W271-2 – Spring 2016 – **Lab 1**

Juanjo Carin, Kevin Davis, Ashley Levato, Minghu Song

January 14, 2016

Contents

Part I: Marginal, Joint, and Conditional Probabilities

Question 1

In a team of data scientists, 36 are expert in machine learning, 28 are expert in statistics, and 18 are awesome. 22 are expert in both machine learning and statistics, 12 are expert in machine learning and are awesome, 9 are expert in statistics and are awesome, and 48 are expert in either machine learning or statistics or are awesome. Suppose you are in a cocktail party with this group of data scientists and you have an equal probability of meeting any one of them.

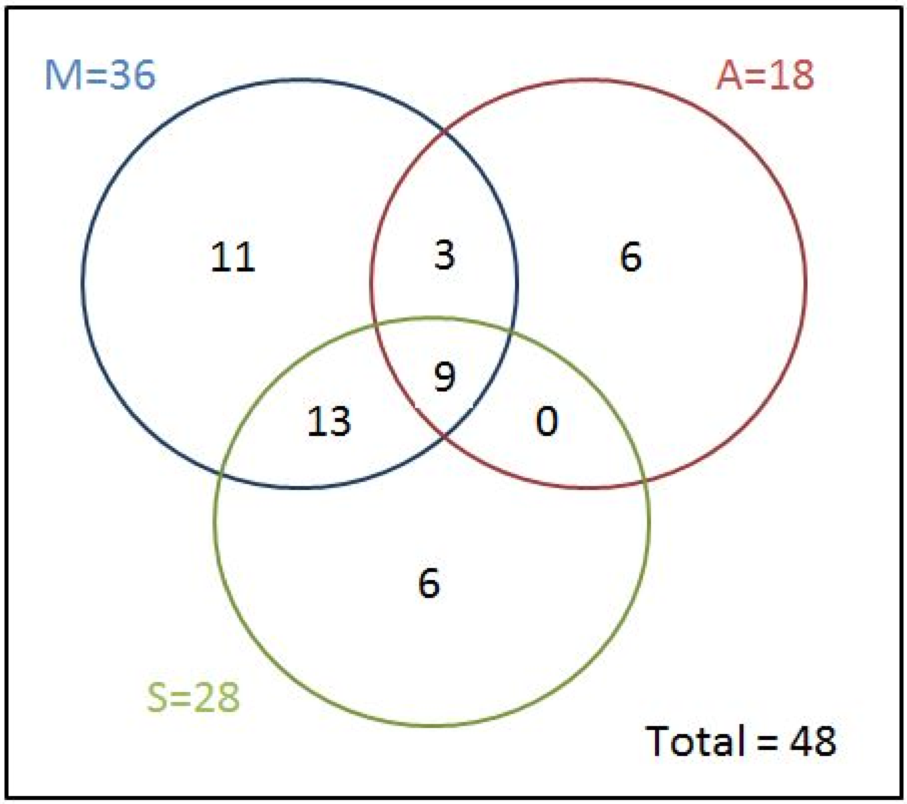

Figure 1: Venn Diagram

1. **What is the probability of meeting a data scientist who is an expert in both machine learning and statistics and is awesome?**

Let N be the size of the team of data scientists (i.e., the sample size), and M, S, A the event that a data scientist is either a machine learning expert, a statistics expert, or awesome.

$$\Pr(\mathbf{M} \cap \mathbf{S} \cap \mathbf{A}) = \Pr(M) + \Pr(S \cap A) - \Pr(M \cup (S \cap A))$$

$$= \Pr(M) + \Pr(S \cap A) - \Pr((M \cup S) \cap (M \cup A))$$

$$= \Pr(M) + \Pr(S \cap A) - ((\Pr(M \cup S) + \Pr(M \cup A) - \Pr((M \cup S) \cup (M \cup A))))$$

$$= \Pr(M) + \Pr(S \cap A) - (\Pr(M) - \Pr(S) + \Pr(M \cap S)) - (\Pr(M) - \Pr(A) + \Pr(M \cap A)) + \Pr(M \cup S \cup A)$$

$$= \Pr(M \cap S) + \Pr(M \cap A) + \Pr(M \cap A) - \Pr(M) - \Pr(S) - \Pr(A) + \Pr(M \cup S \cup A)$$

$$= \frac{22}{N} + \frac{12}{N} + \frac{9}{N} - \frac{36}{N} - \frac{28}{N} - \frac{18}{N} + \frac{48}{N} = \frac{(22+12+9)-(36-28-18)+48}{N} = \frac{43-82+48}{N}$$
$$= \frac{\mathbf{9}}{\mathbf{N}} = \frac{\mathbf{9}}{\mathbf{48}}$$

2. **Suppose you meet a data scientist who is an expert in machine learning. Given this information, what is the probability that s/he is not awesome?**

$$\Pr(\mathbf{A^c}|\mathbf{M}) = 1 - \Pr(A|M) = 1 - \frac{\Pr(A \cap M)}{\Pr(M)} = 1 - \frac{\frac{12}{N}}{\frac{36}{N}} = 1 - \frac{12}{36} = 1 - \frac{1}{3} = \frac{\mathbf{2}}{\mathbf{3}} = \mathbf{0.6667}$$

3. **Suppose the you meet a data scientist who is awesome. Given this information, what is the probability that s/he is an expert in either machine learning or statistics?**

$$\Pr(\mathbf{M} \cup \mathbf{S}|\mathbf{A}) = \frac{\Pr((M \cup S) \cap A)}{\Pr(A)} = \frac{\Pr((M \cap A) \cup (S \cap A))}{\Pr(A)}$$
$$= \frac{\Pr(M \cap A) + \Pr(S \cap A) - \Pr((M \cap A) \cap (S \cap A)}{\Pr(S)} = \frac{\Pr(M \cap A) + \Pr(S \cap A) - \Pr(M \cap S \cap A)}{\Pr(S)}$$
$$= \frac{\frac{12}{N} + \frac{9}{N} - \frac{9}{N}}{\frac{18}{N}} = \frac{12+9-9}{18} = \frac{12}{18}$$
$$= \frac{\mathbf{2}}{\mathbf{3}} = \mathbf{0.6667}$$

Alternatively, if the ambiguity in the question is intending to mean the individual can not be experts in both machine learning and statisitcs given that s/he is awesome then the probability would be $= \frac{\mathbf{3}}{\mathbf{18}}$.

Question 2

Suppose for events A and B, $\Pr(A) = p \leq \frac{1}{2}, \Pr(B) = q$, where $\frac{1}{4} < q < \frac{1}{2}$. These are the only information we have about the events.

1. **What are the maximum and minimum possible values for $\Pr(A \cup B)$?**

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

The maximum value of $\Pr(A \cup B)$ occurs when $A \cap B$ is the smallest set possible. In this case, since $\Pr(A) \leq \frac{1}{2}$ and $\Pr(B) < \frac{1}{2}$, it would be $A \cap B = \varnothing$ (if A and B were disjoint sets, which might be the case), so:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = \Pr(A) + \Pr(B) - \Pr(\varnothing) = p + q - 0 = p + q$$

which could have a maximum value close to 1 (i.e., $A \cup B \approx \Omega$), in case $\Pr(A) = \frac{1}{2}$ and $\Pr(B) \approx \frac{1}{2}$.

The minimum value of $\Pr(A \cup B)$ occurs when $A \cup B$ is the largest set possible, A or B. In this case, since $\Pr(A)$ does not have a lower bound, that would happen when $A \subseteq B$ and (consequently) $A \cap B = A$, which would lead to:

$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A) = \Pr(B) = q$

whose minimum value is greater than $\frac{1}{4}$.

In summary,

$\mathbf{\frac{1}{4}} < \Pr(\mathbf{A} \cup \mathbf{B}) < \mathbf{1}$

2. **What are the maximum and minimum possible values for $\Pr(A|B)$?**

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

If $B \subseteq A$ (which would imply that the lower bound for p is also $\frac{1}{4}$), then:

$\Pr(A|B) = \frac{\Pr(B)}{\Pr(B)} = 1$

As seen in the previous part, since $\Pr(A) \leq \frac{1}{2}$ and $Pr(B) < \frac{1}{2}$, it might occur that $A \cap B = \varnothing$, and hence:

$\Pr(A|B) = \frac{\Pr(\varnothing)}{\Pr(B)} = \frac{0}{q} = 0$

irrespective of the value of q.

