STAT 330
Personal Notes

Chapter 2.1: Introduction

Review for
- probability models;
- random variables.

PROBABILITY MODELS

A "probability model" is used to describe a random experiment.

It consists of 3 components:

① "Sample space" — the collection of all possible outcomes of a random experiment.
- we denote the sample space by "$\underline{\underline{S}}$".

eg tossing a coin twice,

$S = \{(H,H), (H,T), (T,H), (T,T)\}$

② "Event" — a subset of the sample space, S.
- we usually use capitals (eg A, B, C) to denote events.

eg A: 1st toss is a tail. (with S as earlier)

$\Rightarrow A = \{(T,H), (T,T)\}$

③ "Probability function" — a function of events P which satisfies the following:

① $0 \le P(A) \le 1$ for any event A;

② $P(S) = 1$;

③ P satisfies "countable additivity";

ie if A_i, A_j are pairwise, mutually exclusive events, ie if $A_i \cap A_j \neq \emptyset$, $i \neq j$, then this always holds:

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

If $A_1, \dots, A_n$ are pairwise mutually exclusive events, then necessarily

$$\boxed{P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).}$$

* "finite additivity".

Proof. If $P(\emptyset) = 0$, then consider

$A_1, \dots, A_n$, and $A_i = \emptyset$ $\forall i > n$.

Then
$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i).$$

④ From the above 3 properties, we can derive the following:

① $\boxed{P(\emptyset) = 0.}$

Proof. Let $A_1 = S$, and $A_2, A_3, \dots = \emptyset$.

$A_1, A_2, \dots$ are pairwise mutually exclusive, so we know

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = P(A_1) + \sum_{i=2}^{\infty} P(A_i) = P(S) + \sum_{i=2}^{\infty} P(\emptyset)$$

$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(S)$.

Thus
$$P(S) = P(S) + \sum_{i=2}^{\infty} P(\emptyset),$$

and $0 \le P(\emptyset) \le 1$ by def^n.

It follows that we must have that $P(\emptyset) = 0$, as needed. ∎

② Let $\bar{A}$ be the "complementary" event of A; ie $A \cup \bar{A} = S$ & $A \cap \bar{A} = \emptyset$.

* convention; we use $\bar{A}$ for complementary events.

Then $\boxed{P(A) + P(\bar{A}) = 1.}$

Proof. Let $A_1 = A$, & $A_2 = \bar{A}$, and $A_i = \emptyset$ $\forall i \ge 3$.

By def^n, A_i are pairwise & mutually exclusive events.

Thus,
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

$P(A \cup \bar{A}) = P(A) + P(\bar{A}) + 0 + 0 + \cdots$

$= P(S) = P(A) + P(\bar{A})$.

$= 1$

proof follows. ∎

③ If A_1 & A_2 are mutually exclusive, then necessarily

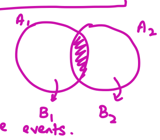
$$\boxed{P(A_1 \cup A_2) = P(A_1) + P(A_2).}$$

Proof. Similar to ③.

④ In general,

$$\boxed{P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).}$$

Proof. Let $B_1 = A_1 \setminus (A_1 \cap A_2)$; & $B_2 = A_2 \setminus (A_1 \cap A_2)$.



Then B_1, B_2 & $A_1 \cap A_2$ are mutually exclusive pairwise events.

Note that $A_1 \cup A_2 = A_2 \cup B_1$. A_2 & B_1 are mutually exclusive; thus

$P(A_1 \cup A_2) = P(A_2 \cup B_1) = P(A_2) + P(B_1)$.

Then, since $A_1 = B_1 \cup (A_1 \cap A_2)$, & these two events are mutually exclusive, thus

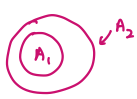
$P(A_1) = P(B_1) + P(A_1 \cap A_2)$.

Therefore $P(B_1) = P(A_1) - P(A_1 \cap A_2)$, and so

$P(A_1 \cup A_2) = P(A_2) + P(B_1) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$. ∎

⑤ If $A_1 \subseteq A_2$, then $\boxed{P(A_1) \le P(A_2).}$

Proof.



Let $B_1 = A_2 \setminus A_1$. By def^n $P(B_1) \ge 0$.

A_1 & B_1 are mutually exclusive; thus

$P(A_2) = P(A_1 \cup B_1) = P(A_1) + P(B_1) \ge P(A_1)$

as needed. ∎

CONDITIONAL PROBABILITY: $P(A|B)$

1. Let A, B be two events, where $P(B)>0$. Then, the "conditional probability" of A given B is

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

INDEPENDENCE OF TWO EVENTS

1. Let A & B be two events. We say A & B are "independent" iff

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

eg Tossing a coin twice. Let

A = 1^{st} toss is H = $\{(H,T),(H,H)\}$

B = 2^{nd} toss is H. = $\{(T,H),(H,H)\}$.

Then $P(A) = \frac{2}{4} = \frac{1}{2}$ & $P(B) = \frac{1}{2}$, &

$P(A\cap B) = \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2} = P(A)P(B)$.

Thus A & B are independent.

2. When A & B are independent, then necessarily

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

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

RANDOM VARIABLE / r.v.

1. A "random variable" is a function from the sample space to $\mathbb{R}$; ie

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

2. We usually denote these via X, Y, etc.

3. For $x\in\mathbb{R}$, we write

$$\boxed{X\le x := \{w: X(w)\le x\},}$$

which is an event.

eg Toss a coin twice.

Denote $X :=$ # of heads in these 2 tosses.

Then

$\{X=0\} = \{(T,T)\}$

$\{X=1\} = \{(H,T),(T,H)\}$

$\{X=2\} = \{(H,H)\}$

Then

$$\{X\le x\} = \begin{cases} x<0 \Rightarrow \phi \\ 0\le x<1 \Rightarrow \{X=0\} \\ 1\le x<2 \Rightarrow \{X=0\}\cup\{X=1\} \\ 2\le x \Rightarrow S \end{cases}$$

CUMULATIVE DISTRIBUTION FUNCTION / CDF

1. The c.d.f. of a random variable X, denoted as $F(x)$, is defined to be

$$\boxed{F(x) = P(X\le x) \quad \forall x\in\mathbb{R}.}$$

2. Some properties of $F(x)$:

① $F(x)$ is a non-decreasing function; ie if $\boxed{x_1\le x_2, \text{ then } F(x_1)\le F(x_2).}$

Proof. Let $A = X\le x_1$, $B = X\le x_2$. Then $A\subseteq B$, so $P(X\le x_1)\le P(X\le x_2)$. ☒

② $\boxed{\lim_{x\to+\infty} F(x) = 1, \quad \lim_{x\to-\infty} F(x) = 0.}$

③ $F(x)$ is a "right-continuous" function; ie

$$\boxed{\lim_{x\to a^+} F(x) = F(a).}$$

④ $\boxed{P(a<X\le b) = F(b) - F(a).}$

Proof. Let $A = \{X\le b\}$, & $B = \{X\le a\}$, and let $C = \{a<X\le b\}$.

Then $B\cup C = A$ & $B\cap C = \phi$.

Thus

$P(X\le b) = P(X\le a) + P(a<X\le b)$,

and rearranging gives the desired result. ☒

⑤ $\boxed{P(X=a) = F(a) - \lim_{x\to a^-} F(x).}$

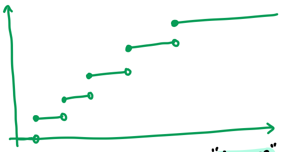
↳ if F is continuous at a, $P(X=a)=0$.

if F is discontinuous at a (ie if it is not left-cts), $P(X=a)\ne 0$.

Chapter 2.2: Discrete Random Variables

We say X is a "discrete random variable" if it can only take a finite or countable number of values.

CDF OF A DRV



1. Where there are "jumps", is where X can take values.
2. For a d.r.v, its cdf is a right continuous step function.

PROBABILITY FUNCTION: $f(x)$

The "probability function" of a drv X is defined to be

$$\boxed{f(x) = \mathbb{P}(X=x) = \begin{cases} >0, & \text{if } X \text{ can take value } x \\ =0, & \text{otherwise} \end{cases}}$$

SUPPORT [OF A DRV]

The "support" of X is

$$\boxed{A = \{x : f(x) > 0\}},$$

ie all the possible values X can take.

PROPERTIES OF $f(x)$

1. $\boxed{f(x) \geq 0 \quad \forall x \in \mathbb{R}.}$
2. $\boxed{\sum_{x \in A} f(x) = 1.}$

Proof. let $A_j = \{X = x_j\}$. Then

$$1 = \mathbb{P}(S) = \mathbb{P}\left(\bigcup_{i=1}^{|A|} A_i\right) = \sum_{i=1}^{|A|} \mathbb{P}(A_i) = \sum_{i=1}^{|A|} \mathbb{P}(X = x_i) = \sum_{i=1}^{|A|} f(x_i). \quad \blacksquare$$

COMMONLY USED DRV

1. Bernoulli r.v.: $X \sim \text{Bern}(p)$.
   - X can only take $0, 1$ as possible values
   - $S = \{\omega_1, \omega_2\}$, and we assign $X(\omega_1) = 0$, $X(\omega_2) = 1$
   - $p = \mathbb{P}(X=1)$, $1-p = \mathbb{P}(X=0)$
   - $A = \{0, 1\}$
   - $f(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \\ 0, & \text{otherwise} \end{cases}$.
   - we can verify $f(x) \geq 0$ & $\sum_{x \in A} f(x) = 1$.

   eg Toss a coin, $X = \#$ of heads you get. $X \sim \text{Ber}(p)$, $p =$ probability we get a head.

2. Binomial r.v.: $X \sim \text{Bin}(n, p)$.
   - toss a coin n times
   - let $X := \#$ of heads
   - assumptions:
     1. different tosses are independent
     2. $\mathbb{P}(\text{head})$ is fixed
   - support of $X = \{0, 1, \dots, n\}$
   - Probability function:

     $f(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

3. Geometric r.v.: $X \sim \text{Geo}(p)$.
   - $X := \#$ of failures before the 1^{st} success
   - $A = \{0, 1, \dots\}$
   - $f(x) = \mathbb{P}(X = x) = (1-p)^x p$.

   (we can prove f is a probability function).

4. Negative binomial r.v.: $X \sim \text{NegBin}(r, p)$.
   - $X := \#$ of failures before the r^{th} success
   - $A = \{0, 1, 2, \dots\}$

5. Poisson r.v.: $X \sim \text{Poi}(\mu)$.
   - $X = \#$ of events in a certain time period.
   - $A = \{0, 1, 2, \dots\}$
   - $f(x) = \dfrac{e^{-\mu} \mu^x}{x!}$

Chapter 2.3: Continuous Random Variables

1. If the collection of the possible values of X is an interval or $\mathbb{R}$, then we say X is a "continuous random variable".

2. Note: If X is a c.r.v, its cdf $F(x)$ is a continuous function, and $F(x)$ is differentiable almost everywhere.
   - it is not differentiable for at most a finite/countable # of points.

3. The pdf is defined to be

$$\boxed{f(x) = \begin{cases} F'(x), & \text{if } F(x) \text{ is diff at } x \\ 0, & \text{otherwise} \end{cases}}$$

SUPPORT

The "support" of a crv X is $\boxed{A = \{x : f(x) > 0\}}$.

PROPERTIES OF $f(x)$

1. $\boxed{f(x) \geq 0 \quad \forall x \in \mathbb{R}}$.

2. $\boxed{\int_{-\infty}^{\infty} f(x)\,dx = 1}$.

Proof. $\int_{-\infty}^{\infty} f(x)\,dx = \lim_{x\to\infty} F(x) - \lim_{x\to-\infty} F(x)$
$= 1 - 0$
$= 1.$ ∎

3. $\boxed{F(x) = \int_{-\infty}^{x} f(t)\,dt}$

4. $\boxed{P(X=x) = 0}$.

* For crv, $f(x) \neq P(X=x)$.

5. $\boxed{P(a < X \leq b) = \int_a^b f(x)\,dx}$.

