

15.2: Some Random Selection Problems

In the unit on [Random Selection](#), we develop some general theoretical results and computational procedures using MATLAB. In this unit, we extend the treatment to a variety of problems. We establish some useful theoretical results and in some cases use MATLAB procedures, including those in the unit on random selection.

The Poisson Decomposition

In many problems, the individual demands may be categorized in one of m types. If the random variable T_i is the type of the i th arrival and the class $\{T_i : 1 \leq i\}$ is iid, we have *multinomial trials*. For $m = 2$ we have the Bernoulli or binomial case, in which one type is called a success and the other a failure.

Multinomial trials

We analyze such a sequence of trials as follows. Suppose there are m types, which we number 1 through m . Let E_{ki} be the event that type k occurs on the i th component trial. For each i , the class $\{E_{ki} : 1 \leq k \leq m\}$ is a partition, since on each component trial exactly one of the types will occur. The type on the i th trial may be represented by the type random variable

$$T_i = \sum_{k=1}^m k I_{E_{ki}}$$

we assume

$$\{T_i : 1 \leq i\} \text{ is iid, with } P(T_i = k) = P(E_{ki}) = p_k \text{ invariant with } i$$

In a sequence of n trials, we let N_{kn} be the number of occurrences of type k . Then

$$N_{kn} = \sum_{i=1}^n I_{E_{ki}} \text{ with } \sum_{k=1}^m N_{kn} = n$$

Now each $N_{kn} \sim \text{binomial}(n, p_k)$. The class $\{N_{kn} : 1 \leq k \leq m\}$ cannot be independent, since it sums to n . If the values of $m - 1$ of them are known, the value of the other is determined. If $n_1 + n_2 + \cdots + n_m = n$, the event

$$\{N_{1n} = n_1, N_{2n} = n_2, \dots, N_{mn} = n_m\}$$

is one of the

$$C(n; n_1, n_2, \dots, n_m) = n! / (n_1! n_2! \cdots n_m!)$$

ways of arranging n_1 of the E_{1i} , n_2 of the E_{2i} , \dots , n_m of the E_{mi} . Each such arrangement has probability $p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$, so that

$$P(N_{1n} = n_1, N_{2n} = n_2, \dots, N_{mn} = n_m) = n! \prod_{k=1}^m \frac{p_k^{n_k}}{n_k!}$$

This set of joint probabilities constitutes the *multinomial distribution*. For $m = 2$, and type 1 a success, this is the binomial distribution with parameter (n, p_1) .

A random number of multinomial trials

We consider, in particular, the case of a random number N of multinomial trials, where $N \sim \text{Poisson}(\mu)$. Let N_k be the number of results of type k in a random number N of multinomial trials.

$$N_k = \sum_{i=1}^N I_{E_{ki}} = \sum_{n=1}^{\infty} I_{\{N=n\}} N_{kn} \text{ with } \sum_{k=1}^m N_k = N$$

Poisson decomposition

Suppose

$$N \sim \text{Poisson}(\mu)$$

$$\{T_i : 1 \leq i\} \text{ is iid with } P(T_i = k) = p_k, 1 \leq k \leq m$$

$$\{N, T_i : 1 \leq i\} \text{ is independent}$$

Then

$$\text{Each } N_k \sim \text{Poisson}(\mu p_k)$$

$$\{N_k : 1 \leq k \leq m\} \text{ is independent.}$$

—□

The usefulness of this remarkable result is enhanced by the fact that the sum of independent Poisson random variables is also Poisson, with μ for the sum the sum of the μ_i for the variables added. This is readily established with the aid of the generating function. Before verifying the propositions above, we consider some examples.

Example 15.2.1 A shipping problem

The number N of orders per day received by a mail order house is Poisson (300). Orders are shipped by next day express, by second day priority, or by regular parcel mail. Suppose 4/10 of the customers want next day express, 5/10 want second day priority, and 1/10 require regular mail. Make the usual assumptions on compound demand. What is the probability that fewer than 150 want next day express? What is the probability that fewer than 300 want one or the other of the two faster deliveries?

