For $n = 0$ there is only one term in the product and that one is equal to $p(\lambda_0)$; the region is defined by $0 \leqq \lambda_0 \leqq 1$. The formula reduces in this case to the definition of the distribution of x_0. The general case follows inductively by the use of the conditional probability formula (6). $\left(\text{For example: } P(x_0 = \lambda_0, x_1 = \lambda_1) = P(x_0 = \lambda_0)P(x_1 = \lambda_1 \mid x_0 = \lambda_0) = p(\lambda_0) \frac{1}{1 - \lambda_0} p\left(\frac{\lambda_1}{1 - \lambda_0}\right).\right)$

From (7) it is possible in principle to calculate the densities of the distributions of x_n and of S_n. Thus for example the density q_n of the distribution of x_n is found by integrating out the λ_j with $j < n$ from (7), so that

$$q_n(\lambda_n) = \int \cdots \int \prod_{i \leqq n} \frac{1}{1 - \sum_{j<i} \lambda_j} p\left(\frac{\lambda_i}{1 - \sum_{j<i} \lambda_j}\right) d\lambda_0 \cdots d\lambda_{n-1}, \tag{8}$$

where the integration is extended over the region defined by $\lambda_j \geqq 0$ $(0 \leqq j \leqq n)$, $\sum_{j \leqq n} \lambda_j \leqq 1$. Similarly $V_n(t) = P(S_n < t)$ is given by

$$V_n(t) = \int \cdots \int \prod_{i \leqq n} \frac{1}{1 - \sum_{j<i} \lambda_j} p\left(\frac{\lambda_i}{1 - \sum_{j<i} \lambda_j}\right) d\lambda_0 \cdots d\lambda_n, \tag{9}$$

$(0 \leqq t \leqq 1)$ where the domain of integration is defined by $\lambda_j \geqq 0$ $(0 \leqq j \leqq n)$, $\sum_{j \leqq n} \lambda_j < t$.

Working with integrals of the type (8) and (9) is often greatly facilitated by the substitution $\mu_i = \sum_{j \leqq i} \lambda_j$, $(\lambda_i = \mu_i - \mu_{i-1})$, $0 \leqq i \leqq n$. The Jacobian of this linear change of variables is identically one. The domain of integration used in (9) is defined in terms of the μ's by $0 \leqq \mu_0 \leqq \mu_1 \leqq \cdots \leqq \mu_n \leqq t \leqq 1$, so that

$$V_n(t) = \int_0^t d\mu_n \int_0^{\mu_n} d\mu_{n-1} \cdots \int_0^{\mu_1} d\mu_0 \prod_{i \leqq n} \frac{1}{1 - \mu_{i-1}} p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right). \tag{10}$$

[1] A summation or a product extended over an empty set of indices will, as is customary, be interpreted as 0 or 1 respectively. Since throughout this paper only non-negative indices are considered, whenever the notation indicates a negative index the quantity to which it is attached is to be interpreted as 0.

Hence the density of the distribution of S_n is

$$v_n(t) = \int_0^t d\mu_{n-1} \int_0^{\mu_{n-1}} d\mu_{n-2} \cdots \int_0^{\mu_1} d\mu_0 \prod_{i<n} \frac{1}{1-\mu_{i-1}} p\left(\frac{\mu_i - \mu_{i-1}}{1-\mu_{i-1}}\right) \cdot \frac{1}{1-\mu_{n-1}} p\left(\frac{t-\mu_{n-1}}{1-\mu_{n-1}}\right). \tag{11}$$

For later purposes it is more convenient to set $t = \mu_n$ in (11) and to express $v_n(\mu_n)$ as a multiple (and not as an iterated) integral; then

$$v_n(\mu_n) = \int \cdots \int \prod_{i \leqq n} \frac{1}{1-\mu_{i-1}} p\left(\frac{\mu_i - \mu_{i-1}}{1-\mu_{i-1}}\right) d\mu_0 \cdots d\mu_{n-1}, \tag{12}$$

where the domain of integration is defined by $0 \leqq \mu_0 \leqq \mu_1 \leqq \cdots \leqq \mu_{n-1} \leqq \mu_n \leqq 1$. The integrals (8) and (12) are explicitly evaluated below for a special case.

