Solutions to Review Problems for Final Exam

1. Three cards are in a bag. One card is red on both sides. Another card is white on both sides. The third card is red on one side and white on the other side. A card is picked at random and placed on a table. Compute the probability that if a given color is shown on top, the color on the other side is the same as that of the top.

   ***Solution**:* Each card has a likelihood of 1/3 of being picked.

   Assume for definiteness that the top of the picked card is red. Let T_r denote the event that the top of the picked car shows red and B_r denote the event that the bottom of the card is also red. We want to compute

$$\Pr(B_r \mid T_r) = \frac{\Pr(T_r \cap B_r)}{\Pr(T_r)}. \tag{1}$$

   Note that

$$\Pr(T_r \cap B_r) = \frac{1}{3}, \tag{2}$$

   since there is only one card for which both sides are red.

   To compute $\Pr(T_r)$ observe that there are three equally likely choices out of six for the top of the card to show red; thus,

$$\Pr(T_r) = \frac{1}{2}. \tag{3}$$

   Hence, using (2) and (3), we obtain from (1) that

$$\Pr(B_r \mid T_r) = \frac{2}{3}. \tag{4}$$

   Similar calculations can be used to show that

$$\Pr(T_w) = \frac{1}{2}, \tag{5}$$

   and

$$\Pr(B_w \mid T_w) = \frac{2}{3}. \tag{6}$$

   Let E denote the event that a card showing a given color on the top side will have the same color on the bottom side. Then, by the law of total probability,

$$\Pr(E) = \Pr(T_r)\cdot\Pr(B_r \mid T_r) + \Pr(T_w)\cdot\Pr(B_w \mid T_w), \tag{7}$$

so that, using (2), (4), (5) and (6), we obtain from (7) that

$$\Pr(E) = \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{2}{3} = \frac{2}{3}.$$

□

2. An urn contains 10 balls: 4 red and 6 blue. A second urn contains 16 red balls and a number b of blue balls. A single ball is drawn from each urn. The probability that both balls are the same color is 0.44. Determine the value of b.

Solution*:* Let R_1 denote the event that the ball drawn from the first urn is red, B_1 denote the event that the ball drawn from the first urn is blue, R_2 denote the event that the ball drawn from the second urn is red, and B_2 denote the event that the ball drawn from the second urn is blue. We are interested in

$$E = (R_1 \cap R_2) \cup (B_1 \cap B_2),$$

the event that both balls are of the same color. We observe that $R_1 \cap R_2$ and $B_1 \cap B_2$ are disjoint events; thus,

$$\Pr(E) = \Pr(R_1 \cap R_2) + \Pr(B_1 \cap B_2), \tag{8}$$

where R_1 and R_2 are independent events as well as B_1 and B_2. It then follows from (8) that

$$\Pr(E) = \Pr(R_1) \cdot \Pr(R_2) + \Pr(B_1) \cdot \Pr(B_2), \tag{9}$$

where

$$\Pr(R_1) = \frac{4}{10}, \quad \Pr(R_2) = \frac{16}{16+b}, \quad \Pr(B_1) = \frac{6}{10}, \quad \text{and} \quad \Pr(B_2) = \frac{b}{16+b};$$

thus, substituting into (9),

$$\Pr(E) = \frac{4}{10} \cdot \frac{16}{16+b} + \frac{6}{10} \cdot \frac{b}{16+b}. \tag{10}$$

We are given that $\Pr(E) = 0.44$; combining this information with (10) yields

$$\frac{32+3b}{16+b} = 2.2. \tag{11}$$

Solving (11) for b yields $b = 4$. □

3. A blood test indicates the presence of a particular disease 95% of the time when the disease is actually present. The same test indicates the presence of the disease 0.5% of the time when the disease is not present. One percent of the population actually has the disease. Calculate the probability that a person has the disease given that the test indicates the presence of the disease.

