Solutions to Review Problems for Exam 3

1. Suppose that a book with n pages contains on average λ misprints per page. What is the probability that there will be at least m pages which contain more than k missprints?

   ***Solution**:* Let Y denote the number of misprints in one page. Then, we may assume that Y follows a Poisson(λ) distribution; so that

   $$\Pr[Y = r] = \frac{\lambda^r}{r!}\, e^{-\lambda}, \quad \text{for } r = 0, 1, 2, \ldots$$

   Thus, the probability that there will be more than k missprints in a given page is

   $$\begin{aligned} p &= \sum_{r=k+1}^{\infty} \Pr[Y = r] \\ &= \sum_{r=k+1}^{\infty} \frac{\lambda^r}{r!}\, e^{-\lambda}. \end{aligned} \tag{1}$$

   Next, let X denote the number of the pages out of the n that contain more than k missprints. Then, $X \sim \text{Binomial}(n, p)$, where p is as given in (1). Then the probability that there will be at least m pages which contain more than k missprints is

   $$\Pr[X \geqslant m] = \sum_{\ell=m}^{n} \binom{n}{\ell} p^{\ell}(1-p)^{n-\ell},$$

   where

   $$p = \sum_{r=k+1}^{\infty} \frac{\lambda^r}{r!}\, e^{-\lambda}.$$

   □

2. Suppose that the total number of items produced by a certain machine has a Poisson distribution with mean λ, all items are produced independently of one another, and the probability that any given item produced by the machine will be defective is p.

   Let X denote the number of defective items produced by the machine.

   (a) Determine the marginal distribution of the number of defective items, X.

Solution: Let N denote the number of items produced by the machine. Then,

$$N \sim \text{Poisson}(\lambda), \tag{2}$$

so that

$$\Pr[N = n] = \frac{\lambda^n}{n!}\, e^{-\lambda}, \quad \text{for } n = 0, 1, 2, \ldots$$

Now, since all items are produced independently of one another, and the probability that any given item produced by the machine will be defective is p, X has a conditional distribution (conditioned on $N = n$) that is Binomial(n, p); thus,

$$\Pr[X = k \mid N = n] = \begin{cases} \dbinom{n}{k} p^k (1-p)^{n-k}, & \text{for } k = 0, 1, 2, \ldots, n; \\ \\ 0 & \text{elsewhere.} \end{cases} \tag{3}$$

Then,

$$\begin{aligned} \Pr[X = k] &= \sum_{n=0}^{\infty} \Pr[X = k, N = n] \\ &= \sum_{n=0}^{\infty} \Pr[N = n] \cdot \Pr[X = k \mid N = n], \end{aligned}$$

where $\Pr[X = k \mid N = n] = 0$ for $n < k$, so that, using (2) and (3),

$$\begin{aligned} \Pr[X = k] &= \sum_{n=k}^{\infty} \frac{\lambda^n}{n!}\, e^{-\lambda} \cdot \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{e^{-\lambda}}{k!}\, p^k \sum_{n=k}^{\infty} \lambda^n \frac{1}{(n-k)!} (1-p)^{n-k}. \end{aligned} \tag{4}$$

Next, make the change of variables $\ell = n - k$ in the last summation in (4) to get

$$\Pr[X = k] = \frac{e^{-\lambda}}{k!}\, p^k \sum_{\ell=0}^{\infty} \lambda^{\ell+k} \frac{1}{\ell!} (1-p)^{\ell},$$

so that

$$\begin{aligned}\Pr[X=k] &= \frac{(\lambda p)^k}{k!}\ e^{-\lambda}\sum_{\ell=0}^{\infty}\frac{1}{\ell!}[\lambda(1-p)]^\ell \\ &= \frac{(\lambda p)^k}{k!}\ e^{-\lambda}e^{\lambda(1-p)} \\ &= \frac{(\lambda p)^k}{k!}\ e^{-\lambda p},\end{aligned}$$

which shows that

$$X \sim \text{Poisson}(\lambda p). \tag{5}$$

□

(b) Let Y denote the number of non–defective items produced by the machine. Show that X and Y are independent random variables.

***Solution**:* Similar calculations to those leading to (5) show that

$$Y \sim \text{Poisson}(\lambda(1-p)), \tag{6}$$

since the probability of an item coming out non–defective is $1-p$.

