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.

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.

$$\begin{aligned}
\Pr(\mathbf{M} \cap \mathbf{S} \cap \mathbf{A}) &= \Pr(M) + \Pr(S \cap A) - \Pr(M \cup (S \cap A)) \\
&= \Pr(M) + \Pr(S \cap A) - \Pr((M \cup S) \cap (M \cup A)) \\
&= \Pr(M) + \Pr(S \cap A) - ((\Pr(M \cup S) + \Pr(M \cup A) - \Pr((M \cup S) \cup (M \cup A)))) \\
&= \Pr(M) + \Pr(S \cap A) - (\Pr(M) - \Pr(S) + \Pr(M \cap S)) - (\Pr(M) - \Pr(A) + \Pr(M \cap A)) + \Pr(M \cup S \cup A) \\
&= \Pr(M \cap S) + \Pr(M \cap A) + \Pr(M \cap A) - \Pr(M) - \Pr(S) - \Pr(A) + \Pr(M \cup S \cup A) \\
&= \frac{22}{N} + \frac{12}{N} + \frac{9}{N} - \frac{36}{N} - \frac{28}{N} - \frac{18}{N} + \frac{48}{N} = \frac{(22+12+9) - (36-28-18) + 48}{N} = \frac{43 - 82 + 48}{N} \\
&= \mathbf{\frac{9}{N}}
\end{aligned}$$

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?**

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

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?**

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

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 $\frac{\theta}{10}$.

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

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

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

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

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

I tried 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:

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

(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:

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

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

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

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

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

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

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

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:

$$\begin{aligned}
\mathbf{E(X)} &= \int_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} x \frac{dF_X(x)}{dx} dx + \sum_{x_k} x_k \Pr(X = x_k) \\
&= \int_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} \frac{2x^2}{kA^2} dx + 0 \cdot \Pr(X=0) + \frac{\theta}{10} \cdot \Pr\left(X = \frac{\theta}{10}\right) + \theta \cdot \Pr(X=\theta)
\end{aligned}$$

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

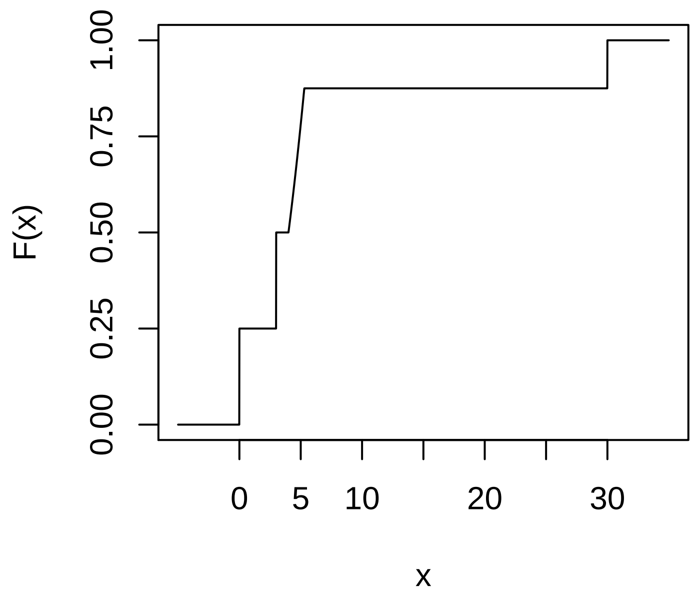

Figure 1: Approximate aspect of $F_X(x)$ (for $A = 2, k = 8, \theta = 30$)

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

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

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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):

$$Var(X_D) = \sum_{x_k}(x_k - E(X_D))^2 \Pr(x_k)$$
$$= \left(0 - \frac{(k+40)\theta}{20(k+2)}\right)^2 \frac{k}{2(k+2)} + \left(\frac{\theta}{10} - \frac{(k+40)\theta}{20(k+2)}\right)^2 \frac{k}{2(k+2)} + \left(\theta - \frac{(k+40)\theta}{20(k+2)}\right)^2 \frac{2}{k+2}$$
$$= \frac{(k+40)^2\theta^2}{400(k+2)^2}\cdot\frac{k}{2(k+2)} + \left(\frac{2\theta(k+2) - \theta(k+40)}{20(k+2)}\right)^2\cdot\frac{k}{2(k+2)} + \left(\frac{20\theta(k+2) - \theta(k+40)}{20(k+2)}\right)^2\cdot\frac{2}{k+2}$$
$$= \frac{k\theta^2(k+40)^2}{800(k+2)} + \frac{4k\theta^2(k+2)^2 + k\theta^2(k+40)^2 - 4k\theta^2(k+2)(k+40)}{800(k+2)^2}$$
$$+ \frac{1600\theta^2(k+2)r + 4\theta^2(k+40)^2 - 80\theta^2(k+2)(k+40)}{800(k+2)^3}$$
$$= \frac{2\theta^2(k+2)(k+40)^2 + 4\theta^2(k+400)(k+2)^2 - 4\theta^2(k+20)(k+2)(k+40)}{800(k+2)^3}$$
$$= \frac{\theta^2(k+40)((k+40) - 2(k+20))}{400(k+2)^2} + \frac{\theta^2(k+400)}{200(k+2)}$$
$$= \frac{\theta^2(k+400)}{200(k+2)} - \frac{k\theta^2(k+40)}{400(k+2)^2} = \frac{\theta^2((k+400)(k+2) - k(k+40))}{400(k+2)^2}$$
$$= \frac{\theta^2(k^2 + 2k + 400k + 800 - k^2 - 40k)}{400(k+2)^2} = \frac{\theta^2(362k+800)}{400(k+2)^2} = \frac{\theta^2(181k+400)}{200(k+2)^2}$$

