Lecture 1

1.1 Statistical Methods (ST551, ST552, ST552)

Descriptive Methods — Summary Tables, Histograms, Bar/Pie Chart

Statistical Inference Methods — Estimation, Confidence Interval, Hypothese Testing

↑

Why do these methods work? { ~~Why are some better?~~ ~~when a method does not work.~~

↑

Statistical Theories (ST 562, ST 563)

↑

Based on theory of Probability and random variables (r.v.) (ST 561)

What is probability? From gambling?

Gambling shows that there has been an interest in quantifying the ideas of probability for a long time, but exact mathematical descriptions arose much later

About 400 years or longer

Two schools of thoughts:

1. Frequentists: probability = relative frequency of occurrence
2. Subjectivists: probability = degree of belief

Actually, 3. Bayesians: include expert knowledge as a prior.

1, 2 → Similar axiomatic /ˌæksiəˈmætik/ foundation

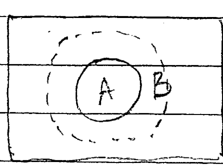
1.2 Set: a collection of elements (objects) with common properties

"S" : universal set ↔ all elements under consideration

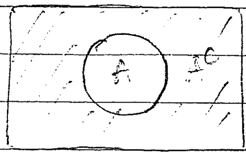
Example: $S = \{\text{integer}\} = \{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}$

$= \{x : x \text{ is an integer}\}$

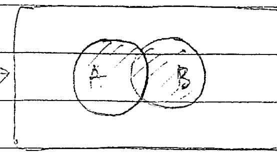
Concepts: subset $A \subseteq S$, $A \subseteq B$



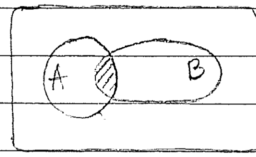
Venn Diagram



Complement
Union
Intersection



"either"



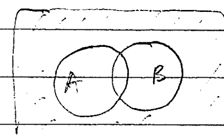
"both"

Set $A_1, A_2, \cdots, A_n, \cdots$

$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$ "at least one set"

$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$ " all sets"

DeMorgan's Law: $\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$



$\bigcup_{i=1}^{\infty} A_i, \quad \bigcap_{i=1}^{\infty} A_i$

Examples: (1) $\bigcup_{i=1}^{\infty} [1/i, 1] = \bigcup_{i=1}^{\infty} (1/i, 1] = (0, 1]$

(2) $\bigcap_{i=1}^{\infty} [0, 1/i] = \bigcap_{i=1}^{\infty} [0, 1/i) = \{0\}$

(3) $\bigcap_{i=1}^{\infty} (0, 1/i) = \phi$ (empty set)

Disjoint: A & B are disjoint (mutually exclusive)

$\Leftrightarrow$ no common elements between A and B

$\Leftrightarrow$ $A \cap B = \phi$

$\{A_i\}_{i=1}^{n}$ or $\{A_i\}_{i=1}^{\infty}$ pairwise disjoint $\Leftrightarrow$ $A_i \cap A_j = \phi$ for all $i \neq j$.

more restrictive than $\bigcap_{i=1}^{n} A_i = \phi$ or $\bigcap_{i=1}^{\infty} A_i = \phi$

Partition: $\{A_i\}_{i=1}^{n}$ or $\{A_i\}_{i=1}^{\infty}$ is a partition of S

$\Leftrightarrow$ $\{A_i\}_{i=1}^{n}$ or $\{A_i\}_{i=1}^{\infty}$ are pairwise disjoint and $\bigcup_i A_i = S$

Example: $\{(\frac{1}{1+i}, \frac{1}{i}]\}_{i=1}^{\infty}$ is a partition of $(0, 1]$

Lecture 2

1.3 Probability

S = a set of all possible outcomes of a random experiment
• "Sample Space"
E.g., Toss a coin once: $S = \{H, T\}$
Roll a die once: $S = \{1, 2, 3, 4, 5, 6\}$

$A \subseteq S$ an "appropriate" subset of the sample space S
• "Event" E.g., $A = \{2, 4, 6\}$

If the outcome of a random experiment is in A, we say "event A occurred".

Assume some random experiment can be performed over and over again under identical conditions (replicability)
A: an event, what is $P(A)$? probability of an event A

N trials: N_A = # of times A occurs in these N trials
N_A/N = relative frequency of A
As $N \to \infty$, expect $N_A/N \to$ some p $(0 \le p \le 1) \Leftrightarrow P(A)$

Example: $A = \{2, 4, 6\}$, $N_A/N \to 1/2$.

From frequentist's viewpoint: $P(A) = \lim_{N\to\infty} N_A/N$.
Therefore, Definition 1.3.4
(i) $N_A/N \ge 0 \Rightarrow P(A) \ge 0$
(ii) $N_S/N = 1 \Rightarrow P(S) = 1$
(iii) A and B disjoint: $N_{A\cup B}/N = N_A/N + N_B/N$.
$\Rightarrow P(A\cup B) = P(A) + P(B)$ if disjoint
More generally, $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ if $\{A_i\}_{i=1}^{\infty}$ pairwise disjoint
"Countable Additivity"

Three axioms of probability.

Note: If S is finite or countably infinite, the A can be any subset of S; however, if S is uncountably infinite, e.g., $S = [0, 1]$, we can not assign probility to every subset of S such that 1.3.4 still holds. Instead, we restrict to Borel sets (Definition 1.3.2)

Thm 1.3.1 Event A, B; empty set ϕ;

(i) $p(\phi) = 0$

(ii) $p(A^c) = 1 - p(A)$

proof. $p(A \cup A^c) = p(S) = 1$;

$p(A \cup A^c) = p(A) + p(A^c)$. ▫

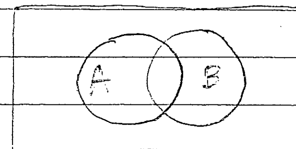
(iii) $p(B \cap A^c) = p(B) - p(B \cap A)$

(iv) If $A \subseteq B$, then $p(A) \le p(B)$

(v) $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

proof: $p(A \cup B) = p(A \cap B^c) + p(B \cap A^c) + p(A \cap B)$

as $(A \cup B) = (A \cap B^c) \cup (B \cap A^c) \cup (A \cap B)$ and these three sets pairwise disjoint



So, $p(A \cup B) = p(A) - p(A \cap B) + p(B) - p(B \cap A) + p(A \cap B)$

$= p(A) + p(B) - p(A \cap B)$

(vi) $p(A) = \sum_{i=1}^{\infty} p(A \cap B_i)$ if $\{B_i : i \ge 1\}$ is a partition of S.

(vii) $p(A \cap B) \ge p(A) + p(B) - 1$

$\boxed{\text{Prove by yourself (vii) from (v)}}$

Application of (vii).

$[l_1, u_1]$ is a 95% confidence interval for θ_1

$[l_2, u_2]$ 95% C.I. θ_2

$\Leftrightarrow$ $p(l_1 \le \theta_1 \le u_1) = 95\% = 0.95$

$p(l_2 \le \theta_2 \le u_2) = 95\% = 0.95$

then $p(l_1 \le \theta_1 \le u_1, l_2 \le \theta_2 \le u_2) \ge 0.95 + 0.95 - 1 = 0.9$

so we get a 90% simultaneous C.I. for (θ_1, θ_2)

2 Children: $P(\text{at least a boy}) = 3/4$

Lecture 3

$P(\text{younger is a boy}) = 1/2$

Finite sample spaces with equally likely outcomes:
relative simple but important.

$S = \{s_1, s_2, \cdots, s_N\}$ and $p(\{s_i\}) = 1/N$ for $i = 1, 2, \cdots, N$.
Event $A \subseteq S$: $P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} (1/N) = \#\text{elements in } A / N$

Calculation of $P(A) \Leftrightarrow$ Counting number of elements in A
Sounds simple, but can be complicated in some cases!

Let's have some fun:

Example 1. Draw 5 cards at random without replacement from a 52-card deck, what is the probability of drawing "3 of a kind"?
"3 of a kind": 3 cards of same rank + 2 unmatched cards
e.g. $\{5, 5, 5, Q, 10\}$

Counting Rules:

Q1. With replacement or without replacement?
Example: Draw 2 cards from a 52-card deck, $P(\text{both hearts})$?
W. replacement: $\#(\text{both hearts}) = 13 \times 13$
W.o. replacement: $\#(\text{both hearts}) = 13 \times 12$
☆ Important question: what is sample space? $\{(\text{card } 1, \text{card } 2)\}$
W. replacement: $\#S = |S| = 52 \times 52$
W.o. replacement: $\#S = 52 \times 51$
sample space changes
So,

$$P(\text{both hearts}) = \begin{cases} (13\times13)/(52\times52) = (1/16) & \text{w. replacement} \\ (13\times12)/(52\times51) < (1/16) & \text{w.o. replacement} \end{cases}$$

Q2. Does order matter?
Example: Roll a dice twice $|S| = 6^2 = 36$
$p(1 \text{ on first}, 2 \text{ on second}) = 1/36$
$p(\text{one } 1 \text{ and one } 2) = 2/36$

Rule 1: Combination — unordered sampling without replacement
# of ways to choose r elements from a set of n elements

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Rule 2: Permutation — ordered sampling without replacement

\# of ways to sample an r-sequence from a set of n elements

$${}^{n}P_{r} = \frac{n!}{(n-r)!} = n \times (n-1) \times \cdots \times (n-r+1)$$

Note: ${}^{n}P_{r} = \binom{n}{r} \times r!$ Why is that? Order $\leftrightarrow r!$

Example: 5 equally qualified candidates to fill 2 positions (president, vice president)

$n=5, r=2$

A, B, C, D, E $\rightarrow$ P1, P2

${}^{n}P_{r} = {}^{5}P_{2} = 5 \times 4 = 20$

① without replacement

② ordered

$n = ?$

$p = ?$

P1	P2
A	B
A	C
A	D
A	E
B	A
...	

What if 2 positions are the same?

A B = B A

$$\binom{5}{2} = \frac{5 \times 4}{2 \times 1} = 10$$

Example: How many 7-digit telephone numbers are possible?

_ _ _ _ _ _ _

0 0
⋮ ⋮
9 9

$10 \times 10 \times \cdots \times 10 = 10^7$

(ordered with replacement)

What if there is no repeated digit? $10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 = {}^{10}P_{7}$

Go back to the card drawing problem:

(a) $S = \{(\text{card}1, \text{card}2, \cdots, \text{card}5)\}$ (unordered + without replacement)

$\# S = \binom{52}{5} = 52 \times 51 \times 50 \times 49 \times 48 / (5 \times 4 \times 3 \times 2 \times 1) =$

(b) A = "3 of a kind".

_ _ _ _ _

b1 # for the kind $= \binom{13}{1} = 13$ (2, 3, 4, ⋯, 10, J, Q, K, A)

b2 # for 3 cards of that kind $= \binom{4}{3} = 4$ (spades, hearts, diamonds, clubs)

b3 # for the other 2 kinds $= \binom{12}{2} = 66$

b4 # for the one other kind = 4

b5 # for the other other kind = 4

Example "3 of a kind"

b1: A

b2: (♢, ♡, ♧)

b3: (J, 5)

b4: (J − ♢)

b5: (5 − ♡)

$\#A = 13 \times 4 \times 66 \times 4 \times 4 = 54{,}912$

(c) $P(A) = \#A / \#S = 0.02113 \approx 2\%$

Example 2: Multiple choice quiz:

10 questions, 4 possible answers each.

Guess on all 10 questions, P(at least 5 correct)?