Next, observe that $Y = N - X$ and compute the joint probability

$$\begin{aligned}\Pr[X=k, Y=\ell] &= \Pr[X=k, N=k+\ell] \\ &= \Pr[N=k+\ell]\cdot\Pr[X=k \mid N=k+\ell] \\ &= \frac{\lambda^{k+\ell}}{(k+\ell)!}\ e^{-\lambda}\cdot\binom{k+\ell}{k}p^k(1-p)^\ell\end{aligned}$$

by virtue of (2) and (3). Thus,

$$\begin{aligned}\Pr[X=k, Y=\ell] &= \frac{\lambda^{k+\ell}}{k!\ \ell!}\ e^{-\lambda}\ p^k(1-p)^\ell \\ &= \frac{\lambda^{k+\ell}}{k!\ \ell!}\ e^{-\lambda}\ p^k(1-p)^\ell,\end{aligned}$$

where

$$e^{-\lambda} = e^{-[p+(1-p)]\lambda} = e^{-p\lambda}\cdot e^{-(1-p)\lambda}.$$

Thus,

$$\begin{aligned}\Pr[X=k, Y=\ell] &= \frac{(p\lambda)^k}{k!}\ e^{-p\lambda}\cdot\frac{[(1-p)\lambda]^\ell}{\ell!}\ e^{-(1-p)\lambda} \\ &= p_X(k)\cdot p_Y(\ell),\end{aligned}$$

in view of (5) and (6). Hence, X and Y are independent. □

3. Suppose that the proportion of color blind people in a certain population is 0.005. Estimate the probability that there will be more than one color blind person in a random sample of 600 people from that population.

   ***Solution**:* Set $p = 0.005$ and $n = 600$. Denote by Y the number of color blind people in the sample. The, we may assume that $Y \sim \text{Binomial}(n, p)$. Since p is small and n is large, we may use the Poisson approximation to the binomial distribution to get
$$\Pr[Y = k] \approx \frac{\lambda^k}{k!}\, e^{-\lambda},$$
   where $\lambda = np = 3$.

   Then,
$$\begin{aligned} \Pr[Y > 1] &= 1 - \Pr[Y \leqslant 1] \\ &\approx 1 - e^{-3} - 3e^{-3} \\ &\approx 0.800852. \end{aligned}$$
   Thus, the probability that there will be more than one color blind person in a random sample of 600 people from that population is about 80%. □

4. An airline sells 200 tickets for a certain flight on an airplane that has 198 seats because, on average, 1% of purchasers of airline tickets do not appear for departure of their flight. Estimate the probability that everyone who appears for the departure of this flight will have a seat.

   ***Solution**:* Set $p = 0.01$, $n = 200$ and let Y denote the number of ticket purchasers that do not appear for departure. Then, we may assume that $Y \sim \text{Binomial}(n, p)$. We want to estimate the probability $\Pr[Y > 2]$. Using the Poisson(λ), with $\lambda = np = 2$, approximation to the distribution of Y we get
$$\begin{aligned} \Pr[Y \geqslant 2] &= 1 - \Pr[Y \leqslant 1] \\ &\approx 1 - e^{-2} - 2e^{-2} \\ &\approx 0.594. \end{aligned}$$
   Thus, the probability that everyone who appears for the departure of this flight will have a seat is about 59.4%. □

5. Let X denote a positive random variable such that $\ln(X)$ has a Normal(0, 1) distribution.

   (a) Give the pdf of X and compute its expectation.
   ***Solution**:* Set $Z = \ln(X)$, so that $Z \sim \text{Normal}(0,1)$; thus,

$$f_Z(y) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}, \qquad \text{for } z \in \mathbb{R}. \tag{7}$$

Next, compute the cdf for X,

$$F_X(x) = \Pr(X \leqslant x), \qquad \text{for } x > 0,$$

to get

$$\begin{aligned} F_X(x) &= \Pr[\ln(X) \leqslant \ln(x)] \\ &= \Pr[Z \leqslant \ln(x)] \\ &= F_Z(\ln(x)), \end{aligned}$$

so that

$$F_X(x) = \begin{cases} F_Z(\ln(x)), & \text{for } x > 0; \\ 0 & \text{for } x \leqslant 0. \end{cases} \tag{8}$$

Differentiating (8) with respect to x, for $x > 0$, we obtain

$$f_X(x) = F_Z'(\ln(x)) \cdot \frac{1}{x},$$

so that

$$f_X(x) = f_Z(\ln(x)) \cdot \frac{1}{x}, \tag{9}$$

where we have used the Chain Rule. Combining (7) and (9) yields

$$f_X(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\, x}\, e^{-(\ln x)^2/2}, & \text{for } x > 0; \\ 0 & \text{for } x \leqslant 0. \end{cases} \tag{10}$$