$$Var(X_C) = \int_{x=-\infty}^{\infty} x^2 f_C(x)dx - (E(X_C))^2$$
$$= \int_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} \frac{4x^3}{A^2(k-2)}dx - (E(X_C))^2 = \left[\frac{x^4}{A^2(k-2)}\right]_{x=A\sqrt{k/2}}^{A\sqrt{k-1}} - (E(X_C))^2$$
$$= \frac{A^4(k-1)^2 - A^4\frac{k^2}{4}}{A^2(k-2)} - (E(X_C))^2 = \frac{A^2(k^2+1-2k-\frac{k^2}{4}}{k-2} - (E(X_C))^2$$
$$= \frac{A^2(3k^2+4-8k)}{4(k-2)} - \frac{16A^2}{9(k-2)^2}\left((k-1)^3 + \frac{k}{2} - 2\left(\frac{k}{2}\right)^{\frac{3}{2}}(k-1)^{\frac{3}{2}}\right)$$

To get the full expression for $Var(X)$ we made use of http://www.wolframalpha.com (the function `fractions {MASS}` could also be used):

$$\mathbf{Var(X)} = \frac{\theta^2(181k+400)}{400k(k+2)} - \frac{A^2(k(k(37k-32\sqrt{2k(k-1)}-66)+32\sqrt{2k(k-1)}+44)+8)}{72(k-2)k} + \frac{9\theta(k-2)^2(k-1)^3(k+40)-80A(k+2)}{720(k-2)(k-1)^3k^2}$$

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:

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

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

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

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)

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

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.**

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

$$= bx^2 + (a - 2b\mu)x + b(\sigma^2 + \mu^2)$$

Hence:

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

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:

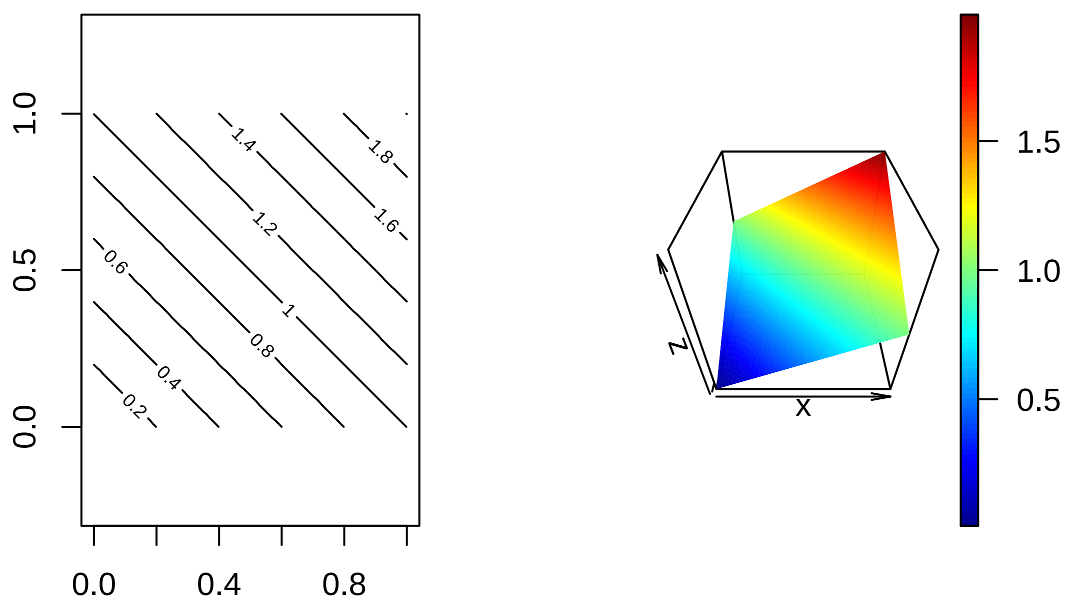

Figure 2: Contour plot and 3D plot of X, Y, Z

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 $(0, z)$.

