Interesting Probability Problems

Jonathan Mostovoy - 1002142665
University of Toronto

August 19, 2016

Contents

1 Chapter 1 Questions

1a) 1.8.17

A deck of 52 cards contains four aces. If the cards are shuffled and distributed in a random manner to four players so that each player receives 13 cards, what is the probability that all four aces will be received by the same player?

Answer: Since $\exists$ agents $i = 1, \dots, 4$ the probability that agent i will receive 4 aces is $\binom{13}{4}$. Since $\exists$ 4 agents, it $\implies$ $\exists$ $4\binom{13}{4}$ scenarios where one agent receives all 4 aces. Therefore, and since $\exists$ there are $\binom{52}{4}$ ways of arranging the four aces in a deck of 52 cards, the probability that all four aces will be received by the same player is:

$$Pr(\text{4 aces recieved by same agent}) = \frac{4\binom{13}{4}}{\binom{52}{4}} = 4\frac{\binom{48}{13,13,13,9}}{\binom{52}{13,13,13,13}} = \frac{44}{4165} \approx 0.0105642$$

1b) 1.12.12

Let $A_1, \dots, A_n$ be n arbitrary events. Show that the probability that exactly one of these n events will occur is:

$$\sum_{i=1}^{n} Pr(A_i) - 2\sum_{i<j}^{n} Pr(A_i \cap A_j) + 3\sum_{i<j<k}^{n} Pr(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} n Pr(A_1 \cap A_2 \cap \dots \cap A_n)$$

Proof.

$$\text{Let } \mathscr{P}(n) = \bigcup_{i=1}^{n} Pr(A_i \backslash (A_1, A_2, \dots, A_{i-1}, A_{i+1}, \dots, A_n)), \text{ which we shall prove:}$$

$$= \sum_{i=1}^{n} Pr(A_i) - 2\sum_{i<j}^{n} Pr(A_i \cap A_j) + 3\sum_{i<j<k}^{n} Pr(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} n Pr(A_1 \cap A_2 \cap \dots \cap A_n)$$

Our plan is to form induction on the function $\mathscr{P}$(n). We begin at $n = 2$. We define an alternate definition of the set A and B as $A = \{x : x \in A \backslash B \text{ or } A \cap B\}$ and $B = \{y : y \in B \backslash A \text{ or } A \cap B\}$. Both sets for which all x or y in A or B are trivially disjoint to one other since $(A \backslash B) \cap (A \cap B) = \emptyset$. Thus, $Pr(A) = Pr(A \backslash B) + Pr(A \cap B)$. From here, we find $Pr(A) + Pr(B) = Pr(A \backslash B) + Pr(B \backslash A) + 2Pr(A \cap B)$. Now since all three sets are disjoint, we can replace our additions to unions to yield:

$$Pr(A) + Pr(B) - 2Pr(A \cap B) = \mathscr{P}(2) = Pr(A \backslash B) \cup Pr(B \backslash A)$$

We now have our inductive hypothesis assuming validity of our Theorem for $\mathscr{P}(n)$, thus for $k = n+1$:

$$\mathscr{P}(n+1) = \bigcup_{i=1}^{n+1} Pr(A_i \backslash (A_1, A_2, \dots, A_{i-1}, A_{i+1}, \dots, A_n, A_{n+1}))$$

But by our inductive hypothesis, we know what $\mathscr{P}(n)$ is, thus all we need to do is add on the remove all intersections of A_{n+1} and add back it's own probability (in a sense, take away the overlaps that A_{n+1} could have on everything else) $\forall$ terms in our $\mathscr{P}(n)$ formula, i.e.:

$$\mathscr{P}(n+1) = \sum_{i=1}^{n} Pr(A_i) + Pr(A_{n+1}) - 2\sum_{i<j}^{n} Pr(A_i \cap A_j) - 2\sum_{i=1}^{n} Pr(A_i \cap A_{n+1}) + \dots$$

$$+(-1)^{n+1}nPr(A_1 \cap A_2 \cap \cdots \cap A_n) + (-1)^{n+1}nPr(A_2 \cap A_3 \cap \cdots \cap A_{n+1} + \ldots$$
$$+(n+1)(-1)^{n+2}Pr(A_1 \cap \cdots \cap A_{n+1})$$

$$= \sum_{i=1}^{n+1} Pr(A_i) - 2\sum_{i<j}^{n+1} Pr(A_i \cap A_j) + 3\sum_{i<j<k}^{n+1} Pr(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+2}(n+1)Pr(A_1 \cap A_2 \cap \cdots \cap A_{n+1})$$

□

2 Chapter 2 Questions

2a) 2.1.14

A machine produces defective parts with three different probabilities depending on its state of repair. If the machine is in good working order, it produces defective parts with probability 0.02. If it is wearing down, it produces defective parts with probability 0.1. If it needs maintenance, it produces defective parts with probability 0.3. The probability that the machine is in good working order is 0.8, the probability that it is wearing down is 0.1, and the probability that it needs maintenance is 0.1. Compute the probability that a randomly selected part will be defective.

Answer: We first define the notation: g, w, b, d as "good working order", "wearing down", "needs maintenance (bad)" and "production of defective parts" respectively. We summarize the information given as: $Pr(d|g) = .2,\ Pr(d|w) = .1,\ Pr(d|b) = .3,\ Pr(g) = .8,\ Pr(w) = .1 = Pr(b)$. We now recall the "Law of Total Probability", which states: given events $B_1, \ldots, B_k$ form a partition of the space S and $Pr(B_j) > 0$ for $j = 1, \ldots, k$. Then, for every event A in S:

$$Pr(A) = \sum_{j=1}^{k} Pr(B_j)Pr(A|B_j)$$

Since g, w & b must be independent and sum to 1, our answer is a direct application of the Law of Total Probability:

$$\mathbf{Pr(d)} = \sum_i Pr(i)Pr(d|i),\ (j = g, w, b) = (.8)(.2) + (.1)(.1) + (.1)(.3) = \mathbf{.2}$$

2b) 2.2.20

Suppose that $A_1, \ldots, A_k$ form a sequence of k independent events. Let $B_1, \ldots, B_k$ be another sequence of k events such that for each value of j, $(j = 1, \ldots, k)$, either $B_j = A_j$ or $B_j = A_j^c$. Prove that $B_1, \ldots, B_k$ are also independent events. Hint: Use an induction argument based on the number of events B_j for which $B_j = A_j^c$.