It is possible from (8) to find the kth moment $M_k^{(n)}$ of x_n, $M_k^{(n)} = \int_0^1 \lambda_n^k q_n(\lambda_n)\, d\lambda_n$. Write

$$\alpha_k = \int_0^1 \lambda^k p(\lambda)\, d\lambda, \qquad \beta_k = \int_0^1 (1-\lambda)^k p(\lambda)\, d\lambda.$$

Clearly $M_k^{(n)}$ is obtained from (8) upon multiplication by λ_n^k and integration with respect to λ_n.

$$M_k^{(n)} = \int \cdots \int \lambda_n^k \prod_{i \leqq n} \frac{1}{1-\sum_{j<i}\lambda_j} p\left(\frac{\lambda_i}{1-\sum_{j<i}\lambda_j}\right) d\lambda_0 \cdots d\lambda_n. \tag{13}$$

It is advantageous once again to write $\mu_i = \sum_{j \leqq i} \lambda_j$. The resulting integral may be written in the iterated form as follows:

$$M_k^{(n)} = \int_0^1 d\mu_0 \int_{\mu_0}^1 d\mu_1 \cdots \int_{\mu_{n-1}}^1 d\mu_n \prod_{i \leqq n} \frac{1}{1-\mu_{i-1}} \cdot p\left(\frac{\mu_i - \mu_{i-1}}{1-\mu_{i-1}}\right) \cdot (\mu_n - \mu_{n-1})^k. \tag{14}$$

Consider separately the innermost integral

$$J = \int_{\mu_{n-1}}^1 p\left(\frac{\mu_n - \mu_{n-1}}{1-\mu_{n-1}}\right) (\mu_n - \mu_{n-1})^k \frac{d\mu_n}{1-\mu_{n-1}}.$$

Writing $\lambda = (\mu_n - \mu_{n-1})/(1-\mu_{n-1})$ this becomes

$$J = \int_0^1 p(\lambda)\lambda^k (1-\mu_{n-1})^k\, d\lambda = \alpha_k (1-\mu_{n-1})^k.$$

Hence

$$M_k^{(n)} = \alpha_k \int_0^1 d\mu_0 \cdots \int_{\mu_{n-2}}^1 d\mu_{n-1} \prod_{i<n} \frac{1}{1-\mu_{i-1}} p\left(\frac{\mu_i - \mu_{i-1}}{1-\mu_{i-1}}\right) \cdot (1-\mu_{n-1})^k. \tag{15}$$

The innermost integral this time is

$$J' = \int_{\mu_{n-2}}^{1} p\left(\frac{\mu_{n-1} - \mu_{n-2}}{1 - \mu_{n-2}}\right)(1 - \mu_{n-1})^k \frac{d\mu_{n-1}}{1 - \mu_{n-2}}.$$

Write $\lambda = (\mu_{n-1} - \mu_{n-2})/(1 - \mu_{n-2})$; then $(1 - \mu_{n-1}) = (1 - \lambda)(1 - \mu_{n-2})$ and

$$J' = \int_0^1 p(\lambda)(1 - \lambda)^k(1 - \mu_{n-2})^k \, d\lambda = \beta_k(1 - \mu_{n-2})^k.$$

Hence, finally,

$$(16)\qquad \begin{aligned} M_k^{(n)} = \alpha_k\beta_k \int_0^1 d\mu_0 \cdots \int_{\mu_{n-3}}^1 d\mu_{n-2} \prod_{i<n-1} \\ \cdot \frac{1}{1 - \mu_{i-1}} p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right)\cdot(1 - \mu_{n-2})^k. \end{aligned}$$

Observe now that the right member of (16) (except for the factor β_k) may be obtained from (15) upon replacing n by $n - 1$. In other words $M_k^{(n)} = \beta_k M_k^{(n-1)}$. Since $M_k^{(0)} = \alpha_k$, it follows that

$$(17)\qquad M_j^{(n)} = \alpha_k\beta_k^n, \qquad n = 0, 1, 2, \cdots.$$

Instead of calculating similarly the moments $\int_0^1 \mu_n^k v_n(\mu_n)\, d\mu_n$ of S_n it is more convenient to calculate the quantities

$$N_k^{(n)} = \int_0^1 (1 - \mu_n)^k v_n(\mu_n)\, d\mu_n .$$

The moments themselves may be obtained from the N's by simple combinatorial formulas.

