Practice Problems (up to Midterm 1)

1.1: Set theory

E.g. coin tossing $\rightarrow S = \{H, T\}$

SAT scores $\rightarrow S = \{200, \dots, 800\}$

Heights $\rightarrow S = (0, ?)$

Proof. $A \cap B = B \cap A$

$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$

$= \{x \mid x \in B \text{ and } x \in B\}$

$= B \cap A$. ∎

Proof. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Show $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$:

$x \in A \cap (B \cup C) \Rightarrow x \in A$ and $x \in B \cup C$

$x \in B \cup C \Rightarrow x \in B$ or $x \in C$

$\Rightarrow x \in A \cap B$ or $x \in A \cap C$

$\Rightarrow x \in (A \cap B) \cup (A \cap C)$.

Similarly, show $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$:

$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$ and

$A \cap C = \{x \mid x \in A \text{ and } x \in C\}$

Thus, $(A \cap B) \cup (A \cap C)$ implies

$(x \in A \text{ and } x \in B)$ or $(x \in A \text{ and } x \in C)$

$\Rightarrow x \in A \cap (B \cup C)$. ∎

Proof. Show $\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c$.

$\omega \notin \bigcup_i A_i \Leftrightarrow \forall i, \omega \notin A$

$\Leftrightarrow \forall i, \omega \in A^c \Leftrightarrow \omega \in \bigcap_i A_i^c$. ∎

note: $\bigcup_{i=1}^{\infty} A_i := \{x \mid x \in A_i \text{ for some } i\} := \exists$ some i s.t. $x \in A_i$.

1.2: Probability Theory Basics

1.2.1: Axiomatic Foundations

Proof. Show that if $A_1, A_2, \dots \in \mathcal{B}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{B}$.

Let $A, B \in \mathcal{B}$. Then $A^c \in \mathcal{B}, B^c \in \mathcal{B}$.

So, $A^c \cup B^c \in \mathcal{B} \Rightarrow (A \cap B)^c \in \mathcal{B}$

$\Rightarrow A \cap B \in \mathcal{B}$. ∎

Eg (σ-algebra). Let $S = \mathbb{R}$, $\mathcal{B} :=$ σ-algebra containing all sets of the form $[a,b], (a,b], [a,b), (a,b)$, $a, b \in \mathbb{R}$.

$\mathcal{B}' :=$ σ-algebra containing all sets of the form $[a,b)$ for $a, b \in \mathbb{Q}$.

Proof. Let $a \in \mathbb{Q}$. Show $\{a\} \in \mathcal{B}'$.

$\{a\} = \lim_{n \to \infty} \bigcap_{i=1}^{n} [a, a + \frac{1}{n})$, $n \in \mathbb{N}$

$\Rightarrow \{a\}$ closed under countable intersection

$\Rightarrow \{a\} \in \mathcal{B}'$. ∎

Proof. Let $\{a\} \in \mathbb{R}$. Show $\{a\} \in \mathcal{B}'$.

Let a_n, b_n be a sequence of rational numbers s.t. $a_n \uparrow a$ and $b_n \downarrow a$.

Then, $\{a\} = \lim_{n \to \infty} \bigcap_{i=1}^{n} [a_n, b_n)$, closed under countable intersection.

$\Rightarrow \{a\} \in \mathcal{B}'$. ∎

Ex (Defining Prob.'s): Tossing fair coin: $S = \{H, T\} \Rightarrow P(\{H\}) := P(\{T\}) := \frac{1}{2}$.

Tossing an unfair coin: $\begin{cases} P(\{H\}) = q \\ P(\{T\}) = 1 - q \end{cases}$ for some $q \in [0, 1]$.

1.2.2: Calculus of Probabilities

Proof. Show $A \subset B \Rightarrow P(A) \leq P(B)$.

$P(B) = P(A) + \underbrace{P(B \cap A^c)}_{\geq 0} \geq P(A)$

$\Rightarrow P(B) \geq P(A)$. ∎

Proof. Show Bonferroni's inequality, i.e. $P(A \cap B) \geq P(A) + P(B) - 1$:

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

$\Rightarrow P(A \cap B) = P(A) + P(B) \underbrace{- P(A \cup B)}_{\leq 1}$

$\Rightarrow P(A \cap B) \geq P(A) + P(B) - 1$. ∎

Proof. Show more general Bonferroni's inequality: $P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$:

$P\left(\bigcup_{i=1}^{n} A_i^c\right) \leq \sum_{i=1}^{n} P(A_i^c)$, Boole's inequality

$1 - P\left(\bigcap_{i=1}^{n} A_i\right) \leq n - \sum_{i=1}^{n} P(A_i)$, Demorgan's extension: $\bigcup_i A_i^c = \left(\bigcap A_i\right)^c$ and complement rule: $P(A_i^c) = 1 - P(A_i)$

$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$. ∎

1.2.3: Counting

Ex (Lottery ticket): From the numbers $1, 2, \dots, 44$ a person may pick any six for their ticket

$n = 44, r = 6$

a. ordered w/o replacement $\left(P_r^n = \frac{n!}{(n-r)!}\right)$.

$P_6^{44} = \frac{44!}{38!} = 44 \times 43 \times 42 \times 41 \times 40 \times 39 = 5{,}082{,}517{,}440$

b. ordered w/ replacement $(n^r) = 44^6 = 7{,}256{,}313{,}856$

c. unordered w/o replacement $\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{44!}{6!\,38!} = \frac{44 \times 43 \times 42 \times 41 \times 40 \times 39}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 7{,}059{,}052$

d. unordered w/ replacement $\binom{n+r-1}{r} = \frac{(n+r-1)!}{(n-1)!} = \frac{49!}{6!\,43!} = \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 13{,}983{,}316$

1.2.4: Enumerating Options

Ex (Poker): Suppose the events do not depend on order, so we use the unordered outcomes.

a) How many 5-card hands can be chosen from 52-deck cards?

unordered, sampling w/o replacement

$$\binom{52}{5} \ (*)$$

$\Rightarrow P(\{\text{(random) 5-card hand}\}) = 1/\binom{52}{5} = 2{,}598{,}960.$

note: (b)-(d) assume a 5-card hand, i.e. "out of a" 5-card hand

b) $P(\{4 \text{ aces}\}) = 48/\binom{52}{5}$ $\quad \underset{52}{\underline{A}}\ \underset{51}{\underline{A}}\ \underset{50}{\underline{A}}\ \underset{49}{\underline{A}}\ \underline{*}$

where 48: # ways to choose 5^{th} card

c) $P(\{4 \text{ of a kind}\}) = \frac{1}{(*)} 48 \cdot 13$ $\quad \underline{8D}\ \underline{8S}\ \underline{8H}\ \underline{8C}\ \underline{*}$

where 13: # ways to choose denomination (2, 3, …, A).

d) $P(\{\text{exactly one pair}\}) = \frac{1}{(*)} 13 \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot 4^3$

remaining denominations — what suit

$\underline{2H}\ \underline{2D}\ \underbrace{\underline{*}\ \underline{*}\ \underline{*}}_{13-1=12}$

Ex (sampling w/ replacement): Assume "uniform sampling"; $r=2$ items from $n=3$ choices w/ replacement, i.e.

ordered	(1,1)	(2,2)	(3,3)	(1,2) (2,1)	(1,3) (3,1)	(2,3) (3,2)
unordered	{1,1}	{2,2}	{3,3}	{1,2}	{1,3}	{2,3}
Prob.	1/9	1/9	1/9	2/9	2/9	2/9

1.3: Conditional Probabilities and Independence

Ex (Four Aces): $P(\{4 \text{ aces}\}) = 1/\binom{52}{4} = \frac{4\times3\times2\times1}{52\times51\times50\times49}$

$= P(1^{st}\text{ ace}) P(2^{nd}\text{ ace} \mid 1^{st}\text{ ace}) \cdot P(3A \mid A1, A2) \cdot P(4A \mid A1, A2, A3)$

$= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49}.$

Ex (Three Prisoners): Three prisoners A, B, C; one of them chosen to be pardoned

Let A, B, C be pardon events, i.e. $P(A)=P(B)=P(C)=1/3$

Let W: event that warden tells A that B dies

$$P(A|W) = \frac{P(W|A)P(A)}{P(W)} = \frac{P(W|A)P(A)}{P(W|A)P(A) + P(W|C)P(C)} = \frac{(\frac{1}{3})\cdot(\frac{1}{2})}{\frac{1}{3}(\frac{1}{2}) + \frac{1}{3}(1)} = \dots = \frac{1}{3}.$$

Alternatively,

$$P(A|B^c) = \frac{P(A\cap B^c)}{P(B^c)} \overset{A\subset B^c}{=} \frac{P(A)}{P(B^c)} = \frac{P(A)}{1-P(B)} = \frac{1/3}{2/3} = 1/2.$$

Ex (Dice Rolling): $S = \{1, \dots, 6\}$. Here, $\tilde{S} = \{(r_1, \dots, r_4)\} \in \{1, \dots, 6\}^4$

$P(\{\text{throwing at least one 6 in 4 rolls}\}) = 1 - P(\text{no 6 in 4 rolls})$

$= 1 - \prod_{i=1}^{4} (\text{no 6 in role } i)$

$= 1 - (1-\frac{1}{6})^4$

$= 1 - (\frac{5}{6})^4 = 0.518.$

Proof. Show $A \perp\!\!\!\perp B \Rightarrow A \perp\!\!\!\perp B^c$

$P(A\cap B^c) = P(A) - P(A\cap B)$

$\overset{\text{ind}}{=} P(A) - P(A)P(B)$

$= P(A)(1-P(B))$

$= P(A)P(B^c).$ ∎

1.4: Random variables

Ex (RV's): Experiments — RV

1) toss 2 Dice — X = sum of 2 numbers

$S = \{1,\dots,6\}\times\{1,\dots,6\}.$

suppose $s = ((s_1, s_2)) \Rightarrow X := X(s) = s_1 + s_2$

2) toss coin 25 times — X = # heads

$S = \{(c_1, \dots, c_{25})\} \Rightarrow X := X(s) = \sum_{i=1}^{25} \mathbb{1}(c_i = H)$

Ex (Three coin tosses): Let X: number of heads out of tossing coin 3 times

$s \in S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

s	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(s)$	3	2	2	1	2	1	1	0

$\Rightarrow$ the range of X, denoted $\mathcal{X}$, is associated w/ the following prob's:

$\mathcal{X}$	0	1	2	3
P_X	1/8	3/8	3/8	1/8

Ex (Dist. of a RV): Let X: # heads after tossing coin 50 times

$$\Rightarrow \mathcal{X} = \{0, \dots, 50\} \Rightarrow P_X(X=i) = \frac{\binom{50}{i}}{2^{50}} \quad \forall i \in \mathcal{X}.$$

1.5: Distribution Functions

Ex (Tossing for a Head): Let p := prob. that a coin turns head

X := # independent tosses (trials) on which first head occurs

$\Rightarrow X \sim \text{Geo}(p)$

$$\Rightarrow X \in \{1, 2, \dots\} \Rightarrow P(X=x) = \begin{cases} (1-p)^{x-1} p, & \forall x \in \{1, 2, \dots\} \\ 0, & \text{otherwise} \end{cases}$$

$\Rightarrow F_X(x) = P(X \le x), \ x \ge 0$

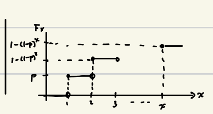
$= \sum_{i=1}^{x} (1-p)^{i-1} p$

change-of-var (FS)

$= p \sum_{k=0}^{x-1} (1-p)^k$, letting $k = i-1$

$= p \frac{1-(1-p)^x}{1-(1-p)}$

$= 1-(1-p)^x, \ x \ge 1 \quad (*)$



recall: (Geometric series)

Prop. $\forall r \ne 1 \in \mathbb{R}$, $\sum_{k=0}^{n-1} r^k = \frac{1-r^n}{1-r}$

Prop. $\forall |r| < 1$, i.e. $r \in (-1, 1)$, $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$; otherwise, series diverges.

Proof. Show that $(*)$ is a cdf:

i) $\lim_{x\to-\infty} F(x) = 0$ since $F_X(x) = 0 \ \forall x < 1.$

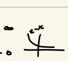
$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} 1-(1-p)^x = 1$ for $x \in \mathbb{Z}$.

ii) $\sum_{i=1}^{x} (1-p)^{i-1} p$ contains more positive terms as x increases

iii) $\underbrace{\lim_{\epsilon \downarrow 0} F_X(x+\epsilon) = F_X(x)}_{\text{convention (FS)}}$. ∎

Ex (continuous cdf): Let $F_X(x) = \frac{1}{1+e^{-x}}$ which satisfies

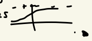
i) $\lim_{x\to-\infty} F_X(x) = 0$ since $\lim_{x\to-\infty} e^{-x} = \infty$

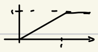


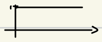
$\lim_{x\to\infty} F_X(x) = 1$ since $\lim_{x\to\infty} e^{-x} = 0$

ii) $\frac{d}{dx}F_X(x) = \frac{d}{dx}(1+e^{-x})^{-1} = \frac{e^{-x}}{(1+e^{-x})^2} > 0$ 

$\Rightarrow$ increasing $\Rightarrow$ nondecreasing.

iii) log. function is (right/left) continuous 

<u>Ex (RV CDF mixture)</u>: Let $X \sim$ Uniform s.t. $F_X(x) = \begin{cases} x, & x \in [0,1) \\ 0, & x \le 0 \\ 1, & x \ge 1 \end{cases}$ 

$F_Y(y) = \begin{cases} 0, & y < 0 \\ 1, & y \ge 0 \end{cases}$ 

Then, $F_Z(z) = F_X(z) + F_Y(z)$ can be shown to be a valid cdf.

<u>Ex (Identically Distributed RV's)</u>: Toss fair coin 3 times

Let $X :=$ # heads, $Y :=$ # tails

$\Rightarrow X \overset{d}{=} Y$ since $P(X=i) = P(Y=i)\ \forall i \in S$.

But, $X \neq Y$ since $\nexists$ sample points s.t. $X(s) = Y(s)$.

1. Do problems 1.1, 1.2.

(Hint: In problem 1.2, recall the set difference operation: $A \setminus B := \{x | x \in A \text{ and } x \notin B\}$.)

1.1) For each of the following experiments, describe the sample space

a. Toss a coin four times: Each toss has two possible outcomes: H or T.

$\boxed{S = \{HHHH, HHHT, HHTH, HTHH, THHH, HHTT, \dots, TTTT\}}$.

S is a set of all sequences of length 4; there are $2^4 = 16$ total number of outcomes

b. count the number of insect-damaged leaves on a plant:

S consists of all possible nonnegative integers, i.e.

$\boxed{S = \{0, 1, 2, 3, \dots, N\}}$ where N: total number of leaves on the plant

$\Leftrightarrow \boxed{S = \{0, 1, 2, 3, \dots\}}$ if N not specified

c. Measure the lifetime (in hours) of a particular brand of light bulb

Lifetime is a positive real-valued measurement: $\boxed{S = (0, \infty)}$

d. Record the weights of 10-day old rats

If only one rat is measured, the sample space is $\boxed{S = (0, \infty)}$

Suppose a group of n rats is measured. Then, $\underline{S = (0, \infty)^n}$, the n-dimensional positive real space.

e. observe the proportion of defectives in a shipment of electronic components

If the shipment has N components, $\boxed{S = \{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\}}$ $\Leftrightarrow$ $\boxed{S = [0, 1]}$ if N is very large or unspecified.

1.2) verify the following identities; Recall: $A \setminus B := \{x | x \in A \text{ and } x \notin B\}$

a. $A \setminus B = A \setminus (A \cap B) = A \cap B^c$

$x \in A \setminus (A \cap B) \Leftrightarrow x \in A$ and $(x \notin A \cap B)$

$\Leftrightarrow x \in A$ and $(x \in (A \cap B)^c)$

$\Leftrightarrow x \in A$ and $(x \in A^c$ or $x \in B^c)$, Demorgan's

$\Leftrightarrow (x \in A$ and $x \in A^c)$ or $(x \in A$ and $x \in B^c)$, Distributive

$\Leftrightarrow x \in \emptyset$ or $(x \in A$ and $x \in B^c)$

$\Leftrightarrow x \in A$ and $x \in B^c$

$\Leftrightarrow x \in A \cap B^c$

$\Leftrightarrow x \in A$ and $x \notin B$

$\Leftrightarrow x \in A \setminus B$. ∎

$A \setminus B = A \setminus (A \cap B) = A \cap B^c$

Proof. $A \setminus B = \{x : x \in A, x \notin B\}$

Also, $A \setminus (A \cap B) = \{x : x \in A, x \notin (A \cap B)\}$

If $x \in A$ and $x \notin (A \cap B)$, then necessarily $x \notin B$. Hence,

$A \setminus (A \cap B) = \{x : x \in A, x \notin B\} = A \setminus B$.

Moreover,

$A \cap B^c = \{x : x \in A, x \in B^c\} = \{x : x \in A, x \notin B\}$.

Thus, $A \setminus B = A \setminus (A \cap B) = A \cap B^c$. ∎

b. $B = (B \cap A) \cup (B \cap A^c)$

$x \in (B \cap A) \cup (B \cap A^c) \Leftrightarrow (x \in B$ and $x \in A)$ or $(x \in B$ and $x \notin A)$

$\Leftrightarrow x \in B$ and $(x \in A$ or $x \in A^c)$, Distributive

$\Leftrightarrow x \in B$ and $x \in \emptyset$

$\Leftrightarrow x \in B$. ∎

Proof. $B = (B \cap A) \cup (B \cap A^c)$

Let $x \in B$. Then either $x \in A$ or $x \in A^c$.

- If $x \in A$, then $x \in B \cap A$
- If $x \in A^c$, then $x \in B \cap A^c$

Thus, $x \in (B \cap A) \cup (B \cap A^c)$.

Conversely, if $x \in (B \cap A) \cup (B \cap A^c)$, then clearly $x \in B$.

Therefore, $B = (B \cap A) \cup (B \cap A^c)$. ∎

c. $B \setminus A = B \cap A^c$

$x \in B \setminus A \Leftrightarrow x \in B$ and $x \notin A$

$\Leftrightarrow x \in B \cap A^c$. ∎

Proof. $B \setminus A = B \cap A^c$.

By definition, $B \setminus A = \{x : x \in B, x \notin A\}$.

On the other hand,

$B \cap A^c = \{x : x \in B, x \in A^c\} = \{x : x \in B, x \notin A\}$

Hence,

$B \setminus A = B \cap A^c$. ∎

d. $A \cup B = A \cup (B \cap A^c)$

$x \in A \cup B \Leftrightarrow x \in A$ or $x \in B$

$\Leftrightarrow x \in A$ or $(x \in B$ and $x \notin A)$

$\Leftrightarrow A \cup (B \cap A^c)$. ∎

Proof. $A \cup B = A \cup (B \cap A^c)$

First, let $x \in A \cup (B \cap A^c)$

- If $x \in A$, then $x \in A \cup B$
- If $x \in B \cap A^c$, then $x \in B$, so $x \in A \cup B$

Thus,

$A \cup (B \cap A^c) \subseteq A \cup B$.

Conversely, let $x \in A \cup B$.

- If $x \in A$, then clearly $x \in A \cup (B \cap A^c)$
- If $x \in B$ and $x \notin A$, then $x \in B \cap A^c$, hence $x \in A \cup (B \cap A^c)$.

Therefore,

$A \cup B \subseteq A \cup (B \cap A^c)$

Combining both inclusions,

$A \cup B = A \cup (B \cap A^c)$. ∎

2. Approximately one-third of all human twins are identical (one-egg), and two-thirds are fraternal (two-egg) twins. Identical twins are necessarily the same sex, with male and female being equally likely. Among fraternal twins, approximately one-fourth are both female, on-fourth are both male, and half are one male and one female. Finally, among all U.S. births, approximately 1 in 90 is a twin birth. Define the following events: $A = \{$a U.S. birth results in twin females$\}$, $B = \{$a U.S. birth results in identical twins$\}$, and $C = \{$a U.S. birth results in twins$\}$.

(i) State, in words, the events $A \cap B \cap C$, $B \setminus A$, $A \setminus B$, $A \cup B$, and $C \setminus (A \cap B)$.

(ii) Find the probabilities of all such events in (i).

Let $A = \{$a U.S. birth results in twin females$\}$

$B = \{$a U.S. birth results in identical twins$\}$

$C = \{$a U.S. birth results in twins$\}$

i) $A \cap B \cap C$: births that are identical twin females ($A, B \subseteq C$, thus, just $A \cap B$)

$B \setminus A$: births that are identical twins but not twin females, i.e. identical twin males

$A \setminus B$: births that are twin females but not identical, i.e. fraternal twin females

$A \cup B$: births that are twin females or identical twins (including their overlap)

$C \setminus A \cap B$: births that are twins but not identical twin females (all twin births except identical FF)

ii) $P(A \cap B \cap C) = P(A)P(B)P(C) = \frac{1}{90}\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) = \frac{1}{540}$

$P(B \setminus A) = P(C) \cdot P(B \setminus A | C)$

$= \frac{1}{90} P(\text{identical}) P(MM | \text{identical}) = \frac{1}{90}\left(\frac{1}{3}\right)\left(\frac{1}{2}\right) = \frac{1}{540}$

$P(A \setminus B) = P(C) \cdot P(A \setminus B | C)$

$= \frac{1}{90} P(\text{fraternal}) P(FF | \text{fraternal}) = \frac{1}{90}\left(\frac{2}{3}\right)\left(\frac{1}{4}\right) = \frac{1}{540}$

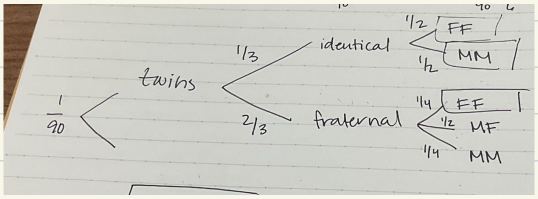
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= \frac{1}{90}\left[\frac{1}{3}\frac{1}{2} + \frac{2}{3}\frac{1}{4}\right] + \left[\frac{1}{90} \cdot \frac{1}{3}\right] - \left[\frac{1}{90} \cdot \frac{1}{3} \cdot \frac{1}{2}\right]$

$= \frac{1}{90}\left(\frac{1}{3}\right) + \frac{1}{90}\left(\frac{1}{3}\right) - \frac{1}{540}$

$= \frac{1}{180}$

$P(C \setminus (A \cap B)) = P(C) - P(A \cap B) = \frac{1}{90} - \left(\frac{1}{540}\right) = \frac{1}{108}$



---

3. Let the sample space S be the real line. Suppose that a sigma algebra $\mathcal{B}$ contains all half-closed intervals of the form $(-\infty, a]$ where a is a rational number. (Note: $(-\infty, a] = \{x | x \leq a\}$). Show that the following sets are elements of $\mathcal{B}$

(i) all singleton sets $\{a\}$ where a is a rational number.

(ii) all singleton sets $\{a\}$ where a is a real number.

(iii) all intervals of the form $(a, b]$, where a and b are real numbers.

(iv) Give an example of an element of $\mathcal{B}$ that is neither empty set, nor S, nor any of the forms mentioned above.

(Hint: This is done by verifying that the set in question can be obtained from known elements of $\mathcal{B}$ via countably many set operations. Use the fact that any real number can be constructed as the limit of a sequence of rational numbers).

Solution: Let $S = \mathbb{R}$ and let $\mathcal{B}$ be the σ-algebra on S that contains every half-closed ray $(-\infty, q]$ with $q \in \mathbb{Q}$.

For any real $b \in \mathbb{R}$, choose a decreasing sequence of rationals $q_n \downarrow b$. Then

$$(-\infty, b] = \bigcap_{n=1}^{\infty} (-\infty, q_n] \in \mathcal{B},$$

since $\mathcal{B}$ is closed under countable intersections. Hence all rays $(-\infty, b]$ with real b are in $\mathcal{B}$.

Also, for any real a,

$$(-\infty, a) = \bigcup_{q < a,\, q \in \mathbb{Q}} (-\infty, q] \in \mathcal{B},$$

since $\mathbb{Q}$ is countable and $\mathcal{B}$ is closed under countable unions.

recall: $\mathbb{R} = \mathbb{Q} \cup \mathbb{Q}^c \Rightarrow$ above solution w.r.t. $\mathbb{R}$ applies for $\mathbb{Q}$.

i) For rational a,

$\underline{\{a\} = (-\infty, a] \setminus (-\infty, a)} = (-\infty, a] \cap ((-\infty, a))^c \in \mathcal{B}$.

ii) The same identity works for any real a:

$\{a\} = (-\infty, a] \setminus (-\infty, a) = (-\infty, a] \cap ((-\infty, a))^c \in \mathcal{B}$

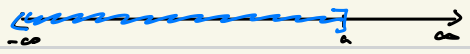
iii) For $a, b \in \mathbb{R}$

$(a, b] = (-\infty, b] \setminus (-\infty, a] = (-\infty, b] \cap ((-\infty, a])^c \in \mathcal{B}$.

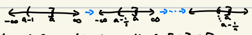
iv) The set of rationals $\mathbb{Q}$ belongs to $\mathcal{B}$ b/c it is a countable union of singletons: $\mathbb{Q} = \bigcup_{q \in \mathbb{Q}} \{q\} \in \mathcal{B}$.

It is neither $\emptyset$ nor S, not a singleton, and not an interval of the form $(a, b]$ or a ray $(-\infty, a]$. Hence $\mathbb{Q}$ is a valid example,

Let $S = \mathbb{R}$ w/ a $\mathcal{B}$ that contains all intervals of the form $(-\infty, a]$ for $a \in \mathbb{Q}$ (rational).



$\Leftrightarrow \{x | x \leq a, a \in \mathbb{Q}\}$.



i) Let $a \in \mathbb{Q}$. I will show that $\{a\} \in \mathcal{B}$

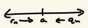
For $n \in \mathbb{N}$, $a - \frac{1}{n} \in \mathbb{Q}$.

Then, $(-\infty, a) = \lim_{n \to \infty} \bigcup_{i=1}^{n} (-\infty, a - \frac{1}{n}) \in \mathcal{B}$.

Since $(-\infty, a] \in \mathcal{B}$,

$\{a\} = (-\infty, a] \setminus (-\infty, a) \in \mathcal{B}$. ∎

ii) Let $a \in \mathbb{R}$, $n \in \mathbb{N}$



Pick rationals $q_n \downarrow a$ (decreasing sequence of rationals converging to a from above, i.e. $q_n > a \ \forall n$.)

Then,

$(-\infty, a] = \lim_{n \to \infty} \bigcap_{i=1}^{n} (-\infty, q_n] \in \mathcal{B}$. (★)

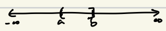
Pick rationals $r_n \uparrow a$ (increasing sequence of rationals converging to a from below, i.e. $r_n < a \ \forall n$.)

Then,

$(-\infty, a) = \lim_{n \to \infty} \bigcup_{i=1}^{n} (-\infty, r_n] \in \mathcal{B}$.

Thus,

$\{a\} = (-\infty, a] \setminus (-\infty, a) \in \mathcal{B}$ ∎



iii) Let $a, b \in \mathbb{R}$ where $a < b$.

From (ii), we know that $(-\infty, a] \in \mathcal{B}$ via (★).

Similarly, $(-\infty, b] \in \mathcal{B}$.

Therefore, $(a, b] = (-\infty, b] \setminus (-\infty, a] \in \mathcal{B}$. ∎

iv) Let $a \in \mathbb{R}$.

From (ii), $\{a\} \in \mathcal{B}$.

Thus, $\underline{\{a\}^c = \mathbb{R} \setminus \{a\} \in \mathcal{B}}$

as $\mathcal{B}$ is closed under complementation. ∎

4. Let $A_1 \subset A_2 \ldots \subset A_n \subset \ldots S$ be an increasing sequence of subsets in a sigma algebra $\mathcal{B}$ associated with a sample space S. The limit of this sequence of subsets is defined as

$$A := \lim_{n\to\infty} A_n := \cup_{i=1}^{\infty} A_i := \{x | x \in A_i \text{ for some } i < \infty\}.$$

Let P be a probability function on $\mathcal{B}$. Use the axioms of probability to show that

(i) The sequence of $P(A_n)$ increases to a finite limit.

(ii) In fact, that limit is equal to $P(A)$.

<u>Solution</u>: Let $(A_n)_{n\geq 1}$ be an increasing sequence of sets:

$$A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots \subset S,$$

and let $A := \lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$.

Define the disjoint sets $B_1 := A_1$, $B_n := A_n \setminus A_{n-1}$ $(n \geq 2)$.

Then, the sets $B_1, B_2, \ldots$ are pairwise disjoint, and we have

$$A_n = \bigcup_{i=1}^{n} B_i, \quad A = \bigcup_{i=1}^{\infty} B_i$$

i) By finite additivity on disjoint unions,

$$P(A_{n+1}) = P(A_n \cup (A_{n+1} \setminus A_n)) = P(A_n) + P(A_{n+1} \setminus A_n) \geq P(A_n).$$

Thus $\{P(A_n)\}$ is nondecreasing. Moreover, since $A_n \subseteq S$,

$$P(A_n) \leq P(S) = 1.$$

Hence, $(P(A_n))$ is a nondecreasing sequence bounded above by 1, so it converges to a finite limit. ∎

ii) By countable additivity on disjoint unions,

$$P(A_n) = \sum_{i=1}^{n} P(B_i), \quad P(A) = \sum_{i=1}^{\infty} P(B_i).$$

Therefore,

$$\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \sum_{i=1}^{\infty} P(B_i) = P(A). \blacksquare$$

I refer to the axioms of probability as AOP(1, 2, or 3).

i) $A_n \subset A_{n+1} \Leftrightarrow x \in A_n \Rightarrow x \in A_{n+1}$

$$\Rightarrow P(A_n) \leq P(A_{n+1}).$$

By AOP(1),

$$P(A_n) \geq 0$$

and AOP(2) asserts

$$P(A_n) \leq P(S) = 1.$$

Thus, the sequence of $P(A_n)$ converges to a finite limit, i.e.

$$\lim_{n\to\infty} P(A_n) = L \text{ for some } L \in [0,1]. \blacksquare$$

ii) Define $B_1 := A_1$, $B_n := A_i \setminus A_{i-1}$, $i \geq 2$.

The sets B_i $(i=1,2,\ldots)$ are disjoint and

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i = A.$$

By AOP(3),

$$P(A) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i).$$

$$\Rightarrow \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \sum_{i=1}^{\infty} P(B_i) = P(A). \blacksquare$$

5. (Counting) Do problems 1.20, 1.21.

1.2) Let E_i : event that day i receives no call, $i = 1, \ldots, 7$. We want

$$P(\text{at least one call each day}) = 1 - P\left(\bigcup_{i=1}^{7} E_i\right)$$

By the inclusion-exclusion principle,

$$P\left(\bigcup_{i=1}^{7} E_i\right) = \sum_{k=1}^{7} (-1)^{k+1} \sum_{1 \leq i_1 < \cdots < i_k \leq 7} P(E_{i_1} \cap \cdots \cap E_{i_k}).$$

If k specific days get no calls, then all 12 calls must fall among the remaining $7-k$ days. Since each call independently chooses a day uniformly,

$$P(E_{i_1} \cap \cdots \cap E_{i_k}) = \left(\frac{7-k}{7}\right)^{12}.$$

There are $\binom{7}{k}$ ways to choose which k days are empty. Hence,

$$P\left(\bigcup_{i=1}^{7} E_i\right) = \sum_{k=1}^{7} (-1)^{k+1} \binom{7}{k} \left(\frac{7-k}{7}\right)^{12}.$$

Therefore,

$$P(\text{at least one call each day}) = 1 - \sum_{k=1}^{7} (-1)^{k+1} \binom{7}{k} \left(\frac{7-k}{7}\right)^{12} = \sum_{k=0}^{7} (-1)^{k} \binom{7}{k} \left(\frac{7-k}{7}\right)^{12}.$$

Equivalently,

$$P = \frac{1}{7^{12}} \sum_{k=0}^{7} (-1)^k \binom{7}{k} (7-k)^{12}.$$

Computing this sum gives

$$P = \frac{3{,}162{,}075{,}840}{13{,}841{,}287{,}201} \approx 0.2285.$$

We enumerate all integer partitions of 12 into 7 positive nonzero integers, then count the number of ways to arrange the calls under the given partition. For example, for the partition $(1,1,1,1,2,3,3)$, we have $\binom{12}{3}$ ways of choosing the first three calls, $\binom{12-3}{3}$ ways of choosing the second three calls, and $\binom{7}{2}$ ways of placing these subsets of calls on a day of the week. Then we have $\binom{9-3}{2}$ ways of choosing the next two calls, and $\binom{7-1}{1}$ ways of chosing the day of the week to place those calls on. Then the rest of the calls must be assigned, one each to the remaining days, so that is 4 options of call for the first day, 3 for the second and so on, giving 4! ways to assign the remaining calls. We use a similar logic and get the following counts for the 7 partitions:

$$\text{Partition } (1,1,1,1,1,1,6): \binom{12}{6}\binom{7}{1} \cdot 6! = 4{,}656{,}960$$

$$\text{Partition } (1,1,1,1,1,2,5): \binom{12}{5}\binom{7}{1} \cdot \binom{7}{2}\binom{6}{1} \cdot 5! = 83{,}825{,}280$$

$$\text{Partition } (1,1,1,1,1,3,4): \binom{12}{4}\binom{7}{1} \cdot \binom{8}{3}\binom{6}{1} \cdot 5! = 139{,}708{,}800$$

$$\text{Partition } (1,1,1,1,2,2,4): \binom{12}{4}\binom{7}{1} \cdot \binom{8}{2}\binom{6}{2}\binom{6}{2} \cdot 4! = 523{,}908{,}000$$

$$\text{Partition } (1,1,1,1,2,3,3): \binom{12}{3}\binom{9}{3}\binom{7}{2} \cdot \binom{6}{2}\binom{5}{1} \cdot 4! = 698{,}544{,}000$$

$$\text{Partition } (1,1,1,2,2,2,3): \binom{12}{3}\binom{7}{1} \cdot \binom{9}{2}\binom{7}{2}\binom{5}{2}\binom{6}{3} \cdot 3! = 1{,}397{,}088{,}000$$

$$\text{Partition } (1,1,2,2,2,2,2): \binom{12}{2}\binom{10}{2}\binom{8}{2}\binom{6}{2}\binom{4}{2}\binom{7}{5} \cdot 2! = 314{,}344{,}800$$

and the total of the above counts is 3,162,075,840, which multiplied by $\frac{1}{7^{12}}$ is approximately .2285 as desired.

1.20) My telephone rings 12 times each week, the calls being randomly distributed among the 7 days. What is the prob. that I get at least one call each day? (A: 0.2285)

We are modeling 12 telephone calls that arrive during the week where

$$\left.\begin{array}{l}\cdot \text{ Each call is a distinct event} \\ \cdot\ \forall \text{ call, decide which of 7 days it happens}\end{array}\right\} \Rightarrow \begin{array}{l}\text{An outcome is a sequence of 12 choices} \\ \text{s.t. each choice is one of 7 days, i.e.} \\ (d_1, d_2, \ldots, d_{12}),\ d_i \in \{1, 2, \ldots, 7\}\end{array}$$

By the Fundamental Thm. of counting (FToC) and supported by the ordered, with replacement scheme,

$$\underline{|\Omega| = 7^{12}} \qquad \underset{\text{choice 1}}{\underline{1\text{–}7}}\ \underset{\text{choice 2}}{\underline{1\text{–}7}} \cdots \underset{\text{choice 12}}{\underline{1\text{–}7}}$$

Let A_j = set of assignments that miss day j. We want assignments in the following assignments: $(A_1 \cup \cdots \cup A_7)^c$.

Then, by inclusion-exclusion principle:

$$\#\{\text{no empty days}\} = \sum_{i=0}^{7} (-1)^i \binom{7}{i} \left(\frac{7-i}{7}\right)^{12}$$

$$= 1 - \sum_{i=1}^{7} (-1)^{i+1} \binom{7}{i} \left(\frac{7-i}{7}\right)^{12} \quad (\star)$$

where
- $(-1)^i \Leftrightarrow$ alternating signs; corrects overcounting, e.g. subtracting assignments that misses one day double-counts assignments that miss two particular days, etc. So, inclusion-exclusion gives the exact correction; Generalizes: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

can prove via induction; instead taking this principle to be true.

- $\binom{7}{i}$: choose i days to exclude for 7 choose i choices

Then, every call must go to one of the remaining $(7-i)$ days

$$\Rightarrow \left(\frac{7-i}{7}\right)^{12} \text{ assignments.}$$

$$\boxed{\text{So, } P\left(\begin{array}{c}\{\text{at least one} \\ \text{call a day}\}\end{array}\right) = \frac{\star}{|\Omega|} = \cdots \underset{\text{technology}}{=} \frac{3162075840}{13841287201} \approx \boxed{0.2285}.} \blacksquare$$

1.21) We have a closet containing n pairs of shoes, i.e. $2n$ individual shoes in total. Suppose $2r$ shoes are chosen at random, with $2r<n$.

The total number of ways to choose $2r$ shoes from $2n$ is $\binom{2n}{2r}$.

To avoid having a matching pair, we must choose at most one shoe from each pair.

- First choose which $2r$ pairs will contribute one shoe each: $\binom{n}{2r}$.
- From each of the chosen pairs, pick exactly one shoe (left/right): 2^{2r}

Thus, the number of favorable outcomes is

$$\binom{n}{2r}2^{2r}$$

Therefore, the required prob. is

$$P=\frac{\binom{n}{2r}2^{2r}}{\binom{2n}{2r}}. \blacksquare$$

1.21) A closet contains n pairs of shoes. If $2r$ shoes are chosen at random ($2r<n$), what is the prob. that there will be no matching pair in the sample? $\left(A: \frac{\binom{n}{2r}2^{2r}}{\binom{2n}{2r}}\right)$

We want $P(\{\text{no matching pairs in sample}\})$

$= P(\{\text{all chosen shoes from different pairs}\})$

We are choosing $2r$ shoes from $2n$ shoes, i.e. $\binom{2n}{2r}$.

Now consider how many ways to choose $2r$ shoes s.t. no full pair is chosen, i.e.

1a) choose which pairs will contribute shoes, i.e. $\binom{n}{2r}$

1b) choose one shoe from each selected pair
$\forall$ of the $2r$ chosen pairs, there are 2 choices, i.e. 2^{2r}

Thus, $\boxed{P(\{\text{no matching pair}\})=\frac{\binom{n}{2r}2^{2r}}{\binom{2n}{2r}}}$

Exam 1 Solutions

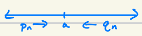
1) Prove the de Morgan's law, i.e. $(A\cup B)^c=(A^c\cap B^c)$ by using the definition of the basic set operations

$(A\cup B)^c=\{x\in\Omega: x\notin(A\cup B)\}=\{x\in\Omega: x\notin A \text{ and } x\notin B\}=\{x\in\Omega: x\in A^c \text{ and } x\in B^c\}=A^c\cap B^c$

Common mistakes: • $\{x\in\Omega: x\notin(A\cup B)\}=\{x\in\Omega: x\notin A \text{ or } x\notin B\}$
• showing only one direction like $(A\cup B)^c\subset(A^c\cap B^c)$

2) Let $\mathcal{B}$:= sigma of subsets of the real line that are generated by sets of the form (a,b) where a,b - rationals

i) Proof. I will show that every singleton, i.e. $\{a\}$ - a: real $\in\mathcal{B}$.



There exists two sequences of rationals $\{p_n\}$ and $\{q_n\}$ s.t. p_n is an increasing sequence that converge to a from below and q_n is a decreasing sequence that converges to a from above.

Since $(p_n,q_n)\in\mathcal{B}$, $\{a\}=\bigcap_{n=1}^{\infty}(p_n,q_n)\in\mathcal{B}$.

ii) Q: Is the set of all real numbers a member of $\mathcal{B}$.

Since $(-n,n)\in\mathcal{B}$ for all $n\in\mathbb{N}$,

$$\mathbb{R}=\bigcup_{n=1}^{\infty}(-n,n)\in\mathcal{B}. \blacksquare$$

Common mistakes: • Let $(S,\mathcal{B})$ be the sigma algebra defined here. We only have $S\subset\mathbb{R}$, and $(a,b)\in\mathcal{B}$ for all $a,b\in\mathbb{Q}$.

• Since it is not guaranteed that $S=\mathbb{R}$, proofs using complement are mostly wrong in (i) and (ii) b/c $(a,b)^c\neq(-\infty,a)\setminus[b,\infty)$ and $\{a\}^c\neq\mathbb{R}\setminus\{a\}$.

• For (i), using $(a-\frac{1}{n},a+\frac{1}{n})$ or something similar is wrong b/c $a+\frac{1}{n}$ is not guaranteed to be a rational

3) Consider an experiment of tossing a coin two times. Let X:= # heads, Y:= # tails (after two tosses). Construct a suitable sample space and define a valid prob. function on the two tosses. Show that X and Y have identical distributions, but the coin tosses are neither independent nor identically distributed.

Solution. The sample space for two coin tosses is

$$\Omega=\{HH,HT,TH,TT\}$$

In order for X and Y to have the same distribution,

$P(HH)=P(TT)$ since recall: $X\stackrel{d}{=}Y := P(X\in A)=P(Y\in A)\ \forall A\in\mathcal{B}$

In order for the two coin tosses to have dependent distributions, we must have

$P(\text{coin 2}=T\mid\text{coin 1}=H)\neq P(\text{coin 2}=T\mid\text{coin 1}=T)$, (#1)

so

$\frac{P(HT)}{P(HT)+P(HH)}\neq\frac{P(TT)}{P(TH)+P(TT)}$ (#2) since recall: $P(A|B)=\frac{P(A\cap B)}{P(B)}$

So the construction must satisfy

$P(TT)P(HH)\neq P(HT)P(TH)$ since $P(HT)[P(TH)+P(TT)]=P(TT)[P(HH)+P(HT)]$

$\Rightarrow P(HT)P(TH)+\cancel{P(HT)P(TT)}=P(TT)P(HH)+\cancel{P(TT)P(HT)}$

$\Rightarrow P(HT)P(TH)=P(TT)P(HH)$

In order for the two coins to not have identical distributions, we must have

$$P(\text{coin 2}=T)\neq P(\text{coin 1}=T)$$

So the construction must satisfy

$P(TT)+P(TH)\neq P(HT)+P(TT) \iff P(TH)\neq P(HT)$

In summary, answer must follow:

1) $P(TT)=P(HH)$, $X\stackrel{d}{=}Y$
2) $P(HT)\neq P(TH)$, non-identical coin tosses
3) $P(HH)P(TT)\neq P(HT)P(TH)$, dependent coin tosses

For example, $P(TT)=P(HH)=\frac{1}{4}$, $P(HT)=\frac{1}{3}$, $P(TH)=\frac{1}{6}$

4) Suppose that we throw 5 identical balls into 3 boxes independently (and in the same random manner), such that every ball will end up in one of the three boxes.

i) Define a sample space of outcomes.
Based on your definition, what are the prob's allocated to elements of sample space

ii) P(no empty box)

iii) P (no box contains more than 3 balls)

i) Since both balls and boxes are identical, an outcome is determined by the (unordered) occupancy triple of box counts (sorted nonincreasing):

$$S = \{(5,0,0), (4,1,0), (3,2,0), (3,1,1), (2,2,1)\}.$$

To assign prob.'s, view the underlying experiment as each of the 5 balls independently choosing one of 3 (temporarily labeled) boxes uniformly, so each microstate has prob. 3^{-5}.

The prob. of a macro-outcome is the number of labeled allocations collapsing to it divided by 3^5.
Then,

$$P(5,0,0) = \frac{3\cdot\binom{5}{5}}{3^5} = \frac{3}{3^5} = \dots = \frac{1}{81}$$

where 3: which of the 3 labeled boxes gets all 5 balls
$\binom{5}{5}$: number of ways to "choose" which balls go into this chosen box

$$P(4,1,0) = \frac{3\cdot\binom{5}{4}\cdot 2}{3^5} = \dots = \frac{30}{243} = \frac{3!\binom{5}{4}}{3^5}$$ where 3! : # distinct ways to arrange (4,1,0).

where 3: Pick which box gets the 4 balls; 3 choices
$\binom{5}{4}$: choose which 4 of the 5 balls go together
2: Pick which of the remaining 2 boxes get the (one ball); 2 choices

$$P(3,2,0) = \frac{6\cdot\binom{5}{3}}{3^5} = \frac{20}{81} = \frac{3!\binom{5}{3}}{3^5} = \frac{3!\binom{5}{3}\binom{2}{2}}{3^5}$$

where 3! : # distinct ways to arrange (3,2,0).
$\binom{5}{3}$: choose which of 3 balls go together

$$P(3,1,1) = \frac{3\cdot\binom{5}{3}\cdot\binom{2}{1}\cdot 2}{3^5} = \frac{20}{81}$$ ☆

where 3: Choose which box gets the three
→ $\binom{5}{3}$: which 3 balls go into this box
$\binom{2}{1}$: assign 2 balls to distinct boxes

$$P(2,2,1) = \frac{3\cdot\binom{5}{2}\cdot\binom{3}{2}\cdot 2}{3^5} = \frac{10}{27}$$ ☆

☆ Can also treat boxes and/or balls as distinguishable; requires more work for (ii) and (iii), but thesese approaches more closely resemble the solution and class practice (notes)

i) To assign prob.'s, view the underlying experiment as each of the 5 balls independently choosing one of 3 (temporarily labeled) boxes uniformly, so each microstate has prob. 3^{-5}.