Proof. We let $n, n \leq k$ denote the number of B_j's where the relation $B_j = A_j^c$ holds, which $\implies \exists$ $k - n$ cases where $B_j = A_j$. For $n = 0$, our desired relation trivially holds because we assume $A_1, \ldots, A_k$ is independent. For higher order cases, we recall the identity of $Pr(A \cap B) = Pr(A) - Pr(A \cap B^c)$ and the fact that if $A_1, \ldots, A_l$ is independent, then so too will $A_1, \ldots, A_v$ where $v \leq l$.

Thus, through induction, we assume $Pr(B_1 \cap \cdots \cap B_k)$ are independent, and n of these B_j's satisfy $B_j = A_j^c$. Therefore, for $s = n+1$, we want to see if we can factor $Pr(B_1 \cap \cdots \cap B_{i-1}, B_i^c \cap B_{i+1} \cap \cdots \cap B_k)$ where it was B_i that now became B_i^c from n to $n+1$. We see this relationship to be true from:

$Pr(B_1 \cap \cdots \cap B_{i-1} \cap B_{i+1} \cap \cdots \cap B_k \cap B_i^c)$
$= Pr(B_1 \cap \cdots \cap B_{i-1} \cap B_{i+1} \cap \cdots \cap B_k) - Pr(B_1 \cap \cdots \cap B_{i-1} \cap B_{i+1} \cap \cdots \cap B_k \cap B_i)$
$= Pr(B_1) \times \cdots \times Pr(B_{i-1}) \times Pr(B_{i+1}) \times \cdots \times Pr(B_k) - Pr(B_1 \cap \cdots \cap B_{i-1} \cap B_{i+1} \cap \cdots \cap B_k \cap B_i)$
$= Pr(B_1) \times \cdots \times Pr(B_{i-1}) \times Pr(B_{i+1}) \times \cdots \times Pr(B_k) - Pr(B_1) \times \cdots \times Pr(B_{i-1}) \times Pr(B_{i+1}) \times \cdots \times Pr(B_k) \times Pr(B_i)$
$= Pr(B_1) \times \cdots \times Pr(B_k) \times (1 - Pr(B_i))$
$= Pr(B_1) \times \cdots \times Pr(B_{i-1}) \times Pr(B_i^c) \times Pr(B_{i+1}) \times \cdots \times Pr(B_k)$ □

3 Chapter 3 Questions

3a) 3.2.13

An ice cream seller takes 20 gallons of ice cream in her truck each day. Let X stand for the number of gallons that she sells. The probability is 0.1 that $X = 20$. If she doesn't sell all 20 gallons, the distribution of X follows a continuous distribution with a p.d.f. of the form:

$$f(x) = \begin{cases} cx & \textit{for } 0 < x < 20 \\ 0 & \textit{otherwise} \end{cases}$$

where c is a constant that makes $Pr(X < 20) = 0.9$. Find the constant c so that $Pr(X < 20) = 0.9$ as described above.

Answer: All we need to remember that the integral of our p.d.f over our sample space must be equal to the the total probability of occurrence, which in this case is equal to .9, but usually 1. Thus:

$$\int_0^{20} cx = .9 \implies 200c = .9 \implies c = \frac{9}{2000} = 0.0045$$

3b) 3.3.17

Prove that the quantile function F^{-1} of a general random variable X has the following three properties that are analogous to properties of the c.d.f.:

1. *F^{-1} is a non-decreasing function of p for $0 < p < 1$.*
2. *Let $x_0 = \lim_{p \to 0,\ p>0} F^{-1}(p)$ and $x_1 = \lim_{p \to 1,\ p<1} F^{-1}(p)$. Then x_0 equals the greatest lower bound on the set of numbers c such that $Pr(X \le c) > 0$, and x_1 equals the least upper bound on the set of numbers d such that $Pr(X \ge d) > 0$.*
3. *F^{-1} is continuous from the left; that is $F^{-1}(p) = F^{-1}(p-)$ for all $0 < p < 1$.*

Proofs:

1. *Proof.* Assume $p, q \in (0, 1)$ and $p \leq q$. Then, by definition of the quantile function and since $F(x)$ is non-decreasing, F^{-1} is non-decreasing since $\forall p, q$:

$$\left(F^{-1}(p) = \min\{x \in \mathbb{R} : F(x) \geq p\}\right) \subseteq \left(F^{-1}(q) = \min\{x \in \mathbb{R} : F(x) \geq q\}\right)$$

□

2. *Proof.* x_0: Let $z_1 > z_2 > z_3 > \dots$ be a decreasing sequence of numbers such that $\lim_{n\to\infty} z_n = 0$.

$$\implies x_0 = \lim_{p\to 0,\ p>0} F^{-1}(p) = \bigcap_{n=1}^{\infty} \min\{x \in \mathbb{R} : F(x) \geq z_n\} = \min\{x \in \mathbb{R} : F(x) \geq c\}$$

for some $c \in \mathbb{R}$ where no other $x \in \mathbb{R}$ is less than c and $F(x) > 0$, which is equivalent to saying c is the least upper bound of all possible numbers where $F(c) > 0 \equiv$ the condition $Pr(X \leq c) > 0$

x_1: Let $y_1 < y_2 < y_3 < \dots$ be an increasing sequence of numbers such that $\lim_{n\to\infty} y_n = 1$

$$\implies x_1 = \lim_{p\to 1,\ p<1} F^{-1}(p) = \bigcup_{n=1}^{\infty} \min\{x \in \mathbb{R} : F(x) \geq y_n\} = \min\{x \in \mathbb{R} : F(x) \geq d\}$$

for some $d \in \mathbb{R}$ where no other $x \in \mathbb{R}$ is greater less than d and $F(x) < 1$, which is equivalent to saying d is the least upper bound of all possible numbers where $F(d) < 1 \equiv$ the condition $Pr(X \geq d) > 0$ □

