

STAT 240

Personal Notes

Marcus Chan (UW 25)

Taught by Ying-Li Qin



Chapter 1: What is Probability?

RANDOM EXPERIMENTS (1.1)

A "random experiment" is the process of obtaining a random observed result.

Random experiments can be split into two types:

① Controlled experiments; and

eg flipping a coin, rolling a die

② Observational studies.

eg # of students taking STAT 240 in F2021

FEATURES OF RANDOM EXPERIMENTS

Note that random experiments have the following common features:

- ① The outcomes/results cannot be predicted with certainty; and
- ② All the possible outcomes are known beforehand with certainty.

SAMPLE SPACE (1.2)

OUTCOME

An "outcome" is an observed result of interest from a random experiment.

eg the number rolled after rolling a die.

SAMPLE SPACE

The "sample space" of a random experiment is the set of all possible distinct outcomes of said experiment.

eg when rolling a 6-sided die:

$$S = \{1, 2, 3, 4, 5, 6\}$$

EVENTS

An "event" of a random experiment is a group or set of outcomes of said experiment; ie subsets of the sample space.

There are two types of events:

① Simple events - consist of one outcome

eg rolling a 1 on a 6-sided die:

$$S = \{1, \dots, 6\} \quad E = \{1\}$$

② Compound events - consist of multiple outcomes

eg rolling an odd number on a 6-sided die:

$$S = \{1, \dots, 6\} \quad E = \{1, 3, 5\}$$

Note that

① Two simple events will never occur simultaneously; eg can never roll a 1 & 3 at the same time with one die.

② A compound event occurs if and only if one of its simple events occurs; and

eg odd # rolled \Leftrightarrow 1 rolled or 3 rolled or 5 rolled (on a 6-sided die)

③ Two compound events can occur simultaneously.

eg 3 rolled \Rightarrow {odd number rolled ($E = \{1, 3, 5\}$) and multiple of 3 rolled ($E = \{3, 6\}$)}

DEFINITIONS OF PROBABILITY (1.3)

💡 "Probability" is a quantitative measure of how likely an event is to occur.

CLASSICAL DEFINITION

💡 The "classical definition" of probability states that each distinct outcome in the sample space is equally likely to occur.

💡 In this case, the probability of an event E is equal to

eg roll a 6-sided die once.

E = number is odd.

$$\Rightarrow E = \{1, 3, 5\}, \quad S = \{1, 2, 3, 4, 5, 6\}.$$

$$\text{So } P(E) = \frac{3}{6} = \frac{1}{2}.$$

RELATIVE FREQUENCY DEFINITION

💡 The "relative frequency" definition of probability states that the probability of an event occurring is the proportion it occurs in a very long series of repetitions of the experiment.

eg rolling a 6-sided die 300 times

\Rightarrow 3 shows up 49 of those 300 times

\Rightarrow so $P(\text{die}=3) \approx \frac{49}{300} \approx \frac{1}{6}$.

SUBJECTIVE PROBABILITY DEFINITION

💡 In the "subjective probability" definition of probability, the probability of an event is determined by an opinion (ie what a person thinks the probability is).

eg the probability of COVID-19 being eradicated by 2022.

💡 Note that this plays a role in fields like "Bayesian Statistics".

DISCRETE PROBABILITY MODELS (1.4)

💡 In discrete probability models:

- ① The sample space S satisfies $|S| \leq |\mathbb{N}|$; ie there are either a finite or countably infinite number of basic events; and
- ② Each probability p_i satisfies $0 \leq p_i \leq 1$; and
- ③ The probabilities of each basic event sum to 1; ie $\sum p_i = 1$.

CLASSIC DISCRETE MODELS (1.5)

💡 In classic discrete models:

- ① The sample space S satisfies $|S| < |\mathbb{N}|$ (ie it is finite); and
- ② All basic events are equally likely to occur;
ie $P(a_1) = \dots = P(a_{|S|}) = \frac{1}{|S|}$.

Chapter 2: Counting Techniques

FULL FACTORIAL: $n!$ (2.1)

💡 The factorial of n , denoted as " $n!$ " and defined to be

$$n! = n(n-1) \dots 1$$

is the number of ways to put n distinguishable objects in a row.

COMBINATIONS: C_n^r OR nC_r (2.2)

💡 " n choose r ", denoted as " C_n^r " or " nC_r ", defined to be

$$C_n^r = \frac{n!}{r!(n-r)!} = \frac{n(n-1) \dots (n-(r-1))}{r!}$$

is the number of ways to select r objects from n distinguishable objects.

PERMUTATIONS: P_n^r OR nP_r (2.3)

💡 " P_n^r " or " nP_r ", defined to be

$$P_n^r = \frac{n!}{(n-r)!} = n(n-1) \dots (n-(r-1)) = C_r^n \cdot r!$$

is the number of ways to select r objects from n distinguishable objects and put them in a row.

GENERALIZATION OF COMBINATIONS (2.4)

💡 we can show the number of ways to arrange n objects in a row, where n_1 objects are of type 1, n_2 objects are of type 2, ..., n_k objects are of type k , where $n_1 + n_2 + \dots + n_k = n$, is

$$\# \text{ of outcomes} = \frac{n!}{n_1! \dots n_k!} = C_n^{n_1} C_{n-n_1}^{n_2} C_{n-n_1-n_2}^{n_3} \dots C_{n_{k-1}+n_k}^{n_k} C_{n_k}^{n_k}$$

eg Roll a die 4 times. Find $P(\text{the sum}=10)$.

Soln. This is equivalent to distributing 10 balls into 4 sections, where each section has at least 1 ball.



9 different spaces for the "dividers", 4 "dividers"

$\Rightarrow C_9^4$ ways of "positioning" the dividers.

But, we exclude the option where one of the sections has 7 balls, ie

Hence $P(\text{event}) = \frac{C_9^4 - 4}{6^4}$, since there are 6^4 outcomes of rolling a 6 sided die twice. \blacksquare

STARS & BARS WITHOUT "EMPTY" SECTIONS

💡 Given n stars, the # of ways to divide them up into k sections with $k-1$ rods without one of the sections containing zero elements is

$$\# = C_{k-1}^{n-1}$$

eg $n=5, k=4$

$$\star | \star | \star | \star$$

STARS & BARS WITH "EMPTY" SECTIONS

💡 Given n stars, the # of ways to divide them up into k sections with $k-1$ rods with one (or more) sections containing zero elements is

$$\# = C_{n+k-1}^{k-1}$$

$$\star | | \star | \star | \star | \star$$

eg $n=5, k=4$

2nd section has no elements.

Chapter 3: Probability Rules

RELATIONS AMONGST EVENTS

(3.1)

EVERY EVENT $\subseteq S$ (THE "CERTAIN" EVENT)

Let A be an event.

Then necessarily

$A \subseteq S = \{\text{the event that always occurs}\}$.

\emptyset (THE "IMPOSSIBLE" EVENT)

We use " \emptyset " to denote the event that never occurs.

UNION OF EVENTS: $A \cup B$

Let A, B be events.

Then " $A \cup B$ " is the event that at least one of the two occurs.



INTERSECTION OF EVENTS: $A \cap B$

Let A, B be events.

Then, " $A \cap B$ " is the event that both A & B occur.

We also denote $A \cap B = AB$.



MUTUALLY EXCLUSIVE / DISJOINT

Let A, B be events.

Then, we say A & B are "mutually exclusive" (or "disjoint") if $A \cap B = \emptyset$.

INCLUSION: $A \subseteq B$

Let A, B be events.

Then, we say " $A \subseteq B$ " if B occurs whenever A occurs; ie

A occurs $\Rightarrow B$ occurs.

COMPLEMENT: $A^c = \bar{A}$

Let A be an event.

Then, \bar{A} is the event such that \bar{A} occurs $\Leftrightarrow A$ does not occur.

PARTITION OF S

Let B_1, \dots, B_n be events.

Then, we say B_1, \dots, B_n form a "partition" of S if

$$B_1 \cup \dots \cup B_n = S \quad \& \quad B_i \cap B_j = \emptyset \quad \forall i \neq j.$$

PROBABILITY RULES (3.2)

A probability function $P: P(S) \rightarrow [0, 1]$ is any function that satisfies the following for any $A, B \subseteq S$:

- ① $P(\emptyset) = 0$;
- ② $P(S) = 1$;
- ③ $P(A) \geq 0 \quad \forall A \subseteq S$; (non-negativity)
- ④ $A \subseteq B \Rightarrow P(A) \leq P(B)$;
- ⑤ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$; & (addition law of probability)
- this generalizes to more variables as well.
- ⑥ $P(A^c) = 1 - P(A)$.

Chapter 4: Conditional Probability and Event Independence

CONDITIONAL PROBABILITY (4.1)

💡 Let A, B be events.
Then, the probability that A happens given B already happens, denoted as " $P(A|B)$ ", is equal to

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

* note $P(B) \neq 0$ necessarily.

INDEPENDENCE [OF TWO EVENTS] (4.2)

💡 Let A, B be events.
Then, we say A & B are "independent" if and only if

$$P(A \cap B) = P(A)P(B).$$

💡 Note that if $P(A), P(B) \neq 0$, then A & B cannot be mutually exclusive (ie $P(A \cap B) = 0$) if they are independent.

💡 If A & B are independent, then

- ① A & B^c are independent;
- ② A^c & B are independent; and
- ③ A^c & B^c are independent.

💡 Note that independence arises from independent random events.

INDEPENDENCE [OF > 2 EVENTS] (4.3)

💡 Let A_1, \dots, A_n be n events.
Then, we say A_1, \dots, A_n are (mutually) independent if

$$P(A_{n_1} A_{n_2} \dots A_{n_k}) = P(A_{n_1}) \dots P(A_{n_k}) \quad \forall \{n_1, \dots, n_k\} \in \binom{\{1, \dots, n\}}{k}.$$

💡 For the $n=3$ case, A_1, A_2 & A_3 are independent if

- ① $P(A_1 A_2) = P(A_1)P(A_2);$
- ② $P(A_1 A_3) = P(A_1)P(A_3);$
- ③ $P(A_2 A_3) = P(A_2)P(A_3);$ and
- ④ $P(A_1 A_2 A_3) = P(A_1)P(A_2)P(A_3).$

A_1, \dots, A_n ARE INDEPENDENT $\Rightarrow P(\prod_{i=1}^n A_i) = \prod_{i=1}^n P(A_i | A_1 \dots A_{i-1})$

(THE MULTIPLICATION FORMULA) (4.4.1)

💡 Let A_1, \dots, A_n be independent events.

Then necessarily

$$P(A_1 \dots A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \dots P(A_n | A_1 \dots A_{n-1}).$$

Proof. Note that for any $k=1, \dots, n$, we have

$$P(A_k | A_1 \dots A_{k-1}) = \frac{P(A_1 \dots A_{k-1} A_k)}{P(A_1 \dots A_{k-1})} = P(A_k).$$

The proof follows trivially. \square

A_1, \dots, A_n PARTITION $S \Rightarrow P(B) = \sum_{i=1}^{\infty} P(A_i)P(B | A_i)$

(TOTAL PROBABILITY FORMULA) (4.4.2)

💡 Let A_1, A_2, \dots form a partition of S , ie we have that

$$A_i A_j = \emptyset \quad \forall i \neq j \quad \text{and} \quad \bigcup_{i=1}^{\infty} A_i = S.$$

Let B be an event. Then necessarily

$$P(B) = P(BS) = \sum_{i=1}^{\infty} P(A_i B) = \sum_{i=1}^{\infty} P(A_i)P(B | A_i).$$

* this also works for finite collections of events as well.

A_1, \dots, A_n PARTITION $S \Rightarrow$

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum_{j=1}^{\infty} P(A_j)P(B | A_j)}$$

(THE BAYES FORMULA) (4.4.3)

💡 Let A_1, A_2, \dots form a partition of S , and let B be such that $P(B) \neq 0$.

Then necessarily, for any $i \in \mathbb{N}$, we have that

$$P(A_i | B) = \frac{P(A_i B)}{P(B)} = \frac{P(A_i)P(B | A_i)}{P(B)} = \frac{\sum_{j=1}^{\infty} P(A_j)P(B | A_j)}{P(B)}.$$

* again, this also generalises to the finite case.

Chapter 5:

Discrete Random Variables and Probability Models

RANDOM VARIABLES (5.1)

RANDOM VARIABLE (RV) (5.1)

Let S be a sample space.

Then, a "random variable" is defined to be some $X: S \rightarrow \mathbb{R}$.

Note that we usually denote random variables by capital letters. (e.g. X, Y, Z , etc.)

DISCRETE [r.v.]

Let $X: S \rightarrow \mathbb{R}$ be a r.v.
Then, we say X is "discrete" if $\text{range}(X) \subseteq \mathbb{N}$.

PROBABILITY MASS FUNCTION (PMF)

Let $X: S \rightarrow \mathbb{R}$ be a r.v.

Then, the "probability mass function" (or pmf) of X is defined to be the function $f: \text{range}(X) \rightarrow [0, 1]$ by $f(x) = P[X=x] \quad \forall x \in \text{range}(X)$.

By construction of f , note that $\sum_{x \in \text{range}(X)} f(x) = 1$.

CUMULATIVE DISTRIBUTION FUNCTION (CDF)

Let $X: S \rightarrow \mathbb{R}$ be a r.v.

Then, the "cumulative distribution function" (or cdf) of X is defined to be the function $F: \mathbb{R} \rightarrow [0, 1]$ by $F(x) = P[X \leq x] \quad \forall x \in \mathbb{R}$.

Properties of cdf:
 ① $F(x_1) \leq F(x_2) \Leftrightarrow x_1 \leq x_2 \quad \forall x_1, x_2 \in \mathbb{R}$; and
 ② $\lim_{x \rightarrow -\infty} F(x) = 0$ & $\lim_{x \rightarrow \infty} F(x) = 1$.

PMF CAN BE OBTAINED BY CDF, AND VICE VERSA

Let $X: S \rightarrow \mathbb{R}$ be discrete.

Then, given the pmf f of X , we can obtain X 's cdf F , and vice versa.

Proof. Let $x \in \text{range}(X)$. See that $f(x) = P[X=x] = P[X \leq x] - P[X \leq x-\epsilon] = F(x) - F(x-\epsilon)$, where $\epsilon > 0$ is such that $\text{range}(X) \cap [x-\epsilon, x] = \{x\}$. (Since X is discrete, such an ϵ will exist.)

FINDING PMF (E1)

Let X be the number of heads after flipping a fair coin n times.

Find the pmf of X .

Solⁿ. See that $\text{range}(X) = \{0, 1, \dots, n\}$.