$\hookrightarrow P(a<X\leq b) = P(a \leq X < b)$
$= P(a<X<b) = P(a \leq X \leq b)$

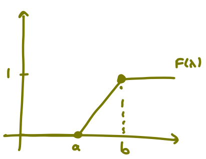
6. $\boxed{f(x) = \lim_{h\to 0} \frac{P(x<X<x+h)}{h}}$.

EXAMPLE 1

Suppose

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a<x<b \\ 1, & x \geq b, \end{cases}$$

ie F is the CDF of $X \sim \text{Unif}([a,b])$.
Find the pdf of $F(x)$.



$$\Rightarrow f(x) = \begin{cases} F'(x) = \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise.} \end{cases}$$

EXAMPLE 2

Suppose the pdf is

$$f(x) = \begin{cases} \frac{\theta}{x^{\theta+1}}, & x \geq 1 \\ 0, & x < 1. \end{cases}$$

Find
i) for what values of θ is f a pdf.
ii) $F(x)$
iii) $P(2<X<3)$ & $P(-2<X<3)$.

i) $f(x) \geq 0 \ \forall x \Rightarrow \theta \geq 0$.

$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{1} f(x)\,dx + \int_1^{\infty} f(x)\,dx$
$= \int_1^{\infty} \frac{\theta}{x^{\theta+1}}\,dx$
$= \left[-x^{-\theta}\right]_1^{\infty} \quad (=1)$.

This is true when $\theta > 0$.

$\Rightarrow \theta > 0$.

ii) $\theta > 0$, $F(x) = \int_{-\infty}^{x} f(t)\,dt$.

if $x \leq 1 \Rightarrow F(x) = \int_{-\infty}^{x} 0\,dt = 0$
if $x > 1 \Rightarrow F(x) = \int_{-\infty}^{x} f(t)\,dt$
$= \int_{-\infty}^{1} f(t)\,dt + \int_1^x f(t)\,dt$
$= \int_1^x f(t)\,dt$
$= \int_1^x \frac{\theta}{t^{\theta+1}}\,dt = \dots = 1 - x^{-\theta}$.

iii) $P(2<X<3) = F(3) - F(2)$
$= (1-3^{-\theta}) - (1-2^{-\theta})$
$= 2^{-\theta} - 3^{-\theta}$.

$P(-2<X<3) = F(3) - F(-2)$
$= 1 - 3^{-\theta} - 0$
$= 1 - 3^{-\theta}$.

GAMMA FUNCTION: $\Gamma(\alpha)$, $\alpha > 0$

1. The gamma function is defined to be

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

2. Properties:

① $\boxed{\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1), \quad \alpha > 1}$

② $\boxed{\Gamma(n) = (n-1)!, \quad n \in \mathbb{N}.}$

③ $\boxed{\Gamma(\tfrac{1}{2}) = \sqrt{\pi}}$

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

1. The "Gamma distribution" is defined by the pdf

$$\boxed{f(x) = \begin{cases} \dfrac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}}$$

2. Verify this is a pdf.

Proof. i) $f(x) \geq 0 \;\; \forall x \in \mathbb{R}$.

ii) $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_0^\infty f(x)\,dx$

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

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

Let $\frac{x}{\beta} = y \Rightarrow x = \beta y, \; dx = \beta\, dy.$

Then

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

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

$$= \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) = 1.$$

WEIBULL DISTN: $X \sim \text{Weibull}(\theta, \beta)$

1. The "Weibull distribution" is defined by the pdf

$$\boxed{f(x) = \begin{cases} \dfrac{\beta}{\theta^\beta} x^{\beta-1} e^{-(\frac{x}{\theta})^\beta}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}}$$

2. Verify f is a pdf.

Proof. Clearly $f(x) \geq 0 \;\; \forall x \in \mathbb{R}$.

Then

$$\int_{-\infty}^\infty f(x)dx = \int_0^\infty \frac{\beta}{\theta^\beta} x^{\beta-1} e^{-(\frac{x}{\theta})^\beta}\, dx.$$

Let $y = (\frac{x}{\theta})^\beta \Rightarrow x = \theta y^{1/\beta}, \; dx = \frac{\theta}{\beta} y^{\frac{1}{\beta}-1} dy.$

Then LHS becomes

$$\int_{-\infty}^\infty f(x)\,dx = \int_0^\infty \frac{\beta}{\theta^\beta} (\theta y^{1/\beta})^{\beta-1} e^{-y} \cdot \frac{\theta}{\beta} y^{\frac{1}{\beta}-1} dy$$

$$= \int_0^\infty y^{1-\frac{1}{\beta}} \cdot y^{\frac{1}{\beta}-1} \cdot e^{-y}\, dy$$

$$= \int_0^\infty e^{-y}$$

$$= \left[-e^{-y}\right]_0^\infty$$

$$= 1.$$

NORMAL DISTRIBUTION

1. The "normal distribution" has pdf

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \mu \in \mathbb{R}, \; \sigma^2 > 0.}$$

2. Steps to verify f is a pdf:

① Check $\int_{-\infty}^\infty f(x)\,dx = 1$ if $\mu = 0, \; \sigma^2 = 1$

② Check $\int_{-\infty}^\infty f(x)\,dx = 1$ when $\mu \in \mathbb{R}, \; \sigma^2 > 0$.

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

$= 2\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx$ (function is symmetrical about the y-axis)

Let $y = \frac{x^2}{2} \Rightarrow x = \sqrt{2y}, \; dx = \frac{\sqrt{2}}{2} y^{-\frac{1}{2}} dy.$ Then

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

$$= \frac{1}{\sqrt{\pi}} \int_0^\infty e^{-y} y^{\frac{1}{2}-1}\, dy$$

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

$$= \frac{1}{\sqrt{\pi}}(\sqrt{\pi}) = 1. \quad \#$$

②: Let $z = \frac{x-\mu}{\sigma}$. $\Rightarrow x = \mu + z\sigma, \; dx = \sigma\, dz$

$$\Rightarrow \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{z^2}{2}} \sigma\, dz$$

$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz$$

$$= 1 \quad (\text{by ①}). \quad \#$$

Chapter 2.3: Expectation and Variance

DRV DEFINITION OF E(X)

1. Suppose X is a drv, with support A & p.f. $f(x)$.
Then the "expectation" of X is given by

$$\boxed{E(X) = \sum_{x\in A} x f(x) \quad \text{if} \quad \sum_{x\in A} |x| f(x) < \infty.}$$

* this needs to be satisfied for the expectation to exist!

CRV DEFINITION OF E(X)

1. Suppose X is a crv, with support A & pdf $f(x)$.
Then the "expectation" of X is given by

$$\boxed{E(X) = \int_{-\infty}^{\infty} x f(x)\,dx \quad \text{if} \quad \int_{-\infty}^{\infty} |x| f(x)\,dx < \infty.}$$

if this is not satisfied, we say $E(X)$ does not exist!

CAUCHY DISTRIBUTION

1. The "Cauchy distribution" is defined by the pdf

$$\boxed{f(x) = \frac{1}{\pi(x^2+1)}.}$$

2. Find $E(X)$.

First, see that

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

$$= 2\int_0^{\infty} \frac{x}{\pi(x^2+1)}\,dx \quad \text{(even function)}$$

$$= \left[\frac{1}{\pi}\ln(x^2+1)\right]_0^{\infty}$$

$$= \infty.$$

In particular, the expectation of X does not exist.

DRV EXAMPLE 1

1. Suppose $f(x) = \frac{1}{x(x+1)}$, $x = 1, 2, \ldots$
See that $A = \{1, 2, \ldots\}$. Note f is a pdf.
Find $E(X)$.

$$E(X) = \sum_{x\in A} |x| f(x) = \sum_{x=1}^{\infty} x \cdot \frac{1}{x(1+x)}$$

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

so the expectation does not exist.

E(X) OF COMMON DISTNS

Bernoulli:

$$E(X) = \sum_{x\in A} x P(X=x) = 0P(X=0) + 1P(X=1) = p.$$

Binomial: Let

$$X_i = \begin{cases} 1, & i^{th} \text{ outcome is success} \\ 0, & \text{otherwise.} \end{cases} \Rightarrow X_i \sim Bern(p).$$

Then

$$X = \sum_{i=1}^{n} X_i$$

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

$$= \sum_{i=1}^{n} p$$

$$= np.$$

EXAMPLE 2

1. Suppose X has pdf

$$f(x) = \begin{cases} \frac{\theta}{x^{\theta+1}}, & x \geq 1 \\ 0, & x < 1, \end{cases}$$

where $\theta > 0$.
For what values of θ does $E(X)$ exist, and find $E(X)$?

Sol^{n}. We want to find θ s.t.

$$\int_{-\infty}^{\infty} |x| f(x)\,dx < \infty.$$

See that

$$\int_{-\infty}^{\infty} |x| f(x)\,dx = \int_1^{\infty} x f(x)\,dx$$

$$= \int_1^{\infty} x \cdot \frac{\theta}{x^{\theta+1}}\,dx$$

$$= \int_1^{\infty} \frac{\theta}{x^{\theta}}\,dx.$$

This is $<\infty$ iff $\theta > 1$.
Then

$$E(X) = \int_1^{\infty} \frac{\theta}{x^{\theta}}\,dx = \int_1^{\infty} \theta x^{-\theta}\,dx$$

$$= \left[\frac{\theta}{-\theta+1} x^{-\theta+1}\right]_1^{\infty}$$

$$= 0 - \frac{\theta}{-\theta+1}$$

$$= \frac{\theta}{\theta-1}.$$

EXPECTATION OF FUNCTIONS OF RV

1. Suppose X is a rv. What is $E[g(X)]$ for a real function g?

① DRV:

$$\boxed{E[g(X)] = \sum_{x\in A} g(x) f(x) \quad \text{if} \quad \sum_{x\in A} |g(x)| f(x) < \infty.}$$

② CRV:

$$\boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx \quad \text{if} \quad \int_{-\infty}^{\infty} |g(x)| f(x) < \infty.}$$

2. "Linearity" property:

$$\boxed{E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)].}$$

VARIANCE: $Var(X)$

The "variance" of X is defined to be

$$\boxed{\begin{aligned} Var(X) &= E[(X-\mu)^2], \quad \mu = E(X). \\ &= E(X^2) - [E(X)]^2 \end{aligned}}$$

MOMENTS

k^{TH} MOMENT ABOUT 0

The "k^{th} moment about 0" is $E(X^k)$.

k^{TH} MOMENT ABOUT MEAN / CENTRAL MOMENT

The "k^{th} moment about mean" is $E((X-\mu)^k)$,

where $\mu = E(X)$.

EXAMPLE 1: $X \sim Poi(\mu)$

Let $X \sim Poi(\mu)$; ie $f(x) = \frac{\mu^x}{x!} e^{-\mu}$, $x = 0, 1, \ldots$

Find $Var(X)$.

$$Var(X) = E(X^2) - [E(X)]^2.$$

$$E(X) = \sum_{x=0}^{\infty} x \cdot \frac{\mu^x}{x!} e^{-\mu}, \qquad x = 0, 1, \ldots$$

$$= \ldots = \mu.$$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \cdot \frac{\mu^x}{x!} e^{-\mu}$$

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

$$= \ldots = \mu^2 + \mu.$$

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

$$= \mu^2.$$

EXAMPLE 2: GAMMA DIST$^{\underline{N}}$

If $X \sim Gamma(\alpha, \beta)$, then

$$\boxed{E(X^k) = \frac{\beta^k \Gamma(\alpha+k)}{\Gamma(\alpha)}.}$$

Proof. Note

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

Thus

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

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

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

$$= \frac{\Gamma(\alpha+k)\beta^k}{\Gamma(\alpha)} \int_0^\infty \underbrace{\frac{x^{(k+\alpha-1)} e^{-\frac{x}{\beta}}}{\Gamma(\alpha+k)\beta^{k+\alpha}}}_{\text{pdf of } X_1 \sim Gamma(\alpha+k, \beta)} dx$$

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

$$= \frac{\Gamma(\alpha+k)\beta^k}{\Gamma(\alpha)}. \qquad \blacksquare$$

$$\boxed{Var(X) = \alpha\beta^2.}$$

Proof.