Solution

Model as a random number of multinomial trials, with three outcome types: Type 1 is next day express, Type 2 is second day priority, and Type 3 is regular mail, with respective probabilities $p_1 = 0.4$, $p_2 = 0.5$, and $p_3 = 0.1$. The $N_1 \sim \text{Poisson}(0.4 \cdot 300 = 120)$, $N_2 \sim \text{Poisson}(0.5 \cdot 300 = 150)$, and $N_3 \sim \text{Poisson}(0.1 \cdot 300 = 30)$. Also $N_1 + N_2 \sim \text{Poisson}(120 + 150 = 270)$.

```
P1 = 1 - cpoisson(120,150)
P1 = 0.9954
P12 = 1 - cpoisson(270,300)
P12 = 0.9620
```

Example 15.2.2 Message routing

A junction point in a network has two incoming lines and two outgoing lines. The number of incoming messages N_1 on line one in one hour is Poisson (50); on line 2 the number is $N_2 \sim \text{Poisson}(45)$. On incoming line 1 the messages have probability $P_{1a} = 0.33$ of leaving on outgoing line a and $1 - p_{1a}$ of leaving on line b. The messages coming in on line 2 have probability $P_{2a} = 0.47$ of leaving on line a. Under the usual independence assumptions, what is the distribution of outgoing messages on line a? What are the probabilities of at least 30, 35, 40 outgoing messages on line a?

Solution

By the Poisson decomposition, $N_a \sim \text{Poisson}(50 \cdot 0.33 + 45 \cdot 0.47 = 37.65)$.

```
ma = 50*0.33 + 45*0.47
ma = 37.6500
Pa = cpoisson(ma,30:5:40)
Pa = 0.9119 0.6890 0.3722
```

VERIFICATION of the Poisson decomposition

$$N_k = \sum_{i=1}^N I_{Eki}.$$

This is composite demand with $Y_k = I_{Eki}$, so that $g_{Y_k}(s) = q_k + sp_k = 1 + p_k(s - 1)$. Therefore,

$$g_{N_k}(s) = g_N[g_{Y_k}(s)] = e = e$$

which is the generating function for $N_k \sim \text{Poisson}(\mu p_k)$.

For any n_1, n_2, \dots, n_m , let $n = n_1 + n_2 + \dots + n_m$, and consider

$$A = \{N_1 = n_1, N_2 = n_2, \dots, N_m = n_m\} = \{N = n\} \cap \{N_{1n} = N_1, N_{2n} = n_2, \dots, N_{mn} = n_m\}$$

Since N is independent of the class of I_{Eki} , the class

$$\{\{N = n\}, \{N_{1n} = n_1, N_{2n} = n_2, \dots, N_{mn} = n_m\}\}$$

is independent. By the product rule and the multinomial distribution

$$P(A) = e^{-\mu} \frac{\mu^n}{n!} \cdot n! \prod_{k=1}^m \frac{p_k^{n_k}}{(n_k)!} = \prod_{k=1}^m e^{-\mu p_k} \frac{p_k^{n_k}}{n_k!} = \prod_{k=1}^m P(N_k = n_k)$$

The second product uses the fact that

$$e^\mu = e^{\mu(p_1+p_2+\dots+p_m)} = \prod_{k=1}^m e^{\mu p_k}$$

Thus, the product rule holds for the class

Extreme values

Consider an iid class $\{Y_i : 1 \leq i\}$ of nonnegative random variables. For any positive integer n we let

$$V_n = \min \{Y_1, Y_2, \dots, Y_n\} \text{ and } W_n = \max \{Y_1, Y_2, \dots, Y_n\}$$

Then

$$P(V_n > t) = P^n(Y > t) \text{ and } P(W_n \leq t) = P^n(Y \leq t)$$

Now consider a random number N of the Y_i . The minimum and maximum random variables are

$$V_N = \sum_{n=0}^{\infty} I_{\{N=n\}} V_n \text{ and } W_N = \sum_{n=0}^{\infty} I_{\{N=n\}} W_n$$

—□

Computational formulas

If we set $V_0 = W_0 = 0$, then

$$F_V(t) = P(V \leq t) = 1 + P(N=0) - g_N[P(Y > t)]$$

$$F_W(t) = g_N[P(Y \leq t)]$$

These results are easily established as follows. $\{V_N > t\} = \bigvee_{n=0}^{\infty} \{N=n\} \{V_n > t\}$. By additivity and independence of $\{N, V_n\}$ for each n

$$P(V_N > t) = \sum_{n=0}^{\infty} P(N=n)P(V_n > t) = \sum_{n=1}^{\infty} P(N=n)P^n(Y > t) \text{ , since } P(V_0 > t) = 0$$

If we add into the last sum the term $P(N=0)P^0(Y > t) = P(N=0)$ then subtract it, we have

$$P(V_N > t) = \sum_{n=0}^{\infty} P(N=n)P^n(Y > t) - P(N=0) = g_N[P(Y > t)] - P(N=0)$$

A similar argument holds for proposition (b). In this case, we do not have the extra term for $\{N=0\}$, since $P(W_0 \leq t) = 1$.

