Lab1

January 28, 2016

1 W271 Applied Regression and Time Series Analysis - Lab 1

1.0.1 Submitted by: Ron Cordell | Subhashini Raghunathan | Lei Yang

1.0.2 Date: 1/28/2016

1.1 Part I: Marginal, Joint, and Conditional Probabilities

1.1.1 Question 1 (6 Points)

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 machine learning or statistics or are awesome. Suppose you have 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 A = Awesome, M = Machine Learning Expert and S = Statistics Expert Given:

$A = 18, S = 28, M = 36$

$(A \cap S) = 9, (A \cap M) = 12, (S \cap M) = 22$

Assuming that there are no team members that don't have these attributes then: $(A \cup S \cup M) = 48$

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

$48 = 18 + 28 + 36 - 9 - 12 - 22 + (A \cap S \cap M)$

$(A \cap S \cap M) = 9$

We can also say that

$M = (M \cap S) + (M \cap A) - (A \cap S \cap M) + m$

$36 = 22 + 12 - 9 + m \Rightarrow m = 11$

$S = (S \cap M) + (S \cap A) - (A \cap S \cap M) + s$

$28 = 22 + 9 - 9 + s \Rightarrow s = 6$

$A = (A \cap M) + (A \cap S) - (A \cap S \cap M) + a$

$18 = 12 + 9 - 9 + a \Rightarrow a = 6$

$Pr(A = a, M = m, S = s) = 9/48$

The probability of meeting a datascientist who is an expert in machine learning, statistics and is awesome is $\dfrac{3}{16}$

2. Suppose you meet a data scientist who is an expert in machine learning. Given this information, what is the probability that they are not awesome? $Pr(\bar{A} \mid M) = Pr(A\bar{M})/Pr(M) = \dfrac{24}{36} = \dfrac{2}{3}$

3. Suppose you meet a data scientist who is awesome. Given this information, what is the probability that they are an expert in either machine learning or statistics? $Pr(M \cup S \mid A) = \dfrac{3+9+0}{18} = \dfrac{12}{18} = \dfrac{2}{3}$

1.1.2 Question 2 (4 Points)

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

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

specify the conditions and use interval notation It depends on whether A and B are independent (overlap). If they are not independent, then:

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

Since $Pr(A)$ can range from 0 to $\frac{1}{2}$ the lowest value for $Pr(A \cup B)$ is

$$0 + \lim_{\delta \to 0}(\frac{1}{4} + \delta) - 0 \times \lim_{\delta \to 0}(\frac{1}{4} + \delta) = \frac{1}{4}$$

At

$$Pr(A) = \frac{1}{2} \text{ and } Pr(B) = q = \lim_{\delta \to 0}(\frac{1}{2} - \delta)$$

we have the high value of

$$\frac{1}{2} + q - (\frac{1}{2}q) = \frac{1}{2} + \frac{1}{2}\lim_{\delta \to 0}(\frac{1}{2} - \delta)$$

If A and B are dependent, $\frac{1}{4} <= Pr(A \cup B) < \frac{5}{8}$

Otherwise, if A and B are independent, then:

$Pr(A \cup B) = Pr(A) + Pr(B)$ so at the low end we would have $0 + \lim_{\delta \to 0}(\frac{1}{4} + \delta)$ and at the high end $\frac{1}{2} + \lim_{\delta \to 0}(\frac{1}{2} - \delta)$

If A and B are independent, $0 \leq Pr(A \cup B) < 1$.

Therefore the full range for is

$$0 \leq Pr(A \cup B) < 1$$

2. What are the maximum and minimum possible values for $Pr(A|B)$? Again, we don't know anything about the independence of A and B.

If A and B are not independent:

$Pr(A \mid B) = \dfrac{Pr(AB)}{Pr(B)}$ and since $Pr(A)$ can be 0 the lower bound is 0

If A is contained in B then the probability is $\lim_{\delta \to 0}(1 - \delta)$.

Therefore the range of $Pr(A|B)$ is

$$0 \leq Pr(A|B) < 1$$

2

1.2 Part II: Random Variables, Expectation, Conditional Expectation

1.2.1 Question 3 (6 Points)

Suppose the life span of a particular server is a continuous random variable, t, with a uniform probability distribution between 0 and k years, 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 $x = A(k-t)^{\frac{1}{2}}$, 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 $\frac{\theta}{10}$

Probability distribution function is: $f(t) = \begin{cases} 0 & t \leq 0 \\ \frac{1}{k} & 0 \leq t \leq k \\ 0 & t > k \end{cases}$

