Homework 1
Due Sep 22nd Midnight

1. Let Y be a random variable with p.d.f. $2ce^{-by}$ for $y \geq 0$, where b, c is some constant.
   (a) Find the condition that b, c need to be satisfied
   (b) Assume $b = 4$, what is the mean and variance of Y
   (c) Compute $\Pr\{Y > 4 \mid Y > 1\}$

   Solution

   (a) $\int_0^\infty \Pr\{Y = y\}dy = \frac{2c}{b} = 1$, thus $b = 2c$
   (b) When $b = 4$, we have $\Pr\{Y = y\} = 4e^{-4y}$, then $\mathrm{E}[Y] = \frac{1}{4}$ and $\mathrm{var}[Y] = \frac{1}{16}$
   (c) $\Pr\{Y > 4 \mid Y > 1\} = \frac{P(Y>4,Y>1)}{P(Y>1)} = \frac{P(Y>4)}{P(Y>1)} = \frac{\int_4^\infty be^{-by}dy}{\int_1^\infty be^{-by}dy} = \frac{e^{-4b}}{e^{-b}} = e^{-3b} = \Pr\{Y > 3\}$

2. Let X be a Poisson random variable with mean 3 , and let $Y = \max(X, 4)$.
   (a) What is the p.m.f. of X
   (b) What is the mean of X
   (c) What is the variance of X
   (d) What is the p.m.f. of Y
   (e) Compute E $[Y]$

   Solution

   (a) $\Pr\{X = n\} = \frac{e^{-3}3^n}{n!}$
   (b) $\mathrm{E}[x] = 3$
   (c) $\mathrm{var}[x] = 3$
   (d) $\Pr\{Y = n\} = \Pr\{X = n\} = \frac{e^{-3}3^n}{n!} \quad \forall n > 4, \quad \Pr\{Y = 4\} = \sum_{n=1}^{4} \Pr\{X = n\} = \frac{131}{8}e^{-3}$
   (e) $\mathrm{E}[Y] = \mathrm{E}[X] + \sum_{n=0}^{3}(4 - n) \cdot \Pr\{X = n\} = 3 + \frac{53}{2}e^{-3}$

3. At 5pm each day, four buses make their way to a bus stop. Each bus would be acceptable to take you home. The time in hours (after 5 pm) that each arrives at the stop is independent with $Y_1, Y_2, Y_3, Y_4 \sim \mathrm{Exp}(\lambda = 6)$ (on average, it takes 1/6 of an hour (10 minutes) for each bus to arrive).
   (a) On Mondays, you want to get home ASAP, so you arrive at the bus stop at 5pm sharp. What is the expected time until the first one arrives?
   (b) On Tuesdays, you have a lab meeting that runs until $5:15$ and are worried you may not catch any bus. What is the probability you miss all the buses?

   Solution

$$(a) f_{Y_{(1)}}(y) = \binom{4}{1-1, 1, 4-1} \cdot \left[1 - e^{-6y}\right]^{1-1} \cdot \left[e^{-6y}\right]^{4-1} \cdot 6e^{-6y} = 4\left[e^{-18y}\right]6e^{-6y} = 24e^{-24y}$$

$$(b) f_{Y_{(4)}}(y) = \binom{4}{4-1, 1, 4-4} \cdot \left[1 - e^{-6y}\right]^{4-1} \cdot \left[e^{-6y}\right]^{4-4} \cdot 6e^{-6y} = \left[1 - e^{-6y}\right]^3 6e^{-6y}$$

$$\mathbb{P}\left(Y_{\max} \leq 0.25\right) = \int_0^{0.25} f_{Y_{\max}}(y)dy = \int_0^{0.25} \left[1 - e^{-6y}\right]^3 6e^{-6y}dy$$

4. Let W_1, W_2, and W_3 be random variables, each of which is greater than 1 with probability 1, and suppose that these random variables have a joint density function. Set $Y_1 = W_1, Y_2 = W_1W_2$, and $Y_3 = W_1W_2W_3$. Observe that $1 < Y_1 < Y_2 < Y_3$ with probability 1.