Special case. In some cases, $N=0$ does not correspond to an admissible outcome (see [Example 14.2.4](#), below, on lowest bidder and [Example 14.2.6](#)). In that case

$$F_V(t) = \sum_{n=1}^{\infty} P(V_n \leq t)P(N=n) = \sum_{n=1}^{\infty} [1 - P^n(Y > t)]P(N=n) = \sum_{n=1}^{\infty} P(N=n) - \sum_{n=1}^{\infty} P^n(Y > t)P(N=n)$$

Add $P(N=0) = P^0(Y > t)P(N=0)$ to each of the sums to get

$$F_V(t) = 1 - \sum_{n=0}^{\infty} P^n(Y > t)P(N=n) = 1 - g_N[P(Y > t)]$$

—□

Example 15.2.3 Maximum service time

The number N of jobs coming into a service center in a week is a random quantity having a Poisson (20) distribution. Suppose the service times (in hours) for individual units are iid, with common distribution exponential (1/3). What is the probability the maximum service time for the units is no greater than 6, 9, 12, 15, 18 hours?

Solution

$$P(W_N \leq t) = g_N[P(Y \leq t)] = e^{20[F_Y(t)-1]} = \exp(-20e^{-t/3})$$

```
t = 6:3:18;
PW = exp(-20*exp(-t/3));
disp([t;PW]')
```

6.0000	0.0668
9.0000	0.3694
12.0000	0.6933
15.0000	0.8739
18.0000	0.9516

Example 15.2.4 Lowest Bidder

A manufacturer seeks bids on a modification of one of his processing units. Twenty contractors are invited to bid. They bid with probability 0.3, so that the number of bids $N \sim \text{binomial}(20, 0.3)$. Assume the bids Y_i (in thousands of dollars) form an iid class. The market is such that the bids have a common distribution symmetric triangular on (150, 250). What is the probability of at least one bid no greater than 170, 180, 190, 200, 210? Note that no bid is *not* a low bid of zero, hence we must use the special case.

Solution

$$P(V \leq t) = 1 - g_N[P(Y > t)] = 1 - (0.7 + 0.3p)^{20} \quad \text{where } p = P(Y > t)$$

Solving graphically for $p = P(Y > t)$, we get

$$p = [23/25 \ 41/50 \ 17/25 \ 1/2 \ 8/25] \text{ for } t = [170 \ 180 \ 190 \ 200 \ 210]$$

Now $g_N(s) = (0.7 + 0.3s)^{20}$. We use MATLAB to obtain

```
t = [170 180 190 200 210];
p = [23/25 41/50 17/25 1/2 8/25];
PV = 1 - (0.7 + 0.3*p).^20;
disp([t;p;PV]')
170.0000    0.9200    0.3848
180.0000    0.8200    0.6705
190.0000    0.6800    0.8671
200.0000    0.5000    0.9612
210.0000    0.3200    0.9896
```

Example 15.2.5 Example 15.2.4 with a general counting variable

Suppose the number of bids is 1, 2 or 3 with probabilities 0.3, 0.5, 0.2, respectively.

Determine $P(V \leq t)$ in each case.

Solution

The minimum of the selected Y 's is no greater than t if and only if there is at least one Y less than or equal to t . We determine in each case probabilities for the number of bids satisfying $Y \leq t$. For each t , we are interested in the probability of one or more occurrences of the event $Y \leq t$. This is essentially the problem in [Example 7](#) from "[Random Selection](#)", with probability $p = P(Y \leq t)$.

```
t = [170 180 190 200 210];
p = [23/25 41/50 17/25 1/2 8/25]; % Probabilities Y <= t are 1 - p
gN = [0 0.3 0.5 0.2];           % Zero for missing value
PV = zeros(1,length(t));
for i=1:length(t)
    gY = [p(i), 1 - p(i)];
    [d, pd] = gendf(gN, gY);
```

```
PV(i) = (d>0)*pd';
end
disp([t;PV]')
170.0000    0.1451
180.0000    0.3075
190.0000    0.5019
200.0000    0.7000
210.0000    0.8462
```

% Selects positions for $d > 0$ and
% adds corresponding probabilities

Example 15.2.4 may be worked in this manner by using `gN = ibinom(20,0.3,0:20)`. The results, of course, are the same as in the previous solution. The fact that the probabilities in this example are lower for each t than in Example 15.2.4 reflects the fact that there are probably fewer bids in each case.