In summary,

$\mathbf{0} \leq \Pr(\mathbf{A}|\mathbf{B}) \leq \mathbf{1}$

Part II: Random Variables, Expectation, Conditional Exp.

Question 3

Suppose the life span of a particular server is a continuous random variable, t, with a uniform probability distribution between 0 and k year, where $k \leq 10$ is a positive integer.

The server comes with a contract that guarantees a full or partial refund, depending on how long it lasts. Specifically, if the server fails in the first year, it gives a full refund denoted by θ. If it lasts more than 1 year but fails before $\frac{k}{2}$ years, the manufacturer will pay $x = \$A(k-t)^{1/2}$, where A is some positive constant equal to 2 if $t \leq \frac{k}{2}$. If it lasts between $\frac{k}{2}$ and $\frac{3k}{4}$ years, it pays $\frac{\theta}{10}$.

$$x = \begin{cases} \theta, & \text{if } t \leq 1 \\ 2(k-t)^{1/2}, & \text{if } 1 < t \leq k/2 \\ \theta/10, & \text{if } k/2 < t \leq 3k/4 \\ 0 & \text{otherwise} \end{cases}$$

1. **Given that the server lasts for $\frac{k}{4}$ years without failing, what is the probability that it will last another year?**

$$\Pr\left(\mathbf{t} \leq \mathbf{1} + \frac{\mathbf{k}}{\mathbf{4}} \middle| \mathbf{t} \geq \frac{\mathbf{k}}{\mathbf{4}}\right) = \frac{\Pr\left(\left(t \leq 1 + \frac{k}{4}\right) \bigcap \left(t \geq \frac{k}{4}\right)\right)}{\Pr\left(t \geq \frac{k}{4}\right)} = \frac{\Pr\left(\frac{k}{4} \leq t \leq 1 + \frac{k}{4}\right)}{Pr\left(t \geq \frac{k}{4}\right)}$$

$$= \frac{\Pr\left(t \leq 1 + \frac{k}{4}\right) - \Pr\left(t \leq \frac{k}{4}\right)}{1 - \Pr\left(t \leq \frac{k}{4}\right)} = \frac{\frac{1}{k}\left(1 + \frac{k}{4} - \frac{k}{4}\right)}{1 - \frac{1}{k}\frac{k}{4}} = \frac{\frac{1}{k}}{\frac{3}{4}}$$

$$= \frac{\mathbf{4}}{\mathbf{3k}}$$

2. **Compute the expected payout from the contract, $E(x)$.**

Below are 3 different approaches, all leading to the same result.

The first one is based on https://www.probabilitycourse.com/chapter4/4_3_1_mixed.php. Before computing the expected value of X, we need to compute its *cdf*.

t (the life span of the server) is a continuous variable with the following *pdf*:

$$f_t(t) = \begin{cases} \frac{1}{k} & 0 \leq t \leq \ k \\ 0 & \text{otherwise} \end{cases}$$

But X (the payout or refund) is not continuous:

$$X = g(t) = \begin{cases} \theta & 0 \le t \le 1 \\ A\sqrt{k-t} & 1 \le t \le \dfrac{k}{2} \\ \dfrac{\theta}{10} & \dfrac{k}{2} \le t \le \dfrac{3k}{4} \\ 0 & \text{otherwise} \end{cases}$$

(The value of X for $t = 1$ is $A\sqrt{k-1}$, which is not necessarily equal to θ; the value of X for $t = k/2$ is $A\sqrt{k/2}$, which again is not necessarily equal to $\theta/10$.)

First we compute the probability that X takes its possible discrete values:

$\Pr(X = 0) = \Pr\left(t \ge \dfrac{3k}{4}\right) = 1 - F_t\left(\dfrac{3k}{4}\right) = 1 - \dfrac{1}{k}\dfrac{3k}{4} = 1 - \dfrac{3}{4} = \dfrac{1}{4}$

$\Pr\left(X = \dfrac{\theta}{10}\right) = \Pr\left(\dfrac{k}{2} \le t \le \dfrac{3k}{4}\right) = F_t\left(\dfrac{3k}{4}\right) - F_t\left(\dfrac{k}{2}\right) = \dfrac{1}{k}\dfrac{3k}{4} - \dfrac{1}{k}\dfrac{k}{2} = \dfrac{3}{4} - \dfrac{1}{2} = \dfrac{3-2}{4} = \dfrac{1}{4}$

$\Pr(X = \theta) = \Pr(0 \le t \le 1) = F_t(1) - F_t(0) = \dfrac{1}{k} - \dfrac{0}{k} = \dfrac{1}{k}$

For $A\sqrt{\dfrac{k}{2}} \le X \le A\sqrt{k-1}$, we can compute the *cdf* of X as follows:

$$\begin{aligned} F_X(x) &= \Pr(X \le x) = \Pr\left(A\sqrt{k-t} \le x\right) \\ &= \Pr\left(k - t \le \left(\tfrac{x}{A}\right)^2\right) = \Pr\left(t \ge k - \left(\tfrac{x}{A}\right)^2\right) = 1 - \Pr\left(t \le k - \left(\tfrac{x}{A}\right)^2\right) \\ &= 1 - \int_{t=0}^{k-\left(\frac{x}{A}\right)^2} \frac{1}{k} dt = 1 - \frac{1}{k}\left(k - \left(\frac{x}{A}\right)^2\right) = \frac{x^2}{kA^2} \end{aligned}$$

So the overall expression for $F_X(x)$ is:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \dfrac{1}{4} & 0 \le x < \dfrac{\theta}{10} \\ \dfrac{1}{2} & \dfrac{\theta}{10} \le x \le A\sqrt{\dfrac{k}{2}} \\ \dfrac{x^2}{kA^2} & A\sqrt{\dfrac{k}{2}} \le x \le A\sqrt{k-1} \\ 1 - \dfrac{1}{k} & A\sqrt{k-1} \le x < \theta \\ 1 & x \ge \theta \end{cases}$$

Let's check that:

$$\int_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} \frac{dF_X(x)}{dx} dx + \sum_{x_k} \Pr(X = x_k) = 1$$

$$\int_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} \frac{2x}{kA^2} dx + \Pr(X = 0) + \Pr\left(X = \frac{\theta}{10}\right) + \Pr(X = \theta) = \left[\frac{x^2}{kA^2}\right]_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} + \frac{1}{4} + \frac{1}{4} + \frac{1}{k}$$

$$= \frac{(k-1) - \frac{k}{2}}{k} + \frac{1}{2} + \frac{1}{k} = \frac{\frac{k}{2} - 1}{k} + \frac{1}{2} + \frac{1}{k} = \frac{1}{2} - \frac{1}{k} + \frac{1}{2} + \frac{1}{k} = 1$$

Now we can compute the expected value of X as:

$$\mathbf{E(X)} = \int_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} x \frac{dF_X(x)}{dx} dx + \sum_{x_k} x_k \Pr(X = x_k)$$

$$= \int_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} \frac{2x^2}{kA^2} dx + 0 \cdot \Pr(X=0) + \frac{\theta}{10} \cdot \Pr\left(X = \frac{\theta}{10}\right) + \theta \cdot \Pr(X = \theta)$$

$$= \left[\frac{2x^2}{3kA^2}\right]_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} + 0 \cdot \frac{1}{4} + \frac{\theta}{10} \cdot \frac{1}{4} + \theta \cdot \frac{1}{k} = \mathbf{\frac{2A}{3k}\left((k-1)^{\frac{3}{2}} - \left(\frac{k}{2}\right)^{\frac{3}{2}}\right) + \frac{\theta}{40} + \frac{\theta}{k}}$$

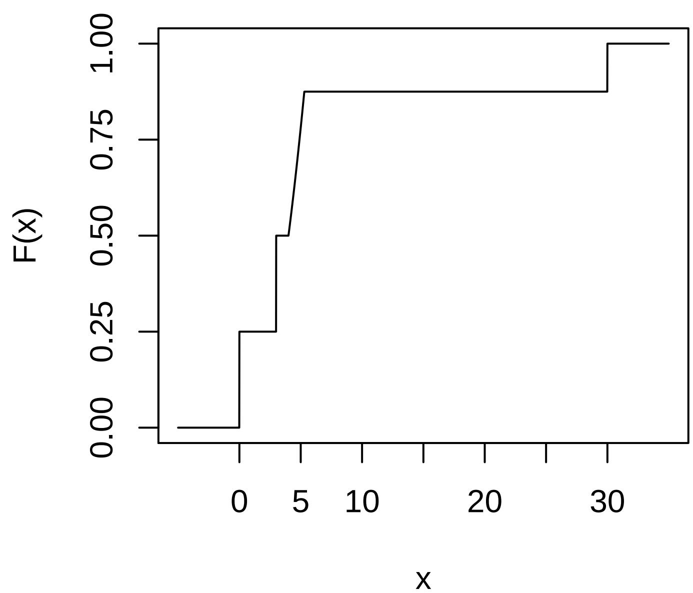

Figure 2: Approximate aspect of $F_X(x)$ (for $A = 2, k = 8, \theta = 30$)