   Solution: Let D denote the event that a person selected at random from the population has the disease. Then,

$$\Pr(D) = 0.01. \tag{12}$$

   Let P denote the event that the blood test is positive for the existence of the disease. We are given that

$$\Pr(P \mid D) = 0.95 \tag{13}$$

   and

$$\Pr(P \mid D^c) = 0.005. \tag{14}$$

   We want to compute

$$\Pr(D \mid P) = \frac{\Pr(D \cap P)}{\Pr(P)}, \tag{15}$$

   where

$$\Pr(D \cap P) = \Pr(D) \cdot \Pr(P \mid D),$$

   by the multiplication rule of probability; so that,

$$\Pr(D \cap P) = 0.0095, \tag{16}$$

   by virtue of (12) and (13), and

$$\Pr(P) = \Pr(D) \cdot \Pr(P \mid D) + \Pr(D^c) \cdot \Pr(P \mid D^c), \tag{17}$$

   by the law of total probability.

   Substituting the values in (12), (13) and (14) into (17) yields

$$\Pr(P) = (0.01) \cdot (0.95) + (0.99) \cdot (0.005) = 0.01445. \tag{18}$$

   Substituting the values in (16) and (18) into (15) yields

$$\Pr(D \mid P) \doteq 0.6574.$$

   Thus, it a person tests positive, there is about a 66% chance that she or he has the disease. □

4. A study is being conducted in which the health of two independent groups of ten policyholders is being monitored over a one-year period of time. Individual participants in the study drop out before the end of the study with probability 0.2 (independently of the other participants). What is the probability that at least 9 participants complete the study in one of the two groups, but not in both groups?

   ***Solution**:* Let X denote the number of participants in the first group that drop out of the study and Y denote the number of participants in the second group that drop out of the study. Then X and Y are independent binomial$(10, 0.2)$ random variables. Then, $A = (X \leqslant 1)$ is the event that at least 9 participants complete the study in the first group, and $B = (Y \leqslant 1)$ is the event that at least 9 participants complete the study in the second group. We want to compute the probability of the event

   $$E = (A \cap B^c) \cup (A^c \cap B),$$

   so that

   $$\Pr(E) = \Pr(A \cap B^c) + \Pr(A^c \cap B), \tag{19}$$

   since $A \cap B^c$ and $A^c \cap B$ are disjoint events.

   Next, use the independence of A and B, given that X and Y are independent random variables, to get from (19) that

   $$\Pr(E) = \Pr(A) \cdot (1 - \Pr(B) + \Pr(B) \cdot (1 - \Pr(A)), \tag{20}$$

   where

   $$\begin{aligned} \Pr(A) &= \Pr(X \leqslant 1) \\ &= \Pr(X = 0) + \Pr(X = 1) \\ &= (0.8)^{10} + 10 \cdot (0.2) \cdot (0.8)^9; \end{aligned}$$

   so that,

   $$\Pr(A) \doteq 0.3758. \tag{21}$$

   Similarly,

   $$\Pr(B) \doteq 0.3758. \tag{22}$$

   Substituting the results in (21) and (22) into (20) yields

   $$\Pr(E) \doteq 0.4691.$$

   Thus, the probability that at least 9 participants complete the study in one of the two groups, but not in both groups, is about 47%. □

5. Suppose that $0 < \rho < 1$ and let $p(k) = C\rho^k$, for $k = 0, 1, 2, 3, \ldots$, and some constant $C > 0$.

   (a) Find the value of C so that p is the probability mass function (pmf) for a random variable.

   Solution: Compute

$$\sum_{k=0}^{\infty} C\rho^k = C\sum_{k=0}^{\infty} \rho^k = C \cdot \frac{1}{1-\rho},$$

since the geometric series $\displaystyle\sum_{n=0}^{\infty} \rho^n$ converges to $\dfrac{1}{1-\rho}$, given that $0 < \rho < 1$. Therefore, since we want p to be a pmf, we must have that

$$\frac{C}{1-\rho} = 1,$$

from which we get that

$$C = 1 - \rho.$$

□

   (b) Let X denote a discrete random variable with pmf p with the value of C found in part (a). Compute $\Pr(X > 1)$.