$S = \{$ all sequences of answers for the 10 questions $\}$

$A = \{$ the ------ with 5 or more correct $\}$

$\#S = 4 \times 4 \times \cdots \times 4 = 4^{10}$

$\#A = ?$ $A = A_5 \cup A_6 \cup A_7 \cup A_8 \cup A_9 \cup A_{10}$

A_k: the sequences with exactly k correct answers

they are pairwise disjoint

$\#A = \#A_5 + \#A_6 + \cdots + \#A_{10}$

$\#A_5 = \binom{10}{5} \times 1^5 \times 3^5$

which 5 correct? 5 correct 5 incorrect

$\#A_6 = \binom{10}{6} \times 1^6 \times 3^4$ ...... $\#A_k = \binom{10}{k} \times 3^{10-k}$

$\#A = \sum_{k=5}^{10} \binom{10}{k} 3^{10-k} = 81{,}922$

$P(A) = 81{,}922 / \binom{10}{5} = 0.0781$

Note: Binomial distribution

Example 3: Record the birthday of all students in ST561. Chance of at least two have the same birthday?

$S = \{(bd1, bd2, \ldots, bd30)\}$ 30 students

$\#S = 365 \times 365 \times \cdots \times 365 = 365^{30}$

$A = \{$ at least two have the same birthday $\}$

$= A_2 \cup A_3 \cup \cdots \cup \cdots A_{30}$ A_k: exactly k have the same birthday

Complicated!

$A^c = \{$ all 30 students have different birthdays $\}$

$\#A^c = 365 \times 364 \times \cdots \times (365-30+1) = {}^{365}P_{30}$

$P(A) = 1 - P(A^c) = 1 - \#A^c / \#S \approx 70.6\%$

Lecture 4 Conditional Probability & Independence

Conditional probability (Def. 1.4.1)

$A, B \subseteq S$

Cond. prob. of the event A given the other event B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{given } P(B) > 0$$

$\boxed{P(A^c|B) = 1 - P(A|B)}$

Example for intuition: toss a die $S = \{1,2,3,4,5,6\}$

A: ≤ 3

B: even $\quad A \cap B = \{2\}$

[prob. of A occurs given B occurs]

$P(A) = 3/6 = 0.5$

$P(A|B) = P(A \cap B)/P(B) = \frac{1}{6} / \frac{1}{2} = 1/3$

sample space becomes smaller

repeat the toss infinity times: $\boxed{2}$, 3, 6, 1, 5, 4, $\boxed{2}$, $\boxed{2}$, 6, 1, 5, 3, 5, ...

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N} \qquad P(A|B) = \lim_{N \to \infty} \frac{N_{A \cap B}}{N_B}$$

Insert Axioms of conditional prob. here!

Bayes's Theorem (Thm. 1.4.3)

If we know $P(A \cap B)$ & $P(B)$, then $P(A|B) = P(A \cap B)/P(B)$

$P(A|B)$ & $P(B)$, then $P(A \cap B) = P(A|B)P(B)$

$P(B|A)$ & $P(A)$, then $P(A \cap B) = P(B|A)P(A)$

$\Rightarrow P(A|B)P(B) = P(B|A)P(A)$

$$\Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

It is useful because sometimes $P(B|A)$ is much easier to compute than $P(A|B)$.

Furthermore, $P(B) = P(B \cap A) + P(B \cap A^c)$

$= P(B|A)P(A) + P(B|A^c)P(A^c)$

(Theory of total probability)

Example[1]: ELISA test for HIV antibody

If antibody exists, ELISA detects it with prob. 0.997

... does not exist, ... does not detect it with prob. 0.985

(Enzyme-linked immunosorbent assay)

/ˈenzaɪm/ /ˌimjunəuˈsɔːbənt/

$p(\text{test}+ \mid \text{antibody present}) \equiv$ sensitivity

$p(\text{test}- \mid \text{antibody absent}) \equiv$ specificity

Question of interest: assume 1% of population have AIDS antibody. Now randomly select a person from the whole population, ELISA gives a positive result. What's the probability that he/she really has the AIDS antibody?

A: antibody present (+) $\quad A^c$: −

B: ELISA positive result (test +) $\quad B^c$: test −

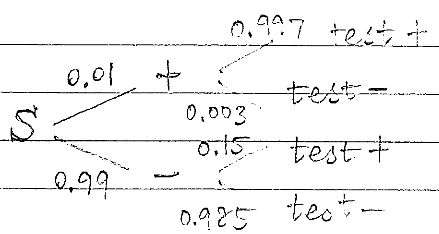
$p(\text{test}+ \mid +) = 0.997$, $p(\text{test}- \mid -) = 0.985$, $p(+) = 0.01$

What is $p(+ \mid \text{test}+)$?

Note: In this situation, $p(\text{test}+ \mid +)$ and $p(\text{test}- \mid -)$ are both high, indicating this is a reliable test, but people are still interested in the question $p(+ \mid \text{test}+)$ because they really want to know whether they have antibody but not just the test result.

$$p(+ \mid \text{test}+) = \frac{p(\text{test}+ \mid +)\,p(+)}{p(\text{test}+)} = \frac{p(\text{test}+ \mid +)\,p(+)}{p(\text{test}+ \mid +)\,p(+) + p(\text{test}+ \mid -)\,p(-)}$$

$$= \frac{0.997 \times 0.01}{0.997 \times 0.01 + (1-0.985) \times (1-0.01)} = \frac{0.997 \times 0.01}{0.997 \times 0.01 + 0.015 \times 0.99} \approx 0.4017 \; (!!)$$



Try to increase either (same amount)

$\begin{cases} p(\text{test}+ \mid +) \\ \text{or} \\ p(\text{test}- \mid -) \end{cases}$ yourselves

to see which affects $p(+ \mid \text{test}+)$ more strongly? Why?

Order: $+/- \rightarrow \text{test}+/-$

So $\text{test}+/- \rightarrow +/-$

~~more difficult.~~

Bayes's Rule:

$$p(A \mid B) = \frac{p(B \mid A)\,p(A)}{p(B \mid A)\,p(A) + p(B \mid A^c)\,p(A^c)}$$

More generally: $S = A_1 \cup A_2 \cup \cdots \cup A_k$ partition

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + \cdots + P(A_k)P(B|A_k)}$$

Example 2: Three bags of currency:

Bag 1: 4 − \$20, 1 − \$10, 2 − \$2, 3 − \$1 ⇒ \$97

2: 1 − \$20, 6 − \$10, 3 − \$2 ⇒ \$86

3: 1 − \$20, 1 − \$2, 8 − \$1 ⇒ \$30

(a) Three bags are identical in appearance. Would you pay \$75 for the opportunity to select one bag at random?

$$\frac{97+86+30}{3} - 75 = -4 \Rightarrow \text{No!}$$

(b) Suppose rules changed. Select a bag at random and take a bill out of that bag, before you decide whether to pay the \$75 for it. Suppose it is a \$20 bill, would you pay for it?

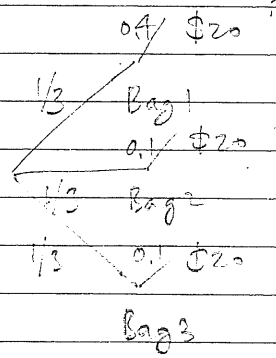
$$P(\text{Bag}1 \mid \$20) = \frac{P(\$20 \mid \text{Bag}1)\,P(\text{Bag}1)}{\sum_k P(\$20 \mid \text{Bag}\,k)\,P(\text{Bag}\,k)} = \frac{0.4 \times \frac{1}{3}}{(0.4+0.1+0.1)\times\frac{1}{3}} = \frac{2}{3}$$

$$P(\text{Bag}2 \mid \$20) = P(\text{Bag}3 \mid \$20) = 1/6$$

$$\frac{2}{3}\times 97 + \frac{1}{6}\times 86 + \frac{1}{6}\times 30 - 75 = 9 \Rightarrow \text{Yes!}$$

Note: (a) Totally random.

(b) With prior knowledge, more advantages.



Order: Bag → \$20

· We want \$20 → Bag

Lecture 5

For a fixed event B with $P(B)>0$, $P(A|B)$ conditional probability function satisfies that:

(i) $P(S|B)=1$

(ii) $P(A|B)\geq 0$

(iii) $P(\cup_{i=1}^{\infty} A_i|B)=\sum_{i=1}^{\infty} P(A_i|B)$ for $\{A_i\}_{i=1}^{\infty}$ pairwise disjoint.

Prove them yourselves as an exercise!

So, $P(\cdot|B)$ satisfies all the properties of a probability function.

Then Thm 1.3.1 holds for $P(\cdot|B)$:

$P(A^c|B)=1-P(A|B)$

$P(A_1\cup A_2|B)=P(A_1|B)+P(A_2|B)-P(A_1\cap A_2|B)$

Independence:

Events A and B are independent if $P(A|B)=P(A)$

"A does not depend on B" in the common sense.

A — raining in Corvallis, OR ~~independent unless "butterfly effect"!~~

B — traffic jam in New York

Events A and B: $P(A\cap B)=P(A)P(B)$ more general definition

Example 1: Roll a die twice

$P(\text{second } 1 \mid \text{first } 1)$

$$=\frac{P(\{1,1\})}{P(\text{first } 1)}=\frac{1/36}{1/6}=\frac{1}{6}=P(\text{second } 1)$$

For the case performing independent trials, they are independent.

Mutually independent:

Events $A_1, A_2, \cdots, A_k$ are mutually independent if

$P(\cap_{j=1}^{\ell} A_{i_j})=\prod_{j=1}^{\ell} P(A_{i_j})$

for any subcollection $\{A_{i_1}, A_{i_2}, \cdots, A_{i_\ell}\}$ of $\{A_1, \cdots, A_k\}$

Note, Mutually independent $\underset{\not\Leftarrow}{\Longrightarrow}$ pairwise independent.

Example 2: Toss a coin twice

$A = \{H \text{ on } 1st\}$

$B = \{T \text{ on } 2nd\}$

$C = \{1H, 1T\} = \{HT, TH\}$

$S = \{HH, HT, TH, TT\}$ $\qquad p(A) = p(B) = p(C) = \frac{1}{2}$

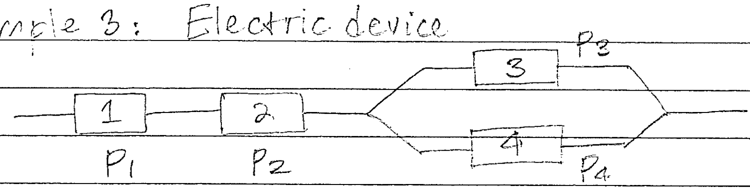
$p(AB) = p(\{HT\}) = 1/4$

$p(AC) = p(\{HT\}) = 1/4 \quad \Rightarrow$ pairwise independent

$p(BC) = p(\{HT\}) = 1/4$

But $p(A \cap B \cap C) = p(\{HT\}) = \frac{1}{4} \neq p(A)\,p(B)\,p(C) \Rightarrow$ NOT mutually indep.

Example 3: Electric device



1, 2, 3, 4 work independently (mutually independently)

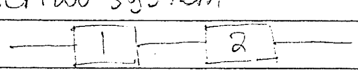
$p(\text{device work}) = p(1,2,3 \text{ work or } 1,2,4 \text{ work})$

$= p(1,2,3 \text{ work}) + p(1,2,4 \text{ work}) - p(1,2,3,4 \text{ work})$

$= p_1 p_2 p_3 + p_1 p_2 p_4 - p_1 p_2 p_3 p_4$

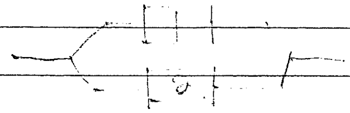
$= p_1 p_2 (p_3 + p_4 - p_3 p_4)$ $\qquad$ Reliability Analysis

Serial system



$p_1 p_2 = 0.9^2 = 0.81$

Parallel system



$p_1 + p_2 - p_1 p_2 = 0.9 + 0.9 - 0.81 = 0.99$

In general, parallel system is more reliable than serial system given the same components.

