

A PROBABILISTIC MODEL FOR THE SURVIVABILITY OF CELLS

ILAN ADLER,*

HYUN-SOO AHN** AND

RICHARD M. KARP,*** *University of California, Berkeley*

SHELDON M. ROSS,**** *University of Southern California*

Abstract

Consider n cells, of which some are target cells, and suppose that each cell has a weight. The cells are killed in a sequential manner, with each currently live cell being the next one killed with a probability proportional to its weight. We study the distribution of the number of cells that are alive at the moment when all the target cells have been killed.

Keywords: Target cell; coupon type; Chernoff bound; simulation

2000 Mathematics Subject Classification: Primary 92D25

Secondary 60C05; 37M05

1. Introduction

Consider n cells, with cell i having weight w_i , that are successively killed in the following manner. If S is the set of currently live cells then, independently of the order in which the cells of S^c were killed, $i \in S$ is the next cell to be killed with probability $w_i / \sum_{j \in S} w_j$. Consequently, the probability that the cells will be killed in the order i_1, \dots, i_n is

$$\prod_{j=1}^n \frac{w_{i_j}}{\sum_{k=j}^n w_{i_k}}.$$

Let 1_j be the indicator function of the event that cell j , $j > r$, is alive when the target cells $1, \dots, r$ have all been killed. We are interested in the properties of $N = \sum_{j=r+1}^n 1_j$, the number of surviving cells when all the target cells have been killed. A possible application for this model is the case in which the target cells are cancerous while the nontarget cells are healthy cells. The model can also be viewed within the framework of the classic coupon collector problem, where $n - N$ represents the number of distinct types of coupon that must be collected before all of the types $1, \dots, r$ have been collected.

In Section 2, we determine formulae for the mean and variance of N and derive simple bounds on the mean for some special cases. In Section 3, we derive a lower bound for

Received 3 November 2004; revision received 18 April 2005.

* Postal address: Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720, USA. Email address: adler@ieor.berkeley.edu

** Current address: Department of Operations and Management Science, Ross School of Business, University of Michigan, Ann Arbor, MI 48109, USA. Email address: hsahn@umich.edu

*** Postal address: Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA. Email address: karp@cs.berkeley.edu

**** Postal address: Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089, USA. Email address: smross@usc.edu

Research supported by the National Science Foundation Grant DMI-9901053 with the University of California.

$P(N \geq k)$ and present a computational procedure as well as an efficient simulation procedure for estimating $P(N \geq k)$. In Section 4, we discuss the asymptotic behavior of the mean and distribution of N for the special case in which all the target cells have the same weight and all the nontarget cells have the same weight. We also obtain sharp asymptotic results in this case if we stop when all but a fixed positive fraction of the target cells have been killed. In Section 5, we suppose that the weights are random variables with specified distributions. In the final section, we return to having fixed weights and consider the case in which a stage consists of a probe: each new probe enters a given live cell with a probability proportional to the weight of that cell divided by the sum of the weights of all currently live cells. Supposing that a probe in cell i kills that cell with probability p_i , we compute the expected number of probes needed to kill all the cells $1, \dots, r$, and present an efficient simulation procedure for estimating the distribution of this number.

2. The expected value and variance of N

To study N , consider a model in which cell i is killed at time T_i , where T_1, \dots, T_n are independent exponential random variables with respective rates w_1, \dots, w_n . Using the fact that the remaining life of an exponential random variable conditioned to be greater than some other independent random variable is still exponential (which is easily seen by conditioning on the latter random variable and then using the lack-of-memory property of exponential random variables), it follows that the order in which the cells are killed in this continuous-time model is probabilistically the same as in the original model. Consequently, by letting $T = \max(T_1, T_2, \dots, T_r)$ and, for $J \subseteq \{r + 1, \dots, n\}$, letting $T_J = \min_{j \in J} T_j$ and $1_J = \prod_{j \in J} 1_j$, we find that the event $(1_J = 1)$ is equivalent to the event $(T_J > T)$. This leads directly to the following result.

Lemma 1. *Let*

$$a(w) = \int_0^\infty w e^{-wt} \prod_{i=1}^r (1 - e^{-w_i t}) dt \quad \text{and} \quad w(J) = \sum_{j \in J} w_j.$$

Then

$$P(1_J = 1) = a(w(J)). \tag{1}$$

Proof. Since $(1_J = 1)$ is equivalent to $(T_J > T)$,

$$P(1_J = 1) = P(T_J > T) = \int_0^\infty w(J) e^{-w(J)t} \prod_{i=1}^r (1 - e^{-w_i t}) dt.$$

Lemma 1 immediately yields a proposition.