Then

$$P[X=k] = C_n^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = C_n^k \left(\frac{1}{2}\right)^n$$

and so the pmf of X is $f: \{0, \dots, n\} \rightarrow [0, 1]$ given by

$$f(k) = P[X=k] = C_n^k \left(\frac{1}{2}\right)^n \quad \forall k=0, \dots, n.$$

BERNOULLI TRIALS & RELATED RV (5.2)

BERNOULLI TRIALS (5.2.1)

A "Bernoulli trial" focuses on a particular random experiment with only two possible outcomes: success or failure.

We call the random variables and the experiment obtained from Bernoulli trials as "Bernoulli random variables" and a "Bernoulli experiment" respectively.

BERNOULLI RV (5.2.2)

In particular, if B is a Bernoulli rv:

① then $P[B=\text{Success}]$, or $P(B)$, is equal to $P(B) = p$ (where p = probability of success); and

② $P[B=\text{Failure}]$, or $P(B^c)$, is equal to $P(B^c) = 1-p$.

Thus, the pmf of B is

$$f: \{0, 1\} \rightarrow [0, 1] \text{ by } f(0) = 1-p \text{ & } f(1) = p,$$

or equivalently by

$$f(x) = p^x (1-p)^{1-x} \quad \forall x \in \{0, 1\}.$$

BERNOULLI SEQUENCE (5.2.3)

A "Bernoulli sequence" occurs when

- ① we repeat a Bernoulli trial many times;
- ② the results are all independent; and
- ③ the success probability p stays the same.

BINOMIAL DISTRIBUTION: $X \sim \text{Binomial}(n, p)$ / $X \sim \text{Bin}(n, p)$ (5.2.4)

Let X be the rv equal to the number of successes after repeating a Bernoulli trial n times independently, with probability of success p .

Then, we say X follows a binomial distribution, and write $X \sim \text{Binomial}(n, p)$.

In this case, the pmf of X is equal to

$$f: \{0, \dots, n\} \rightarrow [0, 1] \text{ by } f(k) = C_n^k p^k (1-p)^{n-k} \quad \forall k \in \mathbb{N}.$$

GEOMETRIC DISTRIBUTION: $X \sim \text{Geometric}(p)$ / $X \sim \text{Geo}(p)$ (5.2.5)

Repeat independent Bernoulli trials, with success probability p , until a trial is successful.

Let the rv X be equal to the number of failures before the success was reached.

Then, we say X follows a geometric distribution, and write $X \sim \text{Geometric}(p)$.

In this case, the pmf of X is equal to

$$f: \mathbb{N} \rightarrow [0, 1] \text{ by } f(k) = (1-p)^k p \quad \forall k \in \mathbb{N}.$$

Note that $P(X \geq n) = (1-p)^n$ $\forall n \in \mathbb{N}$; and

$$\text{Proof. } P(X \geq n) = \sum_{k=n}^{\infty} (1-p)^k p = (1-p)^n p \sum_{k=0}^{\infty} (1-p)^k = (1-p)^n p \left(\frac{1}{1-(1-p)}\right) = (1-p)^n.$$

② $P(X \geq m+n | X \geq n) = P(X \geq m) \quad \forall m, n \in \mathbb{N}$ (the memory-less property).

$$\text{Proof. } P(X \geq m+n | X \geq n) = \frac{P(X \geq m+n \cap X \geq n)}{P(X \geq n)} = \frac{P(X \geq m+n)}{P(X \geq n)} = \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m = P(X \geq m). \quad \square$$

NEGATIVE BINOMIAL DISTRIBUTION:

$X \sim \text{Negative Binomial}(k, p) / X \sim \text{NB}(k, p)$ (5.2.6)

- Repeat independent Bernoulli trials, with success probability p , until the k^{th} success is reached.
- Let the rv X be the number of failures before the k^{th} success.

Then, we say X follows a negative binomial distribution, and write $X \sim \text{Negative Binomial}(k, p)$.

In this case, the pmf of X is equal to

$$f: N \rightarrow [0, 1] \text{ by } f(n) = C_{n+k-1}^n p^k (1-p)^n \quad \forall n \in N.$$

Proof. See that

$$\begin{aligned} P[X=n] &= P[\text{having } n \text{ failures before } k^{\text{th}} \text{ success}] \\ &= P[n \text{ failures \& } k-1 \text{ successes, followed by } k^{\text{th}} \text{ success}] \\ &= \frac{(n+k-1)!}{n!(k-1)!} (1-p)^n p^{k-1}. \\ \therefore P[X=n] &= C_{n+k-1}^n (1-p)^n p^k. \end{aligned}$$

HYPERGEOMETRIC DISTRIBUTION:

$X \sim \text{Hypergeometric}(N, M, n) / X \sim \text{Hyp}(N, M, n)$ (5.3)

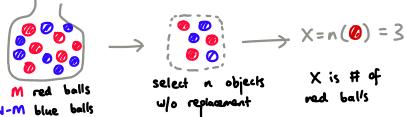
- Suppose we have a collection of N objects; M of one type, and $N-M$ of another (distinct) type.

Randomly select n objects without replacement, where $n \leq \min\{M, N-M\}$.

Let the rv X be the number of objects of the first type in these n objects.

Then, we say X follows a "hypergeometric distribution", and write

$$X \sim \text{Hypergeometric}(N, M, n).$$



In this case, the pmf of X is equal to

$$f: \{0, \dots, n\} \rightarrow [0, 1] \text{ by } f(k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad \forall k=0, \dots, n.$$

VANDERMONDE'S IDENTITY: $\sum_{k=0}^n \binom{M}{k} \binom{N-M}{n-k} = \binom{N}{n}$

Let $n \leq M, N-M$.

Then necessarily

$$\binom{N}{n} = \sum_{k=0}^n \binom{M}{k} \binom{N-M}{n-k}.$$

POISSON DISTRIBUTION:

$X \sim \text{Poisson}(\lambda) / X \sim \text{Poi}(\lambda)$ (5.4)

- In some observational studies, events happen over time or space.

We say such an event follows a Poisson process if the following conditions are satisfied:

- Events in non-overlapping time intervals are independent; $\left\{ \text{independence} \right\}$
- $P[\geq 2 \text{ events in } [t, t+\Delta t]] = o(\Delta t)$, where $\left\{ \text{individuality} \right\}$
- $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$ and $\Delta t \ll t$; and
- $P[\text{one event in } [t, t+\Delta t]] = \lambda \Delta t + o(\Delta t)$, $\lambda \in \mathbb{R}$. $\left\{ \text{homogeneity} \right\}$

Note that we call " λ " in ③ the "intensity parameter".

- Let the rv X be the number of events in $[0, t]$.

Then we say X follows a Poisson distribution, and write

$$X \sim \text{Poisson}(\lambda).$$

In this case, the pmf of X is given by

$$f: N \rightarrow [0, 1] \text{ by } f(x) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}.$$

Proof. First, divide $[0, t]$ into n small intervals:

$$\Delta t, \frac{t}{n}, \dots, \frac{t}{n}$$

Note that $\Delta t \rightarrow 0$ as $n \rightarrow \infty$.

Let the events

$$\begin{aligned} B_1^{(n,x)} &= \text{there are } x \text{ small intervals each with one event;} \\ B_2^{(n)} &= \geq 1 \text{ small interval exists with two or more events.} \end{aligned}$$

Then, see that

$$\begin{aligned} P(B_1^{(n,x)}) &= \binom{n}{x} (P[\text{one event in interval of length } \frac{\Delta t}{n} = \Delta t]) (1-p)^{n-x} \\ &\quad \text{(by binomial distn)} \\ &= \binom{n}{x} (\lambda \frac{\Delta t}{n} + o(\frac{\Delta t}{n}))^x (1 - \lambda \frac{\Delta t}{n} - o(\frac{\Delta t}{n}))^{n-x}. \quad \text{(by point ② of defn)} \end{aligned}$$

Notice that since we want to consider infinitely small periods of time for our Poisson variable, we can deduce that

$$\begin{aligned} P(X=x) &= \lim_{n \rightarrow \infty} P(B_1^{(n,x)}) \\ &= \lim_{n \rightarrow \infty} \left[\binom{n}{x} \left(\lambda \frac{\Delta t}{n} + o(\frac{\Delta t}{n}) \right)^x \left(1 - \lambda \frac{\Delta t}{n} - o(\frac{\Delta t}{n}) \right)^{n-x} \right] \\ &= \lim_{n \rightarrow \infty} \left[\frac{n!}{x!(n-x)!} \left(\lambda \frac{\Delta t}{n} + o(\frac{\Delta t}{n}) \right)^x \left(1 - \lambda \frac{\Delta t}{n} - o(\frac{\Delta t}{n}) \right)^{n-x} \right] \\ &= \lim_{n \rightarrow \infty} \left[\frac{1}{x!} \frac{n(n-1)\dots(n-x+1)}{n^x} \left(\lambda t + no(\frac{\Delta t}{n}) \right)^x \left(1 - \lambda \frac{\Delta t}{n} - o(\frac{\Delta t}{n}) \right)^n \cdot \right. \\ &\quad \left. \left(1 - \lambda \frac{\Delta t}{n} - o(\frac{\Delta t}{n}) \right)^{-x} \right] \\ &= \frac{1}{x!} (1)(\lambda t)^x \lim_{n \rightarrow \infty} \left(1 - \lambda \frac{\Delta t}{n} \right)^n (1) \\ &= \frac{1}{x!} (\lambda t)^x e^{-\lambda t} \quad \text{(using the identity } e^x = \lim_{n \rightarrow \infty} (1 + \frac{x}{n})^n \text{)} \\ \therefore P(X=x) &= \frac{e^{-\lambda t} (\lambda t)^x}{x!}, \end{aligned}$$

as needed \square

Chapter 6: Expectation and Variance

EXPECTED VALUE / EXPECTATION [OF A DISC RV]

(6.1)

Let X be a drv, with $\text{ran}(X) = A$ & pmf $f(x)$. Then, the "expectation" or "expected value" of X , denoted as " $E(X)$ ", is defined to be equal to

$$E(X) = \sum_{x \in A} x f(x). \quad (\text{Def})$$

Note to calculate expectations, we need to:

- ① Identify the rv X involved;
- ② Find the pmf of X ; and
- ③ Compute $E(X)$.

$$X \sim \text{Bernoulli}(p) \Rightarrow E(X) = p \quad (6.2.1)$$

Let $X \sim \text{Bernoulli}(p)$. Then necessarily $E(X) = p$.

$$\begin{aligned} \text{Proof. } E(X) &= \sum_{x \in \{0,1\}} x P(X=x) \\ &= 0P(X=0) + 1P(X=1) \\ &= p. \quad \blacksquare \end{aligned}$$

$$X \sim \text{Binomial}(n, p) \Rightarrow E(X) = np \quad (6.2.2)$$

Let $X \sim \text{Binomial}(n, p)$. Then necessarily $E(X) = np$.

$$\begin{aligned} \text{Proof. } E(X) &= \sum_{k=0}^n k P[X=k] \\ &= \sum_{k=0}^n k \left(\binom{n}{k} p^k (1-p)^{n-k} \right) \\ &= \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{k=0}^n \frac{(n-1)!}{k!(n-k)!} p^k (1-p)^{n-k-1} \\ &= np (1) \quad (\text{by Bin formula}) \\ \therefore E(X) &= np. \quad \blacksquare \end{aligned}$$

$$X \sim \text{Geometric}(p) \Rightarrow E(X) = \frac{1-p}{p} \quad (6.2.3)$$

Let $X \sim \text{Geometric}(p)$. Then necessarily $E(X) = \frac{1-p}{p}$.

$$\begin{aligned} \text{Proof. } E(X) &= \sum_{k=0}^{\infty} k \cdot P(X=k) \\ &= \sum_{k=0}^{\infty} k (1-p)^k p. \\ &= p(1-p) \sum_{k=1}^{\infty} k (1-p)^{k-1}. \end{aligned}$$

Recall the identity $\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \dots + nx^{n-1}, \quad |x| < 1$

Since $|1-p| < 1$, thus

$$\frac{1}{p^2} = \frac{1}{(1-(1-p))^2} = 1 + 2(1-p) + 3(1-p)^2 + \dots = \sum_{k=1}^{\infty} k(1-p)^{k-1},$$

and so

$$E(X) = p(1-p) \left(\frac{1}{p^2} \right) = \frac{1-p}{p}. \quad \blacksquare$$

$$X \sim \text{NB}(k, p) \Rightarrow E(X) = \frac{k(1-p)}{p} \quad (6.2.4)$$

Let $X \sim \text{NB}(k, p)$. Then necessarily $E(X) = \frac{k(1-p)}{p}$.

$$\begin{aligned} \text{Proof. } E(X) &= \sum_{n=0}^{\infty} n \cdot P(X=n) \\ &= \sum_{n=1}^{\infty} n \left(\binom{n+k-1}{n} p^k (1-p)^n \right) \\ &= \sum_{n=1}^{\infty} n \frac{(n+k-1)!}{n! (k-1)!} (1-p)^n p^k \\ &= \sum_{n=1}^{\infty} K \frac{(1-p)^k}{p} \cdot \frac{((n-1)+(k-1)-1)!}{(n-1)! k!} (1-p)^{n-1} p^{k-1} \\ &= \frac{k(1-p)}{p} \sum_{n=1}^{\infty} \binom{(n-1)+(k-1)-1}{n-1} (1-p)^{n-1} p^{k-1} \\ &= \frac{k(1-p)}{p} \sum_{n=0}^{\infty} \binom{n+(k-1)-1}{n} (1-p)^n p^{k-1} \\ &= \frac{k(1-p)}{p} (1) \\ \therefore E(X) &= \frac{k(1-p)}{p}. \quad \blacksquare \end{aligned}$$

$$X \sim \text{Hyp}(N, M, n) \Rightarrow E(X) = \frac{nM}{N} \quad (6.2.5)$$

Let $X \sim \text{Hyp}(N, M, n)$. Then necessarily $E(X) = \frac{nM}{N}$.

$$\begin{aligned} \text{Proof. } E(X) &= \sum_{k=0}^n k \cdot P(X=k) \\ &= \sum_{k=1}^n k \cdot \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \\ &= \frac{n!(N-n)!}{N!} \sum_{k=1}^n k \cdot \frac{M!}{k!(n-k)!} \frac{(N-n)!}{(n-k)!(N-M-n+k)!} \\ &= \frac{M(n!) (N-n)!}{N!} \sum_{k=1}^n \frac{(M-1)!}{(k-1)! (n-k)!} \frac{(N-n)!}{(n-k)!(N-M-n+k)!} \\ &= \frac{M \cdot n!(N-n)!}{N!} \sum_{k=0}^n \frac{(M-1)!}{k!} \frac{(N-n)!}{(n-k-1)!} \\ &= M \frac{n!(N-n)!}{N!} \binom{N-1}{n-1} \quad (\text{by Vandermonde Identity}) \\ &= M \frac{n!(N-n)!}{N!} \cdot \frac{(N-1)!}{(n-1)!(N-n)!} \\ \therefore E(X) &= \frac{Mn}{N}. \quad \blacksquare \end{aligned}$$

$$X \sim \text{Po}(\mu) \Rightarrow E(X) = \mu \quad (6.2.6)$$

Let $X \sim \text{Po}(\mu)$. Then necessarily $E(X) = \mu$.

$$\begin{aligned} \text{Proof. } E(X) &= \sum_{n=0}^{\infty} n \frac{e^{-\mu} \mu^n}{n!} \\ &= \sum_{n=1}^{\infty} \frac{e^{-\mu} \mu^n}{(n-1)!} \\ &= \mu \sum_{n=0}^{\infty} \frac{e^{-\mu} \mu^n}{n!} \\ &= \mu (1) = \mu. \quad \blacksquare \end{aligned}$$

$$E[g(X)] = \sum_{x \in A} g(x) f(x) \quad ((\text{THE LAW OF THE UNCONSCIOUS STATISTICIAN})) \quad (6.3)$$

Let g be a function on the drv X , which has range A and pmf X .

$$\text{Then necessarily } E[g(X)] = \sum_{x \in A} g(x) f(x).$$

Proof. Let $Y = g(X)$, and let $D_Y = \{x \in X : g(x) = y\}$, and let $B = \text{ran}(Y)$.