Example 4: Roll a die 4 times, $p(\text{at least one } 6) = ?$

$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$?

$1 - p(\text{no } 6) = 1 - \left(\frac{5}{6}\right)^4 \doteq 0.518$

$\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6}$?

Lecture 6 Random Variables (r.v.)

A random variable is a function from the sample space to the real number:

(i) $X: S \mapsto \mathbb{R}^1$

(ii) Assign a value to every sample points; some sample points may share the same value

(iii) The value is random, as the sample point is random.

Example for Intuition: Toss a coin twice, $X = \#$ of H's

S	X
HH	2
HT	1
TH	1
TT	0

Only 3 values $\{0, 1, 2\}$

A random variable X is called discrete if X takes only a finite or countably infinite number of distinct values

A continuous r.v. often takes values on an interval or intervals of $\mathbb{R}$.

Example for Intuition: Repeatedly toss a coin until 5 heads. $X = \#$ of tosses. $X = 5, 6, 7, 8, \cdots$ countably infinite

Example for Intuition: Randomly select a student. $X =$ height of the student. $X \in (0, \infty)$ or $[2 \text{ feet}, 8 \text{ feet}]$

Distribution of a discrete random variable:

The randomness is described by the probability mass function (pmf)

$$f_X(x) = P(X = x) = P(\{s : X(s) = x\})$$ an event as a subset of S.

for all $x \in C$ (C be the set of all possible values of X)

For any set A, $P(X \in A) = \sum_{x \in A} f_X(x) = \sum_{x \in A} P(X = x)$

From the Axioms of probability, the pmf must satisfy:

(a) $0 \le f_X(x) \le 1$, for all x
(b) $\sum_{x \in C} f_X(x) = 1$

Cumulative Distribution Function (c.d.f.) F_X or F

$$F_X(t) = P(X \le t) = P(\{s \in S : X(s) \le t\}),\ t \in \mathbb{R}$$

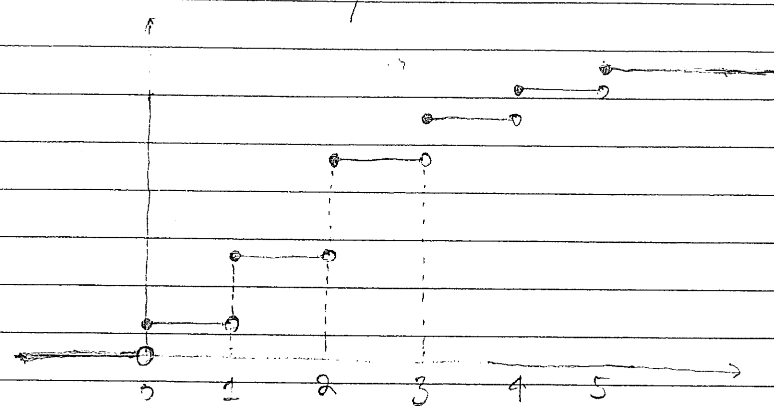
Note (1) CDF is defined for all real number $t \in \mathbb{R}$

(2) For discrete r.v.'s,

$$F(t) = F_X(t) = \sum_{x \le t} P(X = x) = \sum_{x \le t} f(x)$$

Example: X = # wage earners in a randomly selected household in Corvallis

x	$f_X(x)$	$F_X(x)$
0	0.07	0.07
1	0.28	0.35
2	0.37	0.72
3	0.18	0.90
4	0.02	0.98
5	0.02	1.00



Properties of CDF

(1) $F(x) \le F(y)$ for $x \le y$

(2) $\lim_{x \to -\infty} F(x) = 0$

$\lim_{x \to +\infty} F(x) = 1$

(3) $F(x)$ is right-cont. $\Leftrightarrow \lim_{h \to 0+} F(x+h) = F(x)$

(1) step function
(2) Non-decreasing
(3) Right-continuous

Relationship between c.d.f. and p.m.f. for discrete r.v.'s

$$\begin{cases} F(x-) = \lim_{h \to 0+} F(x-h) & \text{left limit} \\ F(x+) = \lim_{h \to 0+} F(x+h) & \text{right limit} \end{cases}$$

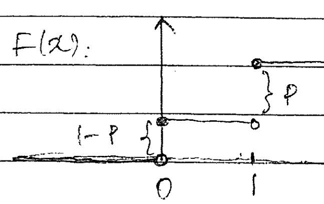
Then $f(x) = P(X = x) = F(x+) - F(x-) = F(x) - F(x-)$

Examples of discrete r.v.'s

(1) Bernoulli r.v. $S = \{\text{Success}, \text{Failure}\}$

$$\begin{cases} X(\text{success}) = 1 \\ X(\text{Failure}) = 0 \end{cases} \qquad f(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \end{cases}$$

$X \sim \text{Bernoulli}(p)$ $F(x)$:



$0 < p < 1$ is a parameter

(2) Poisson r.v. [# calls received at a call center in a day]

$C = \{0, 1, 2, 3, \cdots\}$

$f(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$ for $x \in C$ $\quad$ $\lambda > 0$ is a parameter

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

$$\left[e^t = 1 + \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots \right]$$

Taylor expansion: $f(x) = f(a) + \dfrac{f'(a)}{1!}(x-a) + \cdots + \dfrac{f^{(k)}(a)}{k!}(x-a)^k + \cdots$

Distribution of a continuous random variable:

Continuous r.v. X often takes values within intervals of $\mathbb{R}$.

$P(X = x) = 0$ for any $x \in \mathbb{R}$

X — waiting time at a bus stop

$P(X = 5\text{ min}) = 0$ $\quad$ interested in $P(4\text{ min} \le X \le 5\text{ min})$

probability density function (pdf): $f(x)$ satisfies

(1) $f(x) \ge 0$ for all $x \in \mathbb{R}$

(2) $\int_{\mathbb{R}} f(x)\,dx = 1$

Then $P(X \in A) = \int_{x \in A} f(x)\,dx$.

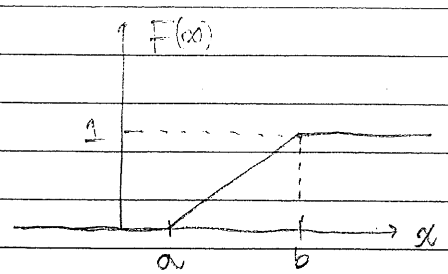
In particular, the cdf $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$

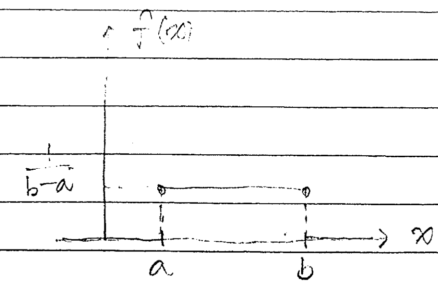
Note: for continuous r.v.'s, the cdf is absolutely continuous.

$\forall \varepsilon > 0, \exists \delta > 0$, for any $|x - y| \le \delta$, $|f(x) - f(y)| < \varepsilon$.

Example 1: $U(a, b)$ uniform distribution

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{o.w.} \end{cases} = \frac{1}{b-a} I(a < x < b)$$





Example 2 : $N(0,1)$ standard normal dist.

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

$$F(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt \qquad \text{No closed form}$$

Remark: (i) some r.v.'s are neither purely discrete or purely continuous.

e.g. zero-added exponential

randomly select a light bulb in a store, X = life time of a light bulb

$P(X=0) = p > 0$ $\qquad$ p = proportion of defective bulbs

$f(x) = e^{-x}/(1-p)$ on $(0, \infty)$

$\boxed{\text{Prove by yourselves } P(X \in \mathbb{R}) = 1}$

(ii) For two r.v.'s X & Y, if $F_X(t) = F_Y(t)$ for all t, say, $X \overset{d}{=} Y$.

This does not mean $X = Y$ (X, Y are identical r.v.'s)

$$X \sim \text{Bernoulli}\left(\tfrac{1}{2}\right) \qquad P(X=x) = \begin{cases} 1/2 & x=0 \\ 1/2 & x=1 \end{cases}$$

$$\text{Let } Y = 1 - X \neq X, \qquad P(Y=y) = \begin{cases} 1/2, & y=0 \\ 1/2, & y=1 \end{cases}$$

Summary: distribution of a discrete r.v. $\leftrightarrow$ pmf

continuous $\leftrightarrow$ pdf $\quad$ > ~~cdf~~

distribution of any r.v. $\leftrightarrow$ cdf $F(x) = P(X \le x)$, $x \in \mathbb{R}$

discrete $\quad f(x) = F(x) - F(x-)$

continuous $\quad f(x) = F'(x) = \dfrac{dF(x)}{dx}$

Lecture 7 Examples of r.v.'s

Discrete distributions: page 33, section 1.7.1

(I) Binomial: n independent Bernoulli trials with parameter p

$S = \{(w_1, w_2, \cdots, w_n) : w_i = S \text{ or } F,\ i = 1, 2, \cdots, n\}$

$X_i = \begin{cases} 1 & \text{if } w_i = S \\ 0 & \text{if } w_i = F \end{cases}$

$Y = \sum_{i=1}^{n} X_i$: # of successes $\sim$ Binomial(n, p)

pmf: $P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, 2, \cdots, n$

$\sqcup \sqcup \sqcup \sqcup \cdots\cdots \sqcup$

① $\binom{n}{y}$ choices of success positions
② p^y prob of success
③ $(1-p)^{n-y}$ failures

Question: is $f(y)$ a pmf $\begin{cases} (1) & f(y) \geq 0 \\ (2) & \boxed{\sum_{y=0}^{n} f(y) = 1} \, ? \end{cases}$

sometime $Y_i \sim$ some continuous r.v. to make it challenging

Example: Y_i — amount of emission on day i

an exceedance occurs if $Y_i > M$, suppose $P(Y_i > M) = p$ for all i.

$X_i = \begin{cases} 1 & \text{if } Y_i > M \\ 0 & Y_i \leq M \end{cases} \qquad X_i \sim \text{Bernoulli}(p)$

$X = \sum_{i=1}^{7} X_i$ = # days with exceedance in a week $\sim$ Binomial$(7, p)$

$P(\text{no exceedance}) = P(X = 0) = \binom{7}{0} p^0 (1-p)^7 = (1-p)^7$

$P(X \leq 2) = \sum_{k=0}^{2} \binom{7}{k} p^k (1-p)^{7-k}$

(II) Geometric: Bernoulli trials with probability p

X: # of trials to first success $\in \{1, 2, 3, 4, \cdots\}$

$f_X(k) = P(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, 3, 4, \cdots\cdots$

$\boxed{\text{Question: is } f_X(k) \text{ a pmf?}}$

Example: $Z_i = \begin{cases} 1 & \text{oil strike in } i\text{-th drill} \\ 0 & \text{o.w.} \end{cases} \qquad P(Z_i = 1) = p$

Oil company drills for oil

$X =$ # drills until the first oil strike occurs $\sim$ Geometric(p)

$P(X=6) = (1-p)^5 p$

(III) Negative Binomial $(NB(p,k))$

$X =$ # trials to get the k-th success in Bernoulli(p) trials $\in \{k, k+1, k+2, \dots\}$

$$P(X=m) = P[\cancel{k}-1 \text{ S in the first } m-1 \text{ trials, S in } m\text{-th trial}]$$

number of trials

$$= \binom{m-1}{k-1} p^{k-1} (1-p)^{m-k} \cdot p \qquad (\text{binomial \& Bernoulli})$$

# of success (fixing # of trials)

$$(\text{fixing \# successes}) = \binom{m-1}{k-1} p^k (1-p)^{m-k}, \quad m = k, k+1, \dots$$

Example (cont'd) $X =$ # drills until 3-rd oil strike occurs $\sim NB(p, 3)$

$$P(X=10) = \binom{10-1}{3-1} p^3 (1-p)^{10-3} = \binom{9}{2} p^3 (1-p)^7$$

(IV) Poisson: one way this arises is an approximation to Binomial when n is large & p is small such that $np \to \lambda$, $n \to \infty$

(e.g. # typos in a page of texts)

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad p = \lambda/n$$

$$= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1-\frac{\lambda}{n}\right)^{-k} \left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}, \ n \to \infty$$

$\hookrightarrow 1$, $n \to \infty$ for fixed k $\qquad 1$, $n \to \infty$

$$\to \frac{\lambda^k}{k!} e^{-\lambda}$$

Possion(λ): $f_Y(k) = P(Y=k) = \frac{\lambda^k}{k!} e^{-\lambda}$

Continuous r.v.'s : page 37, Section 1.7.2

(I) Uniform (a,b) $X \sim U(a,b)$

(II) Normal (Gaussian) $X \sim N(\mu, \sigma^2)$

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

Question: Is $f(x)$ a pdf?