Proposition 1. (i) $E[N] = \sum_{j=r+1}^n a(w_j)$.

(ii) $\text{var}(N) = \sum_{j=r+1}^n a(w_j)(1 - a(w_j)) + 2 \sum_{j=r+1}^{n-1} \sum_{k=j+1}^n [a(w_j + w_k) - a(w_j)a(w_k)]$.

For the special case in which all the target cells have identical weights we have the following corollary.

Corollary 1. Let $w_i = w_1, i = 1, 2, \dots, r$. Then

(i) for $J \subseteq \{r + 1, \dots, n\}$, with $r(J) = w(J)/w_1$,

$$P(1_J = 1) = \sum_{i=0}^r (-1)^i \binom{r}{i} \frac{r(J)}{r(J) + i} = \prod_{i=1}^r \frac{i}{r(J) + i},$$

and

(ii) with $r_j = w_j/w_1$,

$$E[N] = \sum_{j=r+1}^n \sum_{i=0}^r (-1)^i \binom{r}{i} \frac{r_j}{r_j + i} = \sum_{j=r+1}^n \prod_{i=1}^r \frac{i}{r_j + i}.$$

Proof. By Lemma 1, we have

$$\begin{aligned} P(I_J = 1) &= \int_0^\infty w(J)e^{-w(J)t} (1 - e^{-w_1 t})^r dt \\ &= \int_0^\infty w(J)e^{-w(J)t} \sum_{i=0}^r (-1)^i \binom{r}{i} e^{-i w_1 t} dt \\ &= \sum_{i=0}^r (-1)^i \binom{r}{i} \frac{w(J)}{w(J) + i w_1}. \end{aligned}$$

On the other hand, it directly follows from the lack of memory of exponential random variables that

$$P(1_J = 1) = \prod_{i=1}^r \frac{i w_1}{i w_1 + w(J)}.$$

This proves part (i), from which part (ii) follows immediately.

Our next result yields an upper bound for $E[N]$.

Corollary 2. Let $\bar{w}_1 = (1/r) \sum_{i=1}^r w_i$. Then

(i) for $J \subseteq \{r + 1, \dots, n\}$, with $r(J) = w(J)/\bar{w}_1$,

$$P(1_J = 1) \leq \sum_{i=0}^r (-1)^i \binom{r}{i} \frac{r(J)}{r(J) + i} = \prod_{i=1}^r \frac{i}{r(J) + i},$$

and

(ii) with $r_j = w_j/\bar{w}_1$,

$$E[N] \leq \sum_{j=r+1}^n \sum_{i=0}^r (-1)^i \binom{r}{i} \frac{r_j}{r_j + i} = \sum_{j=r+1}^n \prod_{i=1}^r \frac{i}{r_j + i}.$$

Proof. It is easily verified that $\prod_{i=1}^r (1 - e^{-w_i t})$ is a Schur concave function of w_1, \dots, w_r . Therefore,

$$\prod_{i=1}^r (1 - e^{-w_i t}) \leq (1 - e^{-\bar{w}_1 t})^r$$

and the result follows from Lemma 1 and Corollary 1.

3. The distribution of N

Given (1), it is easy to construct an expression for $P(N \geq k)$. However, such an expression involves a number of terms that is exponential with respect to $n - r$, which makes it impractical for computations. We now present some bounds and computational methods.

Proposition 2. *Let*

$$\bar{w}_2 = \frac{1}{n - r} \sum_{j=r+1}^n w_j.$$

Then, for $k = 1, \dots, n - r$,

$$P(N \geq k) \geq \int_0^\infty \sum_{i=1}^r w_i e^{-w_i t} \prod_{j \neq i, j \leq r} (1 - e^{-w_j t}) \sum_{j=k}^{n-r} \binom{n-r}{j} e^{-j \bar{w}_2 t} (1 - e^{-\bar{w}_2 t})^{(n-r-j)} dt.$$

Proof. We can write

$$P(N \geq k) = \int_0^\infty P(N \geq k \mid T = t) dF_T(t), \quad \text{where } F_T(t) = P(T \leq t).$$

However, $P(N \geq k \mid T = t) = P(\text{the } k\text{th largest of } T_{r+1}, \dots, T_n \text{ is greater than } t)$. The result now follows because (see [1]) the order statistic of a vector of independent exponential random variables with rates $\mathbf{r} = (r_1, \dots, r_m)$ is stochastically smaller than the order statistic of a vector of independent exponential random variables with rates $\mathbf{v} = (v_1, \dots, v_m)$ when \mathbf{v} majorizes \mathbf{r} .