Another way would be just applying the *law of the unconscious statistician* (https://en.wikipedia.org/wiki/Law_of_the_unconscious_statistician):

$$E[g(t)] = \int g(t) f_t(t) dt$$

$$\mathbf{E(X)} = \int_{t=0}^{k} g(t) f_t(t) dt$$

$$= \int_{t=0}^{1} \frac{g(x)}{k} dt + \int_{t=1}^{k/2} \frac{g(x)}{k} dt + \int_{t=k/2}^{3k/4} \frac{g(x)}{k} dt + \int_{t=3k/4}^{k} \frac{g(x)}{k} dt$$

$$= \frac{\theta}{k}[t]_{t=0}^{1} + \frac{A}{k}\left[\left(-\frac{2}{3}\right)(k-t)^{\frac{3}{2}}\right]_{t=1}^{k/2} + \frac{\theta/10}{k}[t]_{t=k/2}^{3k/4} + \frac{0}{k}[t]_{t=3k/4}^{k}$$

$$= \frac{\theta}{k} + \frac{2A}{3k}\left((k-1)^{\frac{3}{2}} - \left(\frac{k}{2}\right)^{\frac{3}{2}}\right) + \frac{\theta}{10k}\frac{k}{4} + 0 = \mathbf{\frac{2A}{3k}\left((k-1)^{\frac{3}{2}} - \left(\frac{k}{2}\right)^{\frac{3}{2}}\right) + \frac{\theta}{40} + \frac{\theta}{k}}$$

The last approach is based on http://homepage.stat.uiowa.edu/~nshyamal/22S175/DI.pdf. A *mixed distribution* can be decomposed into a continuous distribution and a discrete one:

$$\mathbf{F(x) = F_C(x) + F_D(x)}$$

each with weights p_C and p_D, respectively.

p_D is the sum of the heights of the jumps:

$F(0) - F(0^-) = \frac{1}{4}$

$F\left(\frac{\theta}{10}\right) - F\left(\frac{\theta^-}{10}\right) = \frac{1}{4}$

$F(\theta) - F(\theta^-) = \frac{1}{k}$

$$\Rightarrow p_D = \frac{1}{2} + \frac{1}{k} = \frac{k+2}{2k} \Rightarrow \mathbf{p_C = 1 - p_D} = \frac{k-2}{2k}$$

$$\Pr(X_D = x) = \begin{cases} \frac{\frac{1}{4}}{\frac{k+2}{2k}} = \frac{1}{4}\frac{2k}{k+2} = \frac{k}{2(k+2)} & x = 0 \\ \frac{\frac{1}{4}}{\frac{k+2}{2k}} = \frac{1}{4}\frac{2k}{k+2} = \frac{k}{2(k+2)} & x = \frac{\theta}{10} \\ \frac{\frac{1}{k}}{\frac{k+2}{2k}} = \frac{1}{k}\frac{2k}{k+2} = \frac{2}{k+2} & x = \theta \end{cases}$$

$$\mathbf{F_C(x) = \frac{F(x) + p_D \cdot F_D(x)}{p_C}}$$

$$\Rightarrow F_C(X_C = x) = \begin{cases} 0 & x < A\sqrt{\frac{k}{2}} \\ \frac{2x^2 - kA^2}{A^2(k-2)} & A\sqrt{\frac{k}{2}} \le x < A\sqrt{k-1} \\ 1 & x \ge A\sqrt{k-1} \end{cases}$$

$$\Rightarrow f_C(X_C = x) = \begin{cases} 0 & x < A\sqrt{\frac{k}{2}} \\ \frac{4x}{A^2(k-2)} & A\sqrt{\frac{k}{2}} \le x < A\sqrt{k-1} \\ 0 & x \ge A\sqrt{k-1} \end{cases}$$

$$E(X_D) = \sum_{x_k} x_k \Pr(X_D = x_k) = 0 \cdot \frac{k}{2(k+2)} + \frac{\theta}{10} \cdot \frac{k}{2(k+2)} + \theta \cdot \frac{2}{k+2}$$

$$= \frac{k\theta}{20(k+2)} + \frac{2\theta}{k+2} = \frac{(k+40)\theta}{20(k+2)}$$

$$E(X_C) = \int_{x=-\infty}^{\infty} x f_C(x) dx = \int_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} \frac{4x^2}{A^2(k-2)} dx = \left[\frac{4x^3}{3A^2(k-2)}\right]_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} = \frac{4A}{3(k-2)}\left((k-1)^{\frac{3}{2}} - \left(\frac{k}{2}\right)^{\frac{3}{2}}\right)$$

$$\mathbf{E(X) = p_D \cdot E(X_D) + p_C \cdot E(X_C)}$$

$$\mathbf{E(X)} = \frac{k+2}{2k} \cdot \frac{(k+40)\theta}{20(k+2)} + \frac{k-2}{2k} \cdot \frac{4A}{3(k-2)}\left((k-1)^{\frac{3}{2}} - \left(\frac{k}{2}\right)^{\frac{3}{2}}\right)$$

$$= \mathbf{\frac{2A}{3k}\left((k-1)^{\frac{3}{2}} - \left(\frac{k}{2}\right)^{\frac{3}{2}}\right)} + \frac{\theta}{\mathbf{40}} + \frac{\theta}{\mathbf{k}}$$

3. **Compute the variance of the payout from the contract.**

The 3rd link in (2) mentions the expression for the variance of a mixed random variable:

$$\mathbf{Var(X) = (p_D \cdot Var(X_D) + p_C \cdot Var(X_C)) + p_D \cdot p_C \cdot (E(X_D) - E(X_C))^2}$$

We already computed $p_D, p_C, E(X_D)$, and $E(X_C)$. We'll compute now the variances of both components of X, omitting the final expression of $Var(X)$ (which is really complex and with a lot of terms):

$$Var(X_D) = \sum_{x_k} (x_k - E(X_D))^2 \Pr(x_k)$$

$$= \left(0 - \frac{(k+40)\theta}{20(k+2)}\right)^2 \frac{k}{2(k+2)} + \left(\frac{\theta}{10} - \frac{(k+40)\theta}{20(k+2)}\right)^2 \frac{k}{2(k+2)} + \left(\theta - \frac{(k+40)\theta}{20(k+2)}\right)^2 \frac{2}{k+2}$$

$$= \frac{(k+40)^2\theta^2}{400(k+2)^2} \cdot \frac{k}{2(k+2)} + \left(\frac{2\theta(k+2) - \theta(k+40)}{20(k+2)}\right)^2 \cdot \frac{k}{2(k+2)} + \left(\frac{20\theta(k+2) - \theta(k+40)}{20(k+2)}\right)^2 \cdot \frac{2}{k+2}$$

$$= \frac{k\theta^2(k+40)^2}{800(k+2)} + \frac{4k\theta^2(k+2)^2 + k\theta^2(k+40)^2 - 4k\theta^2(k+2)(k+40)}{800(k+2)^2}$$

$$+ \frac{1600\theta^2(k+2)r + 4\theta^2(k+40)^2 - 80\theta^2(k+2)(k+40)}{800(k+2)^3}$$

$$= \frac{2\theta^2(k+2)(k+40)^2 + 4\theta^2(k+400)(k+2)^2 - 4\theta^2(k+20)(k+2)(k+40)}{800(k+2)^3}$$

$$= \frac{\theta^2(k+40)((k+40) - 2(k+20))}{400(k+2)^2} + \frac{\theta^2(k+400)}{200(k+2)}$$

$$= \frac{\theta^2(k+400)}{200(k+2)} - \frac{k\theta^2(k+40)}{400(k+2)^2} = \frac{\theta^2((k+400)(k+2) - k(k+40))}{400(k+2)^2}$$

$$= \frac{\theta^2(k^2 + 2k + 400k + 800 - k^2 - 40k)}{400(k+2)^2} = \frac{\theta^2(362k + 800)}{400(k+2)^2} = \frac{\theta^2(181k + 400)}{200(k+2)^2}$$

$$Var(X_C) = \int_{x=-\infty}^{\infty} x^2 f_C(x)dx - (E(X_C))^2$$

$$= \int_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} \frac{4x^3}{A^2(k-2)}dx - (E(X_C))^2 = \left[\frac{x^4}{A^2(k-2)}\right]_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} - (E(X_C))^2$$

$$= \frac{A^4(k-1)^2 - A^4\frac{k^2}{4}}{A^2(k-2)} - (E(X_C))^2 = \frac{A^2(k^2+1-2k-\frac{k^2}{4}}{k-2} - (E(X_C))^2$$

$$= \frac{A^2(3k^2+4-8k)}{4(k-2)} - \frac{16A^2}{9(k-2)^2}\left((k-1)^3 + \frac{k}{2} - 2\left(\frac{k}{2}\right)^{\frac{3}{2}}(k-1)^{\frac{3}{2}}\right)$$