Example 15.2.6 Batch testing

Electrical units from a production line are first inspected for operability. However, experience indicates that a fraction p of those passing the initial operability test are defective. All operable units are subsequently tested in a batch under continuous operation (a "burn in" test). Statistical data indicate the defective units have times to failure Y_i iid, exponential (λ), whereas good units have very long life (infinite from the point of view of the test). A batch of n units is tested. Let V be the time of the first failure and N be the number of defective units in the batch. If the test goes t units of time with no failure (i.e., $V > t$), what is the probability of no defective units?

Solution

Since no defective units implies no failures in any reasonable test time, we have

$$\{N = 0\} \subset \{V > t\} \text{ so that } P(N = 0|V > t) = \frac{P(N = 0)}{P(V > t)}$$

Since $N = 0$ does not yield a minimum value, we have $P(V > t) = g_N[P(Y > t)]$. Now under the condition above, the number of defective units $N \sim \text{binomial}(n, p)$, so that $g_N(s) = (q + ps)^n$. If N is large and p is reasonably small, N is approximately Poisson (np) with $g_N(s) = e^{np(s-1)}$ and $P(N = 0) = e^{-np}$. Now $P(Y > t) = e^{-\lambda t}$; for large n

$$P(N = 0|V > t) = \frac{e^{-np}}{e^{np[P(Y > t) - 1]}} = e^{-npP(Y > t)} = e^{-npe^{-\lambda t}}$$

For $n = 5000$, $p = 0.001$, $\lambda = 2$, and $t = 1, 2, 3, 4, 5$, MATLAB calculations give

```
t = 1:5;
n = 5000;
p = 0.001;
lambda = 2;
P = exp(-n*p*exp(-lambda*t));
disp([t;P]')
1.0000    0.5083
2.0000    0.9125
3.0000    0.9877
4.0000    0.9983
5.0000    0.9998
```

It appears that a test of three to five hours should give reliable results. In actually designing the test, one should probably make calculations with a number of different assumptions on the fraction of defective units and the life duration of defective units. These calculations are relatively easy to make with MATLAB.

Bernoulli trials with random execution times or costs

Consider a Bernoulli sequence with probability p of success on any component trial. Let N be the number of the trial on which the first success occurs. Let Y_i be the time (or cost) to execute the i th trial. Then the total time (or cost) from the beginning to the completion of the first success is

$$T = \sum_{i=1}^N Y_i \text{ (composite "demand" with } N - 1 \sim \text{geometric } p)$$

We suppose the Y_i form an iid class, independent of N . Now $N - 1 \sim \text{geometric } (p)$ implies $g_N(s) = ps/(1 - qs)$, so that

$$M_T(s) = g_N[M_Y(s)] = \frac{pM_Y(s)}{1 - qM_Y(s)}$$

There are two useful special cases:

$$Y_i \sim \text{exponential } (\lambda), \text{ so that } M_Y(s) = \frac{\lambda}{\lambda - s}.$$

$$M_T(s) = \frac{p\lambda}{p\lambda - s} = \frac{p}{p - s/\lambda},$$

which implies $T \sim \text{exponential } (p\lambda)$.

$$Y_i - 1 \sim \text{geometric } (p_0), \text{ so that } g_Y(s) = \frac{p_0}{1 - qs}$$

$$g_T(s) = \frac{p\lambda/(\lambda - s)}{1 - q\lambda/(\lambda - s)} = \frac{p\lambda}{p\lambda - s}$$

so that $T - 1 \sim \text{geometric } (pp_0)$.

Example 15.2.7 Job interviews

Suppose a prospective employer is interviewing candidates for a job from a pool in which twenty percent are qualified. Interview times (in hours) Y_i are presumed to form an iid class, each exponential (3). Thus, the average interview time is 1/3 hour (twenty minutes). We take the probability for success on any interview to be $p = 0.2$. What is the probability a satisfactory candidate will be found in four hours or less? What is the probability the maximum interview time will be no greater than 0.5, 0.75, 1, 1.25, 1.5 hours?

