Fall 2011

Probability Theory

Author:
Todd Gaugler

Professor:
Dr. Stefan Ralescu

December 14, 2011

Contents

Chapter 1

Introduction

1.1 Homework

- Chapter 1: pp(22-26)/ 4,6,7,9,12, 34, 35, 36, 37, 45,46
- We will cover chapters: 1, 3, 4 , 5 , 6 , 7 , 8
- Chapter 3: pp(77-82) 4, 5, 10, 11, 13, 14, 15, 16, 17, 18, 31, 32, 33
- Chapter 4: pp(104-108) 2, 3, 4, 6, 8, 9, 13, 14, 15, 21, 26, 27, 28, 29, 30
- Chapter 5: pp(133-135) 1, 7, 8, 9, 10, 14, 19, 21, 23, 24, 25, 31, 33, 34, 39, 41, 43, 44
- Chapter 6: pp(169-172) 1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 19, 21, 22, 29, 30
- Chapter 7: pp(192-196) 1, 3, 4, 5, 6, 14, 15, 31, 32, 33, 34, 35, 36, 37, 38, 39
- Chapter 8: pp(213- 215) 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 20, 21, 22
- 3 exams, I(30%), II (30%), III (40%)
- Office hours are Wednesday 2 to 3 p.m. in KY 407

1.2 Introductory Probability, Chapter 1

We derive the notion of Probability from that of the idea of a Random experiment, for example tossing a coin, rolling a die, picking a card from a standard deck of cards, and so on.

Ω Is here going to be used as the "sample space" or "probability space" which is the collection of all possible results or outcomes of a particular random experiment. It is a set. We think of a "sample point" as being a result or outcome of a particular experiment. Here an event is a subset of the sample space.

Definition. A non-empty collection of subsets $\mathcal{A}$ of Ω is called a **σ-field** if the following properties are true:

1. If $A \in \mathcal{A} \Rightarrow A' \in \mathcal{A}$ where here, $A^c = A' = \bar{A}$.
2. If we have a sequence, possibly infinite, of elements $A_1, A_2, A_3 \in \mathcal{A}$, this tells us that $\bigcup\limits_{n=1}^{\infty} A_n = 1$ and $\bigcap\limits_{n=1}^{\infty} A_n$ are in $\mathcal{A}$.

Definition. The definition of Probability is as follows: Let Ω be a probability space and let $\mathcal{A}$ be a σ-field on Ω where you have a function $P : \mathcal{A} \to [0, \infty]$ such that:

1. $P(\Omega) = 1$
2. $P(A) \geq 0$
3. If $(A_1, A_2, A_3, ...)$ are mutually disjoint (any two have no intersection), in $\mathcal{A}$, then $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$. This tells us that probability is countable-additive (the sequence may be infinite, but is countable).

Example. Suppose $A_n = \Phi =$ Empty, all $n \geq 1$. Now suppose that $P(\Phi) = P(\Phi) + P(\Phi) + P(\Phi)$... This implies that $P(\Phi) = 0$, since we have (2) and each event has a finite probability. Where the probability of an impossible event must be 0.
Claim. For every event $A \in \mathcal{A}$, $P(A) \leq 1$. This follows from noticing the following:

$$\Omega = A \cup A'$$

Which are obviously disjoint. Thus, take $A_1 = A$, and take $A_2 = A'$ and take $A_3 = A_4 = empty... = \Phi$. Using (3), and recognizing that these sets are mutually disjoint, seeing that $\bigcup_{n=1}^{\infty} A_n = \Omega$, using (3) we know that $1 = P(\Omega) = P(\bigcup_{n=1}^{\infty}) = P(A) + P(A') + ...$ which tells us:

$$1 = P(A) + P(A')$$

and since $P(A') \geq 0$, we know that $P(A) \leq 1$.

For homework, we have the following problem: Show that if $A, B \in \mathcal{A}$ and $A \subseteq B$ then $P(A) \leq P(B)$. This follows again from doing something very similar to what we just did, since we just proved that $A \subseteq \Omega \Rightarrow P(A) \leq P(\Omega)$.
Fact. $P(A \cup B) = P(A) + P(B) - P(AB)$ Where $AB = A \cap B$. For homework: show that $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$.
Lemma 1. *For any finite sequence of events, $A_1, A_2, ...A_n$, we have:*

$$P(A_1 \cup A_2 \cup ... \cup A_n) \leq P(A_1) + P(A_2) + ... + P(A_n)$$

*This is called **Boole's Inequality**.*
Note. If $A_1, A_2, A_3, ...A_n$ are mutually disjoint events, then it is always true that

$$P(A_1 \cup A_2 \cup .. \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$$

This follows from Axiom number (3). The proof of this starts by noticing the obvious fact that

$$P(A \cup B) \leq P(A) + P(B)$$

Inductively, this can continue in the following way:

Proof. Assume that what we are talking about is true for n events. Given $n+1$ events, where we have $A_1, A_2, ...A_n$ and A_{n+1} we see that:

$$P(\bigcup_{1}^{n+1} A_n) = P(\bigcup_{i=1}^{n} A_i \cup A_{n+1}) \leq P(A_1) + P(A_2) + ... + P(A_n) + P(A_{n+1})$$

Since we know that $P(A \cup B) \leq P(A) + P(B)$ from Boole's inequality. This is on page (12) of our textbook. $\square$

The **Strong law of large numbers** dictates that the sample average converges to the mean in the subset of the probability space.

Theorem 2. *If $A_1 \subseteq A_2 \subseteq A_2...$ are events and $\bigcup\limits_{i=1}^{\infty} A_i = A$ then* $\lim_{n\to\infty} P(A_n) = A$

Proof. Let $A_1 = B_1$, $B_2 = A_2 \cap A_1'$, $B_3 = A_3 \cap A_2'$. It is true that $A_n = B_1 \cup B_2 \cup ... \cup B_n$ which says that $P(A_n) = P(B_1) + P(B_2) + ... + P(B_n) = \sum_{i=1}^{n} P(B_i)$. Now, look at the infinite unions of B_i, which looks like $A = B_1 \cup B_2 \cup ...$ which is a countable infinite sequence. And we also know that $P(A) = \sum_{i=1}^{\infty} P(B_i)$. Which says that $\lim \sum_{i=1}^{n} P(B_i) = P(A)$. But since that partial sum is $P(A_n)$, we see that our proof is complete. $\square$

Going back to Boole's inequality, we would like to prove it for a sequence of infinitely many events:

$$A_1, A_2, A_3, ...$$

Claim. $P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$

Proof. Lets assume that $\sum_{i=1}^{\infty}(A_i) < \infty$. Look at $A_1 \cup A_2 \cup ... A_n = A_1 \cup (A_1 \cup A_2) \cup (A_1 \cup A_2 \cup A_3) \cup ... (A_1 \cup A_2 \cup ... \cup A_n)$ Now, it is clear that we have another nested sequence as we did with the last claim. We will rename each term $B_1, B_2, ...B_n$ and claim that $B_1 \subseteq B_2 \subseteq B_2 \subseteq ...$ The union $A = \bigcup\limits_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} iB_i$. By theorem (1) the $\lim_{n\to\infty} P(B_n) = P(A)$ Let's look at

$$P(B_n) \leq P(A_1) + P(A_2) + ... + P(A_n)$$

which follows from Boole's inequality. Also,

$$P(B_n) \leq P(A_1) + P(A_2) + ... + P(A_n) \leq \sum_{i=1}^{\infty} P(A_i) = M < \infty$$

$\square$

In calculus, if we have some $x_n \leq M$ and $\lim_{n\to\infty} x_n = n \Rightarrow x \leq M$. Let $n \to \infty$ in the following double inequality:

$$a_n \leq b_n \leq k \Rightarrow a \leq b \leq M$$

$$P(A) \leq M$$

which is the conclusive result we wanted from Boole's equality.

Chapter 2

Conditional Probability

Recall, $P(A|B) := P(AB)/P(B)$ assuming $P(B) \neq 0$, from which you obtain $P(AB) = P(A|B)P(B) = P(B|A)P(A)$. Also recall that A, B are said to be **independent** if $P(AB) = P(A)P(B)$ (and if $P(B) \neq 0,\ P(A|B) = P(A)$.

Approaching the independence of three events A, B, C, we call these three events "mutually independent" if A, B are independent, A, C are independent, B, C are independent, and $P(ABC) = P(A)P(B)P(C)$. It is worth mentioning that the fourth condition does not necessarily follow from the first three:
Example. $\Omega = \{1, 2, 3, 4\}$ where each outcome is equally likely. Take

$$A = \{1, 2\}, B = \{1, 3\}, C = \{1, 4\}$$

Notice that $\{AB\} = \{AC\} = \{BC\} = \{1\}$. Since the probabilities $P(A) = P(B) = P(C) = P(D) = 1/2$ and thus $P(AB) = P(AC) = P(BC) = 1/4$ and $P(AB) = P(A)P(B), P(AC) = P(A)P(C), P(BC) = P(BC) = P(B)P(C)$. But, $P(ABC) = 1/4$. However, $P(A)P(B)P(C) = 1/8$. Thus, the four condition is also needed, since the conditions are not mutually independant.

$$P(B_n) \leq P(A_1) + P(A_2) + ... + P(A_n)$$

If $\Omega = \bigcup_{i=1}^{\infty} B_i$ where $B_1, B_2, ...$ are disjoint, this is called a **partition** of the probability space Ω. Now given an event A, the **Law of Total Probability** for A is that $P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i)$

Recall that we have the following formula: $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$. By simply replacing B by some B_k, we get $P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{\infty} P(A|B_i)P(B_i)}$. This is known as **Bayes formula**.

2.1 Chapter 1, Question 12

Example. Select a point at random from $\Omega = S$. In talking about the measure of $|S|$, we talk about the area of S, which has a finite area. In making S a probability space, we need

$A \subseteq S$ such that $P(A) = \frac{|A|}{|S|}$. If we were talking about 3 dimensions, we would use area. This is called a uniform probability space.

In taking the unit square, it is clear that $P(\Omega) = 1$, since that is the area of the unit square. using the line $x + y = 1$, the triangle formed by under that line has exactly half of the area contained by the unit square. In comparing the points in $x+y = 1$ and $y = x$, in finding the probability $P(A|B) = \frac{P(AB)}{P(B)} = \frac{|AB|}{B} = \frac{1/4}{1/2} = 1/2$

2.2 Chapter 1, Question 34

Assume that A, B are independent events. This implies that $P(AB) = P(A)P(B)$. We know that $P(AB^c) = P(A) - P(AB)$, and using what we know about A, B, we can factor this into $P(A)(1 - P(B) = P(A)P(B^c)$. From this, we know that also A^c, B and A^c, B^c are independent.

Chapter 3

Discrete Random Variables

Definition. A **Random Variable** X is a function $X : \Omega \to (-\infty, +\infty)$. If the values of the random variable X can be written as finite or an infinite sequence, it is called **discrete**
Example. Let the values look like $x_1, x_2, ...x_n, ...$

There exists something called a probability function, which looks like: $f(x_i) = P\{X = x_i\}, i = 1, 2, ...$ In this case f is called the (discrete) density of X, or the probability function. We can extend this to:

$$f(x) = P\{X = x\}$$

f becomes 0 if x is not a value of the random Variable. We have the following properties:

1. $f(x) \geq 0$
2. $\{x \mid f(x) \neq 0\} = \{x_1, x_2, ...\} =$ countable.
3. $\sum_i f(x_i) = 1$

Example. The $X = Binomial(n, p)$, is the number of successes of a Bernoulli trial with probability of trial being (p) being repeated n times (under independent condition). X is the total number of such successes, and has the following values:

$$X = 1, 2, ...n$$

And its density takes the following form: $f(x) = P\{X = x) = \binom{n}{x} p^x(1-p)^{n-x}$.
Example. The Poisson random variable: let $\lambda > 0$ be given:

$$X \mapsto x = 0, 1, 2, 3, 4, ...$$

And the following is also true:

$$f(x) = P\{X = x\} = e^{\lambda} \cdot \frac{\lambda^x}{x!}, \quad \sum_{x=0}^{\infty} f(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$$

Theorem 3. ***The Poisson Approximation Theorem*** *says that if n is Large, and p is small, (for example, $n = 500, p = .03$), then with $\lambda = n \cdot p$, we have*

$$\binom{n}{x} p^x (1-p)^{n-x} \approx e^{-\lambda} \frac{\lambda^x}{x!}$$

More formally, Let $p_n \in (0,1)$ such that $\lim_{n \to \infty} np_n$ exists and is equal to $\lambda > 0$. Then, if $x \geq 0$ is an integer, $\binom{n}{x} p_n^x (1-p)n)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}$

Example. Let $X = Geometric(p)$ where $0 < p < 1$ if X has values $0, 1, 2, ...$ and density

$$f(x) = P\{X = x\} = pq^x$$

Where $q = 1 - p$. This is geometric, since

$$\sum_{x=0}^{\infty} f(x) = 1 \quad \text{and} \quad \sum_{x=0}^{\infty} pq^x = p(1 + q + q^2 + q^3 + ...)$$

and since $1 + q + q^2 + q^3 + ... = \frac{1}{1-q}$, and since $1 - q = p$, this is really $\frac{1}{p}$ and this sum is 1.

3.1 Calculating Probabilities

We have the following formula: Given $A \subseteq \mathbb{R}$, and looking for $P\{X \in A\}$, we calculate this using:

$$P\{X \in A\} = \sum_{\forall x_i \in A} f(x_i)$$

which is a monumentally important formula.

Example. Given $A = [a, b]$, $P\{a \leq X \leq b\} = \sum_{\text{all } a \leq x_i \leq b} f(x)$. Similarly, given $A = (a, b]$,

$$P\{a < X \leq b\} = \sum_{\text{all } q < x_i \leq b} f(x_i)$$

If you took $A = (-\infty, x]$, then

$$F(x) := P\{X \leq x\} = \sum_{\text{all } x_i \leq x} f(x_i) = \sum_{\forall t \leq x} f(t)$$

And $F(x)$ is called the cumulative distribution function, abbreviated $(c.d.f)$ of X. Be sure to notice the following:

$$F(x) = \sum_{\forall t \leq x} f(t)$$

For any discrete random variable X, we have two functions $f(x)$ and $F(x)$, the density function and the c.d.f, which are related as above. Regarding $F(x)$, we can say:

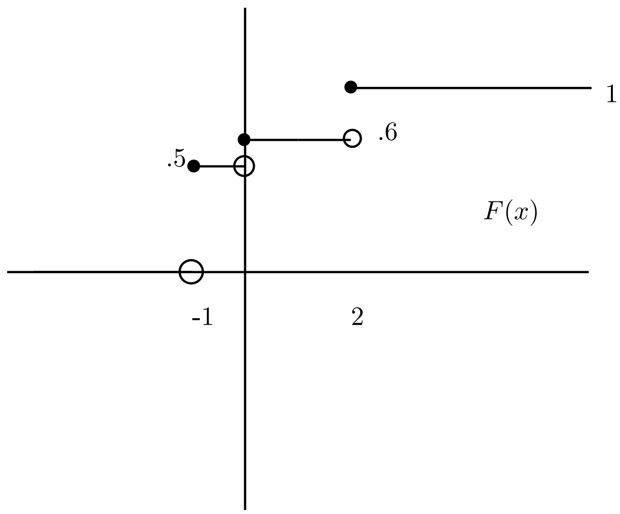
1. $0 \leq F(x) \leq 1$

2. $F(x)$ is a step function (a constant function that has a 'jump'), nondecreasing, right-continuous, with jumps exactly at the values of the random variable X.

Example. Say $X = \{-1, 0, 2\}$ and $f(0) = .5, f(0) = .1, f(2) = .4)$ This can be represented by the picture below. The following are also true:

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1$$

Also notice that $F(X)$ never goes down. For homework, suppose that X is a discrete random variable, and we know $F(x)$, the distribution function. How do we find $f(x)$?



3.2 Random Vectors of Discrete Random Variables

Suppose that $X_1, X_2, ...X_r$ are discrete random variables. Putting them together in a vector,

$$(X_1, X_2, X_3, ...X_r) = \mathbb{X}$$

which is a discrete random vector of r component. A value of the random vector would be written as follows:

$$\hat{x} = (x_1, x_2, ...x_r)$$

Notice that $\mathbb{X} = \hat{x}$ if and only if $X_1 = x_1, X_2 = x_2, ..., X_r = x_r$. Now, let's introduce the density of the random vector: the **Density** of the random vector $\mathbb{X}$ is a function $f(\hat{x})$ is defined as $P\{\mathbb{X} = \hat{x}\}$ which we know to be equal to $P\{X_1 = x_1, X_2 = x_2, ...X_r = x_r\}$ This density $f(\hat{x})$ is also called **the joint density** of the random variables $X_1, X_2, ...X_r$.

3.3 Chapter 3, Question 10

Let $X = Geom(p)$, $x = 0, 1, 2, 3, ...$ and $f(x) = P(X = x) = pq^x$. We have

$$Y = \begin{matrix} X \text{ if } X < M \\ M \text{ if } X \geq M \end{matrix}$$

Assume that $M \geq 0$, and is an integer. In other words, Y is equal to the minimum between (X, M). For example, take $M = 4$. Thus $Y = x$ if $x = 0, 1, 2, 3$ and $Y = 4$ if $x \geq 4$. In finding the density of Y, note that the possible values of Y are $0, 1, 2, 3, 4$. Thus,

$$\begin{aligned}
g(0) &= P(Y = 0) = P(X = 0) = p \\
g(1) &= P(Y = 1) = P(X = 1) = pq \\
g(2) &= P(Y = 2) = P(x = 2) = pq^2 \\
g(3) &= P(Y = 3) = P(x = 3) = pq^3 \\
g(4) &= P(Y = 4) = P(x \geq 4) = 1 - (p + pq + pq^2 + pq^3)
\end{aligned}$$

From this, we should note the following lemma:
Lemma 4. *If $X = Geom(p)$, for any $x \geq 0$ integer, the $P(X \geq x) = q^x$.*

Proof. Notice that

$$P(X \geq x) = P(X = x) + P(X = x + 1) + P(X = x + 2)...$$

Using what we already know, this is equal to:

$$= pq^x + pq^{x+1} + pq^{x+2} + ...$$

Factoring out,

$$pq^x(1 + q + q^2 + q^3 + ...) = pq^x(\frac{1}{1-q}) = pq^x(\frac{1}{p}) = q^x$$

thus, in general for our homework question,

$$g(y) = P(Y = y) = \left| \begin{matrix} pq^q \text{ if } y = 0, 1, ...M - 1 \\ q^M \text{if } y = M \end{matrix} \right.$$

□

Now let's look at the converse of this lemma. We have the following:
Definition. All random variables with non negative integer values will be called 'in $\mathbb{IV}_+$'.
Theorem 5. *If $X \in \mathbb{IV}_+$is a random variable such that for each integer $x \geq 0$,*

$$P\{X \geq x\} = q^x \textit{ for some } 0 < q < 1$$

We can conclude from this that X is geometric, and that $p = 1 - q$ $(X = Geom(p = 1 - q))$.

Proof. We have to look at the density $F(x) = P(X = x)$ and show that $F(x) = pq^x$. Looking at the following union of events:

$$\{X = x\} \cup \{X \geq x + 1\} = \{X \geq x\}$$

Thus,

$$P(X = x) = P(X \geq x) - P(X \geq x + 1)$$
$$\Rightarrow f(x) = P(X = x) = q^x - q^{x+1} = q^x(1 - q) = q^x p$$

since $P(X \geq x) = q^x$ and $P(X \geq x + 1)$ is equal to q^{x+1}. This is also in the textbook. □

3.4 Chapter 3, Question 11(a)

Given that X is geometric, and $Y = X^2$, the values of Y are $0^2, 1^2, 2^2, ...$ Thus we can call $f(y) = P(Y = y) = P(X^2 = y) = P(X = \sqrt{y}) = pq^{\sqrt{y}}$ Looking at part (b) of this question, notice that

$$Y = X + 3 \Rightarrow y = 3, 4, 5, ...$$

Which tells us that $f(y) = P(Y = y) = P(X + 3 = y) = P(X = y - 3) = pq^{y-3}$.

3.5 Chapter 3, Question 14

3.5.1 Necessary Background in Probability Vectors

Recall that if we have: $\mathbb{X} = (X_1, X_2, ...X_r)$ then we have $\hat{x} = (x_1, x_2, ...x_r)$ and

$$f(\hat{x}) = P\{\mathbb{X} = \hat{x}\} = P\{X_1 = x_1, X_2 = x_2, ...X_r = x_r\}$$

called the joint (probability function) density of $X_1, X_2, ...X_r$. Let's look at the simple case in which $r = 2$, and call $X_1 = X, X_2 = Y$. We then have the vector (X, Y) which has the joint density

$$f(x, y) = P(X = x, Y = y)$$

And we also have

$$f_X(x) = P(X = x), \quad f_Y(y) = P(Y = y)$$

which are called the marginal probability functions. Looking at statement (11) in the textbook, we have the fact that:

$$f_X(x) = \sum_{\text{all y}} f(x, y), \quad f_Y(y) = \sum_{\text{all x}} f(x, y)$$

Which is shown on page 62 in the textbook. This implies that the most important thing to have is the joint function, since we can find the marginal functions from the joint function.

Definition. Two discrete random variables X, Y are called independent if the joint density:

$$f(x, y) = f_X(x) \cdot f_Y(y) \text{ for all x, y}$$

Practically, this says that 'the joint is the product of the marginals'. In general, we say that $X_1, X_2, ...X_r$ are independent if

$$f(x_1, x_2, ...f_r) = f_{X_1}(x) \cdot f_{X_2}(x) \cdot ... \cdot f_{X_r}(x)$$

3.5.2 Chapter 3, Question 14

Given X, Y, uniform random variables over$\{0, 1, ...N\}$, we know that:

$$f_X(x) = \frac{1}{n+1} = f_Y(y) \quad \text{where } x, y \in 0, 1, ...N$$

And we know that X, Y are independent.
Remark. Two discrete variables X, Y are independent if and only if $P\{X \in A \text{ and } Y \in B\} = P\{X \in A\} \cdot P\{Y \in B\}$ for any $A \subseteq \mathbb{R}, B \subseteq \mathbb{R}$. This equation 14 on page 64 in our textbook.

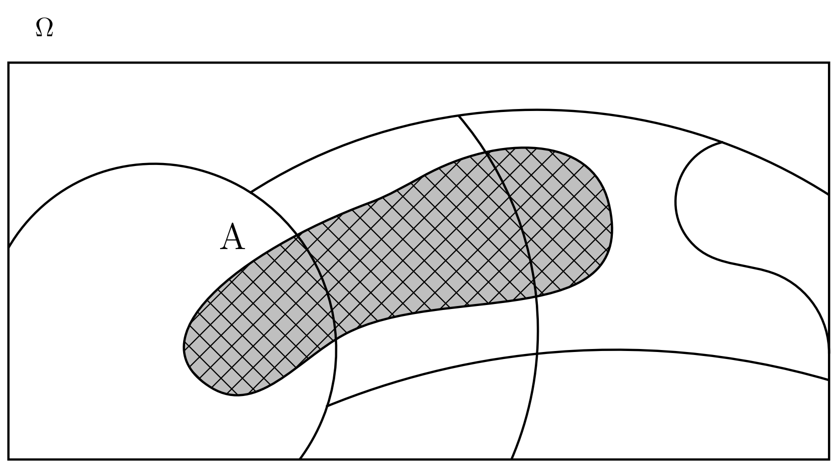
Going back to the question, observe that $P(X \geq Y) = \bigcup_{y=0}^{N} \{X \geq Y, Y = y\}$. And that

$$\Omega = \bigcup_{y=0}^{N} \{Y = y\}$$

which is a partition for Ω. We know that if we have

$$\Omega = \bigcup_k B_k, \quad \text{we can say the following for a set } A: \quad A = \bigcup_k AB_k$$

which can be illustrated by the following picture:



And now we can see that

$$\{X \geq Y, Y = y\} = \{X \geq y, Y = y\}$$

And the event $X \geq y$ is an event that depends only on the random variable X, and the event $Y = y$ is an event that only depends on the variable Y. This is nice, since the first event in our first union was dependent on BOTH X, Y. And, we know that $X \geq y$ means that $X \in [y, +\infty)$, and that $Y = y$ means that $Y \in \{y\}$. The intersection of these events $P\{X \geq Y\}$ can be written as $\bigcup_{y=0}^{N}\{X \in A, Y \in B\}$ where $A = [y, \infty), B = \{y\}$. The probability of the union is the sum of the unions. Proceeding,

$$P\{X \geq Y\} = \sum_{y=0}^{N} P\{X \geq y, Y = y\} = \underbrace{\sum_{y=0}^{N} P\{X \geq y\} \cdot P\{Y = y\}}_{independent} = \frac{1}{N+1}\sum_{y=0}^{N} P(X \geq Y)$$

and since

$$P(X \geq y) = P(X = y) + P(X = y+1) + ... + P(X = N) = (N - y + 1)\frac{1}{N+1}$$

so we have:

$$\frac{1}{N+1}\sum_{y=0}^{N}\frac{N-y+1}{N+1} = \frac{1}{(N+1)^2}\left(\sum_{y=0}^{N}(N+1) - \sum_{y=0}^{N} y\right)$$
$$= \frac{1}{(N+1)^2}\left((N+1)^2 - \frac{N(N+1)}{2}\right) = 1 - \frac{N}{2(N+1)}$$

Try the following as a homework: example 14 on pages 64,65. This is a very important idea for our class.

We are now asked to find $P(X = Y)$. This can be written as:

$$P(X = Y) = \bigcup_{y=0}^{N}\{X = Y, Y = y\} = \bigcup_{y=0}^{N}\{X = y, Y = y\}$$

and since X, Y are independent, the probability would be their product. Thus,

$$P(X = y) = \sum_{0}^{N} P(X = y) \cdot P(Y = y) = \frac{1}{(N+1)^2}\sum_{0}^{N} 1 = \frac{1+N}{(1+N)^2} = \frac{1}{N+1}$$

3.6 The Density of the Sum of Two Independent Random Variables

Let X and Y be independent and discrete. Let $Z = X + Y$. We have:

$$f_{X+Y}(z) = \sum_{\text{over all x}} f_X(x) \cdot f_Y(z - x)$$

Suppose that X, Y are non negative integer values, are in $\mathbb{Y}_+$. This implies that $X+Y \in \mathbb{V}_+$. This ensures that our above formula becomes

$$f_{X+Y}(z) = \sum_{x=0}^{z} f_X(x) \cdot f_Y(z - x)$$

This is called the 'convolution'.

Example. Suppose we have two random variables $X = Poisson(\lambda), Y = Poisson(\lambda)$ and that X, Y are independent. $X + Y = Z$, and it is clear that Z has values $0, 1, 2, ...$ (possion implies in $\mathbb{V}_+$). Using our formula, we have:

$$f_{X+Y}(z) = \sum_{x=0}^{z} f_X(x) \cdot f_Y(z-x), \quad f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad f_Y(z-x) = e^{-\lambda} \cdot \frac{\lambda^{z-x}}{(z-x)!}$$

putting these back into our sum, we have

$$f_{X+Y}(z) = \sum_{x=0}^{n} e^{-2\lambda}\frac{\lambda^z}{x!(z-x)!} = e^{-2\lambda}\lambda^z \sum_{x=1}^{z} \frac{1}{x!(z-x)!}$$

$$= e^{-2\lambda}\lambda^z \frac{1}{z!} \sum_{x=1}^{z} \frac{z!}{x!(z-x)!}$$

and since $\sum \frac{z!}{x!(z-x)!} = \sum z^C x$ which is the binomial expansion where $y = x = 1$, or in other words: $(1+1)^z = 2^z$, so our sum is equal to

$$e^{-2\lambda}\frac{(2\lambda)^z}{z!}$$

3.7 Homework Questions

3.7.1 Chapter 3, Question 15

X, Y are the same independent random variables as before. We would like to calculate the densities of the following:

1. min(X,Y)
2. max(X,Y)
3. $|Y - X|$

Firstly, let

$$Z = min(X, Y) : 0...N, \quad \text{and let } y(z) = \{Z = z\}$$

we will operate under the assumption that

$$P\{Z > k\} \text{ where } k \in 0, 1, ...M$$

and this means that

$$P\{Z > k\} = P\{X \geq k, Y \geq k\}$$

since we know these events are independent, we can multiply them to get

$$P(X \geq k)P(Y \geq k) \text{ which are the same, so this equals } P(X \geq k)^2 = \frac{(N-k+1)^2}{(N+1)^2}$$

notice that

$$\{Z = z\} \cup \{Z \geq z+1\} = \{Z \geq z\} \text{ where the first two probabilities are disjoint}$$

from which we get:

$$P\{Z = z\} = P\{Z \geq z\} - P\{Z \geq z+1\}$$

For $z = 0, 1, ...N-1$,

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

So when $g(N) = P\{Z \geq n\} = \frac{1}{(1+n)^2}$.