Question 4

Continuous random variables X and Y have a joint distribution with probability density function $f(x,y) = 2e^{-x}e^{-2y}$ for $0 < x < \infty, 0 < y < \infty$ and 0 otherwise.

1. **Compute $\Pr(X > a, Y < b)$, where a, b are positive constants and $a < b$.**

$$\Pr(X > a, Y < b) = \int_{x=a}^{\infty}\int_{y=0}^{b} f(x,y)dxdy$$

$$= \int_{x=a}^{\infty}\int_{y=0}^{b} 2e^{-x}e^{-2y}dxdy = 2\left(\int_{x=a}^{\infty} e^{-x}dx\right)\left(\int_{y=0}^{b} e^{-2y}dy\right)$$

$$= 2\left[-e^{-x}\right]_{x=a}^{\infty}\left[-\frac{e^{-2y}}{2}\right]_{y=0}^{b} = \left[e^{-x}\right]_{x=a}^{\infty}\left[e^{-2y}\right]_{y=0}^{b} = \left(0 - e^{-a}\right)\left(e^{-2b} - 1\right)$$

$$= \mathbf{e^{-a}\left(1 - e^{-2b}\right) = e^{-a} - e^{-a-2b}}$$

2. **Compute $\Pr(X < Y)$.**

$$\Pr(X < Y) = \int_{x=0}^{y}\int_{y=0}^{\infty} f(x,y)dxdy$$

$$= \int_{x=0}^{y}\int_{y=0}^{\infty} 2e^{-x}e^{-2y}dxdy = 2\int_{y=0}^{\infty}\left(\int_{x=0}^{y} e^{-x}dx\right)e^{-2y}dy$$

$$= 2\int_{y=o}^{\infty}\left[-e^{-x}\right]_{x=0}^{y} e^{-2y}dy = 2\int_{y=0}^{\infty}\left(1 - e^{-y}\right)e^{-2y}dy = 2\int_{y=0}^{\infty}\left(e^{-2y} - e^{-3y}\right)dy$$

$$= 2\left[-\frac{e^{-2y}}{2} + \frac{e^{-3y}}{3}\right]_{y=0}^{\infty} = 2\left[0 - \left(-\frac{1}{2} + \frac{1}{3}\right)\right] = 2\left(\frac{1}{2} - \frac{1}{3}\right) = 1 - \frac{2}{3}$$

$$= \mathbf{\frac{1}{3}}$$

3. **Compute $\Pr(X < a)$.**

$$\Pr(X < a) = \int_{x=0}^{a}\int_{y=0}^{\infty} f(x,y)dxdy$$
$$= \int_{x=0}^{a}\int_{y=0}^{\infty} 2e^{-x}e^{-2y}dxdy = 2\left(\int_{x=0}^{a} e^{-x}dx\right)\left(\int_{y=0}^{\infty} e^{-2y}dy\right)$$
$$= 2\left[-e^{-x}\right]_{x=0}^{a}\left[-\frac{e^{-2y}}{2}\right]_{y=0}^{\infty} = \left[e^{-x}\right]_{x=0}^{a}\left[e^{-2y}\right]_{y=0}^{\infty} = \left(e^{-a}-1\right)(0-1)$$
$$= \mathbf{1 - e^{-a}}$$

Question 5

Let X be a random variable and x be a real number. A linear function of the squared deviation from x is another random variable, $Y = a + b(X-x)^2$, where a and b are some positive constant.

1. **Find the value of x that minimizes $E(Y)$. Show that your result is really the minimum.**

The *Law of the unconscious statistician* states that:

$$E\left[g(X)\right] = \int g(x)f_X(x)dx$$

So, if we call $\mu = E(X)$ and $\sigma^2 = Var(X)$ (and knowing that $Var(X) = E\left(X^2\right) - \left(E(X)\right)^2 = E\left(X^2\right) - \mu^2$):

$$E(Y) = \int \left(a + b(X-x)^2\right) f(X)dX$$
$$= \int \left(a + b(X^2 + x^2 - 2xX\right) f(X)dX$$
$$= (a + bx^2)\int f(X)dX - 2bx\int Xf(X)dX + b\int X^2 f(X)dX$$
$$= (a+bx^2)1 - 2bx\mu + b(\sigma^2 - \mu^2)$$
$$= bx^2 - 2b\mu x + \left(a + b(\sigma^2+\mu^2)\right)$$

Hence:

$$\frac{dE(Y)}{dx} = 2bx - 2b\mu = 2b(x-\mu) = 2b\left(x - E(X)\right)$$

And consequently:

$$\frac{dE(Y)}{dx} = 0 \Rightarrow \mathbf{x = E(X)}$$

2. **Find the value of $E(Y)$ for the choice of x you found in (1)?**

We just have to substitute in the last expression of $E(Y)$

$$\mathbf{E(Y)} = \int \left(a + b(X-\mu)^2\right) f(X)dX$$
$$= b\mu^2 - 2b\mu^2 + \left(a + b(\sigma^2+\mu^2)\right) = -b\mu^2 + a + b\sigma^2 + b\mu^2$$
$$= \mathbf{a + b\sigma^2}$$

3. **Suppose $Y = ax + b(X - x)^2$. Find the values of x that minimizes $E(Y)$. Show that your result is really the minimum.**

$$\begin{aligned}
E(Y) &= \int \left(ax + b(X-x)^2\right) f(X)dX \\
&= \int \left(ax + b(X^2 + x^2 - 2xX\right) f(X)dX \\
&= (ax + bx^2) \int f(X)dX - 2bx \int Xf(X)dX + b \int X^2 f(X)dX \\
&= (ax + bx^2)1 - 2bx\mu + b(\sigma^2 - \mu^2) \\
&= bx^2 + (a - 2b\mu)x + b(\sigma^2 + \mu^2)
\end{aligned}$$

Hence:

$$\frac{dE(Y)}{dx} = 2bx + (a - 2b\mu) = 2b(x - \mu) + a = 2b\left(x - E(X)\right) + a$$

And consequently:

$$\frac{dE(Y)}{dx} = 0 \Rightarrow \mathbf{x = E(X) - \frac{a}{2b}}$$

Question 6

Suppose X and Y are independent continuous random variables, where both of which are uniformly distributed between 0 and 1. Let random variable $Z = X + Y$.

1. **Choose a value of z between 0 and 2, and draw a graph depicting the region of the $X - Y$ plane for which Z is less than z.**

First let's plot the thre variables:

As shown above (especially in the contour plot on the left), $0 \leq Z \leq 1$, and for a given value of z (let's use $z = 0.8$) the region of the $X - Y$ plane for which $Z < z$ will be a triangle with vertices $(0, 0), (z, 0)$, and $(0, z)$.