3. *Proof.* Let $y_1 < y_2 < y_3 < \dots$ be an increasing sequence of numbers such that $\lim_{n\to\infty} y_n = p$. We immediately see that:

$$F^{-1}(p) = \min\{x \in \mathbb{R} : F(x) \leq p\} = \bigcup_{n=1}^{\infty} \min\{x \in \mathbb{R} : F(x) \leq y_n\}$$

$$\implies F^{-1}(p) = \lim_{n\to\infty} F^{-1}(y_n) = F^{-1}(p-)$$

□

3c) 3.4.4

Suppose that X and Y have a continuous joint distribution for which the joint p.d.f. is defined as follows:

$$f(x, y) = \begin{cases} cy^2 & \text{for } 0 \leq x \leq 2 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Determine:

1. *the value of the constant c*
2. $Pr(X + Y > 2)$
3. $Pr(Y < 1/2)$

4. $Pr(X \leq 1)$
5. $Pr(X = 3Y)$

Answers:

1. $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} cy^2 \; dxdy = 1 = \int_0^1\int_0^2 cy^2 \; dxdy = \int_0^1 2cy^2 \; dy = \frac{2c}{3} = 1 \implies c = \frac{3}{2}$$

2. $$Pr(X+Y>2) = \int_1^2\int_{2-x}^1 \frac{3y^2}{2} \; dxdy = \frac{3}{8}$$

3. $$Pr(Y<\frac{1}{2}) = \int_0^{\frac{1}{2}}\int_0^2 \frac{3y^2}{2} \; dxdy = \frac{1}{8}$$

4. $$Pr(X \leq 1) = \int_0^1\int_0^1 \frac{3y^2}{2} \; dxdy = \frac{1}{2}$$

5. $Pr(X = 3Y) = 0$ since with an n-dimensional continuous probability space, the probability that an $(n-1)$-dimensional event occurs is effectively 0; i.e., since $X = 3Y$ is a line in a 2-dimensional continuous probability space, the probability this happens must be 0.

3d) 3.5.8

Suppose that the joint p.d.f. of X and Y is as follows:

$$f(x,y) = \begin{cases} 24xy & \text{for } x \geq 0,\; y \geq 0 \text{ and } x+y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Are X and Y independent?

Answer: We recall Theorem 3.5.5, which states: when dealing with a joint p.d.f. $f(x,y)$, the random variables, X and Y, will be independent $\iff f(x,y) = h_1(x)h_2(y)$ where $h_i(z)$ is a non-negative function only dependent on z. For every point within the defined triangle, we can define two functions which work, e.g;

$$h_1(x) = \begin{cases} kx & \text{for } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

where $k \in \mathbb{R}$, and

$$h_2(y) = \begin{cases} \frac{24}{k}y & \text{for } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

However, if we choose a point within the the unit square outside of our triangle, we need $f(x,y) = 0$, but $h_1(x) > 0$ and $h_2(y) > 0$ which thus leads to a contradiction which $\implies$ X and Y are NOT independent. Thus, we can conclude the following generalization: If $\exists\, x_i$ within $f(x_1, x_2, \dots)$ whose domain is a function of at least 1 other variable, x_j where $j \neq i$, then $\exists$ some dependency amongst the variables $x_1, x_2, \dots$.

3e) 3.9.18

Let the conditional p.d.f. of X given Y be $g_1(x|y) = \frac{3x^2}{y^3}$ for $0 < x < y$ and 0 otherwise. Let the marginal p.d.f. of Y be $f_2(y)$, where $f_2(y) = 0$ for $y \leq 0$ but is otherwise unspecified. Let $Z = \frac{X}{Y}$. Prove that Z and Y are independent and find the marginal p.d.f. of Z.

Proof. By the definition of conditional probability, the joint p.d.f. of (X, Y) is

$$f(x, y) = \begin{cases} \frac{3x^2 f_2(y)}{y^3} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

We recall that if $Z = \frac{X}{Y}$, we can define the dummy second random variable $W = Y$ so that the Jacobian of our inverse transformation of $x = zw$ and $y = w$ is:

$$J = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} w & z \\ 0 & 1 \end{vmatrix}$$

Thus,

$$g(z, w) = f(x, y)w = f(zw, w)w = \frac{3(zw)^2 f_2(w)w}{w^3} = 3z^2 f_2(w)$$

Where the bounds on our variables were previously established in the question. Thus, since $g(z, w) = f_1(z)f_2(w)$ we may conclude independence. Further, the marginal p.d.f. of Z will be:

$$f_1(z) = \begin{cases} 3z^2 & \text{if } z \in (0, 1) \\ 0 & \text{otherwise} \end{cases}$$

Since $\int_{-\infty}^{\infty} f_2(w)dw = 1$ if $f_2(w)$ is a proper probability function. □

3f) 3.11.26

Let X_1, X_2 be two independent random variables each with p.d.f. $f_1(x) = e^{-x}$ for $x > 0$ and $f_1(x) = 0$ for $x \leq 0$. Let $Z = X_1 - X_2$ and $W = \frac{X_1}{X_2}$.