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

$$Pr(t = \frac{k}{4} + 1 \mid t = \frac{k}{4}) = \int_{t=\frac{k}{4}}^{t=\frac{k}{4}+1} \frac{1}{k} dt = \frac{t}{k} \Big|_{t=\frac{k}{4}}^{t=\frac{k}{4}+1} = \frac{\frac{k}{4}+1}{k} - \frac{\frac{k}{4}}{k} = \frac{1}{k}(\frac{k}{4} + 1) - \frac{1}{k}\frac{k}{4}$$

$$= \frac{1}{4} + \frac{1}{k} - \frac{1}{4} = \frac{1}{k} \text{ for } k \geq \frac{4}{3}$$

When $k < \frac{4}{3}$: since k is an integer, $k = 1$.
Therefore

$$P(\text{server lives 1 year}) = \int_{t=\frac{1}{4}}^{t=1} 1 dt = t \Big|_{\frac{1}{4}}^{1} = 1 - \frac{1}{4} = \frac{3}{4}$$

2. Compute the expected payout from the contract, $E(x)$. $E[x] = \begin{cases} \theta & 0 \leq t \leq 1 \\ A(k-t)^{\frac{1}{2}} & 1 < t \leq \frac{k}{2} \\ \frac{\theta}{10} & \frac{k}{2} < t \leq \frac{3k}{4} \\ 0 & t > \frac{3k}{4} \end{cases}$

For $0 \leq t \leq 1$ we have:

$$E[x] = \theta \times Pr(t \leq 1) = \theta \int_{t=0}^{t=1} \frac{1}{k} dt = \frac{\theta}{k}$$

For $1 < t \leq \frac{k}{2}$ we have:

$$E[x] = \int_{t=1}^{t=\frac{k}{2}} 2(k-t)^{\frac{1}{2}} \frac{1}{k} dt = \frac{2}{k} \int_{t=1}^{t=\frac{k}{2}} (k-t)^{\frac{1}{2}} dt$$

$$= \frac{-4}{3k} (k-t)^{\frac{3}{2}} \Big|_{t=1}^{t=\frac{k}{2}}$$

$$= \frac{4}{3k} \left[(k-1)^{\frac{3}{2}} - (\frac{k}{2}^{\frac{3}{2}}) \right]$$

For $\frac{k}{2} < t < \frac{3k}{4}$ we have:

$$E[x] = \frac{\theta}{10} \int_{t=\frac{k}{2}}^{t=\frac{3k}{4}} \frac{1}{k} dt$$

3

$$= \frac{\theta}{10}\frac{t}{k}\Big|_{\frac{k}{2}}^{\frac{3k}{4}}$$

$$= \frac{\theta}{10}\left(\frac{3}{4} - \frac{1}{2}\right) = \frac{\theta}{40}$$

3. Compute the variance of the payout from the contract. $Var(x) = E[x^2] - E[x]^2$

For $0 \le t \le 1$ we have:

$$E[x^2] = \int_{t=0}^{t=1} \theta^2 \frac{1}{k} dt = \frac{\theta^2}{k}$$

From the previous problem:

$$E[x] = \frac{\theta}{k}$$

Therefore:

$$Var(x) = E[x^2] - E[x]^2 = \frac{\theta^2}{k} - \left(\frac{\theta}{k}\right)^2 = \frac{\theta^2}{k}\left(1 - \frac{1}{k}\right)$$

For $1 < t \le \frac{k}{2}$ we have:

$$E[x^2] = \int_{t=1}^{\frac{k}{2}} 2(k-t)^2 \frac{1}{k} dt = \frac{4}{k}\int_{t=1}^{t=\frac{k}{2}}(k-t)^2 dt$$

$$= \frac{4}{k}\left(\frac{-t^2}{2}\Big|_{t=1}^{t=\frac{k}{2}} + tk\Big|_{1}^{\frac{k}{2}}\right) = \frac{4}{k}\left(\frac{-1}{2}\left(\frac{k^2}{4}-1\right) + k\left(\frac{k}{2}-1\right)\right) = \frac{4}{k}\left(\frac{-k^2}{8} + \frac{1}{2} + \frac{k^2}{2} - k\right)$$

Therefore

$$E[x^2] = \frac{3k}{2} + \frac{2}{k} - 4$$

From earlier results we can calculate $E[x]^2$:

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

$$= \frac{16}{9k^2}\left[(k-1)^3 - 2(k-1)^{\frac{3}{2}}\left(\frac{k}{2}\right)^{\frac{3}{2}} + \left(\frac{k}{2}\right)^3\right]$$

$$= \frac{16}{9k^2}(k-1)^3 - \frac{2^{\frac{7}{2}}}{9k^{\frac{1}{2}}}(k-1)^{\frac{3}{2}} + \frac{2k}{9}$$