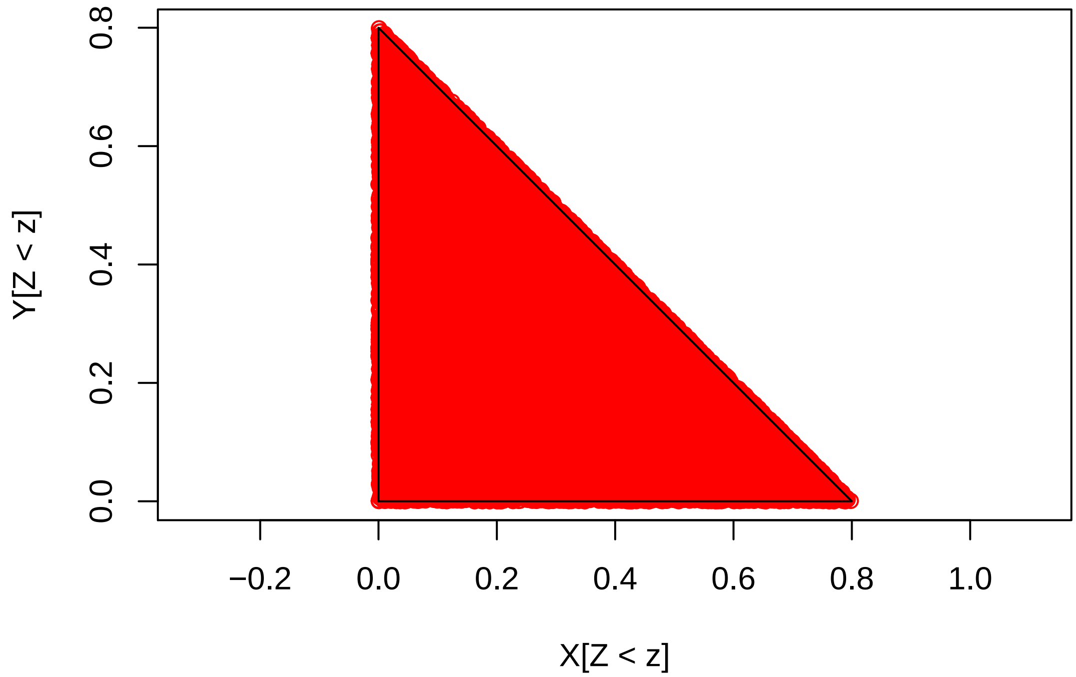

Figure 3: Region of the $X - Y$ plane for which $Z < z$ where $z = 0.8$

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

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:

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

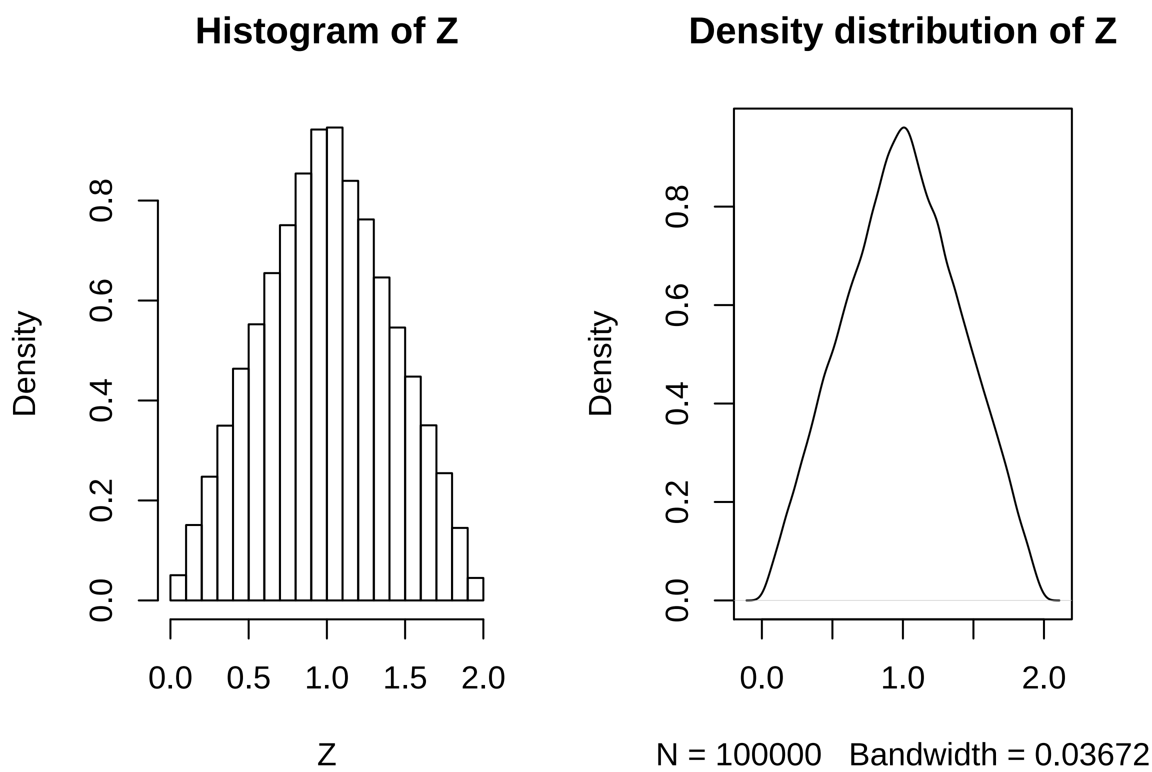

Figure 4: Histogram and (approximate) density distribution plot of Z

Question 7

In a casino, you pay the following game. A pair fair, ordinary 6-faced dices is 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.

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.**

http://www.dehn.wustl.edu/~blake/courses/WU-Ed6021-2012-Summer/handouts/Expected_Value.pdf

First we check how many possible games there are, and the probabilities for each outcome:

Table 1: Possible games

	Die 1=1	**Die 1=2**	**Die 1=3**	**Die 1=4**	**Die 1=5**	**Die 1=6**
Die 2=1	2	3	4	5	6	7
Die 2=2	3	4	5	6	7	8
Die 2=3	4	5	6	7	8	9
Die 2=4	5	6	7	8	9	10
Die 2=5	6	7	8	9	10	11
Die 2=6	7	8	9	10	11	12

From the table above we can derive the following:

$$\Pr(X = x) = \begin{cases} \dfrac{x-1}{36} & x \in \{2, \ldots, 7\} \\ \dfrac{13-x}{36} & x \in \{8, \ldots, 12\} \end{cases}$$

$$\Pr(X = 2) = \Pr(X = 12) = \frac{1}{36}$$

$$\Pr(X = 3) = \Pr(X = 11) = \frac{1}{18}$$

$$\Pr(X = 4) = \Pr(X = 10) = \frac{1}{12}$$

$$\Pr(X = 5) = \Pr(X = 9) = \frac{1}{9}$$

$$\Pr(X = 6) = \Pr(X = 8) = \frac{5}{36}$$

$$\Pr(X = 7) = \frac{1}{6}$$

Let Z be the event that I win (so $\bar{Z}$ is the event that house wins).