1. *Find the joint p.d.f. of X_1 and Z.*
2. *Prove that the conditional p.d.f. of X_1 given $Z = 0$ is:*

$$h_1(x_1|0) = \begin{cases} 2e^{-2x_1} & \textit{for } x_1 > 0 \\ 0 & \textit{otherwise} \end{cases}$$

3. *Find the joint p.d.f. of X_1 and W.*
4. *Prove that the conditional p.d.f. of X_1 given $W = 1$ is:*

$$h_1(x_1|1) = \begin{cases} 4x_1 e^{-2x_1} & \textit{for } x_1 > 0 \\ 0 & \textit{otherwise} \end{cases}$$

5. *Notice that $\{Z = 0\} = \{W = 1\}$, but the conditional distribution of X_1 given $Z = 0$ is not the same as the conditional distribution of X_1 given $W = 1$. This discrepancy is known as the Borel paradox. In light of the discussion that begins on page 146 about how conditional p.d.f.'s are not like conditioning on events of probability 0, show how "Z very close to 0" is not the same as "W very close to 1." Hint: Draw a set of axes for x_1 and x_2, and draw the two sets $\{(x_1, x_2) : |x_1 - x_2| < \epsilon\}$ and $\{(x_1, x_2) : |\frac{x_1}{x_2} - 1| < \epsilon\}$ and see how different they are.*

1. *Answer:* First, let us define the Jacobian from the two equations $X_1 = V$ and $X_2 = V - Z$:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial v} & \frac{\partial x_1}{\partial z} \\ \frac{\partial x_2}{\partial v} & \frac{\partial x_2}{\partial z} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & -1 \end{vmatrix}$$

$\Longrightarrow$

$$g(v, z) = f(x_1, x_2)|J| = |-1|e^{-v}e^{-(v-z)} = e^{-2v}e^{z} \equiv e^{-2x_1}e^{z}$$

2. *Proof.* By definition, $h_1(x_1|0) = \frac{g(x_1,0)}{g_1(0)}$. Thus, we first find $g_1(z)$:

$$g_1(z) = \int_{\max(0,z)}^{\infty} g(x_1, z)dx_1 = \left[-\frac{1}{2}e^{-2x_1}e^{z} \right]_{x_1=\max(0,z)}^{x_1=\infty} = \begin{cases} \frac{1}{2}e^{-z} & \text{if } z \geq 0 \\ \frac{1}{2}e^{z} & \text{if } z < 0 \end{cases}$$

$\Longrightarrow$

$$h_1(x_1|0) = \frac{e^{-2x_1}e^{0}}{\frac{1}{2}e^{0}} = 2e^{-2x_1} \text{ and } 0 \text{ if } x_1 \leq 0$$

□

3. *Answer:* First, let us define the Jacobian from the two equations $X_1 = V$ and $X_2 = \frac{V}{W}$:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial v} & \frac{\partial x_1}{\partial w} \\ \frac{\partial x_2}{\partial v} & \frac{\partial x_2}{\partial w} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ \frac{1}{w} & -\frac{v}{w^2} \end{vmatrix}$$

$\Longrightarrow$

$$g(v, w) = f(x_1, x_2)|J| = |-\frac{v}{w^2}|e^{-v}e^{-(\frac{v}{w})} = \frac{v}{w^2}e^{-v(1+\frac{1}{w})} \equiv \frac{x_1}{w^2}e^{-x_1(1+\frac{1}{w})} \text{ for } v/x_1, w > 0$$

4. *Proof.* By definition, $h_1(x_1|1) = \frac{g(x_1,1)}{g_1(1)}$. Thus, we first find $g_1(w)$:

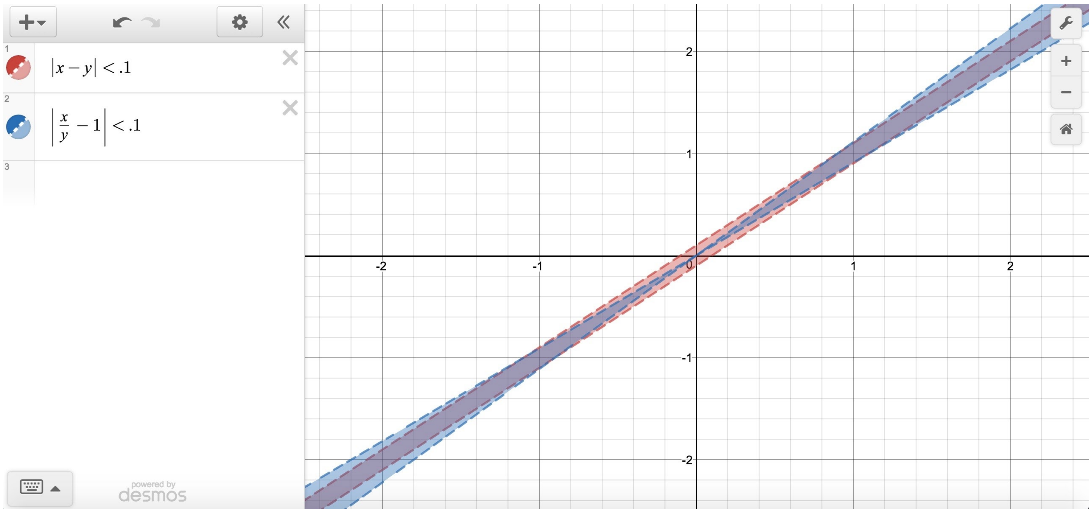
$$g_1(w) = \int_0^{\infty} g(x_1, w)dx_1 = \left[-\frac{1 + x_1}{w^2(1 + \frac{1}{w})^2}e^{-x_1(1+\frac{1}{w})} \right]_{x_1=0}^{x_1=\infty} = \frac{1}{w^2(1 + \frac{1}{w})^2}$$

$\Longrightarrow$

$$h_1(x_1|1) = \left. \frac{\frac{x_1}{w^2}e^{-x_1(1+\frac{1}{w})}}{\frac{1}{w^2(1+\frac{1}{w})^2}} \right|_{w=1} = 4x_1e^{-2x_1} \text{ and } 0 \text{ if } x_1 \leq 0$$

□

5. The difference between $\{(x_1, x_2) : |x_1 - x_2| < \epsilon\}$ and $\{(x_1, x_2) : |\frac{x_1}{x_2} - 1| < \epsilon\}$ can be seen below, which $\Longrightarrow$ $h_1(x_1|w) \neq h_1(x_1|z)$.