In order to compute the expected value of X, use the pdf in (10) to get

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x)\, dx \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-(\ln x)^2/2}\, dx. \end{aligned} \tag{11}$$

Make the change of variables $u = \ln x$ in the last integral in (11) to get $du = \dfrac{1}{x}dx$, so that $dx = e^u du$ and

$$E(X) \;=\; \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{u-u^2/2}\, du. \tag{12}$$

Complete the square in the exponent of the integrand in (12) to obtain

$$E(X) \;=\; e^{1/2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-(u-1)^2/2}\, du. \tag{13}$$

Next, make the change of variables $w = u - 1$ for the integral in (13) to get

$$E(X) \;=\; e^{1/2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-w^2/2}\, dw = \sqrt{e}.$$

□

(b) Estimate $\Pr(X \leq 6.5)$.
Solution: Use the result in (8) to compute

$$\Pr(X \leq 6.5) = F_Z(\ln(6.5)), \qquad \text{where } Z \sim \text{Normal}(0,1).$$

Thus,

$$\Pr(X \leq 6.5) \doteq F_Z(1.8718) \doteq 0.969383,$$

or about 97%. □

6. Forty seven digits are chosen at random and with replacement from $\{0, 1, 2, \ldots, 9\}$. Estimate the probability that their average lies between 4 and 6.

Solution: Let $X_1, X_2, \ldots, X_n$, where $n = 47$, denote the 47 digits. Since the sampling is done without replacement, the random variables $X_1, X_2, \ldots, X_n$ are identically uniformly distributed over the digits $\{0, 1, 2, \ldots, 9\}$; in other words, $X_1, X_2, \ldots, X_n$ is a random sample from the discrete Uniform(10) distribution. Consequently, the mean of the distribution is

$$\mu = \frac{10+1}{2} = 5.5, \tag{14}$$

and the variance is

$$\sigma^2 = \frac{(10+1)(10-1)}{12} = \frac{99}{12} = \frac{33}{4} = 8.25 \tag{15}$$

We would like to estimate

$$\Pr(4 \leqslant \overline{X}_n \leqslant 6),$$

or

$$\Pr(4 - \mu \leqslant \overline{X}_n - \mu \leqslant 6 - \mu),$$

where μ is given in (14), so that

$$\Pr(4 \leqslant \overline{X}_n \leqslant 6) = \Pr(-1.5 \leqslant \overline{X}_n - \mu \leqslant 0.5) \tag{16}$$

Next, divide the last inequality in (16) by $\sigma/\sqrt{n}$, where σ is as given in (15), to get

$$\Pr(4 \leqslant \overline{X}_n \leqslant 6) = \Pr\left(-3.58 \leqslant \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leqslant 1.19\right) \tag{17}$$

For n large (say, $n = 47$), we can apply the Central Limit Theorem to obtain from (17) that

$$\Pr(4 \leqslant \overline{X}_n \leqslant 6) \approx \Pr\left(-3.58 \leqslant Z \leqslant 1.19\right), \quad \text{where } Z \sim \text{Normal}(0, 1). \tag{18}$$

It follows from (18) and the definition of the cdf that

$$\Pr(4 \leqslant \overline{X}_n \leqslant 6) \approx F_z(1.19) - F_z(-3.58), \tag{19}$$

where F_z is the cdf of $Z \sim \text{Normal}(0, 1)$. Using the symmetry of the pdf of $Z \sim \text{Normal}(0, 1)$, we can re–write (19) as

$$\Pr(4 \leqslant \overline{X}_n \leqslant 6) \approx F_z(1.19) + F_z(3.58) - 1. \tag{20}$$

Finally, using a table of standard normal probabilities, we obtain from (20) that

$$\Pr(4 \leqslant \overline{X}_n \leqslant 6) \approx 0.8830 + 0.9998 - 1 = 0.8828.$$

Thus, the probability that the average of the 47 digits is between 4 and 6 is about 88.3%. □

7. Let $X_1, X_2, \ldots, X_{30}$ be independent random variables each having a discrete distribution with pmf:

$$p(x) = \begin{cases} 1/4, & \text{if } x = 0 \text{ or } x = 2; \\ 1/2, & \text{if } x = 1; \\ 0, & \text{otherwise.} \end{cases}$$

Estimate the probability that $X_1 + X_2 + \cdots + X_{30}$ is at most 33.