Solution

$T \sim \text{exponential } (0.2 \cdot 3 = 0.6)$, so that $P(T \leq 4) = 1 - e^{-0.6 \cdot 4} = 0.9093$.

$$P(W \leq t) = g_N[P(Y \leq t)] = \frac{0.2(1 - e^{-3t})}{1 - 0.8(1 - e^{-3t})} = \frac{1 - e^{-3t}}{1 + 4e^{-3t}}$$

MATLAB computations give

```
t = 0.5:0.25:1.5;
Pwt = (1 - exp(-3*t))./(1 + 4*exp(-3*t));
disp([t;Pwt]')
0.5000    0.4105
0.7500    0.6293
1.0000    0.7924
1.2500    0.8925
1.5000    0.9468
```

The average interview time is 1/3 hour; with probability 0.63 the maximum is 3/4 hour or less; with probability 0.79 the maximum is one hour or less; etc.

In the general case, solving for the distribution of T requires transform theory, and may be handled best by a program such as Maple or Mathematica.

For the case of *simple* Y_i we may use approximation procedures based on properties of the geometric series. Since $N - 1 \sim \text{geometric}(p)$.

$$\begin{aligned} g_N 9s) &= \frac{ps}{1 - qs} = ps \sum_{k=0}^{\infty} (qs)^k = ps \left[\sum_{k=0}^n (qs)^k + \sum_{k=n+1}^{\infty} (qs)^k \right] = ps \left[\sum_{k=0}^n (qs)^k + (qs)^{n+1} \sum_{k=0}^{\infty} (qs)^k \right] \\ &= ps \left[\sum_{k=0}^n (qs)^k \right] + (qs)^{n+1} g_N 9s) = g_n(s) + (qs)^{n+1} g_N(s) \end{aligned}$$

Note that $g_n(s)$ has the form of the generating function for a simple approximation N_n which matches values and probabilities with N up to $k = n$. Now

$$g_T(s) = g_n[g_Y(s)] + (qs)^{n+1} g_N[g_Y(s)]$$

The evaluation involves convolution of coefficients which effectively sets $s = 1$. Since $g_N(1) = g_Y(1) = 1$.

$$(qs)^{n+1} g_N[g_Y(s)] \text{ for } s = 1 \text{ reduces to } q^{n+1} = P(N > n)$$

which is negligible if n is large enough. Suitable n may be determined in each case. With such an n , if the Y_i are nonnegative, integer-valued, we may use the *gend* procedure on $g_n[g_Y(s)]$, where

$$g_n(s) = ps + pqs^2 + pq^2s^3 + \dots + pq^n s^{n+1}$$

For the integer-valued case, as in the general case of simple Y_i , we could use *mgd*. However, *gend* is usually faster and more efficient for the integer-valued case. Unless q is small, the number of terms needed to approximate g_n is likely to be too great.

Example 15.2.8 Approximating the generating function

Let $p = 0.3$ and Y be uniformly distributed on $\{1, 2, \dots, 10\}$. Determine the distribution for

$$T = \sum_{k=1}^N Y_k$$

Solution

```
p = 0.3;
q = 1 - p;
a = [30 35 40];          % Check for suitable n
b = q.^a
b = 1.0e-04 *            % Use n = 40
    0.2254    0.0379    0.0064
n = 40;
k = 1:n;
gY = 0.1*[0 ones(1,10)];
gN = p*[0 q.^(k-1)];     % Probabilities, 0 <= k <= 40
gend
Do not forget zero coefficients for missing powers
Enter gen fn COEFFICIENTS for gN  gN
Enter gen fn COEFFICIENTS for gY  gY
Values are in row matrix D; probabilities are in PD.
To view the distribution, call for gD.
sum(PD)                  % Check sum of probabilities
ans = 1.0000
FD = cumsum(PD);         % Distribution function for D
plot(0:100,FD(1:101))    % See Figure 15.2.1
P50 = (D<=50)*PD'
P50 = 0.9497
P30 = (D<=30)*PD'
P30 = 0.8263
```

Figure one is a graph labeled, execution time distribution function. The horizontal axis is labeled, Time, and the vertical axis is labeled, probability. The values on the horizontal axis range from 0 to 100 in increments of 10. The values on the vertical axis range from 0 to 1 in increments of 0.1. There is one plotted distribution function on this graph. It begins in the bottom-left corner, at the point (0, 0), and moves right at a strong positive slope. As the plot moves from left to right, the slope decreases as the function increases. About midway across the graph horizontally, the plot is nearly at the top, at a probability value above 0.9. The plot continues to increase at a decreasing rate until it tapers off to a horizontal line by the point (80, 1), at which it continues and terminates at the top-right corner.