Now for the maximum, call $W = max(X, Y) : 0, 1, ...N$. Looking at

$$P(W \leq k) = P(X \leq k, Y \leq k)$$

we see this is equal to:

$$P(X \leq k)P(Y \leq y) = P(X \leq k)^2$$

where

$$P(X \leq k) = 1 - P(X \geq k+1)$$

using that formula that we have, we know this equals

$$1 - \frac{N-k}{N+1} = \frac{1+k}{N+1} \text{ which when squared is equal to: } \frac{(k+1)^2}{(N+1)^2}$$

so,

$$P(X \leq k)^2 = P(W = k) = \frac{(k+1)^2}{(N+1)^2}$$

We had to make the adjustment;

$$\{W = k\} \cup \{W \leq k-1\} = \{W \leq k\}$$

From this we get that

$$P\{W = k\} = P\{W \leq k\} - P\{W \leq k-1\}$$

which equals

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

Which is equal to

$$\frac{2k+1}{(N+1)^2}$$

Thirdly, we have

$$T = |Y - X| : 0, ...N$$

if we call t one of these values, we have

$$T = t \iff |Y - X| = t \iff Y - X = t \text{ or } Y - X = -t$$

which we can rephrase as

$$Y - X = t \text{ or } X - Y = t$$

Suppose that $t = 1, 2, ...N$. This means that these two joints are then disjoint. which tells us that

$$P\{T = t\} = P\{Y - X = t\} + P\{X - Y = t\}$$

These two probabilities are equal, which tell us the sum is equal to

$$2P\{Y - X = t\}$$

We also need to consider that $t = 0$, which would mean that

$$P\{T = 0\} = P\{X = Y\} = \text{ done in question 14 part (b)}$$

calculating the probability,

$$\{Y - X = t\} = \bigcup_{x=0}^{N} \{Y - X = t, X = x\} = \bigcup_{x=0}^{N} \{Y - x = t, X = x\} = \bigcup_{x=0}^{N-t} \{Y = t + x, X = x\}$$

from which we know

$$P\{Y - X = t\} = \sum_{x=0}^{N-t} P(Y = t + x)P(X = x) = \frac{1}{N+1} \sum_{x=0}^{N-t} P(Y = t + x) = \frac{N - t + 1}{(N+1)^2}$$

thus,

$$P(T = t) = \left| \begin{array}{c} \frac{2(N-t+1)}{(n+1)^2} \text{for } t : 1, 2, ...N \\ 14\text{(b) if } t = 0 \end{array} \right.$$

3.7.2 Chapter 3, Question 16

Given X, Y, we have

$$X = G(p_1), \quad Y = G(p_2)$$

Last time we said that

$$P(X \geq Y) = \bigcup_{y=0}^{n} (X \geq Y, Y = y) = \bigcup_{y=0}^{N} (X \geq y, Y = y)$$

$$\Rightarrow P(X \geq Y) = \sum_{y=0}^{N} P(X \geq y)P(Y = y)$$

and since

$$P(X \geq y) = q_1^y \qquad P(Y = y) = p_2 q^y$$

so we can write this as

$$= p_2 \sum_{y=0}^{N} (q_1 q_2)^y = p_2 \sum_{y=0}^{N} q^y$$

and using the following formula

$$1 + q + q^2 + q^3 + q^4 + ... + q^k = \frac{1 - q^{k+1}}{1 - q}$$

we get

$$= p_2 \frac{1 - q^{N+1}}{1 - q}$$

looking at part b,

$$(X = Y) = \bigcup_{y=0}^{N} (X = Y, Y = y) = \bigcup_{y=0}^{N} (X = y, Y = y)$$

$$P(X = Y) = \sum_{y=0}^{N} p_1 q_1^y p_2 q_2^y = p_1 p_2 \sum_{y=0}^{N} q^y \qquad {}^*$$

which gives us a similar answer of

$$p_1 p_2 \frac{1 - q^{N+1}}{1 - q}$$

3.7.3 Chapter 3, Question 18

Assume that

$$P(X = x, Y = y) = g(x)h(y) \text{ for all x,y} \quad \text{(this is the joint density, h(x,y))}$$

we know that †

$$f_x(x) = \sum_{all\ \ y} h(x, y) = \sum_{all\ \ y} g(x)h(y) = g(x) \sum_{all\ \ y} h(y) = g(x)A$$

looking at part b,

$$P(Y = y) = f_Y(y) = \sum_{all\ \ x} h(x, y) = \sum_{all\ \ x} g(x)h(y) = h(y) \sum_{all\ \ x} g(x) = h(y)B$$

so in summary,

$$f_X(x) = Ag(x), \quad f_Y(y) = Bh(y)$$

Now,

$$1 = \sum_{all\ \ x,\ all\ \ y} h(x, y) = \sum_{all\ \ x\ \ all\ \ y} g(x)h(y) = \left(\sum_{all\ \ x} g(x)\right)\left(\sum_{all\ \ y} h(y)\right) = AB = 1$$

---

* $q = q_1 q_2$

† recall that by convention h is the joint function. The notation here is poor, but makes sense.

Now, multiplying

$$f_X(x)f_Y(y) = ABg(x)h(y) = 1g(x)h(y) = h(x, y)$$

from which, we can conclude that X, Y are independent. Notice that both g and h are constants times the marginal, since:

$$f_X(x)/A = g(x)$$

3.8 The Conclusion of Chapter 3

Recall that if X, Y are independent, the density of the sum $X + Y$ is:

$$f_{X+Y}(z) = \sum_{all\ \ x} f_X(x)f_Y(z - x)$$

Now, for $X \in \mathbb{IV}_+$, we introduce the function $\Phi_x(t)$ for $t \in [-1, +1]$ called **the probability generating function** of the random variable X. It is defined as follows:

$$\Phi_x(t) = \sum_{x=0}^{\infty} P(X = x)t^x = \sum_{x=0}^{\infty} f_X(x)t^x$$

which is a power series in the variable t. Observe that if $|t| \leq 1$, we can look at

$$\sum_{x=0}^{\infty} f_X(x)|t|^x \leq \sum_{x=0}^{\infty} f_X(x) \text{ since } |t| \leq 1.$$

and since

$$\sum_{x=0}^{\infty} f_X(x) = 1$$

we know that the first series must converge from $|t| \leq 1$.

We know the following about power series:

$$\sum_{x=0}^{\infty} a_x t^x = \sum_{x=0}^{\infty} b_x t^x \quad \forall t \in (-\epsilon, \epsilon)$$

implies that all $a_x = b_x$ for $x = 0, 1, ...$
Example. Suppose that X is binomial(n,p). Lets find $\Phi_x(t)$. Well,

$$f_X(x) = \binom{n}{x} p^x q^{n-q} \quad x = 0, 1, 2, ...n$$

so looking at

$$\Phi_x(t) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} t^x = \sum_{x=0}^{n} \binom{n}{x} (pt)^x (q)^{n-x}$$

which is a binomial expansion, which is

$$= (pt + q)^n \text{which comes from the binomial formula}$$

Example. Suppose that $X = Poisson(\lambda)$. Find $\Phi_x(t)$. Well again,

$$f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad \forall x = 0, 1, 2, ...$$

So we know that

$$\Phi_X(t) = \sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!}t^x$$

we can simplify this series to:

$$= e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda t)^x}{x!} = e^{-\lambda}\cdot e^{\lambda t} = e^{\lambda(t-1)} \quad \forall t \in \mathbb{R}$$

For homework, find the probability generating function for $X = Geom(p)$.

We have a few theorems about the probability generating function.

Theorem 6. *Suppose that $X, Y \in \mathbb{IV}_+$ and suppose that $\Phi_X(t) = \Phi_Y(t) \;\forall\; |t| \le 1$. Then,*

$$f_X(x) = f_Y(x) \text{ for all } x \quad (f_X \equiv f_Y)$$

Proof. We know that

$$\sum_{x=0}^{\infty} f_X(x)t^x = \sum_{x=0}^{\infty} f_Y(x)t^x$$

and based on what we know about power series, we know that the coefficients must be equal- thus,

$$f_X(x) = f_Y(x)$$

□

Theorem 7. *If $X, Y \in \mathbb{IV}_+$ and independent, then $\Phi_{X+Y}(t) = \Phi_X(t)\cdot\Phi_Y(t)$ This results easily extends to three or more independent variables. The proof of this is based on the convolution, and is in the book.*

The following is an example of how we can combine both of these theorems and get something interesting.

Example. Let X, Y, where $X = Poisson(\alpha)$, $Y = Poisson(\beta)$ where X, Y are independent. Look at $X + Y$:

$$\Phi_{X+Y}(t) = \Phi_X(t)\cdot\Phi_Y(t) = e^{\alpha(t-1)}e^{\beta(t-1)} = e^{(\alpha+\beta)(t-1)} = \Phi_W(t)$$

where

$$W = Poisson(\alpha + \beta)$$

By theorem number 6, we know that

$$X + Y = Poisson(\alpha + \beta)$$

3.9 Applications of the Theorem and the Probability Generating Function

The following is theorem one on page (75).

3.9.1 Finding $f(x)$ from $F(x)$

Assume that X is a discrete random variable with finitely many values listed in increasing order:

$$x_0 < x_1 < x_2 ... < x_n$$

Suppose that $F(x)$, the cumulative distribution function of X is given. Find the density $f(x)$. We know that

$$F(x) = \sum_{\text{all } t \leq x} f(x)$$

So this tells us that

$$\begin{aligned} F(x_1) &= f(x_1) \\ F(x_2) &= f(x_1) + f(x_2) \\ F(x_3 &= f(x+1) + f(x_2) + f(x_3 \\ &\vdots \end{aligned}$$

subtracting, we can get:

$$\begin{aligned} f(x_1) &= F(x_1) \\ f(x_2) &= F(x_2) - F(x_1) \\ f(x_3) &= F(x_3) - F(x_2) \\ &\vdots \\ f(x_n) &= F(x_n) - F(x_{n-1}) \\ &\vdots \end{aligned}$$

3.9.2 Chapter 3, Question 17

Given $X = G(p_1), Y = G(p_2)$. Looking at

$$Z = min(X, Y) : 0, 1, 2, ... \quad \text{and call the value of the minimum 'z'}$$

Notice that $P(Z \geq z) = P(X \geq z, Y \geq z)$, and since these events are independent (one depends on X, the other on Y) so we can calculate their probabilities via multiplication

$$P(Z \geq z) = P(X \geq z, Y \geq z) = P(X \geq z)P(Y \geq z) = q_1^z q_2^z$$

since $P(X \geq z) = q^z$. Which tells us that this this variable must be the same as the following random variable:

$$Z = Geometric(p = 1 - q = 1 - (1-p_1)(1-p_2)) = 1 - [1 - p_2 - p_3 + p_1 p_2] = p_1 + p_2 - p_1 p_2)$$

Secondly

$$W = X + Y : 0, 1, ...$$

using the convolution formula:

$$f_W(w) = \sum_{x=0}^{w} f_x(x) f_y(w-x) = \sum_{x=0}^{w} p_1 q_1^x p_2 q_2^{w-x} = p_1 p_2 q^w \sum_{x=0}^{w} \left(\frac{q_1}{q_2}\right)^x$$

so we say that say that $\frac{q_1}{q_2} = Q$ We have now two cases:

$$q_1 = q_2 \qquad q_1 \neq q_2$$

the first tells us that $Q = 1$, which would tell us that $p_1 = p_2 = p$, and in this case

$$\sum_{x=0}^{w} (1) = w + 1 \Rightarrow f_W(w) = (w+1) p^2 q^w$$

In the other case, we know that $Q \neq 1$, and as a result,

$$\sum_{x=0}^{w} Q^x = 1 + Q + Q^2 + Q^3 + ... = \frac{Q^{w+1} - 1}{Q - 1} = \frac{1 - Q^{w+1}}{1 - Q}$$

and in this case,

$$f_W(w) = p_1 p_2 q_2^w \frac{\frac{q_1}{q_2}^2 - 1}{\frac{q_1}{q_2} - 1}$$

3.9.3 Chapter 3, Question 31

Suppose that X, Y are independent, and uniform from $\{1, ...N\}$. Let

$$Z = X + Y : 2, 3, ...2N$$

since these are non-negative integer values, we can use the convolution formula, which says:

$$f_{X+Y} = \sum_{x=0}^{z} f_X(x) f_Y(z-x) = \sum_{x=1}^{z} f_X(x) f_Y(z-x)$$

we want to take into account all the cases in which $f_X(x) \neq 0$, and same for f_Y. So for a fixed $z \in \{2, 3, ...2N\}$ and and integer $1 \leq x \leq z$ we should look at

$$f_X(x) f_Y(z-x) \neq 0 \iff 1 \leq x \leq n \text{ and } 1 \leq z - x \leq N$$

solving these inequalities in terms of x, we see that these equations tells us:

$$A = max(1, z - N) \leq x \leq min(N, z - 1) = B$$

so we have the sum that

$$\sum_{x=A}^{B} \frac{1}{N^2} = \frac{B - A + 1}{N^2}$$

which can actually be broken down into two cases, the first of which is that

$$2 \leq z \leq N$$

in this case, $max(1, z - N) = 1$, and $A = 1$. Also, we see that $min(N, z - 1) = z - 1 = B$, so

$$f_Z(z) = \frac{z - 1}{N^2}$$

and in the case where

$$N + 1 \leq z \leq 2N$$

which tells us that $max(1, z - N) = z - N = A$ and $min(N, z - 1) = N = B$. So in this case,

$$f_Z(z) = \frac{2N - 1}{N^2}$$

so our final answer is

$$f_{X+Y}(z) = \begin{cases} \frac{z-1}{N^2} \text{ when } z \in 2, ...N \\ \\ \frac{2N-1}{N^2} \text{ when } z \in N + 1...2N \end{cases}$$

3.9.4 Chapter 3, Question 32

We are looking for the probability generating function $\Phi_x(t)$ for $f_X(x) = \frac{1}{N+1}$

$$\Phi_x(t) = \sum_{x=0}^{N} f_X(x)t^x = \frac{1}{N+1} \sum_{x=0}^{N} t^x \text{ and we know } \sum_{x=0}^{N} t^x = 1 + t + t^2 + ... = \frac{1 - t^{N+1}}{1 - t}$$

so we can say that our answer is 1 if $t = 1$, and $\frac{1-t^{N+1}}{(1-t)(N+1)}$ elsewhere.

3.9.5 Chapter 3, Question 33

Given $X = \mathbb{IV}_+$ with $\Phi_X(t) = e^{\lambda(t^2-1)}, \lambda > 0$. From calculus, we know how to represent e^x for any x. So, we can say that:

$$\sum_{x=0}^{\infty} f_X(x)t^x = \Phi_X(t) = e^{-\lambda}e^{\lambda t^2} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^n t^{2n}}{n!}$$

so,

$$f_X(o)t^0 + f_X(1)t + f_X(2)t^2 + ... = e^{-\lambda}\left[1 + \frac{\lambda t^2}{1!} + \frac{\lambda^2 t^4}{2!}\right] = e^{-\lambda} + \frac{\lambda e^{-\lambda}}{1!}t^2 + \frac{\lambda^2 e^{-\lambda}}{2!}t^4 + ...$$

thus, since the subscripts have to match,

$$\begin{aligned} f_X(0) &= e^{-\lambda} \\ f_X(1) &= 0 \\ f_X(2) &= \frac{\lambda e^{-\lambda}}{1!} \\ f_X(3) &= 0 \\ &\vdots \end{aligned}$$

3.9.6 Exam 1

Our First exam will be on Monday, October 17^{th}.

Chapter 4

Expected Value of Discrete Random Variables

Suppose you have

$$X \to x_1, x_2, ...$$

if you look at

$$E(X) := \sum_{i=1}^{\infty} x_i f(x_i) \text{ provided that } \sum_{i=1}^{\infty} |x_i| f(x_i) < \infty$$

you get something called "the expected value" of X. From calculus, we know that if we have an absolutely convergent series, it must converge. Recall that the opposite is not true.

Example. Suppose that $X : 1, 2, ..n$ and $f_X(x) = \frac{1}{n}$. What is $E(X)$? Well,

$$E(X) = \sum_{x=1}^{n} x \cdot \frac{1}{n} = \frac{n+1}{2}$$

since

$$\sum_{x=1}^{n} = 1 + 2 + 3 + ... + n = \frac{n(n+1)}{2}$$

Now suppose that $X : a_1, a_2, ...a_n$. In this case, we have the same formula over a uniform random variable as before, and get:

$$E(X) = \frac{a_1 + a_2 + ... + a_n}{n}$$

which is intuitively what we think of in terms of an average. We often use μ to denote this value.

Example. Let $X = Binomial(n, p)$. Where $x = (0, 1, ...n)$ which tells us that

$$f_X(x) = \binom{n}{x} p^x q^{n-x}$$

so

$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} = np$$

Example. Let

$$X = Poisson(\lambda) \quad f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

and notice that

$$E(X) = \sum_{x=0}^{\infty} x \cdot e^{-\lambda}\frac{\lambda^x}{x!} = ... = \lambda$$

Example. Let

$$X = Geom(p) \Rightarrow E(X) = \sum_{x=0}^{\infty} xpq^x = ... =?$$

it is a good idea to go and prove these things.

4.1 The Connection between Probability Generating Functions and Expected Values

The connection between $\Phi_X(t)$ and $E(X)$ if $X = \mathbb{IV}_+$ is as follows: recall that we know the following:

$$\Phi_X(t) = \sum_{x=0}^{\infty} f_X(x)t^x \quad \text{converges for } t \in [-1, 1]$$

taking the derivative with respect to t,

$$\Phi_X(t)' = \sum_{x=0}^{\infty} xf_X(x)t^{x-1}, \quad \text{letting t=1}, \quad \Phi_X(1)' = \sum_{x=0}^{\infty} xf_X(x) = E(X)$$

4.1.1 Chapter 4, Question 5

Let us first list the properties of the expected values:

1. If $X = c$, a constant, the expected value is $E(X) = c$.
2. $E(kX) = kE(X)$
3. $E(X + Y) = E(X) + E(Y)$
4. $E(c_1X_1 + c_2X_2 + ...c_nX_n) = c_1E(X_1) + c_2E(X_2) + ... + c_nE(X_n)$
5. If $X \geq Y$, then $E(X) \geq E(Y)$
6. $|E(X)| \leq E|X|$

Theorem 8. *Suppose you have*

$$\mathbb{X} = (X_1, ...X_r)$$

where the X_i are discrete random variables, and you have

$$\hat{f}(x) \textit{ as its joint density}$$

Let $\varphi : \mathbb{R}^r \to \mathbb{R}$. Let

$$Z = \varphi(\mathbb{X})$$

The expected value of Z looks like the following:

$$E(Z) = \sum_{all\ x} \varphi(x)\hat{f}(x)$$

provided if and only if

$$\sum_{all\ x} |\varphi(x)|\hat{f}(x) < \infty$$

This theorem and its proof are discussed well in the book.

We can use this result for our problem. Suppose you allowed $r = 2$. Then,

$$\mathbb{X} = (X_1, X_2)$$

and we let

$$\varphi(x_1, x_2) = x_1 + x_2$$

in this case,

$$Z = X_1 + X_2$$

applying the theorem, we know that

$$E(Z) = \sum_{\text{over all } x_1, x_2} (x_1 + x_2)\hat{f}(x_1, x_2) = \sum_{\text{over all } x_1, x_2} [x_1 f(x_1, x_2) + x_2 f(x_1, x_2)]$$

which we know to be equal to

$$\sum_{\text{over all } x_1, x_2} x_1 f(x_1, x_2) + \sum_{\text{over all } x_1, x_2} x_2 f(x_1, x_2)$$

which we will call,

$$A + B$$

Looking at A, we see that

$$A = \sum_{\text{over all } (x_1)} \left(\sum_{\text{all } (x_2)} x_1 f(x_1, x_2) \right) = \sum_{\text{over all } (x_1)} x_1 \left(\sum_{\text{all } (x_2)} f(x_1, x_2) \right)$$

recognizing the right is the marginal $f_{X_1}(x_1)$, we see

$$A = \sum_{\text{all } (x_1)} x_1 f_{X_1}(x_1) = E(X_1)$$

similarly, $B = E(X_2)$. Notice that as a result of our theorem, letting $r = 1$, we have X, and we allow our function function φ to be $\varphi : \mathbb{R} \to \mathbb{R}$

$$Z = \varphi(x)$$

so

$$E[\varphi(X)] = \sum_{\text{all x}} \varphi(x) f_X(x)$$

another important theorem is:

Theorem 9. *If $|X| \leq M$, some positive constant, the random variable is said to be* ***bounded****. Then, $E(X)$ exists, and*

$$|E(X)| \leq M$$

this can be found on page 88, theorem 3 in the book. The book instead writes that

$$P(|X| \leq M) = 1$$

which is just more appropriate. The two notions are equivalent.

Working back to our question, suppose that X, Y are two random variables such that

$$P(|X - Y|) \leq M = 1$$

which is the same as

$$|X - Y| \leq M$$

for some constant M. We want to show that if Y has finite expectation, then X has finite expectation and $|E(X) - E(Y)| \leq M$. We write $X = (X - Y) + Y$, and from our theorem, we know that $E(X - Y)$ exists, since $X - Y$ is bounded. By assumption $E(Y)$ exists, so we know that X must have $E(X)$ finite. From this, we proceed, noting that

$$E(X - Y) = E(X) - E(Y)^*$$

from theorem (3) in the book on page 88, we can say that

$$|E(X - Y)| \leq M \quad \text{or} \quad |E(X) - E(Y)| \leq M$$

and we are done.

4.1.2 Chapter 4, Question 2

X is of binomial density with parameters $n = 4$ and we have p. We know that

$$X : 0, 1, 2, 3, 4$$

so

$$f(X) = \binom{4}{x} p^x (1-p)^{4-x}$$

Knowing that we want to calculate, $E(sin(\pi X/w))$ we define the function $\varphi(x) = \sin(\pi x/2)$. From one of our theorems,

$$E[sin(\pi X/w)] = \sum_{x=0}^{4} (sin(\pi x/2)) \binom{4}{x} p^x q^{4-x}$$

working through the finite cases, we get the sum

$$\binom{4}{1} pq^3 - \binom{4}{3} p^3 q = 4pq(q^2 - p^2)$$

It turns out that this question is relatively easy once you use that formula when $r = 1$, as we have now done twice.

*to write this, we needed to know that E(X) is finite

4.1.3 Chapter 4, Question

Let X be Poisson with parameter λ, compute the mean of $(1+X)^{-1}$. In other words, find

$$E(\frac{1}{1+X})$$

again using our formula, we say that

$$E(\frac{1}{1+X}) = \sum_{x=0}^{\infty} \varphi(x) e^{-\lambda} \frac{\lambda^x}{x!}$$

by making $\varphi(x) = \frac{1}{1+x}$. Combining $\varphi(x)$ with the $(x!)$ term in our sum, we get

$$= \frac{1}{\lambda} \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^{x+1}}{(x+1)!}$$

changing the notation, we call $x+1=n$. We then have

$$= \frac{1}{\lambda} \sum_{n=1}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} = \frac{e^{-\lambda}}{\lambda}(e^{\lambda}-1) = \frac{1-e^{-\lambda}}{\lambda}$$

If you want, try to see if this can be extended, and find $E(\frac{1}{2+x})$.

4.1.4 An Application of the Probability Generating Function

Last time, we saw a theorem that said if $X \in \mathbb{IV}_+$, this tells us that $E(X) = \Phi'_X(1)$. Let $X = Geometric(p)$. Let us try to now find $E(X)$ using this method. First, we need the probability generating function. This can be done as follows:

$$\Phi_X(t) = \sum_{x=0}^{\infty} f_X(d) t^x = \sum_{x=0}^{\infty} pq^x t^x = \sum_{x=0}^{\infty} p(qt)^x$$

calling $qt = r$, we have

$$= p \sum_{x=0}^{\infty} r^x \text{ which converges to: } = \frac{1}{1-r} \text{ if the absolute value of r is less than one}$$

so, this is doable if

$$q|t| < 1 \iff |t| < \frac{1}{q} \iff -\frac{1}{q} < t < \frac{1}{q}$$

so from this, our probability generating function is

$$\Phi_X(t) = \frac{p}{1-qt}$$

where the geometric function is defined from $\frac{-1}{q}$ to $\frac{1}{q}$. In trying to find the expected value, we take the derivative of our probability generating function:

$$\Phi'_X(t) = [p(1-qt)^{-1}]' = pq(1-qt)^{-2}$$

and

$$\Phi_X''(t) = 2pq^2(1-qt)^{-3}$$

so, looking at the derivative at 1, we have

$$\Phi_X'(1) = pq(1-q)^{-2} = E(X)$$

from this, we can get

$$E(X) = \frac{q}{p}$$

4.2 Moments of a Random Variable

Given X, and some integer $r \geq 1$. We would like to know $E(X^r)$. By the formula we have,

$$\sum_{\text{all x}} x^r f_X(x)$$

which is known as 'the moment of X of order r', or the r^{th} moment.
Theorem 10. *If the r^{th} moment is finite, then all the previous moments were all finite.*

Recall that $\mu := E(x)$. Let us assume that $E(X^2)$ is finite. We have

$$Var(X) := E[(X-\mu)^2]$$

How do you calculate the variance directly from this formula? Looking for a function $\varphi(x)$, we have $\varphi(x) = (x-\mu)^2$, and have the following for the expected value:

$$= \sum_{\text{all x}} (x-\mu)^2 f_X(x)$$

we do have the following formula:

$$Var(X) = E(X^2) - \mu^2$$

This follows from:

$$Var(X) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

Notice that

$$Var(X) \geq 0$$

since it is the expected value of a square. This tells us the following corrolary:
Corollary 11.

$$(E(X))^2 \leq E(X^2)$$

We should note the following about the variance:

1. Recall that if $X = 0$, then $Var(X) = 0$.

2. We say that by definition,
$$\sigma^2 = Var(x)$$
where σ denotes the standard deviation.

3. We know that $Var(aX + b) = a^2 Var(X)$

4. Generally speaking, $Var(X + Y) \neq Var(X) + Var(Y)$.

We now need to talk about the Variance of two random variables. Look at the variance of the sum, $Var(X + Y)$. We will call

$$\mu_x = E(X), \qquad \mu_y = E(Y)$$

we should notice that

$$Var(X + Y) = E(X + Y)^2 - (E(X + Y))^2$$

doing some algebra, we see that this is equal to

$$E(X^2 + 2XY + Y^2) - \left(\mu_x^2 + 2\mu_x\mu_y + \mu_y^2\right)$$

From which we say:

$$= E(X^2) + 2E(EY) + 2E(Y^2) - \mu_x^2 - 2\mu_x\mu_y - \mu_y^2$$

grouping things together, we get

$$= Var(X) + Var(Y) + 2(E(XY) - \mu_x\mu_y) = Var(X) + Var(Y) + 2(E(XY) - E(X)E(Y))$$

where
$$E(XY) - E(X)E(Y) := Cov(X, Y)$$
the covariance of X, Y. From this we, have the following theorem:
Theorem 12. *If X, Y are independent, then*

1. $E(XY) = E(X)E(Y)$

2. $Var(X + Y) = Var(X) + Var(Y)$

Recall that if X, Y are independent, $Var(X+Y) = Var(X)+Var(Y)$ which follows from the fact that if these variables are independent, $Cov(X, Y) = 0$, which implies that $Var(X+Y) = Var(X) + Var(Y)$. This last line can be shown by

$$E(X, Y) = \sum_{\text{all x, ally}} xyf(x, y)$$

which follows from letting $\varphi(x, y) = xy$, and by independence, we know that this is equal to:

$$\sum_{\text{all x, all y}} xyf_X(x)f_Y(y) = \sum_{\text{all x, all y}} (xf_X(x))(yf_Y(y)) = \left(\sum_{\text{all x}} xf_X(x)\right)\left(\sum_{\text{all y}} yf_Y(y)\right) = E(X)E(Y)$$

4.2.1 Announcements

On October 24^{th}, class is canceled. It will be made up on Dec. 14^{th}. Also, on the 19^{th}, please come early, at $5:45$.