3.1. Approximating $P(N \geq k)$

Let $\Phi(k, t) = P(N \geq k \mid T = t)$ and write

$$P(N \geq k) = \int_0^\infty \Phi(k, t) dF_T(t).$$

For a given integer m and an $\varepsilon > 0$, let us construct a sequence t_0, t_1, \dots, t_{m+1} with $t_0 = 0$, $t_{i+1} = t_i + \varepsilon$, $i = 0, \dots, m - 1$, and $t_{m+1} = \infty$. Since $\Phi(k, t)$ is monotonically decreasing in t , we have

$$\sum_{i=0}^m \Phi(k, t_{i+1})(F_T(t_{i+1}) - F_T(t_i)) \leq \int_0^\infty \Phi(k, t) dF(t) \leq \sum_{i=0}^m \Phi(k, t_i)(F_T(t_{i+1}) - F_T(t_i)).$$

Suppose that we can compute $\Phi(k, t)$ (as we will show below). Then the preceding expression can be used to approximate $P(N \geq k)$ to any desired precision, by choosing a sufficiently large m and a sufficiently small ε .

We can compute $\Phi(k, t)$ by first recursively computing

$$\phi_\ell(\ell, i) = P\left(\sum_{j=r+1}^\ell 1(T_j > t) = i\right),$$

where $1(T_j > t)$ is the indicator function of the event $T_j > t$ and $\ell = r + i, \dots, n$, as follows:

$$\begin{aligned} \phi_t(\ell, 0) &= \prod_{j=r+1}^{\ell} (1 - e^{-w_j t}), & \ell = r + 1, \dots, n, \\ \phi_t(\ell, 1) &= e^{-w_{r+1} t}, & \ell = r + 1, \\ \phi_t(\ell, i) &= \phi_t(\ell - 1, i) P(T_\ell \leq t) + \phi_t(\ell - 1, i - 1) P(T_\ell > t) \\ &= \phi_t(\ell - 1, i)(1 - e^{-w_\ell t}) + \phi_t(\ell - 1, i - 1) e^{-w_\ell t}. \end{aligned}$$

Then $\Phi(k, t) = \sum_{i=k}^{n-r} \phi_t(n, i)$.

3.2. Using simulation to compute $P(N \geq k)$

We now show how to efficiently use simulation to estimate $P(N \geq k)$. To begin, generate the values of T_{r+1}, \dots, T_n , order these values, and let Y_i be the value of the i th largest, $i = 1, \dots, n - r$. Then use the conditional expectation estimator

$$P(N \geq k \mid Y_k = y_k) = \prod_{i=1}^r (1 - e^{-w_i y_k}).$$

This yields the following scheme for estimating $P(N \geq k)$, $k = 1, \dots, n - r$.

1. Generate random numbers U_1, \dots, U_{n-r} .
2. Let

$$T_{r+i} = -\frac{1}{w_{i+r}} \log(U_i), \quad i = 1, \dots, n - r.$$

3. Arrange the values T_{r+1}, \dots, T_n in descending order of size and denote the ordered values by Y_1, \dots, Y_{n-r} , respectively.
4. Let

$$\theta_k = \prod_{i=1}^r (1 - e^{-w_i Y_k}), \quad k = 1, \dots, n - r.$$

This scheme should be repeated many times; the average value of θ_k obtained is the estimate of $P(N \geq k)$, $k = 1, \dots, n - r$.

Remark 1. Note that, because the preceding estimator is a monotone function of T_{r+1}, \dots, T_n , antithetic variables can be used for further variance reduction (see [3]).

4. A special case: uniform weights

In this section, we consider the special case in which all the target weights are equal and all the nontarget weights are equal. Specifically, we assume that

$$w_j = \begin{cases} 1 & \text{if } j = 1, \dots, r, \\ w & \text{if } j = r + 1, \dots, n. \end{cases}$$

Let

$$\begin{aligned} \Gamma(a) &= \int_0^\infty e^{-t} t^{a-1} dt, \\ B(a, b) &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1} (1-t)^{b-1} dt. \end{aligned}$$

Then, for $J \subseteq \{r + 1, \dots, n\}$ and $|J| = k$, we have

$$\begin{aligned} P(1_J = 1) &= \int_0^\infty kw e^{-kwt} (1 - e^{-t})^r dt \\ &= kwB(kw, r + 1) \\ &= rB(kw + 1, r). \end{aligned}$$