$|S| = 3^5$ ⇒ prob.'s associated w/ sample space are $\frac{\#\text{outcomes in event}}{|S| = 3^5}$

3^5 ↓ ordered w/ replacement

ii) Find P(no empty box):

Let A_1, A_2, A_3: events that boxes 1, 2, and 3 are empty.

$$P(\text{no empty box}) = 1 - P(\text{at least one empty box}) = \left(P\left(\bigcup_{i=1}^{3} A_i\right)\right)^c = 3^5 - \left[\sum_{i=1}^{3} P(A_i) - [P(A_1\cap A_2) + P(A_2\cap A_3) + P(A_1\cap A_3)] + P(A_1\cap A_2\cap A_3)\right]$$

$$\#\text{non empty} \overset{IE}{=} \#(\text{all assignments}) - \sum_i \#(A_i) + \sum_{i<j} \#(A_i\cap A_j) - \#(A_1\cap A_2\cap A_3)$$

$$= 3^5 - 3\cdot 2^5 + \binom{3}{2}\cdot 1^5 - 0$$

$3\cdot 2^5$:
- box i is empty
- all balls in remaining 2 boxes ⇒ 2^5 assignments
- 3 choices of i

$\binom{3}{2}\cdot 1^5$:
- boxes i and j empty
- all balls go into single remaining box: 1^5 assignments
- $\binom{3}{2} = \frac{3!}{2!\,(1!)} = \frac{3\times2\times1}{2\times1\times1} = 3$ pairs

0: impossible; must place 5 balls

iii) P(no box contains more than 3 balls)

$= P(\max \le 3) = P(3,1,1) + P(2,2,1) + P(3,2,0)$

where $P(3,2,0) = \frac{1}{3^5}\cdot 3!\binom{5}{3}\binom{2}{2}$

Since 3! : # distinct ways to arrange (3,2,0).
$\binom{5}{3}$: choose which of 3 balls go together
→ $\binom{2}{2} = 1$: which remaining two balls go into remaining 2 boxes

$P(3,1,1) = \frac{1}{3^5}\cdot 3\binom{5}{3}\cdot\binom{2}{1}$

Since 3: Choose which box gets the three
→ $\binom{5}{3}$: which 3 balls go into this box
$\binom{2}{1}$: assign 2 balls to distinct boxes

$P(2,2,1) = \frac{1}{3^5}\cdot 3\binom{5}{2}\cdot\binom{3}{2}$

Since 3: choose which box gets 2 balls
→ $\binom{5}{2}$: which two balls go into this box
$\binom{3}{2}$: assign remaining 3 balls other 2 boxes

5) Out of 3 prisoners (A, B, C), one was chosen uniformly at random to be pardoned. A asked the warden: "which among B and C will be executed?", to which the warden responded: "B is to be executed."

Here is the twist: given a correct answer to say either "B" or "C", we happen to know that the warden has the tendency to say "B" w/ prob. p (and "C" w/ prob. 1−p).

Let event W := warden tells A that B dies

Pardoned	Warden tells A who dies	Prob.
A	→ B dies	p
	→ C dies	1−p
B	→ C dies	1
C	→ B dies	1

i) Find P(W)
Let E_A, E_B, E_C be the respective pardon events s.t. $P(E_A) = P(E_B) = P(E_C) = 1/3$.

$P(W|E_A) = p \quad P(W|E_B) = 0 \quad P(W|E_C) = 1$

$$\Rightarrow P(W) \overset{\text{Law of tot. Prob.}}{=} \sum_{g\in\{A,B,C\}} P(W|E_g)P(E_g) = \frac{1}{3}\cdot p + \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot 1 = \frac{p+1}{3}.$$

ii) Find P(A pardoned | W). Comment on whether/not the warden's answer affects our updated knowledge about A's fate, depending on p

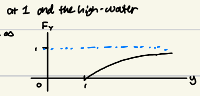
Posterior P(A pardoned | W)

$$= P(E_A|W) \overset{\text{Bayes}}{=} \frac{P(W|E_A)P(E_A)}{P(W)} = \frac{p\cdot 1/3}{(p+1)/3} = \frac{p}{p+1}$$

<u>Remark</u>: the warden's comment is informative unless p = 1/2:

- $p = \frac{1}{2} \Rightarrow P(E_A|W) = \frac{1}{2}\,\frac{2}{3} = \frac{1}{3}$ (no update)
- $p = 1 \Rightarrow P(E_A|W) = \frac{1}{2}$ (A more likely to be pardoned
- $p = 0 \Rightarrow P(E_A|W) = 0$ (A certainly not pardoned)

1.53) A certain river floods every year. Suppose that the low-water mark is set at 1 and the high-water mark Y has distribution function $F_Y(y) = P(Y \le y) = 1 - \frac{1}{y^2}, \quad 1 \le y < \infty$



a. Verify that $F_Y(y)$ is a cdf

i) $\lim_{y\to-\infty} F(y) = 0$ since $F_Y(y) = 0$ for $y < 1$

$\lim_{y\to\infty} F(y) = 1$ since $\lim_{y\to\infty} \frac{1}{y^2} = 0$.

ii) For $y \le 1$, $F_Y(y) = 0$ is constant. For $y > 1$, $\frac{d}{dy}F(y) = \frac{2}{y^3}$ so F_Y is increasing and therefore $\forall\, y$, F_Y is nondecreasing.

iii) F_Y is continuous and hence right continuous.

b. Find the pdf of Y, i.e. $f_Y(y)$

$$f_Y(y) = F'(y) = \frac{d}{dy}\left[1 - \frac{1}{y^2}\right] = \frac{d}{dy}(-y^{-2}) = (-)-2y^{-3} = \frac{2}{y^3}, \quad 1 \le y < \infty$$

$$\text{So, } f_Y(y) = \begin{cases} \frac{2}{y^3}, & 1 \le y < \infty \\ 0, & \text{otherwise} \end{cases}$$

c. If the low-water mark is reset at 0 and we use a unit of msmt that is 1/10 of that given previously, the high-water mark becomes $Z = 10(Y-1)$.

Find $F_Z(z) = P(Z \le z)$
$= P(10(Y-1) \le z)$
$= P(Y - 1 \le \frac{z}{10})$
$= P(Y \le \frac{z}{10} + 1)$
$= 1 - \frac{1}{(\frac{z}{10}+1)^2}.$

$$\text{Thus, } F_Z(z) = \begin{cases} 0 & \text{if } z \le 0 \\ 1 - \frac{1}{(\frac{z}{10}+1)^2} & \text{if } z > 0 \end{cases} \quad \text{since } \begin{cases} y = 1 \Rightarrow z = 10(1-1) = 0 \\ y = \infty \Rightarrow z = 10(\infty) = \infty \end{cases}$$

Practice Problems (up to Midterm 2)

1.6: Density & Mass Functions

<u>Ex(Geometric Prob)</u>: Let $X \sim Geo(p)$, $0<p<1$

s.t. $f_X(x) := P(X=x) := \begin{cases} (1-p)^{x-1}p, & x=1,2,\dots \\ 0, & \text{otherwise} \end{cases}$

where $P(a \le X \le b) = \sum_{k=a}^{b} f_X(k) = \sum_{k=a}^{b} (1-p)^{k-1}p$

and $P(X \le b) = \sum_{k=1}^{b} f_X(k) = F_X(b)$.

<u>Proof (Sketch)</u>. We must be careful in our definition of a pdf in the continuous case

If we naively try to calculate $P(X=x)$ for a continuous RV:

Let X - continuous RV.

Then, since $\{X=x\} \subset \{x-\varepsilon < X \le x\}$ for any $\varepsilon>0$,

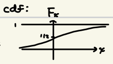
$P(X=x) \le P(x-\varepsilon < X \le x) = F_X(x) - F_X(x-\varepsilon)$ for any $\varepsilon>0$

Therefore,

$0 \le P(X=x) \le \lim_{\varepsilon \downarrow 0} [F_X(x) - F_X(x-\varepsilon)] = 0$ by continuity of F_X ▫

<u>Ex(Logistic Prob's)</u>: Recall the logistic cdf:

$F_X(x) = \frac{1}{1+e^{-x}}$



Then, $f_X(x) = \frac{d}{dx}F_X(x) = \frac{d}{dx}(1+e^{-x})^{-1} = \frac{e^{-x}}{(1+e^{-x})^{-2}}$

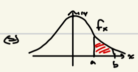
and symmetric since $f_X(x) = f_X(-x)$, i.e.

$f(-x) = \frac{e^{x}}{(1+e^{x})^2} = \frac{e^{x}e^{-2x}}{(1+e^{x})^2 e^{-2x}} = \frac{e^{-x}}{(1+2e^{x}+e^{2x})e^{-2x}} = \frac{e^{-x}}{e^{-2x}+2e^{-x}+1} = \frac{e^{-x}}{(1+e^{-x})^2} = f(x)$.

Then, $P(X \in (a,b)) = P(a<X<b) = F_X(b) - F_X(a)$

$= \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx$

$= \int_a^b f_X(x)dx = \int_a^b \frac{e^{-x}}{(1+e^{-x})^2}dx \iff$



Show that the following functions are cdf's:

a. $F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x)$, $x \in (-\infty,\infty)$

i) $\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x) = \frac{1}{2} + \frac{1}{\pi}\left(-\frac{\pi}{2}\right) = 0$.

$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x) = \frac{1}{2} + \frac{1}{\pi}\left(\frac{\pi}{2}\right) = 1$.

ii) $F'(x) = \frac{d}{dx}\left[\frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x)\right] = \frac{1}{\pi}\underbrace{\frac{1}{1+x^2}}_{>0}$ $\Rightarrow$ increasing

$\Rightarrow$ nondecreasing

iii) $F(x)$ is continuous and therefore right-continuous, i.e.

$\lim_{\varepsilon\downarrow 0} F_X(x+\varepsilon) = F_X(x)$. ∎

b. $F(x) = (1+e^{-x})^{-1}$, $x \in (-\infty,\infty)$

i) $\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} \frac{1}{1+e^{-x}} = 0$ since $\lim_{x\to-\infty} e^{-x} = \infty$

$\lim_{x\to\infty} F(x) = 1$ since $\lim_{x\to\infty} e^{-x} = 0$

ii) $F'(x) = \frac{d}{dx}[(1+e^{-x})^{-1}] = \frac{e^{-x}}{(1+e^{-x})^2} > 0$

$\Rightarrow$ increasing

$\Rightarrow$ nondecreasing

iii) $F(x)$ is (left/right) continuous and

$\lim_{\varepsilon\downarrow 0} F(x+\varepsilon) = F_X(x)$. ∎

c. $F(x) = e^{-e^{-x}}$, $x \in (-\infty,\infty)$

i) $\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} e^{-(e^{-x})} = \lim_{x\to-\infty} \frac{1}{e^{e^{-x}}} = 0$ since $\lim_{x\to-\infty} e^{-x} = \infty$

$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} \frac{1}{e^{e^{-x}}} = 1$ since $\lim_{x\to\infty} e^{-x} = 0$

ii) $F'(x) = \frac{d}{dx}\left[e^{-e^{-x}}\right] = \underbrace{e^{-e^{-x}-x}}_{>0}$ since $\begin{cases} g(f(x)) = e^{-e^{-x}} \\ g(x) = e^{x} \\ f(x) = -e^{-x} \\ \frac{d}{dx}(g(f(x))) = g'(f(x))f'(x) \\ \quad = e^{-e^{-x}}e^{x} \end{cases}$

iii) $F(x)$ is continuous and therefore right-continuous, i.e.

$\lim_{\varepsilon\downarrow 0} F_X(x+\varepsilon) = F_X(x)$. ∎

d. $F(x) = 1-e^{-x}$, $x \in (0,\infty)$

i) $\lim_{x\to-\infty} F(x) = 0$ since $F_X(x) = 0$ for $x \le 0$

$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} 1-e^{-x} = 1$ since $\lim_{x\to\infty} e^{-x} = 0$.

ii) $F'(x) = \frac{d}{dx}[1-e^{-x}] = -\frac{d}{dx}e^{-x} = e^{-x} > 0$

$\Rightarrow$ increasing

$\Rightarrow$ nondecreasing

iii) $F(x)$ is continuous and therefore right-continuous, i.e.

$\lim_{\varepsilon\downarrow 0} F_X(x+\varepsilon) = F_X(x)$. ∎

e. For some $\varepsilon \in (0,1)$: $F_Y(y) = \begin{cases} \frac{1-\varepsilon}{1+e^{-y}}, & y<0, \\ \varepsilon + \frac{(1-\varepsilon)}{1+e^{-y}}, & y \ge 0. \end{cases}$

i) $\lim_{y\to-\infty} F(y) = \lim_{y\to-\infty} \frac{1-\varepsilon}{1+e^{-y}} = 0$ since $\lim_{y\to-\infty} e^{-y} = \infty$ and $1-\varepsilon \in (0,1)$

$\lim_{y\to\infty} F(y) = \lim_{y\to\infty} \varepsilon + \frac{(1-\varepsilon)}{1+e^{-y}} = \varepsilon + (1-\varepsilon) = 1$ since $\lim_{y\to\infty} e^{-y} = 0$.

ii) when $y<0$, $F'(y) = (1-\varepsilon)\frac{d}{dy}(1+e^{-y})^{-1} = \underbrace{\underbrace{(1-\varepsilon)}_{>0}\frac{e^{-y}}{\underbrace{(1+e^{-y})^2}_{>0}}}_{>0}$

when $y \ge 0$, $F'(y) = (1-\varepsilon)\frac{e^{-y}}{(1+e^{-y})^2} > 0$,

and $\lim_{y\to0^+} F(y) = \lim_{y\to0^+}\left[\varepsilon + \frac{(1-\varepsilon)}{1+e^{-y}}\right] = \varepsilon + \frac{(1-\varepsilon)}{2} > \frac{1-\varepsilon}{2} = \lim_{y\to0^-}\left[\frac{1-\varepsilon}{1+e^{-y}}\right] = \lim_{y\to0^-} F(y)$.

Therefore, $F(y)$ is nondecreasing.

iii) For $y<0$, $\lim_{\varepsilon\downarrow 0} F_Y(y+\varepsilon) = F_Y(y)$

$y \ge 0$, $\lim_{\varepsilon\downarrow 0} F_Y(y+\varepsilon) = F_Y(y)$,

i.e. right-continuity holds for the piecewise function $F(y)$. ∎

2.1: Distributions of Functions of RV's

<u>Ex(Binomial transformation)</u>: Let $X \sim Binomial(n,p)$ with

pmf: $f_X(x) := P(X=x) = \binom{n}{x}p^x(1-p)^{n-x}$, $x=0,1,\dots,n$

Now let $Y := g(X) = n-x$

s.t. $Y=n-x$ is also a RV.

Q: What is the (prob.) dist of Y?

Notice that (the support) $Y \in \{0,1,\dots,n\}$

$\Rightarrow$ For $y \in \{0,1,\dots,n\}$, $f_Y(y) = P(Y=y) = \sum_{x \in g^{-1}(y)} f_X(x) = f_X(n-y) = P(X=n-y)$, $Y=n-X \iff X=n-Y$

$= f_X(n-y) = \binom{n}{n-y}p^{n-y}(1-p)^y$

$= \binom{n}{y}(1-p)^y p^{n-y}$ since $\binom{n}{y} = \frac{n!}{y!(n-y)!} = \frac{n!}{(n-y)!\,y!} = \frac{n!}{(n-y)![n-(n-y)]!} = \binom{n}{n-y}$

$\Rightarrow Y \sim Binomial(n, 1-p)$.

<u>Proof (Discrete Transformation)</u>. Let X - discrete ($\mathcal{X}$: countable). Suppose $Y=g(X)$ s.t. $\mathcal{Y} = \{y : y=g(x), x \in \mathcal{X}\}$ and countable.

Then, Y is also a discrete RV:

$\begin{cases} f_Y(y) = P(Y=y) = \sum_{x \in g^{-1}(y)} P(X=x) = \sum_{x \in g^{-1}(y)} f_X(x) \text{ for } y \in \mathcal{Y} \\ \text{and } f_Y(y) = 0 \text{ for } y \notin \mathcal{Y}. \end{cases}$

Finding $f_Y(y)$ involves simply identifying $g^{-1}(y)$ for each $y \in \mathcal{Y}$, and summing the appropriate prob.'s

<u>Proof (Monotonic Transformation Functions)</u>. Let X have cdf $F_X(x)$ where

$\mathcal{X} = \{x : f(x)>0\}$ (support of X and/or f_X) and $\mathcal{Y} = g(\mathcal{X}) := \{y : y=f(x) \text{ for some } x \in \mathcal{X}\}$.

Show that for $Y=g(X) \iff g : \mathcal{X} \to \mathcal{Y}$:

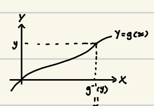
a) If $g\uparrow$ (strictly increasing), then $F_Y(y) = F_X(g^{-1}(y))$ $\forall y \in \mathcal{Y}$

and b) If $g\downarrow$ (strictly decreasing), then $F_Y(y) = 1 - F_X(g^{-1}(y))$ $\forall y \in \mathcal{Y}$:

<u>Proof (a)</u>. If $g\uparrow$ (strictly increasing), where $Y=g(X)$, then

$\{x : g(x) \le y\} = \{x \le g^{-1}(y)\}$

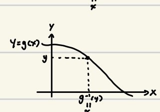
$\Rightarrow F_Y(y) = \int_{\{x \le g^{-1}(y)\}} f_X(x)dx = F_X(g^{-1}(y))$. ▫



<u>Proof (b)</u>. If $g\downarrow$ (strictly decreasing), where $Y=g(X)$, then

$\{x : g(x) \le y\} = \{x \ge g^{-1}(y)\}$

$\Rightarrow F_Y(y) = \int_{\{x \ge g^{-1}(y)\}} f_X(x)dx = 1 - F_X(g^{-1}(y))$. ▫



<u>Ex (Uniform & Exponential)</u>: Let $X \sim Uniform(0,1)$, $f(x) = \begin{cases} 1, & x \in (0,1) \\ 0, & \text{otherwise} \end{cases}$.

Suppose $Y = g(X) := -\log(X)$ for $\begin{cases} \mathcal{X} = (0,1) \\ \mathcal{Y} = (0,+\infty) \end{cases}$

To find the dist. of Y, first solve x for y, i.e.

$y = -\log(x) \Rightarrow \log(x) = -y \Rightarrow x = e^{-y} = g^{-1}(y)$.

$g\downarrow$ (strictly decreasing) since $g'(x) = -\frac{1}{x} < 0$ for $x \in (0,1)$.

For $x \in (0,1)$, $F_X(x) = \int_0^x f_X(x)dx = \int_0^x 1\,dx = x$.

Thus, for $y>0$

$F_Y(y) = 1 - F_X(g^{-1}(y)) = 1 - e^{-y}$.

Hence

$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}(1-e^{-y}) = (-1)(-e^{-y}) = e^{y}$,

so

$f_Y(y) = \begin{cases} e^{y}, & y>0 \\ 0, & y \le 0 \end{cases}$ $\Rightarrow Y \sim Exp(1)$ where <u>recall</u>: Exponential$(\theta) \sim \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)$; $x \ge 0, \theta>0$

<u>Ex(Inverse Gamma)</u>: Let $X \sim Gamma(n,\beta)$ where n,β: shape and scale parameters respectively

Suppose $Y = \frac{1}{X}$, i.e. the inverse gamma RV where

$f_X(x) = \frac{1}{(n-1)!\beta^n}x^{n-1}e^{-x/\beta}$ for $x>0$ $(0,\infty)$

Given $Y = g(X) = \frac{1}{X} \Rightarrow x = g^{-1}(y) = \frac{1}{y}$ for $y \in (0,\infty) = \mathcal{Y}$.

$g\downarrow$ on $\mathcal{X} = (0,\infty)$ and $\frac{d}{dy}[g^{-1}(y)] = \frac{d}{dy}(y^{-1}) = \frac{-1}{y^2}$

2.2) In each of the following find the pdf of Y:

a. $Y = X^2$ and $f_X(x) = 1$, $0<x<1$

$Y = g(X)$ where $g(x) = x^2$ for $\begin{cases} \mathcal{X} = (0,1) \\ \mathcal{Y} = (0,1) \end{cases}$

$F_Y(y) = P(Y \le y)$, $y \in (0,1)$

$= P(X^2 \le y)$

$= P(|X| \le \sqrt{y})$, $x \in (0,1)$

$= P(X \le \sqrt{y})$

$= F_X(\sqrt{y})$

Given $f_X(x) = 1$, $x \in (0,1)$

then $F_X(x) = \int_0^x f_X(t)\,dt = \int_0^x dt = t\Big|_0^x = x$

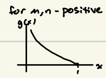
$\Rightarrow F_X(\sqrt{y}) = \sqrt{y}$

$= \sqrt{y}$

$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}y^{\frac{1}{2}} = \frac{1}{2}y^{-\frac{1}{2}} = \boxed{\underbrace{\frac{1}{2\sqrt{y}},\ y \in (0,1)}_{f_Y(y)}}$.

b. $Y = -\log(X)$ and $f_X(x) = \frac{(n+m+1)!}{n!m!}x^n(1-x)^m$, $0<x<1$ for m,n - positive integers

$Y = g(X)$ where $g(x) = -\log(x)$ for $\begin{cases} \mathcal{X} = (0,1) \\ \mathcal{Y} = (0,\infty) \end{cases}$



$y = -\log(x) \Rightarrow \log(x) = -y \Rightarrow x = e^{-y} = g^{-1}(y)$ $\forall y>0$.

$\frac{d}{dy}[g^{-1}(y)] = \frac{d}{dy}(e^{-y}) = -e^{-y}$

and $g\downarrow$ on $\mathcal{X} = (0,1)$.

So,

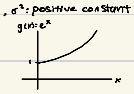
$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$

$= \frac{(n+m+1)!}{n!m!}(e^{-y})^n(1-e^{-y})^m e^{-y}$

$= \boxed{\underbrace{\frac{(n+m+1)!}{n!m!}e^{-y(n+1)}(1-e^{-y})^m,\ y \in (0,\infty)}_{f_Y(y)}}$

c. $Y = e^X$ and $f_X(x) = \frac{1}{\sigma^2}x\exp\left(\frac{-(x/\sigma)^2}{2}\right)$, $0<x<\infty$, σ^2: positive constant

$Y = g(X)$ where $g(x) = e^x$ for $\begin{cases} \mathcal{X} = (0,\infty) \\ \mathcal{Y} = (1,\infty) \end{cases}$



$y = e^x \Rightarrow x = \log(y) = g^{-1}(y)$ $\forall y>1$

$\frac{d}{dy}[g^{-1}(y)] = \frac{d}{dy}[\log(y)] = \frac{1}{y}$

and $g\uparrow$ on $\mathcal{X} = (0,\infty)$.

So,

$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$

$= \frac{1}{\sigma^2}\log(y)\exp\left\{-\frac{(\log(y)/\sigma)^2}{2}\right\}\cdot\frac{1}{y}$

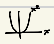
$= \frac{1}{\sigma^2}\frac{\log(y)}{y}\exp\left\{\frac{-1}{2}\left(\frac{\log(y)}{\sigma}\right)^2\right\}$, $y \in (1,\infty)$

we know

$$f_Y(y) = f_X(g^{-1}(y))\left|\tfrac{d}{dy}g^{-1}(y)\right|$$
$$= \frac{1}{(n-1)!\,\beta^n}(1/y)^{n-1}e^{-1/(\beta y)}\frac{1}{y^2}$$
$$= \frac{1}{(n-1)!\,\beta^n}y^{-(n+1)}e^{-1/(\beta y)}$$

<u>Ex (Square Transformation)</u>: Suppose $X \sim Normal(0,1)$ and let $Y = g(X) = X^2 \sim \chi^2_1$.

Here, $g(x) = x^2$ for $x \in \mathbb{R}$ is not a monotonic function



Since $\mathcal{X} = \mathbb{R}$, $\mathcal{Y} = \mathbb{R}_+ \cup \{0\}$.

Thus,

$\forall y \geq 0$, $F_Y(y) = P(Y \leq y) = P(X^2 \leq y)$

$|x| \leq y = \begin{cases} x \leq y \\ x \geq -y \end{cases}$ $\quad -y \quad y$

$|x| \geq y = \begin{cases} x \geq y \\ x \leq -y \end{cases}$ $\quad -y \quad y$

$= P(|X| \leq \sqrt{y})$, $x \in \mathbb{R}$

$= P(X \leq \sqrt{y}) - P(X \leq -\sqrt{y})$ $\quad -\sqrt{y} \quad \sqrt{y}$

$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$

So,

$$f_Y(y) = \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}\left[F_X(y^{1/2}) - F_X(-y^{1/2})\right] = \frac{1}{2\sqrt{y}}f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}}f_X(-\sqrt{y})$$
$$= \frac{1}{2\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-y/2} + \frac{1}{2\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-y/2}$$

since $X \sim N(0,1) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $x \in \mathbb{R}$ and $\chi^2_{df} \overset{d}{\sim} \sum_{i=1}^{df} z_i^2$ for $z_i \overset{iid}{\sim} N(0,1)$.

$$= \frac{1}{\sqrt{2\pi y}}e^{-y/2} \sim \chi^2_1$$
$$\propto y^{\frac{1}{2}-1}e^{-y/2} = Gamma\left(\tfrac{1}{2}, 2\right)$$

<u>Proof (Sketch)</u>. I will prove the following thereom:

Let X - continuous RV w/ cdf $F_X(x)$ and $Y = g(x) = F_X(x)$.
Then, $Y \sim Uniform(0,1)$

<u>Proof.</u> For $y \in (0,1)$,

$P(Y \leq y) = P(F_X(x) \leq y)$

assuming $F_X^{-1}\uparrow$ (i.e. strictly increasing)

$= P(F_X^{-1}(F_X(x)) \leq F_X^{-1}(y))$

$= P(X \leq F_X^{-1}(y))$ (*) requires further discussion

$= F_X(F_X^{-1}(y))$

assuming F_X is continuous

$= y$.

It is easy to check: $P(Y \leq 1) = 1$ and $P(Y \leq 0) = 0$.
Hence, cdf of Y is that of a unif(0,1) RV. $\square$

(*):

- If F_X is strictly increasing at x, for $x \in \mathcal{X}$
  then let $y = F_X(x) \Rightarrow x = F_X^{-1}(y)$
  so $F_X^{-1}(F_X(x)) = F_X^{-1}(y) = x$.
- If F_X is "flat" at x, i.e. $\forall x \in$ some $[x_1, x_2]$
  s.t. $F_X(x') = F_X(x_1)$ $\forall x_1 \in [x_1, x]$
  and $F_X(x) < F_X(x_1)$ $\forall x < x_1$

then
$F_X^{-1}(F_X(x)) = x_1 < \leq x)$

since
$P(X \in [x_1, x]) = 0$
$P(X \in (F_X^{-1}(F_X(x)), x]) = 0$ (**)

So in this scenario
$\forall a \in \mathbb{R}$

$P(F^{-1}(F_X(X)) \leq a \mid F_X$ is flat at $X)$
$= P(F_X^{-1}(F_X(X)) \leq a, X \leq a \mid \cdots) +$
$P(F_X^{-1}(F_X(X)) \leq a, X > a \mid \cdots)$
$=: C + D$, where

$C = P(X \leq a \mid F_X$ is flat at $X)$
b/c $X \leq a \Rightarrow F_X^{-1}(F_X(X)) \leq X \leq a$

and $D \leq P(F_X^{-1}(F_X(X)) < X \mid F_X$ is flat at $X)$
$= P(X \in (F_X^{-1}(F_X(X)), X] \mid F_X$ is flat at $X)$
with prob. 1
$= 0$ due to (**)

combining these two scenarios:

$\forall a$, $P(F_X^{-1}(F_X(X)) \leq a \mid F_X$ flat at $X) = P(X \leq a \mid F_X$ flat at $X)$

$P(F_X^{-1}(F_X(X)) \leq a \mid F_X \uparrow$ at $X) \underset{\text{w/ prob 1}}{=} P(X \leq a \mid F_X \uparrow$ at $X)$

to qualify (*). $\square$

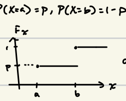
<u>Proof</u>. I will prove the following thereom:

Let $F_X(x)$ be the cdf of a RV and $Y \sim Uniform(0,1)$.
Set $Z = F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}$ $\forall y \in (0,1)$.
Then Z has cdf F_X, i.e. $F_X^{-1}(y) = F_X(x)$

<u>Proof.</u> Suppose a simple case s.t. $X \in \{a, b\}$ where

$P(X=a) = p$, $P(X=b) = 1-p$, $a < b$, and $Y \sim$ Uniform.

Then,



and $F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}$ $\forall y \in (0,1)$
$$= \begin{cases} -\infty, & y \leq 0 \\ a, & 0 < y \leq p \\ b, & p < y \leq 1 \end{cases}$$

See that

$P(F_X^{-1}(y) = a) = P(0 < y \leq p) = F_Y(p) = p$ since $Y \sim Unif(0,1)$

Similarly,

$P(F_X^{-1}(y) = b) = P(p < y \leq 1) = 1-p$,
$P(F_X^{-1}(y) = -\infty) = P(Y = 0) = 0$.

Therefore,

$F_X^{-1}(y) = F_X(x) \Rightarrow F_X^{-1}(y) \overset{d}{=} X$. $\blacksquare$

2.4) Let λ - fixed, positive constant and $f_X(x) = \begin{cases} \frac{1}{2}\lambda e^{-\lambda x}, & x \geq 0 \\ \frac{1}{2}\lambda e^{\lambda x}, & x < 0 \end{cases}$

a. verify that $f(x)$ is a pdf

$f(x)$ is a pdf since it is positive and

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0}\frac{\lambda}{2}e^{\lambda x}dx + \int_0^{\infty}\frac{\lambda}{2}e^{-\lambda x}dx = \frac{\lambda}{2}\left[\frac{1}{\lambda}e^{\lambda x}\right]\Big|_{-\infty}^{0} + \frac{\lambda}{2}\left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^{\infty} = \frac{1}{2} + \frac{1}{2} = 1$$

b. If X is a RV w/ pdf given by $f(x)$, Find $P(X < t)$ for all t. Evaluate all integrals.
Let X be a RV with density $f(x)$.

$$P(X < t) = \begin{cases} \int_{-\infty}^{t}\frac{1}{2}\lambda e^{\lambda x}dx, & t < 0 \\ \int_{-\infty}^{0}\frac{1}{2}\lambda e^{\lambda x}dx + \int_0^t \frac{1}{2}\lambda e^{-\lambda x}dx, & t \geq 0 \end{cases}$$

where $\int_{-\infty}^{t}\frac{1}{2}\lambda e^{\lambda x}dx = \frac{1}{2}e^{\lambda x}\Big|_{-\infty}^{t} = \frac{1}{2}e^{\lambda t}$

and $\int_0^t \frac{1}{2}\lambda e^{-\lambda x}dx = -\frac{1}{2}e^{-\lambda x}\Big|_0^t = \frac{-1}{2}e^{-\lambda t} + \frac{1}{2}$.

Therefore,

$$P(X < t) = \begin{cases} \frac{1}{2}e^{\lambda t}, & t < 0 \\ 1 - \frac{1}{2}e^{-\lambda t}, & t \geq 0 \end{cases}$$

c. Find $P(|X| < t)$ for all t. Evaluate all integrals

$$P(|X| < t) = P(-t < X < t) = \int_{-t}^{0}\frac{1}{2}\lambda e^{\lambda x}dx + \int_0^t \frac{1}{2}\lambda e^{-\lambda x}dx$$
$$= -\int_0^{-t}\frac{1}{2}\lambda e^{\lambda x}dx + \int_0^t \frac{1}{2}\lambda e^{-\lambda x}dx$$
$$= \frac{\lambda}{2}\left[\frac{1}{\lambda}e^{\lambda x}\right]\Big|_{-t}^{0} - \frac{1}{2}e^{-\lambda t} + \frac{1}{2}$$
$$= \frac{1}{2}[1 - e^{-\lambda t}] + \frac{1}{2}[-e^{-\lambda t} + 1]$$
$$= 1 - e^{-\lambda t}$$

a. Let X, Y, Z - continuous RV's w/ the following pdf's:

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \sim N(0,1),$$
$$f_Y(y) = \frac{1}{\sqrt{8\pi}}e^{-y^2/8} = \frac{1}{\sqrt{8\pi}}\exp\left(-\frac{y^2}{8}\right) \sim N(0,4),$$
$$f_Z(z) = \frac{1}{\sqrt{8\pi}}e^{-(z-1)^2/8} = \frac{1}{\sqrt{8\pi}}\exp\left(-\frac{(z-1)^2}{8}\right) \sim N(1,4)$$

*FS Since recall: $Normal(m, \sigma^2)$ has the following pdf: $\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$, $\sigma > 0$.

Let $X \sim N(0,1)$ with previously defined $f_X(x)$

I will show that $X \overset{d}{=} \frac{Y}{2} \iff 2X \overset{d}{=} Y$.

Given $f_X(x)$, $X = \frac{Y}{2} \Rightarrow Y = g(x) = 2X$ for $\begin{cases} x \in \mathbb{R} \\ y \in \mathbb{R} \end{cases}$

Then, $y = 2x \Rightarrow x = \frac{y}{2} = g^{-1}(y)$,

$\frac{d}{dy}[g^{-1}(y)] = \frac{1}{2}y' = \frac{1}{2}$,

and $g(x) = 2x \uparrow$.

So,

$$f_Y(y) = f_X(g^{-1}(y))\left|\tfrac{d}{dy}g^{-1}(y)\right| = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{(-y/2)^2}{2}\right)\frac{1}{2} = \frac{1}{\sqrt{8\pi}}\exp\left(\frac{-y^2}{8}\right)$$

Similarly, to show $X \overset{d}{=} \left(\frac{Z-1}{2}\right) \iff Z \overset{d}{=} 2X+1$,

$Z = g_1(X) = 2X+1$ for $\begin{cases} x \in \mathbb{R} \\ z \in \mathbb{R} \end{cases}$.

Next,

$z = 2x+1 \Rightarrow x = \frac{z-1}{2} = g_1^{-1}(z)$

$\frac{d}{dz}[g_1^{-1}(z)] = \frac{1}{2}\frac{d}{dy}(z-1) = \frac{1}{2}$ and $g_1(x) = 2x+1\uparrow$.

Thus,

$$f(z) = f_X(g_1^{-1}(z))\left|\tfrac{d}{dz}g_1^{-1}(z)\right| = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\left(\frac{z-1}{2}\right)^2}{2}\right)\frac{1}{2} = \frac{1}{\sqrt{8\pi}}\exp\left(-\frac{(z-1)^2}{8}\right)$$

The same process follows to show $X \overset{d}{=} \frac{(1-Z)}{2} \iff Z \overset{d}{=} 1-2X$:

$g_2(x) = z = 1-2x$ for $x \in \mathbb{R}$, $z \in \mathbb{R}$

where

$z = 1-2x \Rightarrow x = \frac{1-z}{2} = g_2^{-1}(z)$, $\frac{d}{dz}[g_2^{-1}(z)] = \frac{-1}{2}$, and $g_2\downarrow$.

So,

$$f(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\left(\frac{1-z}{2}\right)^2}{2}\right)\frac{1}{2} = \frac{1}{\sqrt{8\pi}}\exp\left(\frac{-(z-1)^2}{2}\right).$$

Consequently, I have shown that $X \overset{d}{=} \frac{Y}{2} \overset{d}{=} \frac{(Z-1)}{2} \overset{d}{=} \frac{(1-Z)}{2} \sim f_X(x)$.

b. We know $X \sim N(0,1)$

So, $P(X > 0) = \int_0^{\infty} f_X(x)dx = \frac{1}{2} = \int_{-\infty}^{0} f_X(x)dx$ by symmetry of $f_X(x) \sim N(0,1)$.

Similarly, for $Y \sim N(0,4)$ and $Z \sim N(1,4)$,

$P(X > 0) = P(Y \geq 0) = P(Z \leq 1) = \frac{1}{2}$ by symmetry of the continuous normal pdf and hence equal prob. on either side of the mean.

iii) Let $U \sim \chi^2_1$. I will show that $P(U \leq 1) < P(X \leq 1)$.

We know that $\chi^2_1 = X^2$ for $X \sim N(0,1)$

So, $P(U \leq 1) = P(\chi^2_1 \leq 1)$
$= P(X^2 \leq 1)$
$= P(|X| \leq 1) = P(-1 \leq X \leq 1) = \Phi(1) - \Phi(-1)$

Because $\Phi(-1) > 0$, $\Phi(1) - \Phi(-1) < \Phi(1) = P(X \leq 1)$.

3. This question asks you to prove a theorem in the lecture notes (Theorem 2 in "Probability integral transform" of Section 2.1). Let X be a real-valued random variable with cdf $F_X(x)$. Recall that the inverse function for the (right-continuous) F_X can be defined as follows, for $0 < y < 1$,

$$F_X^{-1}(y) := \inf\{x : F_X(x) \geq y\}.$$

Moreover, in the above definition if a set is empty then its infimum is defined to be $+\infty$ as a convention. We also define $F_X^{-1}(0) := -\infty$. Let U be a uniform random variable in $(0,1)$, and $Z := F_X^{-1}(U)$. Show that Z has the same distribution as that of X in the following two scenarios:

(i) X is a discrete random variable taking values in a finite set $\mathcal{X} = \{a_1, \ldots, a_k\} \subset \mathbb{R}$, for some $k \in \mathbb{N}$.
(ii) X is a continuous random variable.

Let X have cdf F. Define the generalized inverse

$F^{-1}(y) := \inf\{x : F(x) \geq y\}$, $0 < y < 1$,

with $F^{-1}(0) := -\infty$. Let $U \sim Unif(0,1)$ and $Z := F^{-1}(U)$. Show that $Z \overset{d}{=} X$ in:

a) <u>Discrete finite support</u>. Suppose X takes values $a_1 < \cdots < a_k$ w/ prob's $p_i = P(X = a_i) > 0$.

Let $s_i = \sum_{j=1}^{i} p_j$ with $s_0 := 0$. Then $F^{-1}(u) = a_i$ i.f.f. $s_{i-1} < u \leq s_i$.

Since $U \sim uniform(0,1)$,

$P(Z = a_i) = P(s_{i-1} < U \leq s_i) = s_i - s_{i-1} = p_i = P(X = a_i)$,

so Z and X have the same pmf.

b) <u>Continuous</u>. For any $x \in \mathbb{R}$,

$\{F^{-1}(U) \leq x\} = \{U \leq F(x)\}$

since $F^{-1}(u) \leq x$ i.f.f. there exists $x' \leq x$ with $F(x') \geq u$; by monotonicity/right-continuity of F, this is equivalent to $u \leq F(x)$. Hence,

$P(Z \leq x) = P(U \leq F(x)) = F(x)$, i.e.

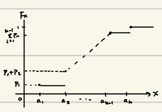
Z has cdf F and therefore $Z \overset{d}{=} X$.

$F_X^{-1}(y) := \inf\{x : F_X(x) \geq y\}$ $\forall y \in (0,1)$
$$= \begin{cases} -\infty, & y \leq 0 \\ a, & 0 < y \leq p \\ b, & p < y \leq 1 \end{cases}$$

i) Suppose $P(X = a_n) = p_n$ $\forall n \in \{1, \ldots, k\}$

where $F_X^{-1}(u) = \inf\{x : F_X(x) \geq u\}$ $\forall u \in (0,1)$

$$F_X^{-1}(u) = \begin{cases} -\infty, & u = 0 \\ a_1, & 0 < u \leq p_1 \\ a_2, & p_1 < u \leq p_1 + p_2 \\ \vdots \\ a_{k-1}, & \sum_{i=1}^{k-2} p_n < u \leq \sum_{i=1}^{k-1} p_n \\ a_k, & \sum_{i=1}^{k-1} p_n < u \leq 1 \end{cases} \iff$$



$$= \begin{cases} -\infty, & u = 0 \\ a_n, & \sum_{i=1}^{n-1} p_n < u \leq \sum_{i=1}^{n} p_n \quad \forall n \in \{1, \ldots, k\} \end{cases}$$

So, $P(F_X^{-1}(U) = a_n) = P\left(\sum_{i=1}^{n-1} p_n < U \leq \sum_{i=1}^{n} p_n\right)$
$= \sum_{i=1}^{n} p_n - \sum_{i=1}^{n-1} p_n$ since $U \sim unif(0,1)$
$= p_n$
$= P(X = a_n)$. Therefore, we have shown that
$Z := F_X^{-1}(U) \overset{d}{=} X$. $\blacksquare$

ii) Let X - continuous RV

$P(Z \leq z) = P(F_X^{-1}(U) \leq z)$ (*)
$= P(F_X(F_X^{-1}(U)) \leq F_X(z))$
$= P(U \leq F_X(z))$ (★)

where (*) $\Rightarrow$ (★) since F_X is monotonic

and (★) $\Rightarrow$ (*): Given $u \leq F(z)$, need to show that $\inf\{x : F(x) \geq u\} \leq z$

since $u \leq F(z)$, $z \in \{x : F(x) \geq u\}$

so $\inf\{x : F(x) \geq u\} \leq z$.

$= F_X(z)$

$= P(X \leq z)$

2.2: Expectation

Ex (Exponential Expectation). Let $X \sim \text{Exp}(\lambda) = \frac{1}{\lambda}e^{-x/\lambda}$, $x \in \mathcal{X} = [0, +\infty)$, $\lambda > 0$.

We can show that $\int_{\mathcal{X}} f_X(x)\,dx = \int_0^\infty \frac{1}{\lambda}e^{-x/\lambda}dx = \frac{1}{\lambda}[e^{-x/\lambda}(-\lambda)]\Big|_0^\infty = -e^{-x/\lambda}\Big|_0^\infty = 0-(-1) = 1.$

$E(X) = \int_0^\infty x\left(\frac{1}{\lambda}e^{-x/\lambda}\right)dx$

↳ IBP: $\int u\,dv = uv - \int v\,du$

let $u = x$, $dv = \frac{1}{\lambda}e^{-x/\lambda}dx$

$du = 1$, $v = \int \frac{1}{\lambda}e^{-x/\lambda}dx = e^{-x/\lambda}\frac{1}{\lambda}(-\lambda) = -e^{-x/\lambda}$

$= -xe^{-x/\lambda}\Big|_0^\infty + \int_0^\infty e^{-x/\lambda}dx$

$= 0 + e^{-x/\lambda}(-\lambda)\Big|_0^\infty$

$= 0 - (-\lambda) = \lambda.$

Ex (Binomial Expectation): Let $X \sim \text{Binomial}(n,p)$

where $f_X(x) = P(X=x) = \binom{n}{x}p^x(1-p)^{n-x}$, $x \in \mathcal{X} = \{0,\dots,n\}$, $n \in \mathbb{N}$, $p \in (0,1)$.

First check that

$\sum_{x\in\mathcal{X}} f_X(x) = \sum_{x=0}^{n}\binom{n}{x}p^x(1-p)^{n-x} = (p+(1-p))^n$ by the Binomial identity: $(x+a)^n = \sum_{k=0}^{n}\binom{n}{k}x^k a^{n-k}$

$= 1.$

Then,

$E(X) = \sum_{x=0}^{n} x\binom{n}{x}p^x(1-p)^{n-x}$

$= \sum_{x=1}^{n} x\binom{n}{x}p^x(1-p)^{n-x}$, $x=0 \Rightarrow$ entire product $=0$

$= \sum_{x=1}^{n} n\binom{n-1}{x-1}p^x(1-p)^{n-x}$, $\quad x\binom{n}{x} = x\cdot\frac{n!}{x!(n-x)!} = \frac{n!}{(x-1)!(n-x)!} = n\cdot\frac{(n-1)!}{(x-1)![(n-1)-(x-1)]!} = n\binom{n-1}{x-1}.$

$= np\sum_{x=1}^{n}\binom{n-1}{x-1}p^{x-1}(1-p)^{n-x}$

letting $y = x-1 \Rightarrow -x = -1-y \Rightarrow x = y+1$

$= np\sum_{y=0}^{n-1}\binom{n-1}{y}p^y(1-p)^{n-1-y}$

$= np(p+\cancel{(1-p)})^{n-1}$

$= np.$

Ex (Cauchy Expectation Nuances): Let $X \sim \text{Cauchy}$; $f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$, $x \in \mathcal{X} = \mathbb{R}$.

check $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty}\frac{1}{\pi}\frac{1}{1+x^2}dx = 1$:

letting $x = \tan(\theta)$, $\frac{dx}{d\theta} = \frac{d}{d\theta}(\tan(\theta))$ (change-of-var) $\Big]\ \frac{d}{d\theta}\tan\theta = \sec^2\theta$

$\Rightarrow dx = \sec^2(\theta)\,d\theta$

For $x = \infty \Rightarrow \theta = \tan^{-1}(\infty) = \pi/2$

$x = -\infty \Rightarrow \theta = \tan^{-1}(-\infty) = -\pi/2$

$= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\frac{1}{1+\tan^2\theta}\sec^2\theta\,d\theta$, $\quad \frac{1}{1+\tan^2\theta} = \sec^2\theta$

$= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}d\theta = \frac{1}{\pi}\left[\frac{\pi}{2}+\frac{\pi}{2}\right] = 1.$

$E|X| = \int_{-\infty}^{\infty}\frac{1}{\pi}\frac{|x|}{1+x^2}dx = \frac{2}{\pi}\int_0^\infty\frac{x}{1+x^2}dx$ by symmetric Cauchy RV

For any $M > 0$:

$\lim_{m\to\infty}\frac{2}{\pi}\int_0^m\frac{x}{1+x^2}dx$

let $u = 1+x^2$

$du = 2x\,dx \Rightarrow \frac{1}{2}du = x\,dx$

$= \lim_{m\to\infty}\frac{2}{\pi}\frac{1}{2}\int_0^m\frac{1}{u}du = \lim_{m\to\infty}\frac{1}{\pi}[\log(1+x^2)]\Big|_0^m$

$= \lim_{m\to\infty}\frac{1}{\pi}\log(1+m^2) = \infty$

$E(X) := \int_{-\infty}^{\infty}x f_X(x)\,dx = \int_{-\infty}^{0}x f_X(x)\,dx + \int_0^\infty x f_X(x)\,dx$

$= \lim_{m\to\infty}\int_{-m}^{0}\cdot + \lim_{n\to\infty}\int_0^n\cdot = \lim_{\substack{m\to+\infty\\ n\to+\infty}}\frac{-1}{2\pi}\log(1+m^2) + \frac{1}{2\pi}\log(1+n^2)$ DNE (undefined)

Remarks:
- a proper expectation exists only when the positive and negative part of integral are finite $\left(\int|x|f_X(x)\,dx < \infty\right)$
- $E(X)$ is undefined since $\int x f_X(x)\,dx$ is not absolutely convergent

Proof (Sketch). For X, Y - real valued RV's w/ expectations that exist,

Show that $E(aX+bY+c) = aE(X)+bE(Y)+c$:

Assume X and Y may be expressed as functions of a RV Z, i.e.