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

$$= \frac{\Gamma(\alpha+2)\beta^2}{\Gamma(\alpha)} - \left(\frac{\Gamma(\alpha+1)\beta}{\Gamma(\alpha)}\right)^2$$

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

$$= \alpha\beta^2$$

Chapter 2.4: Moment Generating Functions

Let X be a r.v. Then the "moment generating function" of X is defined to be

$$\boxed{M(t) = E(e^{tX}).}$$

if it exists for $t \in (-h, h)$ where $h > 0$.

EXAMPLE 1

$\boxed{\text{Let } X \sim Gamma(\alpha, \beta). \text{ Find } M(t).}$

$$M(t) = E(e^{tX})$$
$$= \int_0^\infty e^{tx} \cdot \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha} dx$$
$$= \int_0^\infty \frac{x^{\alpha-1} e^{-(\frac{1}{\beta}-t)x}}{\Gamma(\alpha)\beta^\alpha} dx$$

Approach ①:

let $y = (\frac{1}{\beta} - t)x \Rightarrow x = \frac{1}{\frac{1}{\beta}-t} y$, $dx = \frac{1}{\frac{1}{\beta}-t} dy$

Then

$$M(t) = \int_0^\infty \frac{\left(\frac{1}{\frac{1}{\beta}-t}\right)^{\alpha-1} y^{\alpha-1} e^{-y}}{\Gamma(\alpha)\beta^\alpha} \cdot \frac{1}{\frac{1}{\beta}-t} dy$$

$$= \frac{\left(\frac{\beta}{1-\beta t}\right)^\alpha}{\beta^\alpha} \underbrace{\int_0^\infty \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)} dy}_{\text{gamma function}}$$

$$= \frac{\left(\frac{\beta}{1-\beta t}\right)^\alpha}{\beta^\alpha}(1)$$

$$\boxed{M(t) = \left(\frac{1}{1-\beta t}\right)^\alpha}.$$

(we need $t < \frac{1}{\beta}$ for $M(t) > 0$).

Approach ②:

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

Let $\gamma = \frac{1}{\beta} - t$, $\Rightarrow \frac{1}{\gamma} = \frac{1}{\frac{1}{\beta}-t} = \frac{\beta}{1-\beta t}$.

Then

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

this is the pdf of $Gamma(\alpha, \frac{1}{\gamma})$.

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

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

$$= \frac{\left(\frac{\beta}{1-\beta t}\right)^\alpha}{\beta^\alpha} = \left(\frac{1}{1-\beta t}\right)^\alpha. \quad (\text{where } t < \tfrac{1}{\beta}).$$

EXAMPLE 2

If $X \sim Poi(\mu)$, find $M(t)$.

$$M(t) = E(e^{tX})$$
$$= \sum_{x=0}^\infty e^{tx} \cdot \frac{\mu^x}{x!} e^{-\mu}$$
$$= e^{-\mu} \sum_{x=0}^\infty \frac{(\mu \cdot e^t)^x}{x!}$$
$$= e^{-\mu} e^{\mu \cdot e^t}$$
$$= e^{\mu(e^t - 1)}.$$

This is finite $\forall t \in \mathbb{R}$.

EXAMPLE 3

Find:

① Let $Z \sim N(0,1)$. Find $M_Z(t)$.

② Let $X \sim N(\mu, \sigma^2)$. Find $M_X(t)$.

① $Z \sim N(0,1)$. Then

$$E(e^{tZ}) = \int_{-\infty}^\infty e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$
$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2-2tz}{2}} dz$$
$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2-t^2}{2}} dz$$
$$= e^{\frac{t^2}{2}} \underbrace{\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} dz}_{\text{pdf of } N(t,1)}$$
$$= e^{t^2/2}(1)$$
$$= e^{t^2/2}$$

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

$\Rightarrow X = \mu + \sigma Z$.

$\Rightarrow M_X(t) = e^{\mu t} M_Z(\sigma t)$ (by below)

$$= e^{\mu t} e^{\frac{\sigma^2 t^2}{2}}$$
$$= e^{\mu t + \frac{\sigma^2 t^2}{2}}.$$

MGF OF LINEAR COMBINATIONS OF X

Let mgf of X be $M_X(t)$.

Let $Y = aX + b$. Then

$$\boxed{M_Y(t) = e^{bt} M_X(at).}$$

FINDING MOMENTS FROM MGF

Suppose X has mgf $M(t)$. Then

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

EXAMPLE 1: Gamma(α, β)

$Gamma(\alpha, \beta)$ has mgf $M(t) = (1-\beta t)^{-\alpha}$ $(t < \frac{1}{\beta})$.

Find $E(X)$ & $Var(X)$.

Proof.

$$E(X) = \frac{dM(t)}{dt}\Big|_{t=0}$$
$$= (-\alpha)(-\beta)(1-\beta t)^{-\alpha-1}\Big|_{t=0}$$
$$= \alpha\beta$$

$$E(X^2) = \frac{d^2M(t)}{dt^2}\Big|_{t=0}$$
$$= (-\alpha)(-\alpha-1)(-\beta)^2(1-\beta t)^{-\alpha-1}\Big|_{t=0}$$
$$= \alpha(\alpha+1)\beta^2.$$

$$\Rightarrow Var(X) = E(X^2) - E(X)^2$$
$$= \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2$$
$$= \alpha\beta^2.$$

EXAMPLE 2: Poisson(μ)

Let $X \sim Poi(\mu)$. Find $E(X)$ & $Var(X)$.

Prof. $M(t) = e^{\mu(e^t - 1)}$

$$\Rightarrow m'(t) = \mu e^t \cdot e^{\mu(e^t-1)}$$

$$\Rightarrow E(X) = m'(0) = \mu(1)e^{\mu(0)} = \mu.$$

$$\Rightarrow m^{(2)}(t) = \mu e^t e^{\mu(e^t-1)} + (\mu e^t)^2 e^{\mu(e^t-1)}$$

$$\Rightarrow E(X^2) = m^{(2)}(0) = \mu + \mu^2.$$

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

UNIQUENESS OF MGF

X & Y have the same mgf iff X & Y have the same distribution;

ie they have the same CDF.

EXAMPLE 1

X has mgf $M(t) = e^{\frac{t^2}{2}}$.

① Find mgf of $2X-1$.

② Find $E(Y)$ & $Var(Y)$.

③ What is the distⁿ of Y?

Prof. ① $m_Y(t) = e^{-t} m_X(2t)$

$$= e^{-t} e^{4t^2/2} = e^{-t+2t^2}.$$

② $E(Y) = m_Y'(0) = (t-1)e^{2t^2-t}\big|_{t=0} = -1$

$$E(Y^2) = m_Y''(0) = 4e^{2t^2-t} - (t-1)^2 e^{2t^2-t}\Big|_{t=0} = 5.$$

$$\therefore Var(Y) = E(Y^2) - E(Y)^2 = 5 - 1 = 4.$$

③ mgf of Y is $e^{\mu t} e^{\frac{\sigma^2 t^2}{2}}$

$$\Rightarrow Y \sim N(-1, 4).$$

Chapter 3: Joint Distributions

JOINT CDF (3.1)

JOINT CDF

1. Suppose X & Y are 2 rvs. The "joint cdf" of X & Y is

$$\boxed{F(x,y) = P(X\le x,\ Y\le y), \quad x,y\in\mathbb{R}.}$$

2. This definition can be extended to n rvs $X_1,\dots,X_n$:

$$\boxed{F(x_1,\dots,x_n) = P(X_1\le x_1,\dots,X_n\le x_n).}$$

↳ for this course we focus on joint cdf for 2 rvs.

PROPERTIES OF JOINT CDF

1. If y is fixed, $F(x,y)$ is a non-decreasing function of x.
   If x is fixed, $F(x,y)$ is a non-decreasing function of y.

2. $$\boxed{\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0.}$$
   Idea. $\{X\le x\}\cap\{Y\le y\}\subseteq\{X\le x\}$.

3. $$\boxed{\lim_{x\to\infty} F(x,y) = \lim_{y\to\infty} F(x,y) = 1.}$$
   Idea. If $A_x=\{X\le x\}$ & $B_y=\{Y\le y\}$, then
   $\lim_{x\to\infty} P(A_x)=1$ & $\lim_{y\to\infty} P(B_y)=1$.
   We note
   $\overline{A_x\cap B_y} = \overline{A_x}\cup\overline{B_y}$ (by DeMorgan's law)
   ...

4. $$\boxed{\begin{aligned} F_1(x) &= P(X\le x) = \lim_{y\to\infty} F(x,y) \\ F_2(y) &= P(Y\le y) = \lim_{x\to\infty} F(x,y) \end{aligned}}$$

JOINT DRV (3.2)

1. If both X & Y are discrete r.v., then they are joint discrete r.v.

JOINT PF

1. The "joint pf" of X & Y is given by

$$\boxed{f(x,y) = P(X=x,\ Y=y), \quad x,y\in\mathbb{R}.}$$

2. The support set is given by

$$\boxed{A = \{(x,y):\ f(x,y)>0\}.}$$

PROPERTIES OF JOINT PF

1. $\boxed{f(x,y)\ge 0}$
2. $$\boxed{\sum_{(x,y)\in A} f(x,y) = 1.}$$
3. $$\boxed{\sum_{(x,y)\in C} f(x,y) = P((x,y)\in C)}$$

MARGINAL PF FROM JOINT PF

1. Let $f(x,y)$ be the joint pdf of X & Y. Then the "marginal probability function" of X is

$$\boxed{\begin{aligned} f_1(x) = P(X=x) &= P(X=x,\ Y<\infty) \\ &= \sum_{y\in\mathbb{R}} f(x,y). \end{aligned}}$$

The "marginal probability function" of Y is

$$\boxed{\begin{aligned} f_2(y) = P(Y=y) &= P(X<\infty,\ Y=y) \\ &= \sum_{x\in\mathbb{R}} f(x,y). \end{aligned}}$$

EXAMPLE 1

1. Let X & Y be drv with joint pf

$$f(x,y) = kq^2p^{x+y}, \quad x,y=0,1,\dots, \quad 0<p<1,\ q=1-p.$$

① Find k;
② Find marginal pf of X & Y; &
③ Find $P(X\le Y)$.

Soln.