Thus, if we set $m = n - r$, we obtain

$$\begin{aligned} E[N] &= mrB(w + 1, r), \\ \text{var}(N) &= mrB(w + 1, r) + m(m - 1)rB(2w + 1, r) - m^2r^2B^2(w + 1, r), \end{aligned}$$

and for the distribution of N we have

$$\begin{aligned} P(N \geq k) &= \int_0^\infty P(N \geq k \mid T = t) dF_T(t) \\ &= \int_0^\infty r e^{-t} (1 - e^{-t})^{r-1} \sum_{j=k}^{n-r} \binom{n-r}{j} e^{-wtj} (1 - e^{-wt})^{n-r-j} dt, \end{aligned}$$

which, by using the binomial expansion, collecting terms, and evaluating the resulting integral, yields

$$P(N \geq k) = r \sum_{j=k}^m \binom{m}{j} \sum_{\ell=0}^{m-j} \binom{m-j}{\ell} \sum_{i=0}^{r-1} \binom{r-1}{i} (-1)^{\ell+i} \frac{1}{(\ell + j)w + i + 1}.$$

Alternatively, we could calculate $P(N \geq k)$ by considering

$$\phi(k, r', n') = P(N \geq k \mid \text{there are } r' \text{ target cells and a total of } n' \text{ cells}).$$

The function ϕ is recursively defined as

$$\begin{aligned} \phi(k, 0, n') &= 1, & n' &= r + k, \dots, n, \\ \phi(k, r', r' + k) &= \prod_{i=1}^{r'} \frac{i}{kw + i}, & r' &= 1, \dots, r, \\ \phi(k, r', n') &= \frac{r'}{(n' - r')w + r'} \phi(k, r' - 1, n' - 1) \\ &\quad + \frac{(n' - r')w}{(n' - r')w + r'} \phi(k - 1, r', n' - 1), \end{aligned}$$

and $P(N \geq k) = \phi(r, k, n)$.

Next we develop bounds for $P(N \geq k)$. Let τ denote the time by which all r target cells (which each live for an exponential time with rate 1) have been killed. Imagine that the $m = n - r$ nontarget cells (which each live for an exponential time with rate w) continue to die even after time τ . Let $N(t)$ denote the number of nontarget cells that are alive at time t . Note that $N(t)$ is a binomial random variable with parameters m and e^{-tw} , and that $N = N(\tau)$.

Fix an $A < r$ and let $t = \ln(r/A)$. Note that

$$(1 - e^{-t})^r = (1 - A/r)^r \quad \text{and} \quad e^{-tw} = (A/r)^w.$$

Then

$$\begin{aligned} P(N \geq k) &= P(N \geq k \mid \tau \leq t) P(\tau \leq t) + P(N \geq k \mid \tau > t) P(\tau > t) \\ &\leq P(\tau \leq t) + P(N(t) \geq k) P(\tau > t) \\ &= \left(1 - \frac{A}{r}\right)^r + P(N(t) \geq k) \left(1 - \left(1 - \frac{A}{r}\right)^r\right). \end{aligned} \tag{2}$$

By letting $S = m(A/r)^w$ and $k = (1 + \delta)S$, and applying to (2) a Chernoff bound for a binomial random variable X , namely

$$P(X \geq (1 + \delta) E[X]) \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}}\right)^{E[X]}$$

(see, for example, [2]), we obtain the inequality

$$P(N \geq (1 + \delta)S) \leq \left(1 - \frac{A}{r}\right)^r + \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}}\right)^S \left(1 - \left(1 - \frac{A}{r}\right)^r\right) < e^{-A} + \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}}\right)^S.$$

Finally, we comment on the asymptotic behavior of N as r and m tend to ∞ . By noting that

$$B(w, r) = \Gamma(w)r^{-w} \left(1 - \frac{\Gamma(w)\Gamma(w-1)}{2r}(1 - O(r^{-1}))\right)$$

(see <http://functions.wolfram.com/GammaBetaErf/Beta/06/02/>), we find that

$$rB(w + 1, r) \sim \Gamma(w + 1)r^{-w} \quad \text{as } r \rightarrow \infty.$$

Thus, for $r \rightarrow \infty$,

$$\begin{aligned} E[N] &\sim \Gamma(w + 1)mr^{-w}, \\ \text{var}(N) &\sim E[N] + (\Gamma(2w + 1) - \Gamma^2(w + 1))m^2r^{-2w}. \end{aligned} \tag{3}$$

