

## Solutions to Assignment #12

1. Consider a hypothetical experiment in which there are only three bacteria in a culture. Suppose that each bacterium has a small probability  $p$ , with  $0 < p < 1$ , of developing a mutation in a short time interval. Number the bacteria 1, 2 and 3. Use the symbol  $M$  to denote the given bacterium mutates in the short time interval, and  $N$  to denote that the bacterium did not mutate in that interval.
- (a) List all possible outcomes of the experiment using the symbols  $M$  or  $N$ , for each of the bacteria 1, 2 and 3, to denote whether a bacterium mutated or not, respectively. This will generate triples made up of the symbols  $M$  and  $N$ . What is the probability of each outcome?

*Solution:* For each bacterium there are two possibilities: mutation ( $M$ ) and no mutation ( $N$ ). Thus, since there are three bacteria, there are eight possible outcomes:

$NNN$   
 $NNM$   
 $NMN$   
 $NMM$   
 $MNN$   
 $MNM$   
 $MMN$   
 $MMM$

Since  $P[N] = 1 - p$  and  $P[M] = p$ , and the event that a given bacterium will mutate or not is independent of that for another bacterium, the probabilities for each outcome are given by

Outcome	Probability
$NNN$	$(1 - p)^3$
$NNM$	$p(1 - p)^2$
$NMN$	$p(1 - p)^2$
$NMM$	$p^2(1 - p)$
$MNN$	$p(1 - p)^2$
$MNM$	$p^2(1 - p)$
$MMN$	$p^2(1 - p)$
$MMM$	$p^3$

Table 1: Outcomes and their probabilities

- (b) Let  $X$  denote the number of bacteria that mutate in the short time interval. This defines a discrete random variable. List the possible values for  $X$  and give the probability for each of these values. In other words, give the probability mass function for  $X$ .

*Solution:* The possible values of  $X$  are 0, 1, 2 and 3, and their corresponding probabilities are

$x$	$P[X = x]$
0	$(1 - p)^3$
1	$3p(1 - p)^2$
2	$3p^2(1 - p)$
3	$p^3$

To obtain the probability values in this table, we consider the event  $[X = x]$  and see which outcomes in Table 1 make up this event, and then add their probabilities. For instance, the event  $[X = 1]$  is made up of the outcomes  $NNM$ ,  $NMN$  and  $MNN$ , each of which has probability  $p(1 - p)^2$ , and therefore,  $P[X = 1] = 3p(1 - p)^2$ .  $\square$

- (c) Compute the expected value and variance of  $X$ .

*Solution:*  $E(X) = 0 \cdot P[X = 0] + 1 \cdot P[X = 1] + 2 \cdot P[X = 2] + 3 \cdot P[X = 3]$ , so that

$$\begin{aligned} E(X) &= 3p(1 - p)^2 + 2 \cdot 3p^2(1 - p) + 3 \cdot p^3 \\ &= 3p[(1 - p)^2 + 2p(1 - p) + p^2] \\ &= 3p[1 - 2p + p^2 + 2p - 2p^2 + p^2] \\ &= 3p \end{aligned}$$

The variance of  $X$  is then given by  $\text{var}(X) = \sum_{n=0}^3 n^2 P[X = n] - (3p)^2$ ,

where

$$\begin{aligned} \sum_{n=0}^3 n^2 P[X = n] &= 1^2 \cdot P[X = 1] + 2^2 \cdot P[X = 2] + 3^2 \cdot P[X = 3] \\ &= 3p(1 - p)^2 + 2^2 \cdot 3p^2(1 - p) + 3^2 p^3 \\ &= 3p[(1 - p)^2 + 4p(1 - p) + 3p^2] \\ &= 3p[1 - 2p + p^2 + 4p - 4p^2 + 3p^2] \\ &= 3p[1 + 2p]. \end{aligned}$$

Hence  $\text{var}(X) = 3p(1 + 2p) - (3p)^2 = 3p(1 + 2p - 3p) = 3p(1 - p)$ .  $\square$

2. Repeat the previous problem in the case of four bacteria, each having a probability  $p$  of mutating in a short time interval.