4.2.2 Chapter 3, Question 13

Let X be a nonnegative integer valued random variable. We know that this means that

$$E(\varphi(X)) = \sum_{\text{all x}} \varphi(x) f_X(x)$$

which can lead to the following question: what is $E(X(X-1))$? Well, this would be the following sum :

$$\sum_{\text{all x}} x(x-1) f_X(x) \quad \text{which follows from letting } \varphi(X) = x(x-1)$$

So now looking at $\Phi_X(t)$,

$$\Phi_x(t) = \sum_{x=0}^{\infty} f_X(x) t^x$$

Taking the derivative of this function, (with respect to t)

$$\Phi'_X(t) = \sum_{x=0}^{\infty} x f_X(x) t^{x-1}$$

and taking its second derivative,

$$\Phi''_X(t) = \sum_{x=0}^{\infty} x(x-1) t^{x-2}$$

plugging $t = 1$ in here, you have:

$$\Phi''_X(1) = E[X(X-1)] = E[X^2 - X] = E(X^2) - E(X)$$

also recall that the $E(X)$ for a variable $X \in \mathbb{IV}_+$ is $\Phi'_X(1) = E(X)$, so we have the formula

$$E(X^2) = \Phi'_X(1) + \Phi''_X(1)$$

also recall that

$$Var(X) = E(X^2) - (E(X))^2$$

plugging in what we have,

$$= \Phi'_X(1) + \Phi''_X(1) - [\Phi'_X(1)]^2$$

which works for positive integer valued random variables. Going back to the question and our formula, if we let $\varphi(x) = t^x$ for a fixed $t \in [-1, 1]$ what our formula is really giving us is

$$E(t^X) = \sum_{\text{all x}} f_X(x) t^x = \Phi_X(t)$$

so, $\Phi_X(t) = E(t^X)$. The other parts are also similarly done. Before we do part (b), let us notice the following lemma:

Lemma 13. *Suppose two random variables X, Y are independent. Then, any function $\alpha(X)$ and $\beta(Y)$ are independent. More generally, suppose that you have many random variables that are independent, $X_1, X_2, ..., X_k, X_{k+1}, ..., X_l..X_n$ Thinking of these independent variables in terms of three groups (separated by the '...') then our functions*

$$\alpha(X_1, ..., X_K), \quad \beta(X_{k+1}, .., X_l) \quad \gamma(X_{l+1}...X_n)$$

are independent.

Back to our question, suppose that $X, Y \in \mathbb{IV}_+$ and are independent.

$$\Phi_{X+Y}(t) = E(t^{X+Y}) \qquad \dagger$$

now algebra tells us this is equal to:

$$E(t^X t^Y)$$

which must be independent, since t^X and t^Y are functions as discussed in our above lemma. Since these are independent, we know that:

$$E(t^X t^Y) = E(t^X)E(t^Y)$$

and from a result we just had, this is equal to:

$$= \Phi_X(t)\Phi_Y(t)$$

4.2.3 Chapter 3, Question 9

We want to construct an example of a density that has a finite moment of order r but has no higher finite moment. Let us take some integer $r > 0$. Look at the series:

$$\sum_{x=1}^{\infty} \frac{1}{x^{r+2}}$$

which clearly converges, for if we had some infinite sum:

$$\sum \frac{1}{x^n}$$

which we know converges for $n > 1$, otherwise it diverges.

Let us take some random variable $X : 1, 2, 3, ...$ and let

$$1 < c = \sum_{x=1}^{\infty} \frac{1}{x^{r+2}}$$

now we call

$$f(x) = \frac{1}{cx^{r+2}}$$

† t is again in that same interval

from which we can get

$$\sum_{x=1}^{\infty} f(x) = \frac{1}{c}\sum_{x=1}^{\infty}\frac{1}{x^{r+2}} = 1$$

now the expected value with this sum and these values, letting $\varphi(x) = x^r$ we have:

$$E(X^r) = \sum_{x=1}^{\infty} x^r f_X(x) = \frac{1}{c}\sum_{x=1}^{\infty}\frac{1}{x^2} < \infty$$

now looking at the next moment, using the same formula only using $\varphi(x) = x^{r+1}$, we have:

$$E(X^{r+1}) = \sum_{x=1}^{\infty} x^{r+1} f_X(x) = \frac{1}{c}\sum_{x=1}^{\infty}\frac{1}{x} = \infty$$

thus, all future moments must be infinite.

4.2.4 Chapter 3, Question 6

Let X be a geometrically distributed random variable, and let $M > 0$ be a positive integer. Let

$$Z = min(X, M) : 0, 1, 2, ..M$$

and compute the mean of Z. We know that

$$E(Z) = \sum_{z=0}^{M} z f_Z(z)$$

note that if $0 \leq z < M$, saying that $Z = z$ would be to say that $X = z$. And to say that $z = M$, meaning that $Z = M$, this would mean that $X \geq M$. So, we can split the expected value into the sum:

$$E(Z) = \sum_{z=0}^{M-1} z f_Z(z) + M f_Z(m) = \sum_{z=0}^{M-1} zP(X = z) + M \cdot P(X \geq M)$$

using formulas, we recall that the right hand term is equal to q^M and the left hand term is pq^z. So

$$E(Z) = p\sum_{z=0}^{M-1} zq^z + Mq^m$$

however, we would like to reduce this sum. We can do this by noticing the following:

$$\sum_{n=0}^{N} t^n = 1 + t + t^2 + t^3 + ... + t^N \text{ if } t \neq 1, = \frac{1 - t^{n+1}}{1 - t} = u(t)$$

taking the derivatives on both sides, we have

$$1 + 2t + 3t^2 + ...Nt^{N-1} = u'(t) \quad \text{when } t \neq 1$$

now multiplying both sides by t, we have

$$t + 2t^2 + 3t^3 + ... + Nt^N = t \cdot u'(t) = \sum_{z=0}^{N} zt^z$$

which can then be applied to our answers, which we may not need. This is where we will leave this question.

4.2.5 Chapter 3, Question 21

Suppose that X, Y are random variables such that $\rho(X, Y) = .5$. The is called the **correlation coefficient**, and it is defined as:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{\sigma_x \sigma_y}}$$

and using what we have in the question,

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{\sigma_x \sigma_y}} = \frac{1}{2}$$

and

$$Var(X) = 1 \qquad Var(Y) = 2$$

so, we want to find $Var(X - 2Y)$. the question does **not** tell us that these variables are independent- in fact, they must not be. We proceed in the following way:

$$Var(X - 2Y) = Var(X) + Var(-2Y) + 2Cov(X, -2Y)$$

and since

$$Var(-2Y) = 4Var(Y) = 4(2) = 8$$

and since we have

$$2Cov(X, -2Y) = E(-2XY) - E(X)E(-2Y) = 2E(X)E(Y) - 2E(XY) = -2Cov(X, Y)$$

now since

$$\sigma_x = 1 \qquad \sigma_y = \sqrt{2}$$

we know that

$$Cov(X, Y) = \frac{\sqrt{2}}{2}$$

plugging in,

$$-2Cov(X, Y) = -\sqrt{2}$$

so we see that

$$Var(X - 2Y) = 1 + 8 - 2\sqrt{2}$$

4.3 Chebyshev's Inequality

Chebyshev's Inequality is the following: suppose you had some random variable X with a finite second moment, $E(X^2) < \infty$. This tells us that the variance is finite. Call the expected value of X, $E(X) = \mu$. The inequality looks at the following:

$$P\{|X - \mu| \geq c\} \leq \frac{Var(X)}{c^2}$$

note that this event can be written as the following union:

$$\{|X - \mu| \geq c\} = \{X \leq \mu - c\} \bigcup \{X \geq \mu + c\}$$

so we see that

$$P\{|X - \mu| \geq c\} = P(X \leq \mu - c) + P(X \geq \mu + c)$$

The following is called **Markov's Inequality**: If we have $X \geq 0$, where X is discrete, and if $a > 0$, if $E(X)$ is finite, then

$$P\{X \geq a\} \leq \frac{E(X)}{a}$$

The questions (26-30) belong to these concepts, and should be looked at before next Monday.

4.4 Markov's Inequality

Theorem 14. *let $X \geq 0$ be a discrete random variable with finite expected value. Let $a > 0$. Then,*

$$P(X \geq a) \leq \frac{E(x)}{a}$$

this is Markov's Inequality, and we will now prove it.

Proof. We know that

$$E(X) = \sum_{\text{all x}} xf(x) \qquad \text{converges by assumption}$$

We break this into the following two sums:

$$= \sum_{\text{all } x \geq a} xf(X) + \sum_{\text{all } x < a} xf(x)$$

note that all these terms are non-negative. Continuing, it is clear that:

$$= \sum_{\text{all } x \geq a} xf(X) + \sum_{\text{all } x < a} xf(x) \geq \sum_{\text{all } x \geq a} xf(x)$$

Notice that

$$x \geq a \Rightarrow xf(x) \geq af(x)$$

so, the sum has the following property:

$$\sum_{\text{all } x \geq a} xf(x) \geq a \sum_{\text{all } x \geq a} f(x) = aP(X \geq a)$$

looking at the ends of this sequence, we have:

$$aP(X \geq a) \leq E(X) \quad \Rightarrow \quad P(X \geq a) \leq \frac{E(X)}{a}$$

□

Example. Say that we know that $E(X) = 2$, where $X \geq 0$. We know that

$$P(X \geq 6) \leq \frac{2}{6} = \frac{1}{3}$$

4.5 Chebyshev's Inequality

Theorem 15. *If X is a discrete random variable with a finite second moment, finite mean, and finite variance, if $c < 0$ then*

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

Proof. We can agree that

$$|X - \mu| \geq c \iff (X - \mu)^2 \geq c^2$$

now since $(X - \mu)^2$ is a positive random variable,

$$P(|X - \mu|) = P\{(X - \mu)^2 \geq c^2\}$$

we can now apply Markov's inequality,

$$P\{|X - \mu| \geq c\} = P\{(X - \mu)^2 \geq c^2\} \leq \frac{E[(X - \mu)^2]}{c^2} = \frac{Var(X)}{c^2}$$

so we are done. □

Example. Suppose I know that $X \geq 0$, $E(X) = 2$ and $Var(X) = 2$. What can we say about

$$P(X \geq 6)?$$

since we have the value of μ, we can say the following:

$$P(X \geq 6) = P(X - \mu \geq 4) \leq \frac{Var(X)}{4^2} = \frac{1}{16}$$

and we can see that

$$P(X \leq \mu - c) + P(X \geq \mu + c) = P(|X - \mu| \leq c)$$

4.6 Chapter 4, Question 27

Suppose a bolt manufacturer knows that 5% of his production is defective. He gives a guarantee on his shipment of 10,000 parts by promising to refund the money if more than a are defective. How small can the manufacturer chose a to be and still be assured that he need not give a refund more than 1% of the time?

Let X= the number of defective bolts, or for our purposes, our 'number of successes'. We know that $X = Binomial(n = 10,000, p = .05)$. A refund will be given if and only if $X \geq a$. What we are trying to do is find the smallest possible a for which the probability of a refund is not larger than 1%. For X, what is μ? Well for a binomial random variable, we know that $E(x) = np = 500$. The variance σ^2 is equal to $npq = 475$ Looking at the probability

$$P(X \geq a) = P(X - \mu \geq a - \mu) \leq_{\text{Cheb. Ineq.}} \frac{Var(X)}{(a-\mu)^2}$$

which is true provided that $a - \mu$ is positive. If we impose that $\frac{Var(X)}{(a-\mu)^2} \leq .01$, we can solve for a. Focusing on our inequality,

$$\frac{Var(X)}{(a-\mu)^2} = \frac{475}{(a-500)^2} \leq .01 \Rightarrow 47,500 \leq (a-\mu)^2$$

taking the square root,

$$\sqrt{47,500} \leq a - 500$$

adding 500,

$$a = \sqrt{47,500} + 500 \approx 717.94494$$

so we take $a = 718$, which is the smallest value of a that allows the manager to pay for returns less than 1% of the time.

4.7 Chapter 4, Question 15

Let $X_1, ...X_n$ be independent random variables having a common density with mean μ and variance σ^2. ‡ Looking at the sample mean $\overline{X}_n$, we have:

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

calculating the variance and expected value of the sample mean we have:

$$E(\overline{X}_n) = \mu$$

since the expected value of the sum is the sum of the expected values. In calculating the variance:

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

‡these variables are called 'i.i.d.' since they have the same mean and variance

For $n \geq 2$, we need to show that:

$$E\left(\frac{(X_1 - \overline{X})^2 + ... + (X_n - \overline{X})^2}{n-1}\right) = \sigma^2$$

where here, we are taking the expected value of the **Sample-Variance**. Let us look at:

$$X_i - \overline{X} = (X_i - \mu) - (\overline{X} - \mu)$$

this leads us to:

$$(X_i - \overline{X})^2 = (X_i - \mu)^2 + (\overline{X} - \mu)^2 - 2(X_i - \mu)(\overline{X} - \mu)$$

making this a sum, we have:

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i=1}^{n}(\overline{X} - \mu)^2 - 2\sum_{i=1}^{n}(X_i - \mu)(\overline{X} - \mu)$$

noticing that:

$$\sum_{i=1}^{n}(\overline{X} - \mu) = \sum_{i=1}^{n}\overline{X} - n\mu = n\overline{X} - n\mu = n(\overline{X} - \mu)$$

and that

$$2\sum_{i=1}^{\infty}(X_i - \mu)(\overline{X} - \mu) = 2n(\overline{X} - \mu)\sum_{i=1}^{\infty}(X_i - \mu) = 2n(\overline{X} - \mu)^2$$

plugging in, we get:

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\overline{X} - \mu)^2$$

taking the expected value, we have:

$$E\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2$$

4.8 Chapter 4, Question 29

Given some $X \geq 0 \in \mathbb{V}_+$, given $\Phi_X(t) = E(t^X)$, taking a $0 < t < 1$, we aim to show that

$$P(X \leq x_0) \leq \frac{\Phi_X(t)}{t^{x_0}}$$

It is clear that if $t = 1$, this is clear. Let us assume that $t < 1$ and fixed. Let $M(x) = t^X, x \in \mathbb{R}$. This function is clearly **decreasing**. We can then see that

$$X \leq x_0 \quad \equiv \quad t^{x_0} \leq t^X$$

so since these events are equivalent, we have

$$P(X \leq x_o) = P(t^X \geq t^{x_0})$$

so by Markov's inequality,

$$P(X \leq x_o) = P(t^X \geq t^{x_0}) \leq \frac{E(t^X)}{t^{x_0}} = \frac{\Phi_X(t)}{t^{x_0}}$$

4.9 Chapter 4, Question 30

Let X have a Poisson density with parameter λ. We want to show that:

$$P\left(X \leq \frac{\lambda}{2}\right) \leq \left(\frac{2}{e}\right)^{\lambda/2}; \quad P(X \geq 2\lambda) \leq \left(\frac{e}{4}\right)^{\lambda}$$

We know the following about Poisson random variables of parameter λ:

$$\Phi_X(t) = e^{\lambda(t-1)}$$

Take a look at:

$$P(X \leq \frac{\lambda}{2})$$

from part (a) of question 29, we see that

$$a := P(X \leq \frac{\lambda}{2}) \leq \frac{e^{\lambda(t-1)}}{t^{\lambda/2}}$$

for every $0 < t \leq 1$. Focusing on the function on the right, we do the following:

$$\frac{e^{\lambda(t-1)}}{t^{\lambda/2}} = e^{-\lambda}\frac{e^{\lambda t}}{t^{\lambda/2}} = e^{-\lambda} \cdot \left(\frac{e^{2t}}{t}\right)^{\lambda/2}$$

saying that

$$m(t) = \frac{e^{2t}}{t}$$

we would not like to minimize $m(t)$ on the interval $0 < t \leq 1$. This is strictly from calculus, we proceed by taking the derivative:

$$m'(t) = \frac{2e^{2t} \cdot t - e^{2t}}{t^2} = \frac{e^{2t}(2t-1)}{t^2}$$

so setting this equal to zero, it is clear that

$$m'(t) = 0 \Rightarrow t = \frac{1}{2}$$

so for all t from $0 < t \leq 1$, we know that

$$\frac{e^{2t}}{t} \geq 2e \quad \text{which follows from plugging in } t = .5$$

raising this to the power $\lambda/2$ and multiplying with $e^{-\lambda}$, we put all the pieces back together and get:

$$a := P(X \leq \frac{\lambda}{2}) \leq e^{-\lambda} \cdot (2e)^{\lambda/2} = 2^{\lambda/2} \cdot e^{-\lambda/2}$$

from which we can conclude:

$$P\left(X \leq \frac{\lambda}{2}\right) \leq \left(\frac{2}{e}\right)^{\frac{\lambda}{2}}$$

so we have a fairly good estimate for the probability.

4.10 Chapter 4, Question 26

Given X a random variable with values $1, 2, 3$, where

$$f(1) = f(3) = \frac{1}{18}$$

and

$$f(2) = \frac{16}{18}$$

we need to show that there exists $\delta > 0$ such that

$$P(|X - \mu| \geq \delta) = \frac{Var(x)}{\delta^2}$$

first, we calculate:

$$\mu = 1f(1) + 2f(2) + 3f(3) = 2$$

and

$$\sigma^2 = E(X^2) - \mu = 1^2 f(2) + 2^2 f(2) + 3^2 f(3) - 4 = \frac{1}{9}$$

now looking at

$$|X - 2| \geq \delta$$

for some positive δ, let us first take δ to be 1. What does it mean to say $|X - 2| \geq 1$ in terms of X? It means that X takes the values:

$$X = 1, \quad X = 3$$

so thus,

$$P(|X - 2| \geq 1) = P(X = 1) + P(X = 3) = f(1) + f(3) = \frac{2}{18} = \frac{1}{9} = \frac{Var(X)}{\delta^2}$$

and we are done. Now notice that for any $\delta > 0$, then $|X - 2| \geq \delta$ would mean that $X = 1$ or $X = 3$. So for any $\delta > 0$, the probability of $|X - \mu| \geq \delta$ is $\frac{1}{18} + \frac{1}{18} = \frac{1}{9}$, which we know to be the Variance. Thus, δ must be one.

4.11 Chapter 4, Question 28

Given $X = Poisson(\lambda)$, using Chebyshev's inequality, we know that

$$P(X \leq \mu - c) + P(X \geq \mu + c) = P(|X - \mu| \geq c) \leq \frac{Var(X)}{c^2}$$

we know that for a Poisson random variable, $\mu = \lambda = \sigma^2$, so for this case, we want $\frac{\lambda}{2}$ to be equal to $\lambda - c$, and in solving c, we see that $c = \frac{\lambda}{2}$ and in using as such,

$$P(X \leq \frac{\lambda}{2}) \leq \frac{Var(X)}{c^2} = \frac{\lambda}{\lambda^2/4} = \frac{4}{\lambda}$$

and for part (b), we want the probability that $P(X \geq 2\lambda)$, and since we want $2\lambda = \lambda + c$, we see that $c = \lambda$. Plugging back in again, by Chebuchev's inequality we know that

$$P(X \geq 2\lambda) \leq \frac{Var(X)}{c^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

4.12 The Weak Law of Large Numbers

Suppose that $X_1, X_2, ...X_n$ are independent, identically distributed variables (i.i.d., all with the same f(x). This tells us that μ and σ^2 is the same for all these variables). Remember that $\overline{X}_n$ is the 'sample mean', and is defined to be:

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

Now, let $\epsilon > 0$, fixed. Now, when looking at

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \epsilon) = 0$$

The interpretation is that, if n is large, then $P(|\overline{X} - \mu| \geq \epsilon) \approx 0$, which is the same as saying $P(|\overline{X}_n - \mu < \epsilon)^C \approx 1$. Numerically, saying that $\epsilon = .01$ we say that the distance between $\overline{X}$ and μ being less than %1 is almost %100.

Proof. We know that $E(\overline{X}) = \mu$, and that $Var(\overline{X}) = \frac{\sigma^2}{n}$, so

$$P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{Var(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

and since

$$\lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$$

we see that we have proved the weak law of large numbers. □

4.13 Review for Midterm #1

4.13.1 Chapter 4, Question 6

This question uses Theorem 5, which says for a random variable $X \in \mathbb{V}_+$, the expected value of X exists and is finite if and only if the following series converges:

$$\sum_{x=1}^{\infty} P(X \geq x)$$

and in that case,

$$E(x) = \sum_{x=1}^{\infty} P(X \geq x)$$

So for our question, we have

$$M > 0 \qquad \text{integral,}$$

where $X = Geom(p)$ and let

$$Z = min(X, M) : 0, 1, 2, ..., M$$

so, Z is in $\mathbb{IV}_+$. Looking at

$$\sum_{x=1}^{\infty} P(Z \geq x) = \sum_{x=1}^{M} P(Z \geq x)$$

so this series converges, since it is a finite sum. The minimum,

$$min(X, M) \geq x \quad \equiv \quad X \geq x \quad \text{and} \quad M \geq x$$

since $M \geq x$ for all x,

$$\sum_{x=1}^{\infty} P(Z \geq x) = \sum_{x=1}^{M} P(Z \geq x) = \sum_{x=1}^{M} P(X \geq x)$$

$$= q + q^2 + ... + q^M = \frac{1 - q^{M+1}}{1 - q} - 1$$

4.13.2 Chapter 4, Question 29

We want to show that

$$P(X \geq x_0) \leq \frac{\Phi_x(t)}{t^{x_0}}, \quad t \geq 1$$

now, when $t = 1$, we see that $\Phi_x(t = 1) = 1$, so $t^{x_0} = 1$. Thus, the inequality is true. Let us assume that $t > 1$ and is fixed. Now, we are told that the generating function is finite for all t. We can proceed as follows:

$$v(x) := t^x$$

and we know that $v(x)$ is increasing, since $t > 1$. So,

$$X \geq x_0 \Rightarrow \ t^X \geq t^{x_0}$$

and vice versa, so

$$X \geq x_0 \iff t^X \geq t^{x_0}$$

so,

$$P(X \geq x_0) = P(t^X \geq t^{x_0})$$

so by using Markov's inequality, we have

$$P(X \geq x_0) = P(t^X \geq t^{x_0}) \leq \frac{E(t^X)}{t^{x_0}} = \frac{\Phi_x(t)}{t^{x_0}}$$

4.13.3 Chapter 4, Question 14

We have $E(2X+3Y) = 2a+3b$, where we let $a = E(X),\ b = E(Y)$. If we call

$$\sigma_x^2 = Var(X) \quad \sigma_y^2 = Var(Y)$$

so

$$Var(2X+3Y) = Var(2X) + Var(3Y)$$

since X, Y are independent. Proceeding,

$$Var(2X+3Y) = Var(2X) + Var(3Y) = 4Var(X) + 9Var(Y) = 4\sigma_x^2 + 9\sigma_y^2$$

Recall that

$$Var(X) = E(X^2) - (E(X))^2$$

so for the variance to exist, we needed the second moment to be finite, otherwise it wouldn't exist.

4.13.4 Chapter 3, Question 15(c)

X, Y are independent random variables having the uniform density on $\{0, 1, ...N\}$. We would like to find the density of $|Y-X|$. First, notice that

$$|Y-X| = 0, 1, ...N$$

if we call one of these values z, we then say that

$$Z = |Y-X| = z \quad \text{if and only if} \quad Y-X = z \text{ or } X-Y = z$$

these are disjoint events. I.e.,

$$P(Z=z) = P(Y-X=z) + P(X-Y=z)$$

this is o.k., as long as z ranges from $1, N$. If you take an fix such a z, notice that

$$X-Y \quad \text{and} \quad Y-X$$

have the same density. So, we can double one and drop the other in our calculation.

$$P(Z=z) = 2P(Y-X=z)$$

we can express this event in the following way:

$$Y-X=z \quad \equiv \quad \bigcup_{x=0}^{N}(X=x, Y-X=z) = \bigcup_{x=0}^{N}(X=x, Y=x+z)$$

we want $x+y$ to be at most N, which is equivalent to saying $X \leq N-z$, and in that case,

$$\bigcup_{x=0}^{N}(X=x, Y=x+z) = \bigcup_{x=0}^{N-z}(X=x, Y=x+z)$$

and these events are now disjoint, and the probability of their intersection is their product:

$$P(Y-X=z)=\sum_{x=0}^{N-z}P(X=x)P(Y=x+z)$$

notice that here, x on the right hand event is a **number**. Continuing,

$$P(Y-X=z)=\sum_{x=0}^{N-z}P(X=x)P(Y=x+z)=(N-z+1)\frac{1}{(N+1)^2}$$

so,

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

and when $z=0$,

$$P(Z=0)=P(X=Y)=\text{The answer for 14(b)}$$

4.13.5 Chapter 3, Question 17

Letting $X<Y$ be independent random variables with $X=G(p_1),Y=G(p_2)$, in finding the density of $X+Y$ we use the convolution formula:

$$f_{X+Y}(z)=\sum_{x=0}^{z}f_X(x)f_Y(z-x)$$

and since

$$f_X(x)=p_1q_1^x \quad f_Y(z-x)=p_2q_2^{z-x}$$

combining, and removing things that don't depend on x:

$$f_{X+Y}(z)=\sum_{x=0}^{z}f_X(x)f_Y(z-x)=p_1p_2q_2^z\sum_{x=0}^{z}\left(\frac{q_1}{q_2}\right)^x$$

which gives us two cases. Suppose that $\frac{q_1}{q_2}=1$, which tells us that $q_1=q_2=q,p_1=p_2=p$, so in this case, we have:

$$f_{X+Y}(z)=(z+1)p^zq^z$$

and in case two, where $Q=\frac{q_1}{q_2}\neq 1$, we get

$$f_{X+Y}(z)=p_1p_2q_2^2\left[1+Q+Q^2+Q^3+...\right]=p_1p_2q_2^2\frac{Q^{z+1}-1}{Q-1}=p_1p_2q_2^2\frac{\frac{q_1^{z+1}}{q_1^{z+1}}-1}{\frac{q_1}{q_2}-1}$$

4.13.6 Chapter 3, Question 31

Given X, Y indepenent random variables of uniform distributions of $\{1, 2, ...N\}$, we would like to find the densities of $X + Y$. First, notice that

$$X + Y : 2, 3, ...2N$$

since both X, Y are in $\mathbb{V}_+$, we can use our convolution formula,

$$F_{X+Y}(z) = \sum_{x=1}^{z} f_X(x) f_Y(z-x) = \sum_{x=1}^{z-1} f_X(x) f_Y(z-x) \quad ^{\S}$$

we have already done this question, and will now offer a different approach. Let $N = 5$, and let X, Y be independent uniform variables over $\{1, 2, 3, 4, 5\}$. Notice that

$$\Phi_X(t) = \sum_{x=1}^{5} f_X(x) t^x = \frac{1}{5}(t + t^2 + t^3 + t^4 + t^5)$$

since the sum is finite, this sum converges for all t. Now,

$$\Phi_Y(t) = \frac{1}{5}(t + t^2 + t^3 + t^4 + t^5)$$

since all the values are the same. We know that since the variables are independent,

$$\Phi_{X+Y}(t) = \Phi_X(t) \cdot \Phi_Y(t) = \frac{1}{25}(t^{10} + 2t^9 + 3t^8 + 4t^7 + 5t^6 + 4t^5 + 3t^4 + 2t^3 + t^2)$$

$$\Phi_{X+Y}(T) = \frac{1}{25}(t^2 + 2t^3 + 3t^4 + 4t^5 + 5t^6 + 4t^7 + 3t^8 + 2t^9 + t^{10})$$

recall that

$$\Phi_{X+Y} = \sum_{z=0}^{\infty} f_{X+Y}(z) t^z$$

so matching the coefficients,

$$\begin{aligned} F_{X+Y}(2) &= \frac{1}{25} \quad & F_{X+Y}(3) &= \frac{2}{25} \\ F_{X+Y}(4) &= \frac{3}{25} \quad & F_{X+Y}(5) &= \frac{4}{25} \\ F_{X+Y}(6) &= \frac{5}{25} \quad & F_{X+Y}(7) &= \frac{4}{25} \\ F_{X+Y}(8) &= \frac{3}{25} \quad & F_{X+Y}(9) &= \frac{2}{25} \\ & & F_{X+Y}(10) &= \frac{1}{25} \end{aligned}$$

so in other words, we can match these coefficients to get the densities. You can actually extend this idea to N cases, and can get a nice formula for it.