For asymptotic results related to the distribution of N , we let $t = \ln(r/B)$, and use the fact that

$$\begin{aligned} P(N \leq k) &= P(N \leq k \mid \tau \leq t) P(\tau \leq t) + P(N \leq k \mid \tau > t) P(\tau > t) \\ &\leq P(N(t) \leq k) \left(1 - \frac{B}{r}\right)^r + 1 - \left(1 - \frac{B}{r}\right)^r. \end{aligned} \tag{4}$$

Applying a Chernoff bound for a binomial(m, p) random variable X and an $a > 0$, namely

$$\max(P(X \geq mp + a), P(X \leq mp - a)) \leq e^{-2a^2/m} \tag{5}$$

(see, for example, [2]), first to (2) and then to (4) gives

$$P\left(N \geq m\left(\frac{A}{r}\right)^w + a\right) \leq \left(1 - \frac{A}{r}\right)^r + e^{-2a^2/m} \tag{6}$$

and

$$P\left(N \leq m\left(\frac{B}{r}\right)^w - a\right) \leq e^{-2a^2/m} \left(1 - \frac{B}{r}\right)^r + 1 - \left(1 - \frac{B}{r}\right)^r. \tag{7}$$

Then, if we substitute $a = \delta mA^w/r^w$ into (6) and $a = \delta m/(r^w A^w)$ into (7), let $A < r$ be a nondecreasing unbounded function of r , and let $B = 1/A$, we can conclude that, for a given $\delta > 0$ and as $r, m \rightarrow \infty$,

- (i) if $\liminf mA^w/r^w = \infty$ then $P(N < (1 + \delta)MA^w/r^w) \rightarrow 1$,
- (ii) if $\liminf m/(rA)^w = \infty$ then $P(N > (1 - \delta)m/(rA)^w) \rightarrow 1$.

Remark 2. (a) If, for a fixed α , $0 < \alpha < 1$, we let $A = \alpha r$, then part (i) is satisfied. Hence, by letting $\varepsilon = (1 + \delta)\alpha^w$ we see that, for any $\varepsilon > 0$, $P(N < \varepsilon m) \rightarrow 1$.

(b) It follows from part (ii) that if $m/r^{2w} \rightarrow \infty$, then $P(N > (1 - \varepsilon)m/r^{2w}) \rightarrow 1$ for any $\varepsilon > 0$.

(c) It follows that $P(N \geq k) \leq \min_{A < r} (e^{-A} + e^{-S}(eS/k)^k)$, where $S = m(A/r)^w$.

(d) Finally, we observe from (3) that if $mr^{-w} \rightarrow 0$ (a condition which is satisfied whenever the conditions of parts (i) and (ii) are violated), then $P(N = 0) \rightarrow 1$.

We can obtain sharp asymptotic results if we stop at the first moment the number of surviving target cells has been reduced to a fraction $\varepsilon > 0$ of its original value. Let N_ε be the number of nontarget cells still surviving at that time. We shall prove that N_ε is concentrated around the value $m\varepsilon^w$.

Proposition 3. For all δ greater than 0, as $r \rightarrow \infty$ and $m \rightarrow \infty$, we have

$$P((1 - \delta)m\varepsilon^w \leq N_\varepsilon \leq (1 + \delta)m\varepsilon^w) \rightarrow 1.$$

Proof. We first show that

$$P(N_\varepsilon \leq (1 + \delta)m\varepsilon^w) \rightarrow 1 \quad \text{as } r, m \rightarrow \infty. \tag{8}$$

To show this, let τ_ε denote the first time by which at least $(1 - \varepsilon)r$ target cells have been killed, meaning that $N_\varepsilon = N(\tau_\varepsilon)$. Let γ be such that $0 < \gamma < \delta$, and let $t = -\ln(\varepsilon(1 + \gamma)^{1/w})$. We will prove (8) by showing that, as r and m approach ∞ ,

- (i) $P(\tau_\varepsilon \leq t) \rightarrow 0$ and
- (ii) $P(N(t) > (1 + \delta)m\varepsilon^w) \rightarrow 0$.

As these convergences imply that

$$P(N(\tau_\varepsilon) \leq (1 + \delta)m\varepsilon^w) \geq P(\tau_\varepsilon > t, N(t) \leq (1 + \delta)m\varepsilon^w) \rightarrow 1,$$

(8) will then follow.