① Note $f(x,y)\ge 0 \Rightarrow k\ge 0$.
Then

$$\sum_{(x,y)\in A} f(x,y) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} kq^2p^{x+y} = 1$$

$$\Rightarrow kq^2\Big(\sum_{x=0}^{\infty}p^x\Big)\Big(\sum_{y=0}^{\infty}p^y\Big) = 1$$

$$\Rightarrow \frac{kq^2}{(1-p)^2} = k = 1.$$

② $$f_1(x) = P(X=x) = \sum_{y=0}^{\infty} f(x,y) = \sum_{y=0}^{\infty} q^2p^xp^y = q^2p^x\sum_{y=0}^{\infty}p^y = qp^x, \quad x=0,1,\dots$$

$f_2(x) = qp^y$ by symmetry

③ Let $C=\{(x,y)\mid x\le y\}$. Then

$$P(X\le Y) = P((x,y)\in C) = \sum_{x=0}^{\infty}\sum_{y=x}^{\infty} q^2p^{x+y}$$

$$= \sum_{x=0}^{\infty} q^2p^x \sum_{y=x}^{\infty} p^y$$

$$= \sum_{x=0}^{\infty} q^2p^x\Big(\frac{p^x}{1-p}\Big)$$

$$= q\sum_{x=0}^{\infty}(p^2)^x = q\Big(\frac{1}{1-p^2}\Big)$$

$$= \frac{1}{1+p}.$$

JOINT CRV (3-3)

1. If the joint cdf of (X,Y) can be written as

$$\boxed{F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(t,s)\, dt\, ds}$$

then X & Y are "joint continuous random variables" with joint pdf $f(x,y)$ by

$$\boxed{f(x,y) = \begin{cases} \dfrac{\partial^2 F(x,y)}{\partial x\, \partial y}, & \text{if it exists} \\ 0, & \text{otherwise} \end{cases}}$$

2. The support set of f is given by

$$\boxed{A = \{(x,y) : f(x,y) > 0\}.}$$

PROPERTIES OF f(x,y)

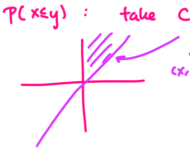
1. $\boxed{f(x,y) \geq 0}$

2. $\boxed{\int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)\, dx\, dy = 1.}$

3. For any region C, we have that

$$\boxed{P((X,Y) \in C) = \iint_{(x,y)\in C} f(x,y)\, dx\, dy.}$$

eg $P(X \leq Y)$: take $C = \{(x,y) \mid x \leq y\}$



$\iint_{(x,y)\in C} f(x,y)\, dx\, dy$

MARGINAL PDF FROM JOINT PDF

The "marginal pdf for X" is given by

$$\boxed{f_1(x) = \int_{-\infty}^{\infty} f(x,y)\, dy,}$$

& the "marginal pdf for Y" is given by

$$\boxed{f_2(y) = \int_{-\infty}^{\infty} f(x,y)\, dx.}$$

EXAMPLE 1

X & Y are joint crv with pdf

$$\boxed{f(x,y) = \begin{cases} x+y, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}}$$

① Show f is a joint pdf;

② Find

a) $P(X \le \frac{1}{3}, Y \le \frac{1}{2})$

b) $P(X \le Y)$

c) $P(X+Y \le \frac{1}{2})$

d) $P(XY < \frac{1}{2})$

③ Find marginal pdfs of X & Y.

Soln. ① We note $f(x,y) \ge 0 \quad \forall x,y \in \mathbb{R}$. ✓

Does $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$?

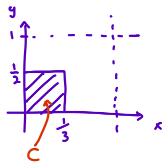
This equals

$$\int_0^1\int_0^1 (x+y)\,dx\,dy = \int_0^1 \left[\left(\frac{x^2}{2}+yx\right)\right]_{x=0}^1 dy$$
$$= \int_0^1 \left(\frac{1}{2}+y\right) dy$$
$$= \left[\frac{1}{2}y + \frac{y^2}{2}\right]_0^1$$
$$= 1 \quad ✓$$

So f is a joint pdf.

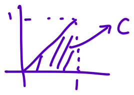
②

a) $P(X<\frac{1}{3}, Y<\frac{1}{2}) = P((X,Y) \in C)$



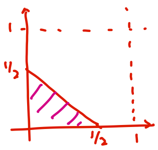
$$= \int_0^{1/2}\int_0^{1/3} x+y\,dx\,dy$$
$$= \int_0^{1/2} \left[\frac{x^2}{2}+yx\right]_0^{\frac{1}{3}} dy$$
$$= \int_0^{1/2} \frac{1}{18} + \frac{y}{3}\,dy$$
$$= \left[\frac{1}{36} + \frac{y^2}{6}\right]_0^{\frac{1}{2}} = \frac{1}{36} + \frac{1}{24} = \frac{5}{72}.$$

b) $P(X<Y) = \int_0^1\int_x^1 x+y\,dx\,dy$



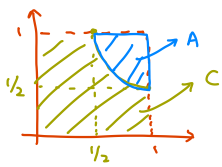
$$= \int_0^1 \left[xy + \frac{y^2}{2}\right]_{y=x}^{y=1} dx$$
$$= \int_0^1 \left(x+\frac{1}{2}\right) - \left(x^2 + \frac{x^2}{2}\right) dx$$
$$= \int_0^1 \left(x+\frac{1}{2} - \frac{3}{2}x^2\right) dx$$
$$= \left[\frac{x^2}{2} + \frac{x}{2} - \frac{x^3}{2}\right]_0^1$$
$$= \frac{1}{2} + \frac{1}{2} - \frac{1}{2} = \frac{1}{2}.$$

c) $P(X+Y \le \frac{1}{2})$



$$= \int_0^{\frac{1}{2}} dx \int_0^{\frac{1}{2}-x} (x+y)\,dy$$
$$= \int_0^{1/2} \left[xy + \frac{y^2}{2}\right]_{y=0}^{y=\frac{1}{2}-x} dx$$
$$= \int_0^{1/2} x\left(\frac{1}{2}-x\right) + \frac{1}{2}\left(\frac{1}{2}-x\right)^2 dx$$
$$= \cdots$$
$$= \frac{1}{24}.$$

d) $P(XY \le \frac{1}{2})$



First, we find

$$P((X,Y) \in A) = \int_{1/2}^1 dx \int_{\frac{1}{2x}}^1 \underline{(x+y)}\,dy$$

this is the pdf.

$$= \int_{1/2}^1 \left[\left(xy + \frac{y^2}{2}\right)\right]_{y=\frac{1}{2x}}^{y=1} dx$$
$$= \int_{\frac{1}{2}}^1 \left(x+\frac{1}{2}\right) - \left(\frac{1}{2} + \frac{1}{8x^2}\right) dx$$
$$= \int_{\frac{1}{2}}^1 \left(x - \frac{1}{8x^2}\right) dx$$
$$= \left[\frac{x^2}{2} + \frac{1}{8x}\right]_{\frac{1}{2}}^1$$
$$= \frac{1}{2} + \frac{1}{8} - \frac{1}{8} - \frac{1}{4} = \frac{1}{4}.$$

Hence $P(XY \le \frac{1}{2}) = 1 - P((X,Y) \in A) = \frac{3}{4}$.

③ Marginal pdf of X:

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \qquad 0<x<1$$
$$= \int_0^1 (x+y)\,dy$$
$$= \left[\left(xy + \frac{y^2}{2}\right)\right]_0^1$$
$$= x + \frac{1}{2}, \qquad 0<x<1$$

(for other x, $f_1(x) = 0$).

$f_2(y) = y + \frac{1}{2}$, $0<y<1$ by symmetry.

EXAMPLE 2

Let

$$f(x,y) = \begin{cases} ke^{-x-y}, & 0<x<y<\infty \\ 0, & \text{otherwise} \end{cases}$$

be the joint pdf of (X,Y).

① Find k.

② Find

a) $P(X \le \frac{1}{3}, Y \le \frac{1}{2})$

b) $P(X \le Y)$

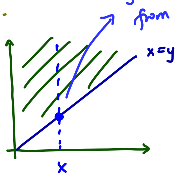
c) $P(X+Y \ge 1)$

③ Find the marginal pdf of X & Y.

④ Find the distribution of $T = X+Y$.

Sol^n. ① $f(x,y) \ge 0 \quad \forall x,y \in \mathbb{R} \Rightarrow k \ge 0.$

given x, y can range from $[x, \infty)$.



Then

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy$$

$$= \int_0^{\infty} dx \int_x^{\infty} ke^{-x-y}\,dy$$

$$= \int_0^{\infty} ke^{-x}\left[-e^{-y}\right]_x^{\infty} dx$$

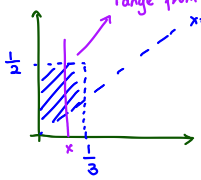
$$= \int_0^{\infty} ke^{-x}e^{-x}\,dx$$

$$= \left[-\frac{k}{2}e^{-2x}\right]_0^{\infty}$$

$$= \frac{k}{2} \quad (=1) \quad \Rightarrow \quad k=2.$$

②

given $x=x$, y can range from $[x, \frac{1}{2}]$



a) $P(X \le \frac{1}{3}, Y \le \frac{1}{2})$

$$= \int_0^{1/3} dx \int_x^{1/2} 2e^{-x-y}\,dy$$

$$= \int_0^{1/3}\left[2e^{-x}(-e^{-y})\right]_x^{1/2} dx$$

$$= \int_0^{1/3} 2e^{-x}(e^{-x} - e^{-\frac{1}{2}})\,dx$$

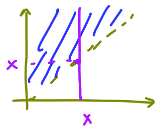
$$= \int_0^{1/3} (2e^{-2x} - 2e^{-\frac{1}{2}}e^{-x})\,dx$$

$$= \left[-e^{-2x} + 2e^{-\frac{1}{2}}e^{-x}\right]_0^{1/3}$$

$$= (-e^{-\frac{2}{3}} + 2e^{-\frac{5}{6}}) - (-1 + 2e^{-\frac{1}{2}})$$

$$= 1 - e^{-\frac{2}{3}} - 2e^{-\frac{1}{2}} + 2e^{-\frac{5}{6}}.$$

b) $P(X \le Y)$

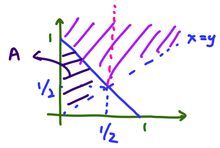


$$= \iint_{x \le y} f(x,y)\,dx\,dy = 1$$

(by construction).

c) $P(X+Y \ge 1)$

First, consider



$$P(X+Y<1) = P((X,Y)\in A)$$

$$= \int_0^{1/2} dx \int_x^{1-x} 2e^{-x-y}\,dy$$

$$= \int_0^{1/2} 2e^{-x}\left[-e^{-y}\right]_x^{1-x} dx$$

$$= \int_0^{1/2} 2e^{-x}(e^{-x} - e^{x-1})\,dx$$

$$= \int_0^{1/2} 2e^{-2x} - 2e^{-1}\,dx$$

$$= \left[-e^{-2x} - 2e^{-1}x\right]_0^{1/2}$$

$$= 1 - e^{-1} - e^{-1} = 1 - 2e^{-1}.$$

$$\therefore P(X+Y \ge 1) = 1 - P(X+Y<1) = 2e^{-1}.$$

③ Marginal pdf of X is

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \quad x>0 \qquad (\text{since support of } X \text{ is } (0,\infty))$$

$$= \int_x^{\infty} f(x,y)\,dy \qquad (\text{since given } x,\ f(x,y)>0 \Leftrightarrow y>x)$$

$$= \int_x^{\infty} 2e^{-x-y}\,dy$$

$$= 2e^{-x}\left[-e^{-y}\right]_x^{\infty}$$

$$f_1(x) = 2e^{-2x}, \quad x>0.$$

marginal pdf of Y is

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)\,dx, \quad y>0$$

$$= \int_0^{y} f(x,y)\,dx \qquad (\text{since given } y,\ f(x,y)>0 \Leftrightarrow 0<x\le y)$$

$$= \int_0^{y} 2e^{-x-y}\,dx$$

$$= 2e^{-y}\left[e^{-x}\right]_0^{y}$$

$$= 2e^{-y}(1-e^{-y}).$$

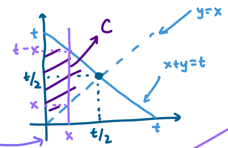
④ $T = X+Y$.

Support of T is $(0,\infty)$.

The CDF of T is

$$F_T(t) = P(T \le t) = 0 \quad \text{if } t \le 0.$$

For $t>0$:



Given a specific x, $y \in [x, t-x]$.

$$F_T(t) = P(T \le t)$$

$$= P(X+Y \le t)$$

$$= P((X,Y) \in C)$$

$$= \int_0^{t/2} dx \int_x^{t-x} f(x,y)\,dy$$

x can range from $[0, t/2]$.

$$= \int_0^{t/2} dx \int_x^{t-x} 2e^{-x-y}\,dy$$

$$= \cdots$$

$$= 1 - e^{-t} - te^{-t}.$$

* this is a very useful technique!

Thus

$$F_T(t) = \begin{cases} 0, & t \le 0 \\ 1-e^{-t}-te^{-t}, & t>0 \end{cases}$$

pdf of T is

$$f_T(t) = \begin{cases} 0, & t \le 0 \\ te^{-t}, & t>0 \end{cases}$$

INDEPENDENCE (3.4)

For any 2 rv, we say X & Y are "independent"
iff

$$\boxed{P(X\in A,\ Y\in B) = P(X\in A)P(Y\in B).}$$

for any $A, B \subseteq \mathbb{R}$.