§since when z=x, F_Y is zero

4.13.7 Chapter 4, Question 14

We have X, Y independent having the uniform density over $\{0, 1, ..N\}$. here, notice that

$$f_X(x) = \frac{1}{N+1} = f_Y(y)$$

for all x, y. Since these variables are independent, looking at the vector (X, Y) its joint can be represented as:

$$joint(X, Y) = f(x, y) = f_X(x)f_Y(y) = \frac{1}{(N+1)^2}$$

as long as x, y are from $0, N$. Now, compare $P(X < Y)$ to $P(X > Y)$. We note that these two probabilities are equal. Let us say:

$$P(X < Y) = P(X > Y) = a$$

for (b), notice that

$$(X = Y) = \bigcup_{y=0}^{N}(X = Y, Y = y) = \bigcup_{y=0}^{N}(X = y, Y = y)$$

and now these events are disjoint, so,

$$P(X = Y) = \sum_{y=0}^{N} P(X = x)P(Y = y) = \sum_{y=0}^{N} \frac{1}{(N+1)^2} = \frac{N+1}{(N+1)^2} = \frac{1}{N+1}$$

now, notice that $\Omega = (X < Y) \bigcup (X = Y) \bigcup (X > Y)$, which are all disjoint events. Now, applying the probabilities,

$$P(\Omega) = 1 = q + \frac{1}{N+1} + a = 2a$$

so,

$$2a = 1 - \frac{1}{N+1} = \frac{N}{N+1} \Rightarrow a = \frac{N}{2(N+1)}$$

so for part (a), we see that

$$P(X \geq Y) = P(X = Y) + P(X > Y) = \frac{1}{N+1} + \frac{N}{2(N+1)} = \frac{N+2}{2(N+1)}$$

thus, we have another solution to this question.

4.13.8 Chapter 4, Question 30

In part (a), $P(X \leq x_0) \leq \mu(t)$ for all $t \in (0, 1)$. In question 29, we have the following idea: so we can eventually get

$$\frac{\Phi_x(t)}{t^{x_0}} = u(t)$$

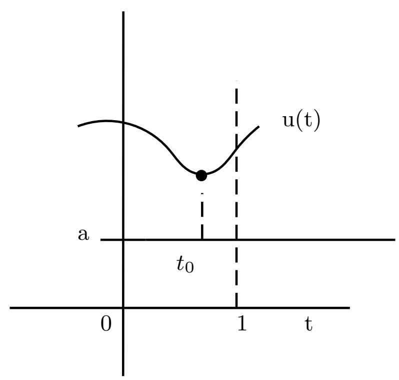


Figure 4.1: The idea is to find the absolute minimum value of $u(t_0)$ and then $a \leq u(t_0)$

Chapter 5

Continuous Random Variables

5.1 Cumulative Distribution Function

We will abbreviate the Cumulative Distribution Function as c.d.f. For each random variable X, there exits a corresponding function F such that:

$$F(x) := P(X \leq x) \quad x \in \mathbb{R}$$

Recall that when when X is discrete, with probability function $f(t) = P(X = t)$, we defined $F(x)$ as follows:

$$F(x) = \sum_{\text{all } t \leq x} f(t)$$

whose graph we could interpret as a step function. So, For a random variable X with c.d.f. $F(X)$, what general properties can we list? Well, it turns out we have the following properties:

1. $0 \leq F(x) \leq 1$
2. $F(x)$ is non-decreasing (if $x_1 < x_2$, then $F(x_2) \leq F(x_2)$)
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

$$\text{for example, take} \quad x_1 \leq x_2 \leq ... \leq x_n \to \infty$$

for this monotonically increasing function, all we need to show is that

$$F(x_n) \to 1$$

and looking at the events

$$\{X \leq x_1\} \subseteq \{X \leq x_2\} \subseteq ... \subseteq \{X \leq x_n\} = A_1 \subseteq A_2 \subseteq ... \subseteq A_n$$

taking their union

$$A = \bigcup_{i=1}^{n} A_i$$

we see that

$$P(X \leq x_n) \to P(A)$$

and since $A = \Omega$, we see that $F(X_n) = P(X \leq x_n) \to P(A) = 1$.

4. $F(x)$ is always right-continuous, meaning that $\lim_{x \to a^+} F(x) = F(a)$. This has an interesting geometric discussion, and is worth looking up in the textbook, Figure 3 of chapter 5. We say that $F(a) - F(a^-)$ is the 'possible size of the jump at a'. We can show without to much trouble that this is really equal to $P(X = a)$. So, our conclusion is that $F(x)$ is continuous on $\mathbb{R}$ if and only if $P(X = a) = 0$ for all a.

Definition. A **density** (with respect to integration) is a function $f(x) \geq 0$ for all $x \in \mathbb{R}$ and such that

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Definition. A random variable X is called **continuous** with density if there exists a density (with respect to integration) $f(x)$ such that the distribution

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad \Rightarrow \quad F'(x) = f(x)$$

We have the formula for an interval $C \in \mathbb{R}$, that

$$P(X \in C) = \int_C f(x)dx$$

For example,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Example. Recall that for variables of the form $Exp(\lambda)$, we have the following density: for $\lambda > 0$,

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases}$$

The **support** for a random variable is the interval such that f on the compliment of it is equal to zero. In this case, the support for our random variable $Exp(\lambda)$ is $[0, \infty]$.

Example. Let $X = Exp(\lambda)$ and let $Y = \sqrt{X}$. We would like to find the density of Y. We first notice that X has corresponding functions f and F. Similarly, we notice that Y has a density function g and cumulative distribution function G. We aim to solve for g. We can immediately see from our definition of Y that $Y \geq 0$, implying that the support of g must be the interval from 0 to ∞. So:

$$G(y) = P(Y \leq y) = 0 \text{ if } y \text{ is negative} \quad \Rightarrow G'(y) = 0 \text{ if } y < 0$$

Immediately, we can say that $g(y) = 0$ if $y \leq 0$. Now, let us take a y that is positive, and fix it. We have the following relationship:

$$G(y) = P(\sqrt{X} \leq y) = P(X \leq y^2) = F(y^2)$$

which gives us a connection between $G(y)$ and $F(y)$. So, since

$$g(y) = G'(y) = F'(y^2) = 2yf(y^2) = 2y\lambda e^{-\lambda y^2}$$

we say that:

$$g(y) = \begin{cases} 2y\lambda e^{-\lambda y^2} \text{ if } y > 0 \\ 0 \text{ if } y \leq 0 \end{cases}$$

One should review pages 110-115 and example 5 on page 117.

5.2 Finding the Density of a Transformation of a Random Variable

The general types of questions we would like to approach are like the following: suppose that we are given some random variable X with a density $f(x)$. Given some function φ, let

$$Y = \varphi(X)$$

If Y is a continuous random variable with density, we would like to be able to find the density of Y, $g(y)$.

Definition. A random variable X is called **Normal**(μ, σ^2) if the density of X is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The 'standard normal' Is $Normal(\mu = 0, \sigma^2 = 1)$, in which case:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Notice that

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$$

The graph of such a density is the well know '**bell curve**', where the area under this curve from $-\infty$ to ∞ is equal to 1.

We have the following facts for a normally distributed variable X:

1. $E(X) = \mu$
2. $Var(X) = \sigma^2$

Example. Suppose that $X = N(\mu, \sigma^2)$, and that $a \neq 0, b$ are two constants. Let $\varphi(x) = ax + b$. Look at

$$Y = \varphi(X) = aX + b$$

Now find the density of Y, called $g(y)$. Notice that X has a density $f(x)$, and a cumulative distribution $F(x)$. The same is true for Y, it has a density $g(y)$ and a cumulative distribution $G(y)$. We know that

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

we also know that

$$G(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a})$$

We are assuming here that $a < 0$, hence making it clearly unnecessary to reverse the inequality. Continuing,

$$P(X \leq \frac{y-b}{a}) = F(\frac{y-b}{a})$$

Thus,

$$G(y) = F(\frac{y-b}{a})$$

Taking the derivative,

$$g(y) = G'(y) = F'(\frac{y-b}{a}) = f(\frac{y-b}{a})\frac{1}{a}$$

notice that when $a > 0$,

$$g(y) = \frac{1}{a} f(\frac{y-b}{a})$$

regardless of the distribution of X- we have not as of yet mentioned the fact that X is a normal random variable. In our example,

$$f(\frac{y-b}{a})\frac{1}{a} = \frac{1}{a} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\frac{y-b}{a}-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}(a\sigma)} e^{-\frac{(y-(a\mu+b))^2}{2(a\sigma)^2}}$$

notice that this density is normal, $N(a\mu + b, a^2\sigma^2)$. We just proved the following theorem:
Theorem 16. *Given* $a > 0, \ \ b \in \mathbb{R}, \ \ X = N(\mu, \sigma^2)$ *then*

$$aX + b = N(a\mu + b, a^2\sigma^2)$$

As an exercise, repeat this process and see what happens if $a < 0$.

5.3 Chapter 5, Question 9

Let X denote the decay time of some radioactive particle and assume that the distribution of X is given by:

$$X = Exp(\lambda)$$

where $\lambda > 0$ and $X \geq 0$ and continuous, with density:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

We aim to find a number t such that $P(X \geq .t) = .9$. Based on the definition of our random variable, we have the following:

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(t)dt$$

and by the properties of our random variable, $F(x) = 0$ if $x \leq 0$, so our integral is really:

$$F(x) = \int_0^x f(t)dt = -e^{-\lambda t}\Big|_{t=0}^{t=x} = 1 - e^{\lambda x}$$

so,

$$F(x) = \begin{cases} 1 - e^{-\lambda x} \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

In our question we are told that $P(X \geq .01) = \frac{1}{2}$, from which we notice that:

$$1 - (1 - e^{-.01\lambda}) = 1 - F(.01) = P(X \geq .01) = \frac{1}{2}$$

so,

$$\lambda = 100\ ln(2)$$

in now finding t such that $P(X \geq t) = .9$, we note that if $X = Exp(\lambda)$,

$$1 - P(X \geq x) = P(X \leq x) = 1 - e^{-\lambda x} \quad \text{if } x \geq 0$$

so,

$$P(X \geq x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$$

thus,

$$e^{-\lambda t} = .9 \Rightarrow -\lambda t = ln(.9) \Rightarrow t = \frac{\ln(.9)}{-\lambda} = \frac{ln(.9)}{-100 \cdot ln(2)}$$

5.4 Chapter 5, Question 7

Pick a point (u, v) uniformly from the square $0 \leq u \leq 1$, $0 \leq v, \leq 1$. Let X b the random variable that assigns to the point (u, v) the number $u + b$. We wish to find the cumulative distribution function of X. This probability space Ω is represented by Figure(5.1).

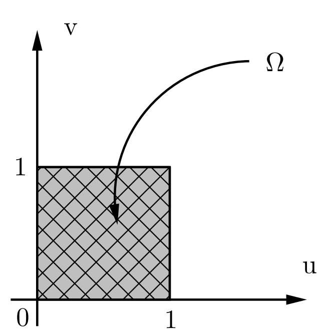


Figure 5.1: Ω

Define:

$$X(u, v) = u + v$$

notice that the values X takes on lie in the following interval:

$$0 \leq X \leq 2$$

So, if $F(x) = c.d.f.$ of X, in view of the above fact,

$$\begin{cases} F(x) = 0 \text{ if and only if } x \leq 0 \\ F(x) = 1 \text{ if and only if } x \geq 2 \end{cases}$$

Fix some x, where $0 < x < 2$. Now look at

$$F(x) = P\{(u, v) \in \Omega \ \mid \ u + v \leq x\}$$

As we did in elementary school, we would now like to graph this line and treat this as an inequality. We now have two cases.

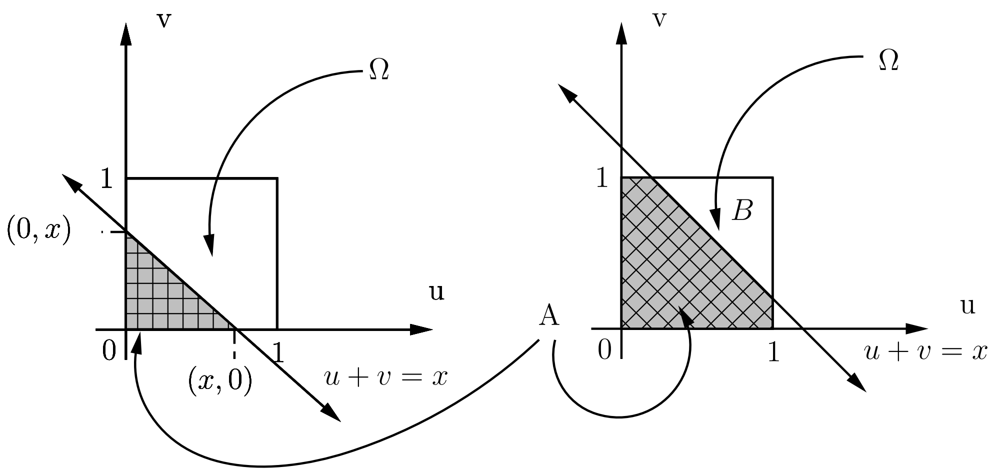


Figure 5.2: Case(1), Case (2)

Fist, let us suppose that $0 < x < 1$. In this case, the line $u + v = x$ can be graphed over our square and the area corresponding to $u + v \leq x < 1$ can be found, and is in fact equal to $F(x)$. In this case,

$$F(x) = x^2/2 = \text{the area of figure A}$$

In the second case, where $1 \leq x \leq 2$, it is easier to calculate the area of triangle B and to then subtract its area from the area of the unit square. The sides of triangle B can be calculated in the following way:

$$1 - (x - 1) = 2 - x$$

In which case,

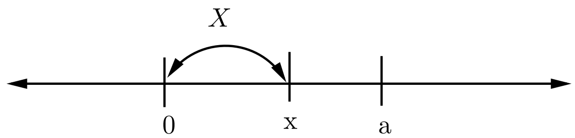
$$F(x) = 1 - B = 1 - \frac{(2 - x)^2}{2}$$

so in summary,

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2}{2} & \text{if } 0 \leq x \leq 1 \\ 1 - \frac{(2-x)^2}{2} & \text{if } 1 \leq x \leq 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

5.5 Chapter 5, Question 10

Let X denote the distance of the point x from the origin where x is chosen uniformly over the interval $[0, a]$.



The density of x is then the following, since it is selected uniformly over the interval $[0, a]$:

$$f(x) = \begin{cases} \frac{x}{a} \text{ if } 0 \leq x \leq a \\ 0 \text{ otherwise} \end{cases}$$

looking at $Y = min(X, \frac{a}{2})$, the question asks us to find the distribution of Y. Notice that:

$$Y = min\left(X, \frac{a}{2}\right) = \begin{cases} X \text{ if } X \leq \frac{a}{2} \\ \frac{a}{2} \text{ if } X \geq \frac{a}{2} \end{cases}$$

so clearly,

$$0 \leq Y \leq \frac{a}{2}$$

Let us call $G(y)$ the cumulative distribution function of Y;

$$G(y) = P(Y \leq y)$$

If $y \leq 0$, this implies that $G(y) = 0$. If $0 < y < \frac{a}{2}$, then

$$G(y) = P(Y \leq y) = P(X \leq y) = \frac{y}{a}$$

which follows from noticing that the following is true in general for X:

$$P(X \leq x) = F(x) = \begin{cases} 0 \text{ if } x < 0 \\ \frac{x}{a} \text{ if } 0 \leq x \leq a \\ 1 \text{ if } x \geq a \end{cases}$$

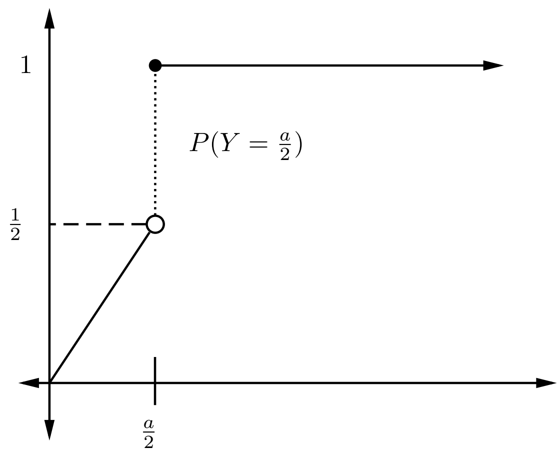
Finally, if $\frac{a}{2} \leq y$, then:

$$G(y) = P(Y \leq y) = 1$$

So, what kind of variable is Y? We notice that it is neither discrete of continuous, it is a "mixture"; it has properties of both discrete and continuous variables. Referring to the graph of $G(y)$ and calculating the probability $P(Y = \frac{a}{2})$, we see that

$$P(Y = \frac{a}{2}) = P(X \geq \frac{a}{2}) = \frac{1}{2}$$

which is the the size of the 'jump' at $Y = \frac{a}{2}$. This is illustrated with the following figure:



Theorem 17. *Let X is a continuous random variable, and $Y = \varphi(X)$, where*

$$\varphi : I_{interval} \to \mathbb{R}$$

is differentiable, strictly increasing or strictly decreasing, and has the property: $f(x) = 0, \forall\, x \notin I$. We have the following:

$$y = \varphi(x) \quad \equiv \quad x = \varphi^{-1}(y)$$

and random variable $Y = \varphi(X)$ is continuous with density:

$$g(y) = f(x) \cdot \left|\frac{dx}{dy}\right|$$

(where here, $x = \varphi^{-1}(y)$). This theorem is stated and proven in the book.

Example. Suppose that $X = Exp(\lambda)$, and let $\beta \neq 0$. Take $\varphi(x) = x^{\frac{1}{\beta}}$. Assume that $\beta < 0$. Taking the derivative,

$$\varphi'(x) = \frac{1}{\beta} x^{\frac{1}{\beta}-1} < 0 \text{ on } (0, +\infty)$$

for $x > o$. So, φ is strictly decreasing on the interval $(0, \infty)$. So,

$$y = \varphi(x) = x^{\frac{1}{\beta}}$$

and we would like to solve for x to get the inverse. In this case,

$$x = y^{\beta} \Rightarrow \frac{dx}{dy} = \beta y^{\beta-1}$$

taking the absolute value,

$$\left|\frac{dx}{dy}\right| = -\beta y^{\beta-1}$$

and using our theorem, we see that the density of Y is :

$$g(y) = f(y^{\beta}) \cdot (-\beta y^{\beta-1}) = -\beta y^{\beta-1} \lambda e^{-\lambda y^{\beta}}$$

Example. Assume that $X = Normal(0, \sigma^2)$. Recall that:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

Let $Y = X^2$. We aim to find the density $g(y)$ of Y. It is natural to think about constructing a function such that:

$$y = \varphi(x) = x^2$$

because in that case, $Y = \varphi(X)$. Notice that X has some density f, and a cumulative distribution F, and similarly Y has corresponding functions g, G where $G' = g$. We would like to connect these functions in a way that allows us to find g. Clearly, since $Y \geq 0$, $G(y) = 0$ if $y \leq 0$. This tells us that we should fix $y > 0$, and look at:

$$G(y) = P(Y = X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y})$$

Thus, we have a connection from G to F. Taking the derivative with respect to y,

$$g(y) = G'(Y) = \frac{1}{2\sqrt{y}} f(\sqrt{y}) + \frac{1}{2\sqrt{y}} f(-\sqrt{y})$$

since f is symmetrical, $f(\sqrt{y}) = f(-\sqrt{y})$, and:

$$g(y) = G'(Y) = \frac{1}{\sqrt{y}} f(\sqrt{y}) = \frac{1}{\sigma\sqrt{2\pi y}} e^{-\frac{y}{2\sigma^2}}$$

Giving us the following density:

$$g(y) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi y}} e^{-\frac{y}{2\sigma^2}} \text{ if } y > 0 \\ 0 \text{ otherwise} \end{cases}$$

Notice that this is the density of $Y = X^2$. Writing this in a different way, we have:

$$g(y) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} y^{-\frac{1}{2}} e^{\left(-\frac{1}{2\sigma^2}\right)y} \text{ if } y > 0 \\ 0 \text{ otherwise} \end{cases}$$

In attempting to extend this example, notice that $Y = X^2 > 0$. So, think of $X \geq 0$ as a continuous variable of the form:

$$X \geq 0 \text{ continuous } \rightsquigarrow f(x) = \begin{cases} cx^{\alpha-1} e^{-\lambda x} \text{ if } x > 0 \\ 0 \text{ otherwise} \end{cases}$$

This leads to a natural transition into what is known as the **Gamma Function**, on which we will connect these two concepts later.

5.6 The Gamma Function

Definition. The **Gamma Function** : For each $\alpha > 0$, the integral

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$$

is finite. We have some values, specifically that:

$$\Gamma(1) = 1, \quad \Gamma(\alpha+1) = \alpha\Gamma(\alpha), \quad \Gamma(n) = (n-1)!$$

Fact. If you take $\alpha > 0, \lambda > 0$, and look at the following:

$$\int_0^\infty x^{\alpha-1}e^{-\lambda x}dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}$$

This can be proven by a change of variable; let $\lambda x = y$, and see that $x = \frac{y}{\lambda}$, which preserves the limits of integration. Also notice that $\lambda dx = dy$, telling us that $dx = \frac{dy}{\lambda}$.
Definition. $X = \Gamma(\alpha, \lambda)$, if $X \geq 0, \alpha, \lambda > 0$, has density

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} \text{ if } x > 0 \\ 0 \text{ otherwise} \end{cases}$$

5.7 Chapter 5, Question 21

Let X be a positive continuous random variable having density f. Find a formula for the density of

$$Y = \frac{1}{(X+1)}$$

Notice that

$$0 < Y = \frac{1}{(X+1)} < 1$$

Call $g(y)$ the density of Y. Notice that $g(y) = 0$ if $y \leq 0$ or $y \geq 1$. So,

$$y = \varphi(x) = \frac{1}{(x+1)}$$

is a strictly decreasing function. We can find the inverse by solving for x,

$$x = \frac{1}{y} - 1$$

so,

$$\frac{dx}{dy} = -\frac{1}{y^2}$$

and using the formula, which says $g(y) = f(x)|\frac{dx}{dy}|$ we get:

$$g(y) = \frac{1}{y^2} \cdot f\left(\frac{1}{y} - 1\right)$$

so in conclusion,

$$g(y) = \begin{cases} \frac{1}{y^2}f(\frac{1}{y} - 1) \quad \text{if } 0 < y < 1 \\ 0 \quad \text{otherwise} \end{cases}$$

5.8 Chapter 5, Question 23

Let X be a random variable uniformly distributed on (a, b). We wish to find a linear function φ such that $Y = \varphi(X)$ is uniformly distributed on $(0, 1)$. Notice that we always have a function:

$$\mu : [a, b] \to [0, 1]$$

defined by:

$$\mu(x) = \frac{x - a}{b - a}$$

This follows from:

$$a \leq x \leq b \iff 0 \leq x - a \leq b - a \iff 0 \leq \frac{x - a}{b - a} \leq 1$$

Now take φ to be this function μ. It defines a **homeomorphism** between these two intervals. Notice that:

$$X = Uniform[a, b] \Rightarrow f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Notice that if

$$y = \frac{x - a}{b - a}$$

in finding the inverse, it is clear that

$$x = (b - a)y + a \Rightarrow \frac{dx}{dy} = b - a > 0$$

so using the formula $g(y) = f(x)|\frac{dx}{dy}|$, we get:

$$g(y) = f((b - a)y + a)(b - a)$$

so we see that

$$g(y) = \begin{cases} 1 & \text{if } y \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

and notice that $g(y)$ is uniform over $[0, 1]$.

Let us assume that X is a continuous random variable with density $f(x)$ and with support (a, b). In other words, $f(x) > 0$ for all $a < x < b$ and $f(x) = 0$ for all $x < a$ or $x > b$. The cumulative distribution function of such a function f would be continuous, and would be **strictly increasing** on the interval (a, b). Past b, the function F is at 1 and before a, the function F is at 0.

Theorem 18. *If we have X, a continuous random variable with density $f(x)$ and with support (a, b), we can call $Y = F(x)$, and note that Y has uniform distribution on $[0, 1]$.*

Proof. First, recognize that

$$0 \leq Y = F(X) \leq 1$$

The distribution function of Y, $G(y)$, is equal to 0 if $y < 0$ and $G(y) = 1$ if $y > 1$. Now take $0 < y < 1$, and say that

$$G(y) = P(F(X) \leq y)$$

Convince yourself that:

$$F(a) \leq b \iff a \leq F^{-1}(b)$$

using this in our problem, we see that

$$G(y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y$$

Putting the pieces back together, we find that $Y = F(x)$ has the distribution function:

$$G(y) = \begin{cases} y & \text{if } 0 < y < 1 \\ 0 & \text{if } y \leq 0 \\ 1 & \text{if } y \geq 1 \end{cases}$$

Which is the distribution function of the uniform random variable on the interval $[0, 1]$. $\square$

Going back to our question, notice that our density $f(x)$ has support (b, a). When you take the cumulative distribution function, we see that

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } \quad a \leq x \leq b \\ 1 & \text{if } x > b \end{cases}$$

Which was interestingly what we took our variable Y to be.

As an exercise, let $X = Exp(\lambda)$. Find φ such that $Y = \varphi(X)$ is Uniform$[0, 1]$. Notice that the distribution function would look like:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

5.9 Exam 2

The date of the second exam will be Wednesday, November the 23^{rd}.

5.10 Chapter 5, Question 19

Let $X, Y = X^2$ be positive continuous random variables having densities f, g respectively. Find f in terms of g and find g in terms of f.

Notice that $X > 0, Y > 0$. Also see that $y = x^2 = \varphi : (0, \infty) \to (0, \infty)$ is one-to-one and onto. Thus the inverse, is $x = \sqrt{y}$. Taking the derivative, $\frac{dx}{dy} = \frac{1}{2\sqrt{y}}$. According to our theorem,

$$g(y) = f(x) \left| \frac{dx}{dy} \right| = f(\sqrt{y}\,) \frac{1}{2\sqrt{y}}$$

which is equivalent to:

$$g(y) = \frac{f(\sqrt{y})}{2\sqrt{y}}$$

if $y > 0$, on our support. In finishing our question, we find f in terms of g, and get

$$f(\sqrt{y}) = 2\sqrt{y}g(y) \quad \Rightarrow \quad f(x) = 2xg(x^2)$$

if $x > 0$, or in other words, is on our support, and we have completed our question. However, how could we have done this question without our theorem? Notice that X has corresponding functions f, F, and Y has corresponding functions g, G. Notice first that

$$G(y) = 0 \quad \text{if } y \leq 0$$

Now take a $y > 0$, and fix it. Thus,

$$G(y) = P(X^2 \leq y) = P(-\sqrt{Y} \leq X \leq \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y}) = F(\sqrt{y})$$

so, if $y > 0$, $G(y) = F(\sqrt{y})$. Now taking the derivative and using the chain rule, we can arrive at our solution.

5.11 The Gamma Function (continued)

Recall that for $\alpha > 0$,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$$

It is easy to show that $\Gamma(1) = 1$, and $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$. For example, through integration by parts:

$$\Gamma(\alpha+1) = \int_0^\infty x^\alpha e^{-x}dx = -\frac{x^\alpha}{e^x}\bigg|_{x=0}^{x=\infty} - \alpha\int_0^\infty x^{\alpha-1}e^{-x}dx = \alpha\Gamma(\alpha)$$

by letting $u = x^\alpha$, $v' = e^{-x}$.

Recall that for $X = \Gamma(\alpha, \lambda)$, if $X \geq 0, \alpha, \lambda > 0$, X has density:

$$f(x) = \begin{cases} \frac{\lambda^x}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} \text{ if } x > 0 \\ 0 \text{ otherwise} \end{cases}$$

Also recall that we have the fact that if $X = Normal(0, \sigma^2)$, $Y = X^2$, the density of Y is

$$g(y) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}}y^{-1/2}e^{-\frac{y}{2\sigma^2}} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

thus when examining $g(y)$, we observe that it is $\Gamma(\alpha = \frac{1}{2}, \lambda = \frac{1}{2\sigma^2})$ By equating the constants of the two random variables, we arrive at the following,

$$\frac{1}{\sigma\sqrt{2\pi}} = \frac{\left(\frac{1}{2\sigma^2}\right)^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} \Rightarrow \frac{1}{\sigma\sqrt{2\pi}} = \frac{1}{\sqrt{2}\sigma\Gamma(\frac{1}{2})} \Rightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

since we have $f(x) = cu(x)\ \ a < x < b$ and $g(x) = du(x),\ a < x < b$, and being on the same support, we can conclude that $c = d$.

Theorem 19. *Let $X_1, ..., X_n$ be independent random variables such that X_i has the gamma density $\Gamma(\alpha_i, \beta)$ for $i = 1, ...n$. Then, $X_1 + X_2 + ... + X_n$ has the gamma density $\Gamma(\alpha, \beta)$ where:*

$$\alpha = \alpha_1 + \alpha_2 + ... + \alpha_n$$

Consider a density f with support on the interval (a, b). Can we produce a continuous random variable X with density exactly equal to f? It turns out that the answer is yes, and we can begin our construction in the following way: Let

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

This function F has an inverse F'. Take a random variable $Y = Uniform[0, 1]$. Let

$$X = F^{-1}(Y)$$

Claim. This random variable X has density $f(x)$ and cumulative distribution function $F(X)$.

For now, the proof of this claim will be skipped. Look at question (45), which uses this theorem to solve the question.