$$\Pr(Z, 2 \text{ on roll } 1) = 0, \Pr(\bar{Z}, 2 \text{ on roll } 1) = \frac{1}{36}$$

$\Pr(Z, 3 \text{ on roll } 1) = 0, \Pr(\bar{Z}, 3 \text{ on roll } 1) = \frac{1}{18}$

$\Pr(Z, 7 \text{ on roll } 1) = \frac{1}{6}, \Pr(\bar{Z}, 7 \text{ on roll } 1) = 0$

$\Pr(Z, 11 \text{ on roll } 1) = \frac{1}{18}, \Pr(\bar{Z}, 11 \text{ on roll } 1) = 0$

$\Pr(Z, 12 \text{ on roll } 1) = 0, \Pr(\bar{Z}, 12 \text{ on roll } 1) = \frac{1}{36}$

When we roll a $4, 5, 6, 8, 9$, or 10, we need to roll again, and the final result will depend on what we obtain in subsequent rolls (the same value, 7, or another):

$\Pr(Z, 4 \text{ on roll } 1) = \Pr(4)\Pr(4) + \Pr(4)\Pr(\text{neither 4 nor 7})\Pr(4) + \dots$

$$\Pr(\text{neither 4 nor 7}) = 1 - \frac{3+6}{36} = 1 - \frac{1}{4} = \frac{3}{4}$$

$$\Pr(Z, 4 \text{ on roll } 1) = \left(\frac{1}{12}\right)^2 \left(1 + \left(\frac{3}{4}\right) + \left(\frac{3}{4}\right)^2 + \dots\right) = \left(\frac{1}{12}\right)^2 \sum_{y=0}^{\infty} \left(\frac{3}{4}\right)^y = \left(\frac{1}{12}\right)^2 \frac{1}{1 - \frac{3}{4}} = \frac{4}{12^2} = \frac{1}{36}$$

$$\Pr(\bar{Z}, 4 \text{ on roll } 1) = \Pr(4)\Pr(7) + \Pr(4)\Pr(\text{neither 4 nor 7})\Pr(7) + \dots = \frac{1}{12}\frac{1}{6}\sum_{y=0}^{\infty}\left(\frac{3}{4}\right)^y = \frac{1}{12}\frac{1}{6}4 = \frac{1}{18}$$

The same applies to $x = 10$.

$\Pr(Z, 5 \text{ on roll } 1) = \Pr(5)\Pr(5) + \Pr(5)\Pr(\text{neither 5 nor 7})\Pr(5) + \dots$

$$\Pr(\text{neither 5 nor 7}) = 1 - \frac{4+6}{36} = \frac{26}{36} = \frac{13}{18}$$

$$\Pr(Z, 5 \text{ on roll } 1) = \left(\frac{1}{9}\right)^2 \frac{1}{1 - \frac{13}{18}} = \left(\frac{1}{9}\right)^2 \frac{18}{5} = \frac{2}{45}$$

$$\Pr(\bar{Z}, 5 \text{ on roll } 1) = \Pr(5)\Pr(7) + \Pr(5)\Pr(\text{neither 5 nor 7})\Pr(7) + \dots = \frac{1}{9}\frac{1}{6}\sum_{y=0}^{\infty}\left(\frac{13}{18}\right)^y = \frac{1}{9}\frac{1}{6}\frac{18}{5} = \frac{1}{15}$$

The same applies to $x = 9$.

$\Pr(Z, 6 \text{ on roll } 1) = \Pr(6)\Pr(6) + \Pr(6)\Pr(\text{neither 6 nor 7})\Pr(6) + \dots$

$$\Pr(\text{neither 6 nor 7}) = 1 - \frac{5+6}{36} = \frac{25}{36}$$

$$\Pr(Z, 6 \text{ on roll } 1) = \left(\frac{5}{36}\right)^2 \frac{1}{1 - \frac{25}{36}} = \left(\frac{5}{36}\right)^2 \frac{36}{11} = \frac{25}{396}$$

$$\Pr(\bar{Z}, 6 \text{ on roll } 1) = \Pr(6)\Pr(7) + \Pr(6)\Pr(\text{neither 6 nor 7})\Pr(7) + \dots = \frac{5}{36}\frac{1}{6}\sum_{y=0}^{\infty}\left(\frac{25}{36}\right)^y = \frac{5}{36}\frac{1}{6}\frac{36}{11} = \frac{5}{66}$$

The same applies to $x = 8$.

$$\Pr(Z) = \sum_{x=2}^{1} 2\Pr(Z, \text{x on roll } 1) = 0 + 0 + \frac{1}{36} + \frac{2}{45} + \frac{25}{396} + \frac{1}{6} + \frac{25}{396} + \frac{2}{45} + \frac{1}{36} + \frac{1}{18} + 0$$

$$= \frac{55 + 88 + 125 + 330 + 125 + 88 + 55 + 110}{1980} = \frac{244}{495} \approx 0.4929 \Rightarrow \Pr(\bar{Z}) = \frac{251}{495} \approx 0.5071$$

What we are asked for is:

$$E[Y|Z] = \sum_{y=1}^{\infty} y \Pr(y|Z)$$

In order to compute $\Pr(y|Z) = \dfrac{\Pr(y, Z)}{\Pr(Z)}$ we need to consider the two main sets of cases: $Y = 1$ and $Y \geq 2$:

$$\Pr(Y = 1|Z) = \frac{\Pr(Y = 1, Z)}{\Pr(Z)} = \frac{\Pr(Z, X = 7) + \Pr(Z, X = 11)}{\Pr(Z)} = \frac{\frac{1}{6} + \frac{1}{18}}{\frac{244}{495}} = \frac{55}{122}$$