$(*) := \begin{cases} X = g_1(Z) \\ Y = g_2(Z)\end{cases}$

Suppose Z is a continuous RV w/ f_Z. Then

$E(aX+bY+c) = E[ag_1(Z)+bg_2(Z)+c]$

$= \int(ag_1(z)+bg_2(z)+c)f(z)\,dz$

$= a\int g_1(z)f(z)\,dz + b\int g_2(z)f(z)\,dz + c\int f(z)\,dz$

$= aE(g_1(Z)) + bE(g_2(Z)) + c = aE(X)+bE(Y)+c.$

The proof proceeds similarly if Z is discrete; incomplete since we assume $(*)$ w/o formality. □

Ex (Linearity of Expectation): Suppose $X \sim \text{Binomial}(n,p)$. Let $Y_i = \begin{cases} 1, & H \text{ at } i^{th} \text{ toss w/ prob } p \\ 0, & \text{otherwise w/ prob } 1-p\end{cases}$; Bernoulli(p) ; $f = p^x(1-p)^{1-x}$, $x = 0,1$, $p \in (0,1)$

Then $E(Y_i) = \sum_{y\in\mathcal{Y}} yP(Y=y) = 1\cdot p + 0\cdot(1-p) = p.$

Let $X = Y_1+\dots+Y_n$. By linearity of expectation

$E(X) = E(Y_1)+\dots+E(Y_n) = np.$

1a) Let X - continuous, nonnegative RV [$f(x) = 0$ for $x < 0$]. Show that $E(X) = \int_0^\infty [1-F(x)]\,dx$:

$\int_0^\infty(1-F(x))\,dx = \int_0^\infty P(X>x)\,dx = \int_0^\infty\int_x^\infty f_X(x)\,dx = \int_0^\infty\left(\int_0^\infty \mathbb{1}(x\le y)f_X(y)\,dy\right)dx$

$= \int_0^\infty\left(\int_0^\infty \mathbb{1}(x\le y)f_X(y)\,dx\right)dy = \int_0^\infty f_X(y)\left(\int_0^\infty\mathbb{1}(x\le y)\,dx\right)dy = \int_0^\infty y f_X(y)\,dy = E(X).$ ∎

1b) Let X - discrete RV whose range is the nonnegative integers. Show that $E(X) = \sum_{k=0}^{\infty}(1-F_X(k))$:

$E(X) = \sum_{k=0}^{\infty}kP(X=k) = \sum_{k=0}^{\infty}\sum_{j=1}^{k}P(X=k)$, letting $k = \sum_{j=1}^{k}1$

$= \sum_{k=0}^{\infty}\sum_{j=1}^{\infty}\mathbb{1}(j\le k)P(X=k) = \sum_{j\ge1}\sum_{k\ge j}P(X=j) = \sum_{j\ge1}1-P(X<j)$

$= \sum_{j\ge1}1-P(X\le j-1) = \sum_{j\ge1}1-F_X(j-1)$, letting $k = j-1$

$= \sum_{k\ge0}1-F_X(k)$

$= \sum_{k=0}^{\infty}1-F_X(k).$ ∎

2.17 A *median* of a distribution is a value m such that $P(X \le m) \ge \frac{1}{2}$ and $P(X \ge m) \ge \frac{1}{2}$. (If X is continuous, m satisfies $\int_{-\infty}^{m} f(x)\,dx = \int_m^\infty f(x)\,dx = \frac{1}{2}$.) Find the median of the following distributions.

(a) $f(x) = 3x^2, \quad 0 < x < 1$ (b) $f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty$

a) $\int_0^m 3x^2\,dx = x^3\Big|_0^m = m^3 \overset{\text{set}}{=} \frac{1}{2} \Rightarrow m = \left(\frac{1}{2}\right)^{1/3} = \underline{0.794}$

b) $X \sim \text{Cauchy}(\theta=0, \sigma=1) = \frac{1}{\pi(1+x^2)}$, $-\infty<x<\infty \Rightarrow \int_{-\infty}^{0}f(x)\,dx = \int_0^\infty f(x)\,dx \Rightarrow \underline{m=0.}$

Formally, $\int_{-\infty}^{m}f(x)\,dx = \frac{1}{\pi}\int_{-\infty}^{m}\frac{1}{1+x^2}dx$ letting $x = \tan(\theta)$, $dx = \sec^2(\theta)\,d\theta$

$\theta = \tan^{-1}(x) \Rightarrow \begin{cases} x=-\infty, \theta = -\pi/2 \\ x = m, \theta = \tan^{-1}(m)\end{cases}$

$= \frac{1}{\pi}\int_{-\pi/2}^{\tan^{-1}(m)}\frac{1}{1+\tan^2(\theta)}\sec^2\theta\,d\theta = \frac{1}{\pi}\left[\theta\Big|_{-\pi/2}^{\tan^{-1}(m)}\right]$

$= \frac{1}{\pi}[\tan^{-1}(m)+\pi/2] \overset{\text{set}}{=} \frac{1}{2} \Rightarrow \tan^{-1}(m) = 0 \Rightarrow m = 0.$ ∎

2.22) Let X have pdf: $f(x) = \frac{4}{\beta^3\sqrt{\pi}}x^2e^{-x^2/\beta^2}$, $0<x<\infty$, $\beta>0$.

a) verify that $f(x)$ is a pdf, i.e. show $\begin{cases}\text{i) nonnegativity}\\ \text{ii) normalization}\end{cases}$
- i) For $x>0$, $\beta>0$, every factor is nonnegative $\Rightarrow f(x) \ge 0$
- ii) show $\int_0^\infty f(x)\,dx = 1$

First, compute $A := \int_0^\infty x^2e^{-x^2/\beta^2}dx$ via IBP

let $u = x$, $dv = xe^{-x^2/\beta^2}dx$

$du = dx$, $v = \int xe^{-x^2/\beta^2}dx$ via substitution

$w = -x^2/\beta^2$

$dw = -2x/\beta^2\,dx \Rightarrow x\,dx = \frac{-\beta^2}{2}dw$

$= \int e^w\left(-\frac{\beta^2}{2}dw\right) = \frac{-\beta^2}{2}\int e^w\,dw = -\frac{\beta^2}{2}e^{-x^2/\beta^2}$

$= uv\Big|_0^\infty - \int_0^\infty v\,du$

where $\lim_{x\to\infty}x\cdot\left(-\frac{\beta^2}{2}e^{-x^2/\beta^2}\right) = 0$ and $uv|_{x=0} = 0 \Rightarrow uv\Big|_0^\infty = 0$

$= -\int_0^\infty\left(\frac{-\beta^2}{2}e^{-x^2/\beta^2}\right)dx = \frac{\beta^2}{2}\int_0^\infty e^{-x^2/\beta^2}dx$; compute integral via substitution

let $y = x/\beta \Rightarrow x = \beta y \Rightarrow dx = \beta\,dy$

$\Rightarrow \int_0^\infty e^{-(\beta y)^2/\beta^2}\beta\,dy = \beta\int_0^\infty e^{-y^2}dy$

$= \beta\cdot\frac{\sqrt{\pi}}{2}$ since $\int_{-\infty}^{\infty}e^{-y^2}dy = \sqrt{\pi}$

Hence,

$A = \frac{\beta^2}{2}\cdot\beta\frac{\sqrt{\pi}}{2} = \frac{\beta^3\sqrt{\pi}}{4}.$

Thus,

$\underline{\int_0^\infty f(x)\,dx = \frac{4}{\beta^3\sqrt{\pi}}A = \frac{4}{\beta^3\sqrt{\pi}}\frac{\beta^3\sqrt{\pi}}{4} = 1.}$ ∎

b) Find $E(X) = \int_0^\infty x f(x)\,dx$ and $V(X) = E(X^2)-[E(X)]^2$ where $E(X^2) = \int_0^\infty x^2f(x)\,dx$

compute $B := \int_0^\infty x^3e^{-x^2/\beta^2}dx$ via IBP

$u = x^2$, $dv = xe^{-x^2/\beta^2}dx$

$du = 2x\,dx$, $v = \int xe^{-x^2/\beta^2}dx = \frac{-\beta^2}{2}e^{-x^2/\beta^2}$ as seen in (a)

$= uv\Big|_0^\infty - \int_0^\infty v\,du = 0 - \int_0^\infty\left(-\frac{\beta^2}{2}e^{-x^2/\beta^2}\right)2x\,dx$

$= \beta^2\int_0^\infty xe^{-x^2/\beta^2}dx$ via substitution

$w = x^2/\beta^2$

$dw = 2x/\beta^2\,dx \Rightarrow x\,dx = \frac{\beta^2}{2}dw$

where $\int_0^\infty xe^{-x^2/\beta^2}dx = \frac{\beta^2}{2}\int_0^\infty e^{-w}dw = \frac{\beta^2}{2}$ since $[-e^{-w}]_0^\infty = 1$

$= \beta^2\cdot\frac{\beta^2}{2} = \frac{\beta^4}{2}.$ So, $\boxed{E(X) = \frac{4}{\beta^3\sqrt{\pi}}B = \frac{4}{\beta^3\sqrt{\pi}}\cdot\frac{\beta^4}{2} = \frac{2\beta}{\sqrt{\pi}}.}$

Now compute $C := \int_0^\infty x^4e^{-x^2/\beta^2}dx$ via IBP

let $u = x^3$, $dv = xe^{-x^2/\beta^2}$

$du = 3x^2dx$, $v = \frac{-\beta^2}{2}e^{-x^2/\beta^2}$ as seen previously

$= uv\Big|_0^\infty - \int_0^\infty v\,du = \int_0^\infty\left(\frac{\beta^2}{2}e^{-x^2/\beta^2}\right)\cdot3x^2dx = \frac{3\beta^2}{2}\int_0^\infty x^2e^{-x^2/\beta^2}dx = \frac{3\beta^2}{2}A = \frac{3\beta^2}{2}\cdot\frac{\beta^3\sqrt{\pi}}{4} = \frac{3\beta^5\sqrt{\pi}}{8}.$

Hence,

$\underline{E(X^2) = \frac{4}{\beta^3\sqrt{\pi}}C = \frac{4}{\beta^3\sqrt{\pi}}\cdot\frac{3\beta^5\sqrt{\pi}}{8} = \frac{3}{2}\beta^2}$

Therefore,

$\boxed{V(X) = E(X^2)-[E(X)]^2 = \frac{3}{2}\beta^2 - \left(\frac{2\beta}{\sqrt{\pi}}\right)^2 = \beta^2\left(\frac{3}{2}-\frac{4}{\pi}\right)}$

Proof (i). Show $X \ge 0$ a.s $\Rightarrow E(X) \ge 0$:

Let X - continuous RV. Then $P(X\ge0) = 1 \Rightarrow P(X<0) = 0$

$\Rightarrow f_X(x) = 0$ for (almost) all $x<0$

Hence,

$E(X) = \int_{-\infty}^{\infty}xf(x)\,dx = \int_0^\infty \underbrace{xf(x)\,dx}_{\ge0} \ge 0.$ ∎

Proof (ii). Show $X \ge Y$ a.s. $\Rightarrow E(X) \ge E(Y)$

$X \ge Y$ a.s. $\Rightarrow X-Y \ge 0$ a.s.

$\Rightarrow E(X-Y) \ge 0$ by (i)

$\Rightarrow E(X)-E(Y) \ge 0$ by linearity

$\Rightarrow E(X) \ge E(Y).$ ∎

2.3: Moments

Proof (variance of a RV). Show that $V(X) = E(X^2) - E^2(X)$:

$$V(X) = E(X - E(X))^2 = E(X^2 - 2XE(X) + E^2(X))$$
$$= E(X^2) - 2E^2(X) + E^2(X) = E(X^2) - E^2(X). \square$$

Ex (optimization problem): Consider the following problem: $\min_{b\in\mathbb{R}} E(X-b)^2$. Then,

$$E(X-b)^2 = E(X^2 - 2Xb + b^2) = E(X^2) - 2bE(X) + b^2$$
$$= b^2 - 2bE(X) + (E(X))^2 + E(X^2) - [E(X)]^2$$
$$= (b - E(X))^2 + \mathrm{var}(X) \geq \mathrm{var}(X) \text{ since } \mathrm{var}(X) \geq 0.$$

The 'location' b which solves the least squares problem is $b = E(X)$.
Thus, $\min_{b\in\mathbb{R}} E(X-b)^2 = \mathrm{var}(X)$

Ex (var. of Exponential RV): Suppose $X \sim EXP(\lambda) = \frac{1}{\lambda}e^{-x/\lambda}$, $x \in \mathcal{X} = (0, +\infty)$, $\lambda > 0$

with $E(X) = \lambda$. Show $V(X) = \lambda^2$

$V(X) = E(X^2) - E^2(X)$ where $E(X^2) =: I = \int_0^\infty x^2 \frac{1}{\lambda} e^{-x/\lambda} dx$ via IBP

let $u = x^2$ $dv = \frac{1}{\lambda}e^{-x/\lambda}$

$du = 2x\,dx$ $v = \int \frac{1}{\lambda} e^{-x/\lambda} dx = -e^{-x/\lambda}$

$= uv - \int v\,du$

$= -x^2 e^{-x/\lambda}\Big|_0^\infty + \int e^{-x/\lambda} 2x\,dx$

where $uv = 0$ since $\lim_{x\to\infty} e^{-x} = 0$ and $uv|_{x=0} = 0$

$$= 2\underbrace{\int x e^{-x/\lambda} dx}_{= J}.$$

$\Rightarrow J = \int_0^\infty x e^{-x/\lambda} dx$ via IBP

let $u = x$ $dv = e^{-x/\lambda}$

$du = dx$ $v = -\lambda e^{-x/\lambda}$

$= \cancel{uv}\Big|_0^\infty - \int_0^\infty v\,du$ (0)

$= \lambda \int_0^\infty e^{-x/\lambda} = -\lambda^2 e^{-x/\lambda}\Big|_0^\infty = \lambda^2$

$\Rightarrow I = 2J = 2\lambda^2$.

Thus, $V(X) = E(X^2) - E^2(X)$
$= 2\lambda^2 - \lambda^2 = \lambda^2$.

Proof (variance of a constant). Show that $\forall\, a, b \in \mathbb{R}$, $V(aX+b) = a^2\mathrm{var}(X)$:

$$V(aX+b) = E[(aX+b)^2 - (aE(X)+b)^2]$$
$$= E[a^2x^2 + 2axb + b^2 - a^2E^2(X) - 2abE(X) - b^2]$$
$$= a^2E(X^2) + 2abE(X) - a^2E^2(X) - 2abE(X)$$
$$= a^2(E(X^2) - E^2(X))$$
$$= a^2 \mathrm{Var}(X). \square$$

Ex (Binomial RV variance): Let $X \sim \mathrm{Binomial}(n,p) = \binom{n}{x} p^x (1-p)^x$, $n \in \mathbb{N}$, $p \in (0,1)$

with $E(X) = np$; show that $V(X) = E(X^2) - E^2(X) = np(1-p)$.

$E(X^2) = \sum_{x=0}^n x^2 f(x) = \sum_{x=0}^n x^2 \binom{n}{x} p^x (1-p)^{n-x}$

where $x\binom{n}{x} = n\binom{n-1}{x-1}$

$= \sum_{x=1}^n xn\binom{n-1}{x-1} p^x (1-p)^{n-x}$,

letting $y = x-1 \Rightarrow x = y+1$

$= \sum_{y=0}^{n-1} n(y+1)\binom{n-1}{y} p^{y+1} (1-p)^{n-y-1}$

$= np \times \left[\sum_{y=0}^{n-1} y\binom{n-1}{y} p^y (1-p)^{n-y-1} + \sum_{y=0}^{n-1}\binom{n-1}{y} p^y (1-p)^{n-y-1}\right]$

where $nx(x+a)^n = \sum_{k=0}^n x\binom{n}{k} x^k a^{n-k}$

$= np[(n-1)p(p+(1-p))^{n-1} + (1)(p + (1-p))^{n-1}]$

$= np[(n-1)p + 1] = n(n-1)p^2 + np$.

Hence,

$V(X) = E(X^2) - E^2(X) = n(n-1)p^2 + np - (np)^2$
$= n^2p^2 - np^2 + np - n^2p^2$
$= np - np^2 = np(1-p). \square$

Equivalently, a faster approach is as follows:

Let $X_i = \begin{cases} 1, & i^{th} \text{ toss } H \\ 0, & \text{otherwise} \end{cases}$; Then, $E(X_i) = \sum_i x_i P(X = x_i) = 1\cdot p + 0(1-p) = p$

and $E(X_i^2) = 1^2 p + 0^2(1-p) = p$.

Also, $V(X_i) = E(X_i^2) - E^2(X_i) = p - p^2 = p(1-p)$.

Note that $X = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n$ are mutually independent.
Later, we will formally show that

$V(X) = V(X_1) \times \cdots \times V(X_n) = n \cdot \mathrm{var}(X_i) = np(1-p)$.

Proof (mgf evaluation). Show that $\forall\, n \in \mathbb{N}$, $E(X^n) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$:

Let $n = 0$. Then $E(X^0) = E(1) = 1$
$\Leftrightarrow M_X(0) = E(e^{0\cdot x}) = E(1) = 1$

$n = 1$. $E(X) = \frac{d}{dt} M_X(t) = \frac{d}{dt}\int e^{tx} f(x)\,dx$ or Σ

$= \int \frac{d}{dt} e^{tx} f(x)\,dx = \int x e^{tx} f(x)\,dx$

$\Rightarrow M_X'(t)\big|_{t=0} = \int x f(x)\,dx = E(X)$.

$n = 2$. $E(X^2) = M_X^{(2)}(t) = \int \frac{d}{dt} x e^{tx} f(x)\,dx = \int x^2 e^{tx} f(x)\,dx$

$\Rightarrow M_X^{(2)}(t)\big|_{t=0} = \int x^2 f(x)\,dx = E(X^2)$.

$\Rightarrow \cdots \Rightarrow$ for any $n \in \mathbb{N}$. $\square$

Ex (Gamma RV): For any $\alpha, \beta > 0$, define Gamma function $:= \Gamma(\alpha) = \frac{1}{\beta^\alpha}\int_0^\infty x^{\alpha-1} e^{-x/\beta} dx = \int_0^\infty y^{\alpha-1} e^{-y} dy$

$\Gamma(\alpha)$ does not depend on β; see via change-of-var, letting $y = x/\beta \Rightarrow x = \beta y \Rightarrow dx = \beta dy$

$$\Gamma(\alpha) = \frac{1}{\beta^\alpha}\int_0^\infty (\beta y)^{\alpha-1} e^{-y} \beta\,dy = \int_0^\infty y^{\alpha-1} e^{-y} dy \quad (*)$$

where $(*)$ does not depend on β and is usually taken to be the definition of $\Gamma(\alpha)$.

Show $E(X) = \alpha\beta$

Let $X \sim \mathrm{Gamma}(\underset{\text{shape}}{\alpha}, \underset{\text{scale}}{\beta})$; $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$, $x \in (0,\infty)$, $\alpha, \beta > 0$

$\Leftrightarrow$ $X \sim \mathrm{Gamma}(\alpha, \lambda = 1/\beta)$; $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$

$f(x)$ is a valid pdf since

$$\int_0^\infty f(x)\,dx = \frac{1}{\Gamma(\alpha)}\frac{1}{\beta^\alpha}\int x^{\alpha-1} e^{-x/\beta} dx = \frac{\Gamma(\alpha)}{\Gamma(\alpha)} = 1.$$

For $X \sim \mathrm{Gamma}(\alpha, \beta)$ with $\Gamma(\alpha) = \int x^{\alpha-1} e^{-x} dx$,

$M_X(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty e^{tx} x^{\alpha-1} e^{-x/\beta} dx$

$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\underbrace{\int_0^\infty x^{\alpha-1} e^{-x(\frac{1}{\beta}-t)} dx}_{I}$

where $\frac{1}{\beta} - t = \frac{1-\beta t}{\beta} \Rightarrow I = \Gamma(\alpha)\left(\frac{\beta}{1-\beta t}\right)^\alpha$, $1 - \beta t > 0$

$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\cdot I = \left(\frac{1}{1-\beta t}\right)^\alpha$, $t < 1/\beta$.

If $t \geq \frac{1}{\beta} \Rightarrow \frac{1}{\beta} - t \leq 0 \Rightarrow M_X(t) = \int_0^\infty x^{\alpha-1} e^{-x(1/\beta - t)} dx = +\infty$

Now,

$E(X) = M_X'(t)\big|_{t=0} = \frac{d}{dt}[(1-\beta t)^{-\alpha}] = \alpha(1-\beta t)^{-\alpha-1}\beta\big|_{t=0} = \alpha\beta$

Note: If $X \sim \mathrm{Gamma}(\alpha, \beta)$, then $\frac{X}{\beta} \sim \mathrm{Gamma}(\alpha, 1)$

2.30 Find the moment generating function corresponding to

(a) $f(x) = \frac{1}{c}$, $0 < x < c$.

(b) $f(x) = \frac{2x}{c^2}$, $0 < x < c$.

(c) $f(x) = \frac{1}{2\beta} e^{-|x-\alpha|/\beta}$, $-\infty < x < \infty$, $-\infty < \alpha < \infty$, $\beta > 0$.

(d) $P(X = x) = \binom{r+x-1}{x} p^r (1-p)^x$, $x = 0, 1, \ldots$, $0 < p < 1, r > 0$ an integer.

a) $E(e^{tX}) = \int_0^c e^{tx}\frac{1}{c}dx = \frac{1}{ct}e^{tx}\Big|_0^c = \frac{1}{ct}e^{tc} - \frac{1}{ct} = \underline{\frac{1}{ct}(e^{tc} - 1)}$

b) $E(e^{tX}) = \int_0^c \frac{2x}{c^2} e^{tx} dx = \frac{1}{c^2}\int_0^c 2x e^{tx} dx$

w/IBP: $u = 2x$ $dv = e^{tx}dx$

$du = 2$ $v = \int e^{tx}dx = \frac{1}{t}e^{tx}dx$

$= \frac{1}{c^2}\left[\frac{2x}{t}e^{tx}\Big|_0^c - \int_0^c \frac{2}{t}e^{tx}dx\right]$

$= \frac{1}{c^2}\left[\frac{2c}{t}e^{tc} - \left(\frac{2}{t^2}e^{tx}\right)\Big|_0^c\right]$

$= \frac{1}{c^2}\left[\frac{2c}{t}e^{tc} - \left(\frac{2}{t^2}e^{tc} - \frac{2}{t^2}\right)\right]$

$= \frac{2e^{tc}}{ct} - \frac{2e^{tc}}{c^2t^2} + \frac{2}{c^2t^2} = \underline{\frac{2}{c^2t^2}(cte^{tc} - e^{tc} + 1)}$

c) $E(e^{tX}) = \int_{-\infty}^\infty \frac{1}{2\beta} e^{-|x-\alpha|/\beta} dx = \int_{-\infty}^\alpha \frac{1}{2\beta} e^{(x-\alpha)/\beta} e^{tx} dx + \int_\alpha^\infty \frac{1}{2\beta} e^{-(x-\alpha)/\beta} e^{tx} dx$

Since $\int_{-\infty}^\infty f(|x-a|)\,dx = \underbrace{\int_{-\infty}^a f(a-x)\,dx}_{x<a} + \underbrace{\int_a^\infty f(x-a)\,dx}_{x>a}$ FS

$= \int_{-\infty}^\alpha \frac{1}{2\beta} e^{x/\beta - \alpha/\beta + tx} dx + \int_\alpha^\infty \frac{1}{2\beta} e^{-x/\beta + \alpha/\beta + tx} dx$

$= \frac{e^{-\alpha/\beta}}{2\beta} e^{x(\frac{1}{\beta}+t)} \frac{1}{(\frac{1}{\beta}+t)}\Big|_{-\infty}^\alpha - \frac{e^{\alpha/\beta}}{2\beta} e^{-x(\frac{1}{\beta}-t)} \frac{1}{(\frac{1}{\beta}-t)}\Big|_\alpha^\infty$ (1)

$= \left(\frac{e^{-\alpha/\beta + \alpha/\beta + \alpha t}}{2\beta(\frac{1}{\beta}+t)} - 0\right) - \left(0 - \frac{e^{\alpha/\beta - \alpha/\beta + \alpha t}}{2\beta(\frac{1}{\beta}-t)}\right)$

$= \frac{e^{\alpha t}}{2\beta(\frac{1}{\beta}+t)} + \frac{e^{\alpha t}}{2\beta(\frac{1}{\beta}-t)}$

$= \frac{e^{\alpha t}}{2+2\beta t} + \frac{e^{\alpha t}}{2-2\beta t}$

$= \frac{(2-2\beta t)e^{\alpha t} + (2+2\beta t)e^{\alpha t}}{(2+2\beta t)(2-2\beta t)}$

$= \frac{2e^{\alpha t} - 2\beta \cancel{t} e^{\alpha t} + 2e^{\alpha t} + 2\beta \cancel{t} e^{\alpha t}}{4 - 4\beta t + 4\beta t - 4\beta^2t^2}$

$= \frac{4e^{\alpha t}}{4 - 4\beta^2t^2}$, $-1/\beta < t < 1/\beta$ s.t. (1) converges, i.e. the integral can only be evaluated appropriately

when $e^{x(\frac{1}{\beta}+t)}\big|_{x=-\infty} \to 0$ iff $\frac{1}{\beta}+t < 0 \Rightarrow t < -\frac{1}{\beta}$

and $e^{-x(\frac{1}{\beta}-t)}\big|_{x=\infty} \to 0$ iff $\frac{1}{\beta} - t < 0 \Rightarrow t > \frac{1}{\beta}$.

d) $E(e^{tX}) = \sum_{x=0}^\infty \binom{r+x-1}{x} p^r (1-p)^x e^{tx}$

$= \sum_{x=0}^\infty \binom{r+x-1}{x} p^r ((1-p)e^t)^x$

$= \sum_{x=0}^\infty \binom{r+x-1}{x} p^r ((1-p)e^t)^x \frac{(1-(1-p)e^t)^r}{(1-(1-p)e^t)^r}$

$= \frac{p^r}{(1-(1-p)e^t)^r} \underbrace{\sum_{x=0}^\infty \binom{r+x-1}{x} (1-(1-p)e^t)^r ((1-p)e^t)^x}_{1}$

where, for $X \sim \mathrm{NegBinom}(r,p)$, $f(x) = \Sigma\binom{r+x-1}{x} p^r (1-p)^x$, $p \in (0,1)$

and above pmf $\mathrm{NegBinom}(r, 1-(1-p)e^t)$ sums to 1

$= \underline{\frac{p^r}{(1-(1-p)e^t)^r}}$, $t < -\log(1-p)$ since $0 < (1-(1-p)e^t) < 1$

b/c $1-(1-p)e^t$ is a prob.

$\Rightarrow 0 < 1 - (1-p)e^t < 1$

$\Rightarrow 1 - (1-p)e^t < 1$ since $(-(1-p)e^t > 0$

$\Rightarrow \log(1-p) + t < \log(1) = 0$

$\Rightarrow t < -\log(1-p)$.

3. Let X be a standard normal variable, i.e., X has pdf $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ for all $x \in (-\infty, +\infty)$. Define a discrete random variable Y by $P(Y = \sqrt{3}) = P(Y = -\sqrt{3}) = 1/6$ and $P(Y = 0) = 2/3$.

(i) Show that X and Y have the same r-th moment, for each $r = 1, 2, 3, 4, 5$.

(ii) Find another random variable Z which has the same r-th moments as X and Y for all $r = 1, \ldots, 5$.

i) For $X \sim N(0,1)$,

$M_X(t) = E(e^{tX}) = \int_{-\infty}^\infty e^{tx}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty \exp(tx - \frac{x^2}{2})\,dx$.

Now complete the square in the exponent:

$tx - \frac{x^2}{2} = \frac{-1}{2}(x^2 - 2tx) = -\frac{1}{2}((x-t)^2 - t^2) = -\frac{(x-t)^2}{2} + \frac{t^2}{2}$.

So,

$M_X(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty \exp\left(-\frac{(x-t)^2}{2} + \frac{t^2}{2}\right)dx = e^{t^2/2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-(x-t)^2/2}dx$

let $u = x - t$ $du = dx$

$= e^{t^2/2}\underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-u^2/2}du}_{\sim N(0,1) \to 1}$

$= e^{t^2/2}$ since

Thus, $M_X(t) = e^{t^2/2}$. Now evaluate $M_X^{(1)}(0), M_X^{(2)}(0), \ldots, M_X^{(5)}(0)$:

$E(X) := M_X'(0) = \frac{d}{dt}e^{t^2/2}\big|_{t=0} = te^{t^2/2}\big|_{t=0} = 0$

$E(X^2) := M_X''(0) = \frac{d}{dt}(te^{t^2/2})\big|_{t=0} = e^{t^2/2} + t(te^{t^2/2})\big|_{t=0} = (1+t^2)e^{t^2/2}\big|_{t=0} = 1$.

$E(X^3) := M_X^{(3)}(0) = \frac{d}{dt}((1+t^2)e^{t^2/2})\big|_{t=0} = 2te^{t^2/2} + (1+t^2)\cdot(te^{t^2/2})\big|_{t=0} = (3t+t^3)e^{t^2/3}\big|_{t=0} = 0$

$E(X^4)$: $M_X^{(4)}(t) = \frac{d}{dt}((3t+t^3)e^{t^2/2}) = (3+3t^2)e^{t^2/2} + (3t+t^3)\cdot(te^{t^2/2}) = (3+6t^2+t^4)e^{t^2/2}$

$\Rightarrow M_X^{(4)}(0) = 3$

$E(X^5) := M_X^{(5)}(0) = 0$ since odd moments vanish by symmetry (as seen above).

For discrete Y given by $P(Y=\sqrt{3}) = P(Y=-\sqrt{3}) = \frac{1}{6}$ and $P(Y=0) = \frac{2}{3}$,

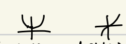
$M_Y(t) = \sum_y e^{yt} f_Y(y) = \frac{1}{6}e^{\sqrt{3}t} + \frac{2}{3}e^0 + \frac{1}{6}e^{-t\sqrt{3}}$

$= \frac{1}{6}(e^{\sqrt{3}t} + e^{-\sqrt{3}t}) + \frac{2}{3}$,

where $\cosh(u) = \frac{1}{2}(e^u + e^{-u})$; $\sinh(u) = \frac{1}{2}(e^u - e^{-u})$

$= \frac{1}{3}\cosh(\sqrt{3}t) + \frac{2}{3}$

Of interest,



$\frac{d}{dt}\cosh(u) = \sinh(u)\cdot u'$, $\frac{d}{dt}\sinh(u) = \cosh(u)\cdot u'$, $\cosh(0) = 1$, and $\sinh(0) = 0$

Given $M_Y(t) = \frac{1}{3}\cosh(\sqrt{3}t) + \frac{2}{3}$, find $E(Y), E(Y^2), \ldots, E(Y^5)$:

$E(Y)$: $M_Y'(t) = \frac{1}{3}(\sqrt{3})\sinh(\sqrt{3}t)\big|_{t=0} = \frac{1}{3}(\sqrt{3})\cdot 0 = 0$

$E(Y^2)$: $M_Y''(t) = \frac{1}{3}(\sqrt{3})^2\cosh(\sqrt{3}t) = \cosh(\sqrt{3}t)\big|_{t=0} = 1$

$E(Y^3)$: $M_Y^{(3)}(t) = \frac{1}{3}(\sqrt{3})^3\sinh(\sqrt{3}t) = \sqrt{3}\sinh(\sqrt{3}t)\big|_{t=0} = 0$

$E(Y^4)$: $M_Y^{(4)}(t) = \frac{1}{3}(\sqrt{3})^4\cosh(\sqrt{3}t) = 3\cosh(\sqrt{3}t)\big|_{t=0} = 3$

$E(Y^5) := M_Y^{(5)}(0) = 0$ since odd moments vanish by symmetry

ii) Define $Z := P(Z=1) = P(Z=-1) = \frac{1}{6}$, $P(Z=2) = P(Z=-2) = \frac{1}{12}$, and $P(Z=0) = \frac{1}{2}$.

Then,

$M_Z(t) = E(e^{tZ}) = \frac{1}{6}(e^t + e^{-t}) + \frac{1}{12}(e^{2t} + e^{-2t}) + \frac{1}{2}$

$= \frac{1}{3}\cosh(t) + \frac{1}{6}\cosh(2t) + \frac{1}{2}$

Next, we can compute $E(Z), E(Z^2), \ldots, E(Z^5)$.

As discussed in (i), the odd moments vanish by symmetry, i.e.

$E(Z) = E(Z^3) = E(Z^5) = 0$, and

$E(Z^2)$: $M_Z''(t) = \frac{1}{3}\cosh(t) + \frac{2}{3}\cosh(2t)\big|_{t=0} = \frac{1}{3} + \frac{2}{3} = 1$

$E(Z^4)$: $M_Z^{(4)}(t) = \frac{1}{3}\cosh(t) + \frac{8}{3}\cosh(2t)\big|_{t=0} = \frac{1}{3} + \frac{8}{3} = 3$.

Therefore, Z contains the same r-th moments as X and Y for $r = 1, \ldots, 5$.

Ex (Binomial RV): Let $X \sim \text{Binomial}(n,p)$; $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \dots, n$

Then
$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x}$$
using the binomial formula
$$= (e^t p + (1-p))^n$$

Then, $E(X^k) = M_X^{(k)}(t)\big|_{t=0}$. ∎

2.3 : Moments (cont.)

Ex (unbounded support): Suppose X and Y have an unbounded support. Let $X \sim f_1$ and $Y \sim f_2$.

Let $Z \sim N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x \in \mathbb{R}$.

Suppose $X = e^Z$. Then, $f_X(x) = f_Z(g^{-1}(x)) \left| \frac{d}{dx} g^{-1}(x) \right|$

where $x = e^z \Rightarrow g^{-1}(x) = \ln(x)$; $\frac{d}{dx} g^{-1}(x) = \frac{1}{x}$.

Thus,
$$f(x) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}(\log(x))^2}, \quad x \in [0, \infty) \sim \text{LogNormal}(0,1)$$

as $\text{LogNormal}(\mu, \sigma^2)$ has pdf $\frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, $x > 0, \sigma > 0$

Let
$$f_2(x) = f_1(x) \cdot (1 + \sin(2\pi \log(x))).$$

Then, it can be shown that
$$E(X^r) = E(Y^r) = e^{r^2/2}, \quad r = 0, 1, 2, \dots$$

Proof (sketch - convergence of mgf $\Rightarrow$ convergence of cdf). $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$

(Beyond scope of class)

is, in fact, the Laplace transform of the pdf $f(x)$.

From the theory of Laplace transform,
$$f(x) \xrightarrow{\mathcal{L}} M_X(t)$$
is one-to-one, provided that $M_X(t)$ is defined for a neighborhood of 0.

Hence f_X is uniquely determined by function M_X.

From here, one can establish a degree of regularity for the inverse map of the Laplace transform. □

Ex (Poisson Approx.): Let $X \sim \text{Binomial}(n,p)$. We know $\begin{cases} E(X) = np \\ V(X) = np(1-p) \end{cases}$

As n gets large, X "behaves" like a Poisson RV.

$Y \sim \text{Poisson}(\lambda)$ if $f_Y(y) = e^{-\lambda} \frac{\lambda^y}{y!}$, $y = 0, 1, \dots$, $E(Y) = V(Y) = \lambda$

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

We will establish that
$$P(X \le x) \simeq P(Y \le x) \text{ as } n \uparrow \infty \ \forall x \in \mathbb{N}.$$

More precisely, we also need $p \downarrow 0$ s.t. $np = \text{constant} \equiv \lambda$, i.e.

assume $\begin{cases} np = \lambda \text{ (constant)} \\ n \uparrow \infty, p \downarrow 0 \end{cases}$

Then,
$$M_X(t) = (pe^t + 1 - p)^n$$

Since $X_n \stackrel{d}{=} Y_1 + \dots + Y_n$ where $Y_i \stackrel{iid}{\sim} \text{Bernoulli}(p)$ s.t. $\begin{cases} P(Y_i = 1) = p \\ P(Y_i = 0) = 1 - p \end{cases}$

$\Rightarrow M_{X_n}(t) = E(e^{tX}) = E(e^{t(Y_1 + \dots + Y_n)}) = E(e^{tY_1}) \cdots E(e^{tY_n})$

and $Y_1, \dots, Y_n$ are independent; thus, $= \prod_{i=1}^{n} E(e^{tY_i}) = (pe^t + (1-p))^n$

Thus,
$$M_{X_n}(t) = (pe^t + (1-p))^n = (1 + p(e^t - 1))^n, \text{ letting } p = \tfrac{1}{n} \cdot np$$
$$= (1 + \tfrac{1}{n}(e^t - 1)np)^n, \text{ letting } np = \lambda$$
$$= (1 + \tfrac{\lambda}{n}(e^t - 1))^n$$
$$\Rightarrow \lim_{n \to \infty} M_{X_n}(t) = \lim_{n \to \infty} (1 + \tfrac{a}{n})^n, \text{ letting } a := \lambda(e^t - 1)$$
$$= e^a \quad \text{since } e^a = \lim_{x \to \infty} (1 + \tfrac{a}{x})^x$$
$$= \exp(\lambda(e^t - 1)) \quad (*)$$

where $(*)$ is exactly the MGF of $Y \sim \text{Poisson}(\lambda)$, i.e.
$$M_Y(t) = e^{\lambda(e^t - 1)}.$$

By the convergence thm. of mgf,

if $X_n \sim \text{Binomial}(n,p)$, $\lambda = np$. Then,

$F_{X_n}(x) \longrightarrow F_Y(x)$ for $Y \sim \text{Poisson}(\lambda)$

as $n \to \infty$ $\forall x$ where F_Y is continuous.

$\Leftrightarrow M_X(t) \to M_Y(t)$ $\forall t$

- For Poisson, F_Y is a step function w/ continuity at $n \notin \mathbb{N}$; for $n \in \mathbb{N}$, $F_Y(y) = F_Y(y + 1/2)$ continuous at $y + 1/2$.
  $\Rightarrow P(X_n \le x) \to P(Y \le x)$ $\forall x \in \mathbb{N}$ (in fact, here $\forall x \in \mathbb{R}$)
- Still holds if $X_n \sim \text{Binomial}(n, p_n)$ where $n \to \infty$, $p_n \to 0$ s.t. $np_n \to \lambda$.

Proof (mgf of Transformed RV). Show that $\forall a, b \in \mathbb{R}$, $M_{aX+b} = e^{bt} M_X(at)$:
$$M_{aX+b} = E(e^{t(aX+b)}) = E(e^{taX} e^{bt})$$
$$= e^{bt} E(e^{taX}) = e^{bt} M_X(at). \ \square$$

2.38 Let X have the negative binomial distribution with pmf
$$f(x) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \dots,$$
where $0 < p < 1$ and $r > 0$ is an integer.

(a) Calculate the mgf of X.

(b) Define a new random variable by $Y = 2pX$. Show that, as $p \downarrow 0$, the mgf of Y converges to that of a chi squared random variable with $2r$ degrees of freedom by showing that
$$\lim_{p \to 0} M_Y(t) = \left(\frac{1}{1-2t}\right)^r, \quad |t| < \frac{1}{2}.$$

a) $M_X(t) = E(e^{tX}) = \dfrac{p^r}{(1 - (1-p)e^t)^r}$, $t < -\log(1-p)$ (as completed above in 2.30d)

b) Let $Y = 2pX$. Then, since $\forall a, b \in \mathbb{R}$, $M_{aX+b}(t) = e^{bt} M_X(at)$
$$M_Y(t) = M_{2pX}(t) = M_X(2pt) = \frac{p^r}{(1 - (1-p)e^{2pt})^r}$$

Then,
$$\lim_{p \to 0} M_Y(t) = \lim_{p \to 0} \frac{p^r}{(1 - (1-p)e^{2pt})^r}$$
$$= \lim_{p \to 0} \left(\frac{p'}{\frac{d}{dp}[1 - (1-p)e^{2pt}] \ (*)} \right)^r$$

by L'hopital's rule: $\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$ indeterminate

where $(*)$: $-((1-p)(2te^{2pt}) + e^{2pt}(-1))$
$$= -e^{2pt}(2t(1-p) - 1)$$
$$= e^{2pt}(1 - 2t(1-p))$$
$$= \lim_{p \to 0} \left(\frac{1}{e^{2pt}(1 - 2t(1-p))} \right)^r$$
$$= \left(\frac{1}{1-2t}\right)^r, \quad |t| < \tfrac{1}{2}.$$

5. Let $n \in \mathbb{N}$, define function f_{X_n} as follows: For $x \in \mathbb{R}$,
$$f_{X_n}(x) = (1/2)\frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}x^2} + (1/2)\frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}(x-1)^2}.$$

(i) Verify that f_{X_n} is a valid pdf for a continuous random variable, which we denote by X_n.

(ii) Show that the sequence of random variables $\{X_n\}$ converges in distribution to a random variable Y. What is the distribution of Y?

For $n \in \mathbb{N}$, define
$$f_{X_n}(x) = \frac{1}{2} \frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}x^2} + \frac{1}{2} \frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}(x-1)^2}, \quad x \in \mathbb{R} \quad (1)$$

where each (equally weighted) term is the density of a normal curve, i.e.

$(2) := \frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}x^2} \sim N(0, \frac{1}{n})$ and $(3) := \frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}(x-1)^2} \sim N(1, \frac{1}{n})$

i) verify that f_{X_n} is a valid pdf, i.e.

i) $f_{X_n}(x) \ge 0$ for all $x \in \mathbb{R}$ and ii) $\int_{-\infty}^{\infty} f_{X_n}(x)\,dx = 1$.

Nonnegativity (i) holds since (1) is comprised of two nonnegative Gaussian densities, i.e.

$(2) \ge 0$ $\forall x \in \mathbb{R}$ and $(3) \ge 0$ $\forall x \in \mathbb{R}$. Thus, $(1) := f_{X_n} \ge 0$ $\forall x \in \mathbb{R}$.

We can also verify (ii), i.e.
$$\int_{-\infty}^{\infty} f_{X_n}(x)\,dx = \frac{1}{2} \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}x^2}}_{1} dx + \frac{1}{2} \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}(x-1)^2}}_{1} = \frac{1}{2} + \frac{1}{2} = 1. \ ∎$$

ii) By definition, the MGF of X_n is
$$M_{X_n}(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_{X_n}(x)\,dx = \frac{1}{2}\sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{n}{2}x^2} dx + \frac{1}{2}\sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{n}{2}(x-1)^2} dx.$$

To compute the first integral $I_1 := \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{n}{2}x^2} dx$,

complete the square in the exponent, i.e.
$$tx - \frac{n}{2}x^2 = -\frac{n}{2}(x^2 - \frac{2t}{n}x) = -\frac{n}{2}\left((x - \frac{t}{n})^2 - \frac{t^2}{n^2}\right) = -\frac{n}{2}(x - \frac{t}{n})^2 + \frac{t^2}{2n}.$$

Therefore, $I_1 = \sqrt{\frac{n}{2\pi}} e^{t^2/2n} \int_{-\infty}^{\infty} e^{-\frac{n}{2}(x - \frac{t}{n})^2} dx$ via substitution:

let $u = x - \frac{t}{n} \Rightarrow du = dx$

where $\int_{-\infty}^{\infty} e^{-\frac{n}{2}u^2} du = \sqrt{\frac{2\pi}{n}}$.
$$\Rightarrow I_1 = \sqrt{\frac{n}{2\pi}} e^{t^2/2n} \sqrt{\frac{2\pi}{n}} = e^{t^2/2n}.$$

Similarly, for the second integral $I_2 := \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{n}{2}(x-1)^2} dx$,
$$tx - \frac{n}{2}(x-1)^2 = t(x-1) + t - \frac{n}{2}(x-1)^2.$$

So,
$$I_2 = e^t \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{t(x-1) - \frac{n}{2}(x-1)^2} dx \quad \text{via substitution}$$
let $s = x - 1$, $ds = dx$
$$= e^t \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{st - \frac{n}{2}s^2} ds$$
$$= e^t \sqrt{\frac{n}{2\pi}} e^{t^2/2n} \sqrt{\frac{2\pi}{n}} \quad \text{by the same above calculation}$$
$$= e^t e^{t^2/2n}. \text{ Thus,}$$
$$M_{X_n}(t) = \frac{1}{2} e^{t^2/2n} + \frac{1}{2} e^t e^{t^2/2n} = \frac{1}{2} e^{t^2/2n} (1 + e^t).$$
$$\Rightarrow \lim_{n \to \infty} M_{X_n}(t) = \lim_{n \to \infty} \frac{1}{2} e^{t^2/2n} (1 + e^t) = \frac{1}{2}(1 + e^t).$$

Now, define

$M_Y(t) := \frac{1}{2}(1 + e^t)$ is the mgf of $Y \sim \text{Bernoulli}(1/2)$

since $E(e^{tY}) = \frac{1}{2} e^{0 \cdot t} \cdot \frac{1}{2} e^{1 \cdot t}$.