The number, Y say, of surviving target cells at time t is a binomial random variable with parameters r and $e^{-t} = \varepsilon(1 + \gamma)^{1/w}$. Hence, with $a = r\varepsilon[(1 + \gamma)^{1/w} - 1]$, we have

$$P(\tau_\varepsilon \leq t) = P(Y \leq \varepsilon r) = P(Y \leq re^{-t} - a) \leq e^{-2a^2/r},$$

where the inequality follows from the Chernoff bound (5). This proves part (i), because a^2/r goes to ∞ as r goes to ∞ .

To prove part (ii), note that $N(t)$ is a binomial random variable with parameters m and $e^{-wt} = \varepsilon^w(1 + \gamma)$. Hence, by letting $b = m\varepsilon^w(\delta - \gamma)$ and again applying the Chernoff bound (5), we obtain

$$\begin{aligned} P(N(t) > (1 + \delta)m\varepsilon^w) &= P(N(t) > me^{-wt} + b) \\ &\leq e^{-2b^2/m}. \end{aligned}$$

This proves part (ii), because b^2/m goes to ∞ as m goes to ∞ . Thus, we have proved (8).

The proof that

$$P(N_\varepsilon \geq (1 - \delta)m\varepsilon^w) \rightarrow 1$$

is similar, and a combination of the results completes the proof of the proposition.

5. The case of random weights

An important variant of the problem arises when we suppose that the target weights are randomized. That is, with W_i being the weight of cell i , suppose that W_1, \dots, W_n are independent random variables with $W_i, i = 1, \dots, n = r + m$, having some specified distribution F_i . Also suppose that, conditional on $W_i = w_i, i = 1, \dots, n$, the probabilistic model for the order in which cells are killed is as before. Let N denote the number of cells $r + 1, \dots, n$ that are alive when all cells $1, \dots, r$ have been killed.

As in the fixed-weights case, it is helpful to use a continuous-time probabilistic experiment to generate the order in which cells are killed. This can be done as follows.

1. Generate the values of n independent random variables from the distributions F_1, \dots, F_n ; call the generated values w_1, \dots, w_n .
2. Generate n independent exponential random variables $T_i, i = 1, \dots, n$, with T_i having rate w_i . Interpret T_i as the time at which cell i is killed.
3. If $T_{i_1} < T_{i_2} < \dots < T_{i_n}$, take i_1, \dots, i_n as the order in which cells are killed.

Denote the Laplace transform of a distribution R by $L_R(t) = \int_0^\infty e^{-xt} dR(x)$. It follows from our construction that T_1, \dots, T_n will be independent random variables, with T_i having distribution function

$$H_i(t) = P(T_i \leq t) = 1 - L_{F_i}(t). \tag{9}$$

Let $T = \max(T_1, \dots, T_r)$ and

$$H(t) = P(T \leq t) = \prod_{i=1}^r H_i(t).$$

We can derive expressions for $E[N]$ and $\text{var}(N)$ as in the fixed-weights case. For instance,

$$\begin{aligned} E[N] &= \sum_{j>r} P(T_j > T) \\ &= \sum_{j>r} \int_0^\infty H(t) dH_j(t). \end{aligned}$$

However, the computational effort expended in doing so seems prohibitive. Probably a better approach is to use simulation. We present two methods of efficiently using simulation to compute $E[N]$.

One way is to generate all the values $W_i, i = 1, \dots, n$; if $W_i = w_i, i = 1, \dots, n$, use the expression given in Proposition 1 as the estimate from that run. Because N can be expressed as a monotone-nondecreasing function of W_1, \dots, W_r and a monotone-nonincreasing function of W_{r+1}, \dots, W_n , (by writing $T_j = E_j/W_j$, where E_1, \dots, E_n are independent exponential random variables, also independent of W_1, \dots, W_n , with rates 1), we can further improve the efficiency of the simulation estimator by using $\sum_{i=1}^r W_i$ and $\sum_{i=r+1}^n W_i$ as control variables. (By improved efficiency we mean that the new estimator would remain unbiased but have a smaller variance. See [3] for details on simulation using control variables.)