For $y \geq 2$:

$$\Pr(Y = y|Z) = \frac{1}{\Pr(Z)} \sum_{x \in \{4,5,6,8,9,10\}} \Pr(x, y, Z) = \frac{495}{244} \left(\left(\frac{1}{12}\right)^2 \left(\frac{3}{4}\right)^{y-2} + \left(\frac{1}{9}\right)^2 \left(\frac{13}{18}\right)^{y-2} + \dots \right)$$

$$= 2\frac{495}{244} \left(\left(\frac{1}{12}\right)^2 \left(\frac{3}{4}\right)^{y-2} + \left(\frac{1}{9}\right)^2 \left(\frac{13}{18}\right)^{y-2} + \left(\frac{5}{36}\right)^2 \left(\frac{25}{36}\right)^{y-2} \right)$$

It can be proved that:

$$\sum_{y=2}^{\infty} y \cdot p^{y-2} = \frac{2 - p}{(1 - p)^2}$$

So:

$$\color{red}{\mathbf{E[Y|Z]}} = 1 \cdot \frac{55}{122} + \frac{495}{122} \left(\left(\frac{1}{12}\right)^2 \frac{2 - \frac{3}{4}}{\left(1 - \frac{3}{4}\right)^2} + \left(\frac{1}{9}\right)^2 \frac{2 - \frac{13}{18}}{\left(1 - \frac{13}{18}\right)^2} + \left(\frac{5}{36}\right)^2 \frac{2 - \frac{25}{36}}{\left(1 - \frac{25}{36}\right)^2} \right)$$

$$= \frac{55}{122} + \frac{495}{122} \left(\frac{1}{144}\frac{5}{4}\frac{16}{1} + \frac{1}{81}\frac{23}{18}\frac{18^2}{25} + \frac{25}{36^2}\frac{47}{36}\frac{36^2}{121} \right) = \frac{55}{122} + \frac{495}{122} \left(\frac{5}{36} + \frac{46}{225} + \frac{1175}{4356} \right) = \frac{9858}{3355} \approx \color{red}{\mathbf{2.9383}}$$

$$E[Y|\bar{Z}] = \sum_{y=1}^{\infty} y \Pr(y|\bar{Z})$$

$$\Pr(y|\bar{Z}) = \frac{\Pr(y, \bar{Z})}{\Pr(\bar{Z})}$$

$$\Pr(Y = 1|\bar{Z}) = \frac{\Pr(Y = 1, \bar{Z})}{\Pr(\bar{Z})} = \frac{\Pr(\bar{Z}, X = 2) + \Pr(\bar{Z}, X = 3) + \Pr(\bar{Z}, X = 12)}{\Pr(Z)} = \frac{\frac{1 + 2 + 1}{36}}{\frac{251}{495}} = \frac{55}{251}$$

For $y \geq 2$:

$$\Pr(Y = y|\bar{Z}) = \frac{1}{\Pr(\bar{Z})} \sum_{x \in \{4,5,6,8,9,10\}} \Pr(x, y, \bar{Z}) = \frac{495}{251} \left(\frac{1}{12}\frac{1}{6} \left(\frac{3}{4}\right)^{y-2} + \frac{1}{9}\frac{1}{6} \left(\frac{13}{18}\right)^{y-2} + \dots \right)$$

$$= 2\frac{495}{251}\left(\frac{1}{12}\frac{1}{6}\left(\frac{3}{4}\right)^{y-2} + \frac{1}{9}\frac{1}{6}\left(\frac{13}{18}\right)^{y-2} + \frac{5}{36}\frac{1}{6}\left(\frac{25}{36}\right)^{y-2}\right)$$

$$\mathbf{E}[\mathbf{Y}|\bar{\mathbf{Z}}] = 1\cdot\frac{55}{251} + \frac{990}{251}\left(\frac{1}{12}\frac{1}{6}\frac{2-\frac{3}{4}}{\left(1-\frac{3}{4}\right)^2} + \frac{1}{9}\frac{1}{6}\frac{2-\frac{13}{18}}{\left(1-\frac{13}{18}\right)^2} + \frac{5}{36}\frac{1}{6}\frac{2-\frac{25}{36}}{\left(1-\frac{25}{36}\right)^2}\right)$$

$$= \frac{55}{251} + \frac{990}{251}\left(\frac{1}{12}\frac{1}{6}\frac{5}{4}\frac{16}{1} + \frac{1}{9}\frac{1}{6}\frac{23}{18}\frac{18^2}{25} + \frac{5}{36}\frac{1}{6}\frac{47}{36}\frac{36^2}{121}\right) = \frac{55}{251} + \frac{990}{251}\left(\frac{5}{18} + \frac{23}{75} + \frac{235}{726}\right) = \frac{52473}{13805} \approx \mathbf{3.8010}$$