Then

$$P[Y=y] = P[g(X)=y] = \sum_{x \in D_Y} P[X=x].$$

Hence

$$\begin{aligned} E(Y) &= \sum_{y \in B} y \cdot P[g(X)=y] \\ &= \sum_{y \in B} y \cdot \sum_{x \in D_Y} P[X=x] \\ &= \sum_{y \in B} \sum_{x \in D_Y} g(x) P[X=x] \\ \therefore E(Y) &= E[g(X)] = \sum_{x \in A} g(x) f(x). \quad \blacksquare \end{aligned}$$

$$E[ag(X) + b] = aE[g(X)] + b; \quad E[ag_1(X) + bg_2(X)] = aE[g_1(X)]$$

$$+ bE[g_2(X)] \quad ((\text{LINEAR PROPERTIES OF EXPECTATION}))$$

Let g, g_1, g_2 be functions on the drv X , and let $a, b \in \mathbb{R}^+$.

Then necessarily

$$\textcircled{1} \quad E[ag(X) + b] = aE[g(X)] + b; \quad \text{and}$$

$$\textcircled{2} \quad E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)].$$

Proof. Follows almost directly from LOUS. \blacksquare

VARIANCE (6.4)

Let X be a drv.

Then, the "variance" of X , denoted as " $\text{Var}(X)$ " or " σ^2 ", is defined to be equal to

$$\text{Var}(X) = E[(X - E[X])^2].$$

* we use " μ " to denote " $E(X)$ ", so

$$\text{Var}(X) = E[(X - \mu)^2].$$

\mathbb{E}_2 $\text{Var}(X)$ is used to measure the "variability" of a sample, ie how "concentrated" the data is.

\mathbb{E}_3 The "standard deviation" of X , denoted as " σ ", is defined to be

$$\sigma = \sqrt{\text{Var}(X)}.$$

$$\text{Var}(X) = \sum_{x \in A} (x - E(X))^2 f(x) = E(X^2) - [E(X)]^2$$

\mathbb{E} (let X be a drv, with $\text{ran}(X) = A$ & pmf f .

Then necessarily

$$\text{① } \text{Var}(X) = \sum_{x \in A} (x - E(X))^2 f(x); \text{ and}$$

$$\text{② } \text{Var}(X) = E(X^2) - [E(X)]^2.$$

Proof. By defn,

$$\text{Var}(X) = E[(X - E(X))^2]$$

$$= \sum_{x \in A} (x - E(X))^2 f(x) \quad (\text{by LOUS}),$$

Showing ①.

$$\text{Then } \sum_{x \in A} (x - E(X))^2 f(x) = \sum_{x \in A} (x^2 - 2xE(X) + E(X)^2) f(x),$$

$$= \sum_{x \in A} x^2 f(x) - 2E(X) \sum_{x \in A} x + E(X)^2$$

$$= \sum_{x \in A} x^2 f(x) - 2E(X) E(X) + E(X) E(X)$$

$$= E(X^2) - [E(X)]^2,$$

showing ②. \square

PROPERTIES OF VARIANCE

\mathbb{E} (let X be a drv. Note the following:

$$\text{① } \text{Var}(X) \geq 0;$$

$$\text{② } E(X^2) \geq (E(X))^2;$$

$$\text{③ } \text{Var}(X) = 0 \Leftrightarrow P(X=c) = 1 \text{ for some constant } c; \text{ and}$$

$$\text{④ } \text{Var}(aX+b) = a^2 \text{Var}(X).$$

Proof. By defn of variance,

$$\text{Var}(X) = \sum_{x \in A} (x - E(X))^2 f(x) = \sum_{x \in A} (x - E(X))^2 P(X=x),$$

and as $P[X=x]$, $(x - E(X))^2 \geq 0 \quad \forall x \in A$. ① follows.

Thus

$$E(X^2) - (E(X))^2 \geq 0 \Rightarrow E(X^2) \geq (E(X))^2, \text{ showing ②.}$$

Next,

$$\text{Var}(X) = 0 \Leftrightarrow \sum_{x \in A} \frac{(x - \mu)^2}{\geq 0} f(x) = 0$$

$$\Leftrightarrow (x - \mu)^2 = 0 \quad \forall x \in A$$

$$\Leftrightarrow x = \mu \quad \forall x \in A \quad (\text{ie } A = \{\mu\})$$

$$\Leftrightarrow P[X=\mu] = 1, \text{ showing ③;}$$

and

$$\text{Var}(aX+b) = E[(aX+b)^2] - (E(aX+b))^2$$

$$= a^2 E(X^2) + 2ab E(X) + b^2 - (aE(X) + b)^2$$

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

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

$$= a^2 \text{Var}(X),$$

showing ④. \square

$$X \sim \text{Bernoulli}(p) \Rightarrow \text{Var}(X) = p(1-p) \quad (6.5.1)$$

\mathbb{E} (let $X \sim \text{Bernoulli}(p)$. Then necessarily $\text{Var}(X) = p(1-p)$.

$$\text{Proof. } \text{Var}(X) = E(X^2) - (E(X))^2$$

$$= \sum_{x=0}^1 x^2 p[X=x] - p^2$$

$$= p - p^2$$

$$\therefore \text{Var}(X) = p(1-p). \quad \square$$

$$X \sim \text{Binomial}(n, p) \Rightarrow \text{Var}(X) = np(1-p) \quad (6.5.2)$$

\mathbb{E} (let $X \sim \text{Binomial}(n, p)$. Then necessarily $\text{Var}(X) = np(1-p)$.

Proof. First, see that

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

$$= E(X(X-1)) - (E(X))^2 + E(X).$$

Then,

$$E(X(X-1)) = \sum_{k=0}^n k(k-1) p[X=k]$$

$$= \sum_{k=2}^n k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=2}^n n(n-1) \cdot \frac{(n-2)!}{(n-2)!(n-k)!} p^k (1-p)^{n-k}$$

$$= n(n-1)p \sum_{k=0}^{n-2} \frac{(n-2)!}{k!(n-k-2)!} p^k (1-p)^{n-2-k}$$

$$= n(n-1)p^2 (1)$$

$$\therefore E(X(X-1)) = n(n-1)p^2.$$

Thus

$$\text{Var}(X) = E(X(X-1)) - (E(X))^2 + E(X)$$

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

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

$$\therefore \text{Var}(X) = np(1-p). \quad \square$$

$$X \sim \text{Geometric}(p) \Rightarrow \text{Var}(X) = \frac{(1-p)^2}{p^2} + \frac{1-p}{p} \quad (6.5.3)$$

\mathbb{E} (let $X \sim \text{Geometric}(p)$. Then necessarily $\text{Var}(X) = \frac{(1-p)^2}{p^2} + \frac{1-p}{p}$.

Proof. See that

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

$$= E(X(X-1)) - (E(X))^2 + E(X).$$

Then

$$E(X(X-1)) = \sum_{n=0}^{\infty} n(n-1) \cdot P[X=n]$$

$$= \sum_{n=2}^{\infty} n(n-1) \cdot (1-p)^{n-2} p^n$$

$$= p(1-p)^2 [(1 \times 2) + (2 \times 3)(1-p) + (3 \times 4)(1-p)^2 + \dots]$$

$$= 2p(1-p)^2 [1 + 3(1-p) + 6(1-p)^2 + \dots]$$

$$= 2p(1-p)^2 \left(\frac{1}{1-(1-p)^2} \right)$$

$$\therefore E(X(X-1)) = \frac{2(1-p)^2}{p^2}.$$

Thus

$$\text{Var}(X) = E(X(X-1)) - (E(X))^2 + E(X)$$

$$= \frac{2(1-p)^2}{p^2} - \left(\frac{1-p}{p} \right)^2 + \left(\frac{1-p}{p} \right)$$

$$= \left(\frac{1-p}{p} \right)^2 + \left(\frac{1-p}{p} \right). \quad \square$$

$$X \sim \text{Poisson}(\mu) \Rightarrow \text{Var}(X) = \mu \quad (6.5.4)$$

\mathbb{E} (let $X \sim \text{Poisson}(\mu)$. Then necessarily $\text{Var}(X) = \mu$.

Proof. $\text{Var}(X) = E(X(X-1)) - (E(X))^2 + E(X).$

Then

$$E(X(X-1)) = \sum_{n=0}^{\infty} n(n-1) P[X=n]$$

$$= \sum_{n=2}^{\infty} n(n-1) \frac{e^{-\mu} \mu^n}{n!}$$

$$= \mu \sum_{n=2}^{\infty} \frac{e^{-\mu} \mu^{n-2}}{(n-2)!}$$

$$= \mu^2 \sum_{n=0}^{\infty} \frac{e^{-\mu} \mu^n}{n!}$$

$$= \mu^2 (1) = \mu^2.$$

$$\text{Hence } \text{Var}(X) = E(X(X-1)) - (E(X))^2 + E(X)$$

$$= \mu^2 - (\mu)^2 + \mu$$

$$= \mu. \quad \square$$

Chapter 7:

Discrete Multivariate Distributions

BIVARIATE DISTRIBUTIONS (7.1)

I: "Bivariate distributions" are probability distributions that deal with two random variables.

JOINT PMF: $(x, y) \sim f(x, y)$

I: Let x, y be drvs. Then, the "joint probability mass function", ie the "joint pmf", of X & Y is the function f defined by

I: In this case, we write

$$(x, y) \sim f(x, y).$$

$$f(x, y) : \text{ran}(X) \times \text{ran}(Y) \rightarrow [0, 1] \text{ by } f(x, y) = P[X=x, Y=y].$$

I: Properties of joint pmf:

- ① $f(x, y) \geq 0$ (by defn of f); and
- ② $\sum_y \sum_x f(x, y) = 1$.

I: In general, for drvs X_1, \dots, X_n , the joint pmf of X_1, \dots, X_n is defined by

$$f: \prod_{i=1}^n \text{ran}(X_i) \rightarrow [0, 1] \text{ by } f(x_1, \dots, x_n) = P[X_1=x_1, \dots, X_n=x_n].$$

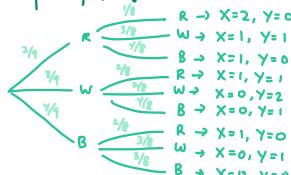
eg: A box contains 2 red, 3 white & 4 black ones. Randomly select 2 balls w/o replacement.

(Let

$$X = \# \text{ of red balls selected}; \quad Y = \# \text{ of white balls selected}.$$

Find the joint pmf of X & Y .

Soln.



Let $f(x, y)$ be the joint pmf of f .

By adding the probabilities for each case, you eventually get that

	0	1	2
0	2/36	8/36	1/36
1	12/36	6/36	0
2	3/36	0	0

MARGINAL PMF: $f_X(x)$, $f_Y(y)$ (7.1.1)

I: Let f be the joint pmf of some drvs X & Y .

Then, the "marginal pmfs" of X & Y , denoted as " f_X " & " f_Y " respectively, is defined to be equal to

$$f_X(x) = P[X=x] = \sum_y P[X=x, Y=y] = \sum_y f(x, y),$$

and

$$f_Y(y) = P[Y=y] = \sum_x P[X=x, Y=y] = \sum_x f(x, y).$$

* we also denote

$$f_X := f_1 \quad \& \quad f_Y := f_2.$$

I: In general, for drvs X_1, \dots, X_n , we have that

$$f_{X_i}(x_i) = P[X_i=x_i] = \sum_{x_1} \dots \sum_{x_{i-1}} \sum_{x_{i+1}} \dots \sum_{x_n} P[X_1=x_1, \dots, X_n=x_n]$$

$$= \sum_{x_1} \dots \sum_{x_{i-1}} \sum_{x_{i+1}} \dots \sum_{x_n} f(x_1, \dots, x_n).$$

* we also denote

$$f_{X_i} := f_i.$$

Note that

- ① $f(x, y)$ determines $f_X(x)$ & $f_Y(y)$; but
- ② We cannot generally find $f(x, y)$ from $f_X(x)$ & $f_Y(y)$.

CONDITIONAL PMF: $f_1(x|y)$ (7.1.2)

I: Let X & Y be drvs.

Then, the "conditional pmf of X given $Y=y$ ", denoted as " $f_1(x|y)$ ", is defined to be

$$f_1(x|y) = \frac{f(x, y)}{f_2(y)} = \frac{P[X=x, Y=y]}{P[Y=y]},$$

given that $f_2(y) > 0$.

Note that:

$$\text{① } f_1(x|y) \geq 0 \quad (\text{as } f(x, y) \geq 0, f_2(y) > 0); \quad \&$$

$$\text{② } \sum_x f_1(x|y) = 1 \quad \forall y \in \text{ran}(Y).$$

$$\text{Proof: } \sum_x f_1(x|y) = \sum_x \frac{f(x, y)}{f_2(y)}$$

$$= \frac{1}{f_2(y)} \sum_x f(x, y)$$

$$= \frac{1}{f_2(y)} f_2(y) \quad (\text{by defn})$$

$$= 1. \quad \blacksquare$$

INDEPENDENT [RANDOM VARIABLES] (7.1.3)

I: Let X & Y be rv, with pmf f .

Then, we say X and Y are "independent" if

$$f(x, y) = P[X=x, Y=y] = P[X=x]P[Y=y] = f_X(x)f_Y(y)$$

for each $x \in \text{ran}(X)$, $y \in \text{ran}(Y)$.

I: So, to show X & Y are not independent, it suffices to find some $x \in \text{ran}(X)$, $y \in \text{ran}(Y)$ such that

$$f(x, y) \neq f_X(x)f_Y(y).$$

eg: $X \sim \text{Geo}(p)$; $Y = X^2$: Show X & Y are not independent.