This implies that

$X_n \xrightarrow{d} Y$ for Bernoulli(1/2) by the mgf convergence theorem.

4. Let $n \in \mathbb{N}$, define function f_{X_n} as follows: For $x \in \{0, 1, \dots, n\}$,
$$f_{X_n}(x) = (1/2)\binom{n}{x}(3n)^{-x}(1 - 1/(3n))^{n-x} + (1/2)\binom{n}{x}(6n)^{-x}(1 - 1/(6n))^{n-x}.$$
and $f_{X_n}(x) = 0$ otherwise.

(i) Verify that f_{X_n} is a valid pmf for a discrete random variable, which we denote by X_n. Derive the moment generating function for X_n.

(ii) Show that the sequence of random variables $\{X_n\}$ converges in distribution to a random variable Y. What is the distribution of Y?

i) Each summand is a binomial pmf: $\text{Bin}(n, \frac{1}{3n})$ and $\text{Bin}(n, \frac{1}{6n})$. Their avg. is a valid pmf.

Recall: For $X \sim \text{Bin}(n,p)$, $M_X(t) = [(1-p) + pe^t]^n$.

Here, the mgf is a mixture of mgf's, i.e.
$$M_{X_n}(t) = \frac{1}{2}\left(1 - \frac{1}{3n} + \frac{e^t}{3n}\right)^n + \frac{1}{2}\left(1 - \frac{1}{6n} + \frac{e^t}{6n}\right)^n = \frac{1}{2}\left(1 + \frac{e^t - 1}{3n}\right)^n + \frac{1}{2}\left(1 + \frac{e^t - 1}{6n}\right)^n.$$

ii) As $n \to \infty$, $(1 + \frac{a}{n})^n \to e^a$. Hence,
$$M_{X_n}(t) \to \frac{1}{2}\exp\left(\frac{e^t - 1}{3}\right) + \frac{1}{2}\exp\left(\frac{e^t - 1}{6}\right),$$

the mgf of the mixture $\frac{1}{2}\text{Poi}(\frac{1}{3}) + \frac{1}{2}\text{Poi}(\frac{1}{6})$.

Therefore,
$X_n \xrightarrow{d} Y$ where Y is the Poisson mixture.

2.4: Tools

Proof (Leibniz's Rule). LHS $= \lim_{\delta\to 0}\frac{1}{\delta}\left(\int_{a(\theta+\delta)}^{b(\theta+\delta)} f(x,\theta+\delta)\,dx - \int_{a(\theta)}^{b(\theta)} f(x,\theta)\,dx\right)$

$$= \lim_{\delta\to0}\frac{1}{\delta}\Big[\underbrace{\int_{a(\theta)}^{b(\theta)}[f(x,\theta+\delta)-f(x,\theta)]\,dx}_{:=A} + \underbrace{\int_{a(\theta+\delta)}^{a(\theta)} f(x,\theta+\delta)\,dx}_{:=B} + \underbrace{\int_{b(\theta)}^{b(\theta+\delta)} f(x,\theta+\delta)\,dx}_{:=C}\Big]$$

where $A = \lim_{\delta\to0}\frac{1}{\delta}\int_{a(\theta)}^{b(\theta)} f(x,\theta+\delta)-f(x,\theta)\,dx$

$= \int_{a(\theta)}^{b(\theta)} \lim_{\delta\to0}\frac{1}{\delta}[f(x,\theta+\delta)-f(x,\theta)]\,dx$

$= \int_{a(\theta)}^{b(\theta)} \frac{\partial f}{\partial\theta}(x,\theta)\,d\theta$ since $\begin{cases}\frac{\partial f}{\partial\theta} \text{ is cont. in } x\\ \text{next theorem}\end{cases}$

$B = \lim_{\delta\to0} -\frac{1}{\delta}\int_{a(\theta)}^{a(\theta+\delta)} f(x,\theta+\delta)\,dx$

$= \lim_{\delta\to0} -\frac{a(\theta+\delta)-a(\theta)}{\delta}\times\frac{1}{a(\theta+\delta)-a(\theta)}\int_{a(\theta)}^{a(\theta+\delta)} f(x,\theta+\delta)\,dx$

$= -\frac{d}{d\theta}a(\theta)\times f(a(\theta),\theta)$

Likewise,

$C = \frac{d}{d\theta}b(\theta)\, f(b(\theta),\theta)$. □

Ex (Exponential): Let $X\sim \text{EXP}(\lambda)$; $f(x)=\frac{1}{\lambda}e^{-x/\lambda}$, $\lambda\in[0,\infty)$, $x\ge 0$

Moments: $E(X^n)=\int_0^\infty \frac{1}{\lambda}x^n e^{-x/\lambda}\,dx$, $n=1,2,\dots$

wish to differentiate

$\frac{d}{d\lambda}E(X^n)=\frac{d}{d\lambda}\int_0^\infty \cdot \overset{(*)}{=} \int_0^\infty \frac{d}{d\lambda}\cdot$

$= \int_0^\infty x^n\left(\frac{-1}{\lambda^2}e^{-x/\lambda}+\frac{1}{\lambda}e^{-x/\lambda}\frac{x}{\lambda^2}\right)dx$

$= \int_0^\infty x^n e^{-x/\lambda}\frac{1}{\lambda^2}\left(\frac{x}{\lambda}-1\right)dx$

$= \int_0^\infty x^{n+1}e^{-x/\lambda}\frac{1}{\lambda^3} - x^n e^{-x/\lambda}\frac{1}{\lambda^2}\,dx$

$= \frac{1}{\lambda^2}E(X^{n+1}) - \frac{1}{\lambda}E(X^n)$

This gives a recursive relation, i.e.

$\underline{E(X^{n+1}) = \lambda E(X^n) + \lambda^2\frac{d}{d\lambda}E(X^n)}$

(Similar identity holds for broad family of dist.'s)

Now, we justify the interchange of $\int$ and $d/d\lambda$, i.e. (*)

where $\frac{d}{d\lambda}f(x,\lambda) = x^n e^{-x/\lambda}\frac{1}{\lambda^2}\left(\frac{x}{\lambda}-1\right)$.

Need to find a dominating function $g(\cdot,\lambda)$ s.t.

1) $\left|\frac{d}{d\theta}f(x,\theta)\big|_{\theta=\theta'}\right| < g(x,\lambda)\ \forall\, \theta'\in(\lambda-\delta_0,\lambda+\delta_0)$

2) g integrable wrt x

$\left|\frac{d}{d\theta}f(x,\theta)\right| = x^n e^{-x/\theta}\frac{1}{\theta^2}\left|\frac{x}{\theta}-1\right|$

$\le x^n e^{-x/\theta}\frac{1}{\theta^2}\left(\frac{x}{\theta}+1\right)$ since $x>0$

$\le x^n e^{-x/(\theta+\delta_0)}\frac{1}{(\theta-\delta_0)^2}\left(\frac{x}{\theta-\delta_0}+1\right) =: g(x,\theta)$

where the last inequality holds $\forall\, x>0, \forall\,\theta>\delta_0$.

Thus, ① holds for the chosen $g(x,\lambda)$.

To verify ②,

$\int g(x,\lambda)\,dx = \int x^{n+1}e^{-x/\lambda+\delta_0}\frac{1}{(\lambda-\delta_0)^3} + \int x^n e^{-x/\lambda+\delta_0}\frac{1}{(\lambda-\delta_0)^2}\,dx$

= Multiple of $(n+1)$-th moment of an exponential RU

+ multiple of n-th moment of another exp. RU

+ ... + " " " " $<+\infty$. □

Ex (Gaussian): Let $X\sim N(0,1)$; $f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $x\in\mathbb{R}$.

$M_X(t)=E(e^{tX})=\int_{-\infty}^{\infty}e^{tx}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx$

where $\frac{d}{dt}M_X(t)=\frac{d}{dt}E(e^{tX}) \overset{(*)}{=} E\left[\frac{d}{dt}(e^{tX})\right]$

Justifying (*) by finding dominating function for

$\left|\frac{d}{dt}e^{tx}e^{-x^2/2}\right| = \left|xe^{tx}e^{-x^2/2}\right| = |x|e^{tx}e^{-x^2/2}$

Let $0<\delta_0<t$:

$\left.\begin{array}{l}\text{if } x\ge0: |x|e^{tx}e^{-x^2/2}\le xe^{(t+\delta_0)x}e^{-x^2/2}\\ \text{if } x<0: |x|e^{tx}e^{-x^2/2}\le(-x)e^{(t-\delta_0)x}e^{-x^2/2}\end{array}\right\} g(x,t)$

Take the RHS as $g(x,t)$ for $x\ge0$ and $x<0$.

Then, $\left|\frac{d}{dt}e^{tx}e^{-x^2/2}\right|_{t=t'}\le g(x,t)\ \forall\, t'\in(t-\delta_0,t+\delta_0)$, i.e.

the domination holds.

Then, can check that $g(\cdot,t)$ is integrable:

$\int_0 xe^{(t+\delta_0)x}e^{-x^2/2} + \int_{-\infty}(-x)e^{(t-\delta_0)x}e^{-x^2/2}$

$<\infty$. □

Ex (Geometric): Let $X\sim\text{Geometric}(\theta)$, $0<\theta<1$; $P(X=x)=\theta(1-\theta)^x$, $x=1,2,\dots$

valid pmf since $\sum_{x=0}^{\infty}\theta(1-\theta)^x = \theta\sum_{x=0}^{\infty}(1-\theta)^x$

where $\sum_{k=0}^{\infty}r^k=\frac{1}{1-r}$, $|r|<1$

and $|1-\theta|<1$ since $0<\theta\le1$

$=\theta\cdot\frac{1}{1-(1-\theta)}=\theta\frac{1}{\theta}=1$

(*) Differentiating both sides wrt θ, assuming we can interchange Σ and $\frac{d}{d\theta}$:

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

$\Rightarrow \underbrace{\sum_{x=0}^{\infty}(1-\theta)^x}_{\frac{1}{1-(1-\theta)}} = \sum_{x=0}^{\infty}\theta x(1-\theta)^{x-1}$

$\frac{1}{1-(1-\theta)} = \frac{1}{1-\theta}\sum_{x=0}^{\infty}\theta x(1-\theta)^x$

$\Rightarrow \frac{1}{\theta}=\frac{1}{1-\theta}E(X) \Rightarrow E(X)=\frac{1-\theta}{\theta}$

Justifying (*),

- $h(x,\theta)=\theta(1-\theta)^x$ and
  $\forall\,\theta\in(0,1)$ $\sum_{x=0}^{k}h(x,\theta)=\theta\sum_{x=0}^{k}(1-\theta)^x=\theta\frac{1-(1-\theta)^{k+1}}{1-(1-\theta)}\to1$ as $k\to\infty$
  since $\sum_{x=0}^{k}r^x=\frac{1-r^{k+1}}{1-r}$

- $\frac{\partial}{\partial\theta}h(x,\theta)=(1-\theta)^x-\theta x(1-\theta)^{x-1}$ continuous in θ,
  so h is continuously diff. in θ

2.39 In each of the following cases calculate the indicated derivatives, justifying all operations.

(a) $\frac{d}{dx}\int_0^x e^{-\lambda t}\,dt$ (b) $\frac{d}{d\lambda}\int_0^\infty e^{-\lambda t}\,dt$

(c) $\frac{d}{dt}\int_t^1 \frac{1}{x^2}\,dx$ (d) $\frac{d}{dt}\int_1^\infty \frac{1}{(x-t)^2}\,dx$

a) verify $\frac{d}{dx}\left[\int_0^x e^{-\lambda t}dt\right]=\frac{d}{dx}\left[\frac{-1}{\lambda}e^{-\lambda t}\Big|_0^x\right]=\frac{d}{dx}\left[\left(\frac{-1}{\lambda}e^{-\lambda x}+\frac{1}{\lambda}\right)\right]=\frac{-1}{\lambda}(-\lambda)e^{-\lambda x}=e^{-\lambda x}$.

b) verify $\frac{d}{d\lambda}\int_0^\infty e^{-\lambda t}dt=\frac{d}{d\lambda}\left[\frac{-1}{\lambda}e^{-\lambda t}\Big|_0^\infty\right]=\frac{d}{d\lambda}\left[0+\frac{1}{\lambda}\right]=\frac{d}{d\lambda}(\lambda^{-1})=\frac{-1}{\lambda^2}$.

c) verify $\frac{d}{dt}\left[\int_t^1\frac{1}{x^2}dx\right]=\frac{d}{dt}\left[\frac{-1}{x}\Big|_t^1\right]=\frac{d}{dt}\left[-1+\frac{1}{t}\right]=-\frac{1}{t^2}$.

d) verify $\frac{d}{dt}\int_1^\infty(x-t)^{-2}dx=\frac{d}{dt}\left[-(x-t)^{-1}\Big|_1^\infty\right]=\frac{d}{dt}\left(\frac{1}{1-t}\right)=\frac{1}{(1-t)^2}$.

• convergence:

<u>Pointwise convergence</u>: for each $\theta \in (0,1)$

$$S_k(\theta) = \sum_{x=0}^{k} \frac{\partial}{\partial\theta} h(x,\theta) = \sum_{x=0}^{k} (1-\theta)^x - \theta x (1-\theta)^{x-1}$$

$$= \frac{1-(1-\theta)^{k+1}}{\theta} + \theta \sum_{x=0}^{k} \frac{\partial}{\partial\theta} (1-\theta)^x$$

$$= \frac{1-(1-\theta)^{k+1}}{\theta} + \theta \frac{d}{d\theta} \sum_{x=0}^{k} (1-\theta)^x$$

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

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

by quotient rule: $f(x) = \frac{u(x)}{v(x)}$; $f'(x) = \frac{v(x)u'(x) - u(x)v'(x)}{(v(x))^2}$

$$= (k+1)(1-\theta)^k$$

so, $S_k(\theta) \to 0$ as $k \to \infty$

<u>Uniform convergence</u>: Take any $[c,d] \in (0,1)$ and check

$$\sup_{\theta\in[c,d]} |S_k(\theta) - \theta| = \sup_{\theta\in[c,d]} (k+1)(1-\theta)^k \to 0 \text{ as } k \to \infty. \quad \square$$

Practice Midterm

Solve the following problems.

1. (15pt) Given three events A, B and C, which are subsets of a sample space $\mathcal{S}$. Suppose that $P(A) = 1/2$, $P(B) = 1/4$ and $P(C) = 1/8$. Assume further that A and B are pairwise independent, A and C are pairwise independent, while B and C are disjoint.

(i) (2pt) Show that A and B and C cannot be mutually independent.

(ii) (8pt) Find $P(A \cup B \cup C)$, $P(A \cup C | A \cap B)$, $P(A \cap B | A \cup C)$, and $P(B \cup C | A^c)$.

(iii) (5pt) Let N be the count of the number of events among the A, B and C that are true. Show that N is a random variable (i.e., by writing N precisely as a function of $s \in \mathcal{S}$). Find the pmf of N.

i) Mutual Ind. requires $P(A\cap B\cap C) = P(A)P(B)P(C) = 1/64$, but since $B \cap C = \emptyset$, $P(A\cap B\cap C) = 0$. Hence not mutually independent.

ii) $A \perp\!\!\!\perp B$: $P(A\cap B) = P(A)P(B) = \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8}$

$A \perp\!\!\!\perp C$: $P(A\cap C) = P(A)P(C) = \frac{1}{2}\cdot\frac{1}{8} = \frac{1}{16}$

$B \cap C = \emptyset \Rightarrow P(B\cap C) = 0$

$$P(A\cup B\cup C) \overset{\text{IE}}{=} P(A) + P(B) + P(C) - P(A\cap B) - P(A\cap C) - P(B\cap C) + P(A\cap B\cap C)$$
$$= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} - \frac{1}{8} - \frac{1}{16} - 0 + 0 = \frac{11}{16}.$$

$$P(A\cup C | A\cap B) = \frac{P((A\cup C)\cap(A\cap B))}{P(A\cap B)}$$

but $(A\cap B) \subseteq A \Rightarrow (A\cup C)\cap(A\cap B) = A\cap B$.

Hence $P(A\cup C | A\cap B) = \frac{P(A\cap B)}{P(A\cap B)} = 1$.

$$P(A\cap B | A\cap C) = \frac{P((A\cap B)\cap(A\cap C))}{P(A\cap C)} = \frac{P(A\cap B\cap C)}{P(A\cap C)} = 0$$

$$P(B\cup C | A^c) = \frac{1}{P(A^c)} \underbrace{P((B\cap A^c)\cup(C\cap A^c))}_{P(B\cap A^c) + P(C\cap A^c)}$$

since B and C disjoint $\Rightarrow B\cap A^c$ and $C\cap A^c$ are disjoint

where $P(B\cap A^c) = P(B) - P(A\cap B) = \frac{4}{16} - \frac{2}{16} = \frac{1}{8}$

$P(C\cap A^c) = P(C) - P(A\cap C) = \frac{2}{16} - \frac{1}{16} = \frac{1}{16}$

$P(A^c) = 1 - P(A) = 1 - \frac{1}{2} = \frac{8}{16}$

$\Rightarrow P(B\cup C | A^c) = \frac{3/16}{1/2} = \frac{3}{8}$.

iii) Define indicator functions $1_A, 1_B, 1_C$ by $1_A(s) = 1$ if $s \in A$ and 0 otherwise.

Then for $s \in S$: $N(s) = 1_A(s) + 1_B(s) + 1_C(s)$.

Thus, N counts how many of A, B, C occur and takes values $\{0,1,2,3\}$, so it is a RV, i.e. a measurable function $S \to \{0,1,2,3\}$.

Now compute the pmf $P_N(k) = P(N=k)$.

a) $P(N=0) = P(\text{none occur}) = 1 - P(A\cup B\cup C) = 1 - \frac{11}{16} = \frac{5}{16}$

b) $P(N=3) = P(A\cap B\cap C) = 0$ b/c $B\cap C = \emptyset$

c) $P(N=1)$: $P(\text{only } A) = P(A\cap B^c\cap C^c) = P(A) - P(A\cap B) - P(A\cap C) + P(A\cap B\cap C) = 5/16$

Similarly, $P(\text{only } B) = \cdots = 2/16$, $P(\text{only } C) = \cdots = 1/16$

$\Rightarrow P(N=1) = \frac{5}{16} + \frac{2}{16} + \frac{1}{16} = \frac{1}{2}$.

d) $P(N=2) = 1 - [P(N=0) + P(N=1) + P(N=3)] = \frac{3}{16}$

So, $\boxed{P(N=0) = \frac{5}{16},\ P(N=1) = \frac{1}{2},\ P(N=2) = \frac{3}{16}}$

2. (5pt) Suppose that 11% of men and 9% of women are left-handed. A person is chosen at random and that person is right-handed. What is the probability that that person is a woman? (Assume that there are the same number of men and women).

Let W: person is a woman

M: person is a man

RH: person is right-handed

LH: person is left-handed

Given: $P(W) = P(M) = 0.5$

$P(LH|M) = 0.11 \Rightarrow P(RH|M) = 1 - 0.11 = 0.89$

$P(LH|W) = 0.09 \Rightarrow P(RH|W) = 1 - 0.09 = 0.91$

$$P(W|RH) = \frac{P(RH|W)P(W)}{P(RH|W)P(W) + P(RH|M)P(M)} = \frac{0.5(0.91)}{0.5(0.91) + 0.5(0.89)} \simeq 0.508$$

3. (20pt) Let function f be defined as follows:

$$f(x) = \begin{cases} x & \text{if } x \in (0,1], \\ 2-x & \text{if } x \in (1,2), \\ 0 & \text{otherwise.} \end{cases}$$

(i) (4pt) Show that f is a valid pdf for a random variable X.

(ii) (6pt) What is the cdf F_X? Find $\mathbb{E}X$ and $\text{Var}(X)$.

(iii) (6pt) Let $Y = (1/\lambda)X - 1$ for $\lambda > 0$. Find the cdf of Y, $\mathbb{E}Y$ and $\text{Var}(Y)$.

(iv) (4pt) Let $Y_n = (1/n^2)X - 1$. What is the limit of the sequence of random variables Y_n as $n \to \infty$? Justify your answer.

i) check that $f(x)$ is a pdf: $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Nonnegativity: on $(0,1]$, $x > 0$, on $(1,2)$, $2-x > 0$. Elsewhere, it is 0. So $f_X \geq 0$.

Normalization:

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^1 x\,dx + \int_1^2 (2-x)\,dx = \frac{x^2}{2}\Big|_0^1 + \left(2x - \frac{x^2}{2}\right)\Big|_1^2$$
$$= \frac{1}{2} + \left(4 - \frac{4}{2} - \left(2 - \frac{1}{2}\right)\right) = \frac{1}{2} + \frac{1}{2} = 1. \blacksquare$$

ii) Find CDF for $x \in \mathbb{R}$: $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt = 0$ for $x \leq 0$

for $\underbrace{0 < x \leq 1}_{f(x)=x}$: $F(x) = \int_0^x t\,dt = \frac{t^2}{2}\Big|_0^x = \frac{x^2}{2}$, $0 < x \leq 1$

for $\underbrace{1 < x < 2}_{f(x)=2-x}$: $F(x) = \int_0^1 t\,dt + \int_1^x (2-t)\,dt$

$$= \frac{1}{2} + \left(2x - \frac{x^2}{2} - \left(2 - \frac{1}{2}\right)\right)$$
$$= -\frac{x^2}{2} + 2x - 1, \quad 1 < x < 2$$

1, $x \geq 2$

Thus, $F_X(x) = \begin{cases} 0 & , x \leq 0 \\ x^2/2 & , 0 < x \leq 1 \\ -x^2/2 + 2x - 1 & , 1 < x < 2 \\ 1 & , x \geq 2 \end{cases}$

<u>Note</u>: can check continuity at $x=1$; both give 1/2. at $x=2$, middle piece gives 1.

Next, $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x \cdot x\,dx + \int_1^2 x(2-x)\,dx$

$$= \int_0^1 x^2\,dx + \int_1^2 (2x - x^2)\,dx$$
$$= \frac{x^3}{3}\Big|_0^1 + \int_1^2 2x\,dx - \int_1^2 x^2\,dx$$
$$= \frac{1}{3} + x^2\Big|_1^2 - \frac{x^3}{3}\Big|_1^2$$
$$= \frac{1}{3} + (4-1) - \left(\frac{8}{3} - \frac{1}{3}\right)$$
$$= \frac{1}{3} + \frac{9}{3} - \frac{7}{3} = 1.$$

Now, $V(X) = E(X^2) - E^2(X)$. So,

$$E(X^2) = \int_0^1 x^2 \cdot x\,dx + \int_1^2 x^2(2-x)\,dx = \int_0^1 x^3\,dx + \int_1^2 (2x^2 - x^3)\,dx$$
$$= \frac{x^4}{4}\Big|_0^1 + \int_1^2 2x^2\,dx - \int_1^2 x^3\,dx$$
$$= \frac{1}{4} + \frac{2x^3}{3}\Big|_1^2 - \frac{x^4}{4}\Big|_1^2$$
$$= \frac{1}{4} + \left(\frac{16}{3} - \frac{2}{3}\right) - \left(\frac{16}{4} - \frac{1}{4}\right)$$
$$= \frac{1}{4} + \left(\frac{14}{3}\right) - \left(\frac{15}{4}\right)$$
$$= \frac{3}{12} + \frac{56}{12} - \frac{45}{12} = \frac{14}{12} = \frac{7}{6}.$$

$$\Rightarrow V(X) = E(X^2) - E^2(X)$$
$$= \frac{7}{6} - 1^2 = \frac{7}{6} - \frac{6}{6} = \frac{1}{6}.$$

iii) Let $Y = g(X) = \frac{1}{\lambda}X - 1$ for $\lambda > 0$.

Here $g \uparrow$ (strictly increasing) in x for $\lambda > 0$.

where $y = \frac{1}{\lambda}x - 1 \Rightarrow y + 1 = \frac{x}{\lambda} \Rightarrow x = \lambda(y+1) = g^{-1}(y)$.

Thus,

$$F_Y(y) = P(Y \le y) = F_X(g^{-1}(y)) = P(X \le \lambda(y+1))$$

where $\mathcal{X} = (0,2)$

$\mathcal{Y} = (-1, \frac{2}{\lambda} - 1)$.

Thus,

$$F_Y(y) = \begin{cases} 0 & y \le -1 \\ \lambda(y+1), & -1 < y < \frac{2}{\lambda} - 1 \\ 1, & y \ge \frac{2}{\lambda} - 1 \end{cases}$$

$E(Y) = \frac{1}{\lambda}E(X) - 1$

by L.O.E

$= \frac{1}{\lambda}(1) - 1 = \frac{1}{\lambda} - 1$

$V(Y) = \left(\frac{1}{\lambda}\right)^2 V(X) = \frac{1}{6\lambda^2}$.

iv) Let $Y_n = \frac{1}{n^2}X - 1$.

Find $\lim_{n\to\infty} Y_n = \lim_{n\to\infty} \frac{1}{n^2}X - 1$,

where $\lim_{n\to\infty} \frac{1}{n^2} = 0$

$= (0)X - 1 = -1$.

4. (10pt) Let function f be given as follows, for $x \in (-\infty, +\infty)$,

$$f(x) = \frac{1}{4}e^{-|x|} + \frac{e^{-x}}{2(1+e^{-x})^2}.$$

(i) (3pt) Show that f is a valid pdf for a random variable X.

(ii) (3pt) Show that the moment generating function exists for a neighborhood of 0 (that is, $M_X(t) < \infty$ for $t \in (-t_0, t_0)$ for some $t_0 > 0$).

(iii) (4pt) For $t \in (-t_0, t_0)$ from part (ii), find the integral expression for $\frac{d}{dt}M_X(t)$, and evaluate it at $t = 0$. Justify your operations.

i) Split the integral into two terms:

1) $\int_{-\infty}^{\infty} \frac{1}{4}e^{-|x|}\,dx = \frac{1}{4}\int_{-\infty}^{\infty} e^{-|x|}\,dx$

where $\int_{-\infty}^{\infty} e^{-|x|}\,dx = 2$

$= \frac{1}{4}(2) = 1/2$

2) $\int_{-\infty}^{\infty} \frac{e^{-x}}{2(1+e^{-x})^2}\,dx$ via u-sub

let $u = 1 + e^{-x}$ $\quad du = -e^{-x}dx$

$-du = e^{-x}dx$

$x = -\infty : u = 1 + e^{\infty} = \infty$

$x = \infty : u = 1 + \frac{1}{e^{\infty}} = 1$

$= \frac{1}{2}\int_{\infty}^{1} \frac{1}{u^2}(-du) = \frac{1}{2}\int_1^{\infty} u^{-2}\,du = \frac{1}{2}\left(\frac{-1}{u}\right)\Big|_1^{\infty}$

$= \frac{1}{2}(0+1) = 1/2$

Thus, f integrates to 1 and all terms $\ge 0 \Rightarrow$ valid pdf.

ii) Recall $f(x) = \frac{1}{4}e^{-|x|} + \frac{e^{-x}}{2(1+e^{-x})^2}$, $x \in \mathbb{R}$.

A simple uniform bound (easy to verify pointwise) is

$$\frac{e^{-x}}{2(1+e^{-x})^2} \le \frac{1}{2}e^{-|x|}$$

So for all x

$$f(x) \le \frac{1}{4}e^{-|x|} + \frac{1}{2}e^{-|x|} = \frac{3}{4}e^{-|x|}!$$

Fix any t with $|t| < \frac{1}{2} \Leftrightarrow -\frac{1}{2} < t < \frac{1}{2} \Leftrightarrow 1 - |t| > \frac{1}{2}$

Then, for all x

$$|e^{tx}f(x)| \le e^{|t|}e^{|x|}f(x) \le \frac{3}{4}e^{|t||x|}e^{-|x|}$$
$$= \frac{3}{4}e^{-(1-|t|)|x|} \le \frac{3}{4}e^{-(1/2)|x|}$$

The function is integrable on $\mathbb{R}$, i.e.

$$\int_{-\infty}^{\infty} |e^{tx}f(x)|\,dx < \infty,$$

so $M_X(t) = \int e^{tx}f(x)\,dx$ exists and is finite for every $|t| < 1/2$.

iii) For $|t| < 1/2$, we want to justify

$$M_X'(t) = \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx}f(x)\,dx = \int_{-\infty}^{\infty} x e^{tx}f(x)\,dx$$

using the same bound as above, i.e. $|t| < 1/2$

$$|xe^{tx}f(x)| \le |x|e^{|t||x|}f(x) \le \frac{3}{4}|x|e^{-(1-|t|)|x|}.$$

The function $|x|e^{-(1/2)|x|}$ is integrable and thus, the dominated convergence holds.

Therefore, we may interchange differentiation & integration.

Thus,

for $|t| < 1/2$, $M_X'(t) = \int_{-\infty}^{\infty} x e^{tx}f(x)\,dx$

Here, $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}f(x)\,dx$

$M_X'(t) = \frac{d}{dt}M_X(t) = \int_{-\infty}^{\infty} x e^{tx}f(x)\,dx$

$M_X'(0) = \int_{-\infty}^{\infty} x f(x)\,dx = E(X)$

where $f = \underbrace{\frac{1}{4}e^{-|x|}}_{f_1(-x) = f_1(x)} + \underbrace{\frac{e^{-x}}{2(1+e^{-x})^2}}_{f_2(-x) = f_2(x)}$

Thus, $f(x) = f(-x)$, i.e. even whereas $(-x) = -x$, i.e. odd.

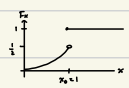
So, $xf(x)$ is odd since $(xf(x))(-x) = (-x)f(-x) = -[xf(x)]$

The integral of an odd integrable function over $\mathbb{R}$ is 0.

Hence, $M_X'(0) = E(X) = 0$.

Midterm 2 Solutions

1) Let F be defined as follows: $F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}x^2, & x \in [0,1) \\ 1, & x \ge 1 \end{cases}$



i) To show F is valid: a) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$

b) $\frac{d}{dx}F(x) = \frac{1}{2}\frac{d}{dx}(x^2) = 2x\frac{1}{2} = x > 0 \;\; \forall x \in \mathcal{X} \Rightarrow$ increasing

$\Rightarrow$ nondecreasing.

c) Right-continuity holds: $\lim_{x \downarrow x_0} F(x) = F(x_0) \;\; \forall x \in \mathcal{X}$ for $x_0 = 1$

ii) $f(x) = \frac{d}{dx}F(x) = x$, $x \in [0,1)$

$\Rightarrow \underline{f(x) = x,\ x \in [0,1) \text{ and } P(X=1) = \frac{1}{2}}$

(mixture of cont. and discrete)

where there is a jump discontinuity in the CDF which correspond to point masses in the distribution (pdf)

Here, $F(1^-) = \lim_{x\to 1^-} \frac{1}{2}x^2 = \frac{1}{2}$ vs. $F(1) = 1$

$\Rightarrow \Delta = F(1) - F(1^-) = 1 - \frac{1}{2} = \frac{1}{2}$.

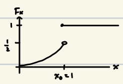
So, even though the derivative $f(x)$ is zero (or undefined) at a jump point from the calculus perspective, from a prob. perspective, we assign a delta mass to the discontinuity

Then, $E(X) = \int_0^1 x \cdot f(x)\,dx + x_0 P(X = x_0)$ where $x_0 = 1$: discontinuity point

$$= \int_0^1 x^2 dx + P(X=1) = \frac{x^3}{3}\Big|_0^1 + \frac{1}{2} = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$$

Similarly, $E(X^2) = \int_0^1 x^3 f(x)\,dx + \frac{1}{2} = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$

$$\Rightarrow V(X) = E(X^2) - E^2(X) = \frac{3}{4} - \left(\frac{5}{6}\right)^2 = \frac{27}{36} - \frac{25}{36} = \frac{1}{18}$$

iii) 

Inverse Function: $\boxed{F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}\ \forall\ y \in (0,1)}$

where $\cdot\ F_X^{-1}(1) := +\infty$ only if $F_X(x) < 1\ \forall x$.

$\cdot\ F_X^{-1}(0) := -\infty$

Recall the definition: $F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}$

For $0 \leq p < 1/2$, we can solve

$$\frac{1}{2}x^2 = p \Rightarrow x^2 = 2p \Rightarrow x = \sqrt{2p}$$

So, the smallest x s.t. $F(x) \geq p$, i.e.

$$F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\} = \sqrt{2p}$$

For $1/2 < p \leq 1$,

for $x < 1$: $F(x) < 1/2$

at $x = 1$: $F(1) = 1$

$\Rightarrow \forall\ p \in (1/2, 1]$, the first time $F(x) \geq p$ is exactly at $x = 1$, i.e.

$$F^{-1}(p) = \inf\{x : F(x) \geq p\} = 1.$$

Thus,

$$F_X^{-1}(p) = \begin{cases} -\infty, & p < 1 \\ \sqrt{2p}, & 0 \leq p < 1/2 \\ 1, & 1/2 \leq p \leq 1 \\ \infty, & p > 1 \end{cases}$$

If $0 \leq x < 1$, since $F(x) = \frac{1}{2}x^2 < \frac{1}{2}$,

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

If $x \geq 1$, since $F(x) = 1$,

$$F^{-1}(F(x)) = F^{-1}(1) = 1$$

Therefore, $F^{-1}(F(x))$ holds for $x \in [0,1]$

2) A function f on $\mathbb{R}$ is called symmetric about a if $f(a+x) = f(a-x)$ for all $x > 0$.

a) Show that if the pdf of a RV X is symmetric around a, and $E(X)$ exists, then $E(X) = a$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{-\infty} (x - a + a) f(x)\,dx$$

$$= a + \int_{-\infty}^{\infty} (x-a) f(x)\,dx$$

$$= a + \int_{-\infty}^{a} (x-a) f(x)\,dx + \int_{a}^{\infty} (x-a) f(x)\,dx$$

let $y = a - x$ change-of-var

$\Rightarrow x = a - y$

$\Rightarrow dx = -dy$

$$= a - \int_0^{\infty} y f(a-y)\,dy + \int_0^{\infty} y f(a+y)\,dy$$

$$= a + \int_0^{\infty} y\,(\underbrace{f(a+y) - f(a-y)}_{0})\,dy$$

$$= a.$$

b) Show that the pdf for an exponential dist. w/ parameter $\lambda > 0$:

$f(x|\lambda) = \frac{1}{\lambda}e^{-x/\lambda}$, $x > 0$ and $f = 0$ otherwise, is not symmetric.

If f is symmetric, it is symmetric around $E(X) = \lambda$. Then, it must hold that:

$$f(\lambda - x) = f(\lambda + x)\quad \forall\ x > 0.$$

However, it is obvious that the equation doesn't hold for $x > \lambda$,

since $f(\lambda - x) = 0$ and $f(\lambda + x) > 0$

Alternatively, and more clearly (to me),

$$f(a+x) = \frac{1}{\lambda}\exp\left(-\frac{(a+x)}{\lambda}\right) = \frac{1}{\lambda}\exp\left(\frac{-a-x}{\lambda}\right)$$

$$f(a-x) = \frac{1}{\lambda}\exp\left(\frac{-(a-x)}{\lambda}\right) = \frac{1}{\lambda}\exp\left(\frac{-a+x}{\lambda}\right)$$

where $f(a+x) \neq f(a-x) \Rightarrow f(x|\lambda)$ is not symmetric

c) Let b be the median of an exponential RV, i.e. b that satisfies

$P(X > b) = P(X < b) = 1/2$. Show that $E(X) > b$:

$$P(X > b) = \int_b^{\infty} \frac{1}{\lambda} e^{-x/\lambda} dx$$

where $\frac{d}{dx}\left[e^{-x/\lambda}\right] = \frac{-1}{\lambda}e^{-x/\lambda}$

$$= \frac{1}{\lambda}(-\lambda)e^{-x/\lambda}\Big|_{x=b}^{x=\infty}$$

$$= -\left(-e^{-b/\lambda}\right)$$

$$= e^{-b/\lambda} = \frac{1}{2}. \Rightarrow e^{b/\lambda} = 2 \Rightarrow \frac{b}{\lambda} = \log 2$$

So, the median is $b = \log(2) \cdot \lambda$.

We know $E(X) = \lambda > \log 2 \cdot \lambda$. ∎

3) construct a sequence of non-identically distributed RV's $(X_n)_{n=1}^{\infty}$ that converges to a RV Y in dist. in the following settings (specify both X_n and Y and briefly justify your answer)

For justification, can prove either the mgf or cdf converges.

requires bounded support, i.e. $\mathcal{X} = \{x : f(x) > 0\}$

i) For each n, X_n is a discrete RV that takes two possible values, and so is Y

<u>Two-Point discrete</u>: Let $X_n \sim \text{Bernoulli}(p_n)$ with $p_n \to p \in [0,1]$ and let $Y \sim \text{Bernoulli}(p)$.

Then, the mgf of X_n converges to mgf of Y, i.e.

$$M_{X_n}(t) = 1 - p_n + p_n e^t \to 1 - p + pe^t = M_Y(t).$$

Hence,

$$X_n \xrightarrow{d} Y.$$

ii) For each n, X_n is an exponential RV, and so is Y.

<u>Exponential</u>: Let $X_n \sim \text{Exp}(\lambda_n)$ with $\lambda_n \to \lambda > 0$ and let $Y \sim \text{Exp}(\lambda)$.

we know $f(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda}$ for $X \sim \text{Exp}(\lambda)$, $x \ge 0$.

$$\Rightarrow F_X(x) = \int_0^x \frac{1}{\lambda} e^{-t/\lambda} dt$$

where $\frac{d}{dt} e^{-t/\lambda} = e^{-t/\lambda}(-1/\lambda)$

$$= \frac{1}{\lambda}\left[(-\lambda) e^{-t/\lambda} \Big|_0^x\right]$$

$$= -e^{-t/\lambda} + 1$$

so, for $x \ge 0$

$$F_{X_n}(x) = 1 - e^{-\lambda_n x} \to 1 - e^{-\lambda x} = F_Y(x)$$

Hence, $X_n \xrightarrow{d} Y$

iii) For each n, X_n is a discrete RV that takes n possible distinct values

<u>n-point discrete</u>: Let X_n be uniform on the grid $\{1/n, 2/n, \dots, n/n\}$:

$$P\left(X_n = \frac{k}{n}\right) = \frac{1}{n} \quad , k = 1, \dots, n.$$

To find $F_{X_n}(x) = P(X_n \le x)$,

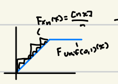
find all $k \in \{1, \dots, n\}$ s.t. $\frac{k}{n} \le x$.

This implies

$$k \le nx$$

$\Rightarrow$ number of such $k = [nx]$

where $[nx] = \max\{k \in \mathbb{Z} \mid k \le nx\}$



Thus, $F_{X_n}(x) = \frac{[nx]}{n}$, i.e.

a step function that jumps at each grid point k/n, with each jump size $1/n$.

For $x \in (0,1)$,

$$\frac{[nx]}{n} \le x \le \frac{[nx]+1}{n}$$

$$\Leftrightarrow \quad 0 \le x - \frac{[nx]}{n} \le \frac{1}{n}$$

$$\Rightarrow |F_{X_n}(x) - x| = \left|\frac{[nx]}{n} - x\right| \le \frac{1}{n} = 0 \text{ as } n \to \infty$$

$$\Rightarrow F_{X_n}(x) = \frac{[nx]}{n} \to x = F_Y(x)$$

$$\Rightarrow X_n \xrightarrow{d} Y \sim \text{Unif}(0,1)$$

(Alternatively, use Poisson approximation from $X_n \sim$ Bernoulli)

4) Let $n > 0$ and function f be given as follows:

$$\text{for } x \in (-\infty, \infty), \quad f(x) = \underbrace{\frac{n}{4} e^{-n|x-1|}}_{f_1(x)} + \underbrace{\frac{\sqrt{n}}{2\sqrt{2\pi}} e^{-\frac{n}{2}x^2}}_{f_2(x)}$$

i) show that f is a valid pdf for a RV X

check two conditions: nonegativity and normalization

clearly $f(x) \ge 0 \;\; \forall x$ b/c both terms are positive

$$\text{compute} \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{n}{4} e^{-n|x-1|} dx + \int_{-\infty}^{\infty} \frac{\sqrt{n}}{2\sqrt{2\pi}} e^{-\frac{n}{2}x^2} dx$$

$$\text{where} \int_{-\infty}^{\infty} f(|x-a|)\,dx = \int_{-\infty}^{a} f(\underbrace{a-x}_{x<a})\,dx + \int_a^{\infty} f(\underbrace{x-a}_{a<x})\,dx$$

$$\Rightarrow \int_{-\infty}^{\infty} \frac{n}{4} e^{-n|x-1|} dx = \frac{n}{4}\left(\underbrace{\int_{-\infty}^{1} e^{-n(\underbrace{1-x}_{x<1})} dx}_{A} + \underbrace{\int_1^{\infty} e^{-n(\underbrace{x-1}_{1<x})} dx}_{B}\right)$$

where let $u = 1-x$ $du = -dx$

$x = -\infty \to u = \infty$
$x = 1 \to u = 0$

$$\Rightarrow \int A = \int_0^{\infty} e^{-nu} du = \frac{-1}{n} e^{-nu} \Big|_{u=0}^{u=\infty} = 0 + \frac{1}{n} = \frac{1}{n}$$

Similarly, let $u = x-1$ and $\int B = \frac{1}{n}$

$$= \frac{n}{4}\left(\frac{1}{n} + \frac{1}{n}\right) = \frac{1}{2}.$$

For $\int f_2(x)\,dx$, recall: $\int_{-\infty}^{\infty} e^{-\frac{n}{2}x^2} dx = \sqrt{\frac{2\pi}{n}}$ (using density of normal dist.)

$$\text{Hence} \int_{-\infty}^{\infty} \frac{\sqrt{n}}{2\sqrt{2\pi}} e^{-\frac{n}{2}x^2} dx = \frac{\sqrt{n}}{2\sqrt{2\pi}} \sqrt{\frac{2\pi}{n}} = \frac{1}{2}$$

ii) show that the mgf exists in a neighborhood of 0, i.e.

show $M_X(t) < \infty$ for $t \in (-t_0, t_0)$ for some $t_0 > 0$

where $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$

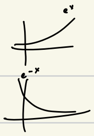
$$|e^{tx}f(x)| = |e^{tx}f_1(x)f_2(x)| = e^{tx}\left[\frac{n}{4}e^{-n|x-1|} + \frac{\sqrt{n}}{2\sqrt{\pi}}e^{-\frac{n}{2}x^2}\right]$$

$$\le e^{(t+t_0)x}\,\frac{n}{4}e^{-n|x-1|} + \frac{\sqrt{n}}{2\sqrt{\pi}}e^{-\frac{n}{2}x^2} \quad \text{for } t_0>0$$

should show cases for $\overset{t<n-a}{x>1}$, $\overset{t>-n+a}{x<1}$ where $t_0 = n-a$, $a<n$.

FS

$$\boxed{\int e^{ax}dx \text{ integrable i.f.f. } a<0}$$



and $\int_{-\infty}^{\infty}|e^{tx}f(x)|\,dx$ is integrable over $\mathbb{R}$ for $t_0>0$. i.f.f. $e^{(t+t_0)x}e^{-n|x-1|}<0$ since

$\Rightarrow M_X(t)<\infty$ for $t\in(-t_0,t_0)$ for some $t_0>0$. ∎

Alternatively, $M_X(t) = \int_{-\infty}^{\infty} e^{tx}f(x)\,dx = I_1(t) + I_2(t)$

where $I_1(t) = \int_{-\infty}^{\infty} e^{tx}\frac{n}{4}e^{-n|x-1|}dx$, splitting at $x=1$

$$= \frac{n}{4}\left(\int_{-\infty}^{1} e^{tx}e^{-n(1-x)}dx + \int_{1}^{\infty} e^{tx}e^{-n(x-1)}dx\right)$$

$$= \frac{n}{4}\left(e^{-n}\underbrace{\int_{-\infty}^{1} e^{(t+n)x}dx}_{a} + e^{n}\underbrace{\int_{1}^{\infty} e^{(t-n)x}dx}_{b}\right)$$

were each integral converges when $|t|<n$: $-n<t<n$ where $|t| = \begin{cases} t, & t\ge 0 \\ -t, & t<0 \end{cases}$ and if $|t|<n$,

then both positive/negative directions of $t<n$

$\Rightarrow t\in(-n,n)$

$\Leftrightarrow -n<t<n$

$$\int_{-\infty}^{1} a = \frac{e^{(t+n)}}{t+n}, \quad \int_{1}^{\infty} b = \frac{e^{(t-n)}}{n-t}$$

Substitute:

$$I_1(t) = \frac{n}{4}\left(e^{-n}\frac{e^{t+n}}{t+n} + e^{n}\frac{e^{t-n}}{n-t}\right) = \frac{ne^t}{4}\left(\frac{1}{t+n}+\frac{1}{n-t}\right) = \frac{n^2e^t}{2(n^2-t^2)}$$

where $I_1(t)$ finite for $|t|<n$.