(i) $f(x) > 0, \ \forall x \in \mathbb{R}$

(ii) $\int_{-\infty}^{\infty} f(x)\,dx \overset{?}{=} 1$

integration by substitution

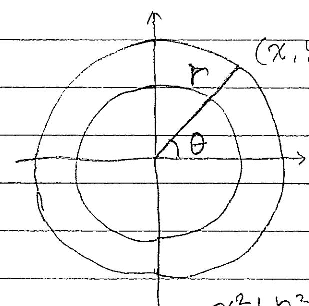
$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \overset{t = \frac{x-\mu}{\sigma}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = I$$

/kɑːtiːzjən/

$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \frac{1}{2\pi} \iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx\,dy$$

Cartesian coordinate system

$$= \frac{1}{2\pi} \int_0^{2\pi} \left[\int_0^{\infty} e^{-\frac{1}{2}r^2} \cdot r\,dr \right] d\theta$$



(Jacobian Matrix)

$$\Delta = \frac{\partial(x,y)}{\partial(r,\theta)} = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

$(x,y) \leftarrow (r,\theta)$

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}$$

$$|\Delta| = r\cos^2\theta + r\sin^2\theta = r$$

$$x^2+y^2 = r^2$$

$$\left\{ \begin{pmatrix} dx = \\ dy = \end{pmatrix} = \left(\text{Jacobian} \right) \begin{pmatrix} dr \\ d\theta \end{pmatrix} \right.$$

anti-derivative ('primitive function)

$$= \frac{1}{2\pi} \int_0^{2\pi} \left(-e^{-\frac{1}{2}r^2} \Big|_0^{\infty} \right) d\theta$$

$$= (2\pi)^{-1} \int_0^{2\pi} 1\,d\theta = 1$$

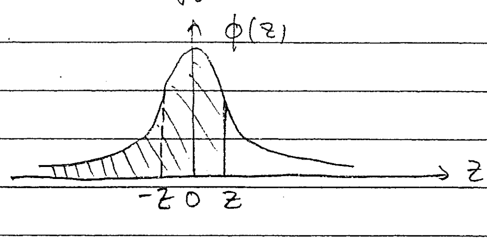
(polar coordinate system)

So, $I^2 = 1 \Rightarrow I = 1$

$$\int f(x,y)\,dx\,dy = \int f(x(u,v), y(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| du\,dv$$

Special case: standard normal $N(0,1)$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \phi(u)\,du \quad \text{no closed form}$$



$\Phi(z) = P(Z \le z)$, for $z \ge 0$

$\Phi(-z) = 1 - \Phi(z)$ $z \ge 0$

(III) Gamma family

$\int f g' dx = fg - \int g f' dx$

Gamma Function $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt, \quad \alpha > 0$

(i) $\Gamma(1) = \int_0^\infty e^{-t} dt = -e^{-t}\Big|_0^\infty = 1$

(ii) For $\alpha > 1$,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt = -\int_0^\infty t^{\alpha-1} de^{-t} \quad \text{(integration by parts)}$$

$$= -t^{\alpha-1} e^{-t}\Big|_0^\infty + \int_0^\infty e^{-t} d t^{\alpha-1}$$

$$= 0 + \int_0^\infty (\alpha-1) t^{\alpha-2} e^{-t} dt = (\alpha-1)\Gamma(\alpha-1)$$

Note: $\lim_{t\to\infty} t^{\alpha-1} e^{-t} = 0$, e.g., $\lim_{t\to\infty} t e^{-t} = \lim_{t\to\infty} \frac{t}{e^t} = \lim_{t\to\infty} \frac{1}{e^t} = 0$

L' Hospital's Rule

In particular, $\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2)$

$$= \cdots = (n-1)(n-2) \times \cdots \times 1\, \Gamma(1)$$

$$= (n-1)!$$

(iii) $\Gamma(\frac{1}{2}) = \int_0^\infty t^{-1/2} e^{-t} dt = \sqrt{\pi} \qquad t = \frac{x^2}{2}$

$$\int_0^\infty \sqrt{2} x^{-1} e^{-\frac{x^2}{2}} x\, dx$$

$$= \int_0^\infty \sqrt{2} e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} dx \, (\sqrt{2}\sqrt{2\pi})$$

$$= \frac{1}{2} \cdot 2\sqrt{\pi} = \sqrt{\pi}$$

$X \sim \text{Gamma}(\alpha, \beta)$ if the density is given by

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & , x > 0 \\ 0 & , x \le 0 \end{cases}$$

$\boxed{\text{Question: is } f(x) \text{ a pdf ?}}$

Special cases:

(a) $\alpha = 1$, $f(x) = \beta^{-1} e^{-x/\beta} I(x>0) \Leftrightarrow X \sim \text{Exp}(\beta)$ (exponential)

$$P(X > s) = \int_s^\infty \beta^{-1} e^{-x/\beta} dx = -e^{-x/\beta}\Big|_s^\infty = e^{-s/\beta}, \quad s > 0$$

$$P(X > s+t \mid X > t) = \frac{P(X > s+t)}{P(X > t)} = \frac{e^{-(s+t)/\beta}}{e^{-t/\beta}} = e^{-s/\beta} = P(X > s)$$

Does not depend on t

Lack of Memory Property

(b) $\alpha = \frac{1}{2}\nu$, $\beta = 2$, $f(x) = \frac{1}{\Gamma(\frac{\nu}{2}) 2^{\nu/2}} x^{\frac{\nu}{2}-1} e^{-x/2} I(x>0) \sim \chi^2_\nu$

ν - degree of freedom

Lecture 8

Chapter 2: Expectation and Moments

Def: Let X be a r.v. with pmf (pdf) $f(x)$, then the expected value of X:

$$\mu = E(X) = \sum_{x \in \mathcal{X}} x f(x), \quad \text{if } X \text{ is discrete}$$
$$\mu = E(X) = \int_{x \in \mathcal{X}} x f(x)\,dx, \quad \text{if } X \text{ is continuous}$$

Note: The expected value exists provided

$$\begin{cases} \sum |x| f(x) < +\infty & \text{(discrete)} \\ \int |x| f(x)\,dx < +\infty & \text{(continuous)} \end{cases}$$

If it fails, we say $E(X)$ does not exist.

① Median of a cont. r.v. X is m such that $P(X \le m) = P(X \ge m) = \frac{1}{2}$

② properties of expectation

Expected value of X = Mean of X (weighted average)

Example 1. $X \sim \text{Poisson}(\lambda)$ $\quad f(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \dots$

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

Trick: always make use of the known pdf/pmf.

Example 2. $X \sim N(0,1)$ $\quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}}\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x)\,dx$$

$-g(x) = g(-x)$ odd function. so, $E(X) = 0$ $\quad \boxed{\int_{-\infty}^{\infty} |x| e^{-\frac{x^2}{2}}\,dx < \infty}$

Fact: If density of X is symmetric about a, and $E(X)$ exists. Then $E(X) = a$

$\hookrightarrow f(a+x) = f(a-x)$

proof: w.o.l.g, assume $a = 0$, $\quad f(x) = f(-x)$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{\infty} g(x)\,dx = 0 \quad \text{as } g(-x) = -g(x)$$

For general a, $Y = X - a$, then the density of Y is symmetric about 0, so $E(Y) = 0 \Rightarrow E(X-a) = 0 \Rightarrow E(X) = a$

Example 3. $X \sim N(\mu, \sigma^2)$

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

① symmetric about $\mu \Leftrightarrow f(\mu+x) = f(\mu-x)$

② $\cancel{\int_{-\infty}^{\infty} |x| e^{-\frac{1}{2}x^2} dx < +\infty}$

$\int_0^{\infty} x e^{-\frac{1}{2}x^2} dx < +\infty$ by $\int_0^{\infty} x e^{-\frac{1}{2}x^2} dx = -e^{-\frac{1}{2}x^2}\Big|_0^{\infty} = 0-(-e^0) = 1$

$\int_{-\infty}^{\infty} |x| e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx < +\infty$

Then $E(X) = \mu$

Example 4. $X \sim$ Cauchy distribution $\quad f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \; -\infty < x < \infty$

① symmetric about $0 \Leftrightarrow f(x) = f(-x)$

but ② $\cancel{\int_{-\infty}^{\infty} \frac{|x|}{1+x^2} dx \text{ or } \int_0^{\infty} \frac{x}{1+x^2} dx = +\infty} \qquad \cancel{\frac{x}{1+x^2} \sim \frac{1}{x}, \; x \to \infty}$

$p>0$: $\int_a^{\infty} \frac{1}{x^p} dx \begin{cases} \text{diverges if } p \le 1 \\ \text{exists if } p > 1 \end{cases}$

Therefore $E(X)$ does not exist.

$X: S \to \mathbb{R}$

Expectations of functions of r.v.'s $\qquad g: \mathbb{R} \to \mathbb{R} \qquad$ composition

X is a r.v., $g: \mathbb{R} \to \mathbb{R}$, then $g(X)$ is a r.v., such as X^2, e^X

$$E[g(X)] = \begin{cases} \sum_x g(x) f(x), & x \text{ discrete} \\ \cancel{\int_x g(x) f(x) dx, \quad x \text{ continuous}} \end{cases}$$

[Actually there is another way!]
(Will learn later)

provided $\begin{cases} \sum_x |g(x)| f(x) < +\infty \\ \int_x |g(x)| f(x) < +\infty \end{cases}$

Example 1: $X \sim$ Poisson (λ), $\quad g(x) = x(x-1)$

$\cancel{E[g(X)] = E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x!} e^{-\lambda}}$

$= \sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!} e^{-\lambda} = \sum_{y=0}^{\infty} \frac{\lambda^{y+2}}{y!} e^{-\lambda} = \lambda^2$

$y = x-2$

Skip it if no time!