→ Let $x=0$, $y=1$.

$$\text{Then } f(0, 1) = P[X=0, Y=1] = P[X=0, X^2=1]$$

= 0 (since this is impossible).

$$\text{But } f_X(0)f_Y(1) = P[X=0]P(Y=1)$$

$$= P[X=0]P(X^2=1)$$

$$= P[X=0]P(X=1)$$

$$= p(1-p).p \quad (\neq 0). \quad \blacksquare$$

I: In general, the rv X_1, \dots, X_n (with pmf f) are independent if

$$f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n) \quad \forall x_i \in \text{ran}(X_i), 1 \leq i \leq n.$$

DISTRIBUTION OF A FUNCTION OF RANDOM VARIABLES (7.2)

To find the pmf for $T = g(X, Y)$, we use the following method:

- ① Evaluate $\text{ran}(T) = \text{ran}(g(X, Y))$;
- ② Find the values of (x, y) such that $g(x, y) |_{X=x, Y=y} = t$ for each $t \in \text{ran}(T)$; then
- ③ Use $f_{X,Y}(x, y)$ to get the respective probabilities of each (x, y) , and "merge" them accordingly under each $t \in \text{ran}(T)$ to obtain the pmf for T .

e.g. A box contains 2 red, 3 white & 4 black ones. Randomly select 2 balls w/o replacement.

Let
 $X = \# \text{ of red balls selected}$;
 $Y = \# \text{ of white balls selected}$.

let $T = 2XY - 1$. Find the pmf of T . (E3)

Soln.

	X		
T	0	1	2
0	-1	-1	-1
1	-1	1	3
2	-1	3	7

From earlier we found the pmf of X, Y in a tabular form:

	x		
y	0	1	2
0	6/36	8/36	1/36
1	12/36	6/36	0
2	3/36	0	0

By "comparing" the two tables, see that

$$\begin{aligned} P[T=-1] &= P(X=0, Y=0) + P(X=0, Y=1) + P(X=0, Y=2) \\ &\quad + P(X=1, Y=0) + P(X=1, Y=2) \\ &= 6/36 + 8/36 + 1/36 + 12/36 + 3/36 = 30/36. \end{aligned}$$

$P[T=1]$, $P[T=3]$ & $P[T=7]$ are calculated similarly.

$X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$, X, Y INDEPENDENT \Rightarrow

$X+Y \sim \text{Bin}(n+m, p)$ (E4)

Let $X \sim \text{Bin}(n, p)$ be independent from $Y \sim \text{Bin}(m, p)$. We can show $X+Y \sim \text{Bin}(n+m, p)$ by a pmf argument.

Proof. Let f be the joint pmf of X & Y .

Since X, Y are indep., thus $f_{X,Y}(x, y) = f_X(x)f_Y(y) \forall x, y$.

Let $T = X+Y$. See that

$$\begin{aligned} P[T=t] &= \sum_{x=0}^t P(X=x)P(Y=t-x) \\ &= \sum_{x=0}^t \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{t-x} p^{t-x} (1-p)^{m-(t-x)} \\ &= \sum_{x=0}^t \binom{n}{x} \binom{m}{t-x} p^t (1-p)^{n+m-t+x} \\ &= p^t (1-p)^{n+m-t} \sum_{x=0}^t \binom{n}{x} \binom{m}{t-x} \\ &= p^t (1-p)^{n+m-t} \binom{n+m}{t} \quad (\text{by Vandermonde's identity}), \end{aligned}$$

and so this suffices to show

$$T \sim \text{Bin}(n+m, p).$$

Let's find the conditional pmf of X given $T=t$ also.

Soln. See that

$$\begin{aligned} P[X=x | T=t] &= \frac{P(X=x, T=t)}{P(T=t)} \\ &= \frac{P(X=x, X+Y=t)}{P(T=t)} \\ &= \frac{P(X=x, Y=t-x)}{P(T=t)} \\ &= \frac{P_X(x) f_Y(t-x)}{P(T=t)} \\ &= \frac{\binom{n}{x} p^x (1-p)^{n-x} \binom{m}{t-x} p^{t-x} (1-p)^{m-(t-x)}}{P(T=t)} \\ &= \frac{\binom{n}{x} p^x (1-p)^{n-x} \binom{m}{t-x} p^{t-x} (1-p)^{m-(t-x)}}{\binom{n+m}{t} p^t (1-p)^{n+m-t}} \end{aligned}$$

$$\therefore P[X=x | T=t] = \frac{\binom{n}{x} \binom{m}{t-x}}{\binom{n+m}{t}}.$$

\Rightarrow see that by construction, this is what a hypergeometric distribution is "finding"!

$X \sim \text{Po}(\mu_1)$, $Y \sim \text{Po}(\mu_2) \Rightarrow X+Y \sim \text{Po}(\mu_1+\mu_2)$ (E5)

Let $X \sim \text{Po}(\mu_1)$ be independent to $Y \sim \text{Po}(\mu_2)$. Then we can similarly show that $X+Y \sim \text{Po}(\mu_1+\mu_2)$.

Proof. Let $T = X+Y$.

$$\begin{aligned} \text{See that } P[T=t] &= \sum_{x=0}^t P[X=x, Y=t-x] \\ &= \sum_{x=0}^t P[X=x] P[Y=t-x] \quad (\text{as } X, Y \text{ are independent}) \\ &= \sum_{x=0}^t \frac{e^{-\mu_1} \mu_1^x}{x!} \frac{e^{-\mu_2} \mu_2^{t-x}}{(t-x)!} \\ &= e^{-\mu_1-\mu_2} \sum_{x=0}^t \frac{\mu_1^x \mu_2^{t-x}}{x!(t-x)!} \\ &= \frac{e^{-\mu_1-\mu_2}}{t!} \sum_{x=0}^t \frac{t!}{x!(t-x)!} \mu_1^x \mu_2^{t-x} \\ &= \frac{e^{-\mu_1-\mu_2}}{t!} \sum_{x=0}^t \binom{t}{x} \mu_1^x \mu_2^{t-x} \\ &= \frac{e^{-\mu_1-\mu_2}}{t!} (\mu_1+\mu_2)^t \quad (\text{by the binomial formula}) \end{aligned}$$

which suffices to show $T \sim \text{Po}(\mu_1+\mu_2)$.

We can similarly find $f_X(x|t)$.

$$\begin{aligned} f_X(x|t) &= \frac{P[X=x, T=t]}{P(T=t)} \\ &= \frac{P(X=x, T=t)}{P(T=t)} \\ &= \frac{P(X=x, X+Y=t)}{P(T=t)} \\ &= \frac{P(X=x, Y=t-x)}{P(T=t)} \\ &= \frac{P(X=x) P(Y=t-x)}{P(T=t)} \quad (\text{by independence of } X, Y) \\ &= \frac{(\frac{e^{-\mu_1} \mu_1^x}{x!})(\frac{e^{-\mu_2} \mu_2^{t-x}}{(t-x)!})}{P(T=t)} \\ &= \frac{(\frac{e^{-\mu_1} \mu_1^x}{x!})(\frac{e^{-\mu_2} \mu_2^{t-x}}{(t-x)!})}{\frac{t!}{x!(t-x)!} \cdot \frac{e^{-\mu_1-\mu_2}}{t!} \cdot \frac{\mu_1^x \mu_2^{t-x}}{(\mu_1+\mu_2)^t}} \\ &= \binom{t}{x} \frac{\mu_1^x \mu_2^{t-x}}{(\mu_1+\mu_2)^t}. \end{aligned}$$

TRINOMIAL DISTRIBUTION:

$$(X_1, X_2, X_3) \sim \text{Tri}(n, p_1, p_2, p_3) \quad (7.3)$$

\exists_1 In a "trinomial distribution":

- ① there are three possible outcomes A, B, C for a trial; and
- ② n trials occur independently.

\exists_2 In particular,

- ① If $P(A) = p_1$, $P(B) = p_2$ & $P(C) = p_3$, then $p_1 + p_2 + p_3 = 1$, and
- ② If $X_1 = \#(A)$, $X_2 = \#(B)$ & $X_3 = \#(C)$, then $X_1 + X_2 + X_3 = n$.

\exists_3 In this case, we write

$$(X_1, X_2, X_3) \sim \text{Trinomial}(n, p_1, p_2, p_3).$$

or

$$(X_1, X_2, X_3) \sim \text{Tri}(n, p_1, p_2, p_3).$$

JOINT PMF [OF TRINOMIAL DISTRIBUTIONS]:

$$f(x_1, x_2, x_3) = \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \quad (7.3.1)$$

\exists_1 Let the rv X_1, X_2, X_3 form a trinomial distribution, with $X_1 + X_2 + X_3 = n$.

Then the joint pmf is necessarily the function f , where

$$\begin{aligned} f(x_1, x_2, x_3) &= P[X_1=x_1, X_2=x_2, X_3=x_3] \\ &= P[X_1=x_1, X_2=x_2] = P[X_1=x_1, X_3=x_3] = P[X_2=x_2, X_3=x_3] \\ &= \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}. \end{aligned}$$

Why? $f(x_1, x_2, x_3) = \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$

of ways to arrange n things w/ x_1 type 1, x_2 type 2 & x_3 type 3 multiplied.

\exists_2 Note that

$$\sum_{x_1=0}^n \sum_{x_2=0}^{n-x_1} P[X_1=x_1, X_2=x_2, X_3=n-x_1] = \sum_{x_1=0}^n \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{n-x_1} \\ = (p_1 + p_2 + p_3)^n \quad (\text{as } p_1 + p_2 + p_3 = 1).$$

MARGINAL PMFS [OF TRINOMIAL DISTRIBUTIONS]:

$$f_{X_1}(x_1) = \sum_{x_2=0}^{n-x_1} f(x_1, x_2, x_3) = \sum_{x_3=0}^{n-x_1} f(x_1, x_2, x_3) \quad (7.3.2)$$

\exists_1 Let X_1, X_2, X_3 form a trinomial distribution, with joint pmf f .

Denote the marginal pmf of f with X_1 as f_{X_1} .

Then

$$f_{X_1}(x_1) = P[X_1=x_1] = \sum_{x_2=0}^{n-x_1} f(x_1, x_2, x_3)$$

and

$$f_{X_1}(x_1) = P[X_1=x_1] = \sum_{x_3=0}^{n-x_1} f(x_1, x_2, x_3),$$

and f_{X_2}, f_{X_3} are defined similarly.

Proof: $f_{X_1}(x_1) = \sum_{x_2=0}^{n-x_1} f_{1,2}(x_1, x_2)$ (where $f_{1,2}(n, x_2) = P[X_1=x_1, X_2=x_2]$)

$$\begin{aligned} &= \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1! x_2! (n-x_1-x_2)!} p_1^{x_1} p_2^{x_2} p_3^{n-x_1-x_2} \\ &= \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \quad (\text{as } x_1+x_2+x_3=n) \\ &= \sum_{x_2=0}^{n-x_1} f(x_1, x_2, x_3), \end{aligned}$$

and the other sum is proved similarly. \square

$$(X_1, X_2, X_3) \sim \text{Tri}(n, p_1, p_2, p_3) \Rightarrow X_i \sim \text{Bin}(n, p_i) \quad (7.3.2)$$

\exists_2 Let X_1, X_2, X_3 form a trinomial distribution.

Then necessarily $X_i \sim \text{Bin}(n, p_i)$. $\forall i \in \{1, 2, 3\}$.

Proof: We prove it for $i=1$; the other cases are similar.

See that

$$\begin{aligned} P[X_1=x_1] &= f_{X_1}(x_1) \\ &= \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1! x_2! (n-x_1-x_2)!} p_1^{x_1} p_2^{x_2} p_3^{n-x_1-x_2} \\ &= p_1^{x_1} \frac{n!}{(n-x_1)! x_1! x_2! (n-x_1-x_2)!} p_2^{x_2} p_3^{n-x_1-x_2} \\ &= p_1^{x_1} \binom{n}{x_1} (p_2 + p_3)^{n-x_1} \quad (\text{by bin formula}) \\ &= p_1^{x_1} (p_1) (1-p_1)^{n-x_1}, \end{aligned}$$

which suffices to show that

$$X_1 \sim \text{Bin}(n, p_1)$$

as needed. \square

CONDITIONAL PMFS [FOR TRINOMIAL DISTRIBUTIONS]:

$$(X_1 | X_3=x_3) \sim \text{Bin}(n-x_3, \frac{p_1}{p_1+p_2}) \quad (7.3.3)$$

\exists_1 Let X_1, X_2, X_3 form a trinomial distribution.

Then necessarily

$$(X_1 | X_3=x_3) \sim \text{Bin}(n-x_3, \frac{p_1}{p_1+p_2}),$$

and the other cases (ie $(X_i | X_j=x_j)$) are defined similarly.

$$\begin{aligned} \text{Proof: } P[X_1=x_1 | X_3=x_3] &= \frac{P[X_1=x_1, X_3=x_3]}{P[X_3=x_3]} \\ &= \frac{\frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}}{\frac{(n-x_3)!}{(n-x_3-x_1)! x_1! x_2!} p_1^{x_1} p_2^{x_2} p_3^{x_3}} \quad (\text{since } X_3 \sim \text{Bin}(n, p_3)) \\ &= \frac{(x_1+x_2)!}{x_1! x_2!} \cdot \frac{p_1^{x_1} p_2^{x_2}}{(p_1+p_2)^{x_1+x_2}} \\ &= \left(\frac{x_1+x_2}{x_2}\right) \left(\frac{p_1}{p_1+p_2}\right)^{x_1} \left(\frac{p_2}{p_1+p_2}\right)^{x_2} \\ &= \binom{n-x_3}{n-x_3-x_1} p_1^{x_1} (1-p_1)^{n-x_3-x_1} \quad (p_1' = \frac{p_1}{p_1+p_2}), \end{aligned}$$

which is sufficient to prove the claim. \square

$$(X_1, X_2, X_3) \sim \text{Tri}(n, p_1, p_2, p_3) \Rightarrow X_1 + X_2 \sim \text{Bin}(n, p_1 + p_2)$$

$$(7.3.4)$$

\exists_1 Let $(X_1, X_2, X_3) \sim \text{Tri}(n, p_1, p_2, p_3)$.