5.12 Chapter 5, Question 39

Let X be an exponentially distributed random variable with parameter λ. Let Y be the integer valued random variable defined in terms of X by $Y = m$ if $m \leq X < m + 1$ where m is a nonnegative integer. How is Y distributed? In other words, Y is the 'floor function' of X, and it takes the following values:

$$Y : 0, 1, 2, ...$$

Let us look at $P(Y \geq m)$. The following are equivalent:

$$P(Y \geq m) \equiv P(X \geq m) = e^{-\lambda m} = q^m$$

where $q = e^{-\lambda}$. $P(X \geq m) = e^{-\lambda m}$ because given $X = Exp(\lambda)$,

$$P(X \geq m) = 1 - P(X \leq m) = 1 - F(x) = 1 - \int_0^a \lambda e^{-\lambda t}dt = \Bigg|_{x=0}^{m} - e^{-\lambda t} = e^{-\lambda m}$$

Then, from what we know about Geometrically distributed random variables,

$$Y = Geom(p = 1 - q = 1 - e^{-\lambda})$$

5.13 Chapter 5, Question 33

Let X be normally distributed with parameters μ and σ^2. We want to find

$$P(|X - \mu| \leq \sigma)$$

We have shown that if X is normal, then $aX + b$ must also be a normally distributed random variable. In this case, assume that $X = N(\mu, \sigma^2)$, thus

$$aX + b = N(a\mu + b, a^2\sigma^2)$$

So, using this knowledge we can transform X to the following variable Z:

$$X = N(\mu, \sigma^2) \rightarrow \frac{X - \mu}{\sigma} = Z = N(0, 1)$$

since

$$\frac{X - \mu}{\sigma} = \frac{1}{\sigma}X + (-\frac{\mu}{\sigma}) = Z$$

where Z is a new random variable with the following distribution and density:

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt, \qquad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

φ is its density and Φ is its cumulative density, where $\Phi(x)$ is the integration of a standard bell curve from $-\infty$ to x. Notice that

$$\Phi(x) + \Phi(-x) = 1$$

Now notice that

$$P(|X - \mu| < \sigma) = P(|\frac{X - \mu}{\sigma}| < 1) = P(|Z| \leq 1) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1$$

since $\Phi(-1) = 1 - \Phi(1)$. We have a table in the back to calculate $\Phi(1)$. Referring to a table in the back of our textbook, we can find an explicit solution to this problem.

5.14 Chapter 5, Question 25

Let $g(x) = x(1 - x)^2,\ 0 \leq x \leq 1$ and $g(x) = 0$ elsewhere. How should g be normalized to make it a density? In other words, what should $c \neq 0$, a constant, be so that $c \cdot g(x)$ is a valid density? Formally, we would like to solve for c so that:

$$1 = \int_0^1 cx(1 - x)^2 dx = c \int_0^1 x(1 - x)^2 dx = c \left(\frac{x^2}{2} - \frac{2x^3}{3} + \frac{x^4}{4} \right) \Bigg|_{x=0}^{x=1} = \frac{c}{12}$$

so thus, $c = 12$, and

$$g^*(x) = \begin{cases} 12x(1 - x)^2 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

5.15 Chapter 5, Question 44

Let Y be uniformly distributed on $(0,1)$. Find a function φ such that $X = \varphi(Y)$ has the density f given by $f(x) = 2x, \quad 0 \leq x \leq 1$ and $f(x) = 0$ elsewhere. Notice that $f(x)$ is strictly increasing on the support $(0,1)$, and

$$F(X) = \int_0^x 2x \, dx = x^2$$

with the properties:

$$F(x) = \begin{cases} 0 & \text{if} \quad 0 \leq 0 \\ x^2 & \text{if} \quad 0 \leq x \leq 1 \\ 1 & \text{elsewhere} \end{cases}$$

By a theorem, if $Y = U(0,1)$ then $F^{-1}(Y) = X$ has density $f(x)$. So to find the inverse of $F : [0,1] \to [0,1]$, we solve for x and see that

$$x = \sqrt{y} = F^{-1}(y)$$

Thus,

$$\varphi(x) = \sqrt{x}, \qquad X = \sqrt{Y}$$

5.16 Chapter 5, Question 45

Let Y be uniformly distributed on $(0,1)$. Find a function φ such that $\varphi(Y)$ has the gamma density $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. First define f such that

$$\Gamma\left(\frac{1}{2}, \frac{1}{2}\right) = f(x) = \begin{cases} \frac{(1/2)^{\frac{1}{2}}}{\sqrt{\pi}} x^{-\frac{1}{2}} e^{-\frac{1}{2}x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

So, for $x > 0$,

$$F(x) = \int_0^x \frac{1}{\sqrt{2\pi t}} e^{-\frac{t}{2}} dt$$

calling $\sqrt{t} = u, t = u^2 \; dt = 2udu$ we have

$$= \int_0^{\sqrt{x}} \frac{1}{u \cdot \sqrt{2\pi}} e^{-\frac{u^2}{2}} 2u \cdot du = 2 \int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

And as we saw in Chapter 5 Question 33, this is equal to:

$$F(x) = 2\Phi(\sqrt{x}) - 1 \quad {}^*$$

Notice that $F^{-1}(Y)$ has density $f(x)$, as defined above. Solving for x, we get

$$y = F(x) \to y = 2\Phi(\sqrt{x}) - 1 \Rightarrow 2\Phi(\sqrt{x}) = y + 1 \Rightarrow \sqrt{x} = \Phi^{-1}\left(\frac{y+1}{2}\right)$$

so,

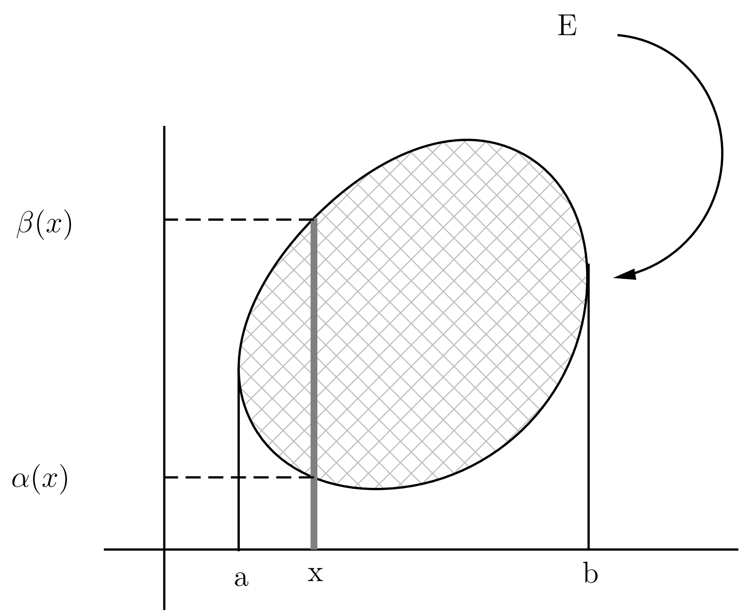
$$x = \left[\Phi^{-1}\left(\frac{y+1}{2}\right)\right]^2 = F^{-1}(y) = \varphi(y)$$

*Here, Φ is the distribution function for a Normally distributed random variable where $\mu = 0, \sigma = 1$

Chapter 6

Jointly Distributed Random Variables

6.1 A Brief Review of the Double Integral



Let us first remark that

$$\int\int_E 1 \; dxdy = Area(E)$$

Recall that in taking the integral of the above area E, we proceed in the following way:

$$\int\int_E f(x,y) \; dxdy = \int_a^b \left(\int_{\alpha(x)}^{\beta(x)} f(x,y)dy \right) dx$$

Example.

$$f(x,y) = xy^2$$

Taking the double integral over E, where E is the triangle bounded by the x, y axes and the line $y = 1 - x$, we have

$$\int\int_e f(x,y)\, dxdy = \int_0^1 \left(\int_0^{1-x} xy^2 dy\right) dx = \int_0^1 \frac{1}{3}(x - 3x^2 + 3x^3 - x^4)dx$$

at which point, we can proceed in the usual manner.

6.2 Multivariate Continuous Distributions

We approach this topic by looking at a simple example: suppose we have a pair (X, Y) where X and Y are continuous random variables. We know that these random variables have corresponding densities $f_X(x), f_Y(y)$ and cumulative distribution functions $F_X(x), F_Y(y)$. We say that together, X, Y have a joint cumulative distribution function,

$$F(x,y) = P\{X \leq x, Y \leq y\}$$

Definition. X, Y are said to be **independent** if and only if for any subset $A \subseteq \mathbb{R},\ B \subseteq \mathbb{R}$,

$$P\{X \in A, Y \in B\} = P\{X \in A\} \cdot P\{Y \in B\}$$

We would like to ask ourselves how the cumulative distribution functions of each variable is related to the cumulative distribution function of the pair (X, Y). Notice the following:
Fact.

$$F_X(x) = \lim_{y\to\infty} F(x,y) \qquad F_y(y) = \lim_{x\to\infty} F(x,y)$$

Notice that armed with $F(x, y)$, you can take limits to compute $F_X(x), F_Y(y)$.
Definition. We say that two random variables X, Y have a **joint** density $f(x, y)$ if $f(x, y) \geq 0$ for all x and y, and:

$$F(x,y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f(u,v)dv\right) du = \int_{-\infty}^{y} \left(\int_{-\infty}^{x} f(u,v)du\right) dv$$

Figure (6.1) represents this graphically.
Note. If $f(x, y)$ is a joint density, then:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

Fact. It is true that X, Y are independent if and only if:

$$f(x,y)O = f_X(x) \cdot f_Y(y)$$

We now arrive naturally at the following questions:

1. How is $f(x, y)$ related to the marginals $f_X(x), f_Y(y)$?

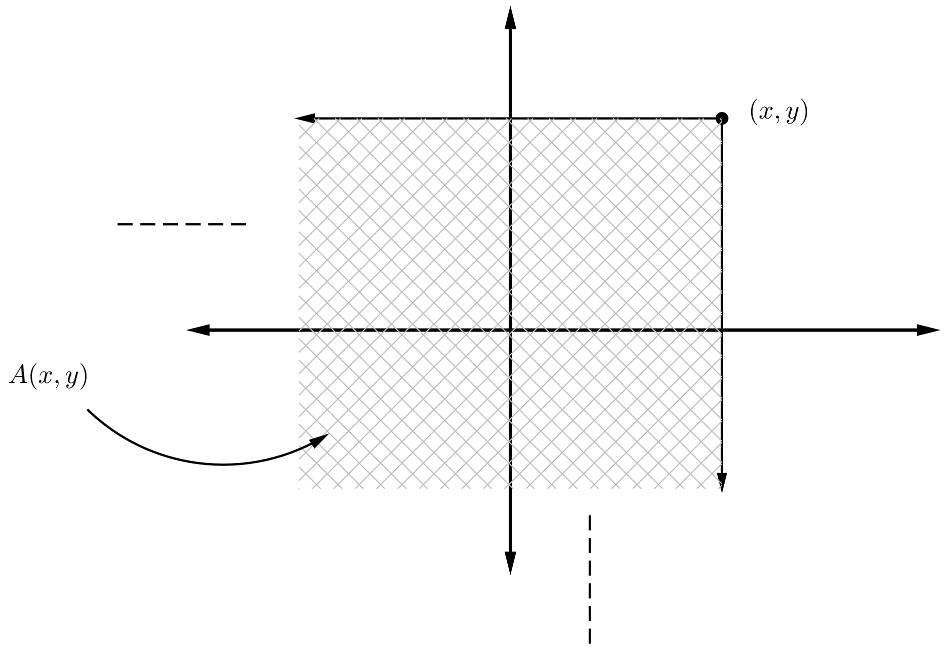


Figure 6.1: Above is a picture of the area A where the shading represents all those points less than (x, y).

2. How are the functions $f(x, y)$, $F(x, y)$ related?

In answering the first question, we say that:

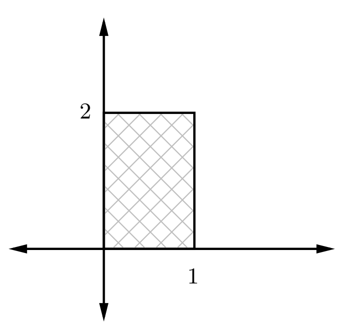
$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Example. Suppose that X, Y are continuous random variables with joint density:

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

First, what is the value of a? Second, what are the marginal densities $f_X(x), f_Y(y)$?

Notice first that the support, S, is equal to the following shaed area:



In solving for a, we claim that:

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dy\ dx = \int\int_S f(x,y)dy\ dx = a\int_0^1 \left(\int_0^2 x^2 y\ dy\right) dx$$

Calculating the following integrals:

$$= \int_0^2 y\ dy\ = \frac{y^2}{2}\bigg|_0^2 = 2 \qquad \int_0^1 x^2 dx = \frac{x^3}{3}\bigg|_0^1 = \frac{1}{3}$$

We see that $a = \frac{3}{2}$. Now looking for the marginal densities, we note that the marginal density $f_X(x) = 0$ if $x \le 0$ or $x \ge 1$. So, fix some $0 < x < 1$ and using our drawing, we see that:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^2 \frac{3}{2}x^2 y\ dy =$$

So in summary,

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

For the other marginal, $f_Y(y)$, we see that if we fix some $0 < y < 2$, then:

$$f_Y(y) = \int_0^1 \frac{3}{2}x^2 y\ dx = \frac{y}{2}$$

and thus:

$$f_Y(y) = \begin{cases} \frac{y}{2} & \text{if } 0 < y < 2 \\ 0 & \text{otherwise} \end{cases} \quad *$$

Now suppose we want to find $P\{X + Y \le 1\}$. In other words, we want to find the points:

$$\{X + Y \le 1\} = \{(X,Y) \in A\}$$

where

$$A = \{(x,y) | x + y \le 1\}$$

One approach is to notice the following:

$$P\{X + Y \le 1\} = \int\int_A f(x,y)\ dx\ dy$$

Which follows from recalling the fact that ff $A \subseteq \mathbb{R}^2$, then

$$P\{(X,Y) \in A\} = \int\int_A f(x,y)\ dx\ dy$$

Which is illustrated in figure (6.2). So, integrating over $\Delta = A \cap S$ where S is the support,

$$\int\int_\Delta = \int_0^1 \left(\int_0^{1-x} x^2 y\ dy\right) dx = \frac{1}{2}\int_0^1 \left(x^4 - 2x^3 + x^2\right) dx = C$$

So, $P\{(X + Y \le 1\} = a \cdot C = \frac{3}{2}C$. Going back to our second question, we claim that:

$$f(x,y) = \frac{\partial^2 F}{\partial x \partial y} = \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right)$$

*in reality you can let y range from $0 \le y \le 2$ without ambiguity.

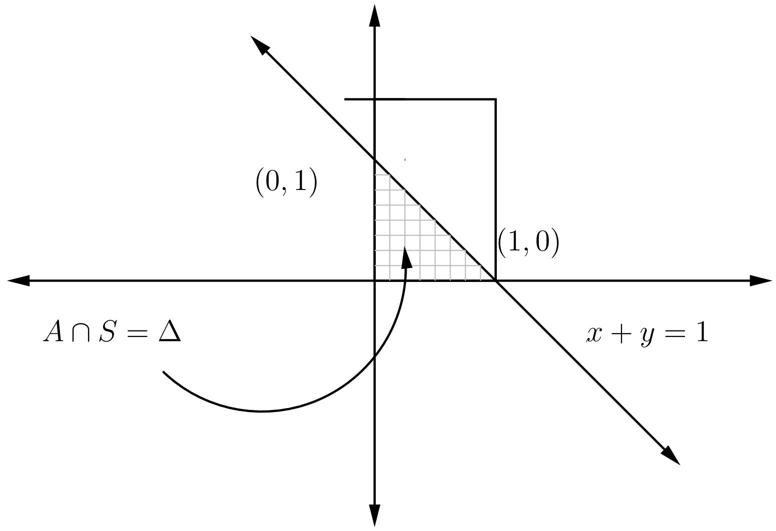


Figure 6.2: Above is a representation of the intersection of A and the support.

Example. On page 144, example 2, we have a pair of continuous random variables (X, Y) with joint density:

$$f(x,y) = ce^{-\frac{(x^2-xy+y^2)}{2}}$$

Notice that the support of this function is all of $\mathbb{R}^2$. So, in finding c, we have to integrate as such:

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ce^{-\frac{(x^2-xy+y^2)}{2}}\ dx\ dy$$

completing the square, we have

$$x^2 - xy + y^2 = (x - \frac{y}{2})^2 + (y^2 - \left(\frac{y}{2}\right)^2) = (x - \frac{y}{2})^2 + \frac{3y^2}{4}$$

So,

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ce^{-\frac{(x^2-xy+y^2)}{2}}\ dx\ dy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-\frac{(x-\frac{y}{2})^2}{2}} e^{-\frac{3y^2}{8}} dx\right) dy$$

now notice that:

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\frac{y}{2})^2}{2}}\ dx$$

takes the form:

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sqrt{2\pi}\sigma$$

so, letting $\sigma = 1$ and $\mu = \frac{y}{2}$:

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\frac{y}{2})^2}{2}}\ dx = \sqrt{2\pi}$$

so what we have is:

$$\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-\frac{(x-\frac{y}{2})^2}{2}} e^{-\frac{3y^2}{8}}\ dx\right) dy = \sqrt{2\pi}\int_{-\infty}^{\infty} e^{-\frac{3y^2}{8}} dy = \sqrt{2\pi}\int_{-\infty}^{\infty} e^{-\frac{(y-\mu)^2}{2\cdot\left(\frac{2}{\sqrt{3}}\right)^2}} dy = \sqrt{2\pi}\cdot\sqrt{2\pi}\cdot\frac{2}{\sqrt{3}} = \frac{4\pi}{\sqrt{3}}$$

6.3 Chapter 6, Question 1

X and Y are continuous random variables with a joint density function $f(x, y)$. We would like to look at the following random variables:

$$W = a + bX$$
$$Z = c + dY$$

Where $b > 0$, $d > 0$ and a, c are constants. We would like to show that if X, Y are independent, then W, Z are also independent. Recall that if we have a continuous random variable X with a density $f(x)$ and if we have a function $\varphi(x)$ that is strictly monotonic, such that:

$$Y = \varphi(X)$$

Then $g(y)$ is the density of Y where:

$$g(y) = f(x)\left|\frac{dx}{dy}\right| = \frac{f(x)}{|\frac{dy}{dx}|}$$

While there is a way to continue in the fashion, we turn to a new method of solving these types of questions (this is a result from pages 166-168). Suppose we have a pair of continuous random variables (X, Y) with joint density $f(x, y)$. We can define random variables U, V by the following functions:

$$U = u(X, Y), \qquad V = v(X, Y)$$

The **Jacobian Transformation Matrix**, denoted J, is determinant of the following matrix:

$$J = det\left(\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}\right)$$

Then, the joint density of (U, V) is:

$$g(u, v) = \frac{1}{|J|} f(x, y)$$

Notice that in applying this to our question, we can say the following:

$$u(x, y) = a + bx$$
$$v(x, y) = c + dy$$

Now looking at the Jacobian, we have the following:

$$J = det\left(\begin{bmatrix} b & 0 \\ 0 & d \end{bmatrix}\right) = bd$$

So the joint of (U, V) is:

$$g(u, v) = \frac{1}{|bd|} f(x, y)$$

Using our definitions of the functions u, v, we have:

$$u = a + bx \to x = \frac{u-a}{b}$$
$$v = c + dy \to y = \frac{v-c}{d}$$

So,

$$g(u,v) = \frac{1}{|bd|} f\left(\frac{u-a}{b}, \frac{v-c}{d}\right)$$

For the second half of this question, suppose that X and Y are independent. This tells us that:

$$f(x,y) = \alpha(x)\beta(y)$$

where α, β are not necessarily the marginals of X, Y. From our answer to part (1), we notice that:

$$g(u,v) = \frac{1}{|bd|}\,\alpha\left(\frac{u-a}{b}\right)\beta\left(\frac{v-c}{d}\right)$$

$g(u,v)$ can be represented as the product of two functions α, β, one of u and one of v, which tells us that U, V are independent. While there is a direct approach to this question involving double integrals, notice that we did not need it to solve this problem.

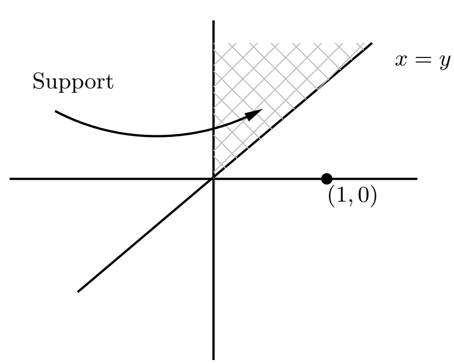
6.4 Chapter 6, Question 7

Let X, Y be continuous random variable with the joint density:

$$f(x,y) = \lambda^2 e^{-\lambda y}, \quad 0 \le x \le y$$

and $f(x,y) = 0$ elsewhere. Let us call $S = 0 \le x \le y$ the support of this density. We have the following:

$$F(x,y) = \int\int_{A(x,y)} f(s,t)dsdt$$



Clearly if (x,y) is in quadrant 2,3 or 4, then $F(x,y) = 0$. So, we focus only quadrant one. We can split this problem into two cases.

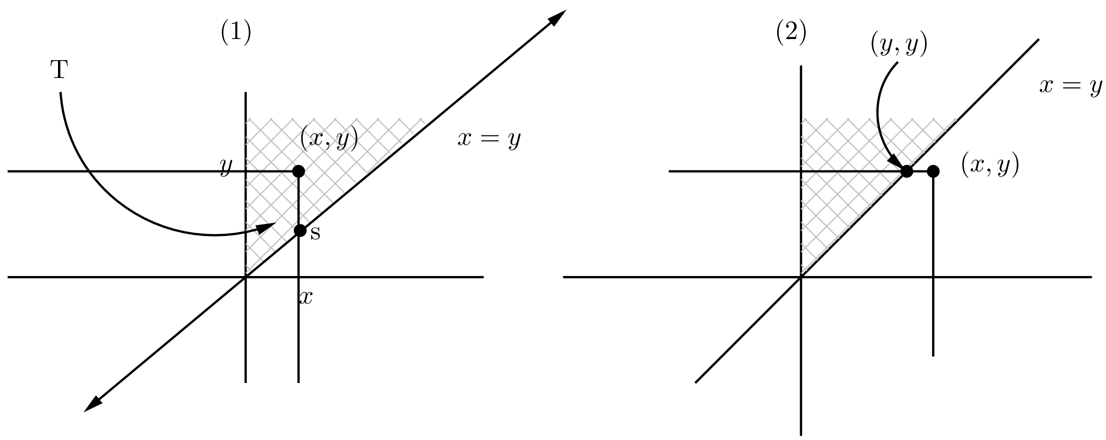


Figure 6.3: Case (1) and (2)

In case one, $(x, y) \in$ Quadrant 1 $\cap S = T$, so:

$$F(x,y) = \int\int_T \lambda^2 e^{-\lambda y} = \int_0^x \left(\int_s^y \lambda^2 e^{-\lambda t} dt\right) ds$$

$$\text{where:} \quad \int_s^y \lambda^2 e^{-\lambda t} dt = -\lambda e^{-\lambda t}\Big|_{t=s}^{t=y} = \lambda e^{-\lambda s} - \lambda e^{-\lambda y}$$

So we have:

$$\int_0^x (\lambda e^{-\lambda s} - \lambda e^{-\lambda y}) ds = \int_0^x \lambda e^{-\lambda s} ds - \int_0^x \lambda e^{-\lambda y} ds$$

$$\int_0^x \lambda e^{-\lambda s} ds = -e^{-\lambda s}\Big|_{s=0}^{s=x} = 1 - e^{-\lambda x}$$

$$\int_0^x \lambda e^{-\lambda y} ds = \lambda s e^{-\lambda y}\Big|_{s=0}^{s=x} = \lambda x e^{-\lambda y}$$

so,

$$F(x,y) = 1 - e^{-\lambda x} - \lambda x e^{-\lambda y}$$

in case two, suppose that $(x, y) \in I \bigcap S^c$. Thanks to the bounds on x and y, we can see that $F(x, y) = F(y, y)$ and from our previous result,

$$F(y,y) = 1 - e^{-\lambda y} - \lambda y e^{-\lambda y}$$

And we have our distribution function for both cases. Looking at our support, since we now that $X \geq 0$, we know that $f_X(x) = 0$ if $x \leq 0$. So, for $x > 0$ and fixed.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy = \int_x^{\infty} \lambda^2 e^{-\lambda y} dy = -\lambda e^{-\lambda y}\Big|_{y=x}^{y=\infty} = \lambda e^{-\lambda x} - 0$$

so,

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for all } x > 0 \\ 0 & \text{otherwise} \end{cases} = Exp(\lambda)$$

Since $Y \geq 0$, this tells us that $f_Y(y) = 0$ for all $y \leq 0$. Fix some $y > 0$. We know that :

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^y \lambda^2 e^{-\lambda y} dx = \lambda^2 y e^{-\lambda y}$$

so,

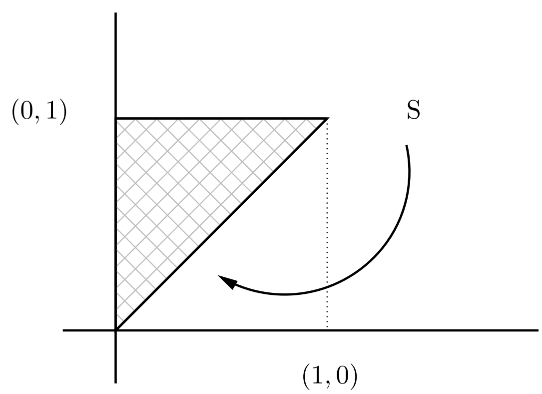
$$f_Y(y) = \begin{cases} \lambda^2 y e^{-\lambda y} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases} \Bigg| \quad = \Gamma(\alpha = 2, \lambda = \lambda)$$

6.5 Chapter 6, Question 8

We have the following:

$$f(x,y) = \begin{cases} c(y-x)^\alpha & \text{if} \quad 0 \leq x < y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The support of this density looks like the following:



Assume that c, α make $f(x,y)$ into a joint density. In this case, $0 \leq X \leq 1$, and

$$f_X(x) = 0 \quad \text{ if } x \leq 0 \text{ or } x \geq 1$$

so, take and fix $0 < x < 1$.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_x^1 f(x,y)dy = \int_x^1 c(y-x)^\alpha dy = c\frac{(y-x)^{\alpha+1}}{\alpha+1}\Bigg|_{y=x}^{y=1}$$

from this, we see that we need $\alpha + 1$ to be positive. If $\alpha + 1$ is positive, then from $0 < x < 1$,

$$f_X(x) = c\frac{(y-x)^{\alpha+1}}{\alpha+1}\Bigg|_{y=x}^{y=1} = c\frac{(1-x)^{\alpha+1}}{\alpha+1} - 0$$

for this to be a density, the following must be true:

$$\iint f(x,y)dydx = 1 = \int_0^1 c\frac{(1-x)^{\alpha+1}}{\alpha+1}dx = -\frac{c(1-x)^{\alpha+1}}{(\alpha+1)(\alpha+2)}\Bigg|_{x=0}^{x=1} = \frac{c}{(\alpha+1)(\alpha+1)}$$

So in summary, we need $c > 0$, $\alpha + 1 > 0$,and $(\alpha > -1)$. In fact, this result tells us that: $c = (\alpha+1)(\alpha+2)$. Looking back at the marginals, we see that:

$$f_X(x) = \begin{cases} (\alpha+2)(1-x)^{\alpha+1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Definition. For $\alpha > 0, \beta > 0$,

$$f(x) = \begin{cases} cx^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

is called a $B(\alpha, \beta)$ density.

Suppose that (X, Y) have the joint density for $f(x, y)$. Let:

$$Z = X + Y$$

We would like to find the density $f_Z(z)$ in terms of $f(x, y)$. We have the following:
Theorem 20.

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)dx$$

As a homework, it is recommended that we memorize this formula and read the proof of its derivation in the book. Notice that if X, Y are independent, then

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$$

Suppose that $X \geq 0$ and $Y \geq 0$. This tells us that $Z \geq 0$. So, for $z > 0$,

$$f_Z(z) = \int_0^z f_X(x)f_Y(z-x)dx$$

As a homework, try the following: take $X = Exp(\lambda)$, $Y = Exp(\lambda)$ where X and Y are independent. Find the density of their sum.

6.6 Chapter 6, Question 21

Let X, Y be independent random variables each having an exponential distribution with parameter λ. We want to find the density of the random variable $Z = \frac{Y}{X}$.