Then, $I_2(t) = \frac{\sqrt{n}}{2\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}e^{-\frac{n}{2}x^2}dx$

completing the square:

$$tx - \frac{n}{2}x^2 = -\frac{n}{2}\left(x^2 - \frac{2t}{n}x\right) = -\frac{n}{2}\left(x-\frac{t}{n}\right)^2 + \frac{t^2}{2n}$$

Hence

$$I_2(t) = \frac{\sqrt{2}}{2\sqrt{2\pi}}e^{\frac{t^2}{2n}}\int_{-\infty}^{\infty} e^{-\frac{n}{2}\left(x-\frac{t}{n}\right)^2}dx$$

$$= \frac{\sqrt{2}}{2\sqrt{2\pi}}e^{t^2/2n}\sqrt{\frac{2\pi}{n}}, \quad a: \int_{-\infty}^{\infty} e^{-\frac{n}{2}x^2}dx = \sqrt{\frac{2\pi}{n}} \Rightarrow \int_{-\infty}^{\infty} e^{-\frac{n}{2}\left(x-\frac{t}{n}\right)^2}dx = \sqrt{\frac{2\pi}{n}}\,?$$

$$= \frac{1}{2}e^{t^2/2n}$$

Thus,

$$M_X(t) = I_1(t) + I_2(t) = \frac{n^2e^t}{2(n^2-t^2)} + \frac{1}{2}e^{\frac{t^2}{2n}}$$

is finite for all $|t|<n$. Therefore, mgf exists in neighborhood of 0.

iii) For $t\in(-t_0,t_0)$ from (ii), find the integral expression for $\frac{d}{dt}M_X(t)$ and evaluate it at $t=0$. Justify your operations, e.g. when interchanging the order of integrals and derivatives

<u>My Initial Approach</u>: $M_X'(t) = \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx}f(x)\,dx \overset{(*)}{=} \int_{-\infty}^{\infty} xe^{tx}f(x)\,dx$

Since $|e^{tx}f(x)| \le e^{(t+t_0)x}\left[\frac{n}{4}e^{-n|x-1|} + \frac{\sqrt{n}}{2\sqrt{\pi}}e^{-\frac{n}{2}x^2}\right]$ for $t_0>0$,

justifying (*):

$$|xe^{tx}f(x)| \le |x|\,e^{(t+t_0)x}\left[\frac{n}{4}e^{-n|x-1|} + \frac{\sqrt{n}}{2\sqrt{\pi}}e^{-\frac{n}{2}x^2}\right]$$

and $\int_{-\infty}^{\infty}|xe^{tx}f(x)|\,dx$ integrable over $\mathbb{R}$

Thus, $\frac{d}{dt}M_X(t) = M_X'(t)$ properly defined

and permits interchanging the order of integrals and derivatives (why?)

Now, $M_X'(t)|_{t=0} = M_X'(0) = \int_{-\infty}^{\infty} xe^{tx}f(x)\,dx\Big|_{t=0} = \int_{-\infty}^{\infty} xf(x)\,dx = E(X)$

where $\int_{\infty}^{\infty} f(x)\,dx = \frac{n}{4}\underbrace{\int_{-\infty}^{\infty} xe^{-n|x-1|}dx}_{a} + \frac{\sqrt{n}}{2\sqrt{2\pi}}\underbrace{\int_{-\infty}^{\infty} xe^{-\frac{n}{2}x^2}dx}_{b}$

recall: even := $f(x)=f(-x)$ vs. odd := $f(-x) = -f(x)$

e.g. $|x|: |x| = |-x|$.

$\hookrightarrow -xe^{-\frac{n}{2}(-x)^2} = -xe^{-\frac{n}{2}x^2}$

where b contains odd integrand $\Rightarrow \int b = 0$

Now, for $\int_{\mathbb{R}} xe^{-n|x-1|}dx$

letting $y = x-1$, $dy = dx$

$$= \int_{\mathbb{R}}(y+1)e^{-n|y|}dy = \underbrace{\int_{\mathbb{R}} ye^{-n|y|}dy}_{\substack{y\text{-odd};\\ \text{odd}\times\text{even}=\text{odd}}} \overset{0}{} + \int_{\mathbb{R}} e^{-n|y|}dy$$

$$= \int_{-\infty}^{0} e^{ny}dy + \int_{0}^{\infty} e^{-ny}dy$$

$$= \frac{1}{n}e^{ny}\Big|_{-\infty}^{0} + \frac{1}{n}e^{-ny}\Big|_{0}^{\infty}$$

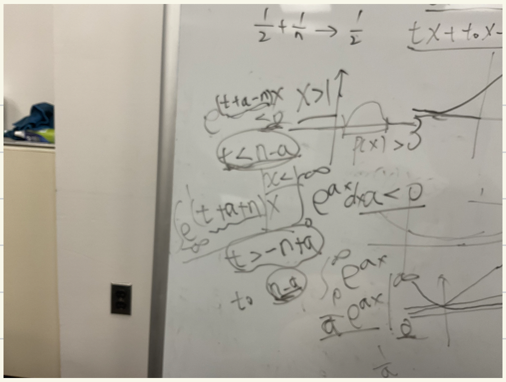
$$= \frac{2}{n}$$

Therefore, $M_X'(0) = \frac{n}{4}\cdot\frac{2}{n} + 0 = \frac{1}{2}$

iv) Let X_n: RV w/ pdf f. What does X_n converge in dist. to as $n\to\infty$?

From (iii), for $|t|<n$,

$$M_X(t) = \frac{n^2e^t}{2(n^2-t^2)} + \frac{1}{2}\exp\left(\frac{t^2}{2n}\right).$$



Practice Problems (up to Final)

3.1: Discrete Distributions

Ex(Uniform): Let $X \sim \text{Unif}(1,N)$. Suppose $P(X=n)=\frac{1}{N}$, $n=1,\dots,N$.

Then $E(X)=\sum_{n\in N} nP(X=n)=\frac{1}{N}(1+\dots+N)\overset{(*)}{=}\frac{1}{N}\frac{N(N+1)}{2}=\frac{N+1}{2}$;

Justifying $(*)$ as given $1+\dots+N$, pair terms from the end, i.e.
$(1+N)+(2+(N-1))+\cdots$

s.t. $\left\{\begin{array}{ll} N/2 \text{ pairs} & , N\text{-even} \\ (N-1)/2 \text{ pairs} + \text{middle term} = \frac{N+1}{2} \text{ pairs} & , N\text{-odd}\end{array}\right] \Rightarrow \frac{N}{2}(N+1)$

Thus, $\sum_{k=1}^{N} k=\frac{N(N+1)}{2}$ (FS)

$E(X^2)=\sum_{n\in N} n^2P(X=n)=\frac{1}{N}(1^2+\dots+N^2)\overset{(*)}{=}\frac{1}{N}\frac{N(N+1)(2N+1)}{6}=\frac{(N+1)(2N+1)}{6}$;

Justifying $(*)$, $(k+1)^3-k^3=3k^2+3k+1$.

Sum this identity for $k=1$ to N, i.e.

$(N+1)^3-1^3=\sum_{k=1}^{N}(3k^2+3k+1)=3\sum_{k=1}^{N}k^2+3\sum_{k=1}^{N}k+\sum_{k=1}^{N}1$

Plugging in $\sum_{k=1}^{N}k=\frac{N(N+1)}{2}$ and $\sum_{k=1}^{N}1=N$, solve for k^2:

FS $\sum_{k=1}^{N}k^2=\frac{N(N+1)(2N+1)}{6}$

$V(X)=E(X^2)-E^2(X)=\frac{(N+1)(2N+1)}{6}-\left(\frac{N+1}{2}\right)$

$=\frac{2(N+1)(2N+1)}{12}-\frac{3(N+1)^2}{12}=\frac{(N+1)}{12}[2(2N+1)-3(N+1)]$

$=\frac{(N+1)}{12}[4N+2-3N-3]=\frac{(N+1)(N-1)}{12}=\frac{(N-1)^2}{12}$

Ex(Hypergeometric): $E(X)=\sum_{x=0}^{K}x\binom{M}{x}\binom{N-M}{K-x}\frac{1}{\binom{N}{K}}$
$\underset{\text{Expectation}}{\shortparallel}$

FS where $x\binom{M}{x}=M\binom{M-1}{x-1}$ and $\frac{\binom{N-1}{K-1}}{\binom{N}{K}}=\frac{\frac{(N-1)!}{(K-1)!(N-K)!}}{\frac{N!}{K!(N-K)!}}=\frac{(N-1)!}{(K-1)!(N-K)!}\cdot\frac{K!(N-K)!}{N!}=\frac{(N-1)!K!}{(K-1)!N!}=\binom{K}{N}$

$=\sum_{x=1}^{K}M\binom{M-1}{x-1}\binom{N-M}{K-x}\left(\frac{N}{K}\binom{N-1}{K-1}\right)^{-1}$

$=\frac{KM}{N}\sum_{x=1}^{K}\binom{M-1}{x-1}\binom{N-M}{K-x}\Big/\binom{N-1}{K-1}$

letting $y=x-1 \Rightarrow -x=-y-1$

$=\frac{KM}{N}\sum_{y=0}^{K-1}\binom{M-1}{y}\binom{(N-1)-(M-1)}{K-1-y}\Big/\binom{N-1}{K-1}$

$=\frac{KM}{N}\sum_{y=0}^{K-1}P(Y=y\mid N-1,M-1,K-1)$

$=\frac{KM}{N}$

Ex(Hypergeometric): Let $N=25$ items, $M=\#$ defective items
we sample K items from the lot, say $K=10$, and find that none are defective.

Q: what can we say about M?

From above, $X\mid N,M,K \sim \text{Hypergeometric}(N,M,K)$

Recall $p(x)=\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$, $x=0,1,\dots,K$.

So, $P(X=0\mid N=25,M=6,K=10)=\frac{\binom{M}{0}\binom{N-M}{K-0}}{\binom{N}{K}}=\frac{\binom{6}{0}\binom{19}{10}}{\binom{25}{10}}=0.028$.

$\Rightarrow P(X=0\mid N=25,K=10,M\geq 6)\leq 0.028$, i.e.
if $M\geq 6$, then the observed event "$X=0$" is highly unlikely

Ex(Bernoulli): Let $X\sim\text{Bernoulli}(p)$, $0\leq p\leq 1$.

$E(X)=\sum_{x\in\mathcal{X}}x\cdot p(x)=0\cdot(1-p)+1\cdot p=p$.

$E(X^2)=\sum_{x\in\mathcal{X}}x^2p(x)=0^2\cdot(1-p)+1^2p=p$

$V(X)=E(X^2)-E^2(X)=p-p^2=p(1-p)$.

Ex(Binomial Expectation): Let $X\sim\text{Binomial}(n,p)$
$\underset{\text{as seen in (2.2)}}{\shortparallel}$

Then,
$E(X)=\sum_{x=0}^{n}x\binom{n}{x}p^x(1-p)^{n-x}$

$=\sum_{x=1}^{n}x\binom{n}{x}p^x(1-p)^{n-x}$, $x=0\Rightarrow$ entire product $=0$

$=\sum_{x=1}^{n}n\binom{n-1}{x-1}p^x(1-p)^{n-x}$
where $x\binom{n}{x}=n\binom{n-1}{x-1}$

$=np\sum_{x=1}\binom{n-1}{x-1}p^{x-1}(1-p)^{n-x}$

letting $y=x-1\Rightarrow -x=-1-y\Rightarrow x=y+1$

$=np\sum_{y=0}^{n-1}\binom{n-1}{y}p^y(1-p)^{n-1-y}$

$=np(p+\cancel{(1-p)})^{n-1}=np$.

Ex(Binomial RV Variance): Let $X\sim\text{Binomial}(n,p)=\binom{n}{x}p^x(1-p)^x$, $n\in\mathbb{N}$, $p\in(0,1)$
$\underset{\text{as seen in (2.3)}}{\shortparallel}$

with $E(X)=np$; show that $V(X)=E(X^2)-E^2(X)=np(1-p)$.

$E(X^2)=\sum_{x=0}^{n}x^2f(x)=\sum_{x=0}^{n}x^2\binom{n}{x}p^x(1-p)^{n-x}$
where $x\binom{n}{x}=n\binom{n-1}{x-1}$

$=\sum_{x=1}^{n}xn\binom{n-1}{x-1}p^x(1-p)^{n-x}$,

letting $y=x-1\Rightarrow x=y+1$

$=\sum_{y=0}^{n-1}n(y+1)\binom{n-1}{y}p^{y+1}(1-p)^{n-y-1}$

$=np\times\left[\sum_{y=0}^{n-1}y\binom{n-1}{y}p^y(1-p)^{n-y-1}+\sum_{y=0}^{n-1}\binom{n-1}{y}p^y(1-p)^{n-y-1}\right]$

where $nx(x+a)^n=\sum_{k=0}^{n}x\binom{n}{k}x^ka^{n-k}$

$=np[(n-1)p(p+(1-p))^{n-1}+(1)(p+(1-p))^{n-1}]$

$=np[(n-1)p+1]=n(n-1)p^2+np$.

Hence,
$V(X)=E(X^2)-E^2(X)=n(n-1)p^2+np-(np)^2$

$=n^2p^2-np^2+np-n^2p^2$

$=np-np^2=np(1-p)$. ∎

3.6 A large number of insects are expected to be attracted to a certain variety of rose plant. A commercial insecticide is advertised as being 99% effective. Suppose 2,000 insects infest a rose garden where the insecticide has been applied, and let X = number of surviving insects.

(a) What probability distribution might provide a reasonable model for this experiment?

(b) Write down, but do not evaluate, an expression for the probability that fewer than 100 insects survive, using the model in part (a).

(c) Evaluate an approximation to the probability in part (b).

a) According to the setup, $X=\#$ surviving insects

We can reasonably assume that $X\sim\text{Binomial}(n=2000,p=0.01)$ since:
- Fixed number of trials $(n=2000)$
- Each trial has same prob. p
- Each trial independent
- Binary outcomes (survive/not)

where p: Prob(survival).

$\Rightarrow P(X=x\mid n,p)=\binom{n}{x}p^x(1-p)^{n-x}$

b) $P(X<100)=\sum_{x=0}^{99}\binom{2000}{x}(0.01)^x(0.99)^{2000-x}$

c) Alternatively, can approximate Binomial(2000, 0.01) by Poisson(λ),

letting $np\to\lambda$, i.e. $2000(0.01)=20\to\lambda$ as $n\to\infty$
(where $n=200$ is reasonably large).

Thus, $P(X<100)\approx P(\text{Poisson}(20)<100)$

$=1-P(\text{Poisson}(20)\geq 100)$

$\approx 1-0$

≈ 1 (same approximating result as when using Normal approx.)

3.13 A *truncated* discrete distribution is one in which a particular class cannot be observed and is eliminated from the sample space. In particular, if X has range $0,1,2,\dots$ and the 0 class cannot be observed (as is usually the case), the 0-*truncated* random variable X_T has pmf

$$P(X_T=x)=\frac{P(X=x)}{P(X>0)},\quad x=1,2,\dots.$$

Find the pmf, mean, and variance of the 0-truncated random variable starting from

(a) $X\sim\text{Poisson}(\lambda)$.

(b) $X\sim\text{negative binomial}(r,p)$, as in (3.2.10).

For any X with support $0,1,\dots$, we have the mean and variance of the 0-truncated X_T:

$E(X^T)=\sum_{x=1}^{\infty}xP(X_T=x)=\sum_{x=1}^{\infty}x\frac{P(X=x)}{P(X>0)}$

$=\frac{1}{P(X>0)}\sum_{x=1}^{\infty}xP(X=x)$

$=\frac{1}{P(X>0)}\sum_{x=0}^{\infty}xP(X=x)=\frac{E(X)}{P(X>0)}$

Similarly, $E(X_T^2)=\frac{E(X^2)}{P(X>0)}$.

Thus, $V(X_T)=E(X_T^2)-E^2(X_T)=\frac{E(X^2)}{P(X>0)}-\left(\frac{E(X)}{P(X>0)}\right)^2$

a. For Poisson(λ), $P(X>0)=1-P(X=0)=1-\frac{e^{-\lambda}\lambda^0}{0!}=1-e^{-\lambda}$.

Therefore, $P(X_T=x)=\frac{e^{-\lambda}\lambda^x}{x!(1-e^{-\lambda})}$, $x=1,2,\dots$

Thus, $E(X_T)=\lambda/(1-e^{-\lambda})$ and
$V(X_T)=\frac{\lambda^2+\lambda}{1-e^{-\lambda}}-\left(\frac{\lambda}{1-e^{-\lambda}}\right)^2$

b. For Negative Binomial(r,p), $P(X>0)=1-P(X=0)$
$=1-\binom{r-1}{0}p^r(1-p)^0=1-p^r$.

Then,
$P(X_T=x)=\frac{\binom{r+x-1}{x}p^r(1-p)^x}{1-p^r}$, $x=1,2,\dots$

$E(X_T)=\frac{r(1-p)}{p(1-p^r)}$

$V(X_T)=\frac{r(1-p)+r^2(1-p)^2}{p^2(1-p^r)}-\left[\frac{r(1-p)}{p(1-p^r)}\right]^2$

3.15 In Section 3.2 it was claimed that the Poisson(λ) distribution is the limit of the negative binomial(r,p) distribution as $r\to\infty$, $p\to 1$, and $r(1-p)\to\lambda$. Show that under these conditions the mgf of the negative binomial converges to that of the Poisson.

The mgf for the negative binomial is $M(t)=\left(\frac{p}{1-(1-p)e^t}\right)^r$. Let $\begin{cases}q=1-p\Rightarrow p=1-q\\ rq\to\lambda \text{ as } r\to\infty,\ q\to 0\end{cases}$

where $\frac{1-q}{1-qe^t}=\left[1+\frac{(1-q)-(1-qe^t)}{1-qe^t}\right]=\left[1+\frac{q(e^t-1)}{1-qe^t}\right]=\left[1+\frac{1}{r}\left(\frac{r(1-p)(e^t-1)}{1-(1-p)e^t}\right)\right]$

where $\frac{r(1-p)(e^t-1)}{1-(1-p)e^t}\longrightarrow\frac{\lambda(e^t-1)}{1}=\lambda(e^t-1)$ as $r\to\infty$, $p\to 1$ and $r(p-1)\to\lambda$

and $\lim_{r\to\infty}\left[1+\frac{1}{r}\left(\frac{r(1-p)(e^t-1)}{1-(1-p)e^t}\right)\right]^r=\exp(\lambda(e^t-1))$, i.e. the mgf of Poisson(λ).

FS: $\lim_{r\to\infty}(1+\frac{a_r}{r})^r=e^{\lim a_r}$ if $a_r\to a$;

generalizes $\lim_{r\to\infty}(1+\frac{a}{r})^r=e^a$

Ex(Binomial RV): let $X \sim \text{Binomial}(n,p)$; $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0,1,\dots,n$

" as seen in (2.3)

then
$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x}$$
using the binomial formula
$$= (e^t p + (1-p))^n$$

then, $E(X^k) = M_X^{(k)}(t)\big|_{t=0}$. ∎

Ex(Poisson Exp.): let $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$.

" variance

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

FS where $\frac{x}{(x-1)!} = \frac{1}{(x-1)!}$ and $\frac{1}{(x-1)!} \lambda^x = \lambda \cdot \frac{\lambda^{x-1}}{(x-1)!}$

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

FS where $e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^x}{x!} + \dots$

$$= e^{-\lambda} \lambda e^{\lambda} = \lambda.$$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!}$$

letting $x^2 = x(x-1) + x$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \Big[\underbrace{x(x-1) \frac{\lambda^x}{x!}}_{S_1} + \underbrace{x \frac{\lambda^x}{x!}}_{S_2} \Big]$$

so $S_1 := e^{-\lambda} \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x!}$

where $\frac{x(x-1)}{x!} = \frac{(x-1)}{(x-1)!} = \frac{1}{(x-2)!}$

and $\frac{1}{(x-2)!} \lambda^x = \lambda^2 \frac{\lambda^{x-2}}{(x-2)!}$

$$= e^{-\lambda} \lambda^2 \underbrace{\sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!}}_{e^{\lambda}} = \lambda^2.$$

similarly,
$$S_2 := e^{-\lambda} \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} = e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda.$$

so, $E(X^2) = \lambda^2 + \lambda$

$$V(X) = E(X^2) - E^2(X) = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

Ex(Poisson Problem): A call operator handles, on average, 5 calls / 3 min.

Find P(no calls in the next minute).

let X : number of calls in the next minute.

Assume $X \sim \text{Poisson}(\lambda)$, $\lambda = 5/3$.

Recall $p(x) = e^{-\lambda} \frac{\lambda^x}{x!}$

then $E(X) = \lambda = 5/3$ and

$$P(X=0 \mid \lambda = 5/3) = e^{-5/3} \frac{(5/3)^0}{0!} = e^{-5/3} = 0.189$$

Ex(Poisson Approx. to Binomial Dist.): A typesetter, on average, makes one error per 500 words typeset.

A typical page has 300 words.

Find P(≤2 errors in a 5-page essay)

let X: number of errors in 5-pages

Assume $X \sim \text{Binomial}(n,p)$ where

$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$ and here,

$n = 300 \times 5 = 1500$

$p = 1/500$

then,
$$P(X \le 2 \mid n=1500, p=\tfrac{1}{500}) = \sum_{x=0}^{2} \binom{1500}{x} \left(\frac{1}{500}\right)^x \left(\frac{499}{500}\right)^{1500-x}$$
$$= 0.4230$$

using the Poisson Approx., i.e. $X \overset{d}{\approx} Y$,

$Y \sim \text{Poisson}(\lambda)$, $\lambda = np = 3$ and $p(y) = e^{-\lambda} \frac{\lambda^y}{y!}$

$\Rightarrow P(X \le 2 \mid \lambda) \approx P(Y \le 2 \mid \lambda)$
$$= \sum_{y=0}^{2} e^{-\lambda} \frac{\lambda^y}{y!} = e^{-3}\left(1 + 3 + \frac{3^2}{2!}\right) = 0.4232$$

Proof (Neg. Binom. Derivation). let $X \sim \text{NegBinom}(r,p)$, $p \in [0,1]$, $r \in \mathbb{N}_+$

count the number of independent Bernoulli(p) trials until obtaining the r-th success

$$p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \dots$$

Let $Y = X - r$, i.e. counting the number of failures which occur before the r-th success of ind. Bernoulli(p) trials

$\Updownarrow$

$-X = -r - Y$

$$P(Y=y \mid p,r) = P(X = y+r \mid p,r) = \binom{y+r-1}{r-1} p^r (1-p)^y \, , \quad y = 0,1,\dots$$

where $\binom{y+r-1}{r-1} = \binom{y+r-1}{y} = (-1)^y \binom{-r}{y}$ FS

since $n = y+r-1 \Rightarrow y = n-r+1$

$$= \frac{(y+r-1)!}{y!(r-1)!} = \frac{(y+r-1)\cdots(r+1)r}{y!} = \frac{r(r+1)\cdots(y+r-1)}{y!} = (-1)^y \frac{(-r)(-r-1)\cdots(-r-y+1)}{y!} = (-1)^y \binom{-r}{y}$$

$$= (-1)^y \binom{-r}{y} p^r (1-p)^y, \quad y = 0,1,\dots \; . \; ∎$$

Proof. (Quadratic Relation): let $X \sim \text{NegBin}(r,p)$ and $Y = X - r$.

Set $m = E(Y) = \frac{r(1-p)}{p} \Rightarrow \frac{1-p}{p} = \frac{m}{r} \Rightarrow \frac{1}{p} = \frac{m+r}{r}$

$\Rightarrow 1-p = \frac{mp}{r}$

Hence $\text{var}(Y) = \frac{E(Y)}{p} = m \cdot \frac{1}{p} = m\left(\frac{m+r}{r}\right) = \frac{m(m+r)}{r} = \frac{1}{r} m^2 + m$. ∎

Proof (Memoryless Property). let $X \sim \text{Geometric}(p)$, $p(x) = p(1-p)^{x-1}$, $x = 1,2,\dots$, $p \in (0,1)$.

FS $P(X > x) = \sum_{k=x+1}^{\infty} p(1-p)^{k-1} = (1-p)^x \Rightarrow P(X > s-t) = (1-p)^{s-t}$, $s > t$

since $\sum_{k=x+1}^{\infty} p(1-p)^{k-1} = p(1-p)^x \sum_{k=x+1}^{\infty} (1-p)^{k-(x+1)}$

letting $j := k-(x+1)$

$$= p(1-p)^x \sum_{j=0}^{\infty} (1-p)^j = p(1-p)^x \frac{1}{1-(1-p)} = p(1-p)^x \frac{1}{p} = (1-p)^x$$

so, if $s > t$,
$$P(X > s \mid X > t) = \frac{P(\{X > t\} \cap \{X > s\})}{P(X > t)} = \frac{P(X > s)}{P(X > t)} = (1-p)^{s-t}$$
$$\Rightarrow P(X > s \mid X > t) = P(X > s-t). \; ∎$$

Proof. Let $X \sim \text{NegBinom}(p,r)$. Let $Y = X - r$. Show $E(Y) = \frac{r(1-p)}{p}$ and $V(Y) = \frac{r(1-p)}{p^2}$.

suppose $Z_1, \dots, Z_r \overset{iid}{\sim} \text{Geom}(p)$ where $Z_i \sim \text{Geom}(p) = \text{NegBinom}(p, r=1)$.

Thus, let $Y = \sum_{i=1}^{r} Z_i$ s.t. $E(Z_i) = \frac{1-p}{p}$, $\text{Var}(Z_i) = \frac{1-p}{p^2}$.

then, by LOE: $E(Y) = \sum_{i=1}^{r} E(Z_i) = r \cdot \frac{1-p}{p}$

Similarly, $V(Y) = \sum_{i=1}^{r} V(Z_i) = r \cdot \frac{1-p}{p^2}$. ∎

3.2: Continuous Distributions

EX (Uniform RV): Let $X \sim \text{unif}([a,b])$; $f(x) = \frac{1}{b-a}$, $a \le x \le b$.

$$\Rightarrow E(X) = \int_a^b x f(x)\,dx = \frac{1}{b-a}\left(\frac{x^2}{2}\Big|_a^b\right) = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}.$$

$$E(X^2) = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{b-a}\left(\frac{x^3}{3}\Big|_a^b\right) = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$

$$\Rightarrow V(X) = \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{4(b^2+ab+a^2) - 3(b^2+2ab+a^2)}{12} = \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}.$$

EX (Gamma Function) = Properties: $\Gamma(\alpha) := \int_0^\infty t^{\alpha-1}e^{-t}\,dt$, $\alpha > 0$

change-of-var $t = x/\beta$, $\beta > 0$:

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

$\Gamma(\alpha+1) = \int_0^\infty t^\alpha e^{-t}\,dt$ via IBP: $\int u\,dv = uv - \int v\,du$

$u = t^\alpha$, $dv = e^{-t}dt$

$du = \alpha t^{\alpha-1}dt$, $v = \int e^{-t}dt = -e^{-t}$

$$= -t^\alpha e^{-t}\Big|_0^\infty + \int_0^\infty e^{-t}\alpha t^{\alpha-1}\,dt$$
$$= \alpha\underbrace{\int e^{-t}t^{\alpha-1}\,dt}_{\Gamma(\alpha)}$$
$$= \alpha\Gamma(\alpha).$$

$\Gamma(\alpha=1) = \int_0^\infty e^{-t}\,dt = e^{-t}\Big|_0^\infty = 1.$

$\Gamma(2) = 1\cdot\Gamma(1) = 1$

$\Gamma(3) = 2\cdot\Gamma(2) = 2!$

$\vdots$

$\Gamma(n) = (n-1)\Gamma(n-1) = \cdots = (n-1)!$

Proof (Gamma RV Moments): Let $X \sim \text{Gamma}(\alpha,\beta)$.

$$E(X) = \int_0^\infty x f(x|\alpha,\beta)\,dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x\cdot x^{\alpha-1}e^{-x/\beta}\,dx$$
$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\Gamma(\alpha+1)\beta^{\alpha+1}$$
$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\alpha\Gamma(\alpha)\beta^{\alpha+1} = \alpha\beta.$$

$$E(X^2) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\Gamma(\alpha+2)\beta^{\alpha+2}$$
$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}(\alpha+1)\Gamma(\alpha)\beta^{\alpha+2} = (\alpha+1)\alpha\beta^2.$$

$$\Rightarrow V(X) = E(X^2) - E^2(X) = (\alpha+1)\alpha\beta^2 - \alpha^2\beta^2$$
$$= \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2$$
$$= \alpha\beta^2.$$

$$M_X(t) = E(e^{tX})$$
$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty e^{tx}x^{\alpha-1}e^{-\frac{x}{\beta}}\,dx$$

where $e^{tx}e^{-x/\beta} = e^{-x(\frac{1}{\beta}-t)} = \exp\left(\frac{-x-\beta t}{\beta}\right)$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty e^{-x(\frac{1}{\beta}-t)}x^{\alpha-1}\,dx \quad (*)$$

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

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\Gamma(\alpha)\left(\frac{1}{\beta}-t\right)^{-\alpha}$$

where $\left(\frac{1}{\beta}-t\right)^{-\alpha} = \left(\frac{1-\beta t}{\beta}\right)^{-\alpha} = \left(\frac{\beta}{1-\beta t}\right)^\alpha$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\Gamma(\alpha)\left(\frac{\beta}{1-\beta t}\right)^\alpha, \quad t < \frac{1}{\beta}$$

for $t < \frac{1}{\beta}$ since $(*)$ converges when $\frac{1}{\beta} - t > 0 \Rightarrow$

where $(*)$ converges when $\frac{1}{\beta} - t > 0 \Rightarrow t < \frac{1}{\beta}$

$$= \left(\frac{1}{1-\beta t}\right)^\alpha, \quad t < \frac{1}{\beta}$$

Proof. Show that $\forall x > 0$ $P(X \le x) = P(Y \ge \alpha)$ for $\begin{cases} X \sim \text{Gamma}(\alpha,\beta) \\ Y \sim \text{Poisson}(x/\beta) \end{cases}$

$$\Leftrightarrow \int_0^x \text{pdf}_{Gam}(t)\,dt = \sum_{y=\alpha}^\infty \text{pmf}_{Poi}(y)$$

consider $P(X \le x) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^x t^{\alpha-1}e^{-t/\beta}\,dt$ via substitution

FS where $d(e^{-t/\beta}) = -\frac{1}{\beta}e^{-t/\beta}dt \Rightarrow e^{-t/\beta} = -\beta\, d(e^{-t/\beta})$

and $\Gamma(\alpha) = (\alpha-1)!$

$$= \frac{1}{(\alpha-1)!\beta^\alpha}\int_0^x t^{\alpha-1}(-\beta)\,d(e^{-t/\beta}) \text{ via IBP} \qquad \int u\,dv = uv - \int v\,du$$

let $u = t^{\alpha-1}$, $dv = (-\beta)\,d(e^{-t/\beta}) = e^{-t/\beta}$

$du = (\alpha-1)t^{\alpha-2}dt$, $v = \int e^{-t/\beta}dt = -\beta e^{-t/\beta}$

$= d(t^{\alpha-1})$

$$= \frac{1}{(\alpha-1)!\beta^\alpha}\left(t^{\alpha-1}\beta e^{-t/\beta}\Big|_x^0 + \beta\int_0^x e^{-t/\beta}\,d(t^{\alpha-1})\right)$$

where $d(t^{\alpha-1}) = (\alpha-1)t^{\alpha-2}$

$$= \underbrace{\frac{-1}{(\alpha-1)!}\left(\frac{x}{\beta}\right)^{\alpha-1}e^{-x/\beta}} + \underbrace{\frac{1}{(\alpha-2)!\beta^{\alpha-1}}\int_0^x e^{-t/\beta}t^{\alpha-2}\,dt}$$

$= -P(Y=\alpha-1)$ for $Y \sim \text{Poisson}(\frac{x}{\beta})$ $+$ $\text{cdf}_{\text{Gamma}(\alpha-1,\beta)}(x)$

$f(y|\lambda) = e^{-\lambda}\frac{\lambda^y}{y!}$, $y = 0, 1, \ldots$

where FS $F_\alpha(x) = F_{\alpha-1}(x) - P(Y=\alpha-1)$ for $\begin{cases} X \sim \text{Gamma} \\ Y \sim \text{Poisson}(\frac{x}{\beta}) \end{cases}$

$= F_{\alpha-2}(x) - P(Y=\alpha-2) - P(Y=\alpha-1)$

$= \cdots$

$= F_1(x) - \sum_{j=1}^{\alpha-1} P(Y=j)$

$= -P(Y=\alpha-1) - P(Y=\alpha-2) + \text{cdf}_{\text{Gamma}(\alpha-2,\beta)}(x)$

$= \cdots$

$= -P(Y=\alpha-1) - \cdots - P(Y=1) + \text{cdf}_{\text{Gamma}(1,\beta)}(x)$

where $\text{cdf}_{\text{Gamma}(1,\beta)}(x) = \frac{1}{\beta}\int_0^x e^{-t/\beta}\,dt$

$= -e^{-t/\beta}\Big|_0^x = 1 - e^{-x/\beta}$

$= -P(Y=\alpha-1) - \cdots - P(Y=1) + 1 - e^{-x/\beta}$

$= -P(Y=\alpha-1) - \cdots - P(Y=1) - P(Y=0) + 1$

$= P(Y \ge \alpha).$

3.23 The *Pareto distribution*, with parameters α and β, has pdf

$$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}}, \quad \alpha < x < \infty, \quad \alpha > 0, \quad \beta > 0.$$

(a) Verify that $f(x)$ is a pdf.
(b) Derive the mean and variance of this distribution.
(c) Prove that the variance does not exist if $\beta \le 2$.

a. $f(x) \ge 0$ $\forall x \in \mathcal{X}$, i.e. nonnegativity holds.

$\int_\alpha^\infty x^{-\beta-1}\,dx = -\frac{1}{\beta}x^{-\beta}\Big|_\alpha^\infty = \frac{1}{\beta\alpha^\beta}$ and thus $f(x)$ integrates to 1.

b. $E(X^n) = \int_\alpha^\infty x^n f(x)\,dx = \int_\alpha^\infty x^n\cdot\frac{\beta\alpha^\beta}{x^{\beta+1}}\,dx = \beta\alpha^\beta\int_\alpha^\infty x^{n-\beta-1}\,dx$,

evaluating $\int_\alpha^\infty x^k\,dx$ where $k = n-\beta-1$ $\begin{cases} \text{integral converges if } k \le -1 \\ \text{integral diverges if } k \ge -1 \end{cases}$

since $\int_\alpha^\infty x^k\,dx = \frac{x^{k+1}}{k+1}\Big|_\alpha^\infty$; to converge as $x \to \infty$, we need $x^{k+1} \to 0$, i.e. $k+1 < \infty \Rightarrow k < -1 \Rightarrow n-\beta-1 < -1 \Rightarrow n < \beta$

$$= \beta\alpha^\beta\left[\frac{x^{n-\beta}}{n-\beta}\Big|_{x=\alpha}^{x=\infty}\right]$$

where $n < \beta \Rightarrow$ convergent case

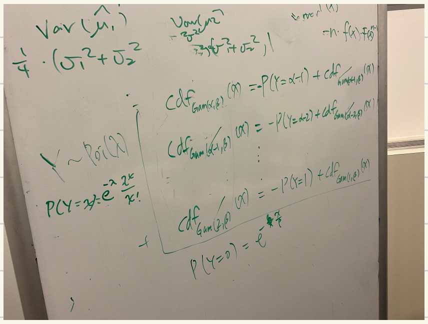
$$= \beta\alpha^\beta\left(0 - \frac{\alpha^{n-\beta}}{n-\beta}\right)$$
$$= \beta\alpha^\beta\cdot\frac{\alpha^{n-\beta}}{\beta-n} = \frac{\beta\alpha^n}{\beta-n}$$

Therefore,

$E(X) = \frac{\beta\alpha}{\beta-1}$, exists if $\beta > 1$

$E(X^2) = \frac{\beta\alpha^2}{\beta-2}$, exists if $\beta > 2$

$\Rightarrow V(X) = E(X^2) - E^2(X) = \frac{\beta\alpha^2}{\beta-2} - \left(\frac{\beta\alpha}{\beta-1}\right)^2$; c. If $\beta \le 2$, the integral of the second moment is infinite. variance expression is only finite when both moments exist, i.e. $\beta > 2$.



<u>Proof (Memoryless - Exponential)</u>: Let $X \sim Gamma(\alpha = 1, \beta > 0)$ s.t.

$$f(x|\beta) = \frac{1}{\beta} e^{-x/\beta}, \ x > 0, \text{ i.e. exponential pdf.}$$

Show Memoryless Prop. for X, i.e.

$$P(X > s | X > t) = P(X > s - t) \text{ for } s > t \geq 0:$$

$$P(X > s | X > t) = \frac{P(X > s \cap X > t)}{P(X > t)} \quad \text{for } s > t \geq 0$$

$$= \frac{P(X > s)}{P(X > t)} = \frac{e^{-s/\beta}}{e^{-t/\beta}}$$

$$= e^{-(s-t)/\beta}$$

$$= P(X > s - t). \blacksquare$$

3.2b - Continuous Dist. (cont.)

<u>Proof (Standard Normal RV)</u>. Let $X \sim N(\mu, \sigma^2)$. For $Z = \frac{1}{\sigma}(X - \mu)$, Show:

$$E(Z) = \frac{1}{\sigma}(E(X) - \mu) = \frac{1}{\sigma}(\mu - \mu) = 0$$

$$V(Z) = V\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2} Var(X) = 1. \blacksquare$$

Show that Z remains a Normal RV:

First look at its cdf s.t. $\forall z \in \mathbb{R}$

$$P(Z \leq z) = P\left(\frac{X - \mu}{\sigma} \leq z\right)$$

$$= P(X \leq \mu + \sigma z)$$

$$= \int_{-\infty}^{\mu + \sigma z} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x - \mu)^2\right) dx$$

via change-of-var and substitution, i.e.

let $x = \mu + \sigma y$ $\Rightarrow y = \frac{x - \mu}{\sigma}$, $\begin{cases} x = -\infty, \ y = -\infty \\ x = \mu + \sigma z, \ y = \frac{\mu + \sigma z - \mu}{\sigma} = z \end{cases}$

$(x - \mu)^2 = (\mu + \sigma y - \mu)^2 = \sigma^2 y^2$,

and $dx = \sigma dy$

$$= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi\cancel{\sigma^2}}} \exp\left(\frac{-1}{2\cancel{\sigma^2}} \cancel{\sigma^2} y^2\right) \cancel{\sigma} dy$$

$$= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} y^2\right) dy$$

and now obtain the pdf, i.e.

$$f_Y(y) = \frac{d}{dz} P(Z \leq z)$$

FS where $f_Y(y) = \frac{d}{dx} P(X \leq x) = \frac{d}{dx} \int_{-\infty}^{x} f(y) dy$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right), \text{ pdf of } N(0,1).$$

$$\Rightarrow Z \sim N(0,1). \blacksquare$$

<u>Proof (Normal Properties)</u>. Show that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x - \mu)^2\right) dx = 1$.

By change-of-var, i.e. let $z = \frac{1}{\sigma}(x - \mu)$,

it suffices to prove the above for $\mu = 0, \sigma = 1$, i.e.

Show: $\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$

The integrand is symmetric; need to allow

$$\int_0^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{\frac{2\pi}{2}} = \sqrt{\frac{\pi}{2}} \Leftrightarrow \left(\int_0^{\infty} e^{-\frac{1}{2}x^2} dx\right)^2 = \pi/2$$

$$\Leftrightarrow \left(\int_0^{\infty} e^{-\frac{1}{2}x^2} dx\right)\left(\int_0^{\infty} e^{-\frac{1}{2}y^2} dy\right) = \pi/2$$

FS where Fubini's Thm: If $f(x,y)$ is integrable on $A \times B$, then

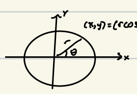
$$\int_A \int_B f(x,y)\, dy\, dx = \int_B \int_A f(x,y)\, dx\, dy = \iint_{A \times B} f(x,y)\, dx\, dy$$

and if $f(x,y) = f_1(x) \cdot f_2(x)$, then

$$\iint f_1(x) f_2(y)\, dx\, dy = \left(\int f_1(x) dx\right)\left(\int f_2(y) dy\right)$$

$$\Leftrightarrow \int_0^{\infty} \int_0^{\infty} \exp\left(-\frac{1}{2}(x^2 + y^2)\right) dx\, dy = \pi/2.$$

Now, another change-of-var, letting $\begin{cases} x = r\cos(\theta) \\ y = r\sin(\theta) \end{cases}$ for $r \geq 0$, $\theta \in [0, \pi/2]$



$(x,y) = (r\cos(\theta), r\sin(\theta))$; $x^2 + y^2 = r^2$

where FS: $dx\, dy = \left|\frac{dx\, dy}{dr\, d\theta}\right| dr\, d\theta = r\, dr\, d\theta$, i.e.

switching from cartesian (x,y) to polar (r, θ) coords.

Since $J = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \Rightarrow |J| = [r(\cos^2\theta + \sin^2\theta)] = r$

$$\Rightarrow dx\, dy = |J|\, dr\, d\theta = r\, dr\, d\theta$$

$$\Leftrightarrow \int_0^{\pi/2} \int_0^{\infty} e^{-\frac{1}{2}r^2} r\, dr\, d\theta$$

let $u = r^2$ $du = 2r\,dr$

$\frac{1}{2} du = r\,dr$

$r = 0 \Rightarrow u = 0$

$r = \infty, \Rightarrow u = \infty$

$$= \int_0^{\pi/2} \int_0^{\infty} e^{-\frac{1}{2}u} \frac{1}{2} du\, d\theta = \int_0^{\pi/2} \left(\frac{1}{2}\left(\frac{-2}{1}\right) e^{-\frac{1}{2}u} \Big|_{u=0}^{u=\infty}\right) d\theta$$

$$= \int_0^{\pi/2} (0 + 1)\, d\theta = \theta \Big|_0^{\pi/2} = \frac{\pi}{2}. \square$$

<u>Proof (Gamma Function Identity)</u>. Show $\Gamma(1/2) = \sqrt{\pi}$:

Let $X \sim N(0,1)$ s.t. $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$, $x \in \mathbb{R}$.

By definition, for $x > 0$, $\sqrt{\pi} = \sqrt{2} \int_0^{\infty} e^{-\frac{1}{2}x^2} dx$

Let $\begin{cases} t = \frac{1}{2}x^2 \Rightarrow x = (2t)^{1/2} \\ dt = x\,dx = (2t)^{1/2} dx \end{cases}$

So,

$$\sqrt{\pi} = \sqrt{2} \int_0^{\infty} e^{-t} (2t)^{-1/2} dt = \int_0^{\infty} t^{-1/2} e^{-t} dt$$

where $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha - 1} e^{-t} dt$, $\alpha > 0$

$$= \Gamma(1/2). \blacksquare$$

<u>Proof (Normal Approx.)</u>: We will later show via CLT that

$$\frac{1}{\sqrt{n}}(X - np) \xrightarrow{d} Normal(0, p_*(1 - p_*)) \text{ if } \begin{cases} n \to \infty \\ p \to p_* \end{cases}$$

$$\Rightarrow \frac{1}{\sqrt{n}}(X - np) \approx Normal(0, p(1-p))$$

$$\Rightarrow X - np \approx Normal(0, np(1-p))$$

$$\Rightarrow X \approx Normal(np, np(1-p)). \blacksquare$$

<u>Ex (Normal Approx.)</u>: Let $X \sim Binomial(n = 25, p = 0.6)$. Then $\begin{cases} E(X) = np = 15 \\ V(X) = np(1-p) = 2.45^2 \end{cases}$

using Normal approx, let $Y \sim N(15, 2.45^2)$.

Then, $P(X \leq 13) \approx P(Y \leq 13) = P\left(Z \leq \frac{13 - 15}{2.45^2}\right) = P(Z \leq -0.82) = 0.206$.

3.25 Suppose the random variable T is the length of life of an object (possibly the lifetime of an electrical component or of a subject given a particular treatment). The *hazard function* $h_T(t)$ associated with the random variable T is defined by

$$h_T(t) = \lim_{\delta \to 0} \frac{P(t \leq T < t + \delta | T \geq t)}{\delta}.$$

Thus, we can interpret $h_T(t)$ as the rate of change of the probability that the object survives a little past time t, given that the object survives to time t. Show that if T is a continuous random variable, then

$$h_T(t) = \frac{f_T(t)}{1 - F_T(t)} = -\frac{d}{dt} \log(1 - F_T(t)).$$

If T is continuous, then

$$P(t \leq T \leq t + \delta | T \geq t) = \frac{P(t \leq T \leq t + \delta, t \leq T)}{P(t \leq T)}$$

$$= \frac{P(t \leq T \leq t + \delta)}{P(t \leq T)}$$

$$= \frac{F_T(t + \delta) - F_T(t)}{1 - F_T(t)}$$

Therefore, from the definition of the derivative,

$$h_T(t) = \lim_{\delta \to 0} \frac{F_T(t + \delta) - F_T(t)}{\delta} = \frac{F_T'(t)}{1 - F_T(t)} = \frac{f_T(t)}{1 - F_T(t)}$$

Also, $\frac{-d}{dt}[\log(1 - F_T(t))] = \frac{-1}{1 - F_T(t)}(-f_T(t)) = h_T(t)$.