*Solution:* In this case, there are 16 outcomes with corresponding probabilities

Outcome	Probability
$NNNN$	$(1 - p)^4$
$NNNM$	$p(1 - p)^3$
$NNMN$	$p(1 - p)^3$
$NNMM$	$p^2(1 - p)^2$
$NMNN$	$p(1 - p)^3$
$NMNM$	$p^2(1 - p)^2$
$NMMN$	$p^2(1 - p)^2$
$NMMM$	$p^3(1 - p)$
$MNNN$	$p(1 - p)^3$
$MNNM$	$p^2(1 - p)^2$
$MNMN$	$p^2(1 - p)^2$
$MNMM$	$p^3(1 - p)$
$MMNN$	$p^2(1 - p)^2$
$MMNM$	$p^3(1 - p)$
$MMMN$	$p^3(1 - p)$
$MMMM$	$p^4$

Table 2: Outcomes and their probabilities

If  $X$  denotes the number of  $M$ 's in the outcomes in Table 2, then the possible values of  $X$  are 0, 1, 2, 3 and 4, and their corresponding probabilities are

$x$	$P[X = x]$
0	$(1 - p)^4$
1	$4p(1 - p)^3$
2	$6p^2(1 - p)^2$
3	$4p^3(1 - p)$
4	$p^4$

For the expected value of  $X$  we have

$$\begin{aligned}
 E(X) &= P[X = 1] + 2 \cdot P[X = 2] + 3 \cdot P[X = 3] + 4 \cdot P[X = 4] \\
 &= 4p(1 - p)^3 + 2 \cdot 6p^2(1 - p)^2 + 3 \cdot 4p^3(1 - p) + 4p^4 \\
 &= 4p[(1 - p)^3 + 3p(1 - p)^2 + 3p^2(1 - p) + p^3]
 \end{aligned}$$

Thus,

$$\begin{aligned} E(X) &= 4p[1 - 3p + 3p^2 - p^3 + 3p(1 - 2p + p^2) + 3p^2 - 3p^3 + p^3] \\ &= 4p[1 - 3p + 3p - 6p^3 + 3p^3 + 6p^2 - 3p^3] \\ &= 4p. \end{aligned}$$

Similarly,

$$\begin{aligned} \sum_{n=0}^4 n^2 P[X = n] &= P[X = 1] + 2^2 \cdot P[X = 2] + 3^2 \cdot P[X = 3] + 4^2 \cdot P[X = 4] \\ &= 4p(1 - p)^3 + 2^2 \cdot 6p^2(1 - p)^2 + 3^2 \cdot 4p^3(1 - p) + 4^2 p^4 \\ &= 4p[(1 - p)^3 + 6p(1 - p)^2 + 9p^2(1 - p) + 4p^3] \\ &= 4p[1 - 3p + 3p^2 - p^3 + 6p(1 - 2p + p^2) + 9p^2 - 9p^3 + 4p^3] \\ &= 4p[1 - 3p + 12p^2 - 6p^3 + 6p - 12p^2 + 6p^3] \\ &= 4p(1 + 3p) \end{aligned}$$

It then follows that

$$\begin{aligned} \text{var}(X) &= \sum_{n=0}^4 n^2 P[X = n] - (E(X))^2 \\ &= 4p(1 + 3p) - (4p)^2 \\ &= 4p(1 + 3p - 4p) \\ &= 4p(1 - p). \quad \square \end{aligned}$$

3. Generalize problems 1 and 2 for the case of  $N$  bacteria, each having a probability  $p$  of mutating in a short time interval.

For this problem it will be helpful to know that the number of different ways of choosing  $m$  bacteria out of  $N$  is given by the combinatorial expression

$$\binom{N}{m} = \frac{N!}{m!(N-m)!},$$

for  $m = 0, 1, 2, \dots, N$ . The symbol  $\binom{N}{m}$  is read “ $N$  choose  $m$ .”

*Note:* The distribution for  $X$  obtained in this problem is called the *binomial* distribution with parameters  $p$  and  $N$ .

*Solution:* Let  $X$  denote the number of bacteria out of the  $N$  that mutate. Then, the event  $[X = m]$  is made up of outcomes consisting of  $m$  “ $M$ s” and  $N - m$  “ $N$ s.” There are  $\binom{N}{m}$  of those outcomes, each having a probability of  $p^m(1-p)^{N-m}$ . It then follows that

$$P[X = m] = \binom{N}{m} p^m (1-p)^{N-m} \quad \text{for } m = 0, 1, 2, \dots, N.$$

As for the expected value of  $X$  and its variance, it appears as if

$$E(X) = Np,$$

and

$$\text{var}(X) = Np(1-p). \quad \square$$

4. In the previous problem you found that the expected number bacteria that mutate in the short time interval is  $pN$ . Denote this value by  $\lambda$ , so that  $\lambda = pN$ . Explore what happens as  $N$  gets larger and larger while  $\lambda$  is kept at a fixed value. In particular, compute  $\lim_{N \rightarrow \infty} P[X = m]$  for any given  $m$ . What do you discover?