   Solution: We are assuming that X has pmf

$$p_X(k) = \begin{cases} (1-\rho)\rho^k, & \text{for } k = 0, 1, 2, \ldots; \\ 0, & \text{elsewhere.} \end{cases} \tag{23}$$

We can use the definition of the pmf of X in (23) to compute

$$\begin{aligned} \Pr(X > 1) &= 1 - \Pr(X \leqslant 1) \\ &= 1 - p(0) - p(1) \\ &= 1 - (1-\rho) - \rho(1-\rho) \\ &= \rho^2. \end{aligned}$$

□

(c) Let X denote a discrete random variable with pmf p with the value of C found in part (a). Compute compute the mgf of X.

Solution: Use the pmf of X in (23) to compute

$$\begin{aligned}\psi_{X}(t) &= E(e^{tX}) \\ &= \sum_{k=0}^{\infty} e^{tk}(1-\rho)\rho^k \\ &= (1-\rho)\sum_{k=0}^{\infty}(e^t)^k\rho^k;\end{aligned}$$

so that,

$$\psi_{X}(t) = (1-\rho)\sum_{k=0}^{\infty}(\rho e^t)^k. \tag{24}$$

Observe that the series on the right–hand side of (24) is a geometric series that converges provided that $\rho e^t < 1$, or

$$e^t < \frac{1}{\rho},$$

from which we get that

$$t < \ln\left(\frac{1}{\rho}\right),$$

or

$$t < -\ln\rho. \tag{25}$$

Thus, if (25) holds true, we obtain from (24) that

$$\psi_{X}(t) = (1-\rho)\frac{1}{1-\rho e^t};$$

so that,

$$\psi_{X}(t) = \frac{1-\rho}{1-\rho e^t}, \qquad \text{for } t < -\ln\rho. \tag{26}$$

□

(d) Let X denote a discrete random variable with pmf p with the value of C found in part (a). Use the mgf of X computed in part (c) to compute the expected value and variance of X.

Solution: Taking the derivatives of the mgf of X given in (26), using the quotient rule and the chain rule, yields

$$\psi'_X(t) = -\frac{1-\rho}{(1-\rho e^t)^2}\cdot(-\rho e^t),$$

which simplifies to

$$\psi'_X(t) = \frac{\rho(1-\rho)e^t}{(1-\rho e^t)^2}, \qquad \text{for } t < -\ln\rho. \tag{27}$$

Similarly, applying the quotient rule and the chain rule to the expression for $\psi'_X(t)$ in (27) yields

$$\psi''_X(t) = \frac{(1-\rho e^t)^2\rho(1-\rho)e^t - \rho(1-\rho)e^t\cdot 2(1-\rho e^t)(-\rho e^t)}{(1-\rho e^t)^4},$$

which simplifies to

$$\psi''_X(t) = \frac{\rho(1-\rho)e^t(1-\rho e^t)[1 - pe^t + 2\rho e^t]}{(1-\rho e^t)^4},$$

or

$$\psi''_X(t) = \frac{\rho(1-\rho)e^t(1+\rho e^t)}{(1-\rho e^t)^3}, \qquad \text{for } t < -\ln\rho. \tag{28}$$

To evaluate the expected value of X we use the expression for $\psi'_X(t)$ in (27) to get

$$E(X) = \psi'_X(0) = \frac{\rho(1-\rho)}{(1-\rho)^2},$$

from which we get that

$$E(X) = \frac{\rho}{1-\rho}. \tag{29}$$

Similarly, we can use (28) to compute the second moment of X:

$$E(X^2) = \psi''_X(0) = \frac{\rho(1-\rho)(1+\rho)}{(1-\rho)^3},$$

from which we get

$$E(X^2) = \frac{\rho(1+\rho)}{(1-\rho)^2}. \tag{30}$$

Finally, to evaluate the variance of X, use (29) and (30) to compute

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{\rho(1+\rho)}{(1-\rho)^2} - \frac{\rho^2}{(1-\rho)^2},$$

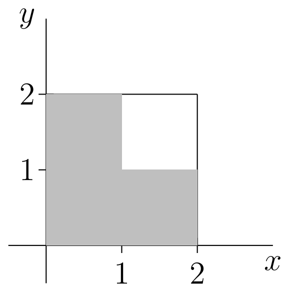

Figure 1: Region for Problem 6

from which we get that

$$\text{Var}(X) = \frac{\rho}{(1-\rho)^2}.$$

$\square$

6. A device runs until either of two components fails, at which point the device stops running. The joint density function of the lifetimes of the two components, both measured in hours, is

$$f(x,y) = \frac{x+y}{8}, \qquad \text{for } 0 < x < 2 \text{ and } 0 < y < 2;$$

and 0 elsewhere.