3.20 Let the random variable X have the pdf

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad 0 < x < \infty.$$

(a) Find the mean and variance of X. (This distribution is sometimes called a *folded normal.*)

(b) If X has the folded normal distribution, find the transformation $g(X) = Y$ and values of α and β so that $Y \sim \text{gamma}(\alpha, \beta)$.

a) $E(X) = \int_0^{\infty} x f(x) dx = \int_0^{\infty} x \cdot \frac{2}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} dx$ via substitution:

let $u = \frac{x^2}{2}$ $du = x\,dx$

$x = 0 : u = 0$

$x = \infty : u = \infty$

$$= \frac{2}{\sqrt{2\pi}} \left[\int_0^{\infty} e^{-u} du = -e^{-u}\Big|_0^{\infty} = 1\right]$$

$$\Rightarrow \underline{E(X) = \frac{2}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}}}$$

Similarly, $E(X^2) = \int_0^{\infty} x^2 f(x) dx = \frac{2}{\sqrt{2\pi}} \underbrace{\int_0^{\infty} x^2 e^{-x^2/2} dx}_{\sqrt{\pi/2}} = \frac{2}{\sqrt{2\pi}} \sqrt{\frac{\pi}{2}} = 1$

Thus, $\underline{V(X) = 1 - \frac{2}{\pi}}$

b) Suppose $Z \sim N(0,1)$ and let

$X = |Z| \Rightarrow X \sim$ Folded Normal$(0,1)$

$$\Rightarrow \underline{Y = X^2 = |Z|^2 = Z^2 \sim \chi_1^2 = Gamma(\alpha = \tfrac{1}{2}, \beta = 2)}.$$

c) Since n is large and p is small, a normal approximation to the binomial is reasonable.

Let $X \sim Binomial(n, p)$ s.t. $E(X) = np$, $V(X) = np(1-p)$.

If $\begin{cases} np \to \infty \\ np(1-p) \to \infty \end{cases}$ then $X \approx Normal(np, np(1-p))$

First compute mean and variance:

$\mu = np = 2000(0.01) = 20$

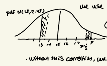
$\sigma^2 = np(1-p) = 2000(0.01)(0.99) = 19.8 \Rightarrow \sigma = \sqrt{19.8}$

Directly,

$$P(X\le 13)=\sum_{x=0}^{13}\binom{25}{x}0.6^x(0.4)^{25-x}=0.267.$$

For better results, apply the continuity correction, i.e.

$$\underset{\in\mathbb{N}}{P(X\le 13)}\approx \underset{\in\mathbb{R}}{P(Y\le 13.5)}=P\left(Z\le\frac{13.5-15}{2.45}\right)=0.271$$



we use $\begin{cases} P(X\le x)\approx P(Y\le x+1/2)\\ P(X\ge x)\approx P(Y\ge x-1/2)\end{cases}$

· without this correction, we tend to underestimate prob.'s at the tails

Now, using the normal approximation: $X\sim Bin(2000, 0.01)\approx Normal(20, \sigma^2=19.8)$,

$$P(X<100)\approx P(Y<100),\quad Y\sim Normal(20, 19.8)$$
$$\approx P(Y<99.5),\ \text{continuity correction}$$
$$=P\left(Z<\frac{99.5-20}{\sqrt{19.8}}\right)\approx 1$$

, i.e. P(fewer than 100 insects survive) is almost certain.

3.2c: Continuous Dist. (cont.)

<u>Proof (Beta n-th moments)</u>: $E(X^n)=\frac{1}{B(\alpha,\beta)}\int_0^1 x^n x^{\alpha-1}(1-x)^{\beta-1}dx$

where $B(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$

$$=\frac{1}{B(\alpha,\beta)}B(\alpha+n,\beta)$$

where $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$$=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cancel{\Gamma(\beta)}}\cdot\frac{\Gamma(\alpha+n)\cancel{\Gamma(\beta)}}{\Gamma(\alpha+\beta+n)}$$

$$=\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+n)}{\Gamma(\alpha)\Gamma(\alpha+\beta+n)}.$$

For $n=1$, $E(X)=\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)}$

where $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$

$$=\frac{\Gamma(\cancel{\alpha+\beta})\alpha\Gamma(\cancel{\alpha})}{\Gamma(\cancel{\alpha})(\alpha+\beta)\Gamma(\cancel{\alpha+\beta})}=\frac{\alpha}{\alpha+\beta}$$

Similarly, $E(X^2)=\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(\alpha+\beta+2)}$

where $\Gamma(\alpha+2)=(\alpha+1)\Gamma(\alpha+1)$

$$=\frac{\Gamma(\cancel{\alpha}+\beta)(\alpha+1)\Gamma(\cancel{\alpha}+1)}{\underbrace{\Gamma(\alpha+1)}_{\alpha}(\alpha+\beta+1)\underbrace{\Gamma(\alpha+\beta+1)}_{(\alpha+\beta)\Gamma(\cancel{\alpha+\beta})}}=\frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

So, $\underline{V(X)}=E(X^2)-E^2(X)=\frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}-\frac{\alpha^2}{(\alpha+\beta)^2}$

$$=\frac{\alpha}{(\alpha+\beta)}\frac{(\alpha+1)(\alpha+\beta)-\alpha(\alpha+\beta+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

$$=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.\ \blacksquare$$

3-3: Exponential Families

<u>Ex (Binomial-Exponential Family)</u>: Let $X\sim Bin(n,p)$. Show that $f(x)=c(\theta)h(x)\langle w(\theta),t(x)\rangle$

$$f(x)=\binom{n}{x}p^x(1-p)^{n-x},\quad x=0,1,\dots,n,\ \theta=p$$

$$=(1-p)^n\binom{n}{x}\left(\frac{p}{1-p}\right)^x$$

FS where $\exp(x\ln(\frac{A}{B}))=\left(\frac{A}{B}\right)^x$

$$=\underbrace{(1-p)^n}_{c(\theta)}\underbrace{\binom{n}{x}}_{h(x)}\exp\Big\{\underbrace{x}_{t_1(x)}\underbrace{\log\frac{p}{1-p}}_{w_1(\theta)}\Big\},\ k=1$$

<u>Ex (Normal-Exponential Family)</u>: Let $X\sim N(\mu,\sigma^2)$.

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right),\quad x\in\mathbb{R},\ \theta=(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_+$$

$$=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-1}{2\sigma^2}(x^2-2x\mu+\mu^2)\right)$$

$$=\underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\mu^2/2\sigma^2}}_{c(\theta)}\exp\Big\{\underbrace{\frac{-1}{2\sigma^2}}_{w_1(\theta)}\underbrace{x^2}_{t_1(x)}+\underbrace{\frac{\mu}{\sigma^2}}_{w_2(\theta)}\underbrace{x}_{t_2(x)}\Big\}$$

<u>Ex (Thm-Binomial)</u>: Let $X\sim Bin(n,p)$. Find $E(X)$ using the following theorem:

Given $f(x|\theta)=c(\theta)h(x)\exp\langle\theta,t(x)\rangle$,

$$E\sum_{i=1}^{k}\frac{\partial w_i(\theta)}{\partial\theta_j}t_i(X)=\frac{\partial}{\partial\theta_j}\log(1/c(\theta)),\ j=1,\dots,k$$

when to use which one?

$$E(t_j(X))=\frac{\partial}{\partial\theta_j}\log(1/c(\theta))$$

How would I apply this one?

Here,

$$f_X(x|p)=\underbrace{(1-p)^n}_{c(\theta)}\underbrace{\binom{n}{x}}_{h(x)}\exp\Big\{\underbrace{x}_{t_1(x)}\underbrace{\log\frac{p}{1-p}}_{w_1(\theta)}\Big\},\ k=1$$

Here, let $\theta=p$ s.t.

$c(\theta)=(1-p)^n$

FS where $\log(x^{-1})=\log(\frac{1}{x})=-\log(x)$

$$\Rightarrow\log\left(\frac{1}{c(\theta)}\right)=\log\left(\frac{1}{(1-p)^n}\right)=-n\log(1-p)$$

Thus,

$$\frac{\partial}{\partial\theta}\log\left(\frac{1}{c(\theta)}\right)=-\frac{\partial}{\partial p}n\log(1-p)=\frac{n}{1-p}$$

where FS $\log(\frac{A}{B})=\log(A)-\log(B)$

and

$$\frac{\partial}{\partial\theta}w_1(\theta)=\frac{\partial}{\partial p}(\log p-\log(1-p))$$

$$=\frac{1}{p}+\frac{1}{1-p}=\frac{(1-p)+p}{p(1-p)}=\frac{1}{p(1-p)}.$$

Applying the thm.:

$$\frac{\partial}{\partial\theta}\log(1/c(\theta))=E\frac{\partial}{\partial\theta}w_1(\theta)t_1(X)\Leftrightarrow\frac{n}{1-p}=E\frac{1}{p(1-p)}X$$

$$\Leftrightarrow n=E\left(\frac{X}{p}\right)$$

$$\Rightarrow \underline{E(X)=np}.\ \blacksquare$$

<u>Ex (Thm-Normal)</u>: Find $E(X)$, $V(X)$ given $f(x)=\underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\mu^2}{2\sigma^2}\right)}_{c(\theta)}\exp\Big\{\underbrace{\frac{-1}{2\sigma^2}}_{w_1(\theta)}\underbrace{x^2}_{t_1(x)}+\underbrace{\frac{\mu}{\sigma^2}}_{w_2(\theta)}\underbrace{x}_{t_2(x)}\Big\}$

$$w(\theta)=\left(\frac{-1}{2\sigma^2},\frac{\mu}{\sigma^2}\right).$$

$$\Rightarrow\log\left(\frac{1}{c(\theta)}\right)=\log\left(\frac{\sqrt{2\pi\sigma^2}}{e^{-\mu^2/2\sigma^2}}\right)=\log(\sqrt{2\pi\sigma^2})+\frac{\mu^2}{2\sigma^2}$$

So,

$$\frac{\partial}{\partial\mu}\log\left(\frac{1}{c(\theta)}\right)=\frac{\partial}{\partial\mu}\left(\log\sqrt{2\pi\sigma^2}+\frac{\mu^2}{2\sigma^2}\right)=\frac{\mu}{\sigma^2}$$

$$\frac{\partial}{\partial\mu}w_2(\theta)=\frac{\partial}{\partial\mu}\left(\frac{\mu}{\sigma^2}\right)=\frac{1}{\sigma^2}$$

Applying the thm.

$$\frac{\partial}{\partial\mu}\log\left(\frac{1}{c(\theta)}\right)=E\left(\frac{\partial}{\partial\mu}w_2(\theta)\right)t_2(X)+\left(\frac{\partial}{\partial\mu}\cancel{w_1(\theta)}\right)t_1(X)$$

$$\frac{\mu}{\sigma^2}=\frac{1}{\sigma^2}EX\Rightarrow EX=\mu.$$

To obtain $Et_2(X)=EX^2$:

$$\frac{\partial}{\partial\sigma}\log\left(\frac{1}{c(\theta)}\right)=\frac{\partial}{\partial\sigma}\left(\log(\sqrt{2\pi\sigma^2})+\frac{\mu^2}{2\sigma^2}\right)$$

where $\log(2\pi\sigma^2)^{1/2}=\frac{1}{2}[\log(2\pi)+\log(\sigma^2)]=\frac{1}{2}\log(2\pi)+\log(\sigma)$

$$\frac{d}{d\sigma}(\log(\sigma))=\frac{1}{\sigma},\ \frac{d}{d\sigma}\left(\frac{\mu^2}{2\sigma^2}\right)=\left(\frac{-\mu^2}{\sigma^3}\right)$$

$$=\frac{1}{\sigma}-\mu^2\frac{1}{\sigma^3}$$

$$\frac{\partial}{\partial\sigma}w_1(\theta)=\frac{\partial}{\partial\sigma}\left(\frac{-1}{2\sigma^2}\right)=\frac{1}{\sigma^3}$$

$$\frac{\partial}{\partial\sigma}w_2(\theta)=\frac{\partial}{\partial\sigma}\left(\frac{\mu}{\sigma^2}\right)=-2\mu\frac{1}{\sigma^3}$$

Applying the thm.

$$\frac{\partial}{\partial\sigma}\log\left(\frac{1}{c(\theta)}\right)=E\left(\frac{\partial}{\partial\sigma}w_1(\theta)\right)t_1(X)+\left(\frac{\partial}{\partial\sigma}w_2(\theta)\right)t_2(X)$$

$$\Rightarrow\frac{1}{\sigma}-\mu^2\frac{1}{\sigma^3}=\frac{1}{\sigma^3}E(X^2)-\frac{2\mu}{\sigma^3}E(X)$$

where $\frac{1}{\sigma}-\mu^2\frac{1}{\sigma^3}=\frac{\sigma^3-\mu^2\sigma}{\sigma\sigma^3}=\frac{\sigma(\sigma^2-\mu^2)}{\sigma\sigma^3}=\frac{\sigma^2-\mu^2}{\sigma^3}$

$$\Rightarrow\sigma^2-\mu^2=EX^2-2\mu^2\Rightarrow EX^2=\sigma^2+\mu^2.$$

3.28 Show that each of the following families is an exponential family.

(a) normal family with either parameter μ or σ known
(b) gamma family with either parameter α or β known or both unknown
(c) beta family with either parameter α or β known or both unknown
(d) Poisson family
(e) negative binomial family with r known, $0<p<1$

a) (i) μ known: $f(x|\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right)$,

$h(x)=1$, $c(\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}I_{(0,\infty)}(\sigma^2)$, $w_1(\sigma^2)=\frac{-1}{2\sigma^2}$, $t_1(x)=(x-\mu)^2$

(ii) σ^2 known: $f(x|\mu)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-x^2}{2\sigma^2}\right)\exp\left(\frac{-\mu^2}{2\sigma^2}\right)\exp\left(\mu\frac{x}{\sigma^2}\right)$

$h(x)=\exp\left(\frac{-x^2}{2\sigma^2}\right)$, $c(\mu)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-\mu^2}{2\sigma^2}\right)$, $w_1(\mu)=\mu$, $t_1(x)=\frac{x}{\sigma^2}$

b) (i) α known: $f(x|\beta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}$

$h(x)=\frac{x^{\alpha-1}}{\Gamma(\alpha)}$, $x>0$, $c(\beta)=1/\beta^\alpha$, $w_1(\beta)=\frac{1}{\beta}$, $t_1(x)=-x$

(ii) β known: $f(x|\alpha)=e^{-x/\beta}\frac{1}{\Gamma(\alpha)\beta^\alpha}\exp((\alpha-1)\log x)$

$h(x)=e^{-x/\beta}$, $x>0$, $c(\alpha)=\frac{1}{\Gamma(\alpha)\beta^\alpha}$, $w_1(\alpha)=\alpha-1$, $t_1(x)=\log x$

(iii) α,β unknown: $f(x|\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}\exp\left((\alpha-1)\log x-\frac{x}{\beta}\right)$

$h(x)=I_{\{x>0\}}(x)$, $c(\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}$, $w_1(\alpha)=\alpha-1$, $t_1(x)=\log x$
$w_2(\alpha,\beta)=-1/\beta$, $t_2(x)=x$

c) (i) α known: $f(x|\beta)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}\exp((\beta-1)\log(1-x))$

$h(x)=x^{\alpha-1}I_{[0,1]}(x)$, $c(\beta)=\frac{1}{B(\alpha,\beta)}$, $w_1(\beta)=\beta-1$, $t_1(x)=\log(1-x)$

(ii) β known: $f(x|\alpha)=\frac{1}{B(\alpha,\beta)}(1-x)^{\beta-1}\exp((\alpha-1)\log x)$

$h(x)=(1-x)^{\beta-1}I_{[0,1]}(x)$, $c(\alpha)=\frac{1}{B(\alpha,\beta)}$, $w_1(\alpha)=\alpha-1$, $t_1(x)=\log x$

(iii) α,β unknown: $f(x|\alpha,\beta)=\frac{1}{B(\alpha,\beta)}\exp((\alpha-1)\log x+(\beta-1)\log(1-x))$

$h(x)=I_{[0,1]}(x)$, $c(\alpha,\beta)=\frac{1}{B(\alpha,\beta)}$, $w_1(\alpha)=\alpha-1$, $t_1(x)=\log x$
$w_2(\beta)=\beta-1$, $t_2(x)=\log(1-x)$

d) $f(x|\theta)=\frac{e^{-\theta}}{x!}\cdot\theta^x=\frac{e^{-\theta}}{x!}\cdot\exp(x\log\theta)$

$h(x)=\frac{1}{x!}I_{\{0,1,2,\dots\}}(x)$, $c(\theta)=e^{-\theta}$, $w_1(\theta)=\log\theta$, $t_1(x)=x$

e) r known: $f(x|p)=\binom{x-1}{r-1}p^r(1-p)^{x-r}$, $0<p<1$

$$=\binom{x-1}{r-1}\left(\frac{p}{1-p}\right)^r\exp(x\log(1-p))$$

$h(x)=\binom{x-1}{r-1}I_{\{r,r+1,\dots\}}$, $c(p)=\left(\frac{p}{1-p}\right)^r$, $w_1(p)=\log(1-p)$, $t_1(x)=x$

Hence, $EX^2 = \sigma^2 + \mu^2 \Rightarrow V(X) = E(X^2) - E^2(X)$
$= \sigma^2 + \mu^2 - \mu^2 = \sigma^2$. ∎

3.4: Location and Scale Families

Proof. Given valid pdf $f(x)$ on $\mathbb{R}$ $\forall \mu \in \mathbb{R}, \sigma > 0$,

Show that $g(x|\mu,\sigma) := \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right), x \in \mathbb{R}$ is a valid pdf on $\mathbb{R}$:

$g \geq 0$ since $f \geq 0$.

checking the integral, $\int_{\mathbb{R}} g(x|\mu,\sigma)\,dx = \int_{\mathbb{R}} \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx$ via change-of-var

let $y = \frac{x-\mu}{\sigma} \Rightarrow x = \mu + \sigma y$
$\Rightarrow dx = \sigma\, dy$

$$= \int_{\mathbb{R}} \frac{1}{\sigma} f(y) \sigma\, dy$$

$$= \int_{\mathbb{R}} f(y)\,dy = 1. \blacksquare$$

EX(1): Let $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$ s.t. $X \sim N(0,1)$.

Then, we obtain the ℓ-s family, i.e.

$$\left\{ g(x|\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{1}{\sigma}(x-\mu)\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \middle| \begin{matrix} \mu \in \mathbb{R} \\ \sigma > 0 \end{matrix} \right\}.$$

EX(2): Let $f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}$; then, we obtain

$$\left\{ g(x|\mu,\sigma) = \frac{1}{\Gamma(\alpha)} \frac{1}{\sigma} \left(\frac{x-\mu}{\sigma}\right)^{\alpha-1} e^{-\frac{x-\mu}{\sigma}} \middle| \begin{matrix} \mu \in \mathbb{R} \\ \sigma > 0 \end{matrix} \right\}$$

NOT a Gamma Function

whereas

$$\left\{ g(x|\sigma) = \frac{1}{\Gamma(\alpha)} \frac{1}{\sigma} \left(\frac{x}{\sigma}\right)^{\alpha-1} e^{-\frac{x}{\sigma}} \middle| \sigma \in \mathbb{R}_+ \right\}$$
$$\equiv \{ \text{Gamma}(\alpha, \sigma) \mid \sigma \in \mathbb{R}_+ \}$$

Recall:
$$\boxed{f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, x > 0}$$

EX(3): Let $f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$, $x = 0, 1, \ldots$

Define: $g(x|\sigma) = \frac{1}{\sigma} e^{-\lambda} \frac{\lambda^{x/\sigma}}{(x/\sigma)!} \;\middle|\; \begin{matrix} x = \sigma y \\ y \in \mathbb{N} \\ \sigma > 0 \end{matrix}$

then g is a valid pmf supported by multiples of σ

NOT a Poisson family but inherits many properties of Poisson Dist.

Proof (sketch): Suppose Y - RV w/ pdf $f(y)$ and $E(Y), V(Y)$ exist

If X is a RV w/ pdf $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, then $\begin{cases} E(X) = \sigma E(Y) + \mu \\ V(X) = \sigma^2 \text{var}(Y) \end{cases}$

From class; why introduce Z?

If $Y \sim f$, let $Z := \mu + \sigma Y \sim \frac{1}{\sigma} f\left(\frac{z-\mu}{\sigma}\right)$
and $Z \overset{d}{=} X$.
$\Rightarrow \begin{cases} EX = EZ = \mu + \sigma EY \\ VX = VZ = \sigma^2 \text{var}(Y) \end{cases}$

3.5: Inequalities and Identities

Proof (Chebychev's Inequality): Suppose $r > 0$ and $g(x) \geq 0$ $\forall x$. Then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx \geq \int_{x: g(x) \geq r} g(x) f(x)\,dx$$
$$\geq \int_{x: g(x) \geq r} r f(x)\,dx$$
$$= r P(g(X) \geq r). \blacksquare$$

Proof (Derivation): Let $g(x) = \frac{(x-\mu)^2}{\sigma^2}$ where $\begin{cases} \mu = E(X) \\ \sigma^2 = V(X) \end{cases}$

Then, $P\left(\frac{(X-\mu)^2}{\sigma^2} \geq r\right) \leq \frac{1}{r} E \frac{(X-\mu)^2}{\sigma^2}$

where $\frac{1}{\sigma^2} E(X-\mu)^2 = \frac{1}{\sigma^2} \cdot \sigma^2 = 1$

$= \frac{1}{r}$

Write $r = t^2$ to obtain

$$P\left(\frac{(X-\mu)^2}{\sigma^2} \geq r\right) = P(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2}$$

and $P(|X-\mu| < t\sigma) \geq 1 - \frac{1}{t^2}$.

Proof. Let $Z \sim N(0,1)$. Show that $P(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}$

$P(|Z| \geq t) = 2P(Z \geq t)$ by symmetry of f_Z

$$= 2\int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$
$$\leq 2\int_t^\infty \frac{1}{\sqrt{2\pi}} \frac{z}{t} e^{-\frac{1}{2}z^2} dz$$
$$= \frac{2}{\sqrt{2\pi}} \frac{1}{t} e^{-\frac{1}{2}z^2}\Big|_\infty^t$$
$$= \sqrt{\frac{2}{\pi}} \frac{1}{t} e^{-\frac{1}{2}t^2}. \blacksquare$$

$\frac{2}{\sqrt{2\pi}\,t} \int_t^\infty z e^{-\frac{1}{2}z^2} dz$ via substitution

$u = z^2$, $du = 2z\,dz$, $\frac{1}{2}du = z\,dz$
$z = t, u = t^2$
$z = \infty, u = \infty$

$$\frac{2}{\sqrt{2\pi}\,t} \int_{t^2}^\infty e^{-\frac{1}{2}u} \frac{1}{2} du = \frac{1}{\sqrt{2\pi}\,t}\left[-2e^{-\frac{1}{2}u}\Big|_{t^2}^\infty\right]$$
$$= 0 + \frac{2}{\sqrt{2\pi}\,t} e^{-\frac{1}{2}t^2}$$

Proof. Show that if $X \sim N(\theta, \sigma^2)$, g-differentiable function s.t. $E|g'(X)| < \infty$, then $E[g(X)(X-\theta)] = \sigma^2 E g'(X)$:

$$\text{LHS} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} g(x)(x-\theta) e^{\frac{-1}{2\sigma^2}(x-\theta)^2} dx$$

Proof.

$$\text{LHS} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} g(x)(x-\theta) e^{-\frac{1}{2\sigma^2}(x-\theta)^2} dx$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} g(x)(-\sigma^2)\, d\left(e^{-\frac{1}{2\sigma^2}(x-\theta)^2}\right)$$
$$\overset{\text{i.b.p.}}{=} \frac{1}{\sqrt{2\pi\sigma^2}} \left(g(x)(-\sigma^2) e^{-\frac{1}{2\sigma^2}(x-\theta)^2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \sigma^2 e^{-\frac{1}{2\sigma^2}(x-\theta)^2} dg(x) \right)$$
$$= \sigma^2 \mathbb{E}\, g'(X). \quad \square$$

$\to 0$ (*) see next page

3.46 Calculate $P(|X - \mu_X| \geq k\sigma_X)$ for $X \sim \text{uniform}(0,1)$ and $X \sim \text{exponential}(\lambda)$, and compare your answers to the bound from Chebychev's Inequality.

For $X \sim$ uniform$(0,1)$, $\mu = 1/2$ and $\sigma^2 = 1/12$; thus,

$$P(|X-\mu| > k\sigma) = P\left(X \leq \tfrac{1}{2} - k\sigma\right) + P\left(X > \tfrac{1}{2} + k\sigma\right)$$
$$= \left(\tfrac{1}{2} - \tfrac{k}{\sqrt{12}}\right) + \left(1 - \tfrac{1}{2} - \tfrac{k}{\sqrt{12}}\right) = 1 - \tfrac{2k}{\sqrt{12}}$$

where if $\frac{1}{2} - k\sigma < 0$ or $\frac{1}{2} + k\sigma > 1$,
then part or all of the prob. mass lies outside $(0,1)$

Hence, when $k \geq \sqrt{3}$ $\begin{cases} \frac{1}{2} - k\sigma \leq 0 \\ \frac{1}{2} + k\sigma \geq 1 \end{cases}$ and the entire prob. region lies w/in $(0,1)$.

$$= \begin{cases} 1 - \frac{2k}{\sqrt{12}}, & k < \sqrt{3} \\ 0, & k > \sqrt{3} \end{cases}$$

Similarly, for $X \sim$ exponential(λ), $\mu = \lambda$, $\sigma^2 = \lambda^2$; then,

$$P(|X-\lambda| \geq k\lambda) = P(X \leq \lambda - k\lambda) + P(X \geq \lambda + k\lambda)$$
$$= \underbrace{P(X \leq \lambda(1-k))}_{\text{left-tail (1)}} + \underbrace{P(X \geq \lambda(1+k))}_{\text{right-tail (2)}}$$

(1) := Exp. Dist. has support for $x \geq 0$, so

if $k > 1$, then $\lambda(1-k) < 0$, outside of the support $\Rightarrow P(X \leq \lambda(1-k)) = 0$
$k \leq 1$, then $\lambda(1-k) \geq 0$, and

$F(x) = 1 - e^{-x/\lambda} \Rightarrow P(X \leq \lambda(1-x)) = 1 - e^{-(1-k)}$

(2) := The right tail is always valid:

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

$$\Rightarrow P(|X-\mu| > k\sigma) = \begin{cases} 1 + e^{-(k+1)} - e^{-k-1}, & k \leq 1 \\ e^{-(k+1)}, & k \geq 1 \end{cases}$$

Chebychev's inequality gives the bound: $P(|X-\mu| > k\sigma) \leq 1/k^2$; it can be shown that:

Comparison of probabilities

k	u(0,1) exact	exp(λ) exact	Chebychev
.1	.942	.926	100
.5	.711	.617	4
1	.423	.135	1
1.5	.134	.0821	.44
$\sqrt{3}$	0	0.0651	.33
2	0	0.0498	.25
4	0	0.00674	.0625
10	0	0.0000167	.01

We see that Chebychev's inequality is quite conservative.

$$E[g(X)(X-\theta)] = \sigma^2 E g'(X)$$

Ex (Applying Stein's Lemma): Let $X \sim N(\theta, \sigma^2)$.

we know $E(X) = \theta$, $E(X^2) = E^2(X) + V(X)$
$= \theta^2 + \sigma^2$

and
$$\begin{aligned} E(X^3) &= EX^2(X-\theta+\theta) \\ &= EX^2(X-\theta) + \theta EX^2 \\ &\text{by Stein's Lemma} \\ &= \sigma^2 E(2X) + \theta(\theta^2+\sigma^2) \\ &= 2\sigma^2\theta + \theta^3 + \theta\sigma^2 \\ &= \theta^3 + 3\sigma^2\theta \end{aligned}$$

and so on for EX^n, $n = 4, 5, \ldots$

Application: • If $X \sim \text{Poisson}(\lambda)$, then $E(X) = V(X) = \lambda$.

Application.

• if $X \sim \text{Poisson}(\lambda)$ then $\mathbb{E}X = \operatorname{var} X = \lambda$

Thus $\mathbb{E}\lambda X^2 = \mathbb{E}X(X-1)^2 \leftarrow g(x) = x^2$

$$\begin{aligned} \lambda(\lambda^2+\lambda) &= \mathbb{E}(X^3 - 2X^2 + X) \\ \lambda^3 + \lambda^2 &= \mathbb{E}X^3 - 2(\lambda^2+\lambda) + \lambda \\ \Rightarrow \quad \mathbb{E}X^3 &= \lambda^3 + 3\lambda^2 + \lambda. \end{aligned}$$

Ex: If $X \sim \text{NegBin}(r, p)$, take $g(x) = r + x$.

Then,

$$\begin{aligned} \mathbb{E}(1-p)(r+X) &= \mathbb{E}\frac{X}{r+X-1}(r+X-1) \\ \Rightarrow \quad (1-p)r + (1-p)\mathbb{E}X &= \mathbb{E}X \\ \Rightarrow \quad \mathbb{E}X &= \frac{(1-p)r}{p}. \end{aligned}$$

4.1: Joint and marginal distributions

Ex (n-dimensional random vector): $X = (X_1, X_2, \ldots, X_n) \in \mathbb{R}$

associated w/ msmt's of a (random) person

where X_1 = temp.
X_2 = weight
X_3 = bp
$\vdots$

Ex (Bivariate vector): consider bivariate RV $(X,Y) \in \mathbb{R}^2$ corresponding to the following

expt (tossing two fair dice): $\begin{cases} X = \text{sum of two dice} \\ Y = \text{absolute diff. of two dice} \end{cases}$

write sample point $s \in S$ as $s = (s_1, s_2)$ for $s_1, s_2 \in \{1, \ldots, 6\}$

Then $\begin{cases} X = s_1 + s_2 \\ Y = |s_1 - s_2| \end{cases}$

Now, we may define prob. of events defined in terms of X and Y:

$$\begin{aligned} P((X,Y) = (5,3)) &:= P(\{(s_1,s_2) \mid s_1+s_2 = 5, |s_1-s_2| = 3\}) \\ &= P(\{(4,1),(1,4)\}) \\ &= P(\{(4,1)\}) + P(\{(1,4)\}) \\ &\text{assuming independence} \\ &= P(s_1=4)P(s_2=1) + P(s_1=1)P(s_2=4) \\ &\text{assuming fair dice} \\ &= (1/6)(1/6) + (1/6)(1/6) = 1/18. \end{aligned}$$

Ex (Joint pmf for (X,Y)): Let $f_{XY}(x,y)$ be the joint pmf for (X,Y); we know

$X \in \{2, \ldots, 12\}$, $Y \in \{0, \ldots, 5\}$

Then $P(X=2) = P((X,Y) \in \{(2,0), (2,1), \ldots, (2,5)\})$
$= \sum_{i=0}^{5} f_{XY}(2,i)$

In fact, $\forall\, x \in \{1, \ldots, 6\}$,
$$P(X=x) = \sum_{y=0}^{5} f_{XY}(x,y)$$

Likewise, the distribution of Y is also completely determined:
$$P(Y=y) = \sum_{x=2}^{12} f_{XY}(x,y)$$

Ex (Joint Dist.): Let $X \sim \text{Bernoulli}(1/2)$
$Y \sim \text{Bernoulli}(1/2)$
$Z = 1 - X$

where $Y \perp\!\!\!\perp X$, $Z \not\perp\!\!\!\perp X$, $Z \perp\!\!\!\perp Y$

Then $X \overset{d}{=} Y \overset{d}{=} Z$

but $\underset{X \perp\!\!\!\perp Y}{(X,Y)} \neq \underset{X \not\perp\!\!\!\perp Z}{(X,Z)}$ in joint dist.

$P((X,Y) = (1,1)) \overset{ind}{=} P(X=1)P(Y=1) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$
$P((X,Z) = (1,1)) = P(X=1, 1-X=1) = 0$

4.5 (a) Find $P(X > \sqrt{Y})$ if X and Y are jointly distributed with pdf
$$f(x,y) = x + y, \quad 0 \le x \le 1, \quad 0 \le y \le 1.$$
(b) Find $P(X^2 < Y < X)$ if X and Y are jointly distributed with pdf
$$f(x,y) = 2x, \quad 0 \le x \le 1, \quad 0 \le y \le 1.$$

a)
$$\begin{aligned} P(X > \sqrt{Y}) &= \int_0^1 \int_{\sqrt{y}}^1 (x+y)\,dx\,dy = \int_0^1 \left(\frac{x^2}{2} + xy \Big|_{x=\sqrt{y}}^{x=1}\right) dy \\ &= \int_0^1 \left(\frac{1}{2} + y\right) - \left(\frac{y}{2} + y^{3/2}\right) dy \\ &= \frac{y}{2} + \frac{y^2}{2} - \frac{y^2}{4} - \frac{2}{5}y^{5/2} \Big|_0^1 \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} - \frac{2}{5} \\ &= \frac{10}{20} + \frac{10}{20} - \frac{5}{20} - \frac{8}{20} \\ &\boxed{= \frac{7}{20}.} \end{aligned}$$

b) $X^2 < Y < X \Rightarrow \begin{cases} X^2 < Y \Rightarrow X < \sqrt{Y} \\ Y < X \Rightarrow X > Y \end{cases}$

$$\begin{aligned} P(X^2 < Y < X) &= \int_0^1 \int_y^{\sqrt{y}} 2x\,dx\,dy \\ &= \int_0^1 \left(x^2 \Big|_y^{\sqrt{y}}\right) dy = \int_0^1 y - y^2\,dy \\ &= \frac{y^2}{2} - \frac{y^3}{3}\Big|_0^1 = \frac{1}{2} - \frac{1}{3} = \frac{3}{6} - \frac{2}{6} = \boxed{\frac{1}{6}}. \end{aligned}$$

4.9 Prove that if the joint cdf of X and Y satisfies
$$F_{X,Y}(x,y) = F_X(x)F_Y(y),$$
then for any pair of intervals (a,b), and (c,d),
$$P(a \le X \le b, c \le Y \le d) = P(a \le X \le b)P(c \le Y \le d).$$

$$\begin{aligned} P(a \le X \le b, c \le Y \le d) &= P(X \le b, c \le Y \le d) - P(X \le a, c \le Y \le d) \\ &= P(X \le b, Y \le d) - P(X \le b, Y \le c) - P(X \le a, Y \le d) - P(X \le a, Y \le c) \\ &= F(b,d) - F(b,c) - F(a,d) - F(a,c) \\ &= F_X(b)F_Y(d) - F_X(b)F_Y(c) - F_X(a)F_Y(d) - F_X(a)F_Y(c) \\ &= P[X \le b][P(Y \le d) - P(Y \le c)] - P(X \le a)[P(Y \le d) - P(Y \le c)] \\ &= P(X \le b)[P(c \le Y \le d)] - P(X \le a)[P(c \le Y \le d)] \\ &= P(a \le X \le b)P(c \le Y \le d) \end{aligned}$$

4.2a: conditional distributions

Ex (conditional prob.): Let (X,Y) = (height, weight) of a (random) person.

Find $P(Y>200\text{ lb} \mid X=6')$

Proof (continuous conditional pdf): Consider the continuous bivariate random vector $(X,Y)\in\mathbb{R}^2$.

$$P(Y\in B\mid X=x) := \lim_{\varepsilon\downarrow 0} P(Y\in B\mid X\in(x-\varepsilon, x+\varepsilon))$$

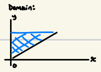
$$= \frac{P(Y\in B\cap X\in(x-\varepsilon,x+\varepsilon))}{P(X\in(x-\varepsilon,x+\varepsilon))} = \frac{\int_B\int_{x-\varepsilon}^{x+\varepsilon} f_{XY}(x,y)\,dx\,dy}{\int_{x-\varepsilon}^{x+\varepsilon} f_X(x)\,dx}$$

Under suitable conditions (see 2.4 - Tools)

$$\lim_{\varepsilon\downarrow 0}\int_B\left(\frac{\int_{x-\varepsilon}^{x+\varepsilon} f_{XY}(x,y)\,dx}{\int_{x-\varepsilon}^{x+\varepsilon} f_X(x)\,dx}\right)dy = \int_B\left(\lim_{\varepsilon\downarrow 0}\frac{\frac{1}{2\varepsilon}\int_{x-\varepsilon}^{x+\varepsilon} f_{XY}(x,y)\,dx}{\frac{1}{2\varepsilon}\int_{x-\varepsilon}^{x+\varepsilon} f_X(x)\,dx}\right)dy = \int_B \frac{f_{XY}(x,y)}{f_X(x)}\,dy$$

Ex: Let $(X,Y)\sim f_{XY}(x,y) = \begin{cases} e^{-y}, & 0<x<y<\infty \\ 0, & \text{otherwise}\end{cases}$

Domain:



$f_{XY}(x,y)\geq 0 \ \forall x,y\in\mathbb{R}^2$ and

$$\int_{\mathbb{R}^2} f_{XY}(x,y)\,dx\,dy = \int_0^\infty\int_0^y e^{-y}\,dx\,dy = \int_0^\infty ye^{-y}\,dy \text{ via IBP}$$

$\int u\,dv = uv - \int v\,du$ LIATE

$u=y \quad du=dy$

$dv=e^{-y}dy \Rightarrow v=\int e^{-y}dy = -e^{-y}$

$$= -ye^{-y}\Big|_0^\infty + \int_0^\infty e^{-y}\,dy = -e^{-y}\Big|_0^\infty = 1;$$ thus, we have a valid joint pdf.

Marginal computation: $f_X(x)=0 \ \forall x\leq 0$.

For $x>0$,
$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_x^\infty f_{XY}(x,y)\,dy = \int_x^\infty e^{-y}\,dy = -e^{-y}\Big|_x^\infty = e^{-x}.$$

For $y>0$,
$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_0^y f_{XY}(x,y)\,dx = \int_0^y e^{-y}\,dx = ye^{-y}$$

and $f_Y(y)=0 \ \forall y\leq 0$

Conditional pdf:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{x-y} \text{ if } y\geq x>0$$

$= 0$ if $y<x$

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{e^{-y}}{ye^{-y}} = \frac{1}{y} \text{ if } x\in(0,y)$$

$= 0$ otherwise

Conditional expectations:

$$E(Y|X=x) = \int y f_{Y|X}(y|x)\,dy = \int_x^\infty ye^{x-y}\,dy = e^x\int_x^\infty ye^{-y}\,dy \text{ via IBP}$$

$$= e^x\left[-ye^{-y}\Big|_x^\infty + \int_x^\infty e^{-y}\,dy\right]$$

$$= e^x\left[xe^{-x} - e^{-y}\Big|_x^\infty\right]$$

$$= e^x[xe^{-x}+e^{-x}] = x+1$$

Similarly, $E(Y^2|X=x) = \int_x^\infty y^2e^{x-y}\,dy = e^x\int_x^\infty y^2e^{-y}\,dy = x^2+2x+2$

Hence, $\operatorname{Var}(Y|X=x) = x^2+2x+2-(x+1)^2 = 1$.

Ex: Given two light bulbs w/ life lengths $X, Z\geq 0$

Let $X, Z \overset{iid}{\sim} \operatorname{Exp}(\lambda)$

where $\begin{cases} X := \text{light bulb 1 is ON, 2 is off} \\ Z := \text{light bulb 2 is turned on as soon as 1 burns out}\end{cases}$

Let $Y=X+Z$ be the time bulb 2 burns out

$$P(Y\leq y|X=x) = P(Z\leq y-x|X=x)$$
$$= P(Z\leq y-x) \text{ since } Z \perp\!\!\!\perp X$$
$$= F_Z(y-x)$$
$$= 1-e^{-\lambda(y-x)} \text{ if } y>x$$

Hence $f_{Y|X}(y|x) = \frac{d}{dy}(1-e^{-\lambda(y-x)}) = \lambda e^{-\lambda(y-x)}$

$= 0$ otherwise

with $f_X(x) = \lambda e^{-\lambda x}\,\mathbb{1}(x>0)$

The joint pdf for (X,Y) is completely determined via

$$f_{XY}(x,y) = f_X(x)f_{Y|X}(y|x)$$
$$= \lambda^2e^{-\lambda y} \text{ if } y>x>0$$
$$= 0 \text{ otherwise}$$

4.2b: Independence

Ex: Suppose $f_{XY}(x,y)$ is given by the table

x \ y	1	2	3
10	1/10	1/5	1/5
20	1/10	1/10	3/10

all numbers sum to 1

$$f_X(10) = f_{XY}(10,1) + f_{XY}(10,2) + f_{XY}(10,3)$$
$$= 1/10 + 1/5 + 1/5 = 1/2$$

$f_Y(3) = 1/5 + 3/10 = 1/2$

But $f_{XY}(10,3) = 1/5 \neq f_X(10)\times f_Y(3)$

$\Rightarrow X, Y$ not independent.

Proof. Let $(X,Y)\sim f_{XY}$. Show that

$X\perp\!\!\!\perp Y$ i.f.f. $\exists$ functions $g(x)$ and $h(y)$ s.t.

$f_{XY}(x,y) = g(x)h(y) \ \forall x,y\in\mathbb{R}$

"only if": when $X\perp\!\!\!\perp Y$, $f_{XY}(x,y) = f_X(x)f_Y(y) = g(x)h(y)$

"if": Given that $f_{XY}(x,y) = g(x)h(y)$,

$$1 = \int_{\mathbb{R}}\int_{\mathbb{R}} f_{XY}(x,y)\,dx\,dy$$
$$= \iint g(x)h(y)\,dx\,dy$$

1. Let f_{XY} denote the pmf or pdf of a bivariate vector (X,Y). f_X denotes the marginal pmf/pdf of X and $f_{Y|X}$ stands for the conditional pmf/pdf of Y given X.

   (i) Let A and B be subsets of X and Y's domains, respectively, such that $P(X\in A)>0$. Produce an example in the discrete setting to demonstrate that
   $$P(Y\in B|X\in A) \neq \sum_{x\in A}\sum_{y\in B} f_{Y|X}(y|x).$$
   It suffices if you simply take X and Y to be binary random variables in your example.

   (ii) Derive a correct expression of $P(Y\in B|X\in A)$ in terms of *only* $f_{Y|X}$ and f_X (do this for the discrete setting, and then proceed to the continuous setting).

   (iii) Let $B=\{y\}$, a singleton. Derive $P(Y\in B|X\in A)$ in terms of *only* $f_{X|Y}$ and f_Y (do this for the discrete setting, and then proceed to the continuous setting).

i) Let $X, Y\in\{0,1\}$ with the following joint pmf:

	Y=0	Y=1
X=0	0.1	0.2
X=1	0.3	0.4

Note that this is a valid pdf:

$f_{XY}(x,y)\geq 0 \ \forall x,y$ and it sums to 1.

Let $A=\{0,1\}$, $B=\{1\}$.

Then $P(Y\in B|X\in A) = P(Y=1) = f_{XY}(0,1)+f_{XY}(1,1) = 0.2+0.4 = 0.6$ (*)

But $\sum_{x\in A}\sum_{y\in B} f_{Y|X}(y|x) = f_{Y|X}(1|0) + f_{Y|X}(1|1)$

$$= \frac{0.2}{0.1+0.2} + \frac{0.4}{0.3+0.4}$$

$= \frac{26}{21} \neq$ (*) and an invalid probability.

iii) Letting $B=\{y\}$,

$$P(Y=y|X\in A) = \frac{P(Y=y, X\in A)}{P(X\in A)}$$

where $f_{XY}(x,y) = f_{X|Y}(x|y)f_Y(y)$

$$= \frac{\sum_{x\in A} P(X=x|Y=y)P(Y=y)}{\sum_{x\in A} P(X\in A)}$$

and $f_X(x) = \sum_{y'} f_{X|Y}(x|y')f_Y(y')$

$$\boxed{P(Y=y|X\in B) = \frac{f_Y(y)\sum_{x\in A} f_{X|Y}(x|y)}{\sum_{x\in A}\sum_{y'} f_{X|Y}(x|y')f_Y(y')}}$$

Now assume X, Y - continuous.

To interpret $P(Y=\{y\}|X\in A)$, we treat this as a conditional density:

$$f_{Y|X\in A}(y) = \frac{f_Y(y)\int_{x\in A} f_{X|Y}(x|y)\,dx}{\int_{x\in A} f_X(x)\,dx}$$

$$\Rightarrow \boxed{f_{Y|X\in A}(y) = \frac{f_Y(y)\int_A f_{X|Y}(x|y)\,dx}{\int_A\int f_{X|Y}(x|y')f_Y(y')\,dy'\,dx}}$$

ii) $P(Y\in B|X\in A) = \frac{P(X\in A, Y\in B)}{P(X\in A)}$

where $P(X\in A, Y\in B) = \sum_{x\in A}\sum_{y\in B} f_{XY}(x,y)$

and $P(X\in A) = \sum_{x\in A}\sum_{y} f_{XY}(x,y) = \sum_{x\in A} f_X(x)$

so,
$$P(Y\in B|X\in A) = \frac{\sum_{x\in A}\sum_{y\in B} f_{XY}(x,y)}{\sum_{x\in A} f_X(x)}$$

where $f_{XY}(x,y) = f_{Y|X}(y|x)f_X(x)$

Thus,
$$P(Y\in B|X\in A) = \frac{\sum_{x\in A}\sum_{y\in B} f_{Y|X}(y|x)f_X(x)}{\sum_{x\in A} f_X(x)}$$

$$\boxed{P(Y\in B|X\in A) = \sum_{x\in A}\left(\frac{f_X(x)}{\sum_{x'\in A} f_X(x')}\cdot\sum_{y\in B} f_{Y|X}(y|x)\right)}$$

i.e. the weighted average of conditional prob.'s.