Example 2: variance of a r.v. X $\quad V(X)$ or $var(X)$

$\mathrm{var}(X) = E(X-\mu)^2 = E(X^2) - \mu^2$ where $\mu = E(X)$

$X \sim \text{Poisson}(\lambda)$, what is $\mathrm{var}(X)$

(Example 1 cont'd)

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} e^{-\lambda} = \sum_{x=1}^{\infty} \frac{x\lambda^x}{(x-1)!} e^{-\lambda} = \sum_{y=0}^{\infty} \frac{(y+1)\lambda^{y+1}}{y!} e^{-\lambda}$$

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

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

$$= \lambda \sum_{z=0}^{\infty} \frac{\lambda^{z+1}}{z!} e^{-\lambda} + \lambda = \lambda^2 + \lambda$$

$E(X) = \lambda$, so $\mathrm{var}(X) = E(X^2) - \lambda^2 = \lambda$

Properties of expectation:

(1) If $X \equiv c$, then $E(X) = c$

(2) $E[c\,g(X)] = c\,E[g(X)]$

(3) $E[a\,g(X) + b\,h(X)] = a\,E[g(X)] + b\,E[h(X)]$ linear operator

(4) $g(y) \le h(y)$ for all $y \Rightarrow E[g(X)] \le E[h(X)]$

Lecture 9 Moment Generating Function

Moments: The r-th moment of a r.v. X is $E(X^r)$ (if $E|X|^r < +\infty$)
e.g., $E(X)$ is the first moment

Central Moments: The r-th central moment of X is $E[(X-\mu)^r]$
e.g., $V(X) = E(X-\mu)^2$ is the second central moment

The first central moment $E(X-\mu) = 0$ always zero

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

$$E(X) = \int_0^\infty x \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}} dx = \int_0^\infty \frac{x^{(\alpha+1)-1} e^{-x/\beta}}{\Gamma(\alpha+1)\beta^{\alpha+1}} dx \; \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \beta$$

$$= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \beta = \alpha\beta \qquad (\Gamma(\alpha+1) = \alpha\Gamma(\alpha))$$

$$V(X) = E(X^2) - (EX)^2$$

$$E(X^2) = \int_0^\infty x^2 \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}} dx = \int_0^\infty \frac{x^{(\alpha+2)-1} e^{-x/\beta}}{\Gamma(\alpha+2)\beta^{\alpha+2}} dx \; \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \beta^2$$

$$= \alpha(\alpha+1)\beta^2$$

$$V(X) = \alpha\beta^2 + \alpha^2\beta^2 - (\alpha\beta)^2 = \alpha\beta^2$$

If $\alpha = 1$, $\exp(\beta)$: $E(X) = \beta$, $V(X) = \beta^2$

Moment Generating Function

$M_X(t) = E(e^{tX})$, also expectation, $t \in (-\infty, \infty)$, may exist for some t, not others

Example 1 $X \sim$ Bernoulli(p) $f(x) = \begin{cases} p, & x=1 \\ 1-p, & x=0 \end{cases}$

$$M_X(t) = E(e^{tX}) = pe^t + (1-p)e^{t\times 0} = pe^t + (1-p), \; \forall t$$

Example 2 $X \sim \exp(p)$ $f(x) = e^{-x}$, $x > 0$

$$M_X(t) = \int_0^\infty e^{tx} e^{-x} dx = \int_0^\infty e^{(t-1)x} dx = \frac{1}{t-1} e^{(t-1)x} \Big|_0^\infty$$

$$= \begin{cases} +\infty, & \text{if } t \geq 1 \\ \frac{1}{1-t}, & \text{if } t < 1 \end{cases}$$

$M_X(t)$ does not exist for $t \geq 1$

Theorem 2.3.2 (page 80)

X has mgf $M_X(t)$, $Y = cX + d$, then $M_Y(t) = e^{td} M_X(tc)$

$M_Y(t) = E(e^{tY}) = E[e^{t(cX+d)}] = E[e^{tcX}]e^{td}$

$= M_X(tc)\, e^{td}$

Example: $X \sim N(0,1)$, $M_X(t) = e^{\frac{t^2}{2}}$;

$Y \sim N(\mu, \sigma^2)$; $(Y = \mu + \sigma X)$, $M_Y(t) = e^{t\mu + \frac{\sigma^2 t^2}{2}}$] skip it!

Why mgf? (discrete)

$\cancel{\frac{d}{dt} M_X(t) = \frac{d}{dt} \sum_x e^{tx} f(x) = \sum_x x e^{tx} f(x)}$

$\Rightarrow \quad \frac{d}{dt} M_X(t) \Big|_{t=0} = \sum_x x f(x) = E(X)$

(similar for continuous)

$\frac{d^2}{dt^2} M_X(t) = \sum_x x^2 e^{tx} f(x) = \sum_x x^2 e^{tx} f(x)$

$\Rightarrow \quad \frac{d^2}{dt^2} M_X(t) \Big|_{t=0} = \sum_x x^2 f(x) = E(X^2)$

$\vdots$

$\frac{d^k M_X(t)}{dt^k} \Big|_{t=0} = E(X^k)$

Theorem (2.3.1) (page 79). $X \sim$ mgf $M_X(t)$, for $|t| < a$ with some $a > 0$, then the r-th moment of X is given by

$\cancel{\frac{d^k M_X(t)}{dt^k} \Big|_{t=0} = E(X^k)}$

$M_X(t) = E(e^{tX}) = E\left[1 + (tX) + \frac{1}{2!} t^2 X^2 + \cdots + \frac{1}{k!} t^k X^k + \cdots\right]$

Example 1: $X \sim$ Bernoulli(p), $M_X(t) = pe^t + 1 - p$

$dM_X(t)/dt = pe^t \rightarrow p$ at $t=0 \Leftrightarrow EX$

$d^2M_X(t)/dt^2 = pe^t \rightarrow p$ at $t=0 \Leftrightarrow EX^2$

$var(X) = p - p^2 = p(1-p)$

Example 2: $X \sim N(\mu, \sigma^2)$ $\quad N(0,1)$ first.

$M_X(t) = \cancel{\frac{1}{\sqrt{2\pi}} \int e^{tx} e^{-\frac{1}{2}x^2} dx}$

$= \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(x-t)^2} dx \; e^{\frac{1}{2}t^2} = e^{\frac{1}{2}t^2}$

↑ pdf of $N(t, 1)$.

$M_X(t) = e^{t\mu} e^{\frac{1}{2}t^2\sigma^2} = e^{t\mu + \frac{1}{2}t^2\sigma^2}$

I always said mean μ, variance σ^2; now they are!

$$\frac{dM_X(t)}{dt} = (\mu+\sigma^2 t)e^{t\mu+\frac{1}{2}\sigma^2 t^2}\Big|_{t=0} = \mu \qquad E(X)=\mu$$

$$\frac{d^2M_X(t)}{dt^2} = \left[\sigma^2 e^{t\mu+\frac{1}{2}\sigma^2t^2} + (\mu+\sigma^2t)^2 e^{t\mu+\frac{1}{2}\sigma^2t^2}\right]\Big|_{t=0} = \sigma^2+\mu^2 = E(X^2)$$

Therefore, $var(X) = \sigma^2$

Example 3. Geometric $f(x) = p(1-p)^{x-1}$, $x = 1,2,3,\ldots$

$$M_X(t) = E[e^{tX}] = \sum_{k=1}^{\infty} e^{tk}p(1-p)^{k-1}$$

$e^t(1-p) < 1 \Leftrightarrow e^t < \frac{1}{1-p} \Leftrightarrow t < \log\frac{1}{1-p}$

$$= pe^t\sum_{k=1}^{\infty}[e^t(1-p)]^{k-1} = pe^t\frac{1}{1-e^t(1-p)} \qquad \left(t<\log\frac{1}{1-p} = -\log(1-p)\right)$$

$$\frac{dM_X(t)}{dt} = \frac{pe^t[1-e^t(1-p)] + pe^t(e^t(1-p))}{[1-e^t(1-p)]^2}\Big|_{t=0} = \frac{p^2+p(1-p)}{p^2} = \frac{1}{p}$$

$E(X) = \frac{1}{p}$

(page 87) A finite mgf determines the distribution uniquely.

Thm. 2.4.1. If $M_X(t) = M_Y(t)$ for all t in an open interval containing 0, then $P(X\in A) = P(Y\in A)$ for all sets A. Therefore $X \stackrel{d}{=} Y$.

postpone!

Example 4. $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2\sim N(\mu_2,\sigma_2^2)$, X_1 & X_2 are independent

$X_1+X_2 \sim$?

$$M_{X_1}(t) = e^{t\mu_1+\frac{1}{2}t^2\sigma_1^2}, \quad M_{X_2}(t) = e^{t\mu_2+\frac{1}{2}t^2\sigma_2^2}$$

$$M_{X_1+X_2}(t) = E[e^{t(X_1+X_2)}] = E(e^{tX_1})E(e^{tX_2}) = e^{t(\mu_1+\mu_2)+\frac{1}{2}t^2(\sigma_1^2+\sigma_2^2)}$$

So, $X_1+X_2 \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

$$\begin{cases} -1 & 0.2 \\ 0 & 0.2 \\ 1 & 0.6 \end{cases}$$

Example 5. $M_X(t) = 0.2e^{-t}+0.2+0.6e^t$, what's the pmf of X?

postpone!

Example 6. $Z_1, Z_2, \ldots, Z_K$ iid $\sim N(0,1)$, $Z_1^2+Z_2^2+\cdots+Z_K^2 \sim \chi_K^2$

$M_X(t) = (1-\beta t)^{-\alpha}$, for Gamma$(\alpha,\beta)$ $\quad \alpha = \nu/2,\ \beta=2$

$M_{\chi_\nu^2}(t) = (1-2t)^{-\nu/2}$

$$M_{Z_1^2}(t) = Ee^{tZ_1^2} = \int e^{tz_1^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{z_1^2}{2}}dz_1 = \frac{1}{\sqrt{2\pi}}\int e^{-\frac{1}{2}(1-2t)z_1^2}dz$$

$$= \int \frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{1-2t}}}e^{-\frac{1}{2}z_1^2/\left(\frac{1}{\sqrt{1-2t}}\right)^2}dz_1\cdot\frac{1}{\sqrt{1-2t}} = \frac{1}{\sqrt{1-2t}} = (1-2t)^{-1/2}$$

$$\int e^{tx}\frac{1}{\Gamma(\alpha)\beta^\alpha}e^{-x/\beta}x^{\alpha-1}dx = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int e^{-(\frac{1}{\beta}-t)x}x^{\alpha-1}dx \qquad \beta^* = \frac{1}{\beta^{-1}-t} = \frac{\beta}{1-\beta t}$$

$$= (\beta^*/\beta)^\alpha = (1-\beta t)^{-\alpha}.$$

Lecture 10 Multivariate r.v.'s Chapter 3

We are often interested in more than 1 r.v. at a time, say, a college student, math, statistics, biology, chemistry
"Could be related"

Bivariate case: Let X_1 & X_2 are r.v.'s defined on S. Then
$\vec{X} = (X_1, X_2)'$ a random vector : $S \to \mathbb{R}^2$
If $g: \mathbb{R}^2 \to \mathbb{R}$, then $g(X_1, X_2)$ is a r.v.

If (X_1, X_2) takes only a finite / countably infinite set of values, then (X_1, X_2) discrete. Joint pmf
$$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$$

Example: 2 draws w.o. replacement from $\{1, 2, 3, 4\}$.
X_1 = smaller #, X_2 = larger # $\quad |S| = \binom{4}{2} = 6$

S		X_1	X_2
1 2	$\longrightarrow$	1	2
1 3		1	3
1 4		1	4
2 3		2	3
2 4		2	4
3 4		3	4

	X_2: 1	2	3	4	$f_{X_1}(x_1)$
X_1: 1	0	1/6	1/6	1/6	1/2
2	0	0	1/6	1/6	1/3
3	0	0	0	1/6	1/6
4	0	0	0	0	
$f_{X_2}(x_2)$	0	1/6	1/3	1/2	

Again, to satisfy the axioms of prob, the joint pmf must satisfy
(i) $f(x_1, x_2) \geq 0$
(ii) $\sum_{\text{all}(x_1, x_2)} f(x_1, x_2) = 1$

Support of $(X_1, X_2) \subseteq \text{supp}(X_1) \times \text{supp}(X_2)$

$\text{supp}(X_1) \times \text{supp}(X_2) = \{(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4)\}$

Marginal prob. function
$$f_{X_1}(x_1) = P(X_1 = x_1) = \sum_{x_2 \in X_2} P(X_1 = x_1, X_2 = x_2) = \sum_{x_2} f(x_1, x_2)$$

Similarly, $f_{X_2}(x_2) = \sum_{x_1} f(x_1, x_2)$
$P(X_1 = 1) = 1/6 + 1/6 + 1/6 = 1/2 = f_{X_1}(1)$

Joint pmf always determines marginal pmf, not vice versa

Example: $X_1, X_2 \sim$ Bernoulli $(\frac{1}{2})$ independent

$Y = 1 - X_1 \sim$ Bernoulli $(\frac{1}{2})$

(X_1, X_2) ~~same marginal~~ $P(X_1=0, X_2=0) = \frac{1}{4}$

(X_1, Y) $P(X_1=0, Y=0) = 0$

Note: Joint dist. has more information than marginal dist.'s

Expectations: $g: \mathbb{R}^2 \to \mathbb{R}^1$

$$E[g(X_1, X_2)] = \sum_{(x_1, x_2)} g(x_1, x_2) f(x_1, x_2)$$

Example: 2 draws w.o. replacement from $\{1,2,3,4\}$

$E(X_1 + X_2) = ?$

① $E(X_1+X_2) = 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} + 7\times\frac{1}{6} = 5$

② $E(X_1+X_2) = E(X_1) + E(X_2) = 1\times\frac{1}{2} + 2\times\frac{1}{3} + 3\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{3} + 4\times\frac{1}{2} = 5$

X_1 & X_2 are continuous.