EQUIVALENT DEFINITIONS OF INDEPENDENCE

1. Let $A = (-\infty, x)$, $B = (-\infty, y)$, $x, y \in \mathbb{R}$:
Then our definition becomes

$$\boxed{F(x,y) = F_1(x)\,F_2(y)}$$

and this is true iff X & Y are independent.

2. Suppose X, Y have joint pf (discrete case) or pdf (continuous case) $f(x,y)$, & marginal pf/pdfs $f_1(x)$ & $f_2(y)$.

Then X & Y are independent iff

$$\boxed{f(x,y) = f_1(x)f_2(y) \quad \forall x, y \in \mathbb{R}.}$$

Proof. Take $\frac{\partial^2}{\partial x \partial y}$ of both sides.

PROPERTIES OF INDEPENDENCE

If X & Y are independent, then $g(X)$ & $h(Y)$ are independent for any g, h.

eg if X & Y are independent, then X^2 & Y^2 are independent.

However the converse is not necessarily true!

eg if X^2 & Y^2 are independent, X & Y may not be independent!

eg Consider $\underline{a}$, ie $P(\underline{a}=a)=1$, $P(\underline{a}\neq a)=0$.
$\underline{a}$ is independent of any rv.
Let

$$X = \begin{cases} 1, & \text{with prob. } 1/2 \\ -1, & \text{with prob. } 1/2 \end{cases}, \qquad Y = X.$$

Then X & Y are not independent, but X^2 & Y^2 are independent
(since $P(X^2=1) = P(Y^2=1) = 1$).

EXAMPLE 1 (DISCRETE)

1. Let X & Y have joint pdf.

$$\boxed{f(x,y) = q^2 p^{x+y}, \quad x, y = 0, 1, \ldots}$$

We showed

$f_1(x) = qp^x$, $x = 0, 1, \ldots$
$f_2(y) = qp^y$, $y = 0, 1, \ldots$.

Thus

$f(x,y) = f_1(x) f_2(y) \quad \forall x, y \in \mathbb{R}$,

and so X & Y are independent.

EXAMPLE 2 (CONTINUOUS)

Let X & Y have joint pdf

$$\boxed{f(x,y) = \begin{cases} x+y, & 0 \le x, y \le 1 \\ 0, & \text{otherwise.} \end{cases}}$$

We showed

$f_1(x) = x + \frac{1}{2}$, $1 \ge x \ge 0$
$f_2(y) = y + \frac{1}{2}$, $1 \ge y \ge 0$.

Thus

$f(x,y) \neq f_1(x) f_2(y)$ if $0<x<1$, $0<y<1$,

and so X & Y are not independent.

FACTORIZATION THEOREM FOR INDEPENDENCE

1. ① We can express

$$\boxed{f(x,y) = g(x)\,h(y)}.$$

② Let A denote the support of (X, Y), let A_1 denote the support of X & let A_2 denote the support of Y.

Then

$$\boxed{A = A_1 \times A_2 = \{(x,y) \mid x \in A_1, y \in A_2\}.}$$

$\iff$ A is a rectangle
$\iff$ the range of X does not depend on the value of y
$\iff$ the range of Y does not depend on the value of x

equiv statements for ②.

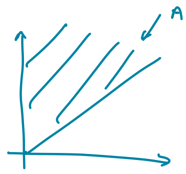
Both conditions are true iff X & Y are independent.

2. Let $f_1(x)$ & $f_2(y)$ be the marginal pfs/pdfs of X & Y, and say they are independent.
Then there exist constants $c, d \in \mathbb{R}$ such that

$$\boxed{\begin{aligned} f_1(x) &= cg(x) \\ f_2(y) &= dh(y) \end{aligned}}.$$

EXAMPLE 1

Suppose the joint support is $0 \le x \le y < \infty$.
This is not a rectangle; thus X & Y cannot be independent by the above.



EXAMPLE 2 : DRV

Let $f(x,y) = \dfrac{\mu^{x+y} e^{-2\mu}}{x!\,y!}$, $x, y = 0, 1, \ldots$

① Are X & Y independent?

② Find the marginal pfs of X & Y.

① $f(x,y) = \dfrac{\mu^x e^{-\mu}}{x!} \dfrac{\mu^y e^{-y}}{y!}$.

let $g(x) = \dfrac{\mu^x e^{-\mu}}{x!}$, $h(y) = \dfrac{\mu^y e^{-\mu}}{y!}$, so $f(x,y) = g(x)h(y)$.

(so condition ① is satisfied).

Then, note the range of X does not depend on the value of Y, and so the second condition holds.

$\therefore$ X & Y are independent.

② $f_1(x) = cg(x)$, for some constant c * corollary of the theorem!

$= c\dfrac{\mu^x e^{-\mu}}{x!}$.

We know $f_1(x) \ge 0 \ \forall x \Rightarrow c \ge 0$.

Then,

$$\sum_{x \in A_1} f_1(x) = \sum_{x=0}^{\infty} c \cdot \frac{\mu^x e^{-\mu}}{x!} = 1$$

$$\iff c \sum_{x=0}^{\infty} \frac{\mu^x e^{-\mu}}{x!} = 1 \qquad \text{pf of Poisson rv.}$$

$$\iff c(1) = 1. \qquad \therefore c = 1.$$

Thus

$f_1(x) = \dfrac{e^{-\mu}\mu^x}{x!}$.

Using a similar proof,

$f_2(y) = d\,h(y) = \dfrac{e^{-\mu}\mu^y}{y!}$.

EXAMPLE 3 : CRV

X & Y have joint pdf

$$f(x,y) = \frac{3}{2}y(1-x^2), \quad -1 \le x \le 1, \quad 0 \le y \le 1.$$

① Are X & Y independent?

② Find the marginal pdf of X & Y.

① Let $g(x) = 1-x^2$ & $h(y) = \frac{3}{2}y$.

$\Rightarrow f(x,y) = g(x)\,h(y)$ (so condition ① is true).

The range of X does not depend on the value of y (so condition ② is true).

② $f_1(x) = c_1 g(x)$, support of $X = [-1,1]$.

$f_1(x) \ge 0 \ \forall x \Rightarrow c_1 \ge 0$.

Then

$$\int_{-\infty}^{\infty} f_1(x)\,dx = c_1 \int_{-1}^{1} (1-x^2)\,dx$$

$$= c_1 \left[x - \frac{x^3}{3}\right]_{-1}^{1}$$

$$= \frac{4}{3}c_1 \quad (=1).$$

$$\therefore c_1 = 3/4.$$

$$\therefore f_1(x) = \frac{3}{4}(1-x^2).$$

$f_2(y) = c_2 h(y)$, support of Y is $[0,1]$.

Then

$$\int_{-\infty}^{\infty} f_2(y)\,dy = c_2 \int_0^1 \frac{3}{2}y\,dy$$

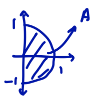
$$= \cdots \quad (=1)$$

$$\therefore c_2 = \frac{4}{3}.$$

$$\therefore f_2(y) = \frac{4}{3}\cdot\frac{3}{2}y = 2y.$$

EXAMPLE 4

Suppose $f(x,y)$ is constant over the region A, say



$$f(x,y) = c_0.$$

$$\Rightarrow \iint_A f(x,y)\,dx\,dy = 1.$$

$$\Rightarrow c_0 \underbrace{\iint_A 1\,dx\,dy}_{\text{area of } A} = 1.$$

$\Rightarrow c_0\left(\frac{\pi}{2}\right) = 1 \quad \therefore c_0 = \frac{2}{\pi}$.

Thus $f(x,y) = \frac{2}{\pi}$.

① Are X & Y independent?

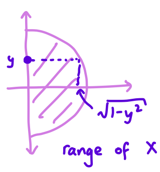
② Find $f_1(x)$ & $f_2(y)$.

① $f(x,y) = \frac{2}{\pi}$. Let $g(x) = \begin{cases} 1, & 0<x<1 \\ 0, & \text{otherwise} \end{cases}$, &

$$h(y) = \begin{cases} \frac{2}{\pi}, & |y|<1 \\ 0, & \text{otherwise} \end{cases}$$

But see that the range of X depends on the value of y.

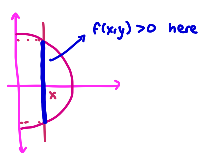
X & Y are not independent.



range of X is $[0, \sqrt{1-y^2}]$.

② $f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$, $\underbrace{0 \le x \le 1}_{\text{support of } X}$

Given $x \in [0,1]$, we know the possible values y can take is $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$.

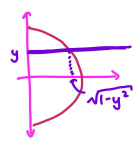


$$\Rightarrow f_1(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{2}{\pi}\,dy = \frac{4}{\pi}\sqrt{1-x^2}.$$

Similarly,

$f_2(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$, $y \in [-1,1]$ support of Y.

Given $y \in [-1,1]$, the possible values x can take is $x \in [0, \sqrt{1-y^2}]$.



$$\Rightarrow f_2(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{\sqrt{1-y^2}} f(x,y)\,dx$$

$$= \frac{2}{\pi}\sqrt{1-y^2}.$$

JOINT EXPECTATION (3.5)

Let $h(x,y)$ be a bivariate function.
Then we define the "joint expectation" of X & Y to be

$$E(h(X,Y)) = \begin{cases} \sum_x \sum_y h(x,y) f(x,y) & (X \,\&\, Y \text{ are joint discrete}) \\ \iint_{\mathbb{R}^2} h(x,y) f(x,y)\,dx\,dy & (X \,\&\, Y \text{ are joint continuous}) \end{cases}$$

PROPERTIES

1. Linearity:

$$\boxed{E(a g(X,Y) + b h(X,Y)) = a E(g(X,Y)) + b E(h(X,Y)).}$$

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

2. If X, Y are independent, then

$$\boxed{\begin{aligned} E(XY) &= E(X)E(Y) \\ E(g(X)h(Y)) &= E(g(X))E(h(Y)) \end{aligned}}$$

more generally, if $X_1, \dots, X_n$ are independent, then

$$\boxed{E\left(\prod_{i=1}^{n} h_i(X_i)\right) = \prod_{i=1}^{n} E(h_i(X_i)).}$$

COVARIANCE: $Cov(X,Y)$

1. The "covariance" of X & Y is defined to be

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

2. If X & Y are independent, then $Cov(X,Y) = 0$.
3. Also note $Cov(X,X) = Var(X)$.

VARIANCE FORMULAS

1. $$\boxed{Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X,Y).}$$
2. $$\boxed{Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) + \sum_{i \neq j} a_i a_j Cov(X_i, X_j).}$$
3. If $X_1, \dots, X_n$ are independent, then

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

EXAMPLE 1

Suppose the joint pf of X & Y is

$$f(x,y) = \frac{\mu^{x+y} e^{-2\mu}}{x!\,y!}, \quad x, y = 0, 1, \dots$$

Find $Var(2X+3Y)$.

Sol^{n}. $Var(2X+3Y) = 4Var(X) + 9Var(Y) + 12Cov(X,Y)$.

Take $X, Y \sim Poi(\mu)$, so that

pf of X, $g(x) = \frac{\mu^x e^{-\mu}}{x!}$, $x = 0, 1, \dots$

pf of Y, $h(y) = \frac{\mu^y e^{-\mu}}{y!}$, $y = 0, 1, \dots$

Note $f(x,y) = g(x)h(y)$, and the support of (X,Y) is independent of the range of X, and so by the factorization theorem X & Y are independent.

$\therefore Cov(X,Y) = 0$, and so

$$\begin{aligned} Var(2X+3Y) &= 4Var(X) + 9Var(Y) \\ &= 4\mu + 9\mu \\ &= 13\mu. \end{aligned}$$

EXAMPLE 2

Suppose X & Y have the joint pdf

$$f(x,y) = \begin{cases} x+y, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find $Var(X+Y)$.

Method #1: First, note that

$$f_1(x) = \begin{cases} x + 1/2, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases} \quad \& \quad f_2(y) = \begin{cases} y + 1/2, & 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$E(X) = \int_0^1 x\left(x + \tfrac{1}{2}\right) dx = \cdots = \tfrac{7}{12}.$$

$$E(X^2) = \int_0^1 x^2\left(x + \tfrac{1}{2}\right) dx = \cdots = \tfrac{5}{12}.$$

By symmetry, $E(Y) = \frac{7}{12}$ & $E(Y^2) = \frac{5}{12}$.