For the continuous case,

replacing PMF's w/ PDF's and sums with integrals

$$P(Y\in B|X\in A) = \frac{\int_A\int_B f_{XY}(x,y)\,dy\,dx}{\int_A f_X(x)\,dx}$$

$$\boxed{P(Y\in B|X\in A) = \frac{\int_A f_X(x)\int_B f_{Y|X}(y|x)\,dy\,dx}{\int_A f_X(x)\,dx}}$$

$$= \Big(\underbrace{\int g(x)dx}_{c}\Big)\Big(\underbrace{\int h(y)dy}_{d}\Big)$$

Now, $f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)dy = g(x)d$

$f_Y(y) = \int_{\mathbb{R}} f_{XY}(x,y)dy = h(y)c.$

Hence,

$$f_X(x)f_Y(y) = g(x)h(y)cd = g(x)h(y) = f_{XY}(x,y). \blacksquare$$

<u>EX</u>: If $f_{XY}(x,y) = \frac{1}{384}x^2y^4e^{-y-(x/2)}$, $x,y>0$

$$= \frac{1}{384}x^2e^{-x/2}y^4e^{-y} \Rightarrow X \perp\!\!\!\perp Y.$$

<u>Proof.</u> Show that the sum of two independent normal RV's is again Normal:

Let $X \sim \text{Normal}(0,1)$. Then

$$M_X(t) = Ee^{tX} = \int_{\mathbb{R}} e^{tx}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx$$

$$= \frac{1}{\sqrt{2\pi}}\int e^{tx-\frac{1}{2}x^2}dx \quad \text{, computing the square}$$

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

$$= e^{\frac{1}{2}t^2}$$

If $X \sim \text{Normal}(\mu, \sigma^2)$, then

$Z := \frac{1}{\sigma}(X-\mu) \sim N(0,1)$, so

$$M_X(t) = Ee^{tX} = Ee^{t(\mu+\sigma Z)} = e^{t\mu}Ee^{t\sigma Z} = e^{t\mu}M_Z(t\sigma) = e^{t\mu+\frac{1}{2}t^2\sigma^2}$$

Now suppose $X \sim \text{Normal}(\mu_1, \sigma_1^2)$

$Y \sim \text{Normal}(\mu_2, \sigma_2^2)$

$X \perp\!\!\!\perp Y$.

Let $Z = X+Y$. Then,

$$M_Z(t) = M_X(t)M_Y(t) \text{, since } X \perp\!\!\!\perp Y$$

$$= e^{t\mu_1+\frac{1}{2}t^2\sigma_1^2}e^{t\mu_2+\frac{1}{2}t^2\sigma_2^2}$$

$$= e^{t(\mu_1+\mu_2)+\frac{1}{2}t^2(\sigma_1^2+\sigma_2^2)}$$

i.e. the mgf of $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

Hence, since a RV's dist. is determined by its mgf,

$$Z \sim \text{Normal}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2). \blacksquare$$

4.3: Bivariate Transformation

<u>First Approach:</u> $M_X(t) = Ee^{tX} = \sum_{x=0}^{\infty} e^{tx}e^{-\theta}\frac{\lambda^x}{x!}$

$$= \sum_{x=0}^{\infty} e^{-\theta}\frac{(e^t)^x}{x!} = e^{\theta(e^t-1)}.$$

Hence $M_{X+Y}(t) = M_X(t)M_Y(t)$ since $X \perp\!\!\!\perp Y$

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

which is the mgf for Poisson$(\lambda+\theta)$.

<u>Second Approach:</u> write $\binom{u}{v} = g\left(\binom{x}{y}\right) := \binom{x+y}{y}$

Then, for $u \geq v$,

$$f_{UV}(u,v) = \sum_{\substack{x,y \\ y=v \\ x+y=u}} f_{XY}(x,y)$$

$$= f_{XY}(u-v,v)$$

$$= f_X(u-v)f_Y(v) \text{ since } X \perp\!\!\!\perp Y$$

$$= e^{-\theta}\frac{\theta^{u-v}}{(u-v)!}e^{-\lambda}\frac{\lambda^v}{v!}$$

$$= e^{-(\theta+\lambda)}\frac{\theta^{u-v}}{(u-v)!}\cdot\frac{\lambda^v}{v!}$$

Obtain $f_U(u)$ by marginalization:

$$f_U(u) = \sum_v f_{UV}(u,v) \text{ since } f_{UV}(u,v) = 0 \text{ if } v>u$$

$$= e^{-(\theta+\lambda)}\sum_{v=0}^{u}\theta^{u-v}\lambda^v\frac{1}{(u-v)!v!}$$

$$= e^{-(\theta+\lambda)}\frac{1}{u!}\sum_{v=0}^{u}\theta^{u-v}\lambda^v\binom{u}{v}$$

$$= e^{-(\theta+\lambda)}\frac{1}{u!}(\theta+\lambda)^u \quad \text{via binomial formula}$$

Hence, $U = X+Y \sim \text{Poisson}(\theta+\lambda)$.

<u>EX</u>: Let $X \sim N(0,1)$, $Y \sim N(0,1)$, $X \perp\!\!\!\perp Y$

Via mgf, we showed $\begin{cases} X+Y \sim N(0,2) \\ X-Y \sim N(0,2) \end{cases}$

Now, via change-of-var formula,

let $\begin{cases} U = X+Y \\ V = X-Y \end{cases}$

the mapping $\begin{cases} u = g_1(x,y) := x+y \\ v = g_2(x,y) := x-y \end{cases}$ is one-to-one and admits the inverse:

$$\begin{cases} x = h_1(u,v) = \frac{1}{2}(u+v) \\ y = h_2(u,v) = \frac{1}{2}(u-v) \end{cases}$$

Hence, $J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}$

Thus, $f_{UV}(u,v) = f_{XY}(\frac{1}{2}(u+v), \frac{1}{2}(u-v)) \cdot \frac{1}{2}$

$$= f_X(\tfrac{1}{2}(u+v))f_Y(\tfrac{1}{2}(u-v)) \cdot \tfrac{1}{2}, \quad X \perp\!\!\!\perp Y$$

$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{1}{2}(u+v))^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{1}{2}(u-v))^2}\cdot\frac{1}{2}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\frac{1}{4}u^2}\frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\frac{1}{4}v^2}$$

4.19 (a) Let X_1 and X_2 be independent n$(0,1)$ random variables. Find the pdf of $(X_1 - X_2)^2/2$.

(b) If $X_i, i = 1,2$, are independent gamma$(\alpha_i, 1)$ random variables, find the marginal distributions of $X_1/(X_1+X_2)$ and $X_2/(X_1+X_2)$.

a) Let $X_1, X_2 \sim N(0,1)$ and independent.

Since both X_1 and X_2 are standard normals and independent:

$$X_1 - X_2 \sim N(0, \text{Var}(X_1-X_2)) = N(0, 1+1) = N(0,2)$$

Now standardize, letting $Z = \frac{X_1-X_2}{\sqrt{2}} \Rightarrow Z \sim N(0,1)$

Then

$$Y = \frac{(X_1-X_2)^2}{2} = \left(\frac{X_1-X_2}{\sqrt{2}}\right)^2 = Z^2 \sim \chi_1^2.$$

So, its pdf is $f(y) = \frac{1}{2^{1/2}\Gamma(1/2)}y^{-1/2}e^{-y/2} = \frac{1}{\sqrt{2}\sqrt{\pi}}y^{-1/2}e^{-y/2} = \frac{1}{\sqrt{2\pi y}}e^{-y/2}$, $y>0$

b) Let $X_1 \sim \text{Gamma}(\alpha_1, 1)$, $X_2 \sim \text{Gamma}(\alpha_2, 1)$, $X_1 \perp\!\!\!\perp X_2$.

Notice that $\frac{X_2}{X_1+X_2} = 1 - \frac{X_1}{X_1+X_2}$

i.e. both X_1 and X_2 complements of each other on $(0,1)$.

So, define the transformation:

$$\begin{cases} Y_1 = \frac{X_1}{X_1+X_2} \in (0,1) \\ Y_2 = X_1+X_2 \in (0,\infty) \end{cases}$$

This change-of-variables maps the random pair (X_1, X_2) to (Y_1, Y_2) with the following inverse transformation:

$Y_2 = X_1+X_2 \Rightarrow X_1+X_2 = Y_2$

$Y_1 = \frac{X_1}{X_1+X_2} \Rightarrow X_1 = Y_1Y_2$

$X_2 = Y_2 - X_1 = Y_2 - Y_1Y_2 = Y_2(1-Y_1)$

$$\begin{cases} X_1 = Y_1Y_2 \\ X_2 = (1-Y_1)Y_2 \end{cases}$$

Find the Jacobian determinant:

$$J = \begin{vmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ -y_2 & 1-y_1 \end{vmatrix} = y_2(1-y_1)+y_1y_2 = y_2$$

Since $X_1 \perp X_2$, the joint PDF is

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{\Gamma(\alpha_1)}x_1^{\alpha_1-1}e^{-x_1}\cdot\frac{1}{\Gamma(\alpha_2)}x_2^{\alpha_2-1}e^{-x_2}$$

Now, substituting $x_1 = y_1y_2$, $x_2 = (1-y_1)y_2$:

$$f_{X_1,X_2}(y_1y_2, (1-y_1)y_2) = \frac{(y_1y_2)^{\alpha_1-1}e^{-y_1y_2}}{\Gamma(\alpha_1)}\cdot\frac{[(1-y_1)y_2]^{\alpha_2-1}e^{-(1-y_1)y_2}}{\Gamma(\alpha_2)}.$$

Computing the joint PDF:

$$f_{Y_1Y_2}(y_1,y_2) = f_{X_1,X_2}(y_1y_2, (1-y_1)y_2)\cdot y_2$$

$$= \frac{y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\cdot y_2^{\alpha_1+\alpha_2-1}e^{-y_2}$$

$$= \underbrace{\left[\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1}\right]}\cdot\underbrace{\left[\frac{1}{\Gamma(\alpha_1+\alpha_2)}y_2^{\alpha_1+\alpha_2-1}e^{-y_2}\right]}$$

$\Rightarrow Y_1 \sim \text{Beta}(\alpha_1, \alpha_2)$, $Y_2 \sim \text{Gamma}(\alpha_1+\alpha_2, 1)$

Thus, $\boxed{\frac{X_1}{X_1+X_2} \sim \text{Beta}(\alpha_1, \alpha_2)}$

It can be shown that $1 - \frac{X_1}{X_1+X_2} = \boxed{\frac{X_2}{X_1+X_2} \sim \text{Beta}(\alpha_2, \alpha_1)}$:

PDF of $X \sim \text{Beta}(\alpha_1, \alpha_2)$ is

$$f_X(x) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}x^{\alpha_1-1}(1-x)^{\alpha_2-1}, \quad 0<x<1$$

Define $Y = 1-X \Rightarrow Y^{-1} = X = 1-Y$, $\left|\frac{dx}{dy}\right| = 1$.

Then $f_Y(y) = f_X(1-y)\cdot 1 = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}(1-y)^{\alpha_1-1}y^{\alpha_2-1} \sim \text{Beta}(\alpha_2, \alpha_1). \blacksquare$

4.23 For X and Y as in Example 4.3.3, find the distribution of XY by making the transformations given in (a) and (b) and integrating out V.

(a) $U = XY$, $V = Y$

(b) $U = XY$, $V = X/Y$

Let $X \sim \text{Beta}(\alpha, \beta)$, $Y \sim \text{Beta}(\alpha+\beta, \gamma)$, $X \perp\!\!\!\perp Y$.

The joint pdf of (X,Y) is

$$f_{XY}(x,y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha+\beta)\Gamma(\gamma)}y^{\alpha+\beta-1}(1-y)^{\gamma-1} \quad \begin{array}{|l} 0<x<1 \\ 0<y<1 \end{array} \quad (*)$$

a) Find the distribution of $U = XY$ by transforming to variables (U,V), computing the joint density $f_{U,V}(u,v)$, and integrating out V to find marginal dist. of U.

Define transformation: $= \begin{cases} U = XY \\ V = Y \end{cases} \Rightarrow$ inverse transformation: $= \begin{cases} X = U/V \\ Y = V \end{cases}$

$$J = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix} = \begin{vmatrix} 1/v & -u/v^2 \\ 0 & 1 \end{vmatrix} = \frac{1}{v}$$

Now substitute $x = (u/v)$ and $y = v$ into $(*)$ and compute

$$f_{U,V}(u,v) = f_X\left(\tfrac{u}{v}\right)\cdot f_Y(v)\cdot\left|\tfrac{1}{v}\right| \text{ since } X \perp\!\!\!\perp Y$$

$$= \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{u}{v}\right)^{\alpha-1}\left(1-\frac{u}{v}\right)^{\beta-1}\right]\cdot\left[\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha+\beta)\Gamma(\gamma)}v^{\alpha+\beta-1}(1-v)^{\gamma-1}\right]\cdot\frac{1}{v}$$

where $\begin{cases} \left(\frac{u}{v}\right)^{\alpha-1}\left(1-\frac{u}{v}\right)^{\beta-1}\cdot\frac{1}{v} = u^{\alpha-1}v^{\beta-1} \\ \left(1-\frac{u}{v}\right)^{\beta-1} = \left(\frac{v-u}{v}\right)^{\beta-1} = (v-u)^{\beta-1}v^{-(\beta-1)} \end{cases}$

$$= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}u^{\alpha-1}(v-u)^{\beta-1}(1-v)^{\gamma-1}, \quad 0<u<v<1 : \begin{cases} \text{constrained by } x\in(0,1),\ \frac{u}{v}\in(0,1) \Rightarrow 0<u<v \\ \text{constrained by } y\in(0,1),\ 0<v<1 \end{cases}$$

Next, $f_U(u) = \int_u^1 f_{UV}(u,v)dv = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}u^{\alpha-1}\int_u^1(v-u)^{\beta-1}(1-v)^{\gamma-1}dv$ via substitution:

letting $y = \frac{v-u}{1-u} \Rightarrow v = y(1-u)+u \Rightarrow dv = (1-u)dy$

when $v = u \Rightarrow y = 0$; $v = 1 \Rightarrow y = 1$

$v-u = y(1-u)$

$1-v = 1-y+yu+u$ $= (1-y)(1-u)$

Rewriting the integral:

$$\int_0^1[y(1-u)]^{\beta-1}[(1-y)(1-u)]^{\gamma-1}(1-u)dy$$

$$= (1-u)^{\beta-1+\gamma-1+1}\int_0^1 y^{\beta-1}(1-y)^{\gamma-1}dy$$

$f_{UV}(u,v)$ factorizes $\Rightarrow U \perp\!\!\!\perp V$, i.e. $(X+Y) \perp\!\!\!\perp (X-Y)$ and

$U \sim N(0,2)$
$V \sim N(0,2)$

__Ex__: Let $X, Y \sim N(0,1)$, $X \perp\!\!\!\perp Y$.

Suppose $\begin{cases} U = X/Y \\ V = |Y| \end{cases}$

note (x,y) and $(-x,-y)$ map to the same (u,v); (x,y) and $(-x,-y)$ produce the same (u,v)

so partition $A = \mathbb{R}^2 = \underbrace{\mathbb{R}\times\mathbb{R}_+}_{A_1} \cup \underbrace{\mathbb{R}\times\mathbb{R}_-}_{A_2} \cup \underbrace{\mathbb{R}\times\{0\}}_{A_3}$

If $(x,y) \in A_3$ $f_{XY}(x,y) = f_X(x) f_Y(y)$ but (u,v) is not defined; thus, not of interest since $P((X,Y)\in A_3) = 0$

__For A_1__: $\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x/y \\ y \end{pmatrix} \Rightarrow \begin{cases} x = uv \\ y = v \end{cases} \Rightarrow \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} uv \\ v \end{pmatrix}$

$J_1 = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$

__For A_2__: $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -uv \\ -v \end{pmatrix} \Rightarrow J_2 = \begin{vmatrix} -v & -u \\ 0 & -1 \end{vmatrix} = v$

Hence, $f_{UV}(u,v) = f_{XY}(h_1(u,v))|J_1| + f_{XY}(h_2(u,v))|J_2|$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(uv)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} |v| + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-uv)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-v)^2} |v|$$

$$= \frac{1}{\pi} e^{-\frac{1}{2}v^2(u^2+1)} v, \text{ for } v > 0 \quad ; = 0 \ \forall v < 0.$$

Thus, $f_U(u) = \int_0^\infty \frac{1}{\pi} e^{-\frac{1}{2}v^2(u^2+1)} v\, dv$

$$= \frac{1}{\pi(u^2+1)} e^{-\frac{1}{2}v^2(u^2+1)} \Big|_{v=\infty}^{v=0}$$

$$= \frac{1}{\pi(u^2+1)} \quad \forall u \in \mathbb{R}$$

$\Rightarrow U = X/Y \sim \text{Cauchy}(1)$.

4.4: Mixture & Hierarchical Models

__Ex__: Insect lays large number of eggs
Each insect survives w/ prob. p
Let X: # eggs themselves

__Q__: How to model behavior of X

Bring in (latent) RV Y which represents the number of eggs laid;
assume $Y \sim \text{Poisson}(\lambda)$
$\underbrace{X|Y}_{\text{mixture of binomial dist.}} \sim \text{Binomial}(Y, p)$

This defines a proper joint dist. for (X,Y) which induces a proper (marginal) dist. for X: $f_{XY}(x,y) = f_Y(y) f_{X|Y}(x|y)$

i.e. $\forall x, y \in \mathbb{N}$, $P(X=x, Y=y) = \underbrace{P(Y=y)}_{\text{Poisson}} \underbrace{P(X|Y)}_{\text{conditionally Binom.}}$

So,

$$P(X=x) = \sum_{y\in\mathbb{N}} P(X=x, Y=y)$$
$$= \sum_{y=0}^{\infty} P(Y=y) P(X=x|Y=y)$$
$$= \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} \binom{y}{x} p^x (1-p)^{y-x}, \quad y \geq x$$
$$= \sum_{y=x}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!} \frac{y!}{(y-x)!x!} p^x (1-p)^{y-x}$$
$$= \frac{\lambda^x p^x}{x!} e^{-\lambda} \sum_{y=x}^{\infty} \frac{\lambda^{y-x}}{(y-x)!} (1-p)^{y-x}$$

letting $t = y - x$

$$= \frac{\lambda^x p^x}{x!} e^{-\lambda} \sum_{t=0}^{\infty} \frac{\lambda^t (1-p)^t}{t!}$$
$$= \frac{\lambda^x p^x}{x!} e^{-\lambda} e^{\lambda(1-p)}$$
$$= \frac{(\lambda p)^x}{x!} e^{-\lambda p} \sim \text{Poisson}(\lambda p) \Rightarrow E(X) = E[E[X|Y]] = E[pY] = pE[Y] = \lambda p$$

__Note__: conditional expectation $E(X|Y)$ is a RV and hence
$E(X|Y) = Yp$
$\Rightarrow E(E(X|Y)) = E(Yp) = pE(Y) = \lambda p$

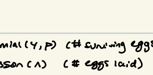
__Remarks__: · Usually, marginal dist. for X may not share the same family as that of the latent Y
· latent variable Y helps to explain the role of parameters λ and p

__Ex__: Let $Y \sim \text{Bernoulli}(p)$; $P(Y=0) = p$
$X|Y=0 \sim \text{Normal}(\mu_1, \sigma_1^2)$
$X|Y=1 \sim \text{Normal}(\mu_2, \sigma_2^2)$

Then
$$P(X \leq x) = P(X \leq x, Y=0) + P(X \leq x, Y=1)$$
$$= P(Y=0)P(X \leq x|Y=0) + P(Y=1)P(X \leq x|Y=1)$$
$$= p F_{N(\mu_1,\sigma_1^2)}(x) + (1-p) F_{N(\mu_2,\sigma_2^2)}(x)$$

Differentiating w.r.t. x:
$$f_X(x) = p f_{N(\mu_1,\sigma_1^2)}(x) + (1-p) f_{N(\mu_2,\sigma_2^2)}(x)$$

The RHS is the pdf of a mixture of normal distribution:



__Ex__: Let
$X|Y \sim \text{Binomial}(Y, p)$ (# surviving eggs)
$Y|\Lambda \sim \text{Poisson}(\Lambda)$ (# eggs laid)
$\Lambda \sim \text{Exponential}(\beta)$

The randomness of Λ captures the variation across the (insect) mothers

$$\forall x \in \mathbb{N}, \ P(X=x) = \sum_y P(X=x, Y=y)$$
$$\sum_{y=0}^{\infty} P(X=x|Y=y)P(Y=y)$$

$$= (1-u)^{\beta+\gamma-1} \cdot B(\beta,\gamma) = (1-u)^{\beta+\gamma-1} \cdot \frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta+\gamma)}$$

$$\Rightarrow f_U(u) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta+\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1}$$

$$\Rightarrow \boxed{U = XY \sim \text{Beta}(\alpha, \beta+\gamma)}$$

b) Find the distribution of $U = XY$ by transforming to variables (U,V), computing the joint density $f_{U,V}(u,v)$, and integrating out $V = X/Y$ to find marginal dist. of U.

From the transformation: $\begin{cases} U = XY \\ V = X/Y \end{cases}$

we can find the inverse transformation: $\begin{cases} X = \sqrt{UV} \\ Y = \sqrt{U/V} \end{cases}$ since: $\left| \begin{array}{l} UV = XY \cdot \frac{X}{Y} = X^2 \Rightarrow X = \sqrt{UV} \\ V = \frac{X}{Y} \Rightarrow Y = \frac{X}{V} = \frac{\sqrt{UV}}{V} = \sqrt{\frac{U}{V}} \end{array} \right.$

$$J = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} \frac{1}{2}\sqrt{v/u} & \frac{1}{2}\sqrt{u/v} \\ \frac{1}{2}\sqrt{1/uv} & -\frac{1}{2}\sqrt{u/v^2} \end{vmatrix} = -\frac{1}{4v} - \frac{1}{4v} = \frac{-1}{2v} \Rightarrow |J| = \frac{1}{2v}.$$

$$f_{U,V}(u,v) = f_{X,Y}(\sqrt{uv}, \sqrt{u/v}) \cdot \frac{1}{2v}$$
$$= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} \left[\sqrt{uv}^{\,\alpha-1} (1-\sqrt{uv})^{\beta-1} \sqrt{\tfrac{u}{v}}^{\,\alpha+\beta-1} \left(1-\sqrt{\tfrac{u}{v}}\right)^{\gamma-1} \right] \frac{1}{2v}, \quad 0<u<v<1/u<1$$

The set $\{0<x<1, 0<y<1\}$ is mapped onto the set $\{0<u<v<\frac{1}{u}, 0<u<1\}$:
constrained by $x \in (0,1)$, $x = \sqrt{uv} \in (0,1) \Rightarrow 0 < uv < 1 \Rightarrow v < 1/u$
constrained by $y \in (0,1)$, $y = \sqrt{u/v} \in (0,1) \Rightarrow 0 < u/v < 1 \Rightarrow u < v$
so, for $u \in (0,1)$, $u < v < 1/u$

Then
$$f_U(u) = \int_u^{1/u} f_{U,V}(u,v)\,dv = \underbrace{\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1}}_{A} \int_u^{1/u} \left(\frac{1-\sqrt{uv}}{1-u}\right)^{\beta-1} \left(\frac{1-\sqrt{u/v}}{1-u}\right)^{\gamma-1} \frac{(\sqrt{u/v})^{\beta}}{2v(1-u)}\,dv$$ via substitution:

letting $z = \frac{\sqrt{u/v}-u}{1-u} \Rightarrow \sqrt{u/v} = (1-u)z + u \Rightarrow v = \frac{u}{[(1-u)z+u]^2}$

$\Rightarrow dz = -\frac{\sqrt{u/v}}{2(1-u)v}dv$ and $\left| \begin{array}{l} v = u \to z = 1 \\ v = 1/u \to z = 0 \end{array} \right.$

$$= A \int z^{\beta-1}(1-z)^{\gamma-1}\,dz, \text{ i.e. the kernel of Beta}(\beta,\gamma)$$
$$= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1} \frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta+\gamma)}$$

$$\Rightarrow f_U(u) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta+\gamma)} u^{\alpha-1}(1-u)^{\beta+\gamma-1}$$

$$\Rightarrow \boxed{U = XY \sim \text{Beta}(\alpha, \beta+\gamma)} \quad \text{(as seen in (a))}$$

4.32 (a) For the hierarchical model

$$Y|\Lambda \sim \text{Poisson}(\Lambda) \quad \text{and} \quad \Lambda \sim \text{gamma}(\alpha, \beta)$$

find the marginal distribution, mean, and variance of Y. Show that the marginal distribution of Y is a negative binomial if α is an integer.

(b) Show that the three-stage model

$$Y|N \sim \text{binomial}(N, p), \quad N|\Lambda \sim \text{Poisson}(\Lambda), \quad \text{and} \quad \Lambda \sim \text{gamma}(\alpha, \beta)$$

leads to the same marginal (unconditional) distribution of Y.

a) Suppose $Y|\Lambda \sim \text{Poisson}(\Lambda)$
$\Lambda \sim \text{Gamma}(\alpha, \beta)$.

For $y = 0, 1, 2, \dots$ the marginal dist. for Y is:
$$f_Y(y) = \int_0^\infty f_{Y|\Lambda}(y|\lambda) f_\Lambda(\lambda)\,d\lambda$$
$$= \int_0^\infty \frac{\lambda^y e^{-\lambda}}{y!} \cdot \frac{\lambda^{\alpha-1}e^{-\lambda/\beta}}{\Gamma(\alpha)\beta^\alpha}\,d\lambda$$
$$= \frac{1}{y!\,\Gamma(\alpha)\beta^\alpha} \underbrace{\int_0^\infty \lambda^{y+\alpha-1} \exp\left(-\lambda\left(1+\tfrac{1}{\beta}\right)\right) d\lambda}_{\text{kernel of the Gamma dist}: \int_0^\infty \lambda^{c-1}e^{-d\lambda} = \frac{\Gamma(c)}{d^c}}$$

FS where $\left| \begin{array}{l} c = y + \alpha \\ d = 1 + \frac{1}{\beta} = \frac{\beta+1}{\beta} \end{array} \right.$

$$= \frac{1}{y!\,\Gamma(\alpha)\beta^\alpha} \cdot \frac{\Gamma(y+\alpha)}{\left(\frac{\beta+1}{\beta}\right)^{y+\alpha}}$$
$$= \frac{\Gamma(y+\alpha)}{y!\,\Gamma(\alpha)} \left(\frac{\beta}{\beta+1}\right)^y \left(\frac{1}{\beta+1}\right)^\alpha$$

FS

$$= \binom{y+\alpha-1}{y} \left(\frac{\beta}{\beta+1}\right)^y \left(\frac{1}{\beta+1}\right)^\alpha.$$

If α is a positive integer, then

$$\boxed{Y \sim \text{NegBin}\left(\alpha, p = \tfrac{1}{1+\beta}\right)}$$

Using the law of total expectation,
$$\boxed{E(Y) = E(E(Y|\Lambda)) = E(\Lambda) = \alpha\beta}$$

Similarly, using the law of total variance,
$$\text{var}(Y) = E[\text{var}(Y|\Lambda)] + \text{var}[E(Y|\Lambda)]$$
$$= E(\Lambda) + \text{var}(\Lambda)$$
$$= \alpha\beta + \alpha\beta^2$$
$$\boxed{\text{var}(Y) = \alpha\beta(1+\beta)}$$

b) Suppose $Y|N \sim \text{Binomial}(N, p)$
$N|\Lambda \sim \text{Poisson}(\Lambda)$
$\Lambda \sim \text{Gamma}(\alpha, \beta)$

Show that this 3-stage model leads to the same marginal (unconditional) dist. of Y

For $y = 0, 1, \dots$
$$P(Y=y|\Lambda=\lambda) = \sum_{n=y}^{\infty} P(Y=y|N=n, \Lambda=\lambda) P(N=n|\Lambda=\lambda)$$

where $\begin{cases} P(Y=y|N=n, \Lambda=\lambda) = \binom{n}{y} p^y (1-p)^{n-y} \\ P(N=n|\Lambda=\lambda) = \frac{e^{-\lambda}\lambda^n}{n!} \end{cases}$

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

where $\binom{n}{y} = \frac{n!}{y!(n-y)!}$

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

via change-of-var: letting $m = n - y$
$\Rightarrow n = m + y$

$$= \frac{p^y e^{-\lambda}\lambda^y}{y!} \sum_{m=0}^{\infty} \frac{[(1-p)\lambda]^m}{m!}$$

$\forall\, y \in \mathbb{N},\ P(Y=y) = P(Y=y,\ 0<\Lambda<\infty)$

$$= \int_0^\infty P(Y=y \mid \Lambda=\lambda)\, f_\Lambda(\lambda)\, d\lambda$$

$$= \int_0^\infty e^{-\lambda}\frac{\lambda^y}{y!}\,\frac{1}{\beta} e^{-\lambda/\beta}\, d\lambda$$

$$= \frac{1}{\beta y!}\int_0^\infty \lambda^y e^{-\lambda(1+\frac{1}{\beta})}\, d\lambda$$

$$= \frac{1}{\beta y!}\Gamma(y+1)\left(\frac{\beta}{1+\beta}\right)^{y+1}$$

$$= \frac{1}{1+\beta}\left(\frac{\beta}{1+\beta}\right)^{y}$$

So $Y \sim \text{NegBinom}(p=\frac{1}{1+\beta},\ r=1) = \text{Geo}(p=\frac{1}{1+\beta})$ (geometric)

Hence, the three-stage hierarchical model is equivalent mixture via (X,Y) (w/ Λ integrated out):

$X \mid Y \sim \text{Binomial}(Y, p)$

$Y \sim \text{NegBinom}(p=\frac{1}{1+\beta},\ r=1)$

Ex (Beta-Binomial (dist'n)): Let $X\mid\theta \sim \text{Binomial}(n,\theta)$, $\theta\in(0,1)$

$\theta \sim \text{Beta}(\alpha,\beta)$

Then $P(X=x) = \int P(X=x\mid\theta)\, f_\theta(\theta)\, d\theta$

$$= \int \binom{n}{x}\theta^x(1-\theta)^{n-x} f_\theta(\theta)\, d\theta$$

$$= \int \binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta$$

$$= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int \theta^{x+\alpha-1}(1-\theta)^{n+\beta-x-1}\, d\theta$$

$$= \frac{n!}{x!(n-x)!}\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

$$= \frac{n!}{x!(n-x)!}\,\frac{\overbrace{(x+\alpha-1)\cdots\alpha}^{x}\ \overbrace{(n-x+\beta-1)\cdots\beta}^{n-x \text{ terms}}}{\underbrace{(n+\alpha+\beta-1)\cdots(\alpha+\beta)}_{n \text{ terms}}}$$

Direct calculation of EX and VX possible but quite complicated

Now $EX = E(E(X\mid\theta))$

$= E(n\theta)$

$= nE\theta$

$= \frac{n\alpha}{\alpha+\beta}$

$\text{Var}\,E[X\mid\theta] = \text{Var}(n\theta) = n^2V(\theta) = \frac{n^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ (*)

$E\text{Var}[X\mid\theta] = E[n\theta(1-\theta)] = nE\theta - nE\theta^2$

$$= \frac{n\alpha}{\alpha+\beta} - \frac{n\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

$$= \frac{n\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} \quad (**)$$

So, $\text{Var}X = (*) + (**)$

$$= \frac{n^2\alpha\beta + n\alpha\beta(\alpha+\beta)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$= \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

recall (Power series expansion): $e^a = \sum_{m=0}^{\infty}\frac{a^m}{m!}$

FS where sum is Taylor expansion of $e^{(1-p)\lambda}$

$$= \frac{p^y e^{-\lambda}\lambda^y}{y!}\cdot e^{(1-p)\lambda}$$

$$= \frac{(p\lambda)^y}{y!}e^{-p\lambda}$$

i.e. the PMF of the Poisson dist. w/ mean $p\lambda$:

$Y\mid\Lambda=\lambda \sim \text{Poisson}(p\lambda)$

Now integrate over the Gamma prior on $\Lambda\sim\text{Gamma}(\alpha,\beta)$:

$$f_Y(y) = \int_0^\infty P(Y=y\mid\lambda)\cdot f_\Lambda(\lambda)\, d\lambda$$

where $\begin{cases} f_\Lambda(\lambda) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\lambda^{\alpha-1}e^{-\lambda/\beta} \\ P(Y=y\mid\lambda) = \frac{(p\lambda)^y}{y!}e^{-p\lambda}\end{cases}$

$$= \frac{p^y}{y!\Gamma(\alpha)\beta^\alpha}\underbrace{\int_0^\infty \lambda^{y+\alpha-1}e^{-\lambda(p+\frac{1}{\beta})}}_{\text{kernel of Gamma dist} \;=\; \Gamma(y+\alpha)\cdot\left(\frac{\beta}{1+p\beta}\right)^{y+\alpha}}$$

$$= \frac{1}{y!\Gamma(\alpha)(p\beta)^\alpha}\cdot\Gamma(y+\alpha)\cdot\left(\frac{p\beta}{1+p\beta}\right)^{y+\alpha}$$

where, again if α is a positive integer, then

$$\boxed{Y\sim \text{NegBinom}(\alpha, \tfrac{1}{1+p\beta})}$$

4.34 (a) For the hierarchy in Example 4.4.6, show that the marginal distribution of X is given by the *beta-binomial distribution*,

$$P(X=x) = \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)}.$$

(b) A variation on the hierarchical model in part (a) is

$$X|P \sim \text{negative binomial}(r, P) \quad\text{and}\quad P\sim\text{beta}(\alpha,\beta).$$

Find the marginal pmf of X and its mean and variance. (This distribution is the *beta-Pascal*.)

a) Suppose

$X\mid P \sim \text{Binomial}(P)$, $i=1,\dots,n$

$P \sim \text{Beta}(\alpha,\beta)$

$EX = E[E(X\mid P)] = E(nP) = n\frac{\alpha}{\alpha+\beta}$

Show that the marginal dist. follows a beta-binomial distribution:

$$P(X=x) = \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)}$$

Start w/ the conditional PMF:

$$P(X=x\mid P=p) = \binom{n}{x}p^x(1-p)^{n-x}$$

We also know for $P\sim\text{Beta}(\alpha,\beta)$:

$$f_P(p) = \frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}$$

Now

$$P(X=x) = \binom{n}{x}\int_0^1 p^x(1-p)^{n-x}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}\, dp$$

$$= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\underbrace{\int_0^1 p^{x+\alpha-1}(1-p)^{n-x+\beta-1}\, dp}$$

FS Beta Function $:= \int_0^1 p^{a-1}(1-p)^{b-1}\, dp = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

where here $a = x+\alpha$, $b = n-x+\beta$

$$= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}. \blacksquare$$

b) Suppose $X\mid P\sim \text{NegBin}(r, P)$

$P\sim\text{Beta}(\alpha,\beta)$

Start with

$$f_X(x) = \int_0^1 f_{X\mid P}(x\mid p)\, f_P(p)\, dp$$

where $\begin{cases} f_{X\mid P}(x\mid p) = \binom{r+x-1}{x}p^r(1-p)^x \\ \text{Beta} := f_P(p) = \frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1}\end{cases}$

$$= \int_0^1\binom{r+x-1}{x}p^r(1-p)^x\cdot\frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1}\, dp$$

$$= \binom{r+x-1}{x}\cdot\frac{1}{B(\alpha,\beta)}\int_0^1 p^{r+\alpha-1}(1-p)^{x+\beta-1}\, dp$$

where $\int_0^1 p^{a-1}(1-p)^{b-1}\, dp = B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

$$= \binom{r+x-1}{x}\cdot\frac{1}{B(\alpha,\beta)}\cdot B(r+\alpha, x+\beta)$$

$$\boxed{f_X(x) = \underbrace{\binom{r+x-1}{x}\cdot\frac{\Gamma(r+\alpha)\Gamma(x+\beta)}{\Gamma(r+\alpha+x)}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{\text{Beta-Pascal Dist.}},\quad x=0,1,\dots}$$

Now use the law of iterated expectation and variance:

$$\boxed{EX = E[E(X\mid P)] = E\left[r\cdot\frac{1-P}{P}\right] = r\cdot E\left[\frac{1-P}{P}\right] = \frac{r\beta}{\alpha-1}}$$

Since $E\left[\frac{1-P}{P}\right] = \int_0^1\left(\frac{1-p}{p}\right)\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}\, dp$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 p^{\alpha-1-1}(1-p)^{\beta+1-1}\, dp$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{\Gamma(\alpha-1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta)} = \frac{\beta}{\alpha-1}$$

Similarly,

$$\boxed{\begin{aligned}\text{Var}X &= E[\text{Var}(X\mid P)] + \text{Var}[E(X\mid P)]\\ &= E\left[\frac{r(1-P)}{P^2}\right] + \text{Var}\left(\frac{r(1-P)}{P}\right)\\ &= r\,\frac{(\beta+1)(\alpha+\beta)}{\alpha(\alpha-1)} + r^2\,\frac{\beta(\alpha+\beta-1)}{(\alpha-1)^2(\alpha-2)}\end{aligned}}$$

Since $E\left[\frac{1-P}{P^2}\right] = \int_0^1\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-2-1}(1-p)^{\beta+1-1}\, dp = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{\Gamma(\alpha-2)\Gamma(\beta+1)}{\Gamma(\alpha+\beta-1)} = \frac{(\beta+1)(\alpha+\beta)}{\alpha(\alpha-1)}$

and $\text{Var}\left(\frac{1-P}{P}\right) = E\left[\left(\frac{1-P}{P}\right)^2\right] - \left[E\left(\frac{1-P}{P}\right)\right]^2 = \frac{\beta(\beta+1)}{(\alpha-2)(\alpha-1)} - \left(\frac{\beta}{\alpha-1}\right)^2 = \frac{\beta(\alpha+\beta-1)}{(\alpha-1)^2(\alpha-2)}$

where $E\left[\left(\frac{1-P}{P}\right)^2\right] = \int_0^1\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\frac{\Gamma(\alpha-2)\Gamma(\beta+2)}{\Gamma(\alpha-2+\beta+2)} = \frac{\beta(\beta+1)}{(\alpha-2)(\alpha-1)}$

4.5: covariance and correlation

Proof. Show that $\text{corr}(X,Y)$ is the covariance of the standardized X and Y:

let $X' = \frac{X-EX}{\sqrt{\text{Var}X}}$, $Y' = \frac{Y-EY}{\sqrt{\text{Var}Y}}$

then $\begin{cases} EX' = EY' = 0 \\ VX' = VY' = 1\end{cases}$

By definition, $\text{corr}(X',Y') = \text{Cov}(X',Y')$

$= E(X'Y')$

$= E\,\frac{(X-EX)}{\sqrt{\text{Var}X}}\,\frac{(Y-EY)}{\sqrt{\text{Var}Y}}$

$= \frac{E(X-EX)(Y-EY)}{\sqrt{\text{Var}X}\,\sqrt{\text{Var}Y}}$

$= \text{Corr}(X,Y)$. ∎

Result: Cauchy-Schwartz Inequality: $E[XY] \le \sqrt{E(X^2)}\sqrt{E(Y^2)}$

$\Leftrightarrow$

$[E(XY)]^2 \le E(X^2)E(Y^2)$

Letting $U = X-\mu_X$, $V = Y-\mu_Y$,

$[E(UV)]^2 \le E(U^2)E(V^2)$

$[\text{Cov}(X,Y)]^2 \le \text{Var}(X)\text{Var}(Y)$

$\frac{[\text{Cov}(X,Y)]^2}{V(X)V(Y)} \le 1$

$\left|\frac{\text{Cov}(X,Y)}{\sqrt{VX}\sqrt{VY}}\right| \le 1$

$|\text{Corr}(X,Y)| \le 1$

$E[XY] \le \sqrt{E(X^2)}\sqrt{E(Y^2)}$

Equality holds i.f.f. $Y = cX$ a.s., i.e.

$P(Y=cX) = 1$, c: constant $\in\mathbb{R}$

Letting $U = X-E_X$, $V = Y-E_Y$,

equality holds i.f.f. $P([Y-EY] = a[X-E_X]) = 1$, a: constant $\in\mathbb{R}$, i.e.

$Y - EY = a(X-EX)$ a.s.

$\Rightarrow \text{Cov}(X,Y) = E[(X-EX)(Y-EY)] = E[(X-EX)a(X-EX)] = aE[(X-EX)^2] = a\,\text{Var}(X)$

$\text{Var}(Y) = E[(Y-EY)^2] = E[a^2(X-EX)^2] = a^2\text{Var}(X)$

$\text{Corr}(X,Y) = \frac{a\,\text{Var}(X)}{\sqrt{\text{Var}(X)}\sqrt{a^2\text{Var}(X)}} = \frac{a}{|a|} = \begin{cases}+1, & a>0\\ -1, & a<0\end{cases}$

$\Rightarrow |\text{Corr}(X,Y)| = 1$

EX: if $\mathrm{corr}(X,Y) \approx 1 \iff Y \approx bX+c,\ b>0$
$\mathrm{corr}(X,Y) \approx -1 \iff Y \approx bX+c,\ b<0$

let $X \sim \mathrm{Unif}(0,1)$, $Z \perp\!\!\!\perp X$
$Z \sim \mathrm{unif}(0, 1/10)$
$Y = X+Z$

$$\begin{aligned} \mathrm{cov}(X,Y) &= EXY - EXEY \\ &= EX(X+Z) - EX(EX+EZ) \\ &\overset{\text{L.O.E}}{=} EX^2 + EXZ - (EX)^2 - (EX)(EZ) \end{aligned}$$

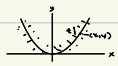
$X \perp\!\!\!\perp Z \Rightarrow EXZ - (EX)(EZ) = 0$

$$\begin{aligned} &= \mathrm{var}X \\ &= 1/12 \end{aligned}$$

Hence, $\mathrm{corr}(X,Y) = \dfrac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}X\,\mathrm{var}Y}} = \dfrac{1/12}{\sqrt{1/12}\sqrt{1/12+1/1200}}$

$$= \frac{1}{\sqrt{1+\frac{1}{100}}} = \left(\frac{100}{101}\right)^{1/2} \approx 1$$

EX: Let $X \sim \mathrm{unif}(-1,1)$
$Z \sim \mathrm{unif}(0, 1/10)$, $Z \perp\!\!\!\perp X$
$Y = X^2+Z$



$$\begin{aligned} \text{then } \mathrm{cov}(X,Y) &= EXY - (EX)(EY) \\ &= EX(X^2+Z) - (EX)(EX^2+EZ) \\ &= EX^3 + EXEZ - (EX)(EX^2) - (EX)(EZ) \\ &= EX^3 - EXEX^2 \end{aligned}$$

since X symmetric around 0

$= 0 - 0 = 0$ does not mean $X \perp\!\!\!\perp Y$; in fact, they are quadratically related

EX: Let $Z = (X,Y) \sim N(\mu, \Sigma)$ $\Big|\ \mu \in \mathbb{R}^2$, Σ positive definite, symmetric $\in \mathbb{R}^{2\times 2}$

$$f_{XY}(x,y) := f_Z(z) := \frac{1}{((2\pi)^2|\Sigma|)^{1/2}} \exp\left\{-\frac{1}{2}(z-\mu)^T\Sigma^{-1}(z-\mu)\right\} \quad (*)$$

Translating (*) into univariate forms (don't need to remember):

$$\begin{aligned} |\Sigma| = \begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix} &= \Sigma_{11}\Sigma_{22} - \Sigma_{21}\Sigma_{22} \\ &= \sigma_X^2\sigma_Y^2 - \sigma_{XY}^2 \quad \text{where } \sigma_{XY}: \mathrm{cov}(X,Y) \\ &= \sigma_X^2\sigma_Y^2(1-\rho_{XY}^2) \end{aligned}$$

where $\rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y} = \mathrm{corr}(X,Y)$

$\mu = (\mu_1, \mu_2)^T \equiv (\mu_X, \mu_Y)^T$

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\begin{pmatrix} \Sigma_{22} & -\Sigma_{21} \\ -\Sigma_{12} & \Sigma_{11} \end{pmatrix}, \quad A^{-1} = \frac{1}{|A|}A$$

$$= \frac{1}{\sigma_X^2\sigma_Y^2(1-\rho_{XY}^2)}\begin{pmatrix} \sigma_Y^2 & -\sigma_{XY} \\ -\sigma_{XY} & \sigma_X^2 \end{pmatrix}$$

$$= \frac{1}{1-\rho_{XY}^2}\begin{pmatrix} \frac{1}{\sigma_X^2} & \frac{-\sigma_{XY}}{\sigma_X^2\sigma_Y^2} \\ \frac{-\sigma_{XY}}{\sigma_X^2\sigma_Y^2} & \frac{1}{\sigma_Y^2} \end{pmatrix}$$

$$\frac{-1}{2}(z-\mu)^T\Sigma^{-1}(z-\mu) = -\frac{1}{2(1-\rho_{XY}^2)}\left\{\frac{1}{\sigma_X^2}(x-\mu_X)^2 - \frac{2\sigma_{XY}}{\sigma_X^2\sigma_Y^2}(x-\mu_X)(y-\mu_Y) + \frac{1}{\sigma_Y^2}(y-\mu_Y)^2\right\}$$

where $2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) = \frac{2\sigma_{XY}}{\sigma_X^2\sigma_Y^2}(x-\mu_X)(y-\mu_Y)$

Hence,

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y(1-\rho_{XY}^2)^{1/2}} \times \exp\left\{\frac{-1}{2(1-\rho_{XY}^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

4.6: Multivariate distributions

Ex (Multinomial dist.): $\begin{cases} m \text{ dependent trials} \\ n \text{ possible outcomes w/} \\ \text{"cell prob.'s" } (p_1, \dots, p_n),\ \sum_{i=1}^n p_i = 1 \end{cases}$

Let $(X_1, \dots, X_n)$: counts of each of n outcomes
$X_1 + \dots + X_n = m$

then $f(x_1, \dots, x_n) = \binom{m}{x_1, \dots, x_n} p_1^{x_1} \cdots p_n^{x_n} = \dfrac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n}$

Marginal distributions:

$$f_{X_n}(x_n) = \sum_{x_1+\dots+x_{n-1}=m-x_n} \frac{m!}{x_1!\cdots x_{n-1}!\,x_n!} p_1^{x_1}\cdots p_{n-1}^{x_{n-1}} p_n^{x_n}$$

$$= p_n^{x_n}\frac{m!}{x_n!(m-x_n)!} \sum_{\substack{x_1+\dots+x_{n-1} \\ = m-x_n}} \frac{(m-x_n)!}{x_1!\cdots x_{n-1}!} p_1^{x_1}\cdots p_{n-1}^{x_{n-1}}$$

$$= \frac{m!}{x_n!(m-x_n)!} p_n^{x_n}(p_1+\dots+p_{n-1})^{m-x_n}$$

via multinomial formula: $\displaystyle\sum_{\substack{x_1,\dots,x_n \\ x_1+\dots+x_n=m}} \binom{m}{x_1 \cdots x_n} p_1^{x_1}\cdots p_n^{x_n} = (p_1+\dots+p_n)^m$

$$= \frac{m!}{x_n!(m-x_n)!} p_n^{x_n}(1-p_n)^{m-x_n}$$

$\Rightarrow X_n \sim \mathrm{Binomial}(m, p_n)$.