Therefore:

$$Var(X) = E[X^2] - E[X]^2 = \frac{3k}{2} + \frac{2}{k} - 4 - \left[\frac{16}{9k^2}(k-1)^3 - \frac{2^{\frac{7}{2}}}{9k^{\frac{1}{2}}}(k-1)^{\frac{3}{2}} + \frac{2k}{9}\right]$$

$$= \frac{3k}{2}k + \frac{2}{k} - 4 - \frac{16}{9k^2}(k-1)^3 + \frac{2^{\frac{7}{2}}}{9k^{\frac{1}{2}}}(k-1)^{\frac{3}{2}} - \frac{2k}{9}$$

$$Var(X) = \frac{23}{18}k + \frac{2}{k} - \frac{16}{9k^2}(k-1)^3 + \frac{2^{\frac{7}{2}}}{9k^{\frac{1}{2}}}(k-1)^{\frac{3}{2}} - 4$$

For $\dfrac{k}{2} < t < \dfrac{3k}{4}$ we have:

$$E[x^2] = \int_{t=\frac{k}{2}}^{t=\frac{3k}{4}} \left(\frac{\theta}{10}\right)^2 \frac{1}{k} dt = \frac{\theta^2}{100}\left[\frac{3k}{4} - \frac{k}{2}\right] = \frac{\theta^2}{100}\frac{1}{4}k = \frac{\theta^2}{400}$$

Using E[X] from previous results we have:

$$Var(x) = E[x^2] - E[x]^2 = \frac{\theta^2}{400} - \left(\frac{\theta}{40}\right)^2 = \frac{3}{1600}\theta^2$$

1.2.2 Question 4 (9 Points)

Continuous random variables X and Y have a joint distribution with probability density function

$$f(x,y) = 2e^{-x}e^{-2y} \begin{cases} 0 \le x \le \infty \\ 0 \le y \le \infty \end{cases}$$

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

$$Pr(X > a, Y < b) = \int_{y=0}^{y=b} \int_{x=a}^{x=\infty} 2e^{-x}e^{-2y} dx dy = \int_{y=0}^{y=b} e^{-2y} dy \int_{x=a}^{x=\infty} 2e^{-x} dx$$

$$= -\frac{1}{2}e^{-2y}\Big|_{y=0}^{y=b} 2(-e^{-x})\Big|_{x=a}^{\infty} = (-e^{-2y})\Big|_{y=0}^{y=b}(-e^{-x})\Big|_{x=a}^{\infty}$$

$$= (-e^{-2b} + \frac{1}{2})(0 + e^{-a})$$

$$= e^{-a} - e^{-a-2b}$$

2. Compute $P(X < Y)$

$$Pr(X < Y) = \int_{y=0}^{y=\infty} \int_{x=0}^{x=y} 2e^{-x}e^{-2y} dx dy$$

$$= 2\int_{y=0}^{y=\infty} \left(-e^{-y} + 1\right)e^{-2y} dy = 2\int_{y=0}^{y=\infty} -e^{-3y} + e^{-2y} dy$$

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

3. Compute $P(X < a)$

$$Pr(X < a) = \int_{x=0}^{x=a} \int_{y=0}^{y=\infty} 2e^{-x}e^{-2y} dx dy = e^{-x}\Big|_{x=0}^{x=a} e^{-2y}\Big|_{y=0}^{y=\infty} = 1 - e^{-a}$$

1.2.3 Question 5 (6 Points)

1. Find the value of x that minimizes E(Y). Show that your result is really the minimum. We
first expand the definition of Y:

$$Y = a + b(X - x)^2 = a + bx^2 + bX^2 - 2bxX$$

then $E(Y)$ becomes:

$$E(Y) = a + bx^2 + bE(X^2) - 2bxE(X)$$

to find the x that minimums $E(Y)$, we take derivative with respect to x:

$$\frac{dE(Y)}{dx} = 2bx - 2bE(X)$$

set it to 0 and we will find the x that minimizes $E(Y)$:

$$x = E(Y)$$

Geometrically, by subtracting $E(X)$ from X we move the whole population of X around zero, which will
give the minimum square value of the population.