$$\mathbf{E}[\mathbf{Y}|\mathbf{Z}] < \mathbf{E}[\mathbf{Y}|\bar{\mathbf{Z}}]$$

Let's prove it with a simulation:

```
N <- 100e3 # number of simulations
Y <- rep(0, N) # value of Y (number of rolls for every simulation)
Z <- rep(0, N) # value of Z (0: house wins; 1: I win)
for (i in 1:N) {
    y <- 0 # initialize number of rolls
    x <- sample(seq(6), 1) + sample(seq(6), 1) # simulate two dice
    y <- y+1 # update number of rolls
    first_x <- x # store x (will be used if it's 4, 5, 6, 8, 9, 10)
    # House wins if 2, 3, 12
    # I win if 7, 11
    # Otherwise, keep that value of x (and insert in z)
    z <- ifelse(x==2 | x==3 | x==12, 0,
                ifelse(x==7 | x==11, 1, first_x))
    # While z > 1 (i.e., if nobody wins)
    while (z>1) {
        x <- sample(seq(6), 1) + sample(seq(6), 1) # simulate other roll
        y <- y+1 # update number of rolls
        # House wins if 7
        # I win if same value of 1st roll
        z <- ifelse(x==7, 0, ifelse(x==first_x , 1, first_x))
    }
    # When finished, store values for that simulation
    Y[i] <- y
    Z[i] <- z
}

mean(Z) # E[Z] (=Pr(Z=1) since Z is binary); should be close to 0.493 (244/495)
```

```
## [1] 0.49159
```

```
mean(Y[Z==0]) # E[Y|Z]; should be close to 3.8010
```

```
## [1] 3.802404
```

```
mean(Y[Z==1]) # E[Y|not Z]; should be close to 2.9383
```

```
## [1] 2.925914
```

2. **Suppose it takes \$20 to pay, 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.**

Let W be the payoff:

$$\mathbf{E[W]} = \sum_{\substack{w\in\{0,40,\\60,80,100\}}} \mathbf{w}\Pr(\mathbf{w}) = \sum_{\substack{w\in\{0,40,\\60,80,100\}}} \mathbf{w}\Pr(\mathbf{y}) = \sum_{\substack{w\in\{0,40,\\60,80,100\}}} \mathbf{w}\sum_{\mathbf{Z=0}}^{\mathbf{1}} \mathbf{w}\Pr(\mathbf{y},\mathbf{Z}) = \sum_{\substack{w\in\{0,40,\\60,80,100\}}} \mathbf{w}\sum_{\mathbf{Z=0}}^{\mathbf{1}} \mathbf{w}\Pr(\mathbf{y}|\mathbf{Z})\Pr(\mathbf{Z})$$

We can focus on the event $Z = 1$ since $\bar{Z} \Rightarrow W = 0$:

$$E[W] = \sum_{\substack{w\in\{0,40,\\60,80,100\}}} \mathbf{w}\Pr(\mathbf{y}|\mathbf{Z=1})\Pr(\mathbf{Z=1}) = \Pr(\mathbf{Z=1}) \sum_{\substack{w\in\{40,60,\\80,100\}}} \mathbf{w}\Pr(\mathbf{y}|\mathbf{Z=1})$$

$$= \Pr(Z=1)\sum_{y=1}^{4} w(y)\Pr(y|Z=1)$$

$$= \frac{244}{495}\left(100\cdot\Pr(Y=1|Z=1) + \textstyle\sum_{y=2}^{4} w(y)\frac{495}{122}\left(\left(\frac{1}{12}\right)^2\left(\frac{3}{4}\right)^{y-2}\right)\right)$$

$$= \frac{244}{495}\left(100\cdot\frac{55}{122} + \sum_{y=2}^{4} w(y)\frac{495}{122}\left(\left(\frac{1}{12}\right)^2\left(\frac{3}{4}\right)^{y-2} + \left(\frac{1}{9}\right)^2\left(\frac{13}{18}\right)^{y-2} + \left(\frac{5}{36}\right)^2\left(\frac{25}{36}\right)^{y-2}\right)\right)$$

$$= 100\cdot\frac{110}{495} + 80\cdot 2\left(\left(\frac{1}{12}\right)^2 + \left(\frac{1}{9}\right)^2 + \left(\frac{5}{36}\right)^2\right)$$

$$+60\cdot 2\left(\left(\frac{1}{12}\right)^2\left(\frac{3}{4}\right) + \left(\frac{1}{9}\right)^2\left(\frac{13}{18}\right) + \left(\frac{5}{36}\right)^2\left(\frac{25}{36}\right)\right)$$

$$+40\cdot 2\left(\left(\frac{1}{12}\right)^2\left(\frac{3}{4}\right)^2 + \left(\frac{1}{9}\right)^2\left(\frac{13}{18}\right)^2 + \left(\frac{5}{36}\right)^2\left(\frac{25}{36}\right)^2\right)$$

$$= 100\cdot\frac{110}{495} + \frac{1}{144}\left(160\cdot 1 + 120\cdot\frac{3}{4} + 80\cdot\frac{9}{16}\right)$$

$$= +\frac{1}{81}\left(160\cdot 1 + 120\cdot\frac{13}{18} + 80\cdot\frac{25}{36}\right)$$

$$= +\frac{25}{1296}\left(160\cdot 1 + 120\cdot\frac{25}{36} + 80\cdot\frac{625}{1296}\right)$$