Conditional distributions

$$f_{X_1,\dots,X_{n-1}|X_n}(x_1,\dots,x_{n-1}|x_n) = \frac{f_{X_1,\dots,X_n}(x_1,\dots,x_n)}{f_{X_n}(x_n)}$$

$$= \frac{\frac{m!}{x_1!\cdots x_n!}p_1^{x_1}\cdots p_{n-1}^{x_{n-1}}p_n^{x_n}}{\frac{m!}{x_n!(m-x_n)!}(1-p_n)^{m-x_n}p_n^{x_n}}$$

$$= \frac{(m-x_n)!}{x_1!\cdots x_{n-1}!}\left(\frac{p_1}{1-p_n}\right)^{x_1}\cdots\left(\frac{p_{n-1}}{1-p_n}\right)^{x_{n-1}}$$

Hence $X_1 \cdots X_{n-1} | X_n = x_n \sim \mathrm{Multinomial}(m - x_n \text{ trial, cell prob: } \left(\frac{p_1}{1-p_n}, \dots, \frac{p_{n-1}}{1-p_n}\right))$

4.42 Let X and Y be independent random variables with means μ_X, μ_Y and variances σ_X^2, σ_Y^2. Find an expression for the correlation of XY and Y in terms of these means and variances.

We want $\rho_{XY,Y} = \dfrac{\mathrm{cov}(XY,Y)}{\sigma_{XY}\sigma_Y}$.

First compute the covariance:

$$\begin{aligned} \mathrm{cov}(XY,Y) = E[XY\cdot Y] - E[XY]\cdot E[Y] &= E(XY^2) - EXEYEY,\ X \perp\!\!\!\perp Y \\ &= EXEY^2 - EXEYEY,\ X \perp\!\!\!\perp Y \\ &= \mu_X E(Y^2) - \mu_X\mu_Y^2 \\ &= \mu_X[E(Y^2) - \mu_Y^2] \\ &= \mu_X\sigma_Y^2 \end{aligned}$$

Next, compute

$\mathrm{var}(XY) = E[(XY)^2] - (E(XY))^2$
where $E[(XY)^2] = E(X^2)E(Y^2) = (\sigma_X^2+\mu_X^2)(\sigma_Y^2+\mu_Y^2)$
and $(E(XY))^2 = (\mu_X\mu_Y)^2$

$$\begin{aligned} \Rightarrow \mathrm{var}(XY) &= (\sigma_X^2+\mu_X^2)(\sigma_Y^2+\mu_Y^2) - \mu_X^2\mu_Y^2 \\ &= \sigma_X^2\sigma_Y^2 + \mu_X^2\sigma_Y^2 + \sigma_X^2\mu_Y^2 \end{aligned}$$

Therefore,

$$\boxed{\rho_{XY,Y} = \frac{\mathrm{cov}(XY,Y)}{\sqrt{\mathrm{var}(XY)}\cdot\sigma_Y} = \frac{\mu_X\sigma_Y^2}{\sigma_Y\cdot\sqrt{\sigma_X^2\sigma_Y^2 + \mu_X^2\sigma_Y^2 + \sigma_X^2\mu_Y^2}}}$$

5.3 Let $X_1, \dots, X_n$ be iid random variables with continuous cdf F_X, and suppose $EX_i = \mu$. Define the random variables $Y_1, \dots, Y_n$ by

$$Y_i = \begin{cases} 1 & \text{if } X_i > \mu \\ 0 & \text{if } X_i \le \mu. \end{cases}$$

Find the distribution of $\sum_{i=1}^n Y_i$.

Note that $Y_i \sim$ Bernoulli with

$p = P(Y_i = 1) = P(X_i > \mu) = P(X_i \ge \mu) = 1 - F(\mu)\ \forall i$

Then, $Y_1, \dots, Y_n \overset{iid}{\sim} \mathrm{Bernoulli}(p = 1 - F(\mu))$.

Thus, $\sum_{i=1}^n Y_i \sim \mathrm{Binomial}(n, p = 1 - F(\mu))$

Chapter 5: Properties of a Random Sample

5.1: iid samples

Ex: $X_1,\dots,X_n \overset{iid}{\sim} \text{Exp}(\beta)$, $\beta>0$

Then, we may compute, say

$P(X_1>a, X_2>a_2, \dots, X_n>a_n)$, etc

Proof. Let $X_1,\dots,X_n$ be an n-iid sample from a population with mean μ, variance $\sigma^2<\infty$

$E\bar{X} = E\frac{1}{n}(X_1+\dots+X_n)$

$= \frac{1}{n}(EX_1+\dots+EX_n)$ by LOE

$= \frac{1}{n}(n\mu) = \mu$

$\text{var}\bar{X} = \text{var}\frac{1}{n}(X_1+\dots+X_n)$

$= \frac{1}{n^2}\text{var}(X_1+\dots+X_n)$

$= \frac{1}{n^2}(\text{v}X_1+\dots+\text{v}X_n)$, since X_i - independent

$= \frac{1}{n^2}\, n\cdot\text{var}X_1$

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

$ES^2 = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$

where $\sum(X_i-\bar{X})^2 = \sum(X_i-\mu+\mu-\bar{X})^2$
$= \sum(X_i-\mu)^2 - n(\bar{X}-\mu)^2$

and $E(S^2) = \frac{1}{n-1}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right]$, $\text{v}\bar{X}=\sigma^2/n$

$= \frac{1}{n-1}n\sigma^2(1-\frac{1}{n})$

$= \frac{1}{n-1}n\sigma^2(\frac{n-1}{n}) = \sigma^2$.

Proof (sketch): Show the convolutional identity

Let X,Y - discrete, f_X, f_Y, $Z=X+Y$.

Then

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

$= \sum_{x,y} P(X=x, Y=y)$

$x+y=z \Rightarrow y=z-x$

$= \sum_x f_X(x) f_Y(z-x)$

Suppose $X_1,\dots,X_n \overset{iid}{\sim} F$

$X_1+X_2 \sim F\cdot F$

$(X_1+X_2)+X_3 \sim (F\cdot F)F = F\cdot(F\cdot F)$

$X_1+\dots+X_n \sim \underbrace{F\cdot F\cdots F}_{n\text{ terms}}$

Formal proof uses change-of-vars for mapping $(X,Y)\to(X+Y,X)$:

Let $Z=X+Y$, $X\sim f_X$, $Y\sim f_Y$, X,Y - independent

want to show $f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx$

Define transformation: $U=X+Y$, $V=X$ $\Rightarrow$ inverse transformation: $X=V$, $Y=U-V$

Need to find joint pdf of (U,V) and marginalize over V to get $f_U(u) = f_Z(z)$

$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1\end{vmatrix} \Rightarrow |J|=1$

Since X and Y are independent,

$f_{XY}(x,y) = f_X(x)f_Y(y)$

$\Rightarrow f_{UV}(u,v) = f_X(v)f_Y(u-v)\cdot 1 = f_X(v)f_Y(u-v)$

Now, $f_U(u) = \int_{-\infty}^{\infty} f_{UV}(u,v)\,dv = \int_{-\infty}^{\infty} f_X(v)f_Y(u-v)\,dv$

$\Rightarrow f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx$.

5.2: useful classical facts

Suppose $X_1,\dots,X_n \overset{iid}{\sim} N(\mu,\sigma^2)$

$\bar{X} = \frac{1}{n}(X_1+\dots+X_n)$

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$

In subsequent proofs, it suffices to show for $\sigma=1$

Proof. Show that $\bar{X}$ and S^2 are independent RV's:

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

$= (X_1-\bar{X})^2 + \sum_{i=2}^{n}(X_i-\bar{X})^2$

defining $Y_i := X_i-\bar{X}$, $i=2,\dots,n$

where $\sum_{i=1}^{n}(X_i-\bar{X})=0 \Rightarrow (X_1-\bar{X}) + \sum_{i=2}^{n}(X_i-\bar{X})=0$

$\Rightarrow (X_1-\bar{X}) = -\sum_{i=2}^{n}(X_i-\bar{X}) = -\sum_{i=2}^{n}Y_i$

$= \left(\sum_{i=2}^{n}Y_i\right)^2 + \sum_{i=2}^{n}Y_i^2$

Use change-of-var. formula for the mapping

$(X_1,\dots,X_n)\to(Y_1,\dots,Y_n)$ where $\begin{cases}Y_1=\bar{X}\\ Y_i = X_i-\bar{X},\ i\ge 2\end{cases}$

to find then

$f_{Y_1,\dots,Y_n}(y_1,\dots,y_n) = \left(\frac{n}{2\pi}\right)^{1/2} e^{-ny_1^2/2}\, \frac{n^{1/2}}{(2\pi)^{(n-1)/2}} e^{-\frac{1}{2}\left(\left(\sum_{i=2}^{n}y_i\right)^2 + \sum_{i=2}^{n}y_i^2\right)}$

which factorizes $\Rightarrow Y_1 \perp\!\!\!\perp (Y_2,\dots,Y_n)$

and S^2 is a function of $(Y_2,\dots,Y_n)$ s.t.

$\Rightarrow (n-1)S^2 \perp\!\!\!\perp Y_1 = \bar{X}$

Proof. Show that $(n-1)S^2 \sim \chi^2_{n-1}$:

Let $\bar{X}_k = \frac{1}{k}(X_1+\dots+X_k)$

$S_k^2 = \frac{1}{k-1}\sum_{i=1}^{k}(X_i-\bar{X}_k)^2$

Proof by induction that $(k-1)S_k^2 \sim \chi^2_{k-1}\ \forall k\ge 2$

For $k=2$, $S_2^2 = (X_1-\bar{X}_2)^2 + (X_2-\bar{X}_2)^2$

$= \frac{1}{2}(X_1-X_2)^2$

$X_1-X_2 \sim N(0,2) \Rightarrow \left(\frac{1}{\sqrt{2}}(X_1-X_2)\right)^2 \sim N(0,1)^2 = \chi^2_1$

Suppose now $(k-2)S_{k-1}^2 \sim \chi^2_{k-1}$

we'll be done by proving that $(k-1)S_k^2 \sim \chi^2_k$

5.6 If X has pdf $f_X(x)$ and Y, independent of X, has pdf $f_Y(y)$, establish formulas, similar to (5.2.3), for the random variable Z in each of the following situations.

(a) $Z = X - Y$

(b) $Z = XY$

(c) $Z = X/Y$

c) For $\begin{cases} Z = X/Y \\ W = X\end{cases} \Rightarrow \begin{cases} Y = W/Z \\ X = W\end{cases}$ and $|J| = \begin{vmatrix} 0 & 1 \\ -w/z^2 & 1/z\end{vmatrix} = w/z^2$

so,

$f_{Z,W}(z,w) = f_X(w)f_Y(w/z)\cdot|w/z^2|$

$\Rightarrow \boxed{f_Z(z) = \int_{-\infty}^{\infty}|w/z^2|\, f_X(w)f_Y(w/z)\,dw}$

a) Define transformation $\begin{cases} Z = X-Y \\ W = X\end{cases} \Rightarrow$ inverse transformation $\begin{cases} X = W \\ Y = W-Z\end{cases}$

we have a one-to-one transformation: $(X,Y)\to(W, W-Z)$

Next compute $J = \left[\frac{\partial(X,Y)}{\partial(Z,W)}\right] = \begin{bmatrix}\frac{\partial X}{\partial z} & \frac{\partial X}{\partial w}\\ \frac{\partial Y}{\partial z} & \frac{\partial Y}{\partial w}\end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 1\end{bmatrix} \Rightarrow |J| = 0+1 = |1| = 1$.

Using independence,

$f_{Z,W}(z,w) = f_X(x)f_Y(y)\cdot|J| = f_X(w)\,f_Y(w-z)\cdot|J|$

$= f_X(w)f_Y(w-z)$

Now marginalize (out) w to find

$\boxed{f_Z(z) = \int_{-\infty}^{\infty} f_X(w)f_Y(w-z)\,dw}$

$(\Leftrightarrow f_X(x) = \int_{\mathbb{R}} f_X(x)f_Y(x-z)\,dx$

b) Define transformation $\begin{cases} Z = XY \\ W = X\end{cases} \Rightarrow$ inverse transformation $\begin{cases} X = W \\ Y = Z/W = W^{-1}Z\end{cases}$

$J = \begin{bmatrix}\frac{\partial X}{\partial z} & \frac{\partial X}{\partial w}\\ \frac{\partial Y}{\partial z} & \frac{\partial Y}{\partial w}\end{bmatrix} = \begin{pmatrix} 0 & 1 \\ 1/w & -z/w^2\end{pmatrix} = -1/w \Rightarrow |J| = 1/w$

Then,

$f_{Z,W}(z,w) = f_X(w)\,f_Y(z/w)\cdot\frac{1}{w}$

$\Rightarrow \boxed{f_Z(z) = \int_{-\infty}^{\infty} f_X(w)f_Y(z/w)\,|\tfrac{1}{w}|\,dw}$

5.8 Let $X_1,\dots,X_n$ be a random sample, where $\bar{X}$ and S^2 are calculated in the usual way.

(a) Show that

$$S^2 = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-X_j)^2.$$

Assume now that the X_is have a finite fourth moment, and denote $\theta_1 = EX_i$, $\theta_j = E(X_i-\theta_1)^j$, $j=2,3,4$.

(b) Show that Var $S^2 = \frac{1}{n}(\theta_4 - \frac{n-3}{n-1}\theta_2^2)$.

(c) Find Cov$(\bar{X}, S^2)$ in terms of $\theta_1,\dots,\theta_4$. Under what conditions is Cov$(\bar{X}, S^2) = 0$?

a) $\frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-X_j)^2 = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\bar{X}+\bar{X}-X_j)^2$

where $(X_i-X_j) = (X_i-\bar{X}) - (X_j-\bar{X})$

$= \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[(X_i-\bar{X})^2 + 2(X_i-\bar{X})(X_j-\bar{X}) + (X_j-\bar{X})^2\right]$

$= \frac{1}{2n(n-1)}\left[\sum_{i=1}^{n}n(X_i-\bar{X})^2 - 2\sum_{i=1}^{n}(X_i-\bar{X})\sum_{j=1}^{n}(X_j-\bar{X}) + n\sum_{j=1}^{n}(X_j-\bar{X})^2\right]$

since $\sum_{i=1}^{n}(X_i-\bar{X}) = \sum_j X_j - n\bar{X} = \sum_j X_j - \sum_j X_j = 0$

$= \frac{n}{2n(n-1)}\sum_{i=1}^{n}(X_i-\bar{X})^2 + \frac{n}{2n(n-1)}\sum_{j=1}^{n}(X_j-\bar{X})$

$= \frac{1}{2(n-1)}\cdot 2\sum_{i=1}^{n}(X_i-\bar{X})^2$

$= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = S^2$.

b) Assume now that the X_i's have a finite fourth moment, denoting

$\theta_1 = EX_i$, $\theta_j = E(X_i-\theta_1)^j$, $j=2,3,4$

we proceed via proof by induction

We can assume $\theta_1 = EX_i = 0$ w/o loss of generality b/c both the sample variance and central moments θ_2, θ_4 are translation invariant, i.e. $\begin{cases}\theta_2 = EX_i^2\\ \theta_4 = EX_i^4\end{cases}$

Base case: First show that $\overbrace{\text{var}(S^2) = \frac{1}{n}\left[\theta_4 - \frac{n-3}{n-1}\theta_2^2\right]}^{(*)}\Big|_{n=4} = \frac{1}{4}(\theta_4 - \frac{1}{3}\theta_2^2)$

From a, $S^2 = \frac{1}{2n(n-1)}\sum_i\sum_j(X_i-X_j)^2\Big|_{n=4} = \frac{1}{24}\sum_{i=1}^{4}\sum_{j=1}^{4}(X_i-X_j)^2$

we want to compute $\text{Var}(S^2) = E[(S^2)^2] - (E(S^2))^2$

Assuming $X_i \overset{iid}{\sim}$ with $EX_i=0$, compute $E(S^2)$,

expanding $S^2 = \begin{cases} i=j: (X_i-X_j)^2 = 0 \text{ (4 such terms)} \\ i\ne j: E[(X_i-X_j)^2] = E(X_i^2) + E(X_j^2) - 2EX_iX_j \\ \quad = \theta^2+\theta^2-0 \\ \quad = 2\theta^2 \text{ } (4^2-4 = 12 \text{ such terms})\end{cases}$

$\Rightarrow E(S^2) = \frac{1}{24}(12)(2\theta^2) = \theta^2$.

Now compute $E[(S^2)^2]$,

expanding $(S^2)^2 = \frac{1}{576}\sum_{i,j,k,l}(X_i-X_j)^2(X_k-X_l)^2 =$

i) $i=j$ or $k=l \to$ expression $=0$
$(4\times16 + 4\times16 - 4^2 = 112$ terms)

ii) $i\ne j$: $(X_i-X_j)^4 = X_i^4 - 4X_i^3X_j + 6X_i^2X_j^2 - 4X_iX_j^3 + X_j^4$
(24 terms) where $E(X_i)=0$ and ind. kills odd moments:
$E[(X_i-X_j)^4] = 2\theta_4 + 6\theta_2 = 2(\theta_4+3\theta_2^2)$

iii) some distinct indices of form: $(X_i-X_j)^2(X_j-X_k)^2$;
(96 terms) $(X_i-X_j)^2 = X_i^2+X_j^2-2X_iX_j$
$(X_j-X_k)^2 = X_j^2+X_k^2-2X_jX_k$
$\Rightarrow E[(X_i-X_j)^2(X_j-X_k)^2] = \theta_4 + 3\theta_2$

iv) all distinct indices: $(X_i-X_j)^2(X_k-X_l)^2$
(24 terms) $\Rightarrow E[(X_i-X_j)^2(X_k-X_l)^2] = E[(X_i-X_j)^2]\,E[(X_l-X_k)^2]$ via ind.
$= 2\theta_2^2\, 2\theta_2^2$
$= 4\theta_2^2$

So, $E[(S^2)^2] = \frac{1}{576}[24\cdot(ii) + 96\cdot(iii) + 24(iv)] = \frac{1}{4}\theta_4 + \frac{11}{12}\theta_2^2$

and $V(S^2) = E[(S^2)^2] - (E(S^2))^2$

$= \left(\frac{1}{4}\theta_4 + \frac{11}{12}\theta_2^2\right) - \theta_2^2$

$= \frac{1}{4}(\theta_4 - \frac{1}{3}\theta_2^2)$

Induction Step. Assume that (*) holds for n and establish it for n+1.

Let S_n^2 := variance based on n observations.

Then, $S_{n+1}^2 = \frac{1}{2n(n+1)}\Big[\underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-X_j)^2}_{A} + 2\underbrace{\sum_{k=1}^{n}(X_k-X_{n+1})^2}_{B}\Big]$

where B: total squared difference btwn the new obs. X_{n+1} and each of the n previous ones

$= \frac{1}{2n(n+1)}[A+2B]$

Then $\text{var}(S_{n+1}^2) = \left(\frac{1}{2n(n+1)}\right)^2\text{var}(A+2B)$

where $\text{var}(A+2B) = \text{var}(A) + 4\,\text{var}(B) + 4\,\text{cov}(A,B)$.

By the induction hypothesis,

$S_n^2 = \frac{1}{2n(n-1)}A \Rightarrow A = 2n(n-1)S_n^2$

$\Rightarrow \text{var}(A) = [2n(n-1)]^2\left[\text{var}(S_n^2) = \frac{1}{n}(\theta_4 - \frac{n-3}{n-1}\theta_2^2)\right]$

$= 4n(n-1)^2\left[\theta_4 - \frac{n-3}{n-1}\theta_2^2\right]$

Given $B = \sum_{k=1}^{n}(X_k-X_{n+1})$, assuming $X_k \overset{iid}{\sim}$ $k=1,\dots,n$ and X_{k+1} independent of other X_k,

$\text{var}(B) = \sum_{k=1}^{n}\text{var}((X_k-X_{n+1})^2) + \sum_{k\ne l}\text{cov}[(X_k-X_{n+1})^2, (X_l-X_{n+1})^2]$

$= n(n+1)\theta_4 - n(n-3)\theta_2^2$

Similarly, $\text{cov}(A,B) = E(AB) - E(A)E(B)$

via tedious calculation reduces to

$= 2n(n-1)[\theta_4 - \theta_2^2]$

Then, after simplification,

$\text{var}(S_{n+1}^2) = \frac{1}{n+1}\left[\theta_4 - \frac{n-2}{n}\theta_2^2\right]$.

establishing the induction and verifying the result

c) $\text{cov}(\bar{X}, S^2) = \text{cov}\left(\frac{1}{n}\sum_k X_k, \frac{1}{2n(n-1)}\sum_{i,j}(X_i-X_j)^2\right)$

$= E(\bar{X}\cdot S^2) - E\bar{X}\cdot ES^2$

where $E\bar{X}=0$ by assuming $\theta_1 = EX_i = 0$

$= E(\bar{X}\cdot S^2)$

$= \frac{1}{2n^2(n-1)}E\left[\left(\sum_{k=1}^{n}X_k\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-X_j)^2\right)\right]$

If $k\ne i$ and $k\ne j$, then X_k is independent of both X_i and X_j and since $E(X_k)=0$, we get

$E(X_k(X_i-X_j)^2) = E(X_k)\cdot E[(X_i-X_j)^2] = 0$

So only terms where $k=i$ or $k=j$ contribute non-zero expectation

- $\forall$ fixed pair (i,j), there are 2 k values which contribute ($k=i$ or $k=j$)
- there are $n(n-1)$ total values of (i,j) with $i\ne j$:

P. For $i,j\in\{1,\dots,n\}$,
total # pairs (i,j) w/o restriction $= n\times n = n^2$
There are n-such diagonal pairs where $i=j$
e.g. (1,1), (2,2), etc.
$\Rightarrow$ # pairs where $i\ne j = n^2-n = n(1-n)$.

$\Rightarrow$ # nonzero expectation terms: $2n(n-1)$.

Thus, $\text{cov}(\bar{X}, S^2) = \frac{2n(n-1)}{2n^2(n-1)}EX_i(X_i-X_j)^2$; computing a representative term:

$E[X_i(X_i-X_j)^2] = E[X_i(X_i^2 - 2X_iX_j + X_j^2)]$

where $\begin{cases} E(X_i^3) = \theta_3 \\ E(X_i^2X_j) = E(X_i^2)E(X_j) = \theta_2\cdot 0 = 0 \\ E(X_iX_j^2) = 0\end{cases}$

$= \theta^3$; the same goes for $E[X_j(X_i-X_j)^2]$ since $X_i\sim X_j$

So, $\text{cov}(\bar{X}, S^2) = \frac{2n(n-1)}{2n^2(n-1)}\cdot\theta^3 = \frac{\theta^3}{n}$.

$\Rightarrow \text{cov}(\bar{X}, S^2) = 0$ i.f.f. $\theta_3 = 0$, i.e. the distribution of X_i is symmetric abt the mean.

(recall: $\theta_3 = E(X_i^3)$, i.e. the third central moment and the numerator of skewness)

Use the recursive relation, which can be verified:

$(k-1)S_k^2 = (k-2)S_{k-1}^2 + \left(\frac{k-1}{k}\right)(X_k - \bar{X}_{k-1})^2$

Now proceed:

$X_k - \bar{X}_{k-1} \sim N\left(0, \frac{k}{k-1}\right)$ for $k \geq 2$

$\Rightarrow \sqrt{\frac{k-1}{k}}(X_k - \bar{X}_{k-1}) \sim N(0,1)$

$\Rightarrow \frac{k-1}{k}(X_k - \bar{X}_{k-1})^2 \sim \chi_1^2$

can check that $(X_k - \bar{X}_{k-1}) \perp\!\!\!\perp S_{k-1}$ (by checking joint pdf)

By inductive hypothesis, $(k-2)S_{k-1}^2 \sim \chi_{k-1}^2$

Now $\begin{cases} \chi_{k-1}^2 \equiv \text{Gamma}\left(\frac{k-1}{2}, 2\right) \\ \chi_1^2 \equiv \text{Gamma}\left(\frac{1}{2}, 2\right) \end{cases}$

and sum of two independent Gamma RV's:

$\left(\frac{k-1}{2}, 2\right)$ and $\left(\frac{1}{2}, 2\right)$ is another $\underbrace{\text{Gamma}\left(\frac{k}{2}, 2\right)}_{\equiv \chi_k^2}$ RV (via mgf argument)

Hence $(k-1)S_k^2 \sim \chi_k^2$

5.3: Convergence concepts

<u>Ex (convergence in distribution)</u>: Poisson approximation of Binomial Dist.:

If $X_n \sim \text{Binomial}(n, p_n)$ and $|np_n \to \infty$ as $n \to \infty$

then $X_n \xrightarrow{d} Y$ where $Y \sim \text{Poisson}(\lambda)$.

But if $X_n \sim \text{Binomial}(n, p_n)$ and $|p_n \to p$ as $n \to \infty$

then $\frac{1}{\sqrt{n}}(X_n - np) \xrightarrow{d} Z$ where $Z \sim N(0, p(1-p))$

as a consequence of the CLT

<u>Proof.</u> Let $X_1, \dots, X_n \overset{iid}{\sim}$ with $\left| \begin{array}{l} EX_i = \mu \\ VX_i = \sigma^2 < \infty \end{array} \right.$

Define $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$. Show that $\bar{X}_n \xrightarrow{P} \mu$. (WLLN)

$\forall \varepsilon > 0$, $P(|\bar{X}_n - \mu| \geq \varepsilon)$

$\leq \frac{1}{\varepsilon^2} E|\bar{X}_n - \mu|^2$ by Chebyshev's inequality

$= \frac{1}{\varepsilon^2} \text{Var}\bar{X}_n$

$= \frac{1}{\varepsilon^2} \frac{1}{n}\sigma^2 \to 0$ as $n \to \infty$. ∎

<u>Proof.</u> Show that $X_n \xrightarrow{a.s.} X := P(\lim_{n\to\infty} |X_n - X| = 0) = 1$

$= P(\lim_{n\to\infty} X_n = X) = 1$

$\Leftrightarrow \forall \varepsilon > 0, P(\lim_{n\to\infty} |X_n - X| < \varepsilon) = 1$

Then, letting $\{\varepsilon = \frac{1}{2k}\}_{k \geq 1}$ — a sequence that converges to 0

$P(\lim_{n\to\infty} |X_n - X| < \frac{1}{2k}, \forall k \in \mathbb{N})$

$= 1 - P(\exists k \in \mathbb{N}, \lim_{n\to\infty} |X_n - X| \geq \frac{1}{2k})$

$= 1 - \sum_{k=0}^{\infty} P(\underbrace{\lim_{n\to\infty} |X_n - X| \geq \tfrac{1}{2k}}_{0})$

$= 1$

$\Rightarrow (*)$ is valid

<u>Proof.</u> Show that if $X_n \xrightarrow{a.s.} X$ then $X_n \xrightarrow{P} X$:

If $X_n \xrightarrow{a.s.} X$ then

$\forall \varepsilon > 0, 1 = P(\lim_{n\to\infty} |X_n - X| < \varepsilon)$

$= E\,\mathbb{1}(\lim_{n\to\infty} |X_n - X| < \varepsilon)$

$\leq \lim_{n\to\infty} P(|X_n - X| < \varepsilon)$, Fatou's lemma

Hence $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$. ∎

<u>Ex</u>: useful to use the original def. of RV as a function on $S: s \Rightarrow X(s)$

Let $S = [0,1]$, P uniform dist. on S.

Define $\begin{cases} X_n(s) = s + s^n \\ X(s) = s \end{cases}$

Then $X_n(s) \to X(s)$ $\forall s \in [0,1)$

so $P(s: \lim_{n\to\infty} X_n(s) = X(s)) = P([0,1)) = 1$

even though $P(X_n \neq X) = P(s \neq 0) = 1$

<u>Ex</u>: Let $X(s) = s$ $\forall s \in [0,1]$

Let $X_1(s) = s + \mathbb{1}(s \in [0,1])$

$X_2(s) = s + \mathbb{1}(s \in [0, 1/2])$

$X_3(s) = s + \mathbb{1}(s \in [1/2, 1])$

$X_4(s) = s + \mathbb{1}(s \in [0, 1/4])$

$X_5(s) = s + \mathbb{1}(s \in [1/4, 2/4])$

$X_6(s) = s + \mathbb{1}(s \in [2/4, 3/4])$

$X_7(s) = s + \mathbb{1}(s \in [3/4, 1])$

⋮ etc. s.t. $X_n(s) = s + \mathbb{1}_{[0, 1/n]}(s)$

Then $\forall \varepsilon > 0$, $P(s: |X_n(s) - X(s)| > \varepsilon)$

$= P(\mathbb{1}_{[0,\frac{1}{n}]} = 1) = \frac{1}{n} \to \infty$ as $n \to \infty$

Hence $X_n \xrightarrow{P} X$.

But $P(s: X_n(s) \to X(s)) = 0$ and thus $X_n \not\xrightarrow{a.s.} X$

<u>Ex</u>: Let $X \sim N(0,1)$ and $Y = -X$

Then $X \xrightarrow{d} Y$

but $P(X \neq Y) = 1$ and so $X \not\xrightarrow{a.s.} Y$.

5.10 Let $X_1, \dots, X_n$ be a random sample from a $n(\mu, \sigma^2)$ population.

(a) Find expressions for $\theta_1, \dots, \theta_4$, as defined in Exercise 5.8, in terms of μ and σ^2.

(b) Use the results of Exercise 5.8, together with the results of part (a), to calculate Var S^2.

(c) Calculate Var S^2 a completely different (and easier) way: Use the fact that $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$.

Let $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

a) $\theta_1 = E(X_i) = \mu$

$\theta_2 = \text{Var}(X_i) = \sigma^2$

$\theta_3 = E(X_i - \mu)^3 = E(X_i - \mu)^2(X_i - \mu)$

recall (Stein's lemma): $Eg(X)(X - \theta) = \sigma^2 Eg'(X)$

$= 2\sigma^2 \underbrace{E(X_i - \mu)}_{E(X_i) - E(\mu) = 0}$

$= 0$, a direct result of the symmetry of the normal dist.

$\theta_4 = E(X_i - \mu)^4 = E(X_i - \mu)^3(X_i - \mu) = 3\sigma^2 E(X_i - \mu)^2 = 3\sigma^4$

b) $\text{Var}S^2 = \frac{1}{n}\left(\theta_4 - \frac{n-3}{n-1}\theta_2^2\right)$

$= \frac{1}{n}\left(3\sigma^4 - \frac{n-3}{n-1}\sigma^4\right)$

$= \frac{\sigma^4}{n}\left(3 - \frac{n-3}{n-1}\right)$

$= \frac{\sigma^4}{n}\left(\frac{3(n-1) - (n-3)}{n-1}\right)$

$= \frac{\sigma^4}{n}\left(\frac{2n}{n-1}\right)$

$= \frac{2\sigma^4}{n-1}$

c) Use the fact that $(n-1)\frac{S^2}{\sigma^2} \sim \chi_{n-1}^2$

and $\overbrace{\text{Var}(X \sim \chi_k^2) = 2k}^{\text{FS}} \Rightarrow \text{Var}\chi_{n-1}^2 = 2(n-1)$

so, $\text{Var}\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1) \Rightarrow \left(\frac{(n-1)^2}{\sigma^4}\right)\text{Var}S^2 = 2(n-1)$

and hence

$\text{Var}S^2 = \frac{2(n-1)}{(n-1)^2/\sigma^4} = \frac{2\sigma^4}{n-1}$. ∎

2. Construct a sequence of non-identically distributed random variables $\{X_n\}_{n=1}^{\infty}$ and a random variable Y in the following settings. (Specify both X_n and Y as functions of a sample space S of your choice, and briefly justify your answer).

(i) X_n converges to Y almost surely, where $Y \sim$ Uniform$(0,1)$.

(ii) X_n converges to Y in probability, where $Y \sim$ Uniform(0,1), but X_n does not converge to Y almost surely.

(iii) For each n, X_n is a binary random variable that takes two possible values (0 or 1), and so is Y, and that X_n converges to Y in distribution, but not in probability.

i) Let $Y(\omega) = \omega \sim \text{Uniform}(0,1)$

$X_n(\omega) = Y(\omega) + \frac{1}{n} = \omega + \frac{1}{n} \sim \text{Uniform}(\frac{1}{n}, 1 + \frac{1}{n})$;

notice that X_n is not identically distributed

$\forall \omega \in [0,1]$,

$X_n(\omega) = \omega + \frac{1}{n} \to \omega = Y(\omega)$ as $n \to \infty$

$\Rightarrow X_n \xrightarrow{a.s.} Y$. ∎

ii) Let $Y \sim \text{Uniform}(0,1)$ and define $X_n := \begin{cases} Y, & \text{with prob. } 1 - \frac{1}{n} \\ Y + 1, & \text{with prob. } \frac{1}{n} \end{cases}$

Each X_n is non-identically distributed since the mixture prob.'s differ

To show convergence in probability, we want to compute

$P(|X_n - Y| > \varepsilon)$.

In this construction $\begin{cases} \text{w/ prob. } 1 - \frac{1}{n}, X_n = Y \Rightarrow |X_n - Y| = 0 \\ \text{w/ prob. } \frac{1}{n}, X_n = Y + 1 \Rightarrow |X_n - Y| = 1 \end{cases}$

So the event $|X_n - Y| \geq \varepsilon$ only occurs when $X_n = Y + 1$ w/ prob. $1/n$.

Hence: $P(|X_n - Y| \geq \varepsilon) = \begin{cases} \frac{1}{n} & \text{if } \varepsilon \leq 1 \\ 0 & \text{if } \varepsilon > 1 \end{cases}$

Thus, $\forall \varepsilon > 0$, $P(|X_n - Y| \geq \varepsilon) \leq \frac{1}{n} \to 0$ as $n \to \infty$

$\Rightarrow X_n \xrightarrow{P} Y$. ∎

Define the event where X_n differs from Y, i.e.

$A_n = \{\omega: X_n(\omega) \neq Y(\omega)\}$ s.t. $P(A_n) = \frac{1}{n}$

We examine whether $X_n(\omega) \neq Y(\omega)$ happens infinitely often (i.o.):

$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$, i.e. the harmonic series diverges

Since events A_n are independent,

citing (second) Borel-Cantelli lemma: $\left| \begin{array}{l} \text{If } \sum P(A_n) = \infty \text{ and } A_n \text{ independent} \\ \text{then } P(A_n \text{ i.o.}) = 1 \end{array} \right.$

Thus, $P(X_n \neq Y \text{ i.o.}) = 1 \Rightarrow P(\lim_{n\to\infty} |X_n - Y| = 0)$

$= P(\lim_{n\to\infty} X_n = Y) = 0$

$\Rightarrow X_n \not\xrightarrow{a.s.} Y$. ∎

iii) Define $X_n \sim \text{Bernoulli}(p_n)$ where $p_n = \frac{1}{2} + \frac{(-1)^n}{2n}$. Let $Y \sim \text{Bernoulli}(1/2)$.

Show: $X_n \xrightarrow{d} Y \sim \text{Bernoulli}(1/2)$:

$\forall x < 0, F_{X_n}(x) = 0 = F_Y(x)$

For $x \in [0,1)$:

$F_{X_n}(x) = P(X_n = 0) = 1 - p_n = \frac{1}{2} - \frac{(-1)^n}{2n} \to \frac{1}{2} = F_Y(x)$.

For $x \geq 1$: $F_{X_n}(x) = 1 = F_Y(x)$. ∎

Assuming X_n is independent of Y, then

$P(X_n \neq Y) = P(X_n = 0, Y = 1) + P(X_n = 1, Y = 0)$

$= P(X_n = 0)P(Y = 1) + P(X_n = 1)P(Y = 0)$

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

$= \left(1 - \left(\frac{1}{2} + \frac{(-1)^n}{2n}\right)\right) \cdot \frac{1}{2} + \left(\frac{1}{2} + \frac{(-1)^n}{2n}\right) \cdot \frac{1}{2}$

$= \frac{1}{4} - \frac{(-1)^n}{4n} + \frac{1}{4} + \frac{(-1)^n}{4n} = \frac{1}{2}$

So,

$P(|X_n - Y| \geq \varepsilon = 1/2) = P(X_n \neq Y) = \frac{1}{2} \not\to 0$ as $n \to \infty$

$\Rightarrow X_n \overset{?}{\not\to} Y$. ∎

3. Let $X_1, \dots, X_n, \dots$ be a sequence of independent random variables with a shared mean μ. X_i has variance σ_i for all i. State a sufficient condition according to which you can show that the weak law of large numbers remains true, i.e., $\bar{X}_n := (1/n)(X_1 + \dots + X_n)$ tends to μ in probability.

<u>WLLN</u> := Let $X_1, \dots, X_n \overset{iid}{\sim}$ RV's with $\left| \begin{array}{l} EX_i = \mu \\ \text{Var}X_i = \sigma^2 \end{array} \right.$

then $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{P} \mu$.

A sufficient condition to show WLLN holds true is a finite variance: $\sigma^2 < \infty$.

Applying Chebyshev's inequality:

$\forall \varepsilon > 0, P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2/n}{\varepsilon^2} \to 0$ as $n \to \infty$

$\Rightarrow \bar{X}_n \xrightarrow{P} \mu$. ∎

4. Let $X_1, X_2, \dots$ be a sequence of random variables that converges in probability to a constant a. Assume that $P(X_i > 0) = 1$ for all i. Show that the sequences defined by $Y_i = \sqrt{X_i}$ and $Y_i' = a/X_i$ converge in probability.

Let $X_1, X_2, \dots \xrightarrow{P} a \Leftrightarrow X_i \xrightarrow{P} a$.

Assume that $P(X_i > 0) = 1$ for all i.

In subsequent derivations, I will use the

<u>Continuous Mapping Thm</u> (CMT) := If $X_i \xrightarrow{P} a$ and g is continuous at a then $g(X_i) \xrightarrow{P} g(a)$

and the <u>lemma</u> := If $\left| \begin{array}{l} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} b: \text{constant} \end{array} \right.$ then $X_n + Y_n \xrightarrow{P} aX + b$

Given $X_i \xrightarrow{P} a$,

- $g(x) = \sqrt{x}$ is a continuous function on $(0, \infty)$
- Since $P(X_i > 0) = 1$, $\sqrt{x}$ is well-defined and continuous at a

Thus, by the CMT,

$Y_i = \sqrt{X_i} \xrightarrow{P} \sqrt{a}$

Next, since $X_i \xrightarrow{P} a > 0$,

$Y_i' = \frac{a}{X_i} \xrightarrow{P} \frac{a}{g(X_i)} = \frac{a}{a} = 1$. ∎

5. Let $\{X_n\}$ be sequence of random variables that converges in distribution to a random variable X. Let $\{Y_n\}$ be a sequence of random variables such that for any finite c, $\lim_{n\to\infty} P(Y_n > c) = 1$. Show that for any finite c,

$$\lim_{n\to\infty} P(X_n + Y_n > c) = 1.$$

Proof (Sketch-CLT). $M_Z(t) = Ee^{tZ} = e^{\frac{1}{2}t^2}$; it is enough to

Show that $M_{\frac{1}{\sqrt{n}}(Y_1+\cdots+Y_n)}(t) \longrightarrow M_Z(t)$ as $n \to \infty$

$\forall\ |t| < \delta$ for some $\delta > 0$

Now

$$M_{\frac{1}{\sqrt{n}}(Y_1+\cdots+Y_n)}(t) = \left(M_Y(t/\sqrt{n})\right)^n$$

Apply Taylor Expansion, which is valid for small $|t/(\sqrt{n}\sigma)|$:

$$M_Y(t/\sqrt{n}) = \underbrace{M_Y(0)}_{1} + \underbrace{M_Y^{(1)}(0)}_{EY=0}\frac{t}{\sqrt{n}} + \frac{1}{2}\underbrace{M_Y^{(2)}(0)}_{EY^2=1}\frac{t^2}{n} + o\left(\frac{t^2}{n}\right)$$

where $o\left(\frac{t^2}{n}\right)$ vanishes faster than $\frac{t^2}{n}$ as $n \to \infty$.

$$= \left(1 + \frac{1}{2}\frac{t^2}{n} + o\left(\frac{t^2}{n}\right)\right)^n$$

$$\to \lim_{n\to\infty}\left(1 + \frac{1}{2}\frac{t^2}{n}\right)^{\frac{2n}{t^2}\cdot\frac{t^2}{2}} = e^{t^2/2}. \ \square$$

Let $X_n \xrightarrow{d} X$. Let Y_n: sequence of RV's s.t. $\forall$ finite c, $\lim_{n\to\infty} P(Y_n > c) = 1$.

Show that for any finite c, $\lim_{n\to\infty} P(X_n + Y_n > c) = 1$:

Fact: = If one sequence Y_n diverges to $+\infty$ in prob., then adding any tight or convergent sequence cannot stop the sum from diverging.

Formally,

$Y_n \xrightarrow{P} +\infty$ means

$\forall\ M \in \mathbb{R}$, $P(Y_n > M) \to 1$.

$X_n \xrightarrow{d} X$ implies tightness of $\{X_n\}$, i.e.

$\forall\ \varepsilon > 0\ \exists\ M$ s.t.

$\sup_n P(|X_n| > M) < \varepsilon$, recall: an upper bound b of S is a supremum (of S) $=: \sup b$ if for $z \in$ partially ordered set $(P, \leq)$,

$z \geq x$ for all $x \in S$.

In general, all sequences that converge in distribution are tight.

a lower bound a of S is called an infimum (of S):

Inf a := for $y \in (P, \leq)$,

$y \leq x$ for all $x \in S$.

Proof. want to show

$$P(X_n + Y_n > c) \to 1 \equiv \lim_{n\to\infty} P(X_n + Y_n > c) = 1$$

consider the event decomposition, letting $M \in \mathbb{R}$:

$$\{X_n + Y_n > c\} \supseteq \{Y_n > c + M\} \cap \{|X_n| \leq M\}$$

citing $P(A \cap B) = P(A) - P(A \cap B^c) \geq P(A) - P(B^c)$

$$\geq P(Y_n > c + M) - P(|X_n| > M)$$

controlling $P(|X_n| > M)$ via tightness:

since $X_n \xrightarrow{d} X$, $\forall\ \varepsilon > 0$, $\exists\ M$ s.t.

$\sup_n P(|X_n| > M) < \varepsilon$.

Fix this M.

Thus, the 'bad' tail event of X_n has prob. at most ε.

By assumption,

$P(Y_n > c + M) \to 1$ as $n \to \infty$.

For sufficiently large n,

$P(X_n + Y_n > c) \geq P(Y_n > c + M) - \varepsilon$

$\Rightarrow \liminf_{n\to\infty} P(X_n + Y_n > c) \geq 1 - \varepsilon$

where $\varepsilon > 0$ was arbitrary

$= 1$

Thus, $\lim_{n\to\infty} P(X_n + Y_n > c) = 1$. $\blacksquare$