It follows from (12) that

$$(18)\quad N_k^{(n)} = \int_0^1 d\mu_0 \int_{\mu_0}^1 d\mu_1 \cdots \int_{\mu_{n-1}}^1 d\mu_n \prod_{i \leq n} \frac{1}{1 - \mu_{i-1}} p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right)(1 - \mu_n)^k.$$

The innermost integral in (18) is

$$J'' = \int_{\mu_{n-1}}^1 p\left(\frac{\mu_n - \mu_{n-1}}{1 - \mu_{n-1}}\right)(1 - \mu_n)^k \frac{d\mu_n}{1 - \mu_{n-1}}.$$

Writing $\lambda = (\mu_n - \mu_{n-1})/(1 - \mu_{n-1})$, $(1 - \mu_n)$ becomes $(1 - \lambda)(1 - \mu_{n-1})$, so that

$$J'' = \int_0^1 p(\lambda)(1 - \lambda)^k(1 - \mu_{n-1})^k \, d\lambda = \beta_k(1 - \mu_{n-1})^k.$$

Consequently

$$(19)\qquad \begin{aligned} N_k^{(n)} &= \beta_k \int_0^1 d\mu_0 \cdots \int_{\mu_{n-2}}^1 d\mu_{n-1} \prod_{i<n} \frac{1}{1 - \mu_{i-1}} p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right)\cdot(1 - \mu_{n-1})^k \\ &= \beta_k N_k^{(n-1)}, \end{aligned}$$

so that

$$N_k^{(n)} = \beta_k^{n+1}, \qquad n = 0, 1, 2, \cdots . \tag{20}$$

The additivity of the first moment yields an amusing check on (17) and (20). Since $E(S_n) = E(\sum_{j\leqq n} x_j) = \sum_{j\leqq n} E(x_j)$ (where E denotes expectation, or first moment), it should be true that $1 - N_1^{(n)} = \sum_{j\leqq n} M_1^{(j)}$. In terms of α's and β's this means $1 - \beta_1^{n+1} = \alpha_1 \sum_{j\leqq n} \beta_1^j$, and this in turn reduces to the trivial identity $\alpha_1 = 1 - \beta_1$.

Since $0 \leqq \sum_{j\leqq n} x_j \leqq 1$ with probability 1 for every n, it is clear that the series $\sum_{j\geqq 0} x_j$ converges with probability 1 to a sum x, $0 \leqq x \leqq 1$. Since $E(x_j) = \alpha_1\beta_1^j$ and since $E(x) = \sum_{j\geqq 0} E(x_j)$, it follows that $E(x) = \sum_{j\geqq 0} \alpha_1\beta_1^j = \alpha_1/(1 - \beta_1) = 1$. This implies (since $0 \leqq x \leqq 1$) that x must be equal to 1 with probability 1. In other words it is almost certain that all the gold dust will eventually be distributed.

3. Product representation. Considerable light is shed on some of the above computations (and in fact the moment formulas (17) and (20) are proved anew) by the following considerations. The principle of equitable treatment enunciated in the introductory paragraph was subsequently formalized by the conditional probability relation (6). It may also be formalized by the following (equivalent) procedure. Let $y_0, y_1, y_2, \cdots$ be a sequence of *independent* chance variables each of whose distributions is given by the probability density p; let y_n be interpreted as the proportion, of the amount available to the nth beggar, that he actually receives. In other words

$$x_n = y_n(1 - \sum_{j<n} x_j), \qquad n = 0, 1, 2, \cdots . \tag{21}$$

The first main problem in this formulation is to express the x's in terms of the y's. This is most easily accomplished by an inductive proof of the formula

$$\sum_{j\leqq n} x_j = 1 - \prod_{j\leqq n} (1 - y_j). \tag{22}$$

For $n = 0$, (22) asserts merely that $x_0 = y_0$. The inductive step proceeds as follows:

$$\begin{aligned}\sum_{j\leqq n} x_j &= x_n + \sum_{j\leqq n-1} x_j = y_n(1 - \sum_{j\leqq n-1} x_j) + \sum_{j\leqq n-1} x_j \\ &= y_n \prod_{j\leqq n-1} (1 - y_j) + 1 - \prod_{j\leqq n-1} (1 - y_j) \\ &= 1 - (1 - y_n) \prod_{j\leqq n-1} (1 - y_j) = 1 - \prod_{j\leqq n} (1 - y_j).\end{aligned}$$

From (22) it follows that

$$x_n = y_n \prod_{j<n} (1 - y_j) \tag{23}$$

and

(24) $$R_n = 1 - S_n = 1 - \sum_{j \leqq n} x_j = \prod_{j \leqq n} (1 - y_j).$$

The moment formulas (17) and (20) follow immediately from (23) and (24) respectively.