What is the probability that the device fails during its first hour of operation?

***Solution**:* We want to compute the probability that either component fails within the first hour of operation; that is the probability of the event E given by

$$E = (0 < X < 1) \cup (0 < Y < 1).$$

The event E is pictured as the shaded region in Figure 1.

The probability of E is given by

$$\begin{aligned}
\Pr(E) &= \iint_E f(x,y)\,dxdy \\
&= \int_0^1 \int_0^2 \frac{x+y}{8}\,dydx + \int_1^2 \int_0^1 \frac{x+y}{8}\,dydx \\
&= \frac{1}{8}\left(\int_0^1 \left[xy + \frac{y^2}{2}\right]_0^2 dx + \int_1^2 \left[xy + \frac{y^2}{2}\right]_0^1 dx\right) \\
&= \frac{1}{8}\left(\int_0^1 (2x+2)\ dx + \int_1^2 \left(x + \frac{1}{2}\right)\ dx\right) \\
&= \frac{1}{8}\left(\left[x^2+2x\right]_0^1 + \left[\frac{x^2}{2} + \frac{x}{2}\right]_1^2\right),
\end{aligned}$$

so that $\Pr(E) = \dfrac{1}{8}(3 + (3-1)) = \dfrac{5}{8}$. Thus, the probability that the device fails during its first hour of operation is 62.5%. $\square$

7. Let $M(t)$ denote the number of mutations in a bacterial colony that occur during the interval $[0,t]$, assuming that $M(0) = 0$. Suppose that $M(t)$ has a Poisson(λt) distribution, where $\lambda > 0$ is a positive parameter.

   (a) Give an interpretation for λ.

   Answer: Since we are assuming that $M(t) \sim \text{Poisson}(\lambda t)$, for all $t > 0$, where $\lambda > 0$, the expected number of mutations in the time interval $[0,t]$ is

   $$E(M(t)) = \lambda t, \qquad \text{for } t > 0.$$

   Thus,

   $$\lambda = \frac{E(M(t))}{t}, \qquad \text{for } t > 0,$$

   gives the average number of mutations per unit of time. $\square$

   (b) Compute the probability that no mutations occur in the interval $[0,t]$.

   Solution: Since we are assuming that $M(t) \sim \text{Poisson}(\lambda t)$, for all $t > 0$, where $\lambda > 0$, the pmf of $M(t)$ is given by

   $$p_{_{M(t)}}(k) = \begin{cases} \dfrac{(\lambda t)^k}{k!} e^{-\lambda t}, & \text{if } k = 0,1,2,\ldots; \\ 0, & \text{elsewhere.} \end{cases} \tag{31}$$

Thus, the probability of no mutations in the interval $[0, t]$ is

$$\Pr(M(t) = 0) = p_{_{M(t)}}(0) = e^{-\lambda t}, \quad \text{for } t > 0, \tag{32}$$

where we have used the pmf of $M(t)$ in (31). □

(c) Let T_1 denote the time that the first mutation occurs. Find the distribution of T_1.

***Solution**:* Observe that, for $t > 0$, the event $(T_1 > t)$ is the same as the event $(M(t) = 0)$; since, if $t < T_1$, there have have been no mutations in the time interval $[0, t]$. Consequently,

$$\Pr[T_1 > t] = \Pr[M(t) = 0] = e^{-\lambda t},$$

where we have used (32). Thus,

$$\Pr(T_1 \leqslant t) = 1 - \Pr(T_1 > t) = 1 - e^{-\lambda t}, \quad \text{for } t > 0,$$

by the complement rule of probability. We then have that the cdf of T_1 is

$$F_{_{T_1}}(t) = \begin{cases} 1 - e^{-\lambda t}, & \text{for } t > 0; \\ 0 & \text{for } t \leqslant 0. \end{cases} \tag{33}$$