Then necessarily $X_1 + X_2 \sim \text{Bin}(n, p_1 + p_2)$.

$$\text{Proof: } P[X_1+x_2=t] = \sum_{x_1=0}^t P[X_1=x_1, X_2=t-x_1]$$

$$= \sum_{x_1=0}^t \frac{n!}{x_1! (t-x_1)!(n-x_1)!} p_1^{x_1} p_2^{t-x_1} p_3^{n-x_1}$$

$$= \frac{n!}{t! (n-t)!} \sum_{x_1=0}^t \frac{t!}{x_1! (t-x_1)!} p_1^{x_1} p_2^{t-x_1} p_3^{n-x_1}$$

$$= \binom{n}{t} (p_1 + p_2)^t (p_1 p_2)^{n-t} \quad (\text{by bin formula}).$$

which suffices to prove the claim. \square

MULTINOMIAL DISTRIBUTION:

$$(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k) \quad (7.4)$$

\exists_1 In a "multinomial distribution":

- ① Each trial has k outcomes, say A_1, \dots, A_k (where $k \geq 2$); and

- ② we repeat said trial independently n times.

\exists_2 In this case, we say

$$(X_1, \dots, X_k) \sim \text{Multinomial}(n, p_1, \dots, p_k)$$

or

$$(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k),$$

where $p_i = P(A_i)$. $\forall i \in \{1, \dots, k\}$ and $X_i = \#\{A_i \text{ occurred in the trials}\}$.

JOINT PMF [OF MULTINOMIAL DISTRIBUTIONS]:

$$P[X_1=x_1, \dots, X_k=x_k] = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$$

\exists_1 Let $(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k)$.

Then the joint pmf of X_1, \dots, X_k is given by f , where

$$f(x_1, \dots, x_k) = P[X_1=x_1, \dots, X_k=x_k] = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}.$$

Why? $f(x_1, \dots, x_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$

generalization of combinations multiplying the probabilities out

Week 8:

Expectation and Variance of Multiple Variables

$$E(g(x_1, x_2)) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) f(x_1, x_2)$$

«LAW OF UNCONSCIOUS STATISTICIAN» (8.1)

\square : Let x_1, x_2 be drvs, and let $(x_1, x_2) \sim f(x_1, x_2)$.

Then necessarily

$$E(g(x_1, x_2)) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) f(x_1, x_2).$$

$$E(aX + bY + c) = aE(X) + bE(Y) + c \quad (8.2)$$

\square : Let X, Y be drvs, and let $a, b, c \in \mathbb{R}$.

Then necessarily

$$E(aX + bY + c) = aE(X) + bE(Y) + c.$$

Proof: Let $(x, y) \sim f(x, y)$.

$$\begin{aligned} \Rightarrow E(aX + bY + c) &= \sum_{x,y} (ax + by + c) f(x, y) \\ &= \sum_x \sum_y ax f(x, y) + \sum_x \sum_y by f(x, y) + \sum_x \sum_y c f(x, y) \\ &= a \sum_x \sum_y f(x, y) + b \sum_y \sum_x f(x, y) + c \sum_x \sum_y f(x, y) \\ &\quad [P(X=x)] \quad [P(Y=y)] \quad [] \\ &= aE(X) + bE(Y) + c. \end{aligned}$$

$$X \text{ & } Y \text{ ARE INDEPENDENT} \Rightarrow E[g(x)h(y)] = E[g(x)]E[h(y)]$$

(8.2)

\square : Let X, Y be independent drvs.

Then necessarily

$$E[g(x)h(y)] = E[g(x)]E[h(y)].$$

Proof: Let $(x, y) \sim f(x, y)$.

$$\begin{aligned} E[g(x)]E[h(y)] &= \left(\sum_x g(x)f_x(x) \right) \left(\sum_y h(y)f_y(y) \right) \\ &= \sum_x \sum_y g(x)h(y) f_x(x)f_y(y) \\ &= \sum_x \sum_y g(x)h(y) f(x, y) \quad (\text{since } X, Y \text{ are independent, see 7.1.3}) \\ &= E[g(x)h(y)] \quad (\text{by LOUS}), \end{aligned}$$

as needed. \blacksquare

COVARIANCE: $Cov(X, Y)$ (8.3)

\square : Let X, Y be drvs.

Then, the "covariance" between X & Y , denoted as

" $Cov(X, Y)$ ", is defined to be

$$\begin{aligned} Cov(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

Proof: Let $(x, y) \sim f(x, y)$. Then

$$\begin{aligned} Cov(X, Y) &= \sum_{x,y} (x - E(X))(y - E(Y)) f(x, y) \quad (\text{by LOUS}) \\ &= \sum_{x,y} xy f(x, y) - \sum_x \sum_y E(X)y f(x, y) - \sum_x \sum_y E(Y)x f(x, y) + \sum_{x,y} E(X)E(Y)f(x, y) \\ &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X, Y) \quad (8.3)$$

\square : Let X, Y be drvs.

Then necessarily

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

$$\begin{aligned} \text{Proof: } Var(X+Y) &= E((X+Y)^2) - (E(X+Y))^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - (E(X) + E(Y))^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\ &= (E(X^2) - E(X)^2) + (E(Y^2) - E(Y)^2) + 2E(XY) - E(X)E(Y) \\ &= Var(X) + Var(Y) + 2Cov(X, Y). \end{aligned}$$

$$Cov(X, X) = Var(X)$$

\square : Let X be a drv.

Then necessarily $Cov(X, X) = Var(X)$.

$$\begin{aligned} \text{Proof: } Cov(X, X) &= \frac{2Var(X) - Var(X) - Var(X)}{2} \quad (\text{by rearranging the above equality}) \\ &= Var(X). \end{aligned}$$

X, Y ARE INDEPENDENT $\Rightarrow Cov(X, Y) = 0$

\square : Let X, Y be independent drvs.

Then necessarily $Cov(X, Y) = 0$.

$$\begin{aligned} \text{Proof: } Cov(X, Y) &= E(XY) - E(X)E(Y) \\ &= E(XY) - E(XY) \quad (\text{by 8.2}) \\ &= 0. \end{aligned}$$

\square : However, the converse is not necessarily true!

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X, Y)$$

\square : Let X, Y be drvs, and let $a, b, c \in \mathbb{R}$.

Then necessarily

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X, Y).$$

$$\begin{aligned} \text{Proof: } Var(aX + bY + c) &= E((aX + bY + c)^2) - (E(aX + bY + c))^2 \\ &= E(a^2 X^2 + b^2 Y^2 + c^2 + 2abXY + 2acX + 2bcY) - (aE(X) + bE(Y) + c)^2 \\ &= a^2 E(X^2) + b^2 E(Y^2) + c^2 + 2abE(XY) + 2acE(X) + 2bcE(Y) \\ &\quad - a^2 E(X)^2 - b^2 E(Y)^2 - c^2 - 2abE(X)E(Y) - 2acE(X) - 2bcE(Y) \\ &= a^2 (E(X^2) - E(X)^2) + b^2 (E(Y^2) - E(Y)^2) + 2ab(E(XY) - E(X)E(Y)) \\ &= a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X, Y). \end{aligned}$$

CORRELATION COEFFICIENT: ρ OR $\rho_{x,y}$ (8.4)

Let X, Y be drvs.

Then, the "correlation coefficient" of X & Y , denoted as

" ρ ", is defined to be equal to

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}},$$

i.e.

$$\rho = \frac{E[(X - E(X))(Y - E(Y))]}{\sqrt{E[(X - E(X))^2]} \sqrt{E[(Y - E(Y))^2]}}.$$

$$|\rho| \leq 1; |\rho|=1 \Leftrightarrow Y = aX+b; a>0 \Rightarrow \rho=1, a<0 \Rightarrow$$

$$\rho=-1$$

Let X, Y be drvs, and let ρ be the correlation coefficient of X & Y .

Then necessarily $|\rho| \leq 1$.

Furthermore, $|\rho|=1$ iff $Y = aX+b$ for some $a, b \in \mathbb{R}$.

① If $a > 0$, then $\rho=1$;

② If $a < 0$, then $\rho=-1$.

Proof. Let $S = X+tY$.

Then

$$\begin{aligned} 0 \leq \text{Var}(S) &= \text{Var}(X+tY) \\ &= \text{Var}(X) + t^2 \text{Var}(Y) + 2t \text{Cov}(X, Y), \end{aligned}$$

which is a quadratic function in t .

Since $0 \leq \text{Var}(S)$, it follows that this quadratic has at most one real root.

Let's evaluate the discriminant:

$$\Delta = (2\text{Cov}(X, Y))^2 - 4(\text{Var}(X))(\text{Var}(Y)) \leq 0$$

$$\Rightarrow 4\text{Cov}(X, Y)^2 - 4\text{Var}(X)\text{Var}(Y) \leq 0$$

$$\Rightarrow \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)\text{Var}(Y)} \leq 1$$

$$\Rightarrow |\rho| \leq 1 \quad (\text{as needed}).$$

In particular, $\Delta=0$ iff $\text{Var}(S)=0$, i.e.

$$4\text{Cov}(X, Y)^2 - 4\text{Var}(X)\text{Var}(Y) = 0,$$

i.e. $|\rho|=1$.

In particular, $\exists t \in \mathbb{R} \ni \text{Var}(X+tY)=0$, i.e. that $X+tY=c$ for some $c \in \mathbb{R}$, i.e. that

$$Y = aX+b \quad \text{for some } a, b \in \mathbb{R}.$$

$$\begin{aligned} \text{Then } \rho &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{E[(X-E(X))(aX+b-aE(X)-b)]}{\sqrt{E[(X-E(X))^2]E[(aX+b-aE(X)-b)^2]}} \\ &= \frac{E[a(X-E(X))^2]}{\sqrt{E[(X-E(X))^2]E[(aX+b-aE(X)-b)^2]}} \\ &= \frac{aE[(X-E(X))^2]}{\sqrt{E[(X-E(X))^2]E[(aX+b-aE(X)-b)^2]}} \\ &= \frac{aE[(X-E(X))^2]}{\sqrt{a^2E[(X-E(X))^2]}} \\ &= \frac{aE[(X-E(X))^2]}{|a|E[(X-E(X))^2]} \\ &= \frac{a}{|a|}, \end{aligned}$$

$$\text{so } \rho=1 \Leftrightarrow a>0 \quad \& \quad \rho=-1 \Leftrightarrow a<0. \quad \blacksquare$$

POSITIVE / NEGATIVE CORRELATION

Let X, Y be drvs.

① We say X & Y follow a "positive correlation" if large values of X tend to be associated with large values of Y , and small values of X tend to be associated with small values of Y .

② Conversely, we say X & Y follow a "negative correlation" if large values of X tend to be associated with small values of Y , and small values of X tend to be associated with large values of Y .

$$(X_1, X_2, X_3) \sim \text{Tri}(n, p_1, p_2, p_3) \Rightarrow \rho_{X_1, X_2} = -\sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}.$$

Let $(X_1, X_2, X_3) \sim \text{Tri}(n, p_1, p_2, p_3)$.

Then necessarily

$$\rho_{X_1, X_2} = -\sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}},$$

and a similar result holds for ρ_{X_1, X_3} & ρ_{X_2, X_3} too.

Proof. Recall that

$$(X_1, X_2, X_3) \sim \text{Tri}(n, p_1, p_2, p_3) \Rightarrow X_i \sim \text{Bin}(n, p_i).$$

Hence

$$E(X_i) = np_i, \quad \text{Var}(X_i) = np_i(1-p_i).$$

Then, see that

$$E(X_1 X_2) = \sum_{X_1=0}^n \sum_{X_2=0}^n X_1 X_2 \cdot \frac{n!}{X_1! X_2! (n-X_1-X_2)!} p_1^{X_1} p_2^{X_2} p_3^{n-X_1-X_2}.$$

On the other hand,

$$X_1 + X_2 \sim \text{Bin}(n, p_1 + p_2),$$

so

$$\text{Var}(X_1 + X_2) = n(p_1 + p_2)(1-(p_1 + p_2)).$$

Thus

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \frac{1}{2}(\text{Var}(X_1 + X_2) - \text{Var}(X_1) - \text{Var}(X_2)) \\ &= \frac{1}{2}(n(p_1 + p_2)(1-(p_1 + p_2)) - np_1(1-p_1) - np_2(1-p_2)) \\ &= -np_1 p_2. \end{aligned}$$

Hence

$$\begin{aligned} \rho_{X_1, X_2} &= \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} = \frac{-np_1 p_2}{\sqrt{np_1(1-p_1)np_2(1-p_2)}} \\ &= \frac{-np_1 p_2}{n\sqrt{p_1 p_2(1-p_1)(1-p_2)}} \\ &= -\sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}. \quad \blacksquare \end{aligned}$$

$$(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k) \Rightarrow \text{Cov}(X_i, X_j) = -np_i p_j$$

Let $(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k)$.

Then necessarily $\rho_{X_i, X_j} = -np_i p_j \quad \forall 1 \leq i, j \leq k$.

Proof. Similar to binomial case.

$$(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k) \Rightarrow \rho_{X_i, X_j} = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}$$

Let $(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k)$.

Then necessarily $\rho_{X_i, X_j} = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}$.

Proof. Similar to binomial case. \blacksquare

$$E\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i E(X_i)$$

Let X_1, \dots, X_n be drvs, and let $c_1, \dots, c_n \in \mathbb{R}$.

Then necessarily

$$E\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i E(X_i).$$

Proof. This can be proved via induction pretty easily. \blacksquare

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

Let X_1, \dots, X_n be drvs, and let $c_1, \dots, c_n \in \mathbb{R}$.

Then necessarily

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n c_i X_i\right) &= \sum_{i=1}^n c_i^2 \text{Var}(X_i) + 2 \sum_{i < j} c_i c_j \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n c_i^2 \text{Var}(X_i) + 2(c_1 c_2 \text{Cov}(X_1, X_2) + \dots + c_1 c_n \text{Cov}(X_1, X_n) \\ &\quad + c_2 c_3 \text{Cov}(X_2, X_3) + \dots + c_{n-1} c_n \text{Cov}(X_{n-1}, X_n)). \end{aligned}$$

$$\begin{aligned} \text{Proof.} \quad \text{Var}\left(\sum_{i=1}^n c_i X_i\right) &= E\left(\left(\sum_{i=1}^n c_i X_i\right)^2\right) - E\left[\left(\sum_{i=1}^n c_i X_i\right)\right]^2 \\ &= E\left[\sum_{i=1}^n \sum_{j=1}^n c_i c_j X_i X_j\right] - \left(\sum_{i=1}^n c_i E(X_i)\right)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n c_i c_j E(X_i X_j) - \sum_{i=1}^n \sum_{j=1}^n c_i c_j E(X_i) E(X_j) \\ &= \sum_{i < j} c_i c_j \underbrace{E(X_i X_j) - E(X_i) E(X_j)}_{\text{Var}(X_i)} + \sum_{i < j} c_i c_j \underbrace{(E(X_i X_j) - E(X_i) E(X_j))}_{\text{Cov}(X_i, X_j)} + \sum_{i < j} c_i c_j (E(X_i X_j) - E(X_i) E(X_j)) \\ &= \sum_{i=1}^n c_i^2 \text{Var}(X_i) + 2 \sum_{i < j} c_i c_j \text{Cov}(X_i, X_j). \quad \blacksquare \end{aligned}$$

$$X_1, \dots, X_n \text{ ARE INDEPENDENT} \Rightarrow \text{Var}\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 \text{Var}(X_i)$$

Let X_1, \dots, X_n be drvs, and let $c_1, \dots, c_n \in \mathbb{R}$.