4 Chapter 4 Questions

4a) 4.4.9

Let X be a random variable with mean μ and variance σ^2, and let $\psi_1(t)$ denote the m.g.f. of X for $-\infty < t < \infty$. Let c be a given positive constant, and let Y be a random variable for which the m.g.f. is:

$$\psi_2(t) = e^{c(\psi_1(t)-1)} \text{ for } -\infty < t < \infty$$

Find expressions for the mean and the variance of Y in terms of the mean and the variance of X.

Answer: We first summarize: $\psi_1(0) = 1$, $\psi_1'(0) = \mu$ and $\psi_1''(0) = \sigma^2 + \mu^2$. We compute:

$$\left.\frac{d\psi_2(t)}{dt}\right|_{t=0} = c\psi_1'(0)e^{c(\psi_1(0)-1)} = c\mu$$

.

$$\left.\frac{d^2\psi_2(t)}{dt^2}\right|_{t=0} = c\psi_1''(0)e^{c(\psi_1(0)-1)} + c^2(\psi_1'(0))^2e^{c(\psi_1(0)-1)} = c(\sigma^2+\mu^2) + c^2\mu^2$$

Therefore, Mean $= c\mu$ and Variance $= c(\sigma^2+\mu^2) + c^2\mu^2 - (c\mu)^2 = c(\sigma^2+\mu^2)$.

4b) 4.7.12

Suppose that X and Y are random variables such that $E(Y|X) = aX + b$. Assuming that $Cov(X, Y)$ exists and that $0 < Var(X) < \infty$, determine expressions for a and b in terms of $E(X)$, $E(Y)$, $Var(X)$, and $Cov(X, Y)$.

Answer: We recall $E(E(X_1|X_2)) = E(X_1) \implies E(E(Y|X)) = E(Y) = E(aX+b) = aE(X)+b$. Thus, we have our first equation: $E(Y) = aE(X) + b$.

Next, we apply $E(Xf(X, Y))$ to both sides of the equation, yielding on the left: $E(XE(Y|X)) = E(E(XY|X))$, and by what we had originally noted above, $= E(XY)$, which $= E(X(aX + b)) = E(aX^2 + bX) = aE(X^2) + bE(X)$. Therefore, we have our second equation: $E(XY) = aE(X^2) + bE(X)$. So we must solve the following linear equations:

1. $E(XY) = aE(X^2) + bE(X)$
2. $E(Y) = aE(X) + b$

Which yields the equation:

$$E(Y) - aE(X) = \frac{E(XY) - aE(X^2)}{E(X)}$$

$$\implies E(X)E(Y) - E(XY) = a((E(X))^2) - E(X^2))$$

$$\implies a = \frac{E(XY) - E(X)E(Y)}{E(X^2) - (E(X))^2} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$$

$$\implies b = E(Y) - aE(X) = E(Y) - E(X)\frac{\text{Cov}(X,Y)}{\text{Var}(X)}$$

4c) 4.9.15

Suppose that $X_1, \ldots, X_n$ are random variables for which Var(X_i) has the same value σ^2 for $i = 1, \ldots, n$ and $\rho(X_i, X_j)$ has the same value ρ for every pair of values i and j such that $i \neq j$. Prove that $\rho \geq -\frac{1}{n-1}$.

Proof. Let us first note that if we want to find the variance of the sum of: $Z = X_1 + \cdots + X_n$, then:

$$\text{Var}(Z) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$$

In this scenario, we have $\text{Var}(X_i) = \sigma^2 \ \forall i$ and $\text{Cov}(X_i, X_j) = \rho\sigma^2 \ \forall i \neq j$, which $\implies$

$$Var(Z) = n\sigma^2 + 2\binom{n}{2}\rho\sigma^2 = n\sigma^2 + n(n-1)\rho\sigma^2$$

We next recall $Var(Y) \geq 0 \ \forall Y$. Therefore, and in recalling $\sigma^2, n > 0$:

$$0 \leq n\sigma^2 + n(n-1)\rho\sigma^2 \implies -n\sigma^2 \leq n(n-1)\rho\sigma^2 \implies -1 \leq (n-1)\rho$$

$$\implies \rho \geq -\frac{1}{n-1}$$

□

5 Chapter 5 Questions

5a) 5.4.16

In this exercise, we shall prove that the three assumptions underlying the Poisson process model do indeed imply that occurrences happen according to a Poisson process. What we need to show is that, for each t, the number of occurrences during a time interval of length t has the Poisson distribution with mean λt . Let X stand for the number of occurrences during a particular time interval of length t. Feel free to use the following extension of Eq. (5.4.7): For all real a,

$$\lim_{u \to 0}(1 + au + o(u))^{\frac{1}{u}} = e^a$$

1. *For each positive integer n divide the time interval into n disjoint subintervals of length $\frac{t}{n}$ each. For $i = 1, \dots, n$, let $Y_i = 1$ if exactly one arrival occurs in the i'th subinterval, and let A_i be the event that two or more occurrences occur during the i'th subinterval. Let $W_n = \sum_{i=1}^n Y_i$. For each non-negative integer k, show that we can write $Pr(X = k) = Pr(W_n = k) + Pr(B)$, where $B \subseteq \cup_{i=1}^n A_i$.*

2. *Show that $\lim_{n\to\infty} Pr(\cup_{i=1}^n A_i) = 0$. Hint: Show that $Pr(\cap_{i=1}^n A_i^c) = (1 + o(u))^{\frac{1}{u}}$ where $u = \frac{1}{n}$.*

3. *Show that $\lim_{n\to\infty} Pr(W_n = k) = \frac{e^{-\lambda}(\lambda t)^k}{k!}$. Hint: $\lim_{n\to\infty} \frac{n!}{n^k(n-k)!} = 1$.*

