Math 2a Solution Set 4

4.1 Problem 1

(a) In the first round of tests, exactly two tests are performed. Let A_1 and A_2 denote the samples in the second round of testing, each of which has been pooled from $n/2$ individuals. Let X_i denote the indicator function for the event that the test on A_i is positive. In the second round of testing, $2(X_1 + X_2)$ tests are performed. Let $B_1, \ldots, B_4$ denote the samples at the last stage, each of which has been pooled from $n/4$ individuals. Let Y_i denote the indicator function for the event that the test on B_i is positive. In the third stage, $n/4$ tests are performed for each positive B_i, so the total number of tests performed is $\frac{n}{4}(Y_1 + \cdots + Y_4)$. The total number of tests performed is therefore

$$Z = 2 + 2(X_1 + X_2) + \frac{n}{4}(Y_1 + Y_2 + Y_3 + Y_4).$$

Let p denote the probability that a given individual has the disease, and set $q = 1 - p$. The test on A_i is negative iff all $n/2$ individuals from whom the samples have been pooled do not have the disease. Hence $E(X_i) = P(X_i = 1) = 1 - q^{n/2}$. Similar reasoning gives $E(Y_i) = 1 - q^{n/4}$ for all i. By linearity,

$$E(Z) = 2 + 4(1 - q^{n/2}) + n(1 - q^{n/4}).$$

(b) Without any groupings, n tests are required. We have

$$\begin{aligned} n - E(Z) &= n - 2 - 4(1 - q^{n/2}) - n(1 - q^{n/4}) \\ &= 4q^{n/2} + nq^{n/4} - 6 \\ &= \left(2q^{n/4} + \frac{n}{4}\right)^2 - \frac{n^2}{16} - 6. \end{aligned}$$

Thus $E(Z) > n$ iff

$$q^{n/4} < \frac{1}{2}\left(\frac{n^2}{16} + 6\right)^{1/2} - \frac{n}{8}.$$

(c) Let X_i denote the indicator function for the event that the blood test on group i is positive. The m groups are tested first, and then k tests are run for each group i with $X_i = 1$. Hence the total number of tests performed is $X = m + k(X_1 + \cdots + X_m)$. Each of the m samples tested is pooled from k individuals, so $E(X_i) = P(X_i = 1) = (1 - q^k)$ for all i by the same argument as in part (a). Thus

$$E(X) = m + kE(X_1) + \cdots + kE(X_m) = m + km(1 - q^k) = m + n(1 - q^k).$$

4.2 Exercise 3.2.16

(a) The probability $P(X_1 = k)$ is the probability that the first k cards are not aces and the $(k+1)$st card is an ace. There are 48 cards that are not aces among the 52 total cards in the deck, so

$$P(X_i = k) = P(X_1 = k) = \frac{48}{52} \cdot \frac{47}{51} \cdots \frac{48 - k + 1}{52 - k + 1} \cdot \frac{4}{52 - k} = \frac{\binom{48}{k}}{\binom{52}{k}} \frac{4}{52 - k}. \tag{1}$$

(b) It is clear that $X_1 + \cdots + X_5 = 48$. Since the X_i have the same distribution,

$$E(X_i) = \frac{1}{5}(E(X_1) + \cdots E(X_5)) = \frac{1}{5}E(X_1 + \cdots + X_5) = \frac{1}{5}E(48) = \frac{48}{5}$$

for all i.

(c) Fix some X_i and X_j. Since the X_k are non-negative integers with $X_1 + \cdots + X_5 = 48$, it follows that $X_i = 0$ if $X_j = 48$. Hence $P(X_i = 48, X_j = 1) = 0$. But $P(X_i = 48)$ and $P(X_j = 1)$ are nonzero by (**??**), so $P(X_i = 48, X_j = 1) \neq P(X_i = 48)P(X_j = 1)$. Thus X_i and X_j are not independent.

4.3 Review Exercise 3.22

(a) Assume that the demand X for newspapers follows a Poisson distribution with mean $\mu = 100$. Suppose the paperboy purchases n papers per day. Then he sells $Y_n = \min(X, n)$ papers per day, so his profit is $\pi_n = 25Y_n - 10n$ (in cents). We compute