Suppose X_1, \dots, X_n are independent.

Then necessarily

$$\text{Var}\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 \text{Var}(X_i).$$

Proof. This follows from the fact that

X_i, X_j are independent $\Rightarrow \text{Cov}(X_i, X_j) = 0$,

and applying this to the previous theorem. \blacksquare

INDICATOR FUNCTION TECHNIQUES

USING IFS ON $\text{Bin}(n, p)$

Given: Let $X \sim \text{Bin}(n, p)$. Find $E(X)$ & $\text{Var}(X)$ using indicator variable techniques.

Solⁿ: Let $X_i = 1$ if the i^{th} trial is a success & 0 otherwise, ie $X_i = I[\text{the } i^{\text{th}} \text{ trial is a success}]$.

Then

$$X_i \sim \text{Bernoulli}(p),$$

and since X_1, \dots, X_n are independent, thus

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n E(X_i) \\ &= np, \end{aligned}$$

and

$$\begin{aligned} \text{Var}(X) &= \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} (0) \quad (\text{by independence of } X_1, \dots, X_n) \\ &= \sum_{i=1}^n p(1-p) \end{aligned}$$

$$\therefore \text{Var}(X) = np(1-p).$$

USING IFS ON $\text{Hyp}(N, M, n)$

Given: Suppose there are M red balls & $N-M$ blue balls in a box.

Randomly select n balls without replacement from the box.

Let X be the # of red balls selected, so that

$$X \sim \text{Hyp}(N, M, n).$$

We can find $E(X)$ & $\text{Var}(X)$ using IF techniques.

Solⁿ: Let $X_i = 1$ if the i^{th} selection is a red ball, and 0 otherwise; ie $X_i = I[\text{the } i^{\text{th}} \text{ trial is a success}]$.

Then

$$X_i \sim \text{Bernoulli}\left(\frac{M}{N}\right), \text{ but note the } X_i's \text{ are not independent!}$$

$$\Rightarrow E(X_i) = \frac{M}{N}, \quad \text{Var}(X_i) = \frac{M}{N}(1 - \frac{M}{N}).$$

So

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n E(X_i) \\ &= \frac{nM}{N}. \end{aligned}$$

Then,

$$\begin{aligned} \text{Var}(X) &= \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j). \end{aligned}$$

See that

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \sum_{x_i=0}^1 \sum_{x_j=0}^1 x_i x_j P[X_i=x_i, X_j=x_j] \\ &= 1 \cdot 1 \cdot P[X_i=1, X_j=1] \\ &= P[X_i=1] P[X_j=1 | X_i=1] \quad (\text{by total prob formula}) \\ &= \frac{M}{N} \cdot \frac{M-1}{N-1}, \end{aligned}$$

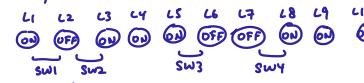
and so

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \\ &= n \cdot \frac{M}{N} \cdot (1 - \frac{M}{N}) + 2 \binom{n}{2} \left[\frac{M}{N} \cdot \frac{M-1}{N-1} - \left(\frac{M}{N} \right)^2 \right] \\ &= n \cdot \frac{M}{N} \cdot \left(\frac{N-M}{N} \right) + n(n-1) \left[\frac{M}{N} \left(\frac{M-1}{N-1} - \frac{M}{N} \right) \right] \\ &= n \cdot \frac{M}{N} \cdot \frac{N-M}{N} + n(n-1) \frac{M}{N} \left(\frac{NM-N-MN+M}{N(N-1)} \right) \\ &= n \cdot \frac{M}{N} \cdot \frac{N-M}{N} + n(n-1) \frac{M}{N} \left(\frac{M-N}{N(N-1)} \right) \\ &= n \cdot \frac{M}{N} \cdot \frac{N-M}{N} \left[1 - \frac{(n-1)}{N-1} \right] \\ &= n \cdot \frac{M}{N} \cdot \frac{N-M}{N} \left[\frac{(N-1)-(n-1)}{N-1} \right] \\ &= n \cdot \frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{N-n}{N-1}. \quad \square \end{aligned}$$

Example 10 Suppose 10 decoration lights are arranged in a row. Each light can be ON with probability 0.7 or OFF with probability 0.3, independent of each other. Two adjacent lights are called a SWITCH if they are ON-OFF or OFF-ON.

Denote $X = \text{number of SWITCHes}$. For $i = 1, \dots, n$, let $X_i = 1$ if light i and $i+1$ form a switch; 0 otherwise. Therefore, $X = \sum_{i=1}^9 X_i$. Find $E(X)$ and $\text{Var}(X)$.

Solⁿ:



See that

$$X_i \sim \text{Ber}(0.7 \times 0.3 + 0.3 \times 0.7) = \text{Ber}(0.42).$$

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^9 X_i\right) \\ &= \sum_{i=1}^9 E(X_i) \\ &= \sum_{i=1}^9 0.42 = 9(0.42). \end{aligned}$$

Next,

$$\begin{aligned} \text{Var}(X) &= \text{Var}\left(\sum_{i=1}^9 X_i\right) \\ &= \sum_{i=1}^9 \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq 9} \text{Cov}(X_i, X_j) \end{aligned}$$

However, note that
→ since any lights that cannot form a switch (ie not directly adjacent to one another) do not have any "overlap",

X_i, X_j are independent if $|j-i| > 1$.

$$\begin{aligned} \text{Thus } \sum_{1 \leq i < j \leq 9} \text{Cov}(X_i, X_j) &= \sum_{i=1}^9 \text{Cov}(X_i, X_{i+1}) \\ &= \sum_{i=1}^9 [E(X_i X_{i+1}) - E(X_i) E(X_{i+1})] \\ &= \sum_{i=1}^9 [(0.7 \times 0.3 \times 0.7 + 0.3 \times 0.7 \times 0.3) - 0.42^2] \\ &= 8 \cdot (0.21 - 0.42^2), \end{aligned}$$

$$\begin{aligned} \text{and so } \text{Var}(X) &= \sum_{i=1}^9 \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq 9} \text{Cov}(X_i, X_j) \\ &= 9(0.42)(1-0.42) + 2 \cdot 8 \cdot (0.21 - 0.42^2). \end{aligned}$$

Chapter 9: Continuous Probability Distributions

CONTINUOUS RV / CRV

Let X be a crv.
We say X is "continuous" if
 $|\text{range}(X)| > 1\text{N}1$.
eg $\text{range}(X) = \mathbb{R}, [0,1], \text{ etc.}$

PROBABILITY DENSITY FUNCTION / PDF: $f(x)$ (9.1.1)

Let X be a crv.
Then, the probability density function of X , ie $f(x)$, describes the "distribution" of probabilities of X .

In particular:

- ① $f(x) \geq 0 \quad \forall x \in (-\infty, \infty)$;
- ② $P[a \leq X \leq b] = \int_a^b f(x) dx$;
- ③ $P[X > b] = \int_b^\infty f(x) dx$;
- ④ $P[X < a] = \int_{-\infty}^a f(x) dx$;
- ⑤ $P[-\infty < X < \infty] = \int_{-\infty}^\infty f(x) dx = 1$.

CUMULATIVE DISTRIBUTION FUNCTION / CDF: $F(x)$ (9.1.2)

Let X be a crv, with pdf $f(x)$.
Then the "cumulative distribution function" of X , denoted by " $F(x)$ ", is defined to be equal to

$$F(x) = P[X \leq x] = \int_{-\infty}^x f(x).$$

Note that

- ① $0 \leq F(x) \leq 1 \quad \forall x \in \mathbb{R}$;
- ② $F(-\infty) = 0$ & $F(\infty) = 1$;
- ③ $F(x) \leq F(x_2) \Leftrightarrow x_1 \leq x_2$;
- ④ If f is cts at x , then $\frac{dF(x)}{dx} = f(x)$.

$X \sim f(x), Y = h(x)$ IS 1-1 \Rightarrow

$$g(y) = f(h^{-1}(y)) \left| \frac{dy}{dy} h^{-1}(y) \right| \quad \text{CHANGE OF}$$

VARIABLES >> (9.1.3)

Let $X \sim f(x)$, and let $Y = h(x)$ such that h has a unique inverse (ie h is 1-1). Then necessarily $Y \sim g(y)$ by

$$g(y) = f(h^{-1}(y)) \left| \frac{dy}{dy} h^{-1}(y) \right|.$$

eg¹ $X \sim f(x) = 3x^2, x \in [0,1]; Y = X^2$.

$$\text{sol}^2. P[Y \leq y] = P[X^2 \leq y] = P[X \leq \sqrt{y}] = F(\sqrt{y}).$$

$$\begin{aligned} \Rightarrow g(y) &= \frac{1}{2y} P[Y \leq y] = \frac{1}{2y} F(\sqrt{y}) \\ &= F'(\sqrt{y}) \cdot \frac{1}{2y} (\sqrt{y})' \\ &= f(\sqrt{y}) \cdot \frac{1}{2y} (2\sqrt{y}) \\ &= 3(\sqrt{y})^2 \cdot \frac{1}{2}\sqrt{y} \\ &= \frac{3}{2}y^{1.5}. \end{aligned}$$

eg² $X \sim f(x) = 3x^2, x \in [0,1]; Y = -\log X$.

$$\begin{aligned} \text{sol}^2. P[Y \leq y] &= P[-\log(x) \leq y] \\ &= P[\log(x) \geq -y] \\ &= P[x \geq e^{-y}] \\ &= 1 - P[x < e^{-y}] \\ &= 1 - P[x \leq e^{-y}] \\ &= 1 - F(e^{-y}). \end{aligned}$$

$$\begin{aligned} \text{Hence } g(y) &= \frac{d}{dy} P[Y \leq y] = -F'(e^{-y}) \cdot \frac{1}{x} e^{-y} \\ &= f(e^{-y}) \cdot \left| \frac{1}{x} e^{-y} \right| \\ &= 3(e^{-y})^2 e^{-y} \\ &= 3e^{-3y}. \quad (y \geq 0) \end{aligned}$$

EXPONENTIAL DISTRIBUTION: $X \sim \text{Exp}(\lambda)$ (9.2.1)

We say the crv $X \sim \text{Exp}(\lambda)$ iff

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0,$$

where $\lambda > 0$.

In particular,

$$F(x) = \int_{-\infty}^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad \forall x \geq 0.$$

Exponential distributions are often used to model waiting times.

eg Say λ is the intensity parameter for a Poisson process.

let X : waiting time for next event

Then $\text{range}(X) = [0, \infty)$ &

$$\begin{aligned} F(x) &= 1 - P[X > x] \\ &= 1 - P[\text{no events in } (0, x)] \\ &= 1 - e^{-\lambda x} \quad (\text{using Poisson dist}) \\ &= 1 - e^{-\lambda x}, \quad \boxed{\text{so } X \sim \text{Exp}(\lambda)}. \end{aligned}$$

Also note the "memory-less property":

$$P[X > t+s \mid X > s] = P[X > t].$$

GAMMA DISTRIBUTION: $X \sim \text{Gam}(\alpha, \beta)$ (9.2.2)

We say $X \sim \text{Gam}(\alpha, \beta)$ iff

$$f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad \forall x \geq 0,$$

where $\alpha, \beta > 0$.

Here, the gamma function Γ is given by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0.$$

Note that

$$\Gamma(\alpha+1) = \alpha \Gamma(\alpha); \quad \text{and}$$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}.$$

Proof: ① can be shown via integration by parts fairly easily.

$$\Gamma(\frac{1}{2}) = \int_0^{+\infty} x^{-\frac{1}{2}} e^{-x} dx.$$

let $x = u^2, \Rightarrow dx = 2u du$.

$$\Rightarrow \Gamma(\frac{1}{2}) = \int_0^\infty 2u e^{-u^2} du.$$

$$\Rightarrow \Gamma(\frac{1}{2}) = 4 \int_0^{+\infty} \int_0^{+\infty} e^{-u^2-v^2} du dv.$$

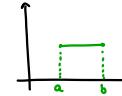
$$\text{let } u = r \cos \theta, v = r \sin \theta \quad \Rightarrow \quad \int_0^{+\infty} \int_0^{+\infty} e^{-u^2-v^2} dr d\theta$$

$$= \pi.$$

$$\Rightarrow \Gamma(\frac{1}{2}) = \sqrt{\pi}. \quad \boxed{\text{so}}$$

UNIFORM DISTRIBUTION: $X \sim \text{Unif}(a, b)$ (9.2.3)

\exists_1 We say $X \sim \text{Unif}(a, b)$ iff
 $f(x) = \frac{1}{b-a} \quad \forall x \in [a, b]$.



\exists_2 In this case,

$$F(x) = \frac{x-a}{b-a} \quad \forall x \in [a, b].$$

* in particular, if $a=0$ & $b=1$, then

$$F(x) = x.$$

F_X IS INCREASING $\Rightarrow F_X(x) \sim \text{Unif}(0, 1)$

let X have cdf $F_X(x)$, and suppose F_X is strictly increasing.

(let $y = F_X(x)$). Then $y \sim \text{Unif}(0, 1)$.

$$\begin{aligned} \text{Proof: } P[Y \leq y] &= P[F_X(x) \leq y] \\ &= P[F_X^{-1}(F_X(x)) \leq F_X^{-1}(y)] \\ &= P[x \leq F_X^{-1}(y)] \\ &= F_X(F_X^{-1}(y)) \\ &= y \\ &= \frac{y-0}{1-0}. \end{aligned}$$

so $y \sim \text{Unif}(0, 1)$. \square

BETA DISTRIBUTION: $X \sim \text{Beta}(\alpha_1, \alpha_2)$ (9.2.4)

\exists_1 Let $\alpha_1, \alpha_2 > 0$. Then, the beta function of α_1, α_2 , denoted as " $B(\alpha_1, \alpha_2)$ ", is defined to be

$$B(\alpha_1, \alpha_2) = \int_0^1 x^{\alpha_1-1} (1-x)^{\alpha_2-1} dx,$$

or equivalently,

$$B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}.$$

\exists_2 Then, we say $X \sim \text{Beta}(\alpha_1, \alpha_2)$ iff

$$f(x) = \frac{1}{B(\alpha_1, \alpha_2)} x^{\alpha_1-1} (1-x)^{\alpha_2-1}, \quad x \in [0, 1].$$

* in particular, when $\alpha_1=1$ & $\alpha_2=1$, then

$$\begin{aligned} f(x) &= \frac{1}{B(1, 1)} x^0 (1-x)^0 \\ &= \frac{\Gamma(1)}{\Gamma(1)\Gamma(1)} \\ &= \frac{1!}{1!0!} \\ &= 1, \end{aligned}$$

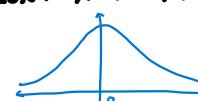
so $F(x) = \int_0^x 1 dx = x$, so $X \sim \text{Unif}(0, 1)$. \square

STANDARD NORMAL DISTRIBUTION: $X \sim N(0, 1)$

(9.2.5)

\exists_1 We say $Z \sim N(0, 1)$ iff

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \forall z \in (-\infty, \infty).$$



\exists_2 In this case, the cdf of Z , denoted by Φ , is equal to

$$\Phi(z) = P[Z \leq z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

* note that $\Phi(0) = \frac{1}{2}$ by symmetry of $f(z)$.