Figure 15.2.1. Execution Time Distribution Function F_D .

The same results may be achieved with mgd, although at the cost of more computing time. In that case, use gN as in Example 15.2.8, but use the actual distribution for Y .

Arrival times and counting processes

Suppose we have phenomena which take place at discrete instants of time, separated by random waiting or interarrival times. These may be arrivals of customers in a store, of noise pulses on a communications line, vehicles passing a position on a road, the failures of a system, etc. We refer to these occurrences as *arrivals* and designate the times of occurrence as *arrival times*. A stream of arrivals may be described in three equivalent ways.

- *Arrival times*: $\{S_n : 0 \leq n\}$, with $0 = S_0 < S_1 < \dots$ a.s. (basic sequence)
- *Interarrival times*: $\{W_i : 1 \leq i\}$, with each $W_i > 0$ a.s. (incremental sequence)

The strict inequalities imply that with probability one there are no simultaneous arrivals. The relations between the two sequences are simply

$$S_0 = 0, S_n = \sum_{i=1}^n W_i \text{ and } W_n = S_n - S_{n-1} \text{ for all } n \geq 1$$

The formulation indicates the essential equivalence of the problem with that of the [compound demand](#). The notation and terminology are changed to correspond to that customarily used in the treatment of arrival and counting processes.

The stream of arrivals may be described in a third way.

- *Counting processes*: $N_t = N(t)$ is the number of arrivals in time period $(0, t]$. It should be clear that this is a random quantity for each nonnegative t . For a given t, ω the value is $N(t, \omega)$. Such a family of random variables constitutes a *random process*. In this case the random process is a *counting process*.

We thus have three equivalent descriptions for the stream of arrivals.

$$\{S_n : 0 \leq n\} \quad \{W_n : 1 \leq n\} \quad \{N_t : 0 \leq t\}$$

Several properties of the counting process N should be noted:

$N(t+h) - N(t)$ counts the arrivals in the interval $(t, t+h]$, $h > 0$, so that $N(t+h) \geq N(t)$ for $h > 0$.

$N_0 = 0$ and for $t > 0$ we have

$$N_t = \sum_{i=1}^{\infty} I_{(0,t]}(S_i) = \max \{n : S_n \leq t\} = \min \{n : S_{n+1} > t\}$$

For any given ω , $N(\cdot, \omega)$ is a nondecreasing, right-continuous, integer-valued function defined on $[0, \infty)$, with $N(0, \omega) = 0$.

The essential relationships between the three ways of describing the stream of arrivals is displayed in

$$W_n = S_n - S_{n-1}, \{N_t \geq n\} = \{S_n \leq t\}, \{N_t = n\} = \{S_n \leq t < S_{n+1}\}$$

This implies

$$P(N_t = n) = P(S_n \leq t) - P(S_{n+1} \leq t) = P(S_{n+1} > t) - P(S_n > t)$$

Although there are many possibilities for the interarrival time distributions, we assume

$$\{W_i : 1 \leq i\} \text{ is iid, with } W_i > 0 \text{ a.s.}$$

Under such assumptions, the counting process is often referred to as a *renewal process* and the interarrival times are called *renewal times*. In the literature on renewal processes, it is common for the random variable to count an arrival at $t = 0$. This requires an adjustment of the expressions relating N_t and the S_i . We use the convention above.

Exponential iid interarrival times

The case of exponential interarrival times is natural in many applications and leads to important mathematical results. We utilize the following propositions about the arrival times S_n , the interarrival times W_i , and the counting process N .

If $\{W_i : 1 \leq i\}$ is iid exponential (λ) , then $S_n \sim \text{gamma}(n, \lambda)$ for all $n \geq 1$. This is worked out in the unit on TRANSFORM METHODS, in the discussion of the connection between the gamma distribution and the exponential distribution.