Look at the transformation

$$(X, Y) \rightarrow (U, V)$$

where we associate f with X, Y and g with U, V, $u = u(x, y) = x$ and $v = v(x, y) = \frac{y}{x}$ We want to find x, y in terms of u, v. We have the following:

$$x = u$$
$$y = u \cdot v$$

So in finding the Jacobian,

$$J = det\left(\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}\right) = det\left(\begin{bmatrix} 1 & 0 \\ -\frac{y}{x^2} & \frac{1}{x} \end{bmatrix}\right) = \frac{1}{x} = \frac{1}{u}$$

so putting all our pieces back into our equation, we get:

$$g(u, v) = \frac{1}{\frac{1}{u}} f(u, uv) = uf(u, uv) = xf(u, uv)$$

In finding the marginal density of V, which is equal to $\frac{Y}{X}$, we have:

$$f_V(v) = \int_0^\infty g(v, u)\, du = \int_0^\infty uf(u, uv)\, du$$

so,

$$f_{\frac{Y}{Z}}(z) = \int_0^\infty xf(x, xz)dx$$

Notice that this is identical to equation 22 on page 151. When X, Y are independent and identically distributed with density $f(x)$, then:

$$f_{\frac{Y}{X}}(z) = \int_0^\infty xf(x)f(xz)dx$$

for all $z > 0$. In our case, $x > 0$, $f(x) = \lambda e^{-\lambda x}$, so:

$$f_{\frac{Y}{X}}(z) = g(z) = \int_0^\infty xf(x)f(xz)dx = \lambda^2 \int_0^\infty xe^{-\lambda x}e^{-\lambda xz}dx = \lambda^2 \int_0^\infty xe^{-\lambda x(z+1)}dx$$

so after evaluating, our conclusion is that:

$$f_{\frac{Y}{X}}(z) = g(z) = \frac{1}{(1+z)^2}, \quad \text{if } z > 0 \text{ and } 0 \text{ otherwise.}$$

6.7 Chapter 6, Question 13

Let X and Y be independent and uniformly distributed on the interval (a, b). We want to find the density of $Z = |Y - X|$. Suppose we have the following: $W = Y - X$. We would like to find $f_W(w)$. Looking at the following transformation:

$$(X, Y) \to (U = X, V = Y - X)$$

we have $f(x, y)$ and we define the density of the other pair $g(u, v)$. We assign the following:

$$u = x$$
$$v = y - x$$

which is equivalent to:

$$x = u$$
$$y = u + v$$

So in finding the Jacobian, we have:

$$J = det\left(\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}\right) = det\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = 1$$

so,

$$g(u, v) = 1 \cdot f(u, u + v)$$

So in finding the marginal density of V, which is equal to $Y - X = W$, we have:

$$f_W(w) = \int_{-\infty}^{\infty} f(u, u + w)\, du$$

and when X, Y are independent,

$$f_W(w) = \int_{-\infty}^{\infty} f(u, u + w)\, du = \int_{-\infty}^{\infty} f_X(x) f_Y(x + w) dx$$

To proceed, we need to solve the following inequalities:

$$a \leq x \leq b \quad \text{and} \quad a \leq x + w \leq b$$

$$\Rightarrow max(a, a - w) \leq x \leq min(b, b - w)$$

This splits us into two cases:

1. Case (1), $w < 0$. In this case, we have the following:

$$f_W(w) = \int_{max(a,a-w)}^{min(b,b-w)} f_X(x) f_Y(x + w) dx = \int_{a-w}^{b} \frac{1}{(b-a)^2} dx = \frac{(b-a) + w}{(b-a)^2}$$

2. Case (2), $w > 0$. In this case, we have the following:

$$f_W(w) = \int_{max(a,a-w)}^{min(b,b-w)} f_X(x) f_Y(x + w) dx = \int_{a}^{b-w} \frac{1}{(b-a)^2} dx = \frac{(b-a) - w}{(b-a)^2}$$

Now $Z = |W|$ has the following density h for some z such that: $0 \leq z \leq b - a$:

$$h(z) = f_W(z) + f_W(-z) = \frac{(b-a) - z}{(b-a)^2} + \frac{(b-a) - z}{(b-a)^2} = \frac{2}{b-a}\left[1 - \frac{z}{b-a}\right]$$

6.8 Finding the Density of the Absolute Value of a Continuous Random Variable

Suppose X is a continuous random variable with density $f(x)$. Letting $Y = |X|$, we wish to find the density $g(y)$ of Y in terms of $f(x)$.

We can suppose that we have $F(x)$, the cumulative distribution function of X, and $G(y)$, the cumulative distribution function of Y. We have:

$$G(y) = P(Y = |X| \leq y) = 0 \quad \text{if } y < 0$$

For $y > 0$,

$$G(y) = P(-y \leq X \leq y) = F(y) - F(-y)$$

taking the derivative, we have:

$$G'(y) = g(y) = \begin{cases} f(y) + f(-y) & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

6.9 Chapter 6, Question 4

Let X, Y be independent random variables having the normal density $N(0, \sigma^2)$. Find $P(X^2 + Y^2 \leq 1)$.

We know that X^2 has a particular density:

$$X^2 = \Gamma\left(\alpha = \frac{1}{2},\ \lambda = \frac{1}{2\sigma^2}\right)$$

(Look at example 12 on page 128 to help verify this fact). Recall that we also have the following: If $X = \Gamma(\alpha_1, \beta)$ and $Y = \Gamma(\alpha_2, \beta)$ (with the same β), where X and Y are independent,

$$X + Y = \Gamma(\alpha_1 + \alpha_2, \beta)$$

Looking at theorem 5 on page 159, we have this result for n random variables. So, we see that:

$$X^2 = \Gamma\left(\frac{1}{2},\ \frac{1}{2\sigma^2}\right) \quad Y^2 = \Gamma\left(\frac{1}{2},\ \frac{1}{2\sigma^2}\right)$$

and since X, Y are independent, we use our result to get:

$$X^2 + Y^2 = \Gamma\left(1, \frac{1}{2\sigma^2}\right)$$

Remember that we have the following definition of Γ-distributed variables:
Definition. $X = \Gamma(\alpha, \lambda)$, if $X \geq 0, \alpha, \lambda > 0$, has density

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \text{ if } x > 0 \\ 0 \text{ otherwise} \end{cases}$$

So, since the first parameter of $X^2 + Y^2$ is 1 its density looks like the following when $X > 0$:

$$X^2 + Y^2 = \frac{\lambda^1}{1} x^0 e^{-\lambda x} = Exp\left(\lambda = \frac{1}{2\sigma^2}\right)$$

and if $T = Exp(\lambda)$, $P(T \geq a) = e^{-\lambda a}$. So,

$$P(X^2 + Y^2 \leq 1) = 1 - P(x^2 + Y^2 > 1) = 1 - e^{-\frac{1}{2\sigma^2}}$$

Theorem 21. *If* $X = \Gamma(\alpha_1, \lambda)$, $Y = \Gamma(\alpha_2, \lambda)$, *then* $X + Y = \Gamma(\alpha_1 + \alpha_2, \lambda)$

Proof. $Z = X + Y$ has a density given to us by the convolution. Thus,

$$g(z) = \int_0^z f_X(x) f_Y(z - x) dx$$

Now, $f_X(x) = c \cdot x^{\alpha_1 - 2} e^{-\lambda x}$ and $f_Y(z - x) = d \cdot (z - x)^{\alpha_2 - 1} e^{-\lambda(z-x)}$ So,

$$g(z) = \int_0^z f_X(x) f_Y(z-x) dx = \int_0^z c d x^{\alpha_1 - 1} (z-x)^{\alpha_2 - 1} e^{-\lambda z}\ dx = c d e^{-\lambda z} \int_0^z x^{\alpha_1 - 1} (z-x)^{\alpha_2 - 1} dx$$

for $z > 0$ fixed. It turns out that:

$$\int_0^1 x^{\alpha - 1}(1 - x)^{\beta - 1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Now call $x = tz$. Now, $dx = zdt$, and we have the following:

$$\int_0^z x^{\alpha_1 - 1}(z - x)^{\alpha_2 - 1} dx = \int_0^1 t^{\alpha_1 - 1} z^{\alpha_1 - 1} \cdot z^{\alpha_2 - 1}(1 - t)^{\alpha_2 - 1} z dt$$

taking out some constants, we have:

$$z^{\alpha_1 + \alpha_2 - 1} \int_0^1 t^{\alpha_1 - 1}(1 - t)^{\alpha_2 - 1} dt = z^{\alpha_1 + \alpha_2 + 1} \cdot \mathbf{b}$$

where $\mathbf{b}$ is a constant we will relate to the Γ function. Putting the pieces back together, we have for $z > 0$:

$$g(z) = \mathbf{b} c d\ z^{\alpha_1 + \alpha_2 - 1} e^{-\lambda z}$$

□

We know two things:

1. $g(z)$ is a density
2. $g(z)$ is a density of a random variable of the type: $\Gamma(\alpha_1 + \alpha_2, \lambda)$.

So, there are some constants that must be the same. By the definition of a Γ random variable, the following are true:

$$c = \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} \qquad d = \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)}$$

and the constant of the density $\Gamma(\alpha_1+\alpha_2,\lambda)$ is:

$$\frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}$$

Since the constants have to agree, we have:

$$\mathbf{b}cd = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_1)}$$

or,

$$\mathbf{b}\cdot\frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}$$

therefore,

$$\mathbf{b} = B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

where $B(\alpha,\beta)$ is called the **Beta Function**.
Definition. The following is the density of a **Beta Density** with parameters α_1, α_2:

$$f(x) = \begin{cases} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\cdot x^{\alpha_1-1}(1-x)^{\alpha_2-1} & 0<x<1 \\ 0 & \text{elsewhere} \end{cases}$$

6.10 Chapter 6, Question 19

Let X, Y be independent random variables each having the normal density $N(0,\sigma^2)$. This tells us that:

$$f_X(x) = f_Y(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-x^2}{2\sigma^2}}$$

We want to show that Y/X and $Y/|X|$ both have the Cauchy Density.

The **Cauchy Density** is the following:

$$f(x) = c\cdot\frac{1}{1+x^2}$$

where $x\in\mathbb{R}$, and the constant is such that f is a valid density. We know that:

$$\int_{-\infty}^{\infty}\frac{dx}{1+x^2} = arctan(x)\Bigg|_{x=-\infty}^{x=\infty} = \frac{\pi}{2}-\left(-\frac{\pi}{2}\right) = \pi$$

so, $c=\frac{1}{\pi}$. Unfortunately, if X is a variable with this density, the expected value $E(X)$ does not exist - recall that $E(X)$ exists if and only if:

$$\int_{-\infty}^{\infty}|x|f(x)dx<\infty$$

This can be shown by noticing the following:

$$\int_0^\infty \frac{x}{1+x^2}dx = \frac{1}{2}ln(1+x^2)\Bigg|_{x=0}^{x=\infty} \qquad \text{diverges at } x=\infty$$

so, $E(X)$ does **not** exist. Back to our question, we introduce the variable $Z=\frac{Y}{X}$. Then, by our formula,

$$f_Z(z) = f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz)dx = \int_{-\infty}^{\infty} |x|\frac{1}{2\pi\sigma^2}e^{-\frac{x^2}{2\sigma^2}(1+z^2)}dx$$

Notice that if $g(-x)=g(x)$,

$$\int_{-\infty}^{\infty} g(x)dx = 2\int_0^\infty g(x)dx$$

so,

$$\int_{-\infty}^{\infty} |x|\frac{1}{2\pi\sigma^2}e^{-\frac{x^2}{2\sigma^2}(1+z^2)}dx = 2\int_0^\infty x\cdot\frac{1}{2\pi\sigma^2}e^{-\frac{x^2}{2\sigma^2}(1+z^2)}dx$$

calling $\frac{1+z^2}{2\sigma^2}=a$, and noticing the following:

$$\int_0^\infty xe^{-ax^2}dx = \frac{e^{-ax^2}}{-2a}\Bigg|_{x=0}^{\infty} = \frac{1}{2a}$$

So,

$$2\int_0^\infty x\cdot\frac{1}{2\pi\sigma^2}e^{-\frac{x^2}{2\sigma^2}(1+z^2)}dx = \frac{1}{2\pi\sigma^2}\cdot 2\cdot\frac{1}{\frac{1+z^2}{\sigma^2}} = \frac{1}{\pi(1+z^2)}$$

so, we have finished our objective.

6.11 Chapter 5, Question 31

Let X have the normal density $N(0,\sigma^2)$. We want to find the density of $Y=|X|$. Associate F,f with X and G,g with Y. Clearly, $G(y)=0$ if $y\le 0$. For $y>0$,

$$G(y) = P(|X|\le y) = P(-y\le X\le y) = F(y)-F(y) \Rightarrow g(y) = f(y)+f(-y)$$

For $Y=|X|$, if $X=N(0,\sigma^2)$ then the density of Y for $y>0$ is:

$$g(y) = \sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sigma}e^{\frac{-y^2}{2\sigma^2}} \qquad \dagger$$

† notice that since $g(y)=f(y)+f(-y)$, the 2 comes out and changes the coefficient to $\sqrt{\frac{2}{\pi}}$

6.12 Chapter 6, Question 29

Let $X_1, ..., X_n$ be independent random variables having a common normal density. We want to show that there exist constants A_n, B_n such that:

$$\frac{X_1 + ... + X_n - A_n}{B_n}$$

has the same density as X_1, $N(\mu, \sigma^2)$. We know that $X_1 + X_2 + ... + X_n = N$ add up to a variable that is normal, based on what we know about the sum of normally distributed random variables. Notice that:

$$X_1 + X_2 + ... + X_n = N(n\mu, n\sigma^2)$$

Since $aU + b$ is normal when U is a normal random variable, $a \neq 0$ and b constants. So, let us notice the following:

$$Y = \frac{X_1 + ... + X_n - A_n}{B_n} = N(\mu, \sigma^2)$$

where now,

$$Y = \frac{1}{B_n}S + \left(-\frac{A_n}{B_n}\right)$$

Thus, we can see that for any $A_n, b_n \neq 0$ Y must be normal with the following parameters:

$$Y = \frac{1}{B_n}S + \left(-\frac{A_n}{B_n}\right) = N(\frac{n\mu}{B_n} - \frac{A_n}{B_n}, \frac{n\sigma^2}{B_n^2})$$

But, we want the following:

$$N(\frac{n\mu}{B_n} - \frac{A_n}{B_n}, \frac{n\sigma^2}{B_n^2}) = N(\mu, \sigma^2)$$

So, we match the following:

$$\frac{n\sigma^2}{B_n^2} = \sigma^2 \Rightarrow B_n = \sqrt{n}$$

$$\mu = \frac{n\mu - A_n}{\sqrt{n}} \Rightarrow A_n = n\mu - \sqrt{n}\mu = (n - \sqrt{n})\mu$$

6.13 Chapter 6, Question 9

We have the following:

$$f(x, y) = ce^{(-x^2 - xy + 4y^2)/2} \quad -\infty < x, y < \infty$$

We would like to chose c such that f is a density, and we would like to find the marginal densities of f. Notice that:

$$x^2 - xy + 4y^2 = (x - \frac{y}{2})^2 + \frac{15y^2}{4}$$

so,

$$f(x,y) = ce^{(-x^2-xy+4y^2)/2} = ce^{-\frac{(x-y/2)^2}{2}}e^{-15y^2/8}$$

In finding the marginal density of Y, we have:

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx = c \cdot e^{-15y^2/8} \cdot \sqrt{2\pi} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-y/2)^2}{2}}}_{N(\mu=\frac{y}{2},\sigma^2=1)} = (c\sqrt{2\pi})e^{-15y^2/8} \cdot 1 \qquad \ddagger$$

now notice that:

$$e^{-15y^2/8} = e^{-\frac{y^2}{2\cdot\frac{4}{15}}}$$

Our variable takes the form:

$$(c\sqrt{2\pi})e^{-15y^2/8} = Normal(0, 4/15)$$

But for a normal variable of this form, the normalizing constant is as follows:

$$\frac{1}{\sqrt{2\pi} \cdot 2/\sqrt{15}}$$

which must be equal to

$$(c\sqrt{2\pi})$$

so, equating them, we solve for c:

$$c(\sqrt{2\pi}) = \frac{\sqrt{15}}{2\sqrt{2\pi}} \Rightarrow c = \frac{\sqrt{15}}{4\pi}$$

and we have the marginal of Y. Alternatively, in finding the marginal for X, notice that:

$$4y^2 - xy + x^2 = (2y - \frac{x}{4})^2 + \frac{15x^2}{16} = 4(y - \frac{x}{8})^2 + \frac{15x^2}{16}$$

and we can proceed in the same was as we did with the marginal for Y. It will again be a normal random variable.

6.14 Chapter 6, Question 30

Let X_1, X_2, X_3 be identically distributed random variables on $(0,1)$, all uniformly distributed. Call $U = X_1 + X_2$, and $V = X_3$. We aim to find the density of the random variable $Y = X_1 + X_2 + X_3$. On page 147, we have the result:

$$f_u(u) = \begin{cases} u & \text{if } 0 < u < 1 \\ 2-u & \text{if } 1 < u < 2 \\ 0 & \text{otherwise} \end{cases}$$

‡The integral evaluates to 1 based on what we know about the Normal density.

We know that U, V are independent, since they are functions of independent random variables. We now call $W = U + V$. The support of w is as follows; $f(w) = o$ if $w \notin [0, 3]$. We now need to use the convolution:

$$f_W(w) = \int_0^w f_U(u) f_V(w-u) du$$

We notice that the support of f_V is from $[0, 1]$. Now notice that:

$$f_U(u) \cdot f_V(w-u) = \begin{cases} f_U(u) & \text{if } 0 < u < 2 \quad \text{and } 0 < w - u < 1 \\ 0 & \text{otherwise} \end{cases}$$

Now let us look at this double inequality. Notice that:

$$0 < u < 2 \quad \text{and } 0 < w - u < 1$$

tells us that:

$$A = max(0, w-1) < u < min(2, w) = B$$

so,

$$f_W(w) = \int_0^w f_U(u) f_V(w-u) du = \int_A^B f_U(u) du$$

Which gives us the following three cases:

1. $0 < w < 1$
2. $1 < w < 2$
3. $2 < w < 3$

We have the following answers:

1. $A = 0, B = w$. In this case,

$$f_W(w) = \int_0^w f_U(u) du = \left. \frac{u^2}{2} \right|_0^w = \frac{w^2}{2}$$

2. $A = w - 1, B = 2$. In this case,

$$f_W(w) = \int_{w-1}^2 f_U(u) du = \int_{w-1}^1 u du + \int_1^w (2-u) du = -w^2 + 3w - \frac{3}{2}$$

3. Case 3 was left as an exercise, but the answer is:

$$f_W(w) = \frac{w^2}{2} - 3w + \frac{9}{2}$$

6.15 Chapter 6, Question 12

Let X, Y have a joint density f:

$$f(x,y) = \begin{cases} (\alpha+1)(\alpha+2)(y-x)^{\alpha} & \text{if } 0 \le x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

The support of $Z = X + Y$ is one of the triangles constructed by the unit square and the line $y = x$. We want to find the density of Z, which is given by:

$$g(z) = \int_{-\infty}^{\infty} f(x, z-x) dx$$

We know that $0 \le X + Y \le 2$, based on what we know about X and Y. So, $g(z) = 0$ if $z \notin [0,2]$. So, fix such a $z \in [0,2]$. We need $0 \le x < y = z - x \le 1$, which happens if and only if:

$$A = max(0, z-1) < x < \frac{z}{2}$$

So we have to deal with the two cases in which $z < 1$, and $z > 1$. Then,

$$g(z) = c \int_{A}^{\frac{z}{2}} (z-2x)^{\alpha} dx = c \cdot \frac{(z-2x)^{\alpha+1}}{-2(\alpha+1)} \Bigg|_{x=A}^{x=z-2}$$

where $c = (\alpha+1)(\alpha+2)$. In case one, we let $0 < z < 1$. In this case, $A = 0$, so:

$$g(z) = \frac{(\alpha+2)}{-2}(z-2x)^{\alpha+1} \Bigg|_{x=1}^{x=\frac{z}{2}} = \frac{\alpha+2}{-2}(0 - z^{\alpha+1})$$

and in case two, where $1 < z < 2$, this tells us that $A = z - 1$ so,

$$g(z) = \frac{(\alpha+2)}{-2}(z-2x)^{\alpha+1} \Bigg|_{x=z-1}^{x=\frac{z}{2}}$$

6.16 Chapter 6, Question 10

Let X and Y be continuous random variables having joint density f. We want to derive a formula for the density of the random variable $Z = Y - X$.

Let us make the following transformation;

$$(X, Y) \to (U, V)$$

using the following functions:

$$u = x \to x = u$$
$$v = y - x \to y = u + v$$

We define $g(u,v)$ to be the joint density of (U,V). Notice that:

$$g(u,v) = f(u, u+v)$$

And that the marginal density of $V = Y - X$, which is the density we want to find. Recall that we have the following formula:

$$g(u,v) = \frac{1}{|J|} f(x,y)$$

So in calculating the Jacobian,

$$J = det\left(\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\right) = 1$$

So,

$$h_V(v) = \int_{-\infty}^{\infty} 1 \cdot f(u, u+v) du$$

6.17 Chapter 6, Question 14

Let X and Y be continuous random variables having joint density f. We want to derive a formula for the density of $Z = aX + bY$, where $b \neq 0$.

As we did in the previous question, we transform $(X,Y) \to (U,V)$ via the following transformations:

$$\begin{aligned} u &= x \\ v &= ax + by \end{aligned}$$

where a, b are non-zero constants. In calculating the Jacobian, we can see that $J = b$. This implies that:

$$f_{aX+bY}(v) = f_V(v) = \frac{1}{|b|} \int_{-\infty}^{\infty} f\left(x, \frac{v - ax}{b}\right) dx$$

6.18 Chapter 6, Question 16

Let X and Y be independent random variables having respective normal densities $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. We want to find the density of $Z = Y - X$. We have the following theorem:

Theorem 22. *Let $X_1, ..., X_n$ be independent random variables such that X_m has the normal density $N(\mu_m, \sigma_m^2)$, where $m = 1, 2, ...n$. Then, $X_1 + ... + X_n$ has the normal density $N(\mu, \sigma^2)$ where:*

$$\mu = \mu_1 + \mu_2 + ... + \mu_n \qquad \sigma^2 = \sigma_1^2 + \sigma_2^2 + \sigma_n^2$$

Using this theorem, the density of Z is as follows:

$$Z = Y - Z = N(\mu_1, \sigma_1^2) - N(\mu_2, \sigma_2^2) = N(\mu_1, \sigma_1^2) + N(-\mu_2, \sigma_2^2) = N(\mu_2 - \mu_1, \sigma_1^2 + \sigma_2^2)$$

6.19 Chapter 6, Question 18

Let X and Y be continuous random variables having joint density f. We would like to derive a formula for the density of $Z = XY$.

We transform the pair of random variables (X, Y) to the pair (U, V) under the following functions:

$$u = x \to x = u$$
$$v = xy \to y = \frac{v}{x} = \frac{v}{u}$$

Recall that we associate the function f with the density of (X, Y) and the function g with the density of (U, V). In calculating the Jacobian, we have:

$$J = det\left(\begin{bmatrix} 1 & 0 \\ y & x \end{bmatrix}\right) = x$$

So, the density $g(u, v)$ is the following:

$$g(u, v) = \frac{1}{|x|} f(x, y) = \frac{1}{|u|} f\left(u, \frac{v}{u}\right)$$

Since we defined V to be: $V = XY$, finding the marginal density of V is the same as finding the density of XY:

$$f_{XY}(z) = f_V(v) = \int_{-\infty}^{\infty} \frac{1}{|u|} f\left(u, \frac{v}{u}\right) du$$

6.20 Chapter 6, Question 22

Let X and Y be independent random variables having respective gamma densities $\Gamma(\alpha_1, \lambda)$ and $\Gamma(\alpha_2, \lambda)$. We want to find the density of $Z = X/(X + Y)$.

We can transform the pair of random variables (X, Y) to the pair (U, V) with the following functions:

$$u = x$$
$$v = \frac{x}{x + y}$$

The steps to find $g(u, v)$, the density of the pair (U, V) can be found using the same method we've used before. We have the following:

$$g(u, v) = \frac{1}{|J|} f(x, y) = \frac{u}{v^2}\ f(u, \frac{u}{v} - u)$$

since $J = \frac{-x}{(x+y)^2}$. So,

$$\begin{aligned} g(u,v) &= \frac{u}{v^2} \cdot f(u) f(\frac{u}{v} - u) \\ &= \frac{u}{v^2} \cdot \frac{\lambda_1^{\alpha}}{\Gamma(a_1)} u^{\alpha_1 - 1} e^{-\lambda u} \cdot \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)} (\frac{u}{v} - u)^{\alpha_2 - 1} e^{-\lambda(\frac{u}{v} - u)} \\ &= \frac{\lambda^{\alpha_1}\lambda^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1+\alpha_2-1} \frac{(1-v)^{\alpha_2-1}}{v^{\alpha_2-1}} e^{-\lambda \frac{u}{v}} \end{aligned}$$

So in finding the density of $\frac{X}{X+Y}$, we can find the density of $g_V(v)$, which is the following:

$$g_V(v) = \int_0^\infty g(u,v) du = (Constant) \cdot \frac{(1-v)^{\alpha_2-1}}{v^{\alpha_2+1}} \cdot \frac{\Gamma(\alpha_1+\alpha_2)}{\left(\frac{\lambda}{v}\right)^{\alpha_1+\alpha_2}}$$

which follows from noticing that:

$$\int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^\alpha} \Rightarrow \int_0^\infty u^{\alpha_1+\alpha_2-1} e^{-\lambda \frac{u}{v}} = \frac{\Gamma(\alpha_1+\alpha_2)}{\left(\frac{\lambda}{v}\right)^{\alpha_1+\alpha_2}}$$

6.21 Test Review

Suppose you have $f(x)$, with support $= (a,b)$. We can find $F(x)$, the cumulative distribution function f, by the following:

$$F(x) = \int_{-\infty}^{x} f(t) dt$$

Now notice that $F(x)$ is strictly increasing on the interval (a,b). So,

$$F(a,b) \to (0,1)$$

must have an inverse. Let $\varphi = F^{-1}$. Then $X = \varphi(U)$ where $U =$Uniform$[0,1]$, and X has the density $f(x)$ and c.d.f. $F(x)$. We have the following example from the formula:

6.22 Chapter 5, Question 44

Given $U = Uniform(0,1)$,

$$f(x) = \begin{cases} 2x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

We want to find a function φ so that $\varphi(U) = X$ has density $f(x)$. We know the following:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$

So,

$$y = F(x) = x^2 : (0,1) \to (0,1)$$

Solving $F(x)$ for x, we have:

$$x = \sqrt{y} = F^{-1}(y)$$

And so, $\varphi(x) = \sqrt{x}$. So, $\sqrt{U}$ has the density $f(x)$.

6.22.1 Chapter 5, Question 43

Let $X = \Gamma(\alpha, \lambda)$. Let $Y = \sqrt{X}$. We would like to find the density of Y, which we will call $g(y)$. We see that:

$$Y = \varphi(X) \quad \varphi(x) = \sqrt{x}$$

is strictly increasing. According to a theorem in the book, since this function φ is strictly increasing,

$$g(y) = \left|\frac{dx}{dy}\right| f(x)$$

Since:

$$y = \varphi(x) = \sqrt{x}, \quad \Rightarrow \quad x = y^2 \Rightarrow \frac{dx}{dy} = 2y$$

So,

$$g(y) = \left|\frac{dx}{dy}\right| f(x) = f(y^2) \cdot 2y$$

and so,

$$g(y) = \begin{cases} 2y \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)}(y^2)^{\alpha-1}e^{-\lambda y^2} = \frac{2\lambda^\alpha}{\Gamma(\alpha)}y^{2\alpha-1}e^{-\lambda y^2} & \text{if} y > 0 \\ 0 \quad \text{otherwise} \end{cases}$$

6.22.2 Chapter 6, Question 5

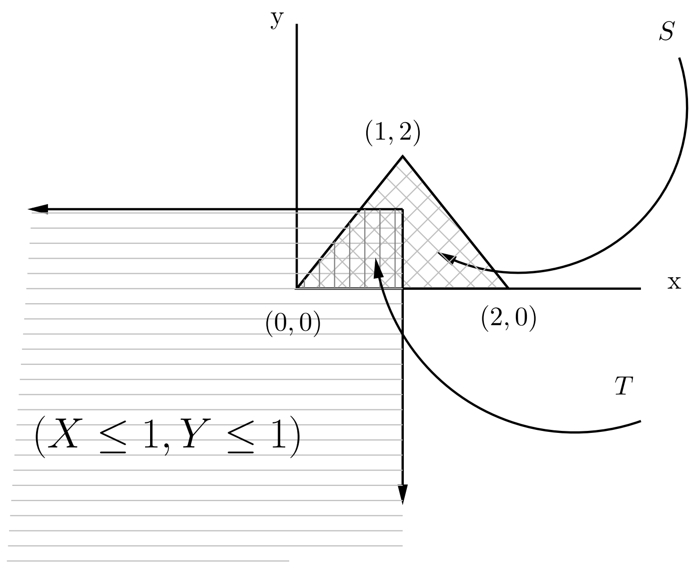
Let X and Y have a joint density f that is uniform over the interior of the triangle with vertices at $(0,0)$, $(2,0)$ and $(1,2)$. Find $P(X \le 1$ and $Y \le 1)$. So, (X, Y) has density:

$$f(x,y) = \begin{cases} \frac{1}{2} & \text{if } (x,y) \in \text{S} \\ 0 & \text{otherwise} \end{cases}$$

Where S is the support. Notice that If we want to integrate a density over space C, it should be clear that all we really need to integrate over is the space $C \cap S$, since the density is 0 outside of the support.