***Solution**:* First, compute the mean, $\mu = E(X)$, and variance, $\sigma^2 = \text{Var}(X)$, of the distribution:

$$\mu = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2\frac{1}{4} = 1. \tag{21}$$

$$\sigma^2 = E(X^2) - [E(X)]^2, \tag{22}$$

where

$$E(X^2) = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2\frac{1}{4} = 1.5; \tag{23}$$

so that, combining (21), (22) and (23),

$$\sigma^2 = 1.5 - 1 = 0.5. \tag{24}$$

Next, let $Y = \displaystyle\sum_{k=1}^{n} X_k$, where $n = 30$. We would like to estimate

$$\Pr[Y \leqslant 33]. \tag{25}$$

By the Central Limit Theorem

$$\Pr\left(\frac{Y - n\mu}{\sqrt{n}\ \sigma} \leqslant\right) \approx \Pr(Z \leqslant z), \qquad \text{for } z \in \mathbb{R}, \tag{26}$$

where $Z \sim \text{Normal}(0, 1)$, $\mu = 1$, $\sigma^2 = 1.5$ and $n = 30$. It follows from (26) that we can estimate the probability in (25) by

$$\Pr[Y \leqslant 33] \approx \Pr(Z \leqslant 0.77) \doteq 0.7794. \tag{27}$$

Thus, according to (27), the probability that $X_1 + X_2 + \cdots + X_{30}$ is at most 33 is about 78%. □

8. Roll a balanced die 36 times. Let Y denote the sum of the outcomes in each of the 36 rolls. Estimate the probability that $108 \leq Y \leq 144$.

   Suggestion: Since the event of interest is $(Y \in \{108, 109, \ldots, 144\})$, rewrite $\Pr(108 \leq Y \leq 144)$ as

   $$\Pr(107.5 < Y \leqslant 144.5).$$

   ***Solution**:* Let $X_1, X_2, \ldots, X_n$, where $n = 36$, denote the outcomes of the 36 rolls. Since we are assuming that the die is balanced, the random variables

$X_1, X_2, \ldots, X_n$ are identically uniformly distributed over the digits $\{1, 2, \ldots, 6\}$; in other words, $X_1, X_2, \ldots, X_n$ is a random sample from the discrete Uniform(6) distribution. Consequently, the mean of the distribution is

$$\mu = \frac{6+1}{2} = 3.5, \tag{28}$$

and the variance is

$$\sigma^2 = \frac{(6+1)(6-1)}{12} = \frac{35}{12}. \tag{29}$$

We also have that

$$Y = \sum_{k=1}^{n} X_k,$$

where $n = 36$.

By the Central Limit Theorem,

$$\Pr(107.5 < Y \leqslant 144.5) \approx \Pr\left(\frac{107.5 - n\mu}{\sqrt{n}\sigma} < Z \leqslant \frac{144.5 - n\mu}{\sqrt{n}\sigma}\right), \tag{30}$$

where $Z \sim \text{Normal}(0, 1)$, $n = 36$, and μ and σ are given in (28) and (29), respectively. We then have from (30) that

$$\begin{aligned} \Pr(107.5 < Y \leqslant 144.5) &\approx \Pr(-1.81 < Z \leqslant 1.81) \\ &\approx F_Z(1.81) - F_Z(-1.81) \\ &\approx 2F_Z(1.81) - 1 \\ &\approx 2(0.9649) - 1 \\ &\approx 0.9298; \end{aligned}$$

so that the probability that $108 \leqslant Y \leqslant 144$ is about 93%. □

9. Let $Y \sim \text{Binomial}(100, 1/2)$. Use the Central Limit Theorem to estimate the value of $\Pr(Y = 50)$.

Solution: We use the so–called continuity correction and estimate

$$\Pr(49.5 < Y \leqslant 50.5).$$

By the Central Limit Theorem,

$$\Pr(49.5 < Y \leqslant 50.5) \approx \Pr\left(\frac{49.5 - n\mu}{\sqrt{n}\sigma} < Z \leqslant \frac{50.5 - n\mu}{\sqrt{n}\sigma}\right), \tag{31}$$

where $Z \sim \text{Normal}(0, 1)$, $n = 100$, and $n\mu = 50$ and

$$\sigma = \sqrt{\frac{1}{2}\left(1 - \frac{1}{2}\right)} = \frac{1}{2}.$$

We then obtain from (31) that

$$\begin{aligned} \Pr(49.5 < Y \leqslant 50.5) &\approx \Pr(-0.1 < Z \leqslant 0.1) \\ &\approx F_Z(0.1) - F_Z(-0.1) \\ &\approx 2F_Z(0.1) - 1 \\ &\approx 2(0.5398) - 1 \\ &\approx 0.0796. \end{aligned}$$