Then

$$Cov(X,Y) = \underbrace{E(XY)}_{\text{we need to find this.}} - E(X)E(Y).$$

$$\begin{aligned} \hookrightarrow E(XY) &= \int_0^1 dx \int_0^1 xy(x+y)\,dy \\ &= \int_0^1 \left[\frac{x^3 y}{3} + \frac{x^2 y^2}{2}\right]_0^1 dy \\ &= \int_0^1 \frac{y}{3} + \frac{y^2}{2}\,dy \\ &= \left[\frac{y^2}{6} + \frac{y^3}{6}\right]_0^1 \\ &= \frac{1}{3}. \end{aligned}$$

Thus

$$\begin{aligned} Var(X+Y) &= Var(X) + Var(Y) + 2Cov(X,Y) \\ &= \cdots \\ &= \frac{20}{144}. \end{aligned}$$

Method #2: Let $T = X+Y \Rightarrow Var(X+Y) = Var(T)$.

Support of T is $[0,2]$.

Consider $F_T(t) = P(X+Y \le t)$.

CORRELATION COEFFICIENT: $\rho(X,Y)$

1. The "correlation coefficient" of X & Y is

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

PROPERTIES OF ρ

1. Note $\boxed{|\rho(X,Y)| \le 1.}$

Proof. Suppose $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$.

Recall the inner product is

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i,$$

& we know $|\langle x,y \rangle| \le \sqrt{\langle x,x \rangle}\sqrt{\langle y,y \rangle}$ (Cauchy's inequality).

We can write ρ in the form "$\frac{\langle x,y \rangle}{\sqrt{\langle x,x \rangle \langle y,y \rangle}}$", so by the inequality, $\therefore |\rho(X,Y)| \le 1$.

2. Note:

① If $\rho(X,Y) = 1 \Rightarrow Y = aX + b$, $a > 0$.

② If $\rho(X,Y) = -1 \Rightarrow Y = aX + b$, $a < 0$.

EXAMPLE 1

Let $Y = Z^2$, $X = Z$, where $Z \sim N(0,1)$.

Then $\rho(Y,Z) = 0$.

EXAMPLE 2

Let

$$f(x,y) = \begin{cases} x+y, & 0 \le x, y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find $\rho(X,Y)$.

$\mathrm{Var}(X) = \frac{11}{144} = \mathrm{Var}(Y)$, $\quad \mathrm{Cov}(X,Y) = \frac{-1}{144}$.

$$\Rightarrow \rho(X,Y) = \frac{-1/144}{11/144} = \frac{-1}{11}.$$

CONDITIONAL DISTRIBUTION (3.6)

JOINT DISCRETE CASE

1. Let X, Y be joint drvs with joint pf $f(x,y)$.
Then the "conditional pf of X given $Y=y$" is

$$\boxed{f_1(x|y) = \frac{f(x,y)}{f_2(y)}, \quad f_2(y) > 0.}$$

The "conditional pf of Y given $X=x$" is

$$\boxed{f_2(y|x) = \frac{f(x,y)}{f_1(x)}, \quad f_1(x) > 0.}$$

2. We can prove $f_1(x|y)$ & $f_2(y|x)$ are pfs.
Proof. ① First, we need to show $f_1(x|y) \geq 0 \quad \forall x \in \mathbb{R}$.
② Then, we need to show $\sum_x f_1(x|y) = 1$.
(Proof for $f_2(y|x)$ is symmetric.)

JOINT CONTINUOUS CASE

1. Let X & Y be joint crvs with joint pdf $f(x,y)$.
Then the "conditional pdf of X given $Y=y$" is

$$\boxed{f_1(x|y) = \frac{f(x,y)}{f_2(y)}, \quad f_2(y) > 0.}$$

The "conditional pdf of Y given $X=x$" is

$$\boxed{f_2(y|x) = \frac{f(x,y)}{f_1(x)}, \quad f_1(x) > 0.}$$

2. We can also show these are pdfs; that is

① $f_1(x|y) \geq 0 \quad \forall x \in \mathbb{R}$, $f_2(y|x) \geq 0 \quad \forall y \in \mathbb{R}$; &
② $\int_{-\infty}^{\infty} f_1(x|y)\,dx = \int_{-\infty}^{\infty} f_2(y|x)\,dy = 1.$

EXAMPLE 1

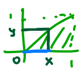
Suppose

$$f(x,y) = \begin{cases} 8xy, & 0<y<x<1 \\ 0, & \text{otherwise.} \end{cases}$$

Find $f_1(x|y)$ & $f_2(y|x)$.

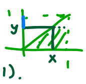
Sol^{n}. $f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^x 8xy\,dy$

$= 4x^3, \quad x \in (0,1)$



$f_2(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_y^1 8xy\,dy$

$= 4y - 4y^3, \quad y \in (0,1).$



Then,

$f_1(x|y) = \frac{f(x,y)}{f_2(y)}, \quad y \in (0,1).$

$= \frac{8xy}{4y-4y^3}, \quad y<x<1.$

this is the support of this conditional dist^{n}. (since otherwise, $f_2(y) = 0$.)

* remember y is fixed.

Similarly,

$f_2(y|x) = \frac{f(x,y)}{f_2(y)}, \quad x \in (0,1)$

$= \frac{8xy}{4x^3}, \quad 0<y<x.$ since otherwise $f_1(x) = 0$.

EXAMPLE 2

Let

$$f(x,y) = \begin{cases} x+y, & 0 \leq x,y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find $f_1(x|y)$ & $f_2(y|x)$.

Sol^{n}. We found $f_1(x) = x + \frac{1}{2}$, $x \in [0,1]$ & $f_2(y) = y + \frac{1}{2}$, $y \in [0,1]$.

$\Rightarrow f_1(x|y) = \frac{f(x,y)}{f_2(y)}, \quad y \in [0,1]$

$= \frac{x+y}{y+\frac{1}{2}}, \quad x \in [0,1].$

Similarly,

$f_2(y|x) = \frac{f(x,y)}{f_1(x)}, \quad x \in [0,1]$

$= \frac{x+y}{x+\frac{1}{2}}, \quad y \in [0,1].$

EXAMPLE 3

Let $f(x,y) = q^2 p^{x+y}$, $x, y = 0, 1, \ldots$, $q = 1-p$.
Find $f_1(x|y)$ & $f_2(y|x)$.

Sol^{n}. We found

$f_1(x) = qp^x, \quad x = 0, 1, \ldots$

$f_2(y) = qp^y, \quad y = 0, 1, \ldots.$

Then

$f_1(x|y) = \frac{f(x,y)}{f_2(y)}, \quad y = 0, 1, \ldots$

$= qp^x = f_1(x), \quad x = 0, 1, \ldots$

&

$f_2(y|x) = \frac{f(x,y)}{f_1(x)}, \quad x = 0, 1, \ldots$

$= qp^y, \quad f_2(y), \quad y = 0, 1, \ldots.$

USING CONDITIONAL DIST^{N} TO FIND INDEPENDENCE

1. X & Y are independent iff

$$\boxed{f_1(x|y) = f_1(x) \text{ or } f_2(y|x) = f_2(y).}$$

USE CONDITIONAL DIST^{N} TO FIND JOINT DIST^{N}

Note by definition,

$$\boxed{f(x,y) = f_1(x|y) \cdot f_2(y) = f_2(y|x) \cdot f_1(x).}$$

EXAMPLE 1: DISCRETE

Let $Y \sim \text{Poi}(\mu)$ & $(X|Y=y) \sim \text{Bin}(y,p)$.
Find $f_1(x)$.

Motivation: Let $Y :=$ # of students going to Tim's in one day.
Let $X :=$ # of female students among y visitors.
We could speculate $X \sim \text{Poi}(\mu p)$?

Note

$f(x,y) = f_1(x|y)\, f_2(y)$

$= \binom{y}{x} p^x (1-p)^{y-x} \cdot \frac{e^{-\mu}\mu^y}{y!}, \quad y = 0, 1, \ldots, \quad x = 0, 1, \ldots, y.$

$= \frac{y!}{x!(y-x)!} p^x (1-p)^{y-x} \cdot \frac{e^{-\mu}\mu^y}{y!},$

$= \frac{1}{x!(y-x)!} p^x (1-p)^{y-x} e^{-\mu} \mu^y.$

Then

$f_1(x) = \sum_y f(x,y)$

Given a particular x, $f(x,y) > 0 \iff x \leq y < \infty$.

$= \sum_{y=x}^{\infty} \frac{1}{x!(y-x)!} p^x (1-p)^{y-x} e^{-\mu} \mu^y$

$= \frac{e^{-\mu} p^x}{x!} \sum_{y=x}^{\infty} \frac{1}{(y-x)!} (1-p)^{y-x} \mu^y$

Let $\ell = y - x$. Then this becomes

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

$= \frac{e^{-\mu} p^x \mu^x}{x!} \underbrace{\sum_{\ell=0}^{\infty} \frac{[\mu(1-p)]^{\ell}}{\ell!}}$ taylor expansion of e^x

$= \frac{e^{-\mu} p^x \mu^x}{x!} [e^{\mu(1-p)}] = \frac{e^{-\mu p} (\mu p)^x}{x!}.$ (So $X \sim \text{Poi}(\mu p)$).

EXAMPLE 2: CONTINUOUS

Let Y have pdf

$$f_2(y) = \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}, \quad y>0$$

ie $Y \sim \text{Gamma}(\alpha, 1)$.

Let the conditional pdf of X given $Y=y$ be

$$f_1(x|y) = ye^{-xy}, \quad x>0, y>0.$$

Find the marginal pdf of X.

Sol$^{\underline{n}}$.

$$f(x,y) = f_1(x|y)\, f_2(y)$$
$$= ye^{-xy} \cdot \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}, \quad x>0, y>0$$
$$= \frac{y^{\alpha}e^{-(x+1)}}{\Gamma(\alpha)}.$$

The support of X is $(0,\infty)$, given $x>0$.

Thus for $x>0$:

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$
$$= \int_0^{\infty} \frac{y^{\alpha}e^{-(x+1)y}}{\Gamma(\alpha)}\,dy$$

Recall for $\text{Gamma}(\alpha,\beta)$, the pdf is $\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}$.

Let $\beta = \frac{1}{x+1}$, $\alpha^* = \alpha+1$. Then

$$f_1(x) = \int_0^{\infty} \underbrace{\frac{y^{\alpha^*-1}e^{-\frac{y}{\beta}}}{\Gamma(\alpha^*)\beta^{\alpha^*}}}_{\text{pdf of Gamma}(\alpha^*,\beta)}\,dy \cdot \frac{\Gamma(\alpha^*)}{\Gamma(\alpha)}\beta^{\alpha^*}$$

* note this gamma trick!

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

$$f_1(x) = \frac{\alpha}{(x+1)^{\alpha+1}}, \quad x>0.$$

CONDITIONAL EXPECTATION (3.7)

1. Note $f_2(y|x)$ is a pf if X & Y are joint discrete, & a pdf if X & Y are joint continuous.

2. So, we can define its expectation wrt $f_2(y|x)$.

3. The "conditional expectation" of $g(Y)$ given $X=x$ is defined by

$$\boxed{E[g(Y)\mid X=x] = \begin{cases} \sum_y g(y) f_2(y|x), & Y \text{ is discrete} \\ \int_{-\infty}^{\infty} g(y) f_2(y|x)\,dy, & Y \text{ is continuous} \end{cases}}$$

we need to check the respective expectations exist first before calculating them.

(ie using the absolute values)

4. We are interested in:

① $E(Y \mid X=x)$;

② $Var(Y \mid X=x) = E(Y^2 \mid X=x) - [E(Y \mid X=x)]^2$;

③ $E(e^{tY} \mid X=x)$.