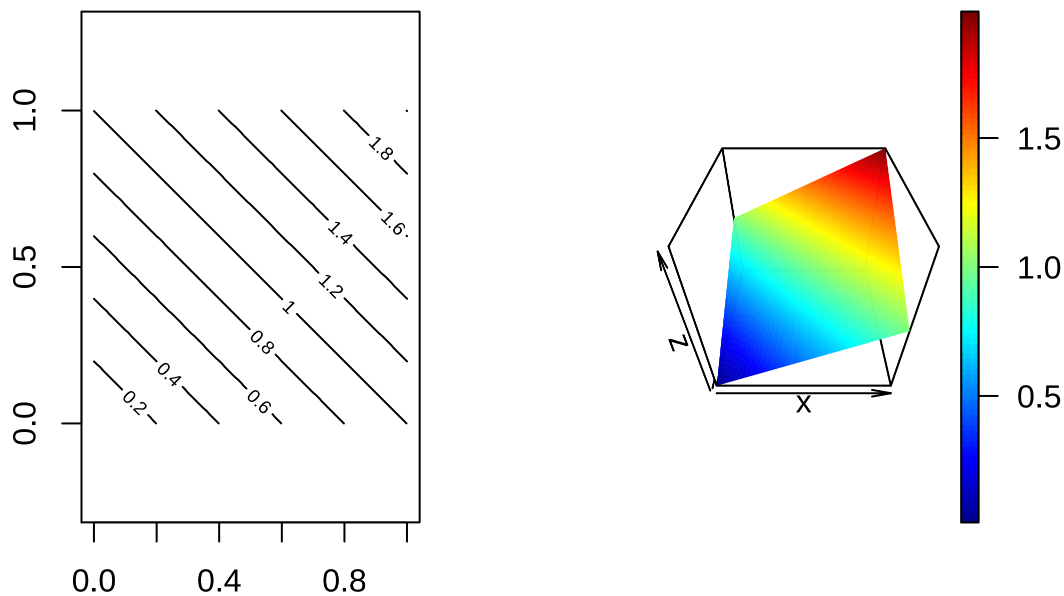

Figure 3: Contour plot and 3D plot of X, Y, Z

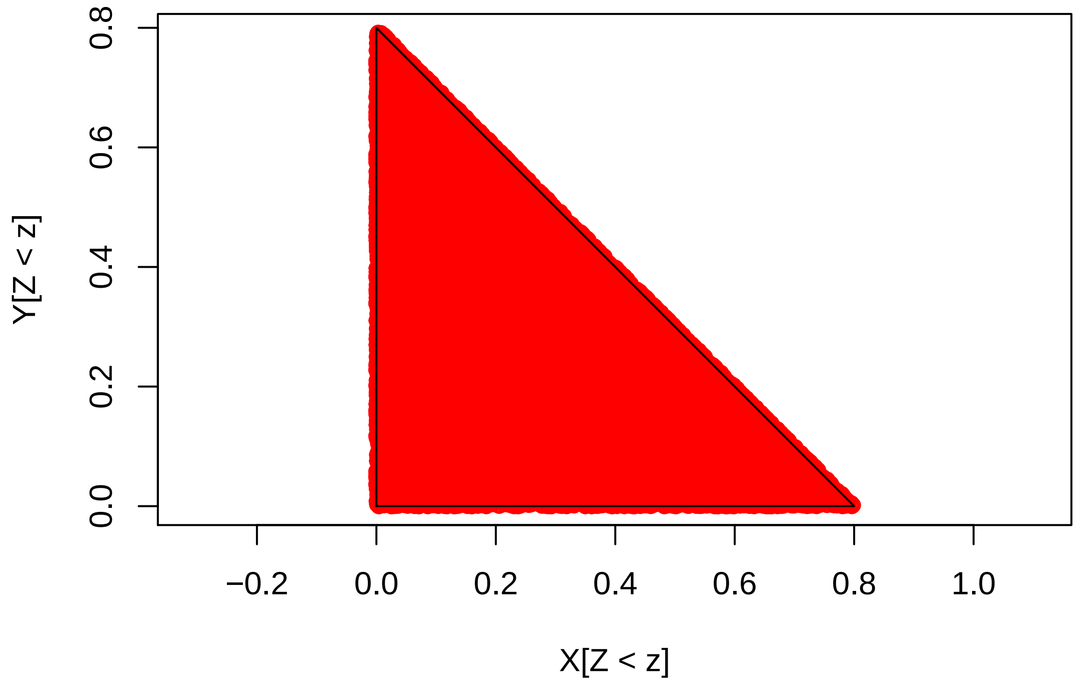

Figure 4: Region of the $X - Y$ plane for which $Z < z$ where $z = 0.8$

2. **Derive the probability density function,** $f(z)$**.**

Note that

$$f_X(x) = f_Y(y) = \begin{cases} 1 & 0 \leq x \leq \ 1 \\ 0 & Otherwise \end{cases}$$

Because X and Y are independent, the probability density function of the sum of X and Y is equal to the convolution of $f_X(x)$ and $f_Y(y)$ described by

$$f_Z(z) = \int_0^1 f_X(z-y)f_Y(y)dy$$

If $0 \leq z \leq 1$

$$f_Z(z) = \int_0^z dy = z$$

And if $1 < z \leq 2$

$$f_Z(z) = \int_{z-1}^1 dy = 2 - z$$

This give us

$$f_Z(z) = \begin{cases} z & 0 \leq z \leq 1 \\ 2-z & 1 < z \leq 2 \\ 0 & Otherwise \end{cases}$$

ALT

There is a theorem that states:

> Let X and Y be two independent random variables with density functions $f_X(x)$ and $f_Y(y)$. Then the sum $Z = X + Y$ is a random variable with density function $f_Z(z)$, where f_Z is the convolution of f_X and f_Y.

$$(f * g)(z) = \int_{x=-\infty}^{+\infty} f_Y(z-x) f_X(x) dx = \int_{y=-\infty}^{+\infty} f_X(z-y) f_Y(y) dy$$

In our case:

$$f_X(x) = f_Y(y) = \begin{cases} 1 & 0 \leq x \leq \ 1 \\ 0 & Otherwise \end{cases}$$

So:

$f_Z(z) = \int_{x=0}^{1} f_Y(z-x) dx$

$f_Y(z-x) \neq 0 \Leftrightarrow 0 \leq z-x \leq 1 \Leftrightarrow z-1 \leq x \leq z$

So for $0 \leq z \leq 1$

$f_Z(z) = \int_{x=0}^{z} dx = z$

While for $1 \leq z \leq 2$

$f_Z(z) = \int_{x=z-1}^{1} dx = 1 - (z-1) = 2 - z$

In summary:

$$\mathbf{f_Z(z)} = \begin{cases} z & 0 \leq z \leq 1 \\ 2-z & 1 < z \leq 2 \\ 0 & Otherwise \end{cases}$$