2. Find the value of E(Y) for the choice of x you found in (1)? Plug in $x = E(X)$ into the formula
of $E(Y)$, we have:

$$E(Y) = a + bE(X)^2 + bE(X^2) - 2bE(X)^2 = a + b(E(X^2) - E(X)^2)$$

and we have

$$E(Y)_{min} = a + bVar(X)$$

**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.** Similar to part #1, we first obtain the new $E(Y)$:

$$E(Y) = ax + bx^2 + bE(X^2) - 2bxE(X)$$

take derivative of $E(Y)$:

$$\frac{dE(Y)}{dx} = a + 2bx - 2bE(X)$$

set it to zero, we have the x that minimums $E(Y)$

$$x = E(X) - \frac{a}{2b}$$

plug x into $E(Y)$, and the minimum is:

$$E(Y)_{min} = aE(X) + bVar(X) - \frac{a^2}{2b}$$

1.2.4 Question 6 (4 Points)

Suppose X and Y are independent continuous random variables where both 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. $f(x) = f(y) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x \leq 1 \\ 0 & x > 1 \end{cases}$

The line x + y = z is shown in brown in the figure above. It has slope = -1 and X and Y intercepts = z. The shaded region in the first quadrant is where X + Y < Z. Since X and Y are defined only in the interval [0,1], X + Y is also defined only for the first quadrant.

2. Derive the probability density function, $f(z)$. The figure above shows the support of the joint probability density function of X and Y. There are 3 cases to consider: 1. $0 < Z < 1$ 2. $1 < Z < 2$ 3. $Z > 2$

For each case, we integrate over the joint probability density function to get $CDF_Z(z)$, then take the derivate to get $f_Z(z)$

Since X and Y are independent, $f_{X,Y}(x, y) = f_X(x) * f_Y(y) = 1$

$Case1 : 0 < Z < 1$

$$F_Z(z) = \int_{x=0}^{z} \int_{y=0}^{z-x} 1 \, dy \, dx$$

$$= \int_{x=0}^{z} \Big|_{y=0}^{z-x} y \, dx = \int_{x=0}^{z} (z - x) dx = \Big|_{x=0}^{z} zx - \Big|_{x=0}^{z} \frac{x^2}{2}$$

$$= z^2 - \frac{z^2}{2} = \frac{z^2}{2}$$

$$\frac{d(F)}{dz} = \frac{d}{dz} \frac{z^2}{2} = z$$

Thus, $f_Z(z) = z$ when $0 < Z < 1$

$Case2 : 1 < Z < 2$

$$F_Z(z) = \int_{x=z-1}^{1} \int_{y=z-x}^{1} 1 \, dy \, dx$$

$$= \int_{x=z-1}^{1} \Big|_{y=z-x}^{1} y \, dx = \int_{x=z-1}^{1} (1 - z + x) dx = (1 - z) \Big|_{x=z-1}^{1} x - \Big|_{x=z-1}^{1} \frac{x^2}{2}$$

$$= (1 - z)(2 - z) + \frac{1}{2}(1 - (Z^2 - 2z + 1)) = 2 - 2z + \frac{z^2}{2}$$

$$\frac{d(F)}{dz} = \frac{d}{dz} 2 - 2z + \frac{z^2}{2} = z - 2$$

Thus, $f_Z(z) = z - 2$ when $1 < Z < 2$

$ Case 3: Z > 2$

This region is outside the support of $f_{X,Y}(x, y)$, hence $f_Z(z) = 0$

[1:21] Just above Lab1Q7.1 was this line: 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.

1.2.5 Question 7 (10 Points)

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. Let's say that $h = \{2, 3, 12\}$, $y = \{7, 11\}$ and $x = \{4, 5, 6, 8, 9, 10\}$ such that on any roll of the dice, the result will be either a value in h, y, or x.

We can calculate probabilities: $P(h) = P(2) + P(3) + P(12), P(y) = P(7) + P(11),$ and $P(x) = P(4) + P(5) + P(6) + P(8) + P(9) + P(10)$

$P(h) = \frac{4}{36} P(y) = \frac{8}{36} P(x) = \frac{24}{36}$ and $P(h) + P(y) + P(x) = 1$

$P(7) = \frac{6}{36}$

The rolls on which the house can win are:

Roll 1: $r_1 \in h$

Roll 2: $[r_2 \in h]$ OR $[r_1 \in x$ AND $r_2 = 7]$

Roll $n > 2$: $r_n \in h$ OR $[r_{n-1} \in x$ AND $r_2 = 7]$ OR $[r_m = x_i$ AND $r_{m+1,\ldots,n-1} = x_j \in x, x_j \neq x_i$ AND $r_n = 7]$

The rolls on which you can win are:

Roll 1: $r_1 \in y$

Roll 2: $[r_2 \in y]$ OR $[r_1 = x_i \in x$ AND $r_2 = x_i]$

Roll $n > 2$: $r_n \in h$ OR $[r_1 = x_i \in x$ AND $r_2 = x_i]$ OR $[r_m = x_i$ AND $r_{m+1,\ldots,n-1} = x_j \in x, x_j \neq x_i$ AND $r_n = x_i]$