A second way to use simulation to estimate $E[N]$ is to generate the value of T (this will usually require generating the values of T_1, \dots, T_r). If $T = t$, use the estimator

$$E[N | T = t] = \sum_{j>r} L_{F_j}(t).$$

Now consider two alternative scenarios. In scenario 1, the weights are independent random variables from the distributions F_1, \dots, F_n . In scenario 2, they are instead from the distributions G_1, \dots, G_n . It then follows from (9), upon using standard coupling results concerning stochastic ordering (see [4]), that if

$$L_{F_i}(t) \leq L_{G_i}(t), \quad i = 1, \dots, r,$$

and

$$L_{F_i}(t) \geq L_{G_i}(t), \quad i = r + 1, \dots, n, \quad \text{for all } t,$$

then the number of surviving normal cells in scenario 1 will be stochastically larger than the number in scenario 2. Similarly, if

$$L_{F_i}(t) \geq L_{G_i}(t), \quad i = 1, \dots, r, \quad \text{and} \quad L_{F_i}(t) \leq L_{G_i}(t), \quad i = r + 1, \dots, n,$$

then the number of surviving normal cells in scenario 1 will be stochastically smaller than the number in scenario 2.

Let $R(w, k)$ be an Erlang distribution with parameters $(k/w, k)$, where k is positive integer (that is, $R(w, k)$ is the distribution of the sum of k independent and identically distributed exponential random variables having rate k/w). Thus, for $W \sim R(w, k)$, we have $E[W] = w$ and $\text{var}(W) = w^2/k$. Since $L_{R(w,k)}(t) = (1 + wt/k)^{-k}$, it follows that $L_{R(w,k)}(t)$ is decreasing in k for all nonnegative w and t . To illustrate the preceding stochastic ordering result, we consider a simulation in which all the target cells have one weight distribution while all the normal cells have another weight distribution. We use the second suggested simulation scheme to estimate $E[N]$, with the following setup.

- The target cell weights have distribution $R(1, K_T)$, with $K_T = 2, 10, 50, \infty$.
- The normal cell weights have distribution $R(\frac{1}{10}, K_N)$, with $K_N = 2, 10, 50, \infty$.
- The number of target cells is $r = 1000$.
- The number of normal cells is $n - r = 10\,000$.

TABLE 1: The expected number of surviving normal cells.

| K_T | $K_N = \infty$ | $K_N = 50$ | $K_N = 10$ | $K_N = 2$ |
|----------|----------------|------------|------------|-----------|
| ∞ | 4768 | 4793 | 4896 | 5327 |
| 50 | 4579 | 4605 | 4714 | 5176 |
| 10 | 4013 | 4046 | 4177 | 4711 |
| 2 | 2482 | 2528 | 2710 | 3452 |

For each of the 16 possible combinations of weight distributions, we simulate $E[N]$, the expected number of surviving normal cells, with a run of 50 000 repetitions. The results are presented in Table 1. The number reported for the case $K_T = K_N = \infty$, which corresponds to having fixed weights, is calculated using the formula for $E[N]$ presented in Section 4. The results of the simulation are clearly consistent with the theoretical observation that the number of surviving cells in this case is stochastically decreasing with the variability of the target weights and stochastically increasing with the variability of the normal weights. It is also interesting to note that, when the variances of both distributions increase simultaneously, it is the variability of the weights of the target cells that has the greater effect. However, this might be partially due to the choice of the specific parameters (r , n , and w) in the test cases.

Now consider the important special case in which $F_i = F, i \leq r$, and $F_i = G, i > r$. That is, all the target cells have weight distribution F and all the normal cells have weight distribution G . Let $A(t) = 1 - L_F(t)$ and $B(t) = 1 - L_G(t)$ be the distributions of the lifetimes of, respectively, target cells and normal cells. Thus, $F_T(x) = A^r(t)$.

Now, for any t , let $K(t)$ be the number of normal cells killed by time t . Then $K(t)$ has the binomial distribution $\text{binomial}(n - r, B(t))$. Note that if $t_1 < t_2$, then $K(t_2)$ stochastically dominates $K(t_1)$. Let $K = n - r - N$ be the number of normal cells killed before all the target cells have been killed. Then, for any a and t ,

$$\begin{aligned} P(K \geq a) &= P(K \geq a \mid T \geq t)P(T \geq t) + P(K \geq a \mid T < t)P(T < t) \\ &\leq P(T \geq t) + P(K(t) \geq a) \\ &= 1 - A^r(t) + P(\text{binomial}(n - r, B(t)) \geq a). \end{aligned}$$

Similarly,

$$\begin{aligned} P(K \leq a) &= P(K \leq a \mid T \leq t)P(T \leq t) + P(K \leq a \mid T > t)P(T > t) \\ &\leq P(T \leq t) + P(K(t) \leq a) \\ &= A^r(t) + P(\text{binomial}(n - r, B(t)) \leq a). \end{aligned}$$

We thus obtain the following bounds:

$$P(K \geq a) \leq \inf_t [1 - A^r(t) + P(\text{binomial}(n - r, B(t)) \geq a)]$$

and

$$P(K \leq a) \leq \inf_t [A^r(t) + P(\text{binomial}(n - r, B(t)) \leq a)].$$

For any particular cumulative distribution functions F and G , these bounds can be estimated numerically with the help of Chernoff bounds for the binomial distribution.