   (a) Determine a formula for the joint density function of Y_1, Y_2, and Y_3 in terms of the joint density function of W_1, W_2, and W_3.

   (b) Suppose that W_1, W_2, and W_3 are independent random variables, each having the density function that equals w^{-2} for $w > 1$ and equals 0 otherwise. Determine the joint density function of Y_1, Y_2, and Y_3.

   (c) (Continued) Are Y_1, Y_2, and Y_3 independent (why or why not)?

   Solution

   (a) Let $w_1, w_2, w_3 > 1$. The equations $y_1 = w_1, y_2 = w_1w_2$, and $y_3 = w_1w_2w_3$ have the unique solution given by $w_1 = y_1, w_2 = y_2/y_1$, and $w_3 = y_3/y_2$. Observe that $\frac{\partial(w_1,w_2,w_3)}{\partial(y_1,y_2,y_3)} = \begin{vmatrix} 1 & 0 & 0 \\ -y_2/y_1^2 & 1/y_1 & 0 \\ 0 & -y_3/y_2^2 & 1/y_2 \end{vmatrix} = \frac{1}{y_1y_2}$. Thus $f_{Y_1,Y_2,Y_3}(y_1, y_2, y_3) = \frac{1}{y_1y_2} f_{W_1,W_2,W_3}(y_1, y_2/y_1, y_3/y_2)$ for $1 < y_1 < y_2 < y_3$, and this joint density function equals zero elsewhere.

   (b) The joint density function of Y_1, Y_2, and Y_3 is given by $f_{Y_1,Y_2,Y_3}(y_1, y_2, y_3 = \frac{1}{y_1y_2y_1^2(y_2/y_1)^2(y_3/y_2)^2} = \frac{1}{y_1y_2y_3^2}$ for $1 < y_1 < y_2 < y_3$, and this joint density function equals zero elsewhere.

   (c) The random variables Y_1, Y_2, and Y_3 are dependent. Their joint density function appears to factor as $\left(\frac{1}{y_1}\right)\left(\frac{1}{y_2}\right)\left(\frac{1}{y_3^2}\right)$, but this reasoning ignores the fact that the range $1 < y_1 < y_2 < y_3$ is not a rectangle with sides parallel to the coordinate axes and hence that the corresponding indicator function does not factor. Alternatively, $P(2 < Y_1 < 3) > 0$ and $P(1 < Y_2 < 2) > 0$, but $P(2 < Y_1 < 3 \text{ and } 1 < Y_2 < 2) = 0$, so Y_1 and Y_2 are dependent.

5. Let the joint distribution of Y_1, Y_2 and Y_3 be multinomial (trinomial) with parameters $n = 100, \pi_1 = 0.2, \pi_2 = 0.35$ and $\pi_3 = 0.45$.

   (a) Justify normal approximation to the distribution of $Y_1 + Y_2 - Y_3$.

   (b) Use normal approximation to determine $P(Y_3 \geq Y_1 + Y_2)$.

   Solution

   (a) Now $Y_1+Y_2+Y_3 = 100$, so $Y_1+Y_2-Y_3 = Y_1+Y_2+Y_1+Y_2-100 = 2(Y_1+Y_2)-100$. (Alternatively, $Y_1 + Y_2 - Y_3 = 100 - 2Y_3$). Observe that $Y_1 + Y_2$ has the binomial distribution with parameters $n = 100$ and $\pi = .55$, which is approximately normally distributed by the Central Limit Theorem, so $Y_1 + Y_2 - Y_3 = 2(Y_1 + Y_2) - 100$ is approximately normally distributed.

   (b) Now $P(Y_3 \geq Y_1 + Y_2) = P(Y_1 + Y_2 - Y_3 \leq 0) = P(2(Y_1 + Y_2) - 100 \leq 0) = P(Y_1 + Y_2 \leq 50)$, while $Y_1 + Y_2$ has mean $100(.55) = 55$ and standard deviation $\sqrt{100(.55)(.45)} \doteq 4.975$. Thus, by the Central Limit Theorem with the half-integer correction **(half-integer correction NOT necessary for full mark)**, $P(Y_3 \geq Y_1 + Y_2) = P(Y_1 + Y_2 \leq 50) \approx \Phi\left(\frac{50+\frac{1}{2}-55}{4.975}\right) \doteq \Phi(-0.9045) = 1 - \Phi(0.9045) \doteq 1 - .8171 = .1829$.