The probability that the we roll m rolls of $x_j \neq x_i = \delta$

The probability that we roll $x_i \in x$ twice is $\sum_i x_i^2 = \frac{100}{1296}$

Roll	house	you
1	4/36	8/36
2	$4/36 + (24/36)(6/36) = 8/36$	$8/36 + 100/1296$
n>2	$8/36 + (6/36)\delta^n$	$8/36 + (100/1296)\delta^n$

From the table we see that the $P(\text{house wins})$ is greater than $P(\text{you win})$ for Rolls > 2

Therefore $E[R = r|\text{ house wins}] < E[R = r|\text{ you win}]$

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. The expected payoff function for you to win: $E(payoff) = P_1 \times 80 + P_2 \times 60 + P_3 \times 40 + P_4 \times 20$, where P_i is the probability that you win on the i-th roll. Let's calculate P_i respectively:

$$P_1 = \frac{8}{36}$$

$$P_2 = \frac{24}{36} \times \frac{24}{36}$$

$$P_3 = \frac{24}{36} \times \frac{6}{36} \times \frac{24}{36}$$

$$P_4 = \frac{24}{36} \times \frac{6}{36} \times \frac{6}{36} \times \frac{24}{36}$$

Note: the probability for any $(i-1)$-th roll without a winner is $1 - \frac{24}{36} - \frac{6}{36} = \frac{6}{36}$ Thus the expected payoff is $E(payoff) = 47.65$

1.3 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.

In [ ]:

1.3.1 Question 8 (10 Points)

Let $Y1, ..., Y_n$ be n random variables, such that any two of them are uncorrelated, and all share the same mean μ and variance σ^2. Let $\overline{Y}$ 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 ai is needed for W to be an unbiased estimator of μ? We know the unbiased estimator of population mean μ is just the sample mean $\overline{Y}$, for W to be unbiased $E(W) = \mu$:

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

$\forall \quad i \in 1, ..., n$, we have $E(Y_i) = \mu$, plug back in:

$$E(W) = \sum_{i=1}^{n} (a_i \mu) = \mu \sum_{i=1}^{n} a_i = \mu$$

thus the restriction on a_i is $\sum_{i=1}^{n} a_i = 1$

2. Find $Var(W)$ From $Var(W) = E(W^2) - E(W)^2$, where

$$E(W^2) = E((\sum_{i=1}^{n} a_i Y_i)^2) = E(\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j Y_i Y_j) = E(\sum_{i=1}^{n} a_i^2 Y_i^2) + E(\sum_{i \neq j} a_i a_j Y_i Y_j)$$

thus

$$E(W^2) = \sum_i a_i^2 E(Y_i^2) + \sum_{i \neq j} a_i a_j E(Y_i Y_j)$$

where $E(Y_i^2) = Var(Y_i) + E(Y_i)^2 = \sigma^2 + \mu^2$, and because any two of Y_i are uncorrelated, we have $E(Y_i Y_j) = E(Y_i)E(Y_j) = \mu^2$, hence

$$E(W^2) = (\sigma^2 + \mu^2) \sum_i a_i^2 + \mu^2 \sum_{i \neq j} a_i a_j = \sigma^2 \sum_i a_i^2 + \mu^2 (\sum_i a_i^2 + \sum_{i \neq j} a_i a_j) = \sigma^2 \sum_i a_i^2 + \mu^2 (\sum_i a_i)^2$$

then we have $E(W)^2 = (\sum_i a_i E(Y_i))^2 = \mu^2 (\sum_i a_i)^2$, then the variance of W is:

$$Var(W) = E(W^2) - E(W)^2 = \sigma^2 \sum_{i=1}^{n} a_i^2$$

3. Given a set of numbers $a1, a2, ...a_n$, the following inequality holds:

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

9

Use this inequality, along with the previous parts of this question, to show that $Var(W) \geq Var(\bar{Y})$ whenever W is unbiased. We say that $\bar{Y}$ is the best linear unbiased estimator (BLUE). From variance definition and property we have $Var(\bar{Y}) = Var(\frac{1}{n} \sum_i Y_i) = \frac{1}{n^2} Var(\sum_i Y_i)$, in which

$$Var(\sum_i Y_i) = \sum i, j = 1^n Cov(Y_i, Y_j) = \sum_i Var(Y_i) + \sum_{i \neq i} Cov(Y_i, Y_j)$$

because any pair of Y_i is uncorrelated, we have $Cov(Y_i, Y_j) = 0 \quad \forall i \neq j$, then $Var(\sum_i Y_i) = \sum_i Var(Y_i) = n\sigma^2$, thus $Var(\bar{Y}) = \frac{1}{n^2} \times n \times \sigma^2 = \frac{1}{n}\sigma^2$
from part 1, when W is unbiased, we have $\sum_i a_i = 1$, then $Var(\bar{Y})$ can be written as $\frac{1}{n}(\sum_i a_i)^2$, because $\sigma^2 \geq 0$, according to the given Cauchy inequality, we have $\sigma^2 \frac{1}{n}(\sum_i a_i)^2 \leq \sigma^2 \sum_i a_i^2$, then

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

whenever W is unbiased, and $\bar{Y}$ is BLUE!