COND. EXPECTATION UNDER INDEPENDENCE

1. If X & Y are independent, then

$$\boxed{\begin{aligned} E[g(Y) \mid X=x] &= E[g(Y)]. \\ E[h(X) \mid Y=y] &= E[h(X)]. \end{aligned}}$$

2. In particular, this implies

$$\boxed{\begin{aligned} &E(Y \mid X=x) = E(Y) \quad \& \\ &Var(Y \mid X=x) = Var(Y). \end{aligned}}$$

SUBSTITUTION RULE

1. Note that

$$\boxed{E[h(X,Y) \mid X=x] = E[h(x,Y) \mid X=x].}$$

this becomes univariate; it is only a function of Y.

eg[1] $E[X+Y \mid X=x]$

$= E[x+Y \mid X=x]$

$= E[x \mid X=x] + E[Y \mid X=x]$

$= x + E[Y \mid X=x]$.

eg[2] $E(XY \mid X=x)$

$= E(xY \mid X=x)$

$= xE(Y \mid X=x)$.

2. See that conditional expectation enjoys all properties of normal expectation.

EXAMPLE 1

$$\boxed{\begin{aligned} &\text{let} \\ &f(x,y) = \begin{cases} 8xy, & 0<y<x<1 \\ 0, & \text{otherwise.} \end{cases} \\ &\text{Find } E(X \mid Y=y) \ \& \ Var(X \mid Y=y). \end{aligned}}$$

Sol^n. We found that $f_1(x|y) = \frac{2x}{1-y^2}$, $0<y<1$, $y<x<1$.

Thus

$$E(X \mid Y=y) = \int_{-\infty}^{\infty} f_1(x|y)\, x\, dx$$

$$= \int_y^1 \frac{2x}{1-y^2}\, x\, dx$$

$$= \left[\tfrac{2}{3}x^3\right]_y^1 \cdot \frac{1}{1-y^2}$$

$$= \frac{2}{3}\left(\frac{1-y^3}{1-y^2}\right), \quad 0<y<1.$$

Then

$$E(X^2 \mid Y=y) = \int_{-\infty}^{\infty} f_1(x|y)\, x^2\, dx, \quad 0<y<1$$

$$= \int_y^1 \frac{2x}{1-y^2}\, x^2\, dx$$

$$= \left[\tfrac{1}{2}x^4\right]_y^1 \cdot \frac{1}{1-y^2}$$

$$= \frac{1}{2} \cdot \frac{1-y^4}{1-y^2} = \frac{1+y^2}{2}.$$

Hence

$$Var(X \mid Y=y) = E(X^2 \mid Y=y) - [E(X \mid Y=y)]^2$$

$$= \frac{1+y^2}{2} - \left(\frac{2}{3}\left(\frac{1-y^3}{1-y^2}\right)\right)^2.$$

EXAMPLE 2

$$\boxed{\begin{aligned} &\text{Suppose } Y \sim Poi(\mu), \ \& \ (X \mid Y=y) \sim Bin(y,p). \\ &\text{Find } E(X \mid Y=y) \ \& \ Var(X \mid Y=y). \end{aligned}}$$

Sol^n. We know $E(X \mid Y=y) = yp$ & $Var(X \mid Y=y) = yp(1-p)$.

$E[g(Y) \mid X]$

1. We define the random variable

$$\boxed{E[g(Y) \mid X] = h(X),}$$

where

$$\boxed{h(x) = E[g(Y) \mid X=x].}$$

(it is a rv since it is a function of X, denoted by $h(X)$).

2. To find $h(X)$, we do

① Find $E[g(Y) \mid X=x] = h(x)$.

② Replace "x" with "X".

eg $Y \sim Poi(\mu)$, $(X \mid Y=y) \sim Bin(y,p)$.

Then to find $E[X \mid Y]$:

① Note $E[X \mid Y=y] = yp$.

② Thus $E[X \mid Y] = Yp$.

DOUBLE EXPECTATION THEOREM

1. Note that

$$\boxed{E[g(Y)] = E\big[\underbrace{E(g(Y) \mid X)}_{\text{this is a function of } X.}\big]}$$

eg $E(Y) = E(E(Y \mid X))$.

2. Note also

$$\boxed{E[g(X,Y)] = E[E(g(X,Y) \mid Y)] = E[E(g(X,Y) \mid X)].}$$

EXAMPLE 1

$\boxed{\text{Let } Y \sim Poi(\mu) \ \& \ (X|Y=y) \sim Bin(y,p). \text{ Find } E(X).}$

Soln. (First method: calculate X's distn.)

We instead use the double exp theorem.

$$E(X) = E[E(X|Y)].$$

Then, note $E(X|Y=y) = yp$, so $E(X|Y) = Yp$.

Thus

$$E[E(X|Y)] = E[pY]$$
$$= pE[Y]$$
$$= \mu p.$$

DOUBLE EXPECTATION THEOREM FOR VARIANCE

Note that

$$\boxed{Var(Y) = E[\underbrace{Var(Y|X)}_{\text{this is a rv, as it is a func of } X}] + Var[E(Y|X)].}$$

* We apply a similar "two-step" method to find $Var(Y|X)$.

ie ① Calculate $Var(Y|X=x)$, & ② Replace x with X.

EXAMPLE 1

$\boxed{\text{Let } Y \sim Poi(\mu), \ (X|Y=y) \sim Bin(y,p). \text{ Find } Var(X).}$

Soln. $Var(X) = E[Var(X|Y)] + Var(E(X|Y))$.

Note $Var(X|Y=y) = yp(1-p)$, & so

$Var(X|Y) = Yp(1-p)$.

We also know $E(X|Y) = Yp$.

Therefore

$$Var(X) = E[Yp(1-p)] + Var(Yp)$$
$$= p(1-p)E[Y] + p^2 Var(Y)$$
$$= p(1-p)(\mu) + p^2(\mu)$$
$$= p\mu.$$

*Note: we showed earlier that $X \sim Poi(\mu p)$, so we would expect $Var(X) = \mu p$.

EXAMPLE 2

$\boxed{\text{Let } X \sim Unif[0,1], \ Y|X=x \sim Bin(10,x). \text{ Find } E[Y] \ \& \ Var(Y).}$

Soln. $E(Y) = E(E(Y|X))$.

Apply the 2-step method to find $E(Y|X)$:

① $E(Y|X=x) = 10x$, so ② $E(Y|X) = 10X$.

Then

$$E(Y) = E(10X) = 10E(X) = 10 \cdot \frac{0+1}{2} = 5.$$

Similarly,

$$Var(Y) = Var[E(Y|X)] + E[Var(Y|X)].$$
$$= Var(10X) + E[Var(Y|X)].$$

Note $Var(Y|X=x) = 10x(1-x)$, so $Var(Y|X) = 10X(1-X)$.

Thus

$$Var(Y) = Var(10X) + E[10X(1-X)]$$
$$= 10^2 Var(X) + 10E[X-X^2]$$
$$= 10^2 Var(X) + 10E[X] - 10E[X^2]$$
$$= 10^2 Var(X) + 10E[X] - 10[Var(X) + E(X)^2]$$
$$= 100 \cdot \frac{1}{12} + 10\left(\frac{1}{2}\right) - 10\left(\frac{1}{12} - \left(\frac{1}{2}\right)^2\right).$$

Soln 2. We could also calculate $M_Y(t) = E(e^{tY})$ first.

Then

$$E(e^{tY}) = E[E(e^{tY}|X)].$$

First note $E(e^{tY}|X=x) = \sum_{y=0}^{x} e^{ty} \cdot \binom{x}{y} p^y (1-p)^{x-y}$

$$= [pe^t + (1-p)]^x.$$

Thus $E(e^{tY}|X) = [pe^t + (1-p)]^X$.

Finally

$$M_X(t) = E[(pe^t + (1-p))^X]$$
$$= \sum_{x=0}^{\infty} (pe^t + (1-p))^x \cdot \frac{\mu^x}{x!} e^{-\mu}$$
$$= e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu(pe^t + (1-p)))^x}{x!}$$
$$= e^{-\mu} \cdot e^a, \quad a = \mu[pe^t + (1-p)]$$
$$= e^{-\mu + \mu[pe^t + (1-p)]}$$
$$= e^{\mu p(e^t - 1)},$$

so that $X \sim Poi(\mu p)$ by uniqueness of MGFs.

We can then use this to calculate $E(X)$ & $Var(X)$.

JOINT MGFS (3.8)

If X & Y are two rvs, then the "joint mgf" of X & Y is

$$\boxed{M(t_1, t_2) = E(e^{t_1 X + t_2 Y}),}$$

if $M(t_1,t_2)$ exists for $|t_1|<h_1$ & $|t_2|<h_2$ for some $h_1, h_2 > 0$.

MARGINAL MGF

Given $M(t_1,t_2)$ is well-defined for $|t_1|<h_1$, $|t_2|<h_2$, then the "marginal mgf" for X is

$$\boxed{M_X(t_1) = M(t_1, 0) = E(e^{t_1 X}), \quad |t_1|<h_1.}$$

The "marginal mgf" for Y is

$$\boxed{M_Y(t_2) = M(0, t_2) = E(e^{t_2 Y}), \quad |t_2|<h_2.}$$

INDEPENDENCE PROPERTY

X & Y are independent iff

$$\boxed{M(t_1,t_2) = M_X(t_1)\, M_Y(t_2)}$$

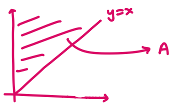
provided M, M_X & M_Y exist.

EXAMPLE 1

Let $f(x,y) = e^{-y}$, $0<x<y<\infty$.
a) Find the joint mgf of X & Y.
b) Are they independent?

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

$$= \iint e^{t_1 x + t_2 y} f(x,y)\, dx\, dy$$



$$= \iint_{(x,y)\in A} e^{t_1 x + t_2 y} \cdot e^{-y}\, dx\, dy$$

$$= \int_0^\infty dx \int_x^\infty e^{t_1 x + (t_2-1)y}\, dy$$

$$= \int_0^\infty e^{t_1 x} dx \underbrace{\int_x^\infty e^{(t_2-1)y}\, dy}$$

this is only finite if $t_2 - 1 < 0$; ie if $t_2 < 1$.

Suppose $t_2<1$. Then

$$M(t_1,t_2) = \int_0^\infty e^{t_1 x}\left[\frac{1}{t_2-1} e^{(t_2-1)y}\right]_x^\infty dx$$

$$= \int_0^\infty \frac{e^{t_1 x}}{1-t_2} e^{(t_2-1)x}\, dx$$

$$= \frac{1}{1-t_2}\underbrace{\int_0^\infty e^{(t_1+t_2-1)x}\, dx}$$

this is only finite if $t_1+t_2-1<0$.

we also need $1-t_2>0 \Rightarrow t_2<1$, for this to be positive.

Suppose $t_1+t_2-1<0$. Then

$$M(t_1,t_2) = \frac{1}{1-t_2}\left[\frac{1}{t_1+t_2-1} e^{(t_1+t_2-1)x}\right]_0^\infty$$

$$= \frac{1}{1-t_2} \cdot \frac{1}{1-t_1-t_2}, \quad t_1, t_2 < 1, \quad t_1+t_2<1.$$

b) Then, see that

$$M_X(t_1) = M(t_1, 0) = \frac{1}{1-t_1}, \qquad 1-t_1>0 \Rightarrow t_1<1$$

$$M_Y(t_2) = M(0,t_2) = \frac{1}{(1-t_2)^2}, \qquad 1-t_2>0 \Rightarrow t_2<1.$$

Then since $M(t_1,t_2) \neq M_X(t_1) M_Y(t_2)$, it follows that X & Y are not independent.

(This can also be seen via the factorization theorem, since the joint support is not a rectangle.)

EXAMPLE 2: ADDITIVITY OF POISSON VARIABLES

Let $X \sim Poi(\mu_1)$, $Y \sim Poi(\mu_2)$.
Assume X & Y are independent.
Then show $X+Y \sim Poi(\mu_1+\mu_2)$.