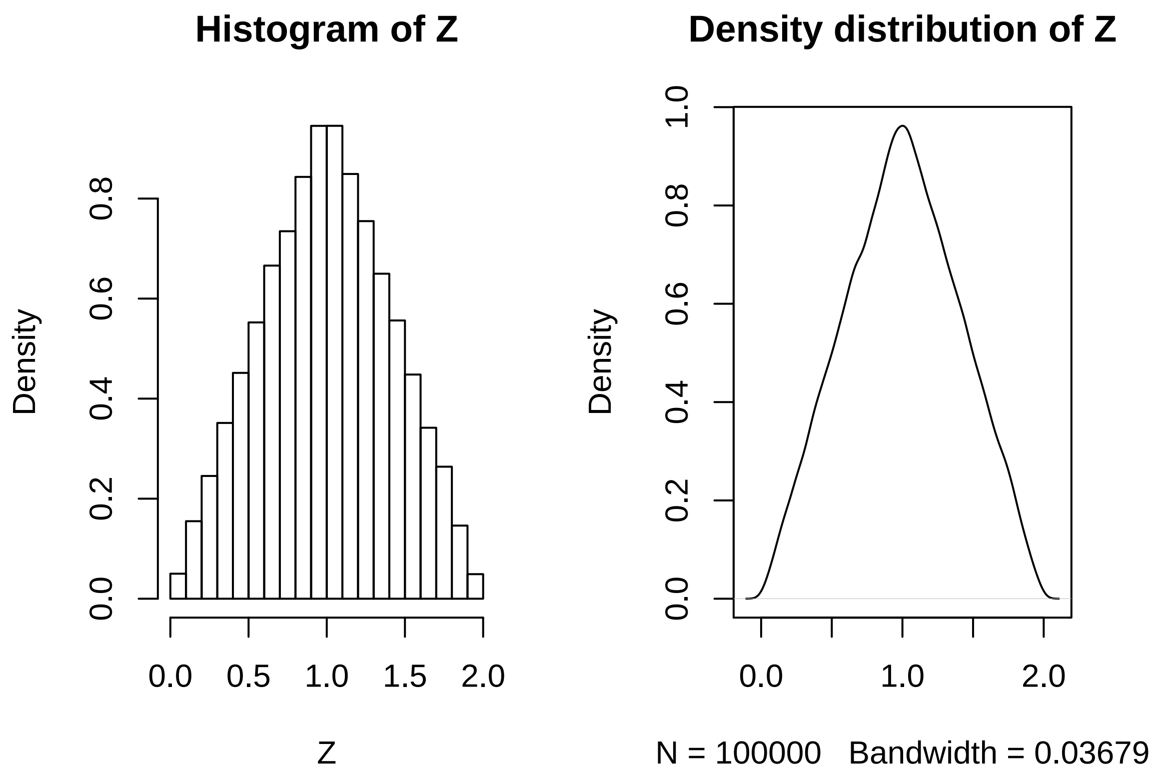

Figure 5: Histogram and (approximate) density distribution plot of Z

Question 7

In a casino, you pay the following game. A pair of fair, ordinary 6-faced dice are rolled. If the sum of the dice is 2, 3, or 12, the house wins. If it is 7 or 11, you win. If it is any other number x, the house rolls the dice again until the sum is either 7 or x. If it is 7, the house wins. If it is x, you win. A game ends if one of the two players wins. Let Y be the number of rolls needed until the game ends.

Table 1: Dice Game Rolls and Outcomes

Sum	Rolls	# of Rolls	Win Outcome
2	(1,1)	1	H
3	(1,2)*	2	H
4	(2,2),(3,1)*	3	
5	(4,1)*,(3,2)*	4	
6	(5,1)*,(5,2)*,(3,3)	5	
7	(6,1)*,(5,2)*,(4,3)*	6	P
8	(6,2)*,(5,3)*,(4,4)	5	
9	(6,3)*,(5,4)*	4	
10	(6,4)*,(5,5)	3	
11	(6,5)*	2	P
12	(6,6)	1	H

1. **Is the expected number of rolls given that you win more than, equal to, or less than the expected number of rolls given that house wins (in a game)? The steps to arrive at your answer numerically need to be clearly shown.**

Y: number of rolls, H: the house wins, P: the player wins, x: sum of dice rolled that does not equal 2, 3, 12, 7, 11 $E[Y|H] = \Sigma y_i \cdot f_{Y|H}(y_i|H) = \frac{4}{36} + \Sigma_{i=2}^{n} y_i \cdot \frac{24}{36} \cdot \frac{6}{36} \cdot p(7^c x^c)^{i-1}$

$E[Y|P] = \Sigma y_i \cdot f_{Y|H}(y_i|H) = \frac{8}{36} + \Sigma_{i=2}^{n} y_i \cdot \frac{24}{36} \cdot p(x)^i \cdot p(7^c x^c)^{i-1}$

where:

$$p(x) = \begin{cases} \frac{3}{36}, & \text{if } x = 4, 10 \\ \frac{4}{36}, & \text{if } x = 5, 9 \\ \frac{5}{36}, & \text{if } x = 6, 8 \end{cases}$$

$$p(7^c x^c) = \begin{cases} \frac{27}{36}, & \text{if } x = 4, 10 \\ \frac{26}{36}, & \text{if } x = 5, 9 \\ \frac{25}{36}, & \text{if } x = 6, 8 \end{cases}$$

To determine if $E[Y|P]$ is more than, less than, or equal to $E[Y|H]$, we can take the difference and see if it is $>$, $<$, or $=$ to zero.

$E[Y|P] - E[Y|H] = \frac{8}{36} + \Sigma_{i=2}^{n} y_i \cdot \frac{24}{36} \cdot p(x)^i \cdot p(7^c x^c)^{i-1} - \frac{4}{36} + \Sigma_{i=2}^{n} y_i \cdot \frac{24}{36} \cdot \frac{6}{36} \cdot p(7^c x^c)^{i-1}$

$= \frac{4}{36} - \frac{24}{36} \Sigma_{i=2}^{n} y_i \cdot p(7^c x^c)^{i-1} (p(x)^i - \frac{6}{36})$

Based on the above equation it will be based on the portion of the equation $(p(x)^i - \frac{6}{36})$ and for all $i = 2, ...n$ it will be less than zero. Therefore, $\mathbf{E[Y|P] < E[Y|H]}$

2. **Suppose it takes \$20 to play, and the payoff is \$100, \$80, \$60, \$40, \$0 if you win in the 1st, 2nd, 3rd, 4th, 5th round, respectively. That is, if you win in the 1st round, you are paid \$100 (so your net profit is \$80), if you win in the 2nd round, you are paid \$80, etc. Derive the expected payoff function of a game.**

M: Monetary payoff $E[M] = E[Y = 1](100) + E[Y = 2](80) + E[Y = 3](60) + E[Y = 4](40) + E[Y = 5](0)$
$= \frac{8}{36} \cdot 100 + \frac{24}{36} \cdot p(x) \cdot 80 + \frac{24}{36} \cdot p(x) \cdot p(7^c x^c) \cdot 60 + \frac{24}{36} \cdot p(x) \cdot p(7^c x^c)^2 \cdot 40 - 20$
where:

$$p(x) = \begin{cases} \frac{3}{36}, & \text{if } x = 4, 10 \\ \frac{4}{36}, & \text{if } x = 5, 9 \\ \frac{5}{36}, & \text{if } x = 6, 8 \end{cases}$$

$$p(7^c x^c) = \begin{cases} \frac{27}{36}, & \text{if } x = 4, 10 \\ \frac{26}{36}, & \text{if } x = 5, 9 \\ \frac{25}{36}, & \text{if } x = 6, 8 \end{cases}$$

$E[M|x = 2, 3, 12] = 0$

$E[M|x = 7, 11] = 100 - 20 = 80$

$E[M|x = 4, 10] = \frac{8}{36} \cdot 100 + \frac{24}{36} \cdot \frac{3}{36} \cdot 80 + \frac{24}{36} \cdot \frac{3}{36} \cdot \frac{27}{36} \cdot 60 + \frac{24}{36} \cdot \frac{3}{36} \cdot \frac{27}{36}^2 \cdot 40 - 20 = 10.4167$

$E[M|x = 5, 9] = \frac{8}{36} \cdot 100 + \frac{24}{36} \cdot \frac{4}{36} \cdot 80 + \frac{24}{36} \cdot \frac{4}{36} \cdot \frac{26}{36} \cdot 60 + \frac{24}{36} \cdot \frac{4}{36} \cdot \frac{26}{36}^2 \cdot 40 - 20 = 12.9035$

$E[M|x = 6, 8] = \frac{8}{36} \cdot 100 + \frac{24}{36} \cdot \frac{5}{36} \cdot 80 + \frac{24}{36} \cdot \frac{5}{36} \cdot \frac{25}{36} \cdot 60 + \frac{24}{36} \cdot \frac{5}{36} \cdot \frac{25}{36}^2 \cdot 40 - 20 = 15.2738$

$E[M] = E[M|x = 4, 10] \cdot p(x = 4, 10) + E[M|x = 5, 9] \cdot (x = 5, 9) + E[M|x = 6, 8] \cdot p(x = 6, 8) + E[M|x = 2, 3, 12] \cdot p(x = 2, 3, 12) + E[M|x = 7, 11] \cdot p(x = 7, 11)$

$= 10.4167\frac{6}{36} + 12.9035\frac{8}{36} + 15.2738\frac{10}{36} + 0\frac{4}{36} + 80\frac{8}{36} =$ **26.624**

Part III: Statistical Estimation and Statistical Inference

In classical statistics, parameters are unknown constants whereas estimators are functions of samples and are random variables. The questions in this section are designed to clarify the relationship between parameters and estimators, and explore the properties that different estimators may have.

Question 8

Let $Y1, \ldots, Yn$ be n random variables, such that any two of them are uncorrelated, and all share the same mean μ and variance σ^2. Let Y be the average Y_i, which is also a random variable.

Define the class of linear estimators of μ by

$$W = \sum_{i=1}^{n} a_i Y_i$$

where the a_i are constants.

1. **What restriction on the a_i is needed for W to be an unbiased estimator of μ?**

$$W = a_1 E(Y_1) + a_2 E(Y_2) + ... + a_n E(Y_n)$$

And $Y_i = \mu \; \forall$ i $\in$ n.