   Alternatively

   (a) Since (Y_1, Y_2, Y_3) follows a multinomial distribution with parameters $n = 100$, $\pi_1 = 0.2$, $\pi_2 = 0.35$, and $\pi_3 = 0.45$, each Y_i is a count of successes over 100 trials with corresponding probabilities of success π_i.

   Means:

$$E(Y_1) = n\pi_1 = 100 \times 0.2 = 20, \quad E(Y_2) = 100 \times 0.35 = 35, \quad E(Y_3) = 100 \times 0.45 = 45$$

Variances:

$$\mathrm{Var}(Y_i) = n\pi_i(1-\pi_i)$$

Specifically:

$$\mathrm{Var}(Y_1) = 100 \times 0.2 \times 0.8 = 16,$$

$$\mathrm{Var}(Y_2) = 100 \times 0.35 \times 0.65 = 22.75,$$

$$\mathrm{Var}(Y_3) = 100 \times 0.45 \times 0.55 = 24.75$$

Covariances: For $i \neq j$, the covariances are given by:

$$\mathrm{Cov}(Y_i, Y_j) = -n\pi_i\pi_j$$

Thus:

$$\mathrm{Cov}(Y_1, Y_2) = -100 \times 0.2 \times 0.35 = -7,$$

$$\mathrm{Cov}(Y_1, Y_3) = -100 \times 0.2 \times 0.45 = -9,$$

$$\mathrm{Cov}(Y_2, Y_3) = -100 \times 0.35 \times 0.45 = -15.75$$

Mean and Variance of $Z = Y_1 + Y_2 - Y_3$:

$$E(Z) = E(Y_1) + E(Y_2) - E(Y_3) = 20 + 35 - 45 = 10$$

Using the variance-covariance formula:

$$\mathrm{Var}(Z) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) + \mathrm{Var}(Y_3) + 2 \cdot \mathrm{Cov}(Y_1, Y_2) - 2 \cdot \mathrm{Cov}(Y_1, Y_3) - 2 \cdot \mathrm{Cov}(Y_2, Y_3)$$

Substituting the values:

$$\mathrm{Var}(Z) = 16 + 22.75 + 24.75 + 2 \times (-7) - 2 \times (-9) - 2 \times (-15.75)$$

$$\mathrm{Var}(Z) = 63.5 - 14 + 18 + 31.5 = 99$$

Since the multinomial distribution involves **a large number of trials ($n = 100$) and the individual probabilities are not too small,** the Central Limit Theorem (CLT) justifies approximating Z with a normal distribution:

$$Z \sim N(10, 99)$$

(b) We are asked to find the probability:

$$P(Y_3 \geq Y_1 + Y_2) = P(Y_3 - Y_1 - Y_2 \geq 0)$$

Define $W = Y_3 - (Y_1 + Y_2)$. From part (a), we know:

$$W = -Z \quad \text{where} \quad Z \sim N(10, 99)$$

Thus:

$$W \sim N(-10, 99)$$

We need:

$$P(W \geq 0) = P\left(N(-10, 99) \geq 0\right)$$

Standardizing the normal variable:

$$P\left(\frac{W+10}{\sqrt{99}} \geq \frac{0+10}{\sqrt{99}}\right) = P\left(Z \geq \frac{10}{\sqrt{99}}\right)$$

$$\frac{10}{\sqrt{99}} \approx 1.01$$

Using standard normal tables:

$$P(Z \geq 1.01) \approx 0.1562$$

Thus:

$$P(Y_3 \geq Y_1 + Y_2) \approx 0.1562$$

- **(a):** The normal approximation is justified by the Central Limit Theorem, since the multinomial distribution involves a large number of trials ($n = 100$).
- **(b):** Using the normal approximation, $P(Y_3 \geq Y_1 + Y_2) \approx 0.1562$.