GENERAL NORMAL DISTRIBUTION: $X \sim N(\mu, \sigma^2)$

\exists_1 We say $X \sim N(\mu, \sigma^2)$ iff

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \forall x \in (-\infty, \infty),$$

where $\mu \in \mathbb{R}$ & $\sigma > 0$.

\exists_2 Note that in this case, we have

$$f(x+\delta) = f(x-\delta).$$

STANDARDIZATION

(let $X \sim N(\mu, \sigma^2)$). Let $Z = \frac{X-\mu}{\sigma}$.

Then see that

$$\begin{aligned} F_Z(z) &= P[Z \leq z] = P\left[\frac{X-\mu}{\sigma} \leq z\right] \\ &= P[X-\mu \leq z\sigma] \\ &= P[X \leq z\sigma + \mu] \\ &= F_X(\mu + z\sigma). \end{aligned}$$

and so

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} F_Z(z) \\ &= \frac{d}{dz} F_X(\mu + z\sigma) \\ &= f_X(\mu + z\sigma) \sigma \\ &= \frac{\sigma}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\mu + z\sigma - \mu)^2}{2\sigma^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \end{aligned}$$

In other words,

$$X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0, 1).$$

PDF/CDF OF MULTIVARIATE CONTINUOUS DISTRIBUTIONS

\exists_1 Let x_1, \dots, x_n be crvs, with joint pdf $f(x_1, \dots, x_n)$. (this is similar to the discrete case).

Then, by definition, we have

$$P(a_1 < x_1 < b_1, \dots, a_n < x_n < b_n) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(x_1, \dots, x_n) dx_n \dots dx_1.$$

$$E(g(x_1, \dots, x_n)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_n \dots dx_1.$$

< LAW OF THE UNCONSCIOUS STATISTICIAN >

\exists_1 Let $(x_1, \dots, x_n) \sim f(x_1, \dots, x_n)$ be crv.

Then necessarily

$$E(g(x_1, \dots, x_n)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_n \dots dx_1.$$

eg $(x, y) \sim f(x, y)$: then

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dy dx.$$

Chapter 10.1:

Expectation and Variance of Continuous Random Variables

EXPECTATION [OF CRV]: $E(X) = \int_{-\infty}^{\infty} x f(x) dx$ (10.1.1)

(at X be a crv, with pdf $f(x)$.
Then

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

E(g(x)) $= \int_{-\infty}^{\infty} g(x) f(x) dx$ (LAW OF UNCONSCIOUS STATISTICIAN) (10.1.1)

(at X be a crv, with pdf $f(x)$.
Then

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

VARIANCE [OF CRV]: $\text{Var}(X) = \int_{-\infty}^{\infty} [x - E(x)]^2 f(x) dx$ (10.1.1.2)

(at X be a crv, with pdf $f(x)$.
Then

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - E(x))^2 f(x) dx.$$

Proof. Follows from Lous as $\text{Var}(X) := E((X - E(X))^2)$. \square

$X \sim \text{Exp}(\lambda) \Rightarrow E(X) = \frac{1}{\lambda}$, $\text{Var}(X) = \frac{1}{\lambda^2}$ (10.1.2.1)

(at $X \sim \text{Exp}(\lambda)$. Then necessarily $E(X) = \frac{1}{\lambda}$ and $\text{Var}(X) = \frac{1}{\lambda^2}$.

Proof. $X \sim \text{Exp}(\lambda) \Rightarrow f(x) = \lambda e^{-\lambda x}$.

Thus

$$E(X) = \int_0^{+\infty} x \lambda e^{-\lambda x} dx$$

$$= \lambda \int_0^{+\infty} x e^{-\lambda x} dx.$$

(at $Y \sim \text{Gamma}(\alpha, \beta)$, so that

$$g(y) = \frac{1}{\beta^{\alpha} \Gamma(\alpha)} y^{\alpha-1} \exp(-\frac{y}{\beta}). \quad y \geq 0, \alpha > 0, \beta > 0$$

Then

$$1 = \int_0^{+\infty} g(y) dy$$

$$= \int_0^{+\infty} \frac{1}{\beta^{\alpha} \Gamma(\alpha)} y^{\alpha-1} \exp(-\frac{y}{\beta}) dy.$$

$$\Rightarrow \beta^{\alpha} \Gamma(\alpha) = \int_0^{+\infty} y^{\alpha-1} \exp(-\frac{y}{\beta}) dy.$$

Let $\beta = \frac{1}{\lambda}$, $\alpha = 2$. The eqn above becomes

$$\int_0^{+\infty} y \exp(-\lambda y) dy = (\frac{1}{\lambda})^2 \Gamma(2)$$

$$= (\frac{1}{\lambda})^2 (1)$$

and so

$$E(X) = \lambda \cdot \int_0^{+\infty} x \exp(-\lambda x) dx = \lambda \cdot \frac{1}{\lambda^2} = \frac{1}{\lambda}.$$

(at $\beta = \frac{1}{\lambda}$, $\alpha = 3$. The eqn above becomes

$$\int_0^{+\infty} y^2 \exp(-\lambda y) dy = (\frac{1}{\lambda})^3 \Gamma(3) = \frac{2}{\lambda^3},$$

and so

$$E(X^2) = 2 \int_0^{+\infty} x^2 \exp(-\lambda x) dx = 2 \cdot \frac{2}{\lambda^3} = \frac{2}{\lambda^2}.$$

Thus

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

$$= \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2$$

$$= \frac{1}{\lambda^2},$$

as required. \square

$X \sim \text{Gamma}(\alpha, \beta) \Rightarrow E(X) = \alpha\beta$, $\text{Var}(X) = \alpha\beta^2$ (10.1.2.2)

(at $X \sim \text{Gamma}(\alpha, \beta)$.

Then necessarily $E(X) = \alpha\beta$ & $\text{Var}(X) = \alpha\beta^2$.

Proof. Recall $X \sim \text{Gamma}(\alpha, \beta) \Rightarrow f(x) = \frac{1}{\beta^{\alpha} \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$, $x \geq 0, \alpha, \beta > 0$. Thus

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

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

(at $u = \frac{x}{\beta} \Rightarrow du = \frac{1}{\beta} dx$

$$\text{so } E(X) = \frac{1}{\beta^{\alpha} \Gamma(\alpha)} \int_{u=0}^{u=+\infty} (\beta u)^{\alpha} e^{-u} (\beta du)$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{+\infty} u^{\alpha} e^{-u} du$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha+1)} \quad (\text{recall } \Gamma(\alpha+1) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx)$$

$$= \alpha \beta.$$

Similarly,

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

Using $u = \frac{x}{\beta}$ again,

$$\Rightarrow E(X^2) = \frac{1}{\beta^{\alpha} \Gamma(\alpha)} \int_{u=0}^{u=+\infty} (\beta u)^{\alpha+1} e^{-u} (\beta du)$$

$$= \frac{\beta^{\alpha+2}}{\Gamma(\alpha+2)} \int_0^{+\infty} u^{\alpha+1} e^{-u} du$$

$$= \frac{\beta^{\alpha+2}}{\Gamma(\alpha+3)} = \beta^2 \alpha \beta.$$

Thus

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

$$= \beta^2 \alpha \beta - (\alpha \beta)^2$$

$$\therefore \text{Var}(X) = \alpha \beta^2. \quad \square$$

$X \sim N(0,1) \Rightarrow E(X) = 0$, $\text{Var}(X) = 1$ (10.1.2.3)

(at $X \sim N(0,1)$.

Then necessarily $E(X) = 0$ & $\text{Var}(X) = 1$.

Proof. Recall $X \sim N(0,1) \Rightarrow f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$, $x \in \mathbb{R}$.

Then

$$E(X) = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \cdot \exp(-\frac{x^2}{2}) dx$$

odd function

$$= 0.$$

Next

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx$$

even function

$$= 2 \int_0^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx$$

(at $u = x^2$ ($x > 0$) $\Rightarrow du = 2x dx \Rightarrow dx = \frac{du}{2x} = \frac{1}{2} u^{-\frac{1}{2}} du$

$$\Rightarrow E(X^2) = 2 \int_{u=0}^{u=+\infty} u \cdot \frac{1}{\sqrt{2\pi}} \exp(-\frac{u}{2}) \frac{1}{2} u^{-\frac{1}{2}} du$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} u^{\frac{1}{2}} \exp(-\frac{u}{2}) du$$

note $U \sim \text{Gamma}(\frac{3}{2}, 2) \Rightarrow \int_0^{+\infty} g(u) = 1 \Rightarrow \int_0^{+\infty} \frac{1}{2^{\frac{3}{2}} \Gamma(\frac{3}{2})} u^{\frac{3}{2}-1} \exp(-\frac{u}{2}) du = 1$

$$\Rightarrow \int_0^{+\infty} u^{\frac{1}{2}} \exp(-\frac{u}{2}) du = \frac{3}{2} \Gamma(\frac{3}{2}).$$

$$\therefore E(X^2) = \frac{1}{\sqrt{2\pi}} (2 \cdot \frac{3}{2} \Gamma(\frac{3}{2}))$$

$$= \frac{1}{\sqrt{2\pi}} (2^{\frac{3}{2}} (\frac{1}{2} \cdot \Gamma(\frac{1}{2})))$$

$$(\Gamma(\frac{3}{2}) = \frac{1}{2} \cdot \Gamma(\frac{1}{2}) = \frac{1}{2} \cdot \sqrt{\pi})$$

$$= 2^{\frac{3}{2}} \cdot \frac{1}{4\pi} \cdot 2^{\frac{3}{2}} \cdot \sqrt{\pi}$$

$$= 1. \quad \square$$

$X \sim N(\mu, \sigma^2) \Rightarrow E(X) = \mu$, $\text{Var}(X) = \sigma^2$

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

Then necessarily $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$.

Proof. Note $z = \frac{x-\mu}{\sigma}$, where $z \sim N(0,1)$.

Thus $X = \mu + \sigma z$, so

$$E(X) = \mu + \sigma E(z) = \mu. \quad \square$$

$$\text{Var}(X) = \sigma^2 \text{Var}(z) = \sigma^2. \quad \square$$

Chapter 10.2:

Moments and Moment Generating Functions

MOMENTS: $E(X^k)$ (10.2.1)

\exists Let X be a rv.
Then, the "kth moment" of X is simply $E(X^k)$.

MOMENT GENERATING FUNCTIONS / MGF:

$$M_X(t) = E(e^{tX})$$

\exists Let X be a rv.
Then, the "moment generating function" of X , denoted as " $M_X(t)$ ", is defined to be $M_X(t) := E(e^{tX})$, provided this expectation exists for all $t \in (-h, h)$, where $h > 0$ (ie is finite).

\exists In particular, if X is discrete, then

$$M_X(t) = E(e^{tX}) = \sum_x e^{tx} f(x).$$

\exists If X is instead continuous, then

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx.$$

\exists Note that a MGF uniquely determines a probability distribution, whether discrete or continuous.

$X \sim \text{Bin}(n, p) \Rightarrow M_X(t) = (e^t p + 1-p)^n, t \in (-\infty, \infty)$ (10.2.3.1)

\exists Let $X \sim \text{Bin}(n, p)$. Then necessarily $M_X(t) = (e^t p + 1-p)^n$, where $t \in (-\infty, \infty)$.

Proof. $M_X(t) = E(e^{tX})$

$$\begin{aligned} &= \sum_{x=0}^n e^{tx} f(x) \\ &= \sum_{x=0}^n e^{tx} \cdot \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (e^t p)^x (1-p)^{n-x} \\ &= (e^t p + 1-p)^n \quad (\text{by binomial formula}) \\ &< +\infty \quad \forall t \in (-\infty, \infty) \end{aligned}$$

$X \sim \text{Gamma}(\alpha, \beta) \Rightarrow M_X(t) = \frac{1}{(1-\beta t)^\alpha}, t < \frac{1}{\beta}$ (10.2.3.2)

\exists Let $X \sim \text{Gamma}(\alpha, \beta)$. Then necessarily $M_X(t) = \frac{1}{(1-\beta t)^\alpha}, \forall t < \frac{1}{\beta}$.

Proof. $M_X(t) = E(e^{tX})$

$$\begin{aligned} &= \int_0^{+\infty} e^{-x} \cdot \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} dx \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^{+\infty} x^{\alpha-1} \exp(-x(\frac{1}{\beta} - t)) dx \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^{+\infty} x^{\alpha-1} \exp(-\frac{x}{\beta(1-t)}) dx \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \cdot (\frac{1}{\beta(1-t)})^{-\alpha} \Gamma(\alpha) \quad (\text{if } \frac{1}{\beta} - t > 0 \text{ using the "gamma dist trick"}) \\ &= \frac{1}{\beta^\alpha} \left(\frac{1-\beta t}{\beta}\right)^{-\alpha} \\ &= \frac{1}{\beta^\alpha} \cdot \frac{\beta^\alpha}{(1-\beta t)^\alpha} \\ &= \frac{1}{(1-\beta t)^\alpha}, \quad t < \frac{1}{\beta}. \end{aligned}$$

What if $t \geq \frac{1}{\beta}$?

Then $M_X(t) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^{+\infty} x^{\alpha-1} \exp(-x(\frac{1}{\beta} - t)) dx$

$\underbrace{\quad}_{-\text{ve}}$ $\underbrace{\quad}_{+\text{ve}}$ ≥ 1

$$\begin{aligned} &\geq \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^{+\infty} x^{\alpha-1} dx \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \left[\frac{x^\alpha}{\alpha} \right]_0^{+\infty} \\ &= +\infty. \quad \blacksquare \end{aligned}$$

$X \sim N(0, 1) \Rightarrow M_X(t) = e^{\frac{t^2}{2}}, t \in (-\infty, \infty)$ (10.2.3.3)

\exists Let $X \sim N(0, 1)$. Then necessarily $M_X(t) = e^{\frac{t^2}{2}}$, for $t \in (-\infty, \infty)$.

$Y = aX + b \Rightarrow M_Y(t) = e^{bt} M_X(at), t \in (-\frac{b}{|a|}, \frac{b}{|a|})$ (10.2.3.4)

\exists Let X be a rv, with MAF $M_X(t)$ $\forall t \in (-h, h)$.