Therefore

$$W = (a_1 + a_2 + ... + a_n)\mu$$

For W to be an unbiased estimator of μ, $a_1 + a_2 + ...a_3 = 1$

2. **Find $Var(W)$.**

$$Var(W) = a_1^2 Var(Y_1) + a_2^2 Var(Y_2) + ... + a_n^2 Var(Y_n)$$

and Var(Y_i) $= \sigma^2 \forall \in$ n. Therefore

$$Var(W) = (a_1^2 + a_2^2 + ... + a_n^2)\sigma^2$$

3. **Given a set of numbes $a_1, \ldots, a_n$, the following inequality holds:**

$$\frac{1}{n}\left(\sum_{i=1}^{n} a_i\right)^2 \leq \sum_{i=1}^{n} a_i^2$$

Use this inequality, along with the previous parts of this question, tho show that $Var(W) \geq Var(\bar{Y})$ whenever W is unbiased. We say that $\bar{Y}$ is the best linear unbiased estimator (BLUE).

When W is unbiased (ie $(a_1 + a_2 + ... + a_n) = 1$), the expression becomes

$$\frac{1}{n} \leq a_1^2 + a_2^2 + ... + a_n^2$$

Multiplying by σ we have

$$Var(\bar{Y}) \leq \frac{\sigma}{n} = \sigma^2(a_1^2 + a_2^2 + ... + a_n^2) = Var(W)$$

or

$$Var(W) \geq Var(\bar{(Y)})$$

Question 9

Let $\bar{Y}$ denote the average of n independent draws from a population distribution with mean μ and variance σ^2. Consider two alternative estimators of μ: $W_1 = \frac{n-1}{n}\bar{Y}$ and $W_2 = k\bar{Y}$, where $0 < k < 1$.

1. **Compute the biases of both W_1 and W_2. Which estimator is consistent?**

$W_1 = \frac{n-1}{n}\bar{Y}; W_2 = k\bar{Y}, where 0 < k < 1$

$Bias(W) = E[W] - \theta$

$Bias(W_1) = E[\frac{n-1}{n}\bar{Y}] - \bar{Y} = \frac{n-1}{n}E[\bar{Y}] - \bar{Y} = \mathbf{\bar{Y}(\frac{n-1}{n} - 1)}$

$Bias(W_2) = E[k\bar{Y}] - \bar{Y} = kE[\bar{Y}] - \bar{Y} = \mathbf{\bar{Y}(k - 1)}$

The W_2 estimator is consistent because it does not depend on n.

2. **Compute $Var(W_1)$ and $Var(W_2)$. Which estimator has lower variance?**

$Var(W) = E[(x - \mu)^2]$

$Var(W_1) = Var(\Sigma\frac{n-1}{n}\bar{Y} = Var(\frac{n-1}{n})Var(\Sigma\bar{Y}) = \Sigma Var(\frac{n-1}{n})\Sigma Var(\bar{Y} = n\sigma^2\Sigma Var(\frac{n-1}{n}) = n\sigma^2\frac{(n-1)^2}{n^2}$ $= \mathbf{\frac{\sigma^2(n-1)^2}{n}}$

$Var(W_2) = Var(\Sigma k\bar{Y} = Var(k\Sigma\bar{Y}) = kVar(\Sigma\bar{Y}) = k\Sigma Var\bar{Y} = k\Sigma\sigma^2 \quad = \mathbf{kn\sigma^2}$

W_2 has a lower variance because $\frac{(n-1)^2}{n} > kn$ for $0 < k < 1$.

Question 10

Given a random sample $Y_1, Y_2, \ldots, Y_n$ from some distribution $F(\cdot)$ with mean μ and variance σ^2, where both μ and σ^2 are unknown parameters.

Let $\bar{Y}$ be the average of the sample. Consider the following estimator for σ^2:

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(Y - \bar{Y}\right)^2$$

1. **Show that $E(\bar{Y}) = E(Y_i)\ \forall i \in 1, 2, \ldots, n$**

$$E(\bar{Y}) = E\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right) = \frac{1}{n}\cdot E\left(\sum_{i=1}^{n}Y_i\right) = \frac{1}{n}\cdot\sum_{i=1}^{n}E(Y_i) = \frac{1}{n}\cdot\sum_{i=1}^{n}\mu = \frac{1}{n}\cdot(n\mu) = \mu$$

Since $Y_1, Y_2, \ldots, Y_n$ are iid random variables, Y_i for all $i \in 1, 2, \ldots, n$ have the same mean μ. In other words, $E(Y_i)\ \forall i \in 1, 2, \ldots, n = \mu$. Therefore, $E(\bar{Y}) = E(Y_i)\ \forall i \in 1, 2, \ldots, n$.

2. **Show that $Var(\bar{Y}) = \frac{1}{n}Var(Y_i)\ \forall i \in 1, 2, \ldots, n$**

$$Var(\bar{Y}) = Var\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right) = \frac{1}{n^2}\cdot Var\left(\sum_{i=1}^{n}Y_i\right) = \frac{1}{n^2}\cdot\left(\sum_{i=1}^{n}Var(Y_i)\right) = \frac{1}{n^2}\cdot\left(\sum_{i=1}^{n}\sigma^2\right) = \frac{1}{n^2}\cdot(n\sigma^2) = \frac{\sigma^2}{n}$$

Since $Y_1, Y_2, \ldots, Y_n$ are iid random variables, Y_i for all $i \in 1, 2, \ldots, n$ have the same variance σ^2. In other words, $Var(Y_i)\ \forall i \in 1, 2, \ldots, n = \sigma^2$. Therefore, $Var(\bar{Y}) = \frac{1}{n}Var(Y_i)\ \forall i \in 1, 2, \ldots, n$.

3. **Compute the expectation of $\widehat{\sigma^2}$ in terms of n and σ^2. In your derivation, make sure make use of the *i.i.d.* property and identify where you use it.**

$$E(\widehat{\sigma^2}) = E\left(\frac{1}{n}\cdot\sum_{i=1}^{n}(Y-\bar{Y})^2\right) = \frac{1}{n}\cdot E\left(\sum_{i=1}^{n}(Y-\bar{Y})^2\right) = \frac{1}{n}\cdot E\left(\sum_{i=1}^{n}(Y_i^2 - 2Y_i^2\bar{Y} + \bar{Y}^2)\right)$$

$$= \frac{1}{n}\cdot\left[E\left(\sum_{i=1}^{n} Y_i^2\right) - E\left(\sum_{i=1}^{n} 2\bar{Y}(Y_i)\right) + E\left(\sum_{i=1}^{n} \bar{Y}^2\right)\right]$$

Since $Var(X) = E\left[(X - E(X)^2)\right] = E(X^2) - E(X)^2$, $E(X^2)$ can be represented as $Var(X) + E(X)^2$. Given a random sample $Y_1, Y_2, \ldots, Y_n$ from some distribution $F(\cdot)$ with mean μ and variance σ^2,

$$E\left(\sum_{i=1}^{n} Y_i^2\right) = \sum_{i=1}^{n} E(Y_i^2) = \sum_{i=1}^{n}\left(Var(Y_i) + E(Y_i)^2\right) = \sum_{i=1}^{n}(\sigma^2 + \mu^2) = n\sigma^2 + n\mu^2$$

Similarly, $E(\bar{Y}^2) = Var(\bar{Y}) + E(\bar{Y})^2 = \dfrac{\sigma^2}{n} + \mu^2$

In addition,

$$E\left(\sum_{i=1}^{n} 2\bar{Y}(Y_i)\right) = E\left(2\bar{Y}\sum_{i=1}^{n} Y_i\right) = E\left(2\bar{Y}\sum_{i=1}^{n} Y_i\right) = E(2\bar{Y}\cdot n\bar{Y}) = 2nE(\bar{Y}^2)$$

Therefore, the above $E(\widehat{\sigma^2})$ can be further derived as:

$$E(\widehat{\sigma^2}) = \frac{1}{n}\cdot\left[n\sigma^2 + n\mu^2 - 2nE(\bar{Y}^2) + nE(\bar{Y}^2)\right] = \frac{1}{n}\cdot\left[n\sigma^2 + n\mu^2 - nE(\bar{Y}^2)\right] = \frac{1}{n}\cdot\left[n\sigma^2 + n\mu^2 - n(\frac{\sigma^2}{n} + \mu^2)\right]$$

$$= \frac{1}{n}\cdot\left[n\sigma^2 - \sigma^2\right] = \mathbf{\frac{n-1}{n}}\boldsymbol{\sigma^2}$$

4. **Is this an unbiased estimator for σ^2?**

No. As shown above, since its expected value is equal to σ^2, it is not an unbiased estimator of the parameter σ^2.