4. *Show that X has the Poisson distribution with mean λt.*

1. *Proof.* Trivially from chapter 1, we know $\{X = k\} = (\{X = k\} \cap A) \cup (\{X = k\} \cap A^c)$ $\forall$ sets A, and the two sets, $(\{X = k\} \cap A)$ and $(\{X = k\} \cap A^c)$ are disjoint. If we let $A = \cup_{i=1}^n A_i$, then $(\{X = k\} \cap (\cup_{i=1}^n A_i)^c) = \{W_n = k\}$. We now trivially note $(\{X = k\} \cap (\cup_{i=1}^n A_i)) \subseteq A$ $\implies Pr(X = k) = Pr(W_n = k) + Pr(B)$ where $B = (\{X = k\} \cap (\cup_{i=1}^n A_i)) \subseteq A$. $\square$

2. *Proof.* First, we note as is stated in the part 1, $\exists n$ disjoint subintervals of length $\frac{t}{n}$ $\implies$ $A_1, \dots, A_n$ are independent and that $Pr(A_i) = Pr(A_j)$ $\forall i, j$. Therefore,

$$Pr(\cap_{i=1}^n A_i^c) = \prod_{i=1}^n Pr(A_i^c) = [Pr(A_1^c)]^n = [1 - Pr(A_1)]^n$$

It was our assumption that $Pr(A_i) = o(\frac{1}{n}) = o(u)$ $\implies$

$$\lim_{n\to\infty} Pr(A) = 1 - \lim_{n\to\infty} (1 + \frac{0}{n} - o(\frac{1}{n}))^n = 1 - e^0 = 0$$

$\square$

3. *Proof.* We recall that if $\exists n$ Bernoulli R.V.'s with the parameter $p = \frac{\lambda t}{n} + o(u)$, and if $W_n = \sum_{i=1}^n Y_i$, then:

$$Pr(W_n = k) = \binom{n}{k}\left(\frac{\lambda t}{n} + o(u)\right)^k \left(1 - \frac{\lambda t}{n} - o(u)\right)^{n-k}$$

Next, we note $\lim_{n\to\infty} n^k(\frac{\lambda t}{n} + o(u))^k = (\lambda t)^k$, and $\lim_{n\to\infty} n^k(1 - \frac{\lambda t}{n} - o(u))^{n-k} = e^{-\lambda t}$. Therefore,

$$\lim_{n\to\infty} Pr(W_n = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \lim_{n\to\infty} \frac{n!}{n^k(n-k)!} = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

$\square$

4. *Proof.* From part 1, we already saw $Pr(X = k) = Pr(W_n = k) + Pr(B)$, since $Pr(X = k) \neq f(n)$ $\implies$

$$Pr(X = k) = \lim_{n\to\infty} Pr(W_n = k) + \lim_{n\to\infty} Pr(B) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} + 1 - \lim_{n\to\infty}(1 + \frac{0}{n} - o(\frac{1}{n}))^n$$

$$= \frac{e^{-\lambda t}(\lambda t)^k}{k!} \text{ By parts 1-3}$$

$\square$

5b) 5.7.24

Review the derivation of the Black-Scholes formula (5.6.18). For this exercise, assume that our stock price at time u in the future is $S_0 e^{\mu u + W_u}$, where W_u has the gamma distribution with parameters αu and β with $\beta > 1$. Let r be the risk-free interest rate.

1. *Prove that* $e^{-ru}E(S_u) = S_0 \iff \mu = r - \alpha\log(\frac{\beta}{\beta-1})$.

2. *Assume that* $\mu = r - \alpha\log(\frac{\beta}{\beta-1})$. *Let R be 1 minus the c.d.f. of the gamma distribution with parameters αu and 1. Prove that the risk-neutral price for the option to buy one share of the stock for the price q at time u is* $S_0 R(c[\beta-1]) - qe^{-ru}R(c\beta)$, *where:*

$$c = \log\left(\frac{q}{S_0}\right) + \alpha u \log\left(\frac{\beta}{\beta-1}\right) - ru$$

3. *Find the price for the option being considered when* $u = 1$, $q = S_0$, $r = 0.06$, $\alpha = 1$, *and* $\beta = 10$.

1. *Proof.* We recall $\psi(t)$ for the Gamma distribution is $= (\frac{\beta}{\beta-t})^\alpha$. Therefore,

$$E(S_u) = E(S_0 e^{\mu u + W_u} = S_0 e^{\mu u} E(e^{W_u}) = S_0 e^{\mu u}(\frac{\beta}{\beta-1})^\alpha$$

Thus, $S_0 = e^{-ru}E(S_u) \iff$

$$S_0 = e^{-ru}S_0 e^{\mu u}(\frac{\beta}{\beta-1})^{\alpha u} \iff \log(1) = -ru + \mu u + \alpha u\log(\frac{\beta}{\beta-1}) \iff \mu = r - \alpha\log(\frac{\beta}{\beta-1})$$

□

2. *Proof.* We recall the value of an option at time u will be $h(S_u)$, where $h(s) = s - q$ if $s > q$ and 0 otherwise. Therefore, $h(S_u) > 0 \iff$

$$W > \log\left(\frac{q}{S_0}\right) + \alpha u \log\left(\frac{\beta}{\beta-1}\right) - ru = c$$

The risk-neutral price of the option is the present value of $E(h(S_u))$, which equals:

$$e^{-ru}E(h(S_)) = e^{-ru}\int_c^\infty \left[S_0 e^{\mu u + W_u} - q\right]\frac{\beta^{\alpha u}}{\Gamma(\alpha u)} w^{\alpha u - 1}e^{-\beta w}dw$$

We split the integrand into two parts at the $-q$. The second integral is then just a constant times the integral of a normal p.d.f., namely,

$$-qe^{-ru}\int_c^\infty \frac{\beta^{\alpha u}}{\Gamma(\alpha u)} w^{\alpha u - 1}e^{-\beta w}dw = -qe^{-ru}R(c\beta)$$

The first integral is:

$$e^{-ru}S_0\int_c^\infty e^{\mu u + W_u}\frac{\beta^{\alpha u}}{\Gamma(\alpha u)} w^{\alpha u - 1}e^{-\beta w}dw = S_0[R(c(\beta-1))]$$