Let $Y = aX + b$.

Then the MAF of Y is necessarily given by

$$M_Y(t) = e^{bt} M_X(at) \quad \forall t \in (-\frac{b}{|a|}, \frac{b}{|a|}).$$

Proof. $M_Y(t) = E[e^{tY}]$

$$\begin{aligned} &= E[e^{t(aX+b)}] \\ &= E[e^{taX+tb}] \\ &= E[e^{taX} e^{tb}] \\ &= e^{tb} E[e^{taX}] \\ &= e^{bt} M_X(at). \end{aligned}$$

In particular, $M_X(at)$ exists for $at \in (-h, h) \Rightarrow t \in (-\frac{b}{|a|}, \frac{b}{|a|})$. \blacksquare

$X \sim N(\mu, \sigma^2) \Rightarrow M_X(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2})$

\exists Let $X \sim N(\mu, \sigma^2)$. Then necessarily $M_X(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2}), t \in (-\infty, \infty)$.

Proof. Note $X = \mu + \sigma Z$, $Z \sim N(0, 1)$.

Hence

$$\begin{aligned} M_X(t) &= e^{\mu t} M_Z(\sigma t) \\ &= e^{\mu t} \left(e^{\frac{\sigma^2 t^2}{2}} \right) \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}}. \quad \blacksquare \end{aligned}$$

$X \sim \text{Gamma}(\alpha, \beta) \Rightarrow \frac{X}{\beta} \sim \text{Gamma}(\alpha, 1)$

\exists Let $X \sim \text{Gamma}(\alpha, \beta)$. Then $\frac{X}{\beta} \sim \text{Gamma}(\alpha, 1)$.

Proof. Let $Y = \frac{X}{\beta}$. See that

$M_Y(t) = E[e^{tY}]$

$$\begin{aligned} &= \frac{1}{(1-\beta t)^\alpha}, \quad \frac{t}{\beta} < 1 \quad (\text{ie } t < 1) \\ &= \frac{1}{(1-t)^\alpha}. \end{aligned}$$

Note $W \sim \text{Gamma}(\alpha, 1) \Leftrightarrow M_W(t) = \frac{1}{(1-t)^\alpha}$.

Therefore $\frac{X}{\beta} \sim \text{Gamma}(\alpha, 1)$, as needed. \blacksquare

JOINT MGF: $M_{X,Y}(t_1, t_2) = E(e^{t_1 X + t_2 Y})$ (10.2.3.5)

\exists Let X and Y be rv. Then, the "joint MGF" of X & Y , denoted as " $M_{X,Y}(t_1, t_2)$ ", is defined to be

$$M_{X,Y}(t_1, t_2) = E(e^{t_1 X + t_2 Y}).$$

provided it exists $\forall t_1 \in (-h_1, h_1), t_2 \in (-h_2, h_2)$.

Note that

$$M_X(t_1) = E(e^{t_1 X}) = E(e^{t_1 X + 0 \cdot Y}) = M_{X,Y}(t_1, 0),$$

and

$$M_Y(t_2) = E(e^{t_2 Y}) = E(e^{0 \cdot X + t_2 Y}) = M_{X,Y}(0, t_2).$$

\Rightarrow this is valid, as $0 \in (-h_1, h_1)$ & $0 \in (-h_2, h_2)$.
This also extends to more variables as well.

Chapter III.1:

Main Properties of MGF

MOMENTS FROM MGF:

$$M_X(k) = 1, \quad M_X^{(k)}(0) = E(X^k) \quad (\text{III.1.1})$$

Let X be a rv with MGF $M_X(t)$, $t \in (-h, h)$.

Then necessarily $M_X(0) = 1$ and

$$M_X^{(k)}(0) = E(X^k), \quad k=1, 2, \dots$$

where

$$M_X^{(k)}(t) = \frac{d^k}{dt^k} M_X(t).$$

Proof. Recall

$$M_X(t) = E(e^{tX}).$$

Thus

$$M_X(0) = E(e^{tX})|_{t=0} = E(t)=1.$$

Case #1: X is a dev.

$$\begin{aligned} \text{Then } M_X^{(k)}(t) &= \frac{d^k}{dt^k} M_X(t) \\ &= \frac{d^k}{dt^k} \sum_x e^{tx} f(x) \\ &= \sum_x x^k e^{tx} f(x). \end{aligned}$$

so

$$M_X^{(k)}(0) = \sum_x x^k f(x) = E(X^k) \quad (\text{by LOUS}). \quad *$$

Case #2: X is a crv.

$$\begin{aligned} \text{Then } M_X^{(k)}(t) &= \frac{d^k}{dt^k} M_X(t) \\ &= \frac{d^k}{dt^k} \int_{-\infty}^{\infty} e^{tx} f(x) dx \\ &= \int_{-\infty}^{\infty} \frac{d^k}{dt^k} e^{tx} f(x) dx \\ &= \int_{-\infty}^{\infty} x^k e^{tx} f(x) dx. \end{aligned}$$

so

$$\begin{aligned} M_X^{(k)}(0) &= \int_{-\infty}^{\infty} x^k f(x) dx \\ &= \int_{-\infty}^{\infty} x^k dx \\ &= E(X^k). \quad \square \end{aligned}$$

JOINT MOMENTS FROM JOINT MGF:

$$E(X^j Y^k) = \frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} M_{X,Y}(t_1, t_2)|_{(t_1, t_2)=(0,0)} \quad (\text{III.1.2})$$

Let X, Y be crvs with joint MGF $M_{X,Y}(t_1, t_2)$, $\forall t_i \in (-h_1, h_1)$, $t_2 \in (-h_2, h_2)$.

Then necessarily

$$E(X^j Y^k) = \frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} M_{X,Y}(t_1, t_2) \Big|_{(t_1, t_2)=(0,0)}.$$

$$\begin{aligned} \text{Proof. } \frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} M_{X,Y}(t_1, t_2) &= \frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1 x + t_2 y} f(x,y) dx dy \\ &= \frac{\partial^j}{\partial t_1^j} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^k}{\partial t_2^k} e^{t_1 x + t_2 y} f(x,y) dx dy \\ &= \frac{\partial^j}{\partial t_1^j} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^k e^{t_1 x + t_2 y} f(x,y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^j}{\partial t_1^j} y^k e^{t_1 x + t_2 y} f(x,y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^j y^k e^{t_1 x + t_2 y} f(x,y) dx dy, \end{aligned}$$

and so

$$\begin{aligned} \frac{\partial^{j+k}}{\partial t_1^j \partial t_2^k} M_{X,Y}(t_1, t_2) \Big|_{(t_1, t_2)=(0,0)} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^j y^k e^{t_1 x + t_2 y} f(x,y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^j y^k f(x,y) dx dy \\ &= E(X^j Y^k) \quad (\text{by LOUS}). \quad \square \end{aligned}$$

Ex: $(X_1, X_2, X_3) \sim \text{Tri}(n, p_1, p_2, p_3) \Rightarrow E(X_1 X_2) = ?$

Solⁿ. See that

$$E(X_1 X_2) = \frac{\partial^2}{\partial t_1 \partial t_2} M_{X_1, X_2}(t_1, t_2)$$

where

$$\begin{aligned} M_{X_1, X_2}(t_1, t_2) &= E(e^{t_1 X_1 + t_2 X_2}) \quad (\forall t_1, t_2 \in \mathbb{R}) \\ &= \sum_{\substack{n-x_1 \\ n-x_2 \geq 0}} e^{t_1 x_1 + t_2 x_2} \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}, \quad x_3 = n - x_1 - x_2 \quad (\text{by LOUS}) \\ &= \sum_{\substack{n-x_1 \\ n-x_2 \geq 0}} \frac{n!}{x_1! x_2! x_3!} (p_1 e^{t_1})^{x_1} (p_2 e^{t_2})^{x_2} p_3^{x_3}, \quad x_3 = n - x_1 - x_2 \\ &= (p_1 e^{t_1} + p_2 e^{t_2} + p_3)^n. \quad (\text{Recall } (a+b+c)^n = \sum_{x_1+x_2+x_3=n} \frac{n!}{x_1! x_2! x_3!} a^{x_1} b^{x_2} c^{x_3}). \end{aligned}$$

Thus

$$\begin{aligned} \frac{\partial^2}{\partial t_1 \partial t_2} M_{X_1, X_2}(t_1, t_2) &= \frac{\partial^2}{\partial t_1 \partial t_2} (p_1 e^{t_1} + p_2 e^{t_2} + p_3)^n \\ &= \frac{\partial}{\partial t_1} (n(p_1 e^{t_1} + p_2 e^{t_2} + p_3)^{n-1} \cdot p_1 e^{t_1}) \\ &= n(n-1)(p_1 e^{t_1} + p_2 e^{t_2} + p_3)^{n-2} \cdot p_1 e^{t_1} \cdot p_1 e^{t_1}, \end{aligned}$$

and so

$$\begin{aligned} E(X_1 X_2) &= \frac{\partial^2}{\partial t_1 \partial t_2} M_{X_1, X_2}(t_1, t_2) \Big|_{(t_1, t_2)=(0,0)} \\ &= n(n-1)(p_1 e^0 + p_2 e^0 + p_3)^{n-2} p_1 e^0 p_1 e^0 \\ &= n(n-1)p_1 p_2. \quad \# \end{aligned}$$

X, Y ARE INDEPENDENT ($\Rightarrow M_{X,Y}(t_1, t_2) = M_X(t_1) M_Y(t_2)$)

<< MGF INDEPENDENCE THEOREM >> (III.1.3)

Let X, Y be rv.

Then X & Y are independent iff

$$M_{X,Y}(t_1, t_2) = M_X(t_1) M_Y(t_2) \quad \forall t_1 \in (-h_1, h_1), t_2 \in (-h_2, h_2).$$

$M_X(t) \equiv M_Y(t) \Leftrightarrow X=Y$ << UNIQUENESS OF MGF >>

(III.1.4)

Let X, Y be rv such that $M_X(t) = M_Y(t) \quad \forall t \in (-h, h)$.

Then necessarily $X=Y$.

In other words, X & Y have the same distribution.

Proof. This follows from injectivity of e^{tX} & e^{tY} . \square

X_1, X_2 ARE INDEPENDENT $\Rightarrow M_{X_1+X_2}(t) = M_{X_1}(t) M_{X_2}(t)$ (III.1.5)

Let X_1 & X_2 be independent rv, and X_i has

MGF $M_{X_i}(t) \quad \forall t \in (-h, h)$.

Then necessarily

$$M_{X_1+X_2}(t) = M_{X_1}(t) M_{X_2}(t).$$

Chapter 11.2:

Convergence in Probability

CONVERGENCE IN PROBABILITY: $X_n \xrightarrow{P} b$

Let (X_n) be a sequence of rvs.

Then, we say X_n "converges in probability" to a constant

b if for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P[|X_n - b| \geq \epsilon] = 0 \text{ or } \lim_{n \rightarrow \infty} [P(|X_n - b| < \epsilon)] = 1,$$

and in this case we write

$$X_n \xrightarrow{P} b. \quad |E(|X|^k)| < \infty \Rightarrow P(|X| \geq c) \leq \frac{E(|X|^k)}{c^k} \quad \forall c, k > 0$$

<< MARKOV'S INEQUALITY >> (11.2.1)

Let X be a rv, and suppose $E(|X|^k)$ is finite for each $k \in \mathbb{N}$.

Then necessarily

$$P(|X| \geq c) \leq \frac{E(|X|^k)}{c^k} \quad \forall c, k > 0.$$

Proof: We just consider the cts case; discrete case is similar.

See that

$$\begin{aligned} \frac{E(|X|^k)}{c^k} &= \frac{1}{c^k} \left[\int_{x:|x| \leq c} |x|^k f(x) dx + \int_{x:|x| > c} |x|^k f(x) dx \right] \\ &= \int_{x:|\frac{x}{c}| \leq 1} |\frac{x}{c}|^k f(x) dx + \int_{x:|\frac{x}{c}| \geq 1} |\frac{x}{c}|^k f(x) dx \\ &\geq 0 + \int_{x:|\frac{x}{c}| \geq 1} |\frac{x}{c}|^k f(x) dx \\ &\geq 0 + \int_{x:|\frac{x}{c}| \geq 1} 1^k f(x) dx \\ &= P\left(\frac{|X|}{c} \geq 1\right) \\ &= P(|X| \geq c). \end{aligned}$$

as needed. \square

INDEPENDENT & IDENTICALLY DISTRIBUTED RANDOM VARIABLES / IID

We say the rvs X_1, \dots, X_n are "independent & identically distributed", or just "IID", if

① each random variable has the same probability distribution (ie $X_1, \dots, X_n \sim f(x)$); and

② the variables are mutually independent.

In this case, we may also write that

$$X_i \stackrel{\text{iid}}{\sim} f(x).$$

X_i ARE IID, $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

$\Rightarrow \bar{X}_n \xrightarrow{P} \mu$ **<< WEAK LAW OF LARGE NUMBERS >>**

Let the rvs X_i be IID, with each $E(X_i) = \mu$ &

$$\text{Var}(X_i) = \sigma^2.$$

Denote

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i.$$

Then necessarily $\bar{X}_n \xrightarrow{P} \mu$.

Proof: First, see that

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{n \cdot \mu}{n} = \mu; \quad \&$$

$$\text{Var}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} (n \sigma^2) = \frac{\sigma^2}{n}$$

Let $\epsilon > 0$. We wish to show

$$\lim_{n \rightarrow \infty} P[|\bar{X} - \mu| \geq \epsilon] = 0.$$

By Markov's Inequality,

$$P[|\bar{X} - \mu| \geq \epsilon] \leq \frac{E[|\bar{X} - \mu|^k]}{\epsilon^k} \quad \forall k > 0.$$

In particular, when $k=2$,

$$\begin{aligned} P[|\bar{X} - \mu| \geq \epsilon] &\leq \frac{E[(\bar{X} - \mu)^2]}{\epsilon^2} \\ &= \frac{E[(\bar{X} - \mu)^2]}{\epsilon^2} \\ &= \frac{\text{Var}(\bar{X})}{\epsilon^2} \\ &= \frac{\sigma^2}{n \epsilon^2} \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

By a squeeze theorem argument, the proof follows. \square

$X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b \Rightarrow X_n + Y_n \xrightarrow{P} a+b, X_n Y_n \xrightarrow{P} ab$

Let X_n and Y_n be rvs such that

$$X_n \xrightarrow{P} a \quad \& \quad Y_n \xrightarrow{P} b.$$

Then necessarily

$$\textcircled{1} \quad X_n + Y_n \xrightarrow{P} a+b; \quad \text{and}$$

$$\textcircled{2} \quad X_n Y_n \xrightarrow{P} ab.$$