$S_n \sim \text{gamma}(n, \lambda)$ for all $n \geq 1$, and $S_0 = 0$, iff $N_t \sim \text{Poisson}(\lambda t)$ for all $t > 0$. This follows the result in the unit DISTRIBUTION APPROXIMATIONS on the relationship between the Poisson and gamma distributions, along with the fact that $\{N_t \geq n\} = \{S_n \leq t\}$.

Remark. The counting process is a Poisson process in the sense that $N_t \sim \text{Poisson}(\lambda t)$ for all $t > 0$. More advanced treatments show that the process has independent, stationary increments. That is

$$N(t+h) - N(t) = N(h) \quad \text{for all } t, h > 0, \text{ and}$$

For $t_1 < t_2 \leq t_3 < t_4 \leq \dots \leq t_{m-1} < t_m$, the class $\{N(t_2) - N(t_1), N(t_4) - N(t_3), \dots, N(t_m) - N(t_{m-1})\}$ is independent.

In words, the number of arrivals in any time interval depends upon the length of the interval and not its location in time, and the numbers of arrivals in nonoverlapping time intervals are independent.

Example 15.2.9 Emergency calls

Emergency calls arrive at a police switchboard with interarrival times (in hours) exponential (15). Thus, the average interarrival time is 1/15 hour (four minutes). What is the probability the number of calls in an eight hour shift is no more than 100, 120, 140?

```
p = 1 - cpoisson(8*15, [101 121 141])
p = 0.0347    0.5243    0.9669
```

We develop next a simple computational result for arrival processes for which $S_n \sim \text{gamma}(n, \lambda)$

Example 15.2.10 Gamma arrival times

Suppose the arrival times $S_n \sim \text{gamma}(n, \lambda)$ and g is such that

$$\int_0^\infty |g| < \infty \text{ and } E[\sum_{n=1}^\infty |g(S_n)|] < \infty$$

Then

$$E[\sum_{n=1}^\infty g(S_n)] = \lambda \int_0^\infty g$$

VERIFICATION

We use the countable sums property (E8b) for expectation and the corresponding property for integrals to assert

$$E[\sum_{n=1}^\infty g(S_n)] = \sum_{n=1}^\infty E[g(S_n)] = \sum_{n=1}^\infty \int_0^\infty g(t) f_n(t) dt \quad \text{where } f_n(t) = \frac{\lambda e^{-\lambda t} (\lambda t)^{n-1}}{(n-1)!}$$

We may apply (E8b) to assert

$$\sum_{n=1}^\infty \int_0^\infty g f_n = \int_0^\infty g \sum_{n=1}^\infty f_n$$

Since

$$\sum_{n=1}^\infty f_n(t) = \lambda e^{-\lambda t} \sum_{n=1}^\infty \frac{(\lambda t)^{n-1}}{(n-1)!} = \lambda e^{-\lambda t} e^{\lambda t} = \lambda$$

the proposition is established.

Example 15.2.11 Discounted replacement costs

A critical unit in a production system has life duration exponential (λ) . Upon failure the unit is replaced immediately by a similar unit. Units fail independently. Cost of replacement of a unit is c dollars. If money is discounted at a rate α , then a dollar

spent t units of time in the future has a current value $e^{-\alpha t}$. If S_n is the time of replacement of the n th unit, then $S_n \sim \text{gamma}(n, \lambda)$ and the present value of all future replacements is

$$C = \sum_{n=1}^{\infty} c e^{-\alpha S_n}$$

The expected replacement cost is

$$E[C] = E\left[\sum_{n=1}^{\infty} g(S_n)\right] \text{ where } g(t) = c e^{-\alpha t}$$

Hence

$$E[C] = \lambda \int_0^{\infty} c e^{-\alpha t} dt = \frac{\lambda c}{\alpha}$$

Suppose unit replacement cost $c = 1200$, average time (in years) to failure $1/\lambda = 1/4$, and the discount rate per year $\alpha = 0.08$ (eight percent). Then

$$E[C] = \frac{1200 \cdot 4}{0.08} = 60,000$$

Example 15.2.12 Random costs

Suppose the cost of the n th replacement in Example 15.2.11 is a random quantity C_n , with $\{C_n, S_n\}$ independent and $E[C_n] = c$, invariant with n . Then

$$E[C] = E\left[\sum_{n=1}^{\infty} C_n e^{-\alpha S_n}\right] = \sum_{n=1}^{\infty} E[C_n] E[e^{-\alpha S_n}] = \sum_{n=1}^{\infty} c E[e^{-\alpha S_n}] = \frac{\lambda c}{\alpha}$$

The analysis to this point assumes the process will continue endlessly into the future. Often, it is desirable to plan for a specific, finite period. The result of [Example 15.2.10](#) may be modified easily to account for a finite period, often referred to as a *finite horizon*.