It follows from (33) that the pdf for T_1 is

$$f_{_{T_1}}(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{for } t > 0; \\ 0 & \text{for } t \leqslant 0, \end{cases}$$

which is the pdf for an exponential distribution with parameter $\beta = 1/\lambda$; thus,

$$T_1 \sim \text{Exponential}(1/\lambda).$$

□

8. A computer manufacturing company conducts acceptance sampling for incoming computer chips. After receiving a huge shipment of computer chips, the company randomly selects 800 chips. If three or fewer nonconforming chips are found, the entire lot is accepted without inspecting the remaining chips in the lot. If four or more chips are nonconforming, every chip in the entire lot is carefully inspected at the supplier's expense. Assume that the true proportion of nonconforming computer chips being supplied is 0.001. Estimate the probability the lot will be accepted.

Solution: Let X denote the number of nonconforming chips found in the random sample of 800. We may assume that the tests of the chips are independent trials. Thus,

$$X \sim \text{binomial}(n, p),$$

where $n = 800$ and $p = 0.001$. We want to estimate $\Pr(X \leqslant 3)$.

Since n is large and p is very small, we may use the Poisson approximation to the binomial distribution to get

$$\Pr(X \leqslant 3) \approx \Pr(Y \leqslant 3), \qquad \text{where } Y \sim \text{Poisson}(0.8).$$

It then follows that

$$\Pr(X \leqslant 3) \approx e^{-0.8} + (0.8)e^{-0.8} + \frac{(0.8)^2}{2}e^{-0.8} + \frac{(0.8)^3}{6}e^{-0.8} \doteq 0.9909;$$

thus, the probability the lot will be accepted is about 99.09%. $\square$

9. A company manufactures a brand of light bulb with a lifetime in months that is normally distributed with mean 3 and variance 1 . A consumer buys a number of these bulbs with the intention of replacing them successively as they burn out. The light bulbs have independent lifetimes. What is the smallest number of bulbs to be purchased so that the succession of light bulbs produces light for at least 40 months with probability at least 0.9772?

Solution: Denote the lifetimes of the light bulbs by $T_1, T_2, T_3, \ldots$, so that

$$T_i \sim \text{normal}(3, 1), \quad \text{for } i = 1, 2, 3, \ldots,$$

are independent random variables measured in months. The total duration of n light bulbs is $\sum_{k=1}^{n} T_k$. We want to find the smallest n so that

$$\Pr\left(\sum_{k=1}^{n} T_k > 40\right) \geqslant 0.9772. \tag{34}$$

Since the T_i's are independent, normally distributed, it follows that

$$\sum_{k=1}^{n} T_k \sim \text{normal}(3n, n),$$

it then follows that

$$\frac{\sum_{k=1}^{n} T_k - 3n}{\sqrt{n}} \sim \text{normal}(0, 1). \tag{35}$$

Using the information in (35), we can write the estimate in (34) as

$$\Pr\left(Z > \frac{40-3n}{\sqrt{n}}\right) \geqslant 0.9772,$$

where $Z \sim \text{normal}(0,1)$, or

$$1 - F_{Z}\left(\frac{40-3n}{\sqrt{n}}\right) \geqslant 0.9772,$$

form which we get

$$1 - F_{Z}\left(-\frac{3n-40}{\sqrt{n}}\right) \geqslant 0.9772. \tag{36}$$

Using the symmetry for the pdf of standard normal distribution we can rewrite (36) as

$$F_{Z}\left(\frac{3n-40}{\sqrt{n}}\right) \geqslant 0.9772. \tag{37}$$

Since the cdf of Z is strictly increasing, it has an inverse function F_{Z}^{-1}. Applying the inverse function to both sides of (37) we obtain

$$\frac{3n-40}{\sqrt{n}} \geqslant F_{Z}^{-1}(0.9772). \tag{38}$$

An approximation for $F_{Z}^{-1}(0.9772)$ on the right–hand side of (38) can be obtained using the `NORM.INV` function in MS Excel. We then get that, approximately, the inequality in (37) is satisfied if

$$\frac{3n-40}{\sqrt{n}} \geqslant 1.9991. \tag{39}$$

To find the smallest value of n for which (39) is satisfied, we solve the inequality in (39) to get $n \geqslant 16$. Thus, the smallest number of bulbs to be purchased so that the succession of light bulbs produces light for at least 40 months with probability at least 0.9772 is 16. □

10. A random walk on the integer points on the x–axis begins at $x = 0$. At each step, the random walker is equally likely to move one unit to the left or to the right. Furthermore, the choice to move to the left or to the right is independent of the choices made in previous steps. Use the central limit theorem to estimate the probability that the random walker will be 10 units or more away from the origin after 100 steps.