Combining these two integrals yields:

$$S_0 R(c[\beta - 1]) - qe^{-ru} R(c\beta)$$

where

$$c = \log\left(\frac{q}{S_0}\right) + \alpha u \log\left(\frac{\beta}{\beta - 1}\right) - ru$$

□

3. Since $q = S_0$ it $\implies$

$$c = \log\left(\frac{q}{S_0}\right) + \alpha u \log\left(\frac{\beta}{\beta - 1}\right) - ru = \log\left(\frac{S_0}{S_0}\right) + \log\left(\frac{10}{9}\right) - 0.06 \approx 0.0453605156$$

From here we substitute c into

$$S_0 R(c[\beta - 1]) - qe^{-ru} R(c\beta) \approx S_0 \left[R(0.0453605156(9)) - e^{-.06} R(0.0453605156(10))\right]$$
$$\approx 0.0599976 S_0$$

6 Non-Textbook Problems

6a) A

For an event B with $P(B) > 0$, define $Q(A) = P(A|B)$ for any event A. Show that Q satisfies Axioms 1-3 of probability and conclude Q is a probability.

Let us first recall the first 3 Axioms of Probability:

1.) For every event A, $Pr(A) \geq 0$

2.) $Pr(S) = 1$

3.) If $A_1, A_2, \ldots$ is a countably infinite sequence of disjoint events, then $Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Pr(A_i)$.

Proof. Since Pr is a probability function, we know that $\forall$ sets X, $Pr(X) \geq 0$. Furthermore, we recall the formal definition of condition probability: $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$. From here we note $\forall x \in (A \cap B), x \in B \implies Pr(A \cap B) \leq Pr(B) \implies \frac{Pr(A \cap B)}{Pr(B)} \leq 1$. Thus, we can conclude the first 2 Axioms hold.

For the 3rd Axiom, We find if $A_1, A_2, \ldots$ is a countably infinite sequence of disjoint events, then $\forall i \neq j$ $(A_i \cap B) \cap (A_j \cap B) = (A_i \cap A_j) \cap B = (\emptyset \cap B) = \emptyset \implies (A_i \cap B)$ and $(A_j \cap B)$ are independent $\implies Pr(\cup_{i=1}^{\infty} A_i \cap B) = \sum_{i=1}^{\infty} Pr(A_i \cap B)$ (steps made due to the distribution law and definition of disjointness respectively). Let us call these findings "Rule Z". Next, $Pr(\cup_{i=1}^{\infty} Q(A_i)) = Pr(\cup_{i=1}^{\infty}(A_i|B)) = \ldots$

$$= \frac{Pr((\cup_{i=1}^{\infty}(A_i)) \cap B)}{Pr(B)}$$

$$= \frac{Pr(A_1 \cap B) \cup (A_2 \cap B) \cup \ldots)}{Pr(B)} \text{ by distribution law}$$

$$= \frac{\sum_{i=1}^{\infty} Pr(A_i \cap B)}{Pr(B)} \text{ by Rule Z}$$

$$= \sum_{i=1}^{\infty} Pr(A_i|B) \text{ and thus completes our proof}$$

□

6b) B

Assume that $X_1, X_2, \ldots$ are an i.i.d. random variables having $E(|X_j|^p) < \infty$ for $1 < p \leq 2$. Let $\mu = E(X_j)$.

1. *Show that $\overline{X}_n = \frac{X_1+\cdots+X_n}{n} \to \mu$ in L^p as $n \to \infty$.*
2. *Show that $\overline{X}_n \to \mu$ almost surely.*

1. First, we recognize the function, $f(X_j) = |X_j|^p$ must be convex for $p \in (1, 2]$. Therefore, we can establish the lower bound from Jensen's Inequality: $E(|\overline{X}_n|^p) \geq |E(\overline{X}_n)|^p$ and by the Theorems of 4.2, $plim(E(\overline{X}_n)) = \mu$. Therefore, from Jensen's and Sec. 4.2, $\overline{X}_n \to c$ s.t. $c \geq \mu$ as $n \to \infty$. We now establish the upper bound by $E(|\overline{X}_n|^p) \leq \sum_{i=1}^{n} E(|X_i|^p) \implies$ (also by 4.2 findings) $\overline{X}_n \to k, k \leq \mu$ Therefore, combining both our upper and lower bounds, we get $\overline{X}_n \to \mu$ as $n \to \infty$
2. By our previous findings, we know each of these upper and lower bounds work for *plim*'s, but they also do work for *a.s.* covergance, therefore, $P(\lim_{n\to\infty} \overline{X}_n = \mu) = 1$ as $n \to \infty$.

6c) C

Two random variables X and Y have bivariate normal distribution if the joint density is:

$$pdf_{X,Y}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left((\frac{x-\mu_x}{\sigma_x})^2 - 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y}) + (\frac{y-\mu_y}{\sigma_y})^2\right)}$$

1. *Compute marginal probability density function of X.*
2. *Show conditional distribution of Y given $X = x$ is $N(\nu, \tau^2)$ and find ν and τ.*
3. *Show that X and Y are independent if and only if $Cov(X, Y) = 0$.*

1. *Proof.* Let $Q(X, Y) = -\frac{1}{2(1-\rho^2)}\left((\frac{x-\mu_x}{\sigma_x})^2 - 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y}) + (\frac{y-\mu_y}{\sigma_y})^2\right)$. We can re-write $Q(X, Y)$ as:

$$Q(X, Y) = \frac{-1}{2(1-\rho^2)}\left[\left((\frac{y-\mu_y}{\sigma_y}) - \rho(\frac{x-\mu_x}{\sigma_x})\right)^2 + (1-\rho^2)(\frac{x-\mu_x}{\sigma_x})^2\right]$$

$$= \frac{-1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\alpha(x)}{\sigma_y\sqrt{1-\rho^2}}\right)^2\right], \quad \text{where } \alpha(x) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$$