$$= \mathbf{\frac{160658}{4829} \approx 33.2694}$$

```r
N <- 100e3 # number of simulations
Y <- rep(0, N) # value of Y (number of rolls for every simulation)
Z <- rep(0, N) # value of Z (0: house wins; 1: I win)
P <- rep(0, N) # payoff
for (i in 1:N) {
    y <- 0 # initialize number of rolls
    x <- sample(seq(6), 1) + sample(seq(6), 1) # simulate two dice
    y <- y+1 # update number of rolls
    first_x <- x # store x (will be used if it's 4, 5, 6, 8, 9, 10)
    # House wins if 2, 3, 12
    # I win if 7, 11
    # Otherwise, keep that value of x (and insert in z)
    z <- ifelse(x==2 | x==3 | x==12, 0,
                ifelse(x==7 | x==11, 1, first_x))
    p <- ifelse(z==1, 100, 0) # if I win 1st time, payoff is 100
    # While z > 1 (i.e., if nobody wins)
    j <- 2 # keep count of rolls
    while (z>1) {
        x <- sample(seq(6), 1) + sample(seq(6), 1) # simulate other roll
        y <- y+1 # update number of rolls
        # House wins if 7
        # I win if same value of 1st roll
        z <- ifelse(x==7, 0, ifelse(x==first_x , 1, first_x))
        # Compute payoff if I win (with less than 5 rolls)
        p <- ifelse(z==1 & j<5, 100-20*(j-1), 0)
        j <- j+1 # update count of rolls
    }
    # When finished, store values for that simulation
    Y[i] <- y
    Z[i] <- z
    P[i] <- p
}

mean(P) # E[P]; should be close to 33.2694
```

```
## [1] 33.0558
```

To calculate the net profit we would just have to subtract \$20 to every game:

$E[\text{net profit}] = E[W - 20] = E[W] - 20 = 13.2694$

(It's positive, so it does not seem likely that the house would choose payoffs so high: usually they are selected so $E[\text{net profit}] < 0$, so the house always wins on the long term.)

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 $Y1, \dots, Yn$ 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 $Y_i = \mu$ ∀ i ∈ 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 numbers $a_1, \dots, a_n$, the following inequality holds:**

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

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?**

$$W_1 = (\frac{n-1}{n})\bar{Y}$$
$$E(W_1) = \frac{n-1}{n}E(\bar{Y})$$
$$= \frac{n-1}{n}\mu$$

Therefore

$$Bias(W_1) = \frac{n-1}{n}\mu - \mu$$
$$nBias(W_1) = -\mu$$
$$Bias(W_1) = \frac{-\mu}{n}$$

Note that as $n \to \infty$, Bias(W_1) $\to 0$.

$$W_2 = k\bar{Y}$$
$$E(W_2) = kE(\bar{Y})$$
$$E(W_2) = k\mu$$

And

$$Bias(W_2) = k\mu - \mu$$
$$Bias(W_2) = \mu(k-1)$$

Given that $0 <$ k < 1, -$\mu <$ Bias(W_2) < 0.
Note that W_1 is consistent, as E(W_1) converges to μ as $n \to \infty$.

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

$$Var(W_1) = (\frac{n-1}{2})^2 Var(\bar{(Y)})$$
$$Var(W_1) = \frac{n-1}{n}\frac{n-1}{n}\frac{\sigma^2}{n}$$
$$Var(W_1) = \left(\frac{(n-1)^2}{n^3}\right)\sigma^2$$

$$Var(W_2) = k^2 Var(\bar{(Y)})$$
$$Var(W_2) = \frac{k^2}{n}\sigma^2$$

Note that by choosing k sufficiently close to 0, the variance of W_2 can be made arbitrarily small. Thus W_2 has the lower variance.

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:

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

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

$$E(\hat{Y}) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i)$$
$$= \frac{1}{n}[E(Y_1) + E(Y_2) + ... + E(Y_n)]$$

Because $Y_1\ \ldots\ Y_n$ are from the same distribution with mean μ,

$$E(\hat{Y}) = \frac{1}{n}[\mu + \mu + ... + \mu]$$

Thus

$$E(Y_1) = E(Y_2) = \cdots = E(Y_n) = E(\hat{Y})$$

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

$$Var(\hat{Y}) = \sum_{i=1}^{n} Var(\frac{Y_i}{n})$$
$$= Var\left(\frac{1}{n}Y_1 + \frac{1}{n}Y_2 + \cdots + \frac{1}{n}Y_n\right)$$
$$= \frac{1}{n^2}Var(Y_1) + \frac{1}{n^2}Var(Y_2) + \cdots + \frac{1}{n^2}Var(Y_n)$$

Since $Y_1\ \ldots\ Y_n$ are from the same distribution with variance σ^2

$$= \frac{1}{n^2}[\sigma^2 + \sigma^2 + ... + \sigma^2]$$

Note that σ^2 occurs n times

$$= \frac{1}{n^2}[n\sigma^2]$$

Thus

$$Var(\hat{Y}) = \frac{1}{n}Var(Y_i)\forall i \in 1, 2, \ldots, n$$

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.**

$$\begin{aligned} E(\widehat{\sigma^2}) &= E\bigg(\frac{(Y_i - \bar{Y})^2}{n}\bigg) \\ nE(\widehat{\sigma^2}) &= E\big(\sum_{i=1}^{n} Y_i^2 - 2Y_i\bar{Y} + \bar{Y}^2\big) \\ &= E\bigg[\sum_{i=1}^{n} Y_i^2\bigg] - E\bigg[2\bar{Y}\sum_{i=1}^{n} Y_i\bigg] + E\bigg[\bar{Y}^2\sum_{i=1}^{n} 1\bigg] \end{aligned}$$

Note that $\sum_{i=1}^{n} Y_i = \mathrm{n}\bar{Y}$

$$\begin{aligned} nE(\widehat{\sigma^2}) &= E\bigg[\sum_{i=1}^{n} Y_i^2\bigg] - nE\bigg[\bar{Y}^2\bigg] \\ &= nE[Y_i^2] - nE[\bar{Y}^2] \\ E(\widehat{\sigma^2}) &= E[Y_i^2] - E[\bar{Y}^2] \end{aligned}$$

Note that for independent and identically distributed random variables $Y_i \ldots Y_n$ with sample mean $\bar{Y}$

$$E[Y^2] = E[\bar{Y^2}] = \frac{1}{n}\sigma^2 - \mu^2$$

Substituting into our equation gives

$$\begin{aligned} E(\widehat{\sigma^2}) &= \big[\sigma^2 + \mu^2\big] - \big[\frac{1}{n}\sigma^2 + \mu^2\big] \\ &= \frac{(n-1)}{n}\sigma^2 \end{aligned}$$

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

Since $E(\widehat{\sigma^2}) \neq \sigma^2$, this is a biased estimator.