Example 15.2.13 Finite horizon

Under the conditions assumed in [Example 15.2.10](#), above, let N_t be the counting random variable for arrivals in the interval $(0, t]$.

$$\text{If } Z_t = \sum_{n=1}^{N_t} g(S_n), \text{ then } E[Z_t] = \lambda \int_0^t g(u) du$$

VERIFICATION

Since $N_t \geq n$ iff $S_n \leq t$. $\sum_{n=1}^{N_t} g(S_n) = \sum_{n=0}^{\infty} I_{(0,t]}(S_n) g(S_n)$. In the result of [Example 15.2.10](#), replace g by $I_{(0,t]}g$ and note that

$$\int_0^{\infty} I_{(0,t]}(u) g(u) du = \int_0^t g(u) du$$

Example 15.2.14 Replacement costs, finite horizon

Under the condition of Example 15.2.11, consider the replacement costs over a two-year period.

Solution

$$E[C] = \lambda c \int_0^t e^{-\alpha u} du = \frac{\lambda c}{\alpha} (1 - e^{-\alpha t})$$

Thus, the expected cost for the infinite horizon $\lambda c / \alpha$ is reduced by the factor $1 - e^{-\alpha t}$. For $t = 2$ and the number in Example 15.2.11, the reduction factor is $1 - e^{-0.16} = 0.1479$ to give $E[C] = 60000 \cdot 0.1479 = 8871.37$

In the important special case that $g(u) = c e^{-\alpha u}$, the expression for $E[\sum_{n=1}^{\infty} g(S_n)]$ may be put into a form which does not require the interarrival times to be exponential.

Example 15.2.15 General interarrival, exponential g

Suppose $S_0 = 0$ and $S_n = \sum_{i=1}^n W_i$, where $\{W_i : 1 \leq i\}$ is iid. Let $\{V_n : 1 \leq n\}$ be a class such that each $E[V_n] = c$ and each pair $\{V_n, S_n\}$ is independent. Then for $\alpha > 0$

$$E[C] = E[\sum_{n=1}^{\infty} V_n e^{-\alpha S_n}] = c \cdot \frac{M_W(-\alpha)}{1 - M_W(-\alpha)}$$

where M_W is the moment generating function for W .

DERIVATION

First we note that

$$E[V_n e^{-\alpha S_n}] = c M_{S_n}(-\alpha) = c M_W^n(-\alpha)$$

Hence, by properties of expectation and the geometric series

$$E[C] = c \sum_{n=1}^{\infty} M_W^n(-\alpha) = \frac{M_W(-\alpha)}{1 - M_W(-\alpha)}, \text{ provided } |M_W(-\alpha)| < 1$$

Since $\alpha > 0$ and $W > 0$, we have $0 < e^{-\alpha W} < 1$, so that $M_W(-\alpha) = E[e^{-\alpha W}] < 1$

Example 15.2.16 Uniformly distributed interarrival times

Suppose each $W_i \sim \text{uniform}(a, b)$. Then (see [Appendix C](#)),

$$M_W(-\alpha) = \frac{e^{-a\alpha} - e^{-b\alpha}}{\alpha(b-a)} \text{ so that } E[C] = c \cdot \frac{e^{-a\alpha} - e^{-b\alpha}}{\alpha(b-a) - [e^{-a\alpha} - e^{-b\alpha}]}$$

Let $a = 1$, $b = 5$, $c = 100$ and $\alpha = 0$. Then,

```
a = 1;
b = 5;
c = 100;
A = 0.08;
MW = (exp(-a*A) - exp(-b*A))/(A*(b - a))
MW = 0.7900
EC = c*MW/(1 - MW)
EC = 376.1643
```

This page titled [15.2: Some Random Selection Problems](#) is shared under a [CC BY 3.0](#) license and was authored, remixed, and/or curated by [Paul Pfeiffer](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.