Thus, since $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$, we now have:

$$f_X(x) = \frac{e^{-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\frac{y-\alpha(x)}{\sigma_y\sqrt{1-\rho}})^2} dy$$

We next recognize that:

$$\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2}(\frac{y-\alpha(x)}{\sigma_y\sqrt{1-\rho^2}})^2} dy \equiv \text{the p.d.f. of the } N(\alpha(x), \sigma_y^2(1-\rho^2) \text{ distribution}$$

Therefore,

$$\xi(x,y) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2}(\frac{y-\alpha(x)}{\sigma_y\sqrt{1-\rho^2}})^2} dy = 1$$

Thus,

$$f_X(x) = \zeta(x)\xi(x,y) = \zeta(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2} \equiv N(\mu_x, \sigma_x^2)$$

□

2. *Proof.* We recall: $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$. Thus, from part 1, we can immediately substitute in $f_X(x) = \zeta(x)$ and $f_{X,Y}(x,y)$:

$$f_{Y|X}(y|x) = \frac{[2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}]^{-1} e^{\frac{-1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\alpha(x)}{\sigma_y\sqrt{1-\rho^2}}\right)^2\right]}}{[\sqrt{2\pi}\sigma_x]^{-1} e^{-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2}}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(\frac{y-\alpha(x)}{\sigma_y\sqrt{1-\rho^2}}\right)^2} \equiv N(\alpha(x), \sigma_y\sqrt{1-\rho^2})$$

Thus, $f_{Y|X}(y|x)$ does $= N(\nu, \tau)$, where $\nu = \alpha(x)$ and $\tau = \sigma_y\sqrt{1-\rho^2}$. □

3. *Proof.* We recall Corollary 3.5.1, which states two variables are independent $\iff f_{X,Y}(x,y) = f_X(x)f_Y(y)$. Using this information, and by the symmetry of x and y in the Bivariate Normal Distribution, we know $f_Y(y) \equiv N(\mu_y, \sigma_y^2)$. Therefore,

$$f_Y(y)f_X(x) = \frac{1}{2\pi\sigma_x\sigma_y} e^{\frac{-1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

However, since $\rho^2 \in [0,1]$, we have a problem with independence. If we let:

$$\mathscr{M}(g(x,y,\rho),\rho) = e^{\frac{1}{1-\rho^2}g(x,y,\rho) - \log(\sqrt{1-\rho^2})}, \text{and}$$

$$g(x,y,\rho) = -(1-\rho^2)\log(2\pi\sigma_x\sigma_y) - \frac{1}{2}\Big((\frac{x-\mu_x}{\sigma_x})^2 - 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y}) + (\frac{y-\mu_y}{\sigma_y})^2\Big)$$

Thus one can see that: $\mathscr{M}(g(x,y,\rho),\rho) = f_{X,Y}(x,y)$. Furthermore, by construction, it is impossible for $\mathscr{M}(g(x,y,\rho),\rho) = f_X(x)f_Y(y)$ unless $\rho = 0$ since otherwise the $\frac{1}{1-\rho^2}$ and $-\log(\sqrt{1-\rho^2})$ terms will be $\neq 1$ and 0 respectively, thereby shifting the density too much for the now non-zero $-2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y})$ to bring back. Also, since $\rho = 0 \iff \text{Cov}(X,Y) = 0 \implies X$ and Y are independent $\iff \text{Cov}(X,Y) = 0$ □

6d) D

Suppose $Y_i \sim$ ind. exponential(μ_i) where $\mu_i = \frac{1}{\beta x_i}$ where $\beta > 0$ and $x_i > 0$.

1. *Find β maximizing:*

$$\prod_{i=1}^{n} pdf_{Y_i}(y_i)$$

2. *Show that the $\hat{\beta}$ found in (1) converges to β in probability.*

3. *Show that $Var(\hat{\beta}) \to 0$.*

1. *Proof.* We know $f_{Y_i}(y_i) = \frac{1}{\beta x_i} e^{-\frac{1}{\beta x_i} y_i}$ Therefore, by standard MLE practices, we have

$$\mathscr{L} = \prod_{i=1}^{n} \frac{1}{\beta x_i} e^{-\frac{1}{\beta x_i} y_i} \implies \log(\mathscr{L}) = \log\left[\frac{1}{\beta^n}\left(\prod_{i=1}^{n}\frac{1}{x_i}\right) e^{-\frac{1}{\beta}\sum_{i=1}^{n}\frac{y_i}{x_i}}\right]$$

$$\text{And Maximizing: } -n\log(\beta) + \log\left(\prod_{i=1}^{n}\frac{1}{x_i}\right) - \frac{1}{\beta}\sum_{i=1}^{n}\frac{y_i}{x_i} \implies$$

$$\frac{\partial \log(\mathscr{L})}{\partial \beta} = -\frac{n}{\beta} + \frac{n}{\beta^2}\sum_{i=1}^{n}\frac{y_i}{x_i} = 0 \implies \hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}$$

□

2. First, we know $E(\hat{\beta}) = E(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}) = \frac{1}{n}E(\sum_{i=1}^{n}\frac{y_i}{x_i}) = \frac{1}{n}\sum_{i=1}^{n}E(\frac{y_i}{x_i}) = \frac{1}{n}(n\beta) = \beta$. By theorems from 4.2, and by Chebyshev's Inequality, and noting $E(Y_i) = 1/\mu_i \implies E(Y_i/x_i) = \frac{1}{x_i \mu_i} = \frac{\beta x_i}{x_i} = \beta$.

3. Since there is only one parameter within an exponential distribution, and $\sigma^2 = \mu^2$ for exponential $\implies \hat{\beta}^2 = \hat{\sigma}^2$. Since we already showed that $plim(\hat{\beta}) = \beta \implies plim(\hat{\beta}^2) = \beta^2$. Next, we apply Chebyshev's Inequality, which results in the fact that Var($\hat{\beta}$) $\to 0$ as $n \to \infty$ else our previous findings would not hold under such assumptions.