Another very important application of (23) and (24) is the following theorem. If the first geometric moment (geometric mean)

$$r = \exp \{E(\log [1 - y_j])\} = \exp \left\{\int_0^1 \log (1 - \lambda)p(\lambda)\, d\lambda\right\}$$

is different from zero $\left(\text{i.e. if } \int_0^1 \log (1 - \lambda)p(\lambda)\, d\lambda \text{ is finite}\right)$ then the limits

$$\lim_{n \to \infty} (x_n/y_n)^{1/n} \quad \text{and} \quad \lim_{n \to \infty} R_n^{1/n}$$

both exist and are both equal to r.

Since according to (23) and (24), $x_n/y_n = R_{n-1}$ the two parts of the conclusion are seen to be equivalent. For the proof take the logarithm of both sides of (24) and divide by n, obtaining

(25) $$\log R_n^{1/n} = \frac{1}{n} \sum_{j \leqq n} \log (1 - y_j).$$

Since, according to the hypotheses stated, the chance variables $\log (1 - y_j)$ are independent and all have the same distribution with a finite expectation, the strong law of large numbers applies to the right side of (25) and (after taking exponentials) yields the desired conclusion.

The result just obtained may be phrased as follows: with probability 1 x_n is asymptotically equal to $r^n y_n$. This statement shows that in an obvious if somewhat crude sense the rate of convergence of $\sum_{j \geqq 0} x_j$ is that (at least) of a geometric series with ratio r. This conclusion is further supported by the behavior of R_n, which again is the sort of thing one expects from a geometric series. (That is: the nth root of the nth remainder of a geometric series always does converge to the common ratio.) As usual, more delicate quantitative results concerning the rate of convergence may be obtained by applying to (25) not merely the law of large numbers but the law of the iterated logarithm.

The product representation of x_n in formula (23) points the way to a generalization of this theory which may be of some interest. In this generalization x_n is still defined by (23) and the y's are still independent, but the distribution of y_j is given by a density p_j, where the p's need not be equal to each other. In terms of random alms this means that the condition of equitable treatment is replaced by the following weaker condition: the probability distribution of the amount that the jth beggar receives depends only on j and on the amount left by the preceding beggars, and in particular does not depend on the sizes of the alms already distributed. Many of the conclusions obtained under the simpler hypotheses carry over to this generalized case with only slight changes. In particular the distribution formulas (7), (8), and (12), and the moment formulas (17) and (20), are changed only to the extent of acquiring an extra subscript due to the difference of the p_j.

4. Applications. (A) The original motivation of the present work was an investigation of the notion of a random mass distribution, and the results obtained may be considered as one possible solution of the problem of defining randomness for mass distributions in the special (discrete) case where the entire mass is concentrated on the non-negative integers. It would be of great interest to extend the results of this note to various continuous cases in which the set of integers is replaced by the unit interval, or the entire real line, or n dimensional Euclidean space. I intend to study some of these extensions at another time; at the moment I merely mention one implication of this statistical point of view.

Considering the sequence $\{x_0, x_1, x_2, \cdots\}$ as a system of weights, the integer n carrying the weight x_n, various questions may be raised concerning properties of the discrete mass distributions so obtained. For example: do the moments $m_k = \sum_{n \geqq 0} n^k x_n$ exist and, if so, what are their averages and dispersions and, more generally, their moments and their distributions? I shall settle here the questions concerning existence and expectation.

The chance variable m_k is non-negative and, even it if is infinite with positive probability, its expectation is defined by $E(m_k) = \sum_{n \geqq 0} n^k E(x_n) = \sum_{n \geqq 0} n^k M_1^{(n)} = \sum_{n \geqq 0} n^k \alpha_1 \beta_1^n$. Since $0 < \beta_1 < 1$, the last written series converges and therefore $E(m_k)$ is finite. This implies that m_k is finite with probability 1.