5. **If not, what function of $\widehat{\sigma^2}$ produce an unbiased estimator?**

In order to produce an unbiased estimator, the above $\widehat{\sigma^2}$ need to be modified to as below:

$$s^2 = \widehat{\sigma^2}\cdot(\frac{n}{n-1}) = \frac{1}{n}\sum_{i=1}^{n}\left(Y-\bar{Y}\right)^2\cdot(\frac{n}{n-1}) = \mathbf{\frac{1}{n-1}\sum_{i=1}^{n}\left(Y-\bar{Y}\right)^2}$$

Question 11

Wooldridge's textbook: Appendix C, Question 4*i*, *ii*, *iii*.

4. **For positive random variables X and Y, suppose the expected value of Y given X is $E[Y|X] = \theta X$. The unknown parameter θ shows how the expected value of Y changes with X.**

(i) Define the random variable $Z = Y/X$. Show that $E(z) = \theta$. [Hint: Use Property CE.2 along with the law of iterated expectations, Property CE.4. In particular, first show that $E[Z|X] = \theta$ and then use CE.4.]

$E[Z|X] = E[\frac{Y}{X}|X] = E[\frac{1}{X}Y|X] = \frac{1}{X}E[Y|X] = \frac{1}{X} \cdot \theta X = \theta$

(ii) Use part (i) to prove that the estimator $W_1 = n^{-1}\Sigma_{i=1}^{n}\frac{Y_i}{X_i}$ is unbiased for θ, where $(X_i, Y_i) : i = 1, 2, ...n$ is a random sample.

$bias(W) = E[W_1] - \theta$

$E[W_1] = E[\frac{1}{n}\Sigma\frac{Y_i}{X_i}] = \frac{1}{n}\Sigma E[\frac{Y_i}{X_i}] = \frac{1}{n}\Sigma E[Z_i] = \frac{1}{n}\Sigma\theta = \frac{\theta}{n}$

$\mathbf{bias(W_1) = E[W_1]} - \theta = \frac{\theta}{\mathbf{n}} - \theta$

$$E(W_1) = E\left[\frac{1}{n}\sum_{i=1}^{n}(\frac{Y_i}{X_i})\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}(\frac{Y_i}{X_i})\right] = \frac{1}{n} \cdot n \cdot E(Z) = \theta$$

Therefore, the estimator W1 is unbiased for θ.

(iii) Explain why the estimator $W_2 = \frac{\bar{(Y)}}{\bar{X}}$, where the overbars denote sample averages, is not the same as W_1. Nevertheless, show that W_2 is also unbiased for θ.

$$E(W_2) = E\left(\frac{\bar{Y}}{\bar{X}}\right) = E\left(\frac{\frac{1}{n} \cdot \sum_{i=1}^{n} Y_i}{\frac{1}{n} \cdot \sum_{i=1}^{n} X_i}\right) = E\left(\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i}\right) = E(\sum_{i=1}^{n} Y_i) \cdot E\left(\frac{1}{\sum_{i=1}^{n} X_i}\right)$$

$$= \sum_{i=1}^{n}(E(Y_i)) \cdot \left(\frac{1}{\sum_{i=1}^{n} E(X_i)}\right) = \sum_{i=1}^{n}(E(E(Y_i|X_i))) \cdot \left(\frac{1}{\sum_{i=1}^{n} E(X_i)}\right)$$

$$= \sum_{i=1}^{n}(E(\theta \cdot X_i)) \cdot \left(\frac{1}{\sum_{i=1}^{n} E(X_i)}\right) = \theta \cdot \sum_{i=1}^{n}(E(X_i)) \cdot \left(\frac{1}{\sum_{i=1}^{n} E(X_i)}\right) = \theta$$

Therefore, W_2 is also unbiased for θ. Although $E(W_1)$ and $E(W_2)$ are all equal to θ, $W_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i}{X_i}\right)$ and $W_2 = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i}$,respectively.

Question 12

Wooldridge's textbook: Appendix C, Question 6.

6. **You are hired by the governor to study whether a tax on liquor has decreased average liquor consumption in your state. You are able to obtain, for a sample of individuals selected at random, the difference in liquor consumption (in ounce) for the years before and after the tax. For person i who is sampled randomly from the population. Y_i denotes the change in liquor consumption. Treat these as a random sample from a Normal(μ, $|sigma^2$) distribution.**

(i) The null hypothesis is that there was no change in average liquor consumption; state this formally in terms of μ

The null hypothesis is also stated as $\mu_\mathbf{0} = \mu_\mathbf{A}$.

(ii) The alternative is that there was a decline in liquor consumption; state the alternative in terms of μ.

The alternative hypothesis is also stated as $\mu_{\mathbf{A}} < \mu_{\mathbf{0}}$.

(iii) Now, suppose your sample size is $n = 900$ and you obtain the estimates $\bar{y} = -32.8$ and $s = 466.4$. Calculate the t statistic for testing H_0 against H_1; obtain the p-value for the test. (Because of the large sample size, just use the standard normal distribution tabulated in Table G.1) Do you reject H_0 at te 5% level? At the 1% level?

$t = \frac{\sqrt{n}\bar{y}}{s} = \frac{\sqrt{900}-32.8}{466.4} = -2.10977$

p(.0179) $<$ .05 The null hypothesis can be rejected at the 5% level.

p(.0179) $>$ .01 The null hypothesis cannot be rejected at the 1% level.

(iv) Would you say that the estimated fall in consumption is large in magnitude? Comment on the practical versus statistical significance of this estimate.

The estimated fall is unlikely to be large in magnitide since the p-values are hovering right around the significance values. Addtionally, as n gets larger you can achieve a statistical significance without any practical significance. In this study since n=900 and the p-values are near the significance values, there is likely very little practical significance.

(v) What has been implicity assumed in your analysis about other determinats of liquor consumption over the two-year period in order to infer causality from the tax change to liquor consumption?

This study is assuming that there are no other correlates or factors that changed over the two year period or that would contribute to a change in liquor sales.

Question 13

Wooldridge's textbook: Appendix C, Question 8. In addition, answer the following questions:

The New York Times (2/5/90) reported three-point shooting performace for the top 10 three-point shooters in the NBA. The following table summarizes these data:

Table 2: Appendix C, Question 8 Table

Player	FGA-FGM
Mark Price	429-188
Trent Tucker	833-345
Dale Ellis	1149-472
Craig Hodges	1016-396
Danny Ainge	1051-406
Byron Scott	676-260
Reggie Miller	416-159
Larry Bird	1206-455
Jon Sundvold	440-166
Brian Taylor	417-157
FGA = field goals attempted	FGM = field goals made

For a given player, the outcome of a particular shot can be modeled as a Bernoulli (zero-one) variable: if Y_i is the outcome of shot i, the $Y_i = 1$ if the shot is made, and $Y_i = 0$ if the shot is missed. Let θ denote the probability of making any particular three-point shot attempt. The natural estimator of θ is $\bar{Y} = \dfrac{FGM}{FGA}$.

(i) Estimate θ for Mark Price.

$\theta = \dfrac{188}{429}$

(ii) Find the statnd deviation of the estimator $\bar{Y}$ in terms of θ and the number of shot attempts n.

(iii) The asymptotic distribution of $\dfrac{(\bar{Y} - \theta)}{se(\bar{Y})}$ is a standard normal, where $se(\bar{Y}) = \sqrt{\bar{Y} \cdot (1 - \bar{Y})/n}$. Use this fact to test $H_0 : \theta = .5$ against $H_1 : \theta < .5$ for Mark Price. Use a 1% significance level.

1. **Define Type I error.**

The Type I error is a false positive or rejecting the null hypothesis when it is true.

2. **What is the probability of Type I error of this test?**
3. **Define Type II error.**

The Type II error is a false negative or accepting the null hypothesis that is actually false.

4. **What is the probability of Type II error when using this decision rule, assuming the "true" population proportion is $\theta^* = 0.45$.**
5. **Define the power of the test (in general terms).**

The power of the test is 1- type II error rate.

6. **Calculate the power of this test, again assuming the "true" population proportion is $\theta^* = 0.45$.**