1.3.2 Question 9 (10 Points)

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: $W1 = (\frac{n-1}{n})\bar{Y}$ and $W2 = k\bar{Y}$, where $0 < k < 1$.

1. Compute the biases of both W_1 and W_2. Which estimator is consistent? The bias B of an estimator e for statistic s is defined as the difference between the expectation of the estimator $E(e)$ and the true value s from the population, thus

$$B(W_1) = E(W_1) - \mu,$$

where

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

, because the sample average $\bar{Y}$ is the unbiased estimation population μ,
thus $B(W_1) = (\frac{n-1}{n})\mu - \mu = -\frac{\mu}{n}$.
Similarly $B(W_2) = E(W_2) - \mu$, where
$E(W_2) = E(k\bar{Y}) = kE(\bar{Y}) = k\mu$, then

$$B(W_2) = k\mu - \mu.$$

As n increases, $B(W_1)$ will decreases to 0, so W_1 is consistent, while W_2 is not.

2. Compute $Var(W_1)$ and $Var(W_2)$. Which estimator has lower variance? In general $Var(X) = E(X^2) - E(X)^2$, so we need $E(W_1^2)$ and $E(W_2^2)$:

$$E(W_1^2) = E((\frac{n-1}{n})^2 \bar{Y}^2) = (\frac{n-1}{n})^2 E(\bar{Y}^2)$$

$$E(W_2^2) = E(k^2 \bar{Y}^2) = k^2 E(\bar{Y}^2)$$

where $E(\bar{Y}^2) = Var(\bar{Y}) + E(\bar{Y})^2$,
then $Var(\bar{Y}) = Var(\frac{1}{n} \sum_i X_i) = \frac{1}{n^2} Var(\sum_i X_i)$.
Because X_i are independent draws, $Var(\sum_i X_i) = \sum_i Var(X_i) = n\sigma^2$, then $Var(\bar{Y}) = \frac{1}{n^2} \times n \times \sigma^2 = \frac{\sigma^2}{n}$
And we can obtain $E(\bar{Y}^2) = \frac{\sigma^2}{n} + \mu^2$, plug it back to get $E(W_1^2)$ and $E(W_2^2)$:

$$E(W_1^2) = (\frac{n-1}{n})^2 \times (\frac{\sigma^2}{n} + \mu^2)$$

$$E(W_2^2) = k^2 \times (\frac{\sigma^2}{n} + \mu^2)$$

Finally we have $Var(W_1)$ and $Var(W_2)$:

$$Var(W_1) = E(W_1^2) - E(W_1)^2 = (\frac{n-1}{n})^2 \times (\frac{\sigma^2}{n} + \mu^2) - (\frac{n-1}{n})^2 \times \mu^2 = \frac{(n-1)^2}{n^3}\sigma^2$$

$$Var(W_2) = E(W_2^2) - E(W_2)^2 = k^2 \times (\frac{\sigma^2}{n} + \mu^2) - (k\mu)^2 = \frac{k^2}{n}\sigma^2$$

To compare the two variations, we compare $\frac{(n-1)^2}{n^3}$ and $\frac{(kn)^2}{n^3}$, thus if $(n-1) > nk$, $Var(W_1) > Var(W_2)$, otherwise $Var(W_1) \leq Var(W_2)$

1.3.3 Question 10 (10 Points)

Given a random sample $Y_1, Y_2, ..., Y_n$ from some distribution $F(.)$ 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:

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

1. **Show that** $E(\bar{Y}) = E(Y_i) \forall i \in 1, 2, ..., n$ $E(\bar{Y}) = E(\frac{1}{n}\sum_{i=1}^{n} Y_i) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i)$, considering $\forall i$ $E(Y_i) = E(Y)$, we have
$E(\bar{Y}) = \frac{1}{n}\sum_{i=1}^{n} E(Y) = \frac{1}{n} \times n \times E(Y) = E(Y_i)$