So, we continue in the following way:

$$P\{X \le 1, Y \le 1\} = \int\int_T \frac{1}{2} dxdy = \frac{1}{2}\int\int_T 1 dxdy = \frac{1}{2} Area(T)$$



Where $T = C \cap S$. In calculating the are of the trapezoid T, we see that the area of $T = \frac{3}{4}$. As a result:

$$P\{X \leq 1, Y \leq 1\} = \frac{1}{2} Area(T) = \frac{3}{8}$$

6.22.3 Chapter 6, Question 16

Let $X$and Y be independent random variables having respective densities $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$.

We know that $Y - X$ is a normal variable, since it is a non-constant linear combination of independent normal variables. Thus,

$$Y - X = N(\mu_2 - \mu_1, \sigma_1^2 + \sigma_2^2)$$

Since we have the following theorem:
Theorem 23. *If $X_1, ...X_k$ are independent random variables,*

$$Var(a_1X_1 + a_2X_2 + ... + a_kX_k + b) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + ... + a_k^2 Var(X_k)$$

Implying that:

$$Var(Y - X) = Var(Y) + Var(X)$$

6.23 Chapter 6, Question 9, Continued

In finding the marginal density of X

$$x^2 - xy + 4y^2 = 4y^2 - xy + x^2 = 4(y - \frac{x}{8})^2 + \frac{15x^2}{16}$$

So,

$$f_1(X) = \int_{-\infty}^{\infty} f(x,y)dy = ce^{-\frac{15x^2}{16}} \underbrace{\int_{-\infty}^{\infty} e^{-2(y-\frac{x}{8})^2} dy}_{\text{must evaluate to a constant `}a\text{'}} = ae^{-\frac{(x-0)^2}{2\cdot\frac{16}{15}}}$$

Since we know that $f_1(x)$ is a density, we recognize that we have a variable that is normal, such that:

$$f_1(x) = N(mean = 0, variance = \sigma^2 = \frac{16}{15})$$

and:

$$a = \frac{1}{\sqrt{2\pi}\frac{4}{\sqrt{15}}}$$

Chapter 7

The Expectations of Continuous Random Variables and the Central Limit Theorem

7.1 Expected Values of Continuous Random Variables

Suppose you have a random variable $X : \Omega \to (-\infty, \infty)$. In section 7.1 of the textbook, there is a good attempt at explaining what the expected value of X should be.

The idea is, for each $\epsilon > 0$, define X_ϵ as follows:
Definition. $X_\epsilon = \epsilon k$ if $\epsilon k \leq X \leq \epsilon(k+1)$

This type of random variable X_ϵ is discrete. In this case, we know that the expected value of X_ϵ:

$$E(X_\epsilon) = \sum \epsilon k \cdot P\{\epsilon k \leq X \leq \epsilon(k+1)\}$$

The expected value of X is then the limit,

$$E(X) = \lim_{\epsilon \to 0^+} E(X_\epsilon)$$

This definition can be seen on page 176.

Let us assume that X is a continuous random variable with a probability density function $f(x)$. The expected value of X,

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad \text{if} \quad \int_{-\infty}^{\infty} |x| f(x) dx < \infty$$

In fact, we can show that:

$$E\left[\varphi(X)\right] = \int_{-\infty}^{\infty} \varphi(x) f(x) dx \quad \text{if} \quad \int_{-\infty}^{\infty} |\varphi(x)| f(x)| dx < \infty$$

For two or more continuous random variables, the following is true:

$$E\left[\varphi(X, Y)\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) f(x, y) dx dy$$

Where $f(x, y)$ is the joint density of X, Y and $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

Example. Let $X = \Gamma(\alpha, \lambda)$. This means that $X \geq 0$. According to our formula,

$$E(x) = \int_{-\infty}^{\infty} x \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} = \frac{\lambda^\alpha}{\Gamma(\alpha} \int_0^\infty x^{(\alpha-1)+1} e^{-\lambda x} dx$$

Using the fact that:

$$\int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}$$

we see that:

$$\int_0^\infty x^{(\alpha-1)+1} e^{-\lambda x} dx = \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}}$$

So,

$$E(X) = \frac{\alpha}{\lambda}$$

This is the first moment of the random variable X. The n^{th} moment of $X = \Gamma(\alpha, \lambda)$ is $E(X^n)$, and using the same technique,

$$E(X^n) = \int_0^\infty x^n \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(n+\alpha)-1} e^{-\lambda x} dx$$

Again using our formula,

$$\int_0^\infty x^{(n+\alpha)-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha+n)}{\lambda^{\alpha+n}}$$

So,

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(n+\alpha)-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha+n)}{\lambda^n \Gamma(\alpha)}$$

As a result,

$$E(X^2) = \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} = \frac{(\alpha+1)\alpha}{\lambda}$$

So,

$$Var(X) = E(X^2) - (E(X))^2 = \frac{(\alpha+1)\alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$$

Thus, $Var(X = \Gamma(\alpha, \lambda)) = \frac{\alpha}{\lambda^2}$.

Example. Suppose that X is a random variable, where $\mu = E(X)$ and $E(X^n)$ is the n^{th} movement of X. Then, $E(X-\mu)^n$ is the n^{th} central moment of X, which we are sometimes interested in finding.

Fix a random variable $X = Normal(\mu, \sigma^2)$. We want to now find the central moments of X, meaning $E(X-\mu)^n$.

Suppose that n is odd. We claim that $E(X-\mu)^n = 0$. Recall that $\frac{X-\mu}{\sigma}$ is the standard normal, which implies that we can represent $X - \mu$ as σZ, where Z is the standard normal. Thus, $(X-\mu)^n = \sigma^n Z^n$. Since σ^n is a constant,

$$E(X-\mu)^n = \sigma^n E(Z^n)$$

Now suppose that $N = 2m + 1$. Then,

$$E(Z^{2m+1}) = \int_{-\infty}^{\infty} x^{2m+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

We can try to use integration by parts, calling $u = x^{2m}$, and $v' = x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Then,

$$v = \frac{-1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad v' = 2mx^{2m-1}$$

in which ase,

$$E(Z^{2m+1}) = \int_{-\infty}^{\infty} x^{2m+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \underbrace{-\frac{1}{\sqrt{2\pi}} \frac{x^{2m}}{e^{\frac{x^2}{2}}} \Bigg|_{-\infty}^{\infty}}_{0} + 2m \int_{-\infty}^{\infty} x^{2m-1} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Repeating the use of L'Hopital's rule, left hand term evaluates to 0. So,

$$E(Z^{2m+1}) = 2m \int_{-\infty}^{\infty} x^{2m-1} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2mE(Z^{2m-1})$$

Continuing in this fashion, we eventually get a string of constants acting on $E(Z)$. Recall that the expected value of the standard normal is equal to 0, and we are done. As another application from the book, look at $E(Z^{2m})$. This is the same as $E(X^2)^m$. We know that if $X = N(0, \sigma^2)$ then $X^2 = \Gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$. This tells us that $Z^2 = \Gamma(\frac{1}{2}, \frac{1}{2})$. By our formula,

$$E(Z^2)^m = \frac{\Gamma(\frac{1}{2} + n)}{\left(\frac{1}{2}\right)^n \Gamma(\frac{1}{2})}$$

So, we get

$$E(Z^{2m}) = \frac{2^m}{\sqrt{\pi}} \Gamma(m + \frac{1}{2})$$

In conclusion, if $X + Normal(\mu, \sigma^2)$,

$$E(X-\mu)^2 = \begin{cases} 0 & \text{if } n = odd \\ \frac{2^m \sigma^{2m}}{\sqrt{\pi}} \Gamma(m + \frac{1}{2}) & \text{if } n = 2m, m \in \mathbb{Z}_+ \end{cases}$$

7.2 The Central Limit Theorem

Fix a sample, $X_1, X_2, X_3, ..., X_n$ identically distributed random variables with finite second moments. Let $\mu = E(X_1) = E(X_2) = ... = E(X_n)$. Also let $\sigma^2 = Var(X_1) = Var(X_2) = ... = Var(X_n)$. We are interested in the following sum:

$$S_n = X_1 + X +_2 + ... + X_n$$

We would like to know what we can say about $P\{S_n \leq x\}$. The central limit theorem will shed some light on this question.

Theorem 24. *For each fixed $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} P\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\} = \Phi(x) = P\{Z \leq x\} = \int_{-\infty}^{x} \varphi(x)dt$$

Where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$.

If $X_i = Bernoulli(p)$, $1 \leq i \leq n$, then

$$S_n = \sum_{i=1}^{n} X_i = Binomial(n, p)$$

As n gets large, look at the following probabilities:

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P\{a \leq S_n \leq b\} = \sum_{\text{all k} \quad a \leq k \leq b} f(k)$$

How do we approximate the probability $P(S_n \leq x)$ if n is large? We will show the following:

$$P(S_n \leq x\} \approx \Phi\left(\frac{x - E(S_n)}{\sqrt{Var(S_n)}}\right)$$

In central-limit theorem considerations, we proceed as follows: if n is large,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \approx Z$$

Notice that $E(S_n) = n\mu$, and that $\sigma\sqrt{n} = \sqrt{Var(S_n)}$. Then,

$$P(S_n \leq x) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) \leq \frac{x - n\mu}{\sigma\sqrt{n}} \approx P\left(Z \leq \frac{x - n\mu}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

Example 11 on page 187 is a useful illustration.

7.3 Final

Our final will be on December 21^{st}, from $6:30 - 8:30$, in KY 431.

7.4 The Beta Function

We have the following:

$$B(\alpha, \beta) = \int_0^\infty x^{\alpha-1}(1-x)^{\beta-1}$$

and we say that the following is the **Beta Density**:

$$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

if $0 < x < 1$ and 0 otherwise is called the 'Beta(α, β) density'. We once proved that:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

7.5 Chapter 7, Question 5

Let X have a Beta density with parameters α, β. We would like to find the moments and variance of X.

We know the following:

$$E(X^n) = c\int_0^1 x^n x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{(n+\alpha)-1}(1-x)^{\beta-1}dx = B(n+\alpha,\beta)$$

using the formula for $B(\alpha, \beta)$, we have the following:

$$B(n+\alpha,\beta) = \frac{\Gamma(n+\alpha)\Gamma(\beta)}{\Gamma(n+\alpha+\beta)}$$

Thus, our answer when multiplied with c is the following:

$$cB(\alpha+n,\beta) = \frac{\Gamma(n+\alpha)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+n}$$

So, wen $n = 1$,

$$E(X) = \frac{\alpha}{\alpha+\beta}$$

Now notice that $\Gamma(\alpha+2) = (\alpha+1)\Gamma(\alpha)$. So,

$$E(X^2) = \frac{(\alpha+1)\alpha}{(\alpha+\beta)(\alpha+\beta+1)}$$

7.6 Chapter 7, Question 4

Let X have exponential density with parameter λ and let X_ϵ be defined in terms of X and $\epsilon < 0$, where $X_\epsilon = k\epsilon$ if $k\epsilon \leq X < (k+1)$, where $k \geq 0$ and is an integer. We would like to find the distribution of X_ϵ/ϵ. We can find $E(X_\epsilon)$, and evaluate its limit as $\epsilon \to 0$.

Notice that saying $X_\epsilon = k\epsilon$ is the same as saying $\frac{X_\epsilon}{\epsilon}$. So,

$$g(k) = P(\frac{X_\epsilon}{\epsilon} = k) = P\{k\epsilon \leq X < (k+1)\epsilon\} = F((k+1)\epsilon) - F(k\epsilon) = e^{-\lambda\epsilon} - e^{-\lambda(k+1)\epsilon}$$

Which follows from the fact that X is exponentially distributed. Notice that:

$$e^{-\lambda\epsilon} - e^{-\lambda(k+1)\epsilon} = e^{-\lambda k\epsilon}(1 - e^{-\lambda\epsilon})$$

So,

$$P(\frac{X_\epsilon}{\epsilon} = k) = pq^k$$

For k taking values from $0, 1, 2, ...,$ and for $q = e^{-\lambda\epsilon}$ and $p = 1 - q$. So, this tells us that $\frac{X_\epsilon}{\epsilon}$ is $Geometric(p = 1 - e^{-\lambda\epsilon})$. Recall that for $Y = Geometric(p)$, $E(Y) = \frac{q}{p}$. using this formula,

$$\frac{E(X_\epsilon)}{\epsilon} = \frac{e^{-\lambda\epsilon}}{1 - e^{-\lambda\epsilon}}$$

Thus,

$$E(X_\epsilon) = \frac{\epsilon}{e^{\lambda\epsilon}1}$$

So,

$$\lim_{\epsilon \to 0} E(X_\epsilon) = \lim_{\epsilon \to 0} \frac{\epsilon}{e^{\lambda\epsilon} - 1} = \frac{1}{\lambda} = E(x)$$

where you get $\frac{1}{\lambda}$ by applying L'Hopital's rule to $\frac{\epsilon}{e^{\lambda\epsilon}-1}$

As an exercise, let

$$A = \int_0^\infty e^{-\frac{x^2}{2}} dx$$

Fix $\alpha > 0$, and let $x = \alpha y$. This is a change of variable. From this, we get $y = \frac{x}{\alpha}$.

1. Use the change of variable to express A as a definite integral of a function y, dy.
2. Using the new definite integral,

$$A = \int_0^\infty e^{-\frac{\alpha^2 y^2}{2}} \alpha dy$$

multiply this integral by $e^{-\alpha^2/2}$ and integrate from 0 to α $d\alpha$. You get the following:

$$A^2 = \int_0^\infty \left(\int_0^\infty e^{-\frac{\alpha^2 y^2}{2}} \alpha dy \right) e^{-\frac{\alpha^2}{2}} d\alpha$$

Switch the order of integration in this integral, and obtain your answer.

7.7 Chapter 7, Question 6

Definition. A continuous random variable $X \geq 0$ is called χ^2 if $X = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. Recall that when $X = N(0, \sigma^2)$, we found that $X^2 = \Gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$. In particular, if $Z_1, Z_2, ...Z_n$ are independent distributed standard normal random variables (i.e., $\sigma = 1$), Notice that

$$Z_1^2, Z_2^2, ...Z_n^2$$

will also be independent with a distribution $\Gamma(\frac{1}{2}, \frac{1}{2})$. Taking their sum,

$$\sum_{i=1}^{n} Z_i^2 = \Gamma(\frac{n}{2}, \frac{1}{2})$$

Let X have a χ distribution with n degrees of freedom. We want to find the mean of $Y = \sqrt{X}$.

We have the following:

$$f(x) = \frac{1/2)^n/2}{\Gamma(\frac{n}{2})} x^{n/2-1} e^{-x/2} \qquad \text{if } x > 0$$

So,

$$E(Y) = E(\sqrt{X}) = \int_0^\infty \sqrt{x} f(x) dx =$$

$$\int_0^\infty x^{1/2} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} dx = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} \int_0^\infty x^{(n+1)/2-1} e^{-x/2} dx$$

$$= \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} \cdot 2^{(n+1)/2} \Gamma(\frac{n+1}{2})$$

$$= \frac{\sqrt{2}\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}$$

Notice that $\Gamma(3/2) = (1/2)\Gamma(1/2) = \frac{\sqrt{\pi}}{2}$.

7.8 The Strong Law of Large Numbers

Suppose that $X_1, X_2, ...X$ are independent identically distributed random variable whose $mean = 0$ and $var = \sigma^2$ is finite. We know that $\overline{X}$ is the following:

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}$$

and is called the **sample mean**. The following is true:

$$E(\overline{X}_n) = 0, \qquad Var(\overline{X}) = \frac{\sigma^2}{n}$$

We proved the weak law of large numbers, which says that for each fixed $\epsilon > 0$,

$$\lim_{n\to\infty} P\{|\overline{X}_n| \geq \epsilon\} = 0$$

The **Strong Law of Large Numbers** says the following:

$$\{\omega \in \Omega | \lim_{n\to\infty} \overline{X}_n(\omega) = 0\} = A$$

Implies that the probability $P(A) = P(\omega \in \Omega) = 1$.

Let $A_1, A_2, A_3, ...$ be an infinite sequence of events in a probability space Ω. Let $E = \{$ all $\omega \in \omega |$ ω belongs to infinitely many events A_i of the sequence$\}$. This set is called A_n 'infinitely often', or A_n**i.o.**.

Under what sufficient conditions can we guarantee that $P(A_n \ \ i.o) = 0$?

Suppose you took the relations $A_n = (|\overline{X}_n| \geq \epsilon)$ for some $\epsilon > 0$. Thinking of the compliment of $A_n i.o.$, would look like the following:

$$(A_n \ \ i.o)^c = \{\omega | \ |\overline{X}_n| < \epsilon, \text{ all } n \text{ that are sufficiently large}\}$$

If we could show that the probability of A_n i.o would be 0, the probability of this set would have to be 1. We have the following lemma:

Lemma 25. *If the series $\sum_{n=1}^{\infty} P(A_n)$ converges, then $P(A_n \ \ i.o) = 0$. This is called the **Borel-Cantelli lemma**.*

7.9 Chapter 7, Question 9

Let X, Y be independent random variables each having an exponential density with parameter λ, and set $Z = max(U_1, U_2)$. Find the mean and variance of Z.

Since $Z = max(X, Y)$, this tells us that $Z \geq 0$. Looking at $P(Z \leq z)$, we have:

$$P(Z \leq z) = P(X \leq z, Y \leq z) = P(X \leq z)P(Y \leq z)$$

Since $P(X \leq z) = 1 - e^{-\lambda z}$, we have:

$$P(X \leq z)P(Y \leq z) = (1 - e^{-\lambda z})^2 = 1 - 2e^{-\lambda z} + e^{-2\lambda z})$$

The density of Z is then:

$$g(z) = (1 - 2e^{-\lambda z} + e^{-2\lambda z})' = 2\lambda e^{-\lambda z} - 2\lambda e^{-2\lambda z}$$

We know the following;

$$E(Z) = \int_0^\infty zg(z)dz = 2\lambda \int_0^\infty (ze^{-\lambda z} - ze^{-2\lambda z})dz$$

This integral can be found through integration by parts of $z(e^{-\lambda z} - e^{-2\lambda z})$. Finish this question as an exercise.

7.10 Chapter 7, Question 15

Let X have the normal density $n(0, \sigma^2)$. Find the mean and variance of the following random variables:

1. $|X|$
2. $|X^2$
3. e^{tX}

Solutions:

1. See Next Chapter
2. We know that if Z is normally distributed, $Z^2 = \Gamma(\frac{1}{2}, \frac{1}{2})$. Also, notice that for $Y = \Gamma(\alpha, \lambda)$, the mean is $\frac{\alpha}{\lambda}$, and $var = \frac{\alpha}{\lambda^2}$

   Call $\frac{X}{\sigma} = Z$. In other words, $X^2 = \sigma^2 Z^2$. Based on what we know of the distribution of X^2 and our formula, $E(Z^2) = 1$ and $Var(Z^2) = 2$. Notice that from the variance, we can find that $E(Z^4) = 3$. So,

   $$E(X^2) = \sigma^2 E(Z^2) = \sigma^2, \qquad Var(X^2) = \sigma^4 Var(Z^2) = 2\sigma^4$$

3. See Next Chapter

Chapter 8

Moment Generating Functions

Definition. For any random variable X, the moment generating function (mfg) is:

$$M_X(t) = E(e^{tX}) \qquad \text{(Provided this expectation exists)}$$

where $t \in \mathbb{R}$. Notice that $M_X(0) = 1$. If $M_X(t)$ exists for some $t \neq 0$, then we can show that the moment generating function $M_X(t)$ exists for all $t \in (-\epsilon, \epsilon)$.

Recall from calculus that:

$$e^u = \sum_{n=0}^{\infty} \frac{u^n}{n!}$$

Replacing u by tX, we have:

$$e^{tX} = \sum_{n=0}^{\infty} \frac{(tX)^n}{n!} = \sum_{n=0}^{\infty} \frac{t^n X^n}{n!}$$

Now taking the expected value, we have the following:

$$E(e^{tX} = E\left(\sum_{n=0}^{\infty} \frac{t^n X^n}{n!}\right) = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!} t^n$$

where the right-most term is a power series. Recall that:

$$f(t) = \sum_{n=0}^{\infty} \frac{f^{(n)}}{n!} t^n$$

So,

$$M_X(t) = \sum_{n=0}^{\infty} \frac{M_X^{(n)}(0)}{n!} t^n = E(X^n) = M_X^{(n)}(0)$$

For all $n \geq 1$.

For Z, a normal variable with the standard distribution, let us find $M_Z(t)$. We have the following:

$$M_Z(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tX} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

For some fixed $t \neq 0$. Notice that:

$$-\frac{x^2}{2} + tx = \frac{-(x^2 - 2tx + t^2) + t^2}{2} = -\frac{(x-t)^2}{2} + \frac{t^2}{2}$$

So,

$$E(e^{tX}) = \int_{-\infty}^{\infty} e^{tX} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx$$

Notice that $\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}}$ is normally distributed with mean=t and $var = 1$, the integral of which is 1. Thus, our answer is simply, $e^{\frac{t^2}{2}}$- i.e., for a normally distributed random variable with the standard distribution, its moment generating function is $e^{\frac{t^2}{2}}$. Using our expansion, we have:

$$e^{\frac{t^2}{2}} = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!}$$

Regarding the moments of the standard random variable, we have the following remarks:

1. The expected value of $Z^{odd} = 0$. This follows from setting our expansion above equal to our definition of $M_X(t)$, and thinking realizing that the coefficients of the power series need to agree. In other words, we have the following:

$$\frac{E(Z^{2n})}{(2n)!} = \frac{t^{2n}}{2^n n!} \Rightarrow E(Z^{2n}) = \frac{(2n)!}{2^n n!}$$

2. Suppose that a, b are two constants. How do we relate the moment generating function, $M_{aX+b}(t)$ and $M_X(t)$? By definition,

$$M_{aX+b} = E(E^{t(aX+b)}) = e^{tb} E(e^{taX}) = e^{tb} M_X(at)$$

So, $M_{aX+b} = e^{tb} M_X(at)$.

If $X = N(\mu, \sigma^2)$, then

$$\frac{X - \mu}{\sigma^2} = Z \Rightarrow X = \sigma Z + \mu$$

So, $M_X(t) = M_{\sigma Z + \mu} = e^{\mu t} M_Z(\sigma t) = e^{\mu t} \cdot e^{\frac{\sigma^2 t^2}{2}}$. So, $X = N(\mu, \sigma) \Rightarrow M_X(T) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$.

Notice that Given e^{tX}, looking for $E(X))$ we have:

$$E(e^{tX}) = M_X(t) = e^{\frac{\sigma^2 t^2}{2}}$$

And that:

$$Var(e^{tX}) = E(e^{2tX}) - [(E(e^{tX})]^2 = M_X(2t) - e^{\sigma^2 t^2} = e^{2\sigma^2 t^2} - e^{\sigma^2 t^2}$$

Which follows from letting $t = 2t$ in our formula for $M_X(t)$.

8.1 Chapter 7, Question 14

Let X be the sine of an angle in radians chosen uniformly from $(-\pi/2, \pi/2)$. Find the mean and variance of X.

Let $\hat{\theta} =$ Uniform$[\,-\pi/2, \pi/2]$. The random variable $X = sin(\hat{\theta})$. We have the following:

$$f(\theta) = \begin{pmatrix} \frac{1}{\pi} & \frac{-\pi}{2} \leq \theta \leq \frac{\pi}{2} \\ 0 & \text{otherwise} \end{pmatrix}$$

We know that $E(X) = E(sin(\hat{\theta})$. Notice that:

$$E(X) = E(\hat{\theta}) = \int_{-\infty}^{\infty} sin(\theta) f(\theta) d\theta = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} sin(\theta) d\theta = \frac{1}{\pi} \cdot \frac{-cos(\theta)}{\pi} \Bigg|_{\theta=-\pi/2}^{\theta=\pi/2} = 0$$

Simultaneously, we want:

$$E(X^2) = E(sin^2(\hat{\theta})) = \int_{-\infty}^{\infty} sin^2(\theta) f(\theta) d\theta = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} sin^2(\theta) d\theta = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1 - cos(2\theta)}{2} d\theta =$$

$$= \frac{1}{\pi} \cdot \frac{\theta - sin(2\theta)/2}{2} \Bigg|_{\theta=-\pi/2}^{\theta=\pi/2} = \frac{\pi}{4} + \frac{\pi}{4}$$

Since $sin^2(\theta) = \frac{1-cos(\theta)}{2}$.

8.2 Chapter 7, Question 31

Let $X_1, X_2, ...$ be independent, identically distributed random variables having mean 0 and finite nonzero variance σ^2. Set $S_n = X_1 + ... X_n$. We want to show that if X_1 has finite third moment, then

$$E(S_n^3) = nE(X_1^3)$$

and

$$\lim_{n\to\infty} E\left(\frac{S_n}{\sigma\sqrt{n}}\right)^3 = 0$$

We here assume that $E(X_1^3)$ is finite. What we really have to look at is $\frac{E(S_n^3)}{\sigma^3 n\sqrt{n}}$. Notice that S_n^3 takes the following form:

$$S_n^3 = (X_1 + X_2 + ... + X_n)^3 = \sum_{i=1}^{n} X_i^3 + \sum_{i\neq j} X_i^2 X_j + \sum_{i\neq j\neq k} X_i X_j X_k$$

Taking the expected value,

$$E(S_n^3) = \sum_{i=1}^{n} E(X_i^3) + \sum_{i\neq j} E(X_i^2 X_j) + \sum_{i\neq j\neq k} E(X_i X_j X_k)$$

$$= \sum_{i=1}^{n} E(X_i^3) + \sum_{i\neq j} E(X_i^2)E(X_j) + \sum_{i\neq j\neq k} E(X_i)E(X_j)E(X_k)$$

Notice that $\sum_{i=1}^{n} E(X_i^3) = nE(X_i^3)$. So,

$$\frac{E(S_n^3)}{\sigma^3 n\sqrt{n}} = \frac{nE(X_1^3}{\sigma^3 n\sqrt{n}} = \frac{E(X_1^3)}{\sigma^3\sqrt{n}} \to 0 \qquad \text{as } n \to \infty$$

From the central limit theorem, we have:

$$\frac{(X_1 + ... + X_n) - n\mu}{\sigma\sqrt{n}} \to^{\mathcal{D}} Z$$

which tells us that:

$$\lim_{n\to\infty} P\left(\frac{(X_1 + ... + X_n) - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \qquad \forall x$$

In other words, we have:

$$\frac{S_n}{\sigma\sqrt{n}} \quad \text{converges to } Z$$

So,

$$E\left(\frac{S_n}{\sigma\sqrt{n}}\right)^3 \to E(Z^3)$$

8.3 Chapter 7, Question 35

Let $X_1, X_2, ...X_{100}$ be independent normally distributed random variables having mean 0 and variance 1. To indicate that they are standard normally distributed variables, we denote them as $Z_1, Z_2, ...Z_{100}$. Notice that $Z_i = \Gamma(\frac{1}{2}, \frac{1}{2})$. Since Z_i^2 is independent from Z_j^2 when $i \neq j$, we have:

$$U = Z_1^2 + ... + Z_{100}^2 = \Gamma(\frac{n}{2}, \frac{1}{2}) = \chi_n^2$$

We then have:

$$E(U) = n \qquad Var(U) = 2n$$

By the central limit theorem,

$$\frac{U - n}{\sqrt{2n}} \approx Z_{\text{standard normal}}$$

Approaching our problem:

1. $P(U \leq 120) = P(\frac{U-100}{10\sqrt{2}} \leq \frac{120-100}{10\sqrt{2}} \approx P(Z \leq 1.41) = \Phi(1.41) = .9207$
2. $P(80 \leq U \leq 120) = P(\frac{80-100}{10\sqrt{2}} \leq \frac{U-100}{10\sqrt{2}} \leq \frac{120-100}{10\sqrt{2}} \approx P(01.41 \leq Z \leq 1.41) = P(|Z| \leq 1.41) = 2\Phi(1.41) - 1 = .8414$

3. We aim to find c such that: $P(100-c \leq I \leq 100+c) = .95$ We change this question to:

$$P(\frac{-c}{10\sqrt{2}} \leq \frac{U-100}{10\sqrt{2}} \leq \frac{c}{10\sqrt{2}}) = .95$$

$$\Rightarrow 2\Phi(\frac{c}{10\sqrt{2}}) - 1 \approx .95 \Rightarrow \Phi(\frac{c}{10\sqrt{2}}) = .975 = \Phi(1.96) \Rightarrow c = 1.96 \cdot 10\sqrt{2} = 27.718$$

8.4 Chapter 7, Question 32

We have $S_n = X_1 + X_2 + ... + X_n$, as in question 31. We know that $E(X_1) = 0$, and that:

$$E(X_1^4) < \infty$$

We have the following:

$$S_n^4 = (X_1 + ... + X_n)^4 = \sum_i X_i^4 + \sum_{i \neq j} X_i^2 X_j^2 + \sum_{i \neq k \neq j} X_i^2 X_j X_k + \sum_{i \neq j \neq k \neq l} X_i X_j X_k X_l + \sum_{i \neq j} X_i^3 X_j$$

Looking at the terms of type $X_i^2 X_j^2$, we would like to find out how many variables we have of that type. We proceed using the counting principal:

1. Fix locations of S_n^4:

$$S_n^4 = \underbrace{(X_1 + ... + X_n)}_{1} \underbrace{(X_1 + ... + X_n)}_{2} \underbrace{(X_1 + ... + X_n)}_{3} \underbrace{(X_1 + ... + X_n)}_{4}$$

   For example, locations 1 and 3.