*Hints:*

- i. For this problem it will be helpful to remember that another expression for the exponential function,  $e^x$ , is given by the limit

$$e^x = \lim_{N \rightarrow \infty} \left(1 + \frac{x}{N}\right)^N \quad \text{for any real value of } x.$$

ii. Also, 
$$\frac{N!}{N^m(N-m)!} = \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{m+1}{N}\right).$$

*Solution:* Suppose that  $pN = \lambda$  is constant. Then,  $p = \frac{\lambda}{N}$ , and

$$\begin{aligned} P[X = m] &= \binom{N}{m} p^m (1-p)^{N-m} \\ &= \frac{N!}{m!(N-m)!} \left(\frac{\lambda}{N}\right)^m \left(1 - \frac{\lambda}{N}\right)^{N-m}. \end{aligned}$$

Thus,

$$P[X = m] = \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{N}\right)^N \frac{N!}{N^m(N-m)!} \left(1 - \frac{\lambda}{N}\right)^{-m}. \quad (1)$$

Observe that for fixed  $m$

$$\lim_{N \rightarrow \infty} \left(1 - \frac{\lambda}{N}\right)^{-m} = 1,$$

and

$$\lim_{N \rightarrow \infty} \frac{N!}{N^m(N-m)!} = \lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{m+1}{N}\right) = 1.$$

It then follows from (1) that

$$\lim_{N \rightarrow \infty} P[X = m] = \frac{\lambda^m}{m!} \lim_{N \rightarrow \infty} \left(1 - \frac{\lambda}{N}\right)^N = \frac{\lambda^m}{m!} e^{-\lambda}.$$

This shows that the distribution of  $X$  tends to a Poisson distribution with parameter  $\lambda = pN$  as  $N \rightarrow \infty$ . Thus, when  $N$  is very large and  $p$  is very small, the distribution of  $X$  can be approximated by a Poisson distribution with parameter  $\lambda = pN$ .  $\square$

5. *Modeling Survival Time after a Treatment*<sup>1</sup>. Consider a group of people who have received a treatment for a disease such as cancer. Let  $T$  denote the *survival time*; that is,  $T$  is the number of years a person lives after receiving the treatment.  $T$  can be modeled as a continuous random variable with probability density function (pdf) given by

$$f_T(t) = \begin{cases} \frac{1}{\beta} e^{-t/\beta} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0, \end{cases}$$

for some positive constant  $\beta$ . This pdf can be used to compute the probability that, after receiving treatment, a patient will survive between  $t_1$  and  $t_2$  years as follows:

$$P[t_1 < T < t_2] = \int_{t_1}^{t_2} f_T(t) dt.$$

- (a) Find the expected value of  $T$ ; that is, compute  $E(T) = \int_{-\infty}^{\infty} t f_T(t) dt$ .

*Solution:* Integrate by parts to get

$$\begin{aligned} E(T) &= \int_0^{\infty} t \frac{1}{\beta} e^{-t/\beta} dt \\ &= -t e^{-t/\beta} \Big|_0^{\infty} + \int_0^{\infty} e^{-t/\beta} dt \\ &= 0 + \left[ -\beta e^{-t/\beta} \right]_0^{\infty} \\ &= \beta. \quad \square \end{aligned}$$

- (b) The survival function,  $S(t)$ , is the probability that a randomly selected person will survive for at least  $t$  years after receiving treatment. Compute  $S(t)$ .

*Solution:* For  $t > 0$ ,

$$S(t) = P[T > t] = \int_t^{\infty} f_T(\tau) d\tau = \int_t^{\infty} \frac{1}{\beta} e^{-\tau/\beta} d\tau.$$

Thus,

$$S(t) = \left[ -e^{-\tau/\beta} d\tau \right]_t^{\infty} = e^{-t/\beta}. \quad \square$$

---

<sup>1</sup>Adapted from problem 7 on p. 427 in *Calculus: Single Variable*, Hughes-Hallet *et al.*, Fourth Edition, Wiley, 2005

- (c) Suppose that a patient has a 70% probability of surviving at least two years. Find  $\beta$ .

*Solution:* Given that  $S(2) = 0.70$ , we have that  $e^{-2/\beta} = 0.7$ . Solving for  $\beta$  we then obtain that  $\beta = -\frac{2}{\ln(0.7)} \doteq 5.6$  years.  $\square$