2. Show that $Var(\bar{Y}) = \frac{1}{n}Var(Y_i)$ $\forall i \in 1, 2, ..., n$, $Var(\bar{Y}) = Var(\frac{1}{n}\sum_i Y_i) = \frac{1}{n^2}Var(\sum_i Y_i)$ where
$Var(\sum_i Y_i) = \sum_{i=1}^{n} Var(Y_i) + \sum_{i \neq j} Cov(Y_i, Y_j)$
because the sample is randomly drawn, applying *i.i.d* property, we have $Cov(Y_i, Y_j) = 0$
and since $\forall i$ $Var(Y_i) = Var(Y)$
thus $Var(\bar{Y}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i) = \frac{1}{n^2} \times n \times Var(Y) = \frac{1}{n}Var(Y_i)$

3. Compute the expectation of $\hat{\sigma^2}$ 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. Let's massage $\hat{\sigma^2}$ first:

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}((Y_i - \mu) - (\bar{Y} - \mu))^2 = \frac{1}{n}\sum_i((Y_i - \mu)^2 + (\bar{Y} - \mu)^2 - 2(Y_i - \mu)(\bar{Y} - \mu))$$

expand the summation, we have:

$$\hat{\sigma^2} = \frac{1}{n}\sum_i(Y_i - \mu)^2 + (\bar{Y} - \mu)^2 - \frac{1}{n}\sum_i(2(Y_i - \mu)(\bar{Y} - \mu))$$

the first summation term is the definition of σ^2, we further expand the last summation:

$$\hat{\sigma^2} = \sigma^2 + (\bar{Y} - \mu)^2 - \frac{2(\bar{Y} - \mu)}{n}\sum_i(Y_i - \mu)$$

with $\sum_i Y_i = n\bar{Y}$:

$$\hat{\sigma^2} = \sigma^2 + (\bar{Y} - \mu)^2 - \frac{2(\bar{Y} - \mu)}{n}(n\bar{Y} - n\mu) = \sigma^2 - (\bar{Y} - \mu)^2$$

now take expectation:

$$E(\hat{\sigma^2}) = E(\sigma^2) - E((\bar{Y} - \mu)^2) = \sigma^2 - E(\bar{Y}^2) - \mu^2 + 2\mu E(\bar{Y})$$

from (1): $E(\bar{Y}) = E(Y_i) = \mu$, we have:

11

$$E(\hat{\sigma^2}) = \sigma^2 + \mu^2 - E(\bar{Y}^2)$$

where $E(\bar{Y}^2) = Var(\bar{Y}) + E(\bar{Y})^2$, from (2), we have:

$$E(\bar{Y}^2) = \frac{\sigma^2}{n} + \mu^2$$

plug back in, the expectation of $\hat{\sigma^2}$:

$$\hat{\sigma^2} = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2$$

4. Is this an unbiased estimator for σ^2? Obviously **it is a biased estimator**, as we see its expectation is a fraction $\frac{n-1}{n}$ of the truth σ^2

5. If not, what function of $\hat{\sigma^2}$ produce an unbiased estimator? To become unbiased, we simply multiply $\frac{n}{n-1}$ on $\hat{\sigma^2}$, and the unbiased estimator is:

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

In [ ]:

1.3.4 Question 11 (5 Points)

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.

1. Define the random variable $Z = Y/X$. Show that $E(Z) = \theta$. Evaluate $E(Z|X) = E(Y/X|X) = E(Y|X)/X = (\theta X)/X = \theta$, thus

$$E(Z) = E(E(Z|X)) = E(\theta) = \theta$$

2. Use part (1) to prove that the estimator $W_1 = \frac{1}{n}\sum_{i=1}^{n}(Y_i/X_i)$ is unbiased for θ, where $(X_i, Y_i) : i = 1, 2, ..., n$ is a random sample.
    Evaluate $E(W_1) = \frac{1}{n}\sum_{i=1}^{n} E((Y_i/X_i))$, from (1) we have

$$E(W_1) = \frac{1}{n}\sum_{i=1}^{n} E(Z_i) = \frac{1}{n} \times n \times \theta = \theta$$

3. Explain why the estimator $W2 = \bar{Y}/\bar{X}$, where the overbars denote sample averages, is not the same as W_1. Nevertheless, show that W_2 is also unbiased for θ.
    Expand $W_2 = \frac{n^{-1}\sum_i Y_i}{n^{-1}\sum_i X_i} = \frac{\sum_i Y_i}{\sum_i X_i}$, the summation can't be further simplified, obviously $W_1 \neq W_2$.
    Evaluate $E(W_2) = E(\frac{\sum_i Y_i}{\sum_i X_i})$, apply expectaion within the summation, we have:

$$E(W_2) = \frac{\sum_i E(Y_i)}{\sum_i E(X_i)} = \frac{\sum_i E(Y)}{\sum_i E(X)} = \frac{E(Y)}{E(X)} = \frac{E(Y|X)}{E(X|X)} = \frac{\theta X}{X} = \theta$$

1.3.5 Question 12 (10 Points)

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, σ^2) distribution.