(B) It has been observed that the logarithms of the sizes of particles such as mineral grains are frequently normally distributed. Kolmogoroff[2] has given an explanation of this phenomenon; the results of the present paper yield an alternative and in some respects simpler explanation. Suppose in fact that the probability of a particle losing a chip the proportion of whose size to the size of the original particle is between λ and $\lambda + d\lambda$ is $p(\lambda)\, d\lambda$. With this stochastic scheme the size of the remaining particle after n chips have been lost is given by R_n. Since, by (25), log R_n is a sum of independent chance variables with the same distributions, the Laplace-Liapounoff theorem may be invoked to show that the distribution of R_n is for large n nearly normal. (It is necessary of course to assume here the finiteness of the second geometric moment, or equivalently of the integral $\int_0^1 \log^2 (1 - \lambda)p(\lambda)\, d\lambda$.) The mean and the variance of each summand of log R_n are

$$a = \int_0^1 \log (1 - \lambda)p(\lambda)\, d\lambda \quad \text{and} \quad b^2 = \int_0^1 [\log (1 - \lambda) - a]^2 p(\lambda)\, d\lambda,$$

[2] A. N. Kolmogoroff, "Ueber das logarithmisch normale Verteilungsgesetz der Dimensionen der Teilchen bei Zerstueckelung," *C. R. (Doklady) Acad. Sci. URSS* (N. S.) Vol. 31(1941), pp. 99–101.

respectively; consequently (by the additivity of the mean and the variance) the corresponding parameters of the distribution of log R_n (and hence of the approximating normal distribution) are given by $(n + 1)a$ and $(n + 1)b^2$ respectively.

(C) A special case of the distributions studied in this paper (namely the case of uniform distribution, $p(\lambda) = 1$) arises in the theory of scattering of neutrons by protons of the same mass. According to Bethe[3]: "In each collision with a proton the neutron will lose energy. As long as the neutron is fast compared to the proton, the probability that the neutron energy lies between E and $E + dE$ after the collision, is $w(E)\,dE = dE/E_0$, where E_0 is the neutron energy before the collision. This means that any value of the final energy of the neutron, between 0 and the initial energy E_0, is equally probable."

To calculate explicitly the distributions it is most convenient to start from (11). If p (with any argument) is replaced by 1 and the terms of the product are distributed, each under its own differential, (11) takes the form

$$v_n(t) = \int_0^t \frac{d\mu_{n-1}}{1 - \mu_{n-1}} \int_0^{\mu_{n-1}} \frac{d\mu_{n-2}}{1 - \mu_{n-2}} \cdots \int_0^{\mu_1} \frac{d\mu_0}{1 - \mu_0}. \tag{26}$$

The value of the iterated integral is easy to obtain: $v_n(t) = (-1)^n(1/n!) \log^n (1 - t)$. Since $v_n(t)$ gives the distribution of the partial sum S_n, the distribution of $R_n = 1 - S_n$ is given by $v_n(1 - t) = (-1)^n(1/n!) \log^n t$.[4] It is possible but not necessary to derive similarly the distribution of x_n. It is simpler to obtain this distribution by exploiting the symmetry of the uniform distribution. Since, according to (23) and (24), x_n and R_n are both products of $n + 1$ uniformly and independently distributed chance variables they have the same distribution, so that the density of the distribution of x_n is also given by $(-1)^n(1/n!) \log^n t$, $n = 0, 1, 2, \cdots$.

The roles of the geometric mean r ($= 1/e$ in case $p = 1$) and of the normal distribution have also been observed in the physical situation. Fermi[5] has expressed the geometric series like behavior of $\sum_{n \geq 0} x_n$ by the statement "$\cdots$ an impact of a neutron against a proton reduces, on the average, the neutron energy by a factor $1/e$," and Bethe[6] remarks that "$\cdots$ the actual values of log E after n collisions form very nearly a Gaussian distribution $\cdots$"

---

[3] H. A. Bethe, "Nuclear Physics, B. Nuclear Dynamics, Theoretical," *Reviews of Modern Physics*, Vol. 9 (1937) p. 120.

[4] This distribution has been calculated by E. U. Condon and G. Breit, "The energy distribution of neutrons slowed by elastic impacts," *Physical Review*, Vol. 49 (1936) pp. 229–231.

[5] Quoted by Condon and Breit, *loc. cit.*

[6] *Loc. cit.*