A nonnegative function $f(x_1, x_2)$ is a joint pdf if

$$\begin{cases} f(x_1, x_2) \geq 0 \text{ for all } x_1, x_2 \\ \iint f(x_1, x_2)\,dx_1\,dx_2 = 1 \end{cases}$$

For any $A \subseteq \mathbb{R}^2$, $P((X_1, X_2) \in A) = \iint_A f(x_1, x_2)\,dx_1\,dx_2$

any $g: \mathbb{R}^2 \to \mathbb{R}$, $E[g(X_1, X_2)] = \iint g(x_1, x_2) f(x_1, x_2)\,dx_1\,dx_2$

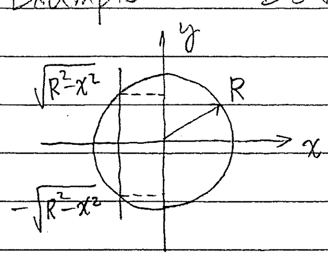
Marginal densities:

$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\,dx_2$ $E(X_1) = \cancel{\iint x_1 f(x_1, x_2)\,dx_1\,dx_2}$

$= \int x_1 f_{X_1}(x_1)\,dx_1$

$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\,dx_1$ $E(X_2) = \int x_2 f_{X_2}(x_2)\,dx_2$

Example 1. Disk $D = \{(x, y): x^2 + y^2 \leq R^2\}$



"Randomly" pick up a point in the disk, i.e.,

$$f(x, y) = \begin{cases} C & x^2+y^2 \leq R^2 \\ 0 & \text{o.w.} \end{cases}$$

(a) Choose c such that $f(x,y)$ is a joint pdf

$$\iint_D f(x,y)\,dx\,dy = 1 \Rightarrow \iint_D c\,dx\,dy = c \times \text{area}(D) = c\pi R^2 = 1$$

$$\Rightarrow \quad c = \frac{1}{\pi R^2}.$$

(b) Find the marginal dist. of X

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}\,dy = \frac{2}{\pi R^2}\sqrt{R^2-x^2} \quad (-R<x<R)$$

(Not uniform, why?)

(c) $E(X)$?

odd function

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_{-R}^{R} x\,\frac{2}{\pi R^2}\sqrt{R^2-x^2}\,dx = 0$$

Example 2.

$$f(x_1,x_2) = \begin{cases} x_1+x_2 & 0<x_1,x_2<1 \\ 0 & \text{o.w.} \end{cases}$$

$$f_{X_1}(x_1) = \int_0^1 (x_1+x_2)\,dx_2 = x_1+\frac{1}{2}, \quad 0<x_1<1$$

$$f_{X_2}(x_2) = x_2+\frac{1}{2}, \quad 0<x_2<1$$

$$P(X_1+X_2\le 1) = \iint_{x_1+x_2\le 1} (x_1+x_2)\,dx_1\,dx_2 = \int_0^1\left[\int_0^{1-x_2}(x_1+x_2)\,dx_1\right]dx_2$$

$$= \int_0^1 \left[(1-x_2)x_2 + \frac{1}{2}x_1^2\Big|_0^{1-x_2}\right]dx_2$$

$$= \int_0^1 \left[x_2 - x_2^2 + \frac{1}{2}(1-2x_2+x_2^2)\right]dx_2$$

$$= \int_0^1 \left[\frac{1}{2}-\frac{1}{2}x_2^2\right]dx_2$$

$$= \frac{1}{2} - \frac{1}{6}x_2^3\Big|_0^1 = \frac{1}{2}-\frac{1}{6} = \frac{1}{3}$$

Lecture 11 → 12

Conditional distributions

(X_1, X_2) discrete. $P(X_2 = x_2 | X_1 = x_1) \triangleq f(x_2|x_1)$

$$= \frac{P(X_2 = x_2, X_1 = x_1)}{P(X_1 = x_1)} = \frac{f(x_1, x_2)}{f_1(x_1)}$$

conditional pmf of X_2 given $X_1 = x_1$.

Likewise, $f(x_1|x_2) = \dfrac{f(x_1, x_2)}{f_2(x_2)}$

(X_1, X_2) continuous, conditional pdf $f(x_1|x_2) = f(x_1, x_2)/f_2(x_2)$

$f(x_2|x_1) = f(x_1, x_2)/f_1(x_1)$

Is $f(x_1|x_2)$ a pmf or pdf? (for fixed x_2) $(f_2(x_2)>0$ or $f_1(x_1)>0)$

(1) $f(x_1|x_2) \geq 0$

(2) $\sum_{x_1} f(x_1|x_2) = \sum_{x_1} f(x_1, x_2) / f_2(x_2) = f_2(x_2)/f_2(x_2) = 1$

Conditional expectations

$E(X_1 | X_2 = x_2) = \sum_{x_1} x_1 f(x_1|x_2)$

or $E(g(X_1) | X_2 = x_2) = \sum_{x_1} g(x_1) f(x_1|x_2)$

Theorem: $E[g(X_1)] = E[E\{g(X_1)|X_2\}]$

2-step procedure where is easier than a direct way sometimes

Example: $\{X_i\}_{i=1}^{n}$ is a sequence of independent Bernoulli variables with parameter p. Conditional on $\sum_{i=1}^{n} X_i$, $Y \sim \text{Poisson}(\sum_{i=1}^{n} X_i)$,

Find $E(Y)$.

$E(Y | \sum_{i=1}^{n} X_i = x) = x$

$E(Y) = E[E(Y|\sum_{i=1}^{n} X_i)] = E[\sum_{i=1}^{n} X_i] = np$

Is $f(x_1|x_2)$ a pdf?

(1) $f(x_1|x_2) \geq 0$

(2) $\int f(x_1|x_2)\,dx_1 = \dfrac{\int f(x_1, x_2)\,dx_1}{f_2(x_2)} = \dfrac{f_2(x_2)}{f_2(x_2)} = 1$

Example: $f(x_1, x_2) = x_1 + x_2, \quad 0 \le x_1, x_2 \le 1$

$$f_1(x_1) = \int_0^1 f(x_1, x_2)\,dx_2 = \int_0^1 x_1\,dx_2 + \int_0^1 x_2\,dx_2 = x_1 + \frac{1}{2}$$

$f(x_2|x_1) = (x_1 + x_2)/(x_1 + \frac{1}{2}), \quad 0 \le x_2 \le 1$ given $0 \le x_1 \le 1$

$$E(X_1 | X_2 = x_2) = \int_0^1 x_1 f(x_1|x_2)\,dx_1 = \int_0^1 \frac{x_1(x_1+x_2)}{x_2 + \frac{1}{2}}\,dx_1$$

$$= \frac{1/3 + x_2/2}{x_2 + 1/2} = \frac{3X_2 + 2}{6X_2 + 3}$$

So, $E(X_1 | X_2 = 0.5) = 3.5/6 = 7/12$.

$$\mathrm{var}(X_1 | X_2 = x_2) \stackrel{\Delta}{=} E[X_1^2 | X_2 = x_2] - [E\{X_1 | X_2 = x_2\}]^2 \stackrel{\Delta}{=} E[(X_1 - \mu_{X_1|X_2})^2 | X_2 = x_2]$$

$$= \int_0^1 \frac{x_1^2(x_1+x_2)}{x_2 + 1/2}\,dx_1 - \left(\frac{3x_2+2}{6x_2+3}\right)^2$$

$$= \frac{\frac{1}{4} + \frac{1}{3}x_2}{x_2 + 1/2} - \frac{(\frac{x_2}{2} + \frac{1}{3})^2}{(x_2 + \frac{1}{2})^2}$$

$$= \frac{1}{(x_2+\frac{1}{2})^2}\left[\left(\frac{1}{3}x_2 + \frac{1}{4}\right)\left(x_2 + \frac{1}{2}\right) - \left(\frac{x_2}{2} + \frac{1}{3}\right)^2\right]$$

$$= \frac{1}{(x_2+\frac{1}{2})^2}\left(\frac{1}{12}x_2^2 + \frac{1}{12}x_2 + \frac{1}{72}\right)$$

What's the relationship between $\mathrm{var}(X_1)$ & $\mathrm{var}(X_1 | X_2 = x_2)$?

Theorem 3.3.1 (page 112)

$$\mathrm{var}(X_1) = E[\mathrm{var}(X_1|X_2)] + \mathrm{var}[E(X_1|X_2)]$$

Proof: $\text{RHS} = E[E(X_1^2|X_2)] - E[\{E(X_1|X_2)\}^2] + E[\{E(X_1|X_2)\}^2] - [E\{E(X_1|X_2)\}]^2$

$= E(X_1^2) - E[\{E(X_1|X_2)\}^2] + E[\{E(X_1|X_2)\}^2] - [E(X_1)]^2$

$= E(X_1^2) - [E(X_1)]^2$

Example: Conditional on $X_1 = x_1$, $X_2 | X_1 = x_1 \sim N(\beta_0 + \beta_1 x_1, x_1^2)$

& Marginally $X_1 \sim N(3, 10)$.

Find $E(X_2)$ & $\mathrm{var}(X_2)$.

$E(X_2 | X_1 = x_1) = \beta_0 + \beta_1 x_1$

$\mathrm{var}(X_2 | X_1 = x_1) = x_1^2$

$E[X_2] = E(\beta_0 + \beta_1 X_1) = \beta_0 + 3\beta_1$

$\mathrm{var}(X_2) = E[\mathrm{var}(X_2|X_1)] + \mathrm{var}[E(X_2|X_1)]$

$= E[X_1^2] + \mathrm{var}[\beta_0 + \beta_1 X_1]$

$= (1 + \beta_1^2)\,\mathrm{var}(X_1) + [EX_1]^2$

$= 10(1 + \beta_1^2) + 9$

Lecture 13 or 14?