1. The null hypothesis is that there was no change in average liquor consumption. State this formally in terms of μ.   Null hypothesis: $H_0 : \mu = 0$

2. The alternative is that there was a decline in liquor consumption; state the alternative in terms of μ.   Alternative hypothesis: $H_1 : \mu < 0$

3. 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. Do you reject H_0 at the 5% level? At the 1% level?   The t statistic:

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{-32.8 - 0}{466.4/\sqrt{900}} = -2.11$$

p-value:

$$p = 0.0176$$

Reject H_0 at the 5% level, and fail to reject H_0 at the 1% level.

4. Would you say that the estimated fall in consumption is large in magnitude? Comment on the practical versus statistical significance of this estimate.   For practical significance since we are using a paired-sample t-test, we evaluate the Cohen's d effect size:

$$d = \frac{\bar{y} - \mu}{s} = 0.070$$

thus there is **little practical significance** from the test.

5. What has been implicitly assumed in your analysis about other determinants of liquor consumption over the two-year period in order to infer causality from the tax change to liquor consumption?   It is implicitly assumed that before and after the tax law is implemented, all other determinants that can affect liquor consumption remain the same, thus the difference is only attribted to tax change.

1.3.6 Question 13 (10 Points)

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

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	416-519

Player	FGA-FGM
Larry Bird	1206-455
Jon Sundvold	440-166
Brian Taylor	417-157

For a given player, the outcome of a particular shot can be modelled as a Bernoulli (zero-one) 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 θ is $\bar{Y} = FGM/FGA$.

1. Estimate θ for Mark Price.

$$\theta_{MP} = \frac{188}{429} = 0.4382$$

2. Find the standard deviation of the estimator $\bar{Y}$ in terms of θ and the number of shot attempts, n. For Bernoulli variable Y_i, the $\bar{Y} = \frac{\sum_i Y_i}{n}$, thus:

$$Var(\bar{Y}) = Var(\frac{\sum_i Y_i}{n}) = \frac{1}{n^2} Var(\sum_i Y_i)$$

because each shot is independent from each other, we have: $Var(\sum_i Y_i) = \sum_i Var(Y_i)$. And for Bernoulli Y_i, $Var(Y_i) = \theta(1-\theta)$, thus:

$$Var(\bar{Y}) = \frac{1}{n^2} \times n \times \theta(1-\theta) = \frac{\theta(1-\theta)}{n}$$

then:

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

3. The asymptotic distribution of $(\bar{Y}-\theta)/se(\bar{Y})$ is standard normal, where $se(\bar{Y}) = \sqrt{\bar{Y}(1-\bar{Y})/n}$. Use this fact to test $H_0 : \theta = .5$ against $H_1 : \theta < .5$ for Mark Price. Use a 1% significance level. For Mark Price: $se(\bar{Y}) = \sqrt{\frac{0.4382(1-0.4382)}{429}} = 0.0240$, and $\frac{\bar{Y}-\theta}{se(\bar{Y})} = \frac{0.4382-0.5}{0.0240} = -2.5786$, p-value $= 0.005$, thus we reject H_0 at 1% level

4. Define Type I error. Type I error is the probability that H_0 is falsely rejected given it's true, $P(\text{reject } H_0|H_0)$.

5. What is the probability of Type I error of this test? At 1% leve, the Type I error is $\alpha = 1\%$

6. Define Type II error. Type II error is the probability of failing to reject H_0 given that H_1 is true, namely $P(\text{fail to reject } H_0|H_1)$

7. What is the probability of Type II error when using this decision rule, assuming the "true" population proportion is $\theta^* = 0.45$. For 1% level to reject H_0, the decision rule: $\frac{\bar{Y}-0.5}{se(\bar{Y})} < Z_{0.01} = -2.3263$, or $\bar{Y} < Z_{0.01} * se(\bar{Y}) + 0.5 = 0.4442$
with true $\theta^* = 0.45$:

$$\beta = P(\bar{Y} \leq 0.4442|\theta = 0.45) = P(Z \leq \frac{0.4442 - 0.45}{se(\bar{Y})}) = 0.4040$$

8. Define the power of the test (in general terms). Power of a test is the probability that H_0 is rejected given that H_1 is true.

9. Calculate the power of this test, again assuming the "true" population proportion is $\theta^* = 0.45$. The **power of the test** (given $\theta^* = 0.45$) is: $1 - \beta = 0.5960$

In [ ]: