An Introduction to Probability Theory and Its Applications

WILLIAM FELLER (1906 - 1970)

Eugene Higgins Professor of Mathematics

Princeton University

VOLUME I

THIRD EDITION

Revised Printing

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

30

ISBN 0 471 25708-7

Library of Congress Catalog Card Number: 68-11708
Printed in the United States of America

Printed and bound by Quinn - Woodbine, Inc.

are either no descendants or else a great many descendants (the corresponding probabilities being x and $1 - x$).

5. THE TOTAL PROGENY IN BRANCHING PROCESSES

We now turn our attention to the random variable[10]

$$\mathbf{Y}_n = 1 + \mathbf{Z}_1 + \cdots + \mathbf{Z}_n \tag{5.1}$$

which equals the total number of descendants up to and including the nth generation and also including the ancestor (zero generation). Letting $n \to \infty$ we get the size of the total progeny which may be finite or infinite. Clearly, for each n the random variable $\mathbf{Y}_n$ is well defined and we denote by R_n the generating function of its probability distribution. Since $\mathbf{Y}_1 = 1 + \mathbf{Z}_1$ we have $R_1(s) = sP(s)$. A recursion formula for R_n can be derived by the argument of the preceding section, the only difference being that to obtain $\mathbf{Y}_n$ we must add the progenitor to the sum of the progenies of the $\mathbf{Z}_1$ members of the first generation. Accordingly

$$R_n(s) = sP(R_{n-1}(s)). \tag{5.2}$$

From this recursion formula it is theoretically possible to calculate successively $R_1, R_2, \ldots,$ but the labor is prohibitive. Fortunately it is possible to discuss the asymptotic behavior of R_n by the geometric argument used in the preceding section to derive the extinction probability x.

First we note that for each $s < 1$, need to have nontrivial branch process

$$R_2(s) = sP(R_1(s)) < sP(s) = R_1(s) \tag{5.3}$$

and by induction it follows that $R_n(s) < R_{n-1}(s)$. Accordingly $R_n(s)$ decreases monotonically to a limit $\rho(s)$, and the latter satisfies

$$\rho(s) = sP(\rho(s)) \qquad 0 < s < 1. \tag{5.4}$$

From the continuity theorem of XI,6 we know that as limit of probability generating functions ρ is the generating function of a sequence of non-negative numbers ρ_k such that $\sum \rho_k \leq 1$.

It follows from (5.4) that for fixed $s < 1$ the value $\rho(s)$ is a root of the equation

$$t = sP(t). \tag{5.5}$$

---

[10] This section was inspired by I. J. Good, *The number of individuals in a cascade process*, Proc. Cambridge Philos. Soc., vol. 45 (1949), pp. 360–363.

We show that this root is unique. For that purpose we denote again by x the smallest positive root of $x = P(x)$ (so that $x \leq 1$). We observe that $y = sP(t)$ (with s fixed) is a convex function of t and so its graph intersects the line $y = t$ in at most two points. But for $t = 0$ the right side in (5.5) is greater than the left, whereas the inequality is reversed when $t = x$, and also when $t = 1$; thus (5.5) has exactly one root between 0 and x, and no root between x and 1. Accordingly, $\rho(s)$ is uniquely characterized as this root, and we see furthermore that $\rho(s) < x$. But $\rho(1)$ is obviously a root of $t = P(t)$, and since x is the smallest root of this equation it is clear that $\rho(1) = x$. In other words, ρ is an honest probability generating function if, and only if, $x = 1$. We can summarize these findings as follows.

Let ρ_k be the probability that the total progeny consists of k elements.
(a) $\sum \rho_k$ equals the extinction probability x (and $1 - x$ equals the probability of an infinite progeny).
(b) The generating function $\rho(s) = \sum \rho_k s^k$ is given by the unique positive root of (5.5), and $\rho(s) \leq x$.

We know already that with probability one *the total progeny is finite whenever $\mu \leq 1$.* By differentiation of (5.4) it is now seen that its *expectation equals $1/(1 - \mu)$ when $\mu < 1$ and is infinite when $\mu = 1$.*

Examples. (*a*) In example (4.*a*) we had $P(s) = q/(1-ps)$, and (5.5) reduces to the quadratic equation $pt^2 - t + qs = 0$ from which we conclude that

$$\rho(s) = \frac{1 - \sqrt{1 - 4pqs}}{2p}. \tag{5.6}$$

(This generating function occurred also in connection with the first-passage times in XI,3.)

(*b*) *Busy periods.* We turn to a more detailed analysis of the queuing problem mentioned in example (3.*d*). Suppose for simplicity that customers can arrive only one at a time and only at integral-valued epochs.[11] We assume that the arrivals are regulated by Bernoulli trials in such a way that at epoch n a customer arrives with probability p, while with probability $q = 1 - p$ no arrival takes place. A customer arriving when the server is free is served immediately, and otherwise he joins the queue (waiting line). The server continues service without interruption as long as there are customers in the queue requiring service. We suppose finally that the

[11] Following a practice introduced by J. Riordan we use the term epoch for points on the time axis because the alternative terms such as time, moment, etc., are overburdened with other meanings.

successive service times are independent (integral-valued) random variables with a common distribution $\{\beta_k\}$ and generating function $\beta(s) = \sum \beta_k s^k$.

Suppose then that a customer arrives at epoch 0 and finds the server free. His service time starts immediately. If it has duration n, the counter becomes free at epoch n provided that no new customer arrives at epochs $1, 2, \ldots, n$. Otherwise the service continues without interruption. By *busy period* is meant the duration of uninterrupted service commencing at epoch 0. We show how the theory of branching process may be used to analyze the duration of the busy period.

The customer arriving at epoch 0 initiates the busy period and will be called ancestor. The first generation consists of the customers arriving prior to or at the epoch of the termination of the ancestor's service time. If there are no such direct descendants the process stops. Otherwise the direct descendants are served successively, and during their service times their direct descendants join the queue. We have here a branching process such that *the probability x of extinction equals the probability of a termination of the busy period, and the total progeny consists of all customers (including the ancestor) arriving during the busy period.* Needless to say, only queues with $x = 1$ are feasible in practice.

To apply our results we require the generating function $P(s)$ for the number of direct descendants. By definition this number is determined by the random sum $\mathbf{X}_1 + \cdots + \mathbf{X}_\mathbf{N}$ where the $\mathbf{X}_j$ are mutually independent and assume the values 1 and 0 with probabilities p and q, while $\mathbf{N}$ is the length of the ancestor's service time. Thus in the present situation $P(s) = \beta(ps+q)$, and hence $\mu = p\sigma$ where $\sigma = \beta'(1)$ is the expected duration of the service time. It follows that *the busy period is certain to terminate only if $p\sigma \leq 1$. The expected number of customers during a busy period is finite only if $p\sigma < 1$.* In other words, congestion is guaranteed when $p\sigma = 1$, and long queues must be the order of the day unless $p\sigma$ is substantially less than 1.

(*c*) *Duration of the busy period.* The preceding example treats the number of customers during a busy period, but the actual duration of the busy period is of greater practical interest. It can be obtained by the elegant device[12] of considering time units as elements of a branching process. We say that the epoch n has no descendants if no customer arrives at epoch n. If such a customer arrives and his service time lasts r time units, then the epochs $n + 1, \ldots, n + r$ are counted as direct descendants of the epoch. Suppose that at epoch 0 the server is free. A little reflection now shows that

[12] It is due to I. J. Good. See the discussion following Kendall's paper quoted in example (3.*d*).

so, the actual epochs at which that customer is served the branching process originated by the epoch 0 either does not come off at all or else lasts exactly for the duration of the uninterrupted service time initiated by a new customer. The generating function for the number of direct descendants is given by

$$P(s) = q + p\beta(s). \tag{5.7}$$

The root x gives the probability of a termination of the busy period. The total progeny equals 1 with probability q while with probability p it equals the duration of the busy period commencing at epoch 0. The duration of the busy period itself has obviously the generating function given by $\beta(\rho(s))$. Not just $p/\ldots$ because the period needs to be busy, exclude case of no arrivals

total progeny in epoch branching process

6. PROBLEMS FOR SOLUTION

1. The distribution (1.1) of the random sum S_N has mean $E(N)E(X)$ and variance $E(N)\,\mathrm{Var}\,(X) + \mathrm{Var}\,(N)E^2(X)$. Verify this (*a*) using the generating function, (*b*) directly from the definition and the notion of conditional expectations.

2. *Animal trapping* [example (1.*b*)]. If $\{g_n\}$ is a geometric distribution, so is the resulting distribution. If $\{g_n\}$ is a logarithmic distribution [cf. (2.8)], there results a logarithmic distribution with an added term.

3. In N Bernoulli trials, where N is a random variable with a Poisson distribution, the numbers of successes and failures are stochastically independent variables. Generalize this to the multinomial distribution (*a*) directly, (*b*) using multivariate generating functions. [Cf. example IX,(1.*d*).] multinomial gen. f. factors

4. *Randomization.* Let N have a Poisson distribution with mean λ, and let N balls be placed randomly into n cells. Show without calculation that the probability of finding exactly m cells empty is $\binom{n}{m} e^{-\lambda m/n}[1-e^{-\lambda/n}]^{n-m}$.

5. *Continuation.*[13] Show that when a fixed number r of balls is placed randomly into n cells the probability of finding exactly m cells empty equals the coefficient of $e^{-\lambda}\lambda^r/r!$ in the expression above. (*a*) Discuss the connection with moment generating functions (problem 24 of XI, 7). (*b*) Use the result for an effortless derivation of II,(11.7). $s=\lambda$

6. *Mixtures of probability distributions.* Let $\{f_i\}$ and $\{g_i\}$ be two probability distributions, $\alpha > 0$, $\beta > 0$, $\alpha + \beta = 1$. Then $\{\alpha f_i + \beta g_i\}$ is again a probability distribution. Discuss its meaning and the connection with the urn models of V,2. Generalize to more than two distributions. Show that such a mixture can be a compound Poisson distribution.

7. Using generating functions show that in the branching process $\mathrm{Var}\,(X_{n+1}) = \mu\,\mathrm{Var}\,(X_n) + \mu^{2n}\sigma^2$. Using conditional expectations prove the equivalent

see Prob 1

[13] This elegant derivation of various combinatorial formulas by randomizing a parameter is due to C. Domb, *On the use of a random parameter in combinatorial problems*, Proceedings Physical Society, Sec. A., vol. 65 (1952), pp. 305–309.

6. $E s^X = \sum g_j E s^{X_j} = \sum g_j e^{\lambda_j(s-1)}$