Solution: Let S_k denote the distance (either $+1$ or -1) that the random walker travels at the k^{th} step, for $k = 1, 2, 3, \ldots$. Then, S_k has a pmf given by

$$p_{S_k}(s) = \begin{cases} \dfrac{1}{2}, & \text{if } s = -1; \\[2ex] \dfrac{1}{2}, & \text{if } s = +1; \\[2ex] 0, & \text{elsewhere.} \end{cases} \tag{40}$$

We assume that $S_1, S_2, S_3, \ldots$ are mutually independent.

Using the pmf in (40) we compute the expected value of each S_k to be

$$E(S_k) = (-1)\frac{1}{2} + (+1)\frac{1}{2} = 0;$$

so that,

$$\mu = 0. \tag{41}$$

Similarly, the second moment of each S_k is

$$E(S_k^2) = (-1)^2\frac{1}{2} + (+1)^2\frac{1}{2} = 1.$$

Consequently, the variance of each S_k is

$$\text{Var}(S_k) = E(S_k^2) - [E(S_k)]^2 = 1,$$

where we have used (41). Thus,

$$\sigma^2 = 1. \tag{42}$$

The location of the random walker after n steps is then the random variable

$$X_n = \sum_{k=1}^{n} S_k, \qquad \text{for } n = 1, 2, 3, \ldots, \tag{43}$$

since the random walker starts at $x = 0$.

We would like to use the central limit theorem to estimate the probability

$$\Pr(|X_n| \geqslant 10), \qquad \text{where } n = 100, \tag{44}$$

and X_n is given in (43).

First, use the complement rule of probability to rewrite (44) as

$$\Pr(|X_n| \geqslant 10) = 1 - \Pr(|X_n| < 10),$$

or

$$\Pr(|X_n| \geqslant 10) = 1 - \Pr(|X_n| \leqslant 9), \tag{45}$$

since X_n is a discrete random variable with integer values.

We estimate the right–most probability in (45)

$$\Pr(|X_n| \leqslant 9) = \Pr(-9 \leqslant X_n \leqslant 9),$$

or, using the continuity correction,

$$\Pr(|X_n| \leqslant 9) = \Pr(-9.5 < X_n \leqslant 9.5), \tag{46}$$

for $n = 100$.

We can rewrite (46) as

$$\Pr(|X_n| \leqslant 9) = \Pr\left(-0.95 < \frac{X_n - n\mu}{\sqrt{n}\sigma} \leqslant 0.95\right), \tag{47}$$

where we have used (41) and (42), and set $n = 100$.

Applying the central limit theorem, we obtain from (47) that

$$\Pr(|X_n| \leqslant 9) \approx \Pr\left(-0.95 < Z \leqslant 0.95\right), \quad \text{where } Z \sim \text{normal}(0,1), \tag{48}$$

and $n = 100$.

It follows from (48) that

$$\Pr(|X_n| \leqslant 9) \approx F_{Z}(0.95) - F_{Z}(-0.95) \quad \text{where } Z \sim \text{normal}(0,1), \tag{49}$$

and $n = 100$.

Using the `NORM.DIST` function in MS Excel we obtain from (49) the estimate

$$\Pr(|X_n| \leqslant 9) \approx 0.8289 - 0.1711,$$

or

$$\Pr(|X_n| \leqslant 9) \approx 0.6578. \tag{50}$$

Combining (45) and (50) we obtain the estimate

$$\Pr(|X_n| \geqslant 10) \approx 1 - 0.6578,$$

or

$$\Pr(|X_n| \geqslant 10) \approx 0.3422.$$

Thus, the probability that the random walker will be 10 units or more away from the origin after 100 steps is about 34.22%. □