$$\begin{aligned}
E(Y_n) &= \sum_k kP(Y_n = k) \\
&= \sum_{k=1}^{n-1} kP(\min(X,n) = k) + nP(\min(X,n) = n) \\
&= \sum_{k=1}^{n-1} kP(X = k) + nP(X \geq n) \\
&= \sum_{k=1}^{n-1} kP(X = k) + n - nP(X < n) \\
&= n + \sum_{k=1}^{n-1} kP(X = k) - n\sum_{k=0}^{n-1} P(X = k) \\
&= n + e^{-\mu}\sum_{k=1}^{n-1} \frac{\mu^k}{(k-1)!} - ne^{-\mu}\sum_{k=0}^{n-1} \frac{\mu^k}{k!} \\
&= n + \mu e^{-\mu}\sum_{k=1}^{n-1} \frac{\mu^{k-1}}{(k-1)!} - ne^{-\mu}\sum_{k=0}^{n-1} \frac{\mu^k}{k!} \\
&= n + (\mu - n)e^{-\mu}\sum_{k=1}^{n-2} \frac{\mu^k}{k!} - ne^{-\mu}\frac{\mu^{n-1}}{(n-1)!}.
\end{aligned}$$

For $n = \mu = 100$,

$$E(Y_n) = n - ne^{-\mu}\frac{\mu^{n-1}}{(n-1)!} \approx 96.0319$$

By linearity,

$$E(\pi_n) = 25E(Y_n) - 10n \approx 1400.3475$$

for $n = 100$.

(b) As shown in Figure 1 below, $E(\pi_n) = 25E(Y_n) - 10n$ increases to a maximum value of 1402.9150 at $n = 103$, then decreases approximately linearly for larger n. (The reason for this behavior is that $E(Y_n)$ rapidly approaches $E(X) = \mu$ as n increases, and so $E(\pi_{n+1}) \approx E(\pi_n) - 10$ for $E(Y_n) \approx \mu$.)

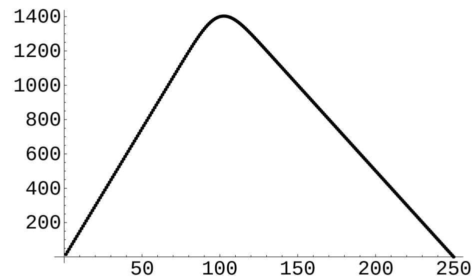

Figure 1: The long-term profit $E(\pi_n)$ as a function of n.

4.4 Exercise 3.3.20

(a) In case (1), the profit is 200, 100, 0, or -100 (in units of hundreds of dollars), each with probability $1/4$. The probability that the profit is at least 80 is therefore $1/2$.

(b) Let X_i be the profit from stock i, and set $X = X_1 + \cdots + X_{100}$. The mean μ of each X_i is given by

$$\mu = E(X_i) = \frac{1}{4}(2+1+0-1) = \frac{1}{2}.$$

The standard deviation σ of each X_i is given by

$$\sigma^2 = E((X-\mu)^2) = \frac{1}{4}\left[\left(\frac{3}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + \left(-\frac{3}{2}\right)^2\right] = \frac{5}{4}$$

Using the normal approximation,

$$\begin{aligned} P(X \geq a) &= P(a \leq X \leq \infty) \\ &= P\left(\frac{a-100\mu}{10\sigma} \leq \frac{X-100\mu}{10\sigma} \leq \infty\right) \\ &\approx \Phi(\infty) - \Phi\left(\frac{a-100\mu}{10\sigma}\right) \\ &= 1 - \Phi\left(\frac{a-100\mu}{10\sigma}\right). \end{aligned}$$

for any a. Hence the probability that the profit in case (2) is at least 80 is given by

$$P(X \geq 80) \approx 1 - \Phi\left(\frac{80-100\mu}{10\sigma}\right) = 1 - \Phi(2.68) = .0037.$$

4.5 Problem 5

Label the cameras $C_1, \ldots, C_{10}$, and let X_i denote the indicator for the event that camera C_i is never used throughout the 20 vacations. Since C_i is not used on any fixed vacation with probability $9/10$, it follows that $P(X_i = 1) = (9/10)^{20}$. Hence $E(X_i) = P(X_i = 1) = (9/10)^{20}$ for all i. Let X denote the number of cameras that were never used throughout the 20 vacations. Then $X = X_1 + \cdots + X_{10}$, so

$$E(X) = E(X_1 + \cdots + X_{10}) = E(X_1) + \cdots + E(X_{10}) = 10\left(\frac{9}{10}\right)^{20} = 1.2158.$$

4.6 Problem 6

Let U_1 and U_2 denote the two break points of the stick. Assuming that U_1 and U_2 are independent and uniformly distributed on $[0, 1]$, the average length of the middle portion of the stick is given by

$$\int_0^1 \int_0^1 |U_2 - U_1| \; dU_1 \, dU_2 = \int_0^1 \int_0^{U_2} (U_2 - U_1) \, dU_1 \, dU_2 - \int_0^1 \int_{U_2}^1 (U_2 - U_1) \, dU_1 \, dU_2 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

4.7 Bonus Problem

The solution will be posted separately.