Thus,

$$\Pr(Y = 50) \approx 0.08,$$

or about 8%. □

10. Let $Y \sim \text{Binomial}(n, 0.55)$. Find the smallest value of n such that, approximately,

$$\Pr(Y/n > 1/2) \geqslant 0.95. \tag{32}$$

***Solution**:* By the Central Limit Theorem,

$$\frac{\dfrac{Y}{n} - 0.55}{\sqrt{(0.55)(1 - 0.55)}/\sqrt{n}} \xrightarrow{D} Z \sim \text{Normal}(0, 1) \text{ as } n \to \infty. \tag{33}$$

Thus, according to (32) and (33), we need to find the smallest value of n such that

$$\Pr\left(Z > \frac{0.5 - 0.55}{(0.4975)/\sqrt{n}}\right) \geqslant 0.95,$$

or

$$\Pr\left(Z > -\frac{\sqrt{n}}{10}\right) \geqslant 0.95. \tag{34}$$

The expression in (34) is equivalent to

$$1 - \Pr\left(Z \leqslant -\frac{\sqrt{n}}{10}\right) \geqslant 0.95,$$

which can be re-written as

$$1 - F_{Z}\left(-\frac{\sqrt{n}}{10}\right) \geqslant 0.95, \tag{35}$$

where F_{Z} is the cdf of $Z \sim \text{Normal}(0, 1)$.
By the symmetry of the pdf for $Z \sim \text{Normal}(0, 1)$, (35) is equivalent to

$$F_{Z}\left(\frac{\sqrt{n}}{10}\right) \geqslant 0.95. \tag{36}$$

The smallest value of n for which (36) holds true occurs when

$$\frac{\sqrt{n}}{10} \geqslant z^*, \tag{37}$$

where z^* is a positive real number with the property

$$F_{Z}(z^*) = 0.95. \tag{38}$$

The equality in (38) occurs approximately when

$$z^* = 1.645. \tag{39}$$

It follows from (37) and (39) that (32) holds approximately when

$$\frac{\sqrt{n}}{10} \geqslant 1.645,$$

or $n \geqslant 270.6025$. Thus, $n = 271$ is the smallest value of n such that, approximately,

$$\Pr(Y/n > 1/2) \geqslant 0.95.$$

□

11. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean λ. Thus, $Y = \sum_{i=1}^{n} X_i$ has a Poisson distribution with mean $n\lambda$. Moreover, by the Central Limit Theorem, $\overline{X} = Y/n$ has, approximately, a Normal$(\lambda, \lambda/n)$ distribution for large n. Show that $u(Y/n) = \sqrt{Y/n}$ is a function of Y/n which is essentially free of λ.

    ***Solution**:* We will show that, for large values of n, the distribution of

    $$2\sqrt{n}\left(\sqrt{\frac{Y}{n}} - \sqrt{\lambda}\right) \tag{40}$$

    is independent of λ. In fact, we will show that, for large values of n, the distribution of the random variables in (40) can be approximated by a Normal$(0, 1)$ distribution.

    First, note that by the Law of Large Numbers,

    $$\frac{Y}{n} \xrightarrow{\text{Pr}} \lambda, \qquad \text{as } n \to \infty.$$

    Thus, for large values of n, we can approximate $u(Y/n)$ by its linear approximation around λ

    $$u(Y/n) \approx u(\lambda) + u'(\lambda)\left(\frac{Y}{n} - \lambda\right), \tag{41}$$

    where

    $$u'(\lambda) = \frac{1}{2\sqrt{\lambda}}. \tag{42}$$

    Combining (41) and (42) we see that, for large values of n,

    $$\sqrt{\frac{Y}{n}} - \sqrt{\lambda} \approx \frac{1}{2\sqrt{n}} \cdot \frac{\frac{Y}{n} - \lambda}{\sqrt{\frac{\lambda}{n}}}. \tag{43}$$

    Since, by the Central Limit Theorem,

    $$\frac{\frac{Y}{n} - \lambda}{\sqrt{\frac{\lambda}{n}}} \xrightarrow{D} Z \sim \text{Normal}(0, 1) \text{ as } n \to \infty, \tag{44}$$

it follows from (43) and (44) that

$$2\sqrt{n}\left(\sqrt{\frac{Y}{n}}-\sqrt{\lambda}\right) \stackrel{D}{\longrightarrow} Z \sim \text{Normal}(0,1) \text{ as } n \to \infty,$$

which was to be shown. $\square$