2. From $1, 2, ...n$ we pick a pair (i, j) such that $i \neq j$. For example, 5 and 12 (suppose that $n \geq 12$). Then, we're thinking of X_5 from location location 1, and X_{12} from location 2, which when multiplied, gives us a term: $X_5^2 \cdot X_{12}^2$. So, if we can find the number of choices we have in step 1, and the number of choices for step two, we can calculate the number of total possibilities we have.

In step 1, we have ($4^C 2$ possibilities (since there are four locations). In step 2, we have $n^C 2$ choices. So, by the multiplication rule in S_n^4, we have $4^C 2 \times n^C 2$ terms of the type $X_i^2 X_j^2$ where $i \neq j$. Reducing our equation, we have:

$$6 \cdot \frac{n(n-1)}{2}$$

So,

$$E(S_n^4) = nE(X_1^4) + 3n(n+1)\sigma^4$$

Since $E(X_i^2)E(X_j^2) = \sigma^4$.

8.5 Chapter 7, Question 34

Let $X_1, X_2, ...X_n$ be independent normally distributed random variables having mean 0 and variance σ^2. Recall that:

$$X_i^2 = \Gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$$

from which we know: $E(X_i^2) = \sigma^2$, and $Var(X_i^2) = 2\sigma^4$.

8.6 Chapter 7, Question 37

Twenty numbers are rounded off to the nearest integer and then added. Assume that the individual round-off errors are independent and uniformly distributed over $U = (-\frac{1}{2}, \frac{1}{2})$. Find the probability that the given sum will differ from the of the original twenty numbers by more than 3.

In other words, $|\text{sum} - \text{rounded sum}| > 3$. Looking at *sum- rounded sum*, we have the sum of the errors, which is equal to $\epsilon_1 + \epsilon_2 + ... + \epsilon_{20}$. We know that all ϵ_i are distributed uniformly over our interval U. Notice that the mean of these variables is $\mu = 0$, since the average of the values of U is 0. The variance, $\sigma^2 = \frac{1}{12}$. Using the central limit theorem (since $n = 20$) we can say that:

$$\frac{\epsilon_1 + \epsilon_2 + ... + \epsilon_{20}}{\frac{1}{\sqrt{12}}\sqrt{20}} \approx Z$$

Where $1.291 = \frac{1}{\sqrt{12}}\sqrt{20}$. So,

$$P\{|\epsilon_1 + ... + \epsilon_{20}\} \approx P\{|Z| > \frac{3}{1.29} = 2.32\}$$

Recall that:

$$P\{|Z| \leq a\} = 2\Phi(a) = 1$$

So,

$$P\{|Z| > \frac{3}{1.29} = 2.32\} = 2 - 2\Phi(2.32) = .0204$$

8.7 Chapter 7, Question 38

A fair coin is tossed until 100 heads appear. Find the probability that at least 226 tosses will be necessary.

Let:

$$X_i = \begin{cases} 1 & \text{if Heads on the } i^{th} \text{ toss} \\ 1 & \text{otherwise} \end{cases}$$

So, if we make n tosses, the total number of heads is equal to $X_1 + X_2 + ... + X_n = S_n$. Let $N \geq 1$ be a random variable such that $S_N = 100$, and that $S_{N-1} < 100$. What we now look for is: $P\{N \geq 226\}$. Forcing $N \geq 226$, we see that the sum:

$$X_1 + X_2 + .. + X_{225} < 100$$

Observe that this statement is equivalent to the statement that $P\{N \geq 226\}$. So, we want:

$$P\{X_1 + X_2 + ... + X_{225} < 100\}$$

Notice that $X_1 + X_2 + ... + X_{225}$ is binomial, where $n = 225, p = \frac{1}{2}$. Thus, the mean is $\mu = np = 225 \cdot \frac{1}{2}$ and the variance is $\sigma^2 = np^2 = 225 \cdot \frac{1}{4}$. So, by the central limit theorem,

$$\frac{X_1 + X_2 + ... + X_{225} - 112.5}{7.5} \approx Z_{s.n.}$$

We then have:

$$P(X_1 + X_2 + ... + X_{225} < 100) \approx P(Z < -.167) = .0475$$

which comes from the use of a table.

8.8 Chapter 7, Question 39

Using the same set up as question 38, we want to find the probability that $N = 226$.

Here, we can say that $(N = 226) \cup (n \geq 227) = (N \geq 226)$. So,

$$P(N = 226) = P(N \geq 226) - P(N \geq 227$$

which can be calculated in the same manner as question 38.

Alternatively, notice that from the independence of our result, we have:

$$\{N = 226\}\{X_1 + X_2 + ... + X_{225}\} = 99 \quad \text{and } X_{226} = 1 = P(X_1 + X_2 ... + X_{225} = 99)P(X_{226} = \frac{1}{2})$$

Which would be: $\binom{225}{99} \cdot (\frac{1}{2}^{226})$. If you would like, look at Stirling's formula, which has applications to our solution.

8.9 Chapter 8, Question 3

Let X have a Poisson distribution with parameter λ. We are asked to find the mean and variance of X by using the moment generating function.

We have the following:

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} \cdot e^{\lambda e^t}$$

Now, we take the derivative, and say:

$$M'_X(t) = e^{-\lambda} \cdot e^{\lambda e^t} \lambda e^t$$

Now, looking at the following:

$$E(X) = M'_X(0) = e^{-\lambda} \cdot e^{\lambda} \lambda = \lambda$$

Notice that $M'_X(t) = \lambda e^{-\lambda} e^{t+\lambda e^t}$. Now taking the second derivative,

$$M''_X(t) = \lambda e^{-\lambda} e^{t+\lambda e^t}(1 + \lambda e^t)$$

Again using properties of the moment generating function,

$$E(X^2) = M''_X(0) = \lambda(1 + \lambda)$$

Calculating the variance,

$$Var(X) = E(X^2) = E(X)^2 = \lambda(1 + \lambda) = \lambda^2 = \lambda$$

8.10 A Relationship Between $M_X(t)$ and $\Phi_X(t)$

Remember that for $X \in \mathbb{IV}_+$, for which we have the following functions:

$$\Phi_X(t) = \sum_{x=0}^{\infty} f(x) t^x = E(t^X) \qquad M_X(t) = E(e^{tX})$$

Can we find a relationship between these two generating functions?

Formally, for any positive number $a > 0$, can be represented as $a = e^{ln(a)}$. Doing this for t^X, we have:

$$t^X = e^{ln(t)^X} = e^{X ln(t)}$$

This tells us that:

$$\Phi_X(t) = E(e^{ln(t) \cdot X}) = M_X(ln(t))$$

Calling $ln(t) = s$, then $e^s = t$ and we arrive at the following:

$$M_X(s) = \Phi_X(e^s)$$

8.11 Chapter 8, Question 5

Let X be a continuous random variable having the density:

$$f_X(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < \infty$$

This is called a 'double exponential' distribution. We want to show that:

$$M_X(t) = \frac{1}{(1-t^2)}, \qquad -1 < t < 1$$

Fix some $t \neq 0$. Then,

$$E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}\frac{1}{2}e^{-|x|}dx = \int_{-\infty}^{0} e^{tx}\frac{1}{2}e^{x}dx + \int_{0}^{\infty} e^{tx}\frac{1}{2}e^{-x}dx = \frac{1}{2}\left(\int_{-\infty}^{0} e^{(t+1)x}dx + \int_{0}^{\infty} e^{(t-1)x}dx\right)$$

Clearly, these integrals would not be finite if $t = -1, 1$. Notice that if $t \neq \pm 1$, then the anti derivatives of the following are true:

$$\int e^{(t+1)x} = \frac{e^{(t+1)x}}{t+1} \qquad \int e^{(t-1)x} = \frac{e^{(t-1)x}}{t-1}$$

Notice that for $a \neq 0$,

$$\lim_{x\to\infty} e^{ax} = \text{ finite if and only if } a > 0 \text{ (in which case the limit} = 0)$$

Similarly,

$$\lim_{x\to\infty} e^{ax} \text{ finite if and only if } a < 0 \text{ (in which case the limit} = 0)$$

So, we have the following evaluations:

$$= \frac{1}{2}\frac{e^{(t+1)x}}{(t+1)}\Bigg|_{x=-\infty}^{\infty} + \frac{1}{2}\frac{e^{(t-1)x}}{(t-1)}\Bigg|_{0}^{\infty} = \frac{1}{2(t+1)} - \frac{1}{2(t-1)} = \frac{1}{(1-t^2)}$$

which followed from assuming that $-1 < t < 1$. We know that we have the following power series expansion of the moment generating function:

$$M_X(t) = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!}t^n$$

In our case, for the double exponential,

$$M_X(t) = \frac{1}{(1-t^2)}$$

Recall that if $0 \geq q < 1$ then:

$$\frac{1}{1-q} = 1 + q + q^2 + q^3 + \ldots$$

Taking $t^2 = q$, and assuming that $0 < t^2 = q < 1$:

$$M_X(t) = \frac{1}{(1-t^2)} = 1 + t^2 + t^4 + t^6 + ...$$

Noticing that:

$$M_X(t) = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!} t^n = 1 + \frac{E(X)}{1!}t + \frac{E(X^2)}{2!}t^2 + \frac{E(X^3)}{3!}t^3 + ..$$

Matching up the coefficients, clearly $E(X^{2k+1}) = 0$ for all $k \geq 0$. We end up with the following:

$$\frac{E(X^2)}{2!} = 1 \quad \frac{E(X^4)}{4!} = 1 \quad \frac{E(X^6)}{6!} = 1...$$

Which tells us that $E(X^{2n}) = (2n)!$, e.g., $E(X^{100}) = 100!$.

8.12 Properties of Moment Generating Functions

We have the following Important properties for moment generating functions:

1. If X, Y are two random variables such that for all $t \in (-\epsilon, \epsilon)$ $M_X(t) = M_Y(t)$, then X and Y have the same distribution.
   Example. For a Poisson random variable X with parameter λ, the moment generating function is $M_X(t) = e^{\lambda(e^t-1)}$. If for some random variable Y, $M_Y(t) = e^{3(e^t-1)}$ then we can say that since $e^{3(e^t-1)} = M_X(t)$ when $X = Poisson(3)$, we have that $Y = Poisson(\lambda = 3)$.

2. If random variables $X_1, ..., X_n$ are independent, then

   $$M_{X_1+X_2+...+X_N}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot ... \cdot M_{X_n}(t)$$

   This follows from seeing that:

   $$E\left[e^{t(x_1+x_2+...+x_n)}\right] = E(e^{tx_1} \cdot e^{tx_2} \cdot ... \cdot e^{tx_n}) = E(e^{tx_1}) \cdot E(e^{xt_2}) \cdot ... \cdot E(e^{tx_n})$$

   where the last step follows from the independence of X_i, X_j when $i \neq j$.
   Example. Let $X = \Gamma(\alpha, \lambda)$. $M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$ for all $t < \lambda$. Notice that if you take $\alpha = 1$, then $\Gamma(1, \lambda) = Exp(\lambda)$, so $= Exp(\lambda)$ and $M_X(t) = \frac{\lambda}{\lambda - t}$. Read the proof of this in the textbook.

Suppose one is asked to prove the following result: If $X_1 = \Gamma(\alpha_1, \lambda), X_2 = \Gamma(\alpha_2, \lambda), ..., X_n = \Gamma(\alpha_n, \lambda)$ are independent, then $X_1 + X_2 + ... + X_n$ is $\Gamma(\alpha_1 + ... + \alpha_n, \lambda)$. We have the following method to prove this statement using our two properties:

We know that $M_{X_i}(t) = \left(\frac{\lambda}{\lambda = t}\right)_i^\alpha$ for all $t < \lambda$. We also know that:

$$M_{X_1+X_2+...+X_n} = M_{X_1}(t) \cdot M_{X_2}(t) \cdot ... \cdot M_{X_n}(t)$$

Since the $X's$ are independent. Substituting,

$$\cdot M_{X_2}(t) \cdot ... \cdot M_{X_n}(t) = \left(\frac{\lambda}{\lambda = t}\right)_1^{\alpha} \cdot \left(\frac{\lambda}{\lambda = t}\right)_2^{\alpha} \cdot ... \cdot \left(\frac{\lambda}{\lambda = t}\right)_n^{\alpha}$$

Doing some algebra, we have:

$$\left(\frac{\lambda}{\lambda = t}\right)_1^{\alpha} \cdot \left(\frac{\lambda}{\lambda = t}\right)_2^{\alpha} \cdot ... \cdot \left(\frac{\lambda}{\lambda = t}\right)_n^{\alpha} = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1 + \alpha_2 + ... + \alpha_n = M_Y(t)}$$

where $Y = \Gamma(\alpha_1 + \alpha_2 + ... + \alpha_n, \lambda)$. By property 1, the sum $X_1 + X_2 + ... + X_n$ has the same distribution as Y, and we are done.

8.13 Chapter 8, Question 8

Let X be a random variable such that $M_X(t)$ is finite for all t. We would like to use the same argument as in the proof of Chebyshev's Inequality to show that:

$$P(X \geq x) \leq e^{-tx} M_X(t) = \frac{M_X(t)}{e^{tx}}, \qquad t \geq 0$$

First, let us note that if $t = 0$, the moment generating function of t is 1, and we know that the probability of anything is ≤ 1. So, we assume that $t > 0$. Notice that the following statements are equivalent:

$$X \geq x \iff tX \geq tx \iff e^{tX} \geq e^{tX}$$

So, for $t > 0$,

$$P(X \geq x) = P(e^{tX} \geq e^{tx})$$

By Markov's inequality,

$$P(X \geq x) = P(e^{tX} \geq e^{tx}) \leq \frac{E(e^{tX})}{e^{tx}} = \frac{M_X(t)}{e^{tx}}$$

And we have finished our exercise.

Now if we suppose that x is fixed and call:

$$\frac{M_T(x)}{e^{tx}} = g(t)$$

It follows that:

$$P(X \geq x) \leq min(g(t)) \quad \text{for all } t \geq 0$$

This is discussed in problem 9, which we will discuss next class.

8.14 Chapter 8, Question 6

Let X have a binomial distribution with parameters n and p. We want to find $dM_X(t)/dt$ and $d^2M_X(t)/dt^2$.

Since $X = Binnomial(n,p)$, we know that:

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx}(n^C x)p^x q^{n-x} = \sum_{x=0}^{n}(n^C x)(pe^t)^x q^{n-x}$$

Since

$$E(\varphi(X)) = \sum_x \varphi(x)f(x)$$

and in our case, $\varphi(x) = e^{tx}$. Recalling that

$$\sum_{x=0}^{\infty}(n^C x)a^x b^{n-x} = (a+b)^n,$$

we have:

$$\sum_{x=0}^{n}(n^C x)(pe^t)^x q^{n-x} = (pe^t + q)^n$$

Once we arrive here, we can take some derivatives and solve for the mean and variance.

It also may be helpful to notice that the probability generative function for a Binomial random variable is as follows:

$$\varphi_X(t) = (pt+q)^n$$

Using our connection between the moment generating and probability generating functions, we could have arrived at the same answers (which can be found as formula (3) on page 198).

8.15 Chapter 8, Question 7

Let $X_1, X_2, ...X_n$ be independent, identically distributed random variables such that $M_X(t)$ is finite for all t. We want to show using moment generating functions that:

$$E(X_1 + X_2 + ...X_n)^3 =$$

Calling $M_{X_1}(t) = m(t)$. Let

$$M(t) = M_{X_1+X_2+...+X_n}(t) = [m(t)]^n$$

We know that:

$$E(X_1 + X_2 + ...X_n)^3 = M'''(0)$$

So, taking derivatives, we have:

$$\begin{gathered} M'(t) = n(m(t))^{n-1}m'(t) \\ M''(t) = n(n-1)(m(t))^{n-2}(m'(t))^2 + n(m(t))^{n-1} \cdot m''(t) \\ M'''(t) = n(n-1)(n-1)(m(t))^{n-3}[m'(t)]^3 + n(n-1)(m(t))^{n-2}2m'(t) \cdot m''(t) \\ +n(n-1)[m(t)]^{n-2}m'(t)m''(t) + +n[m(t)]^{n-1} \cdot m'''(t) \end{gathered}$$

So,

$$E(X_1+X_2+...X_n)^3 = n(n-1)(n-1)[1]^0E(X_1^3)+2n(n-1)E(X_1)E(X_1^2)+n(n-1)E(X_1)E(X_1^2)+nE(X_1^3)$$

Since notice that $m'(0) = E(X_1)$, $m''(0) = E(X_1^2)$ and $m''' = E(X_1^3)$. This gives us the result we were looking for.

8.16 Chapter 7, Question 36

A runner attempts to pace off 100 meters for an informal race. His paces are independently distributed with $\mu = .97$ and standard deviation $\sigma = .1$ meter. Find the probability that his 100 paces will differ from 100 meters by no more than 5 meters.

Let $X = pace$. We have $X_1, X_2, ...X_{100}$. These are naturally identically distributed independent random variables. We are looking for the probability of the following event:

$$P\{-5 \le X_1 + X_2 + ... + X_{100} - 100 \le 5\}$$

By the central limit theorem for the sum,

$$\frac{X_1 + X_2 + ... + X_100 - n\mu}{\sigma \cdot \sqrt{n}} = \frac{X_1 + X_2 + ... + X_100 - 97}{1} \approx Z_{s.n}$$

So, we can change our event to the following:

$$\begin{aligned} &P\{-5 \le X_1 + X_2 + ... + X_{100} - 100 \le 5\} \\ &\qquad = P\{-2 \le X_1 + X_2 + ... + X_{100} - 97 \le 8\} \approx P\{-2 \le Z \le 8\} \\ &\qquad\qquad = \Phi(8) - \Phi(-2) \approx .0288 \end{aligned}$$

8.17 Announcement for the Final

We should be comfortable with setups like the following:

$$\underbrace{(X,Y)}_{f(x,y)} \longmapsto \underbrace{(U,V)}_{g(u,v)}$$

Now find a formula for the density of $X \cdot Y$, or $X + Y$, or $\frac{X}{Y}$, etc.

Example. Suppose $X = Exp(\lambda)$. We want to find the density of $Y = \frac{1}{1+X}$. Notice that $0 < Y \leq 1$. We want to find something along the lines of:

$$g(y) = \begin{cases} \text{... if } 0 < y \leq 1 \\ 0 \text{ otherwise} \end{cases}$$

8.18 Review for the Final

8.18.1 Chapter 8, Question 9

Let X have a gamma distribution with parameters α, λ. Use the following result:

$$A = P\{X \geq a\} \leq e^{-ta} M_X(t)$$

to show that $P(X \geq 2\alpha/\lambda) \leq (2/e)^\alpha$.

If $X = \Gamma(\alpha, \lambda)$, and $X \geq 0$, we know that $M_X(t) = (\frac{\lambda}{\lambda - t})^\alpha$ for all $t < \lambda$. In our case, $0 \leq t < \lambda$:

$$g(t) = e^{-\alpha t} \left(\frac{\lambda}{\lambda - t})^\alpha \right)$$

So, we see that when we set $g'(t) = 0$, we get that $\frac{\alpha}{\lambda - t} = a$, which is the same as saying that $t = \lambda - \frac{\alpha}{a} = \lambda$. For our choice of a, which is $2\alpha/\lambda$, we see that $t_0 = \frac{\lambda}{2} \in (0, \lambda)$. Using a chart ,we can see that this value truly is the minimum value of $g(t)$ on our interval. So, $A \leq g(\frac{\lambda}{2})$. Notice that Since $a\frac{\lambda}{2} = \alpha$.

8.18.2 Exam Question Number 6

A point is chosen at random uniformly in the space$x = (0, 1)$ and $y = (0, 3)$. We let $X(u, v) = u + v$. When $0 < x < 2$, we want to find the density.

First, we can find the distribution. Fix some $0 < x < 2$. Let $F(X)$ be the probability that $F(x) = P\{X \leq x\}$. This event, $\{X \leq \alpha\} = \{(u, v) \mid X(u, v) \leq x\} = \{(u, v) \mid u + v \leq x\}$ Notice that:

$$F(x) = P(\delta_x) = \frac{|\delta_x|}{|\Omega|} = \frac{x^2}{12} \Rightarrow f(X) = F'(x) = \frac{x}{6}$$

8.18.3 General Question

Suppose that X and Y are independent and $X = Geom(p)$ and $Y = U(0, 1, 2, 3, 4, 5)$. We would like to find $P(Y \leq X)$. We can express this event as follows:

$$\{Y \leq X\} = \bigcup_{x=0}^{\infty} \{Y \leq X, X = x\} = \bigcup_{x=0}^{\infty} \{Y \leq x, X = x\} = A_x$$

This union is the following sum:

$$P(Y \leq X) = \sum_{x=0}^{\infty}(A_x) = P(A_0) + P(A_1) + P(A_2) + P(A_3) + P(A_4) + P(A_5) + P(A_6) + ...$$

Notice that starting with $P(A_5)$, that $P(A_x) = 1$, since $Y \leq 5$ by definition. But, we also have the following:

$$P(A_0) = P(Y \leq 0, X = 0) = P(Y \leq 0)P(X = 0) = \frac{1}{6} \cdot p$$
$$P(A_1) = P(Y \leq 1, X = 1) = P(Y \leq 1)P(X = 1) = \frac{2}{6} \cdot pq$$
$$P(A_2) = P(Y \leq 2, X = 2) = P(Y \leq 2)P(X = 2) = \frac{3}{6} \cdot pq^2$$

And so on. Finishing in this fashion, we can find our sum, and we can be done. The tail of this sum is simply:

$$pq^5 + pq^6 + pq^7 + ... = pq^5(1 + q + q^2 + ...) = pq^5 \frac{1}{1-q} = q^5$$

8.18.4 Chapter 6, Question 11

Let X and Y be independent continuous random variables having the indicated marginal densities. Find the density of $Z = X + Y$ where X is uniform on (0,1) and Y is $Exp(\lambda)$. Before we begin, notice that $X \geq 0$ and $Y \geq 0$, so $Z = X + Y \geq 0$. Thus, we could use the convolution:

$$f_Z(z) = \int_0^z f_X(x) f_Y(z-x)dx$$

Notice that

$$f_Y(z-x) = \lambda e^{-\lambda(z-x)}$$

We have to distinguish cases based on where z is. We have the following cases:

1. Suppose that $0 < z < 1$. Now,

$$f_Z(z) = \int_0^z 1 \cdot \lambda e^{-\lambda(z)} e^{\lambda x} dx = e^{-\lambda z} \int_)^z \lambda e^{\lambda x} dx =$$

   Since $\int_0^z e^{\lambda x} = \frac{e^{\lambda z}-1}{\lambda}$, we have:

$$f_Z(Z) = \frac{e^{\lambda z} - 1}{\lambda} \cdot \lambda e^{-\lambda z} = 1 - e^{\lambda z}$$

2. Suppose that $z > 1$. We now split our function to the following:

$$f_Z(z) = \int_0^1 1 \cdot \lambda e^{-\lambda z} e^{\lambda x} dx + \int_1^z 0 \; dx = \int_0^1 \lambda e^{-\lambda z} e^{\lambda x} dx$$

   And since we have the answer to case one, we take that answer when $z = 1$, and we have:

$$f_Z(z) = \lambda e^{-\lambda z} \cdot \frac{(e^{\lambda} - 1)}{\lambda} = (e^{\lambda} - 1)e^{-\lambda z}$$

So,

$$f_Z(z) = \begin{cases} 1 - e^{-\lambda z} \quad \text{if } z < 1 \\ (e^{\lambda} - 1)e^{-\lambda z} \quad \text{if } z > 1 \\ 0 \quad \text{if } z < 0 \end{cases}$$

8.18.5 Chapter 7, Question 15

Let $X = Normal(0, \sigma^2)$. This tells us that $E(X) = 0$ and $E(X^2) = \sigma^2$. We would like to find $Var(|X|)$. Recall that we have $E(|X|) = \sqrt{\frac{2}{\pi}}$. Notice now that:

$$Var(|X|) = E(|X|^2) - E(|X|)^2 = E(X^2) = (\sigma\sqrt{\frac{2}{\pi}})^2 = \sigma^2 - \frac{2}{\pi}\sigma^2 = (1 - \frac{2}{\pi})\sigma^2$$

8.18.6 Information About the Exam

Question one is similar to what we just did with $X \leq Y$. The second is to find the density of the minimum of discrete random variables. By the way, $min(U, V) \geq a)$ is $U \geq a, V \geq q$. When you deal with the minimum, you deal with greater than or equal. The third is finding the density of some function of x. One was given in our second exam. X is exponential, what is $Y = X/X + 1$, for example. We apply a theorem to do these questions. The next is a moment-generating function question, you have a MGF of some function and you are asked to find the mean, variance, and other moments. Then there is a problem about finding the mean of the Gamma distributed variable. There is then a question about what we just did with the convolution, and in general if you want to find the density of $5X - Y$, there is a way to do it by taking $(X, Y) \to (U, V)$. Using the Jacobian and integrating you can get what you are looking for. If the joint density $f(x, y) > 0$ if $(x, y) \in S$ and 0 otherwise, when you write $g(u, v) = \frac{1}{|J|}f(u, v)$ you get something, and you will need to explain where the pairs are in S, i.e., where they aren't zero.

E.G., $g(u, v) = \frac{1}{3}f(3u - v, u + v)$. Suppose that you know that $f(x, y) \neq 0$ when $x \in (0, 1), y \in (0, 2)$. So, we would want:

$$0 < 3u - v < 1$$
$$0 < u + v < 2$$

This needs to be taken into account. Suppose that $X + Exp(\lambda)$, and $Y = U(0, 1)$. We can eventually find that: $g(u, v) = f(u, u - v)$ and this gives us the conditions $u > 0$ and $0 < u - v < 1$, which reduces to $u - 1 < v < u$. Finding the marginal requires integrating the joint density. Notice that we would fix v, so this means that $u > 0$ and $v < u < v + 1$, for a fixed v. If we would have $v > 0$, then

$$g_V(v) = \int_v^{v+1} g(u, v)du = \int_v^{v+1} \lambda e^{-\lambda u}du = -e^{\lambda u}\Big|_{u=v}^{u=v+1} = e^{-\lambda v} - e^{-\lambda v + \lambda}$$

There is one about moment generating functions again, related to the idea that if X is gamma alpha 1 lamaba Y is alpha 2 lamabda and they are independent, there is something similar that we did similar to this in class a few days ago.

There will be on on the cetnral limit theorem.

$$\{\sqrt{x_1} + \sqrt{X_2} + ... + \sqrt{X_n} \leq x\} \approx \Phi(something)$$

We have to find $E(\sqrt{X_1})$ and $Var(\sqrt{X_2})$. There were some questions like this in the book. 34 and 35 dealt with squares, but we here have radicals.