6. The probe model

Suppose now that whereas a probe will hit a live cell i with probability equal to w_i divided by the sum of weights of all currently live cells, the probe only kills the cell with probability p_i . The results of the previous sections clearly are applicable, with w_k replaced with $p_k w_k$, $k = 1, \dots, n$. Thus, for $j > r$,

$$P(1_j = 1) = \int_0^\infty w_j p_j e^{-p_j w_j t} \prod_{i=1}^r (1 - e^{-p_i w_i t}) dt$$

and

$$E[N] = \sum_{j=r+1}^n \int_0^\infty w_j p_j e^{-p_j w_j t} \prod_{i=1}^r (1 - e^{-p_i w_i t}) dt.$$

An additional random variable of interest in this model is R , the number of probes needed to kill all the target cells $1, \dots, r$.

Proposition 4. *The expectation of R is*

$$E[R] = \sum_{i=1}^n \frac{1}{p_i} - \sum_{j=r+1}^n \int_0^\infty w_j e^{-p_j w_j t} \prod_{i=1}^r (1 - e^{-p_i w_i t}) dt.$$

Proof. Imagine that probing does not end when the cells $1, \dots, r$ have all been killed, but continues until all n cells have been killed, and let Q denote the number of probes required to do so. Also, let R_j denote the number of probes of cell j after cells $1, \dots, r$ have all been killed. Then

$$E[R] = E[Q] - \sum_{j=r+1}^n E[R_j] = \sum_{i=1}^n \frac{1}{p_i} - \sum_{j=r+1}^n P(1_j = 1) \frac{1}{p_j},$$

which completes the proof.

We now show how to efficiently use simulation to estimate $P(R < k + r)$. Suppose that probes of the cells $i, i \geq 1$, occur at times distributed according to independent Poisson processes with rates $w_i, i \geq 1$, with each probe of cell i being a killing probe with probability p_i or a nonkilling probe with probability $1 - p_i$. Then T_1, \dots, T_n , the times of death of cells $1, \dots, n$, are independent exponential random variables with respective rates $p_1 w_1, \dots, p_n w_n$. Let $T = \max(T_1, \dots, T_r)$. Because the processes of nonkilling probes are independent of those of killing probes, it follows that, conditional on $\mathbf{T} = (T_1, \dots, T_n)$, the number of nonkilling probes of live cells by time \mathbf{T} is Poisson distributed with mean

$$\sum_i w_i (1 - p_i) \min(T_i, T).$$

Consequently, conditional on \mathbf{T} , we have $R \stackrel{D}{=} n - N + W$, where W is a Poisson random variable, independent of N , with mean

$$\sum_i w_i (1 - p_i) \min(T_i, T),$$

and ‘ $\stackrel{D}{=}$ ’ denotes equality in distribution.

It follows from this that

$$P(R < k + r \mid T) = P(W < k + r - (n - N) \mid T).$$

Therefore, we propose the following method of estimating $P(R < k + r)$ for each $k \geq 1$.

1. Generate independent exponential random variables T_1, \dots, T_n with respective rates $p_1 w_1, \dots, p_n w_n$.
2. Let $T = \max(T_1, \dots, T_r)$.
3. Let $n - N = r + \sum_{j=r+1}^n 1(T_j < T)$.
4. Let $a = \sum_i w_i(1 - p_i) \min(T_i, T)$.
5. Let

$$\text{est}(k) = \begin{cases} 0 & \text{if } k + r \leq n - N, \\ \sum_{j=0}^m e^{-a} \frac{a^j}{j!} & \text{if } k + r > n - N, \end{cases}$$

where $m = k + r - (n - N) - 1$.

The estimate of $P(R < k + r)$ from this run is $\text{est}(k)$.

Acknowledgement

We thank the referee for helpful comments, including the suggestion to consider the case of random weights.

References

- [1] MARSHALL, A. W. AND OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.
- [2] MOTWANI, R. AND RAGHAVAN, P. (1995). *Randomized Algorithms*. Cambridge University Press.
- [3] ROSS, S. M. (2002). *Simulation*, 3rd edn. Academic Press, Burlington, MA.
- [4] SHAKED, M. AND SHANTHIKUMAR, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, Boston, MA.