Independence:

Events A & B are independent if $P(A|B)=P(A)$ or $P(A\cap B)=P(A)P(B)$

Two r.v.'s X and Y are independent if

$$P(X\in A \mid Y\in B) = P(X\in A) \quad \text{for all sets } A \,\&\, B$$

$$\Leftrightarrow \quad P(X\in A \cap Y\in B) = P(X\in A)\, P(Y\in B) \qquad (1)$$

If (X,Y) has joint pmf or pdf, then other equivalent statements are

$$\begin{cases} f(x|y) = f_X(x), \text{ for all } x, y \\ \text{or } \quad f(x,y) = f_X(x) f_Y(y), \text{ for all } x, y. \qquad (2) \end{cases}$$

pf (discrete): $(1)\Rightarrow(2)$ Choose $A=\{x\}$ & $B=\{y\}$

$(2)\Rightarrow(1)$ $P(X\in A, Y\in B) = \sum_A \sum_B f(x,y)$

$= \sum_A \sum_B f_X(x) f_Y(y) = \sum_A f_X(x) \sum_B f_Y(y) = P(X\in A)P(Y\in B)$

Example: X_1, X_2 independent $\sim N(\mu_1,\sigma_1^2)$, $N(\mu_2,\sigma_2^2)$, respectively. Then

$$f(x_1,x_2) = f_1(x_1) f_2(x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

If X_1 and X_2 are independent, their marginals $\Rightarrow$ joint

What if you do not know marginals and want to check if they are independent?

Lemma. Let (X_1,X_2) have joint p.m.f. (p.d.f.) $f(x_1,x_2)$. Then X_1 & X_2 are independent $\Leftrightarrow$ $\exists\, g$ & h, s.t. $f(x_1,x_2) = g(x_1)h(x_2)$, $\forall (x_1,x_2)$

pf: $\Rightarrow$ If X_1 & X_2 independent, then $g=f_1$ & $h=f_2$.

$\Leftarrow$ $f(x_1,x_2) = g(x_1)h(x_2)$,

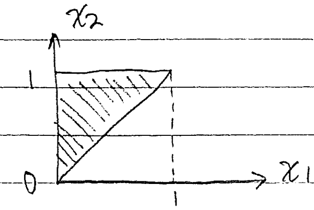
$$f_1(x_1) = \int f(x_1,x_2)\,dx_2 = g(x_1)\int h(x_2)\,dx_2 = C_1 g(x_1)$$

$$f_2(x_2) = \int f(x_1,x_2)\,dx_1 = h(x_2)\int g(x_1)\,dx_1 = C_2 h(x_2)$$

So, $\quad f(x_1,x_2) = \frac{1}{C_1} f_1(x_1) \frac{1}{C_2} f_2(x_2)$,

by $\iint f(x_1,x_2)\,dx_1 dx_2 = 1 \Leftrightarrow C_1C_2 = 1$, so $f(x_1,x_2) = f_1(x_1) f_2(x_2)$

Example: $f(x_1,x_2)=\begin{cases} 8x_1x_2, & 0<x_1<x_2<1 \\ 0, & \text{otherwise} \end{cases}$



Are X_1 & X_2 independent? Seems so?

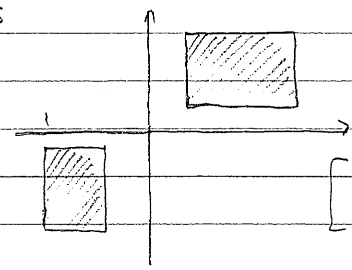
$f_1(x_1)=\int_{x_1}^{1} 8x_1x_2\,dx_2 = 8x_1 \frac{1}{2}x_2^2\Big|_{x_1}^{1} = 4x_1(1-x_1^2), \quad 0<x_1<1$

$f_2(x_2)=\int_{0}^{x_2} 8x_1x_2\,dx_1 = 8x_2\frac{1}{2}x_1^2\Big|_{0}^{x_2} = 4x_2^3, \quad 0<x_2<1$

$f(x_1,x_2) \neq f_1(x_1)f_2(x_2)$, NOT Independent. But why?

Point 1: $f(x_1,x_2)=8x_1x_2 I(0<x_1<x_2<1) \neq g(x_1)h(x_2)$

2: The support must be rectangular for two continuous independent r.v.'s



Sometimes it's a good way to check dependence!

[$f(x_1,x_2)=g(x_1)h(x_2)$ for all possible x_1, x_2. not only for the region $0<x_1<x_2<1$]

Properties of independent r.v.'s:

$X_1 \perp X_2 \Rightarrow E[g(X_1)h(X_2)] = E[g(X_1)]E[h(X_2)]$

pf: $E[g(X_1)h(X_2)]$

$= \iint g(x_1)h(x_2)f(x_1,x_2)\,dx_1dx_2 = \iint g(x_1)f_1(x_1)\,h(x_2)f_2(x_2)\,dx_1dx_2$

$= \int g(x_1)f_1(x_1)dx_1 \int h(x_2)f_2(x_2)dx_2 = E[g(X_1)]E[h(X_2)]$

(i) $M_{X_1+X_2}(t)=M_{X_1}(t)M_{X_2}(t)$ if $X_1\perp X_2$

$E[e^{t(X_1+X_2)}] = E(e^{tX_1}e^{tX_2}) = E(e^{tX_1})E(e^{tX_2}) = M_{X_1}(t)M_{X_2}(t)$

(ii) $X_1 \perp X_2 \Rightarrow E(X_1X_2)=E(X_1)E(X_2)$

$var(X_1X_2) = var(X_1)\,var(X_2)$?

$var(X_1X_2) = E(X_1^2X_2^2) - [E(X_1X_2)]^2 = E(X_1^2)E(X_2^2) - E^2(X_1)E^2(X_2)$

$var(X_1)var(X_2) = [E(X_1^2)-E^2(X_1)][E(X_2^2)-E^2(X_2)]$

But $var(aX_1+bX_2) = a^2 var(X_1) + b^2 var(X_2)$

Example: $X_1 \sim$ Poisson (λ_1), $X_2 \sim$ Poisson (λ_2)

$X_1 \perp X_2$ then $X_1+X_2 \sim$ Poisson $(\lambda_1+\lambda_2)$

$M_{X_1}(t)=e^{\lambda_1(e^t-1)}$, $M_{X_2}(t)=e^{\lambda_2(e^t-1)} \Rightarrow M_{X_1+X_2}(t)=e^{(\lambda_1+\lambda_2)(e^t-1)}$

is the mgf of Poisson $(\lambda_1+\lambda_2)$

(iii) $X_1 \perp X_2 \Rightarrow Y_1 = g(X_1) \perp Y_2 = h(X_2)$

Example: $X_1 \sim \Gamma(\alpha_1, \beta)$, $X_2 \sim \Gamma(\alpha_2, \beta)$ $(X_1 \perp X_2)$

$M_1(t) = (1-\beta t)^{-\alpha_1}$, $M_2(t) = (1-\beta t)^{-\alpha_2}$

$M_{X_1+X_2}(t) = (1-\beta t)^{-(\alpha_1+\alpha_2)}$

$\Rightarrow X_1 + X_2 \sim \Gamma(\alpha_1+\alpha_2, \beta)$

$X_1 \sim \chi^2_{p_1} \sim \Gamma(\frac{p_1}{2}, 2)$, $X_2 \sim \chi^2_{p_2} \sim \Gamma(\frac{p_2}{2}, 2)$, $X_1 \perp X_2$

$X_1 + X_2 \sim \Gamma((p_1+p_2)/2, 2) \sim \chi^2_{p_1+p_2}$

Lecture 15 Covariance & Correlation

Covariance of X_1 and X_2: $\mathrm{cov}(X_1,X_2) = E[(X_1-EX_1)(X_2-EX_2)]$
$= E(X_1X_2) - E(X_1)E(X_2)$

If $X_1=X_2$, $\mathrm{cov}(X_1,X_1) = \mathrm{var}(X_1)$

Properties:

(i) $\mathrm{cov}(X_1,X_2) = \mathrm{cov}(X_2,X_1)$

(ii) $\mathrm{cov}(c,X_2) = 0$

(iii) $\mathrm{cov}(X_1+X_2,Y) = \mathrm{cov}(X_1,Y) + \mathrm{cov}(X_2,Y)$

(iv) $\mathrm{cov}(cX_1,X_2) = c\,\mathrm{cov}(X_1,X_2) = \mathrm{cov}(X_1,cX_2)$

(v) $\mathrm{var}(X_1+X_2) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + 2\mathrm{cov}(X_1,X_2)$

pf: (v). $\mathrm{var}(X_1+X_2) = E[(X_1+X_2) - (EX_1+EX_2)]^2$
$= E[(X_1-EX_1)^2] + E[(X_2-EX_2)^2] + 2E[(X_1-EX_1)(X_2-EX_2)]$
$= \mathrm{var}(X_1) + \mathrm{var}(X_2) + 2\mathrm{cov}(X_1,X_2)$

(vi) $X_1 \perp X_2 \Rightarrow \mathrm{cov}(X_1,X_2)=0$

Remark 1: $X_1\perp X_2$ (or $\mathrm{cov}(X_1,X_2) \overset{"}{=} 0$ suffices here)
$\Rightarrow \mathrm{var}(X_1+X_2) = \mathrm{var}(X_1)+\mathrm{var}(X_2)$

Remark 2 $\mathrm{cov}(X_1,X_2) \not\Rightarrow X_1 \perp X_2$

Example: $X_1 \sim N(0,1)$, $X_2 = X_1^2$
$\mathrm{cov}(X_1,X_2) = E(X_1X_2) - (EX_1)(EX_2) = EX_1^3 - EX_1EX_1^2 = 0-0=0$
But X_1 and X_2 are not independent. (?)

Correlation Coefficient of (X_1,X_2) is given by

$$\rho_{X_1,X_2} = \frac{\mathrm{cov}(X_1,X_2)}{\sqrt{\mathrm{var}(X_1)}\sqrt{\mathrm{var}(X_2)}} = E\left[\left(\frac{X_1-\mu_1}{\sigma_1}\right)\left(\frac{X_2-\mu_2}{\sigma_2}\right)\right] \qquad \begin{array}{l}\mathrm{var}(X_1)=\sigma_1^2\\ \mathrm{var}(X_2)=\sigma_2^2\end{array}$$

ρ_{X_1,X_2} is a measure of "linear" correlation

(quadratic function)

(i) $-1 \le \rho_{X_1,X_2} \le 1$

pf: $h(v) = E[(X_1-\mu_1) + v(X_2-\mu_2)]^2 = \sigma_1^2 + v^2\sigma_2^2 + 2v\,\mathrm{cov}(X_1,X_2) \ge 0$

$a(x+\frac{b}{2a})^2 + c - \frac{b^2}{4a} \ge 0$

$ax^2+bx+c \ge 0$ for all x So, $4\mathrm{cov}^2(X_1,X_2) \le 4\sigma_1^2\sigma_2^2$

$\Leftrightarrow \begin{cases} a>0 \\ b^2-4ac\le 0 \end{cases}$ or $\begin{cases} a=0 \\ b=0,\ c\ge 0 \end{cases}$ $\Leftrightarrow \left|\frac{\mathrm{cov}(X_1,X_2)}{\sigma_1\sigma_2}\right| \le 1$

(ii) $\rho_{X_1,X_2}=0$ says (X_1,X_2) are uncorrelated $\Leftrightarrow \mathrm{cov}(X_1,X_2)=0$

(ii) $\rho_{X_1,X_2} = 1 \iff X_2 = aX_1 + b, \quad a > 0$

$\qquad\qquad\quad = -1 \iff X_2 = aX_1 + b, \quad a < 0$

pf: $\Leftarrow$
$$\rho_{X_1,X_2} = \frac{cov(X_1,X_2)}{\sqrt{var(X_1)\,var(X_2)}} = \frac{a\,var(X_1)}{\sqrt{var(X_1)\,a^2 var(X_1)}} = 1$$

as $cov(X_1,X_2) = cov(X_1, aX_1+b) = a\,cov(X_1,X_1) = a\,var(X_1)$

$\Rightarrow$
$$var\left(\frac{X_1}{\sigma_1} - \rho_{X_1,X_2}\frac{X_2}{\sigma_2}\right) \geq 0$$

$$1 + \rho_{X_1,X_2}^2 - 2\,cov\left(\frac{X_1}{\sigma_1}, \rho_{X_1,X_2}\frac{X_2}{\sigma_2}\right) = 1 + \rho_{X_1,X_2}^2 - 2\rho_{X_1,X_2}^2 = 1 - \rho_{X_1,X_2}^2 \geq 0$$

$\iff |\rho_{X_1,X_2}| \leq 1$

When $|\rho_{X_1,X_2}| = 1$, $\quad var\left(\frac{X_1}{\sigma_1} - \rho_{X_1,X_2}\frac{X_2}{\sigma_2}\right) = 0$

$$\iff \frac{X_1}{\sigma_1} = \rho_{X_1,X_2}\frac{X_2}{\sigma_2} + c$$

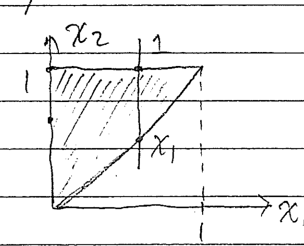
$$\Rightarrow X_1 = aX_2 + b \quad \text{with} \quad a = \rho_{X_1,X_2}\frac{\sigma_1}{\sigma_2}$$

Note: ρ_{X_1,X_2} measures only linear correlation.