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

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

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 ounces) 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. H_0: $\mu = 0$.
ii. H_a: $\mu < 0$.
iii.

$$t = \frac{\bar(y) - \mu}{s/\sqrt{n}} = \frac{-32.8}{466.4/\sqrt{900}} \approx -2.11$$

p = Pr(z ≤ -2.11) = .017

Because 0.01 < p < 0.05, we can reject H_0 at the 5% level but not at the 1% level.

iv. Given our estimate of μ, we are estimating the tax was associated with a 32.8 oz decrease in alcohol consumption. From a health perspective, this is less than a 1 oz decrease in consumption of liquor per week, a fairly insignificant difference.
v. This analysis assumes that other factors that influence alcohol consumption over the period of interest have remained constant.

Question 13

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

8. **For a given player, the outcome of a particular shot can be modeled as a Bernoulli variable: if Y_i is the outcome of shot i, then 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 u is $\bar{Y}$ = FGM/FGA.**

i.

$$\theta = 188/429 \approx 0.438$$

ii. For each trial

$$Var(Y_i) = \theta(1 - \theta)$$

Thus for the sample mean we have

$$Var(\bar{Y}) = \frac{\theta(1 - \theta)}{n}$$

$$sd(\bar{Y}) = \sqrt{\frac{\theta(1 - \theta)}{n}}$$

iii. For $\theta = 0.5$, we have $t = (\bar{Y} - 0.5)/se(\bar{Y})$

$$se = \sqrt{0.438(1 - .438)/429} \approx 0.024$$

$$t = (.438 - 0.5)/0.024 \approx -2.58$$

$$p \approx Pr(Z \leq -2.58) \approx 0.004$$

At the 1% significance level, we reject the null hypotheses that $\theta = 0.5$

1. **Define Type I error.**

Rejecting the null hypothesis when it is true.

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

0.01

3. **Define Type II error.**

Accepting the null hypothesis when it is false.

4. **What is the probability of Type II error when using this decision rule, assuming the "true" population proportion is $\theta^* = 0.45$.**

At the 1% significance level, Z_{crit} = -2.33. And

$$-2.33 = (\theta^*_{crit} - 0.5)/0.24$$
$$\theta^*_{crit} = .444.$$
$$P(\theta^*_{crit} \geq .444 | \mu = 0.45 \text{ and } \sigma = 0.024) = Pr(Z \geq -.25)) \approx 0.60.$$

The type II error rate in this scenario is 0.60.

5. **Define the power of the test (in general terms).**

The power of a test is the probability that we correctly reject the null hypothesis, ie. Pr(Reject | Null is False).

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

The power is 1 - 0.65 = 0.40

```r
plot(cars)
```

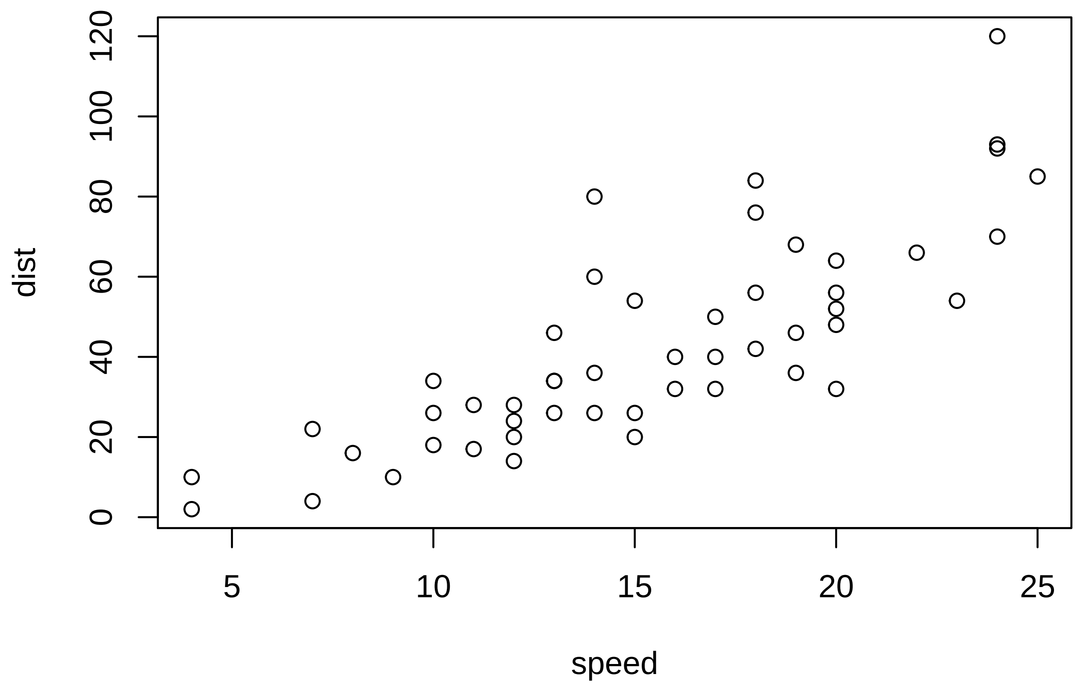

Figure 5: Some plot