Soln. Let $Z = X+Y$. Then

$$M_Z(t) = E(e^{tZ})$$
$$= E(e^{t(X+Y)})$$
$$= E(e^{tX+tY}).$$
$$= M_X(t)\, M_Y(t), \quad \text{since } X \text{ \& } Y \text{ are independent}$$
$$= e^{\mu_1(e^t-1)} \cdot e^{\mu_2(e^t-1)}$$
$$= e^{(\mu_1+\mu_2)(e^t-1)},$$

so by uniqueness of mgfs thus $Z = (X+Y) \sim Poi(\mu_1+\mu_2)$, as needed. ☒

MULTINOMIAL DISTRIBUTION (3.9)

💡₁ We say $(X_1, \dots, X_k)$ has a "multinomial distribution" if it has joint pf

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

where $x_i = 0, 1, \dots, n$ & $\sum_{i=1}^{k} x_i = n$,
& $0 < p_i < 1$ & $\sum_{i=1}^{k} p_i = 1$,
and write

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

💡₂ If $k=2$, then if

$$\boxed{\begin{aligned} X_1 &:= \#\text{ of successes} \\ X_2 &:= \#\text{ of failures}, \end{aligned}}$$

then (X_1, X_2) follows a multinomial distribution with p_1 being the success probability.

JOINT MGF

💡 Suppose $(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, \dots, p_k)$.
Then the joint mgf is

$$\boxed{\begin{aligned} m(t_1, \dots, t_k) &= E(e^{t_1 X_1 + \cdots + t_k X_k}) \\ &= (p_1 e^{t_1} + \cdots + p_{k-1} e^{t_{k-1}} + p_k)^n \end{aligned}}$$

MARGINAL DISTRIBUTION

💡 Suppose $(X_1, \dots, X_n) \sim \text{Mult}(n, p_1, \dots, p_k)$.
Then

$$\boxed{X_i \sim \text{Bin}(n, p_i).}$$

for $i = 1, \dots, k$.

Why? Suppose we aggregate the outcomes to be

"success" = i^{th} outcome
"failure" = anything else.

This would give a "binomial" distⁿ.

SUMMATION OF X_i

💡 Let $T = X_i + X_j$, where $(X_1, \dots, X_n) \sim \text{Mult}(n, p_1, \dots, p_k)$.
Then

$$\boxed{T \sim \text{Bin}(n, p_i + p_j).}$$

Why? We can use a similar proof to the above.
Let

Success := i^{th} or j^{th} outcomes
failure := anything else.

Alternatively, we can find the MGF of T:

$$m_T(t) = E(e^{t(X_i + X_j)})$$
$$= E(e^{tX_i + tX_j})$$

WLOG, set $i=1$ & $j=2$.
In particular, see that

$$\underbrace{m(t, t, 0, \dots, 0)}_{\text{mgf of } (X_1, \dots, X_k)} = E(e^{t_1 X_1 + t_2 X_2 + 0 + \cdots + 0})$$
$$= E(e^{t_1 X_1 + t_2 X_2}).$$

Then

$$m(t, t, 0, \dots, 0) = (p_1 e^t + p_2 e^t + 0 + \cdots + 0 + \underline{(1 - p_1 - p_2 - 0 - \cdots - 0)})^n$$
$$= ((p_1 + p_2)e^t + (1 - p_1 - p_2))^n,$$

↳ since $p_k = 1 - p_1 - \cdots - p_{k-1}$

which is the MGF of $\text{Bin}(n, p_1 + p_2)$.
Thus, by uniqueness, the conclusion follows.

JOINT MOMENTS

💡₁ For $(X_1, \dots, X_n) \sim \text{Mult}(n, p_1, \dots, p_k)$, note

$$\boxed{\begin{aligned} E(X_i) &= np_i \\ \text{Var}(X_i) &= np_i(1 - p_i) \\ \text{Cov}(X_i, X_j) &= -np_i p_j. \end{aligned}}$$

We can do

$$\text{Var}(X_i + X_j) = \text{Var}(X_i) + \text{Var}(X_j) + 2\text{Cov}(X_i, X_j).$$

$$\therefore \text{Cov}(X_i, X_j) = \frac{1}{2}[\text{Var}(\underbrace{X_i + X_j}_{\sim \text{Bin}(n, p_i + p_j)}) - \text{Var}(\underbrace{X_i}_{\sim \text{Bin}(n, p_i)}) - \text{Var}(\underbrace{X_j}_{\sim \text{Bin}(n, p_j)})]$$

$$= \frac{1}{2}[n(p_i + p_j)(1 - p_i - p_j) - np_i(1 - p_i) - np_j(1 - p_j)]$$
$$= \dots$$
$$= -np_i p_j.$$

💡₂ Note $\text{Cov}(X_i, X_j) < 0$, so $\rho(X_i, X_j) < 0$.

Why? Note $\sum_{i=1}^{k} X_i = n$.
So when X_i increases, necessarily X_j decreases
(to keep the sum $= n$.)

$X_i \mid X_i + X_j = t$

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

$$\boxed{(X_i \mid X_i + X_j = t) \sim \text{Bin}\left(t, \frac{p_i}{p_i + p_j}\right).}$$

Why? → we have t independent trials of "$X_i + X_j = t$".
Then success probability is $\frac{\text{success of } X_i}{\text{success of } X_i + X_j}$.

$X_i \mid X_j = x_j$

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

$$\boxed{X_i \mid X_j = x_j \sim \text{Bin}\left(n - x_j, \frac{p_i}{1 - p_j}\right).}$$

Why? If $X_j = x_j$, then consider

success = i^{th} outcome
failure = everything else but the j^{th} outcome.

Then we have $n - x_j$ independent trials with "outcomes" $1, \dots, j-1, j+1, \dots, n$.

$$\text{sucess} = \frac{\text{success of } X_i}{\text{success of everything but } X_j} = \frac{p_i}{1 - p_j}.$$

BIVARIATE NORMAL DISTRIBUTION (3.10)

1. Let X_1, X_2 be joint continuous rvs with joint pdf

$$\boxed{f(x_1,x_2) = \frac{1}{2\pi\underbrace{|\Sigma|^{1/2}}_{\text{determinant of } \Sigma}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}}$$

where

$$\boxed{x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} E(X_1) \\ E(X_2) \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}}$$

& $|\rho| < 1$.

* $|\Sigma| = (1-\rho^2)\sigma_1^2\sigma_2^2$.

Then $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ follows a "bivariate normal distribution",

and we write $X \sim BVN(\mu, \Sigma)$.

2. We call
- $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ the "mean vector"; &
- Σ the "covariance matrix".

JOINT MGF

1. Note

$$\boxed{\begin{aligned} M(t_1,t_2) &= E(e^{t_1X_1 + t_2X_2}) \\ &= E(e^{t^TX}) \\ &= e^{t^T\mu + \frac{1}{2}t^T\Sigma t}, \end{aligned}}$$

where $t = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$.

MARGINAL DISTRIBUTIONS

1. Note

$$\boxed{M_{X_1}(t_1) = M(t_1, t_2=0) = \exp\left\{\mu_1 t_1 + \frac{\sigma_1^2}{2}t_1^2\right\},}$$

so $X_1 \sim N(\mu_1, \sigma_1^2)$ (by uniqueness of MGFs).

Similarly,

$$\boxed{M_{X_2}(t_2) = M(t_1=0, t_2) = \exp\left\{\mu_2 t_2 + \frac{\sigma_2^2}{2}t_2^2\right\}}$$

so $X_2 \sim N(\mu_2, \sigma_2^2)$.

CONDITIONAL DISTRIBUTION

1. Note

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

In particular, $(X_1 | X_2 = x_2)$ is normal.

Similarly,

$$\boxed{(X_1 | X_2 = x_2) = N\left(\mu_1 + \frac{\rho\sigma_1(x_2-\mu_2)}{\sigma_2}, \ \sigma_1^2(1-\rho^2)\right).}$$

$Cov(X_1, X_2)$

1. Note

$$\boxed{Cov(X_1, X_2) = \rho\sigma_1\sigma_2.}$$

Proof. Note

$E(X_1X_2) = E[E(X_1X_2|X_1)]$ (by double expectation theorem)

$\hookrightarrow$ rv of X_1.

To find $E(X_1X_2|X_1)$, we use the 2-step method.

① $E(X_1X_2 | X_1 = x_1) = E(x_1X_2 | X_1 = x_1)$ (by the substitution rule)

$$\begin{aligned} &= x_1 E(X_2 | X_1 = x_1) \\ &= x_1\left[\mu_2 + \frac{\rho\sigma_2(x_1-\mu_1)}{\sigma_1}\right] \\ &= x_1\mu_2 + \frac{\rho\sigma_2(x_1-\mu_1)x_1}{\sigma_1}. \end{aligned}$$

② Replace $x_1 \to X_1$ to get

$$E(X_1X_2|X_1) = \mu_2X_1 + \frac{\rho\sigma_2(X_1-\mu_1)X_1}{\sigma_1}.$$

Thus

$$\begin{aligned} E(X_1X_2) &= E\left(\mu_2X_1 + \frac{\rho\sigma_2(X_1-\mu_1)X_1}{\sigma_1}\right) \\ &= \mu_2E(X_1) + \frac{\rho\sigma_2}{\sigma_1}E(X_1^2 - \mu_1X_1) \\ &= \mu_2(\mu_1) + \frac{\rho\sigma_2}{\sigma_1}\underbrace{E(X_1^2 - \mu_1^2)}_{Var(X_1)} \\ &= \mu_1\mu_2 + \frac{\rho\sigma_2}{\sigma_1}\sigma_1^2 \\ &= \mu_1\mu_2 + \rho\sigma_1\sigma_2. \end{aligned}$$

Hence

$$\begin{aligned} Cov(X_1, X_2) &= E(X_1X_2) - E(X_1)E(X_2) \\ &= \mu_1\mu_2 + \rho\sigma_1\sigma_2 - \mu_1\mu_2 \\ &= \rho\sigma_1\sigma_2. \end{aligned}$$

2. So

$$\boxed{\rho(X_1,X_2) = \frac{Cov(X_1,X_2)}{\sqrt{Var(X_1)Var(X_2)}} = \frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2} = \rho.}$$

3. So
- the diagonal elements of Σ are variances of X_1 & X_2;
- the non-diagonal elements of Σ is $Cov(X_1, X_2)$; &
- ρ is $\rho(X_1, X_2)$.

INDEPENDENCE OF X_1 & X_2

1. Note if $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim BVN(\mu, \Sigma)$, then

$$\boxed{\rho = 0 \iff X_1 \text{ \& } X_2 \text{ are independent.}}$$

* It is not true that if X_1 & X_2 are normally distributed, then X_1 & X_2 are independent iff $\rho(X_1,X_2) = 0$!
We need the stronger condition that $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim BVN(\mu, \Sigma)$; ie the joint pdf of X_1 & X_2 is normal.
We cannot make this conclusion if we only know the marginal pdfs are normal.

c^TX

1. Let $c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$, where c_1 & c_2 are constants.

Then if $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim BVN(\mu, \Sigma)$, then

$$\boxed{c^TX = c_1X_1 + c_2X_2 \sim N(c^T\mu, c^T\Sigma c).}$$

* Again, this does not hold if we only know $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, but we don't know if $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim BVN(\mu, \Sigma)$.

$AX+b \sim BVN(A\mu+b, A\Sigma A^T)$

If $A \in M_{2\times 2}(\mathbb{R})$ & $b \in \mathbb{R}^2$, then

$$\boxed{Y = AX+b \sim BVN(A\mu+b, \ A\Sigma A^T)}$$

if $X \sim BVN(\mu, \Sigma)$.

χ^2- DISTRIBUTION

1. χ^2 is defined by

$$\chi_1^2 = Z^2, \quad \text{where } Z \sim N(0,1).$$

$$\chi_n^2 = \sum_{i=1}^{n} Z_i^2, \quad \text{where } Z_i \sim N(0,1) \text{ and are iid.}$$

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

$$\boxed{\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi_1^2.}$$

$(X-\mu)^T \Sigma^{-1} (X-\mu) \sim \chi_2^