Example: $X_1 \sim N(0,1)$, $X_2 = X_1^2$, $cov(X_1,X_2) = 0 \Rightarrow \rho_{X_1,X_2} = 0$

but X_1 and X_2 are perfectly quadratically correlated

Example: $f(x_1,x_2) = \begin{cases} 2 & 0 < x_1 < x_2 < 1 \\ 0 & o.w. \end{cases}$



$$f_1(x_1) = \int_{x_1}^{1} 2\,dx_2 = 2(1-x_1), \quad 0 < x_1 < 1$$

$$f_2(x_2) = \int_0^{x_2} 2\,dx_1 = 2x_2, \quad 0 < x_2 < 1$$

$$EX_1 = \int_0^1 x_1\,2(1-x_1)\,dx_1 = 1 - 2\cdot\frac{1}{3} = \frac{1}{3}, \quad EX_2 = \int_0^1 x_2\cdot 2x_2\,dx_2 = \frac{2}{3}$$

$$E(X_1X_2) = \int_0^1\int_0^{x_2} 2x_1x_2\,dx_1\,dx_2 = \int_0^1 2x_2\,\frac{1}{2}x_2^2\,dx_2 = \int_0^1 x_2^3\,dx_2 = \frac{1}{4}$$

$$cov(X_1,X_2) = E(X_1X_2) - (EX_1)(EX_2) = \frac{1}{4} - \frac{2}{9} = 1/36$$

$$Var(X_1) = E(X_1^2) - \frac{1}{9} = \int_0^1 2x_1^2(1-x_1)\,dx_1 - \frac{1}{9} = \left(\frac{2}{3} - \frac{2}{4}\right) - \frac{1}{9} = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$$

$$var(X_2) = E(X_2^2) - \frac{4}{9} = \int_0^1 2x_2^3\,dx_2 - \frac{4}{9} = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}.$$

$$\rho_{X_1,X_2} = \frac{1/36}{\sqrt{\frac{1}{18}\cdot\frac{1}{18}}} = \frac{1/36}{1/18} = \frac{1}{2}$$

Lecture 16 Bivariate Normal

Bivariate Normal Distribution $(X_1, X_2) \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right)$ if

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 \right] \right\} \quad (*)$$

$-\infty < x_1, x_2 < \infty, \ \sigma_1 > 0, \ \sigma_2 > 0, \ -1 < \rho < 1$

Note: if $\rho = 0$, i.e., X_1 and X_2 are uncorrelated,

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ -\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 \right] \right\},$$

which is a product of $N(\mu_1, \sigma_1^2)$ & $N(\mu_2, \sigma_2^2)$ p.d.f.'s

$\Rightarrow$ X_1 and X_2 are independent

We show $(*)$ is a pdf: $\iint f(x_1, x_2)\, dx_1\, dx_2 = 1$

Let $u_1 = (x_1 - \mu_1)/\sigma_1$ and $u_2 = (x_2 - \mu_2)/\sigma_2$

then $\iint f(x_1, x_2)\, dx_1\, dx_2 = \iint \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} (u_1^2 - 2\rho u_1 u_2 + u_2^2) \right\} du_1\, du_2$

$= \iint \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left((u_1 - \rho u_2)^2 + u_2^2 - \rho^2 u_2^2 \right) \right\} du_1\, du_2$

$= \frac{1}{\sqrt{2\pi}} \iint \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} (u_1 - \rho u_2)^2 \right\}}_{\text{pdf of } N(\rho u_2,\ 1-\rho^2)} du_1 \exp\left\{ -\frac{1}{2} u_2^2 \right\} du_2$

$= \frac{1}{\sqrt{2\pi}} \int \exp\left\{ -\frac{1}{2} u_2^2 \right\} du_2 = 1$

Bivariate normal has many interesting properties:

(i) Marginal of X_i is $N(\mu_i, \sigma_i^2)$, $i = 1, 2$

pf: $f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1$

$f(x_1, x_2) = g(x_1, x_2)\, h(x_2)$, with

$$g(x_1, x_2) = \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\,\sigma_1} \exp\left\{ -\frac{1}{2(1-\rho^2)\sigma_1^2} \left[(x_1 - \mu_1) - \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2) \right]^2 \right\}}_{N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2),\ (1-\rho^2)\sigma_1^2\right)}$$

$$h(x_2) = \frac{1}{\sqrt{2\pi}\,\sigma_2} \exp\left\{ -\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 \right\}$$

so $f_{X_2}(x_2) = \int_{-\infty}^{\infty} g(x_1, x_2)\, h(x_2)\, dx_1 = h(x_2) \sim N(\mu_2, \sigma_2^2)$

(ii) Conditional distribution of $X_1 | X_2 = x_2 \sim N(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2))$

$X_2 | X_1 = x_1 \sim N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2))$

pf:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = g(x_1, x_2)$$

(iii) $\mathrm{corr}(X_1, X_2) = \rho$. Let $Z_1 = \frac{X_1 - \mu_1}{\sigma_1}$ & $Z_2 = \frac{X_2 - \mu_2}{\sigma_2} \sim N(0,1)$

$$\rho_{X_1, X_2} = E\left[\left(\frac{X_1 - \mu_1}{\sigma_1}\right)\left(\frac{X_2 - \mu_2}{\sigma_2}\right)\right] = E(Z_1 Z_2)$$

$$E(Z_1 Z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \iint z_1 z_2 e^{-\frac{1}{2}z_2^2} e^{-\frac{1}{2(1-\rho^2)}(z_1 - \rho z_2)^2} dz_1 dz_2$$

$$= \int \frac{1}{\sqrt{2\pi}} z_2 e^{-\frac{1}{2}z_2^2} \underbrace{\int \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} z_1 e^{-\frac{1}{2(1-\rho^2)}(z_1 - \rho z_2)^2} dz_1}_{\rho z_2} dz_2$$

$$= \int \frac{1}{\sqrt{2\pi}} \rho z_2^2 e^{-\frac{1}{2}z_2^2} dz_2$$

$$= \rho E(Z_2^2) = \rho$$

BVN always have normal marginals, but normal marginals does not imply BVN

Linear regression: $(X_1, X_2) \sim$ BVN

$$E(X_2 | X_1 = x_1) = \mu_2 + \rho\sigma_2\left(\frac{X_1 - \mu_1}{\sigma_1}\right) = \rho\frac{\sigma_2}{\sigma_1}X_1 + (\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1) = \beta_1 x_1 + \beta_0$$

$$\begin{cases} \beta_1 = \rho\frac{\sigma_2}{\sigma_1} = \frac{\mathrm{cov}(X_1, X_2)}{\sigma_1\sigma_2}\frac{\sigma_2}{\sigma_1} = \frac{\mathrm{cov}(X_1, X_2)}{\sigma_1^2} \\ \beta_0 = \mu_2 - \beta_1\mu_1 \end{cases}$$

In the least square estimation

$$\min \sum_{i=1}^{n}(X_{2i} - \beta_0 - \beta_1 X_{1i})^2 \qquad \begin{cases} \sum_{i=1}^{n}(X_{2i} - \beta_0 - \beta_1 X_{1i}) = 0 \\ \sum_{i=1}^{n}(X_{2i} - \beta_0 - \beta_1 X_{1i})X_{1i} = 0 \end{cases}$$

$$\hat{\beta}_1 = \frac{\sum(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sum(X_{1i} - \bar{X}_1)^2} = \frac{\text{sample cov}(X_1, X_2)}{\text{sample var}(X_1)}$$

$\hat{\beta}_0 = \bar{X}_2 - \hat{\beta}_1\bar{X}_1$ similar

More than two r.v.'s

$\vec{X} = (X_1, \cdots, X_n)^T$ each X_i is a r.v.

If each X_i is discrete, the dist. of $\vec{X}$ is described by its joint pmf

$$f(x_1, \cdots, x_n) = P(X_1 = x_1, \cdots, X_n = x_n)$$

For any $A \subseteq \mathbb{R}^n$,

$$P(X \in A) = \sum_{(x_1, \cdots, x_n) \in A} f(x_1, \cdots, x_n)$$

If each X_i is cont., $f(x_1, \cdots, x_n)$ is a joint pdf,

$$P(X \in A) = \int \cdots \int_A f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n$$

Example 1. Roll 3 dice. X_i = outcome of i-th roll

$$f(x_1, x_2, x_3) = 1/6^3 = 1/216, \quad \text{for } 1 \le x_1, x_2, x_3 \le 6$$

Example 2. $f(x_1, x_2, x_3) = e^{-x_1 - x_2 - x_3}, \quad x_1 > 0, x_2 > 0, x_3 > 0$

Marginal pmf/pdf of X_i

$$f_{X_i}(x_i) = \sum_{x_j, j \ne i} f(x_1, \cdots, x_n)$$

or

$$f_{X_i}(x_i) = \int \cdots \int_{x_j, j \ne i} f(x_1, \cdots, x_n)\, dx_1 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_n$$

Sum/integration out other variables

Ex 1 (cont.) $f_1(x_1) = \sum_{x_2} \sum_{x_3} 1/216 = \sum_{x_2=1}^{6} \sum_{x_3=1}^{6} \frac{1}{216} = \frac{1}{6}, \quad x_1 = 1, \cdots, 6$

Ex 2 (cont.) $f_2(x_2) = \int_0^\infty \int_0^\infty e^{-x_1 - x_2 - x_3}\, dx_1\, dx_3 = e^{-x_2}, \quad x_2 > 0$

Marginal of 2 r.v.'s (X_i, X_j)

$$\begin{cases} f_{i,j}(x_i, x_j) = \sum_{x_l, l \ne i,j} f(x_1, \cdots, x_n) \\ f_{i,j}(x_i, x_j) = \int \cdots \int_{x_l, l \ne i,j} f(x_1, \cdots, x_n)\, dx_1 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_{j-1}\, dx_{j+1} \cdots dx_n \end{cases}$$

Ex 1 (cont.) $f_{1,2}(x_1, x_2) = \sum_{x_3=1}^{6} f(x_1, x_2, x_3) = 1/36$

2 $\quad f_{2,3}(x_2, x_3) = \int_0^\infty e^{-x_1 - x_2 - x_3}\, dx_1 = e^{-x_1 - x_2}, \quad x_1 > 0, x_2 > 0$

Conditional pmf/pdf

$$f(x_1 \mid x_2, \cdots, x_n) = \frac{f(x_1, \cdots, x_n)}{f_{2, \cdots, n}(x_2, \cdots, x_n)}$$

$$f(x_1, x_2 \mid x_3, \cdots, x_n) = \frac{f(x_1, \cdots, x_n)}{f_{3, \cdots, n}(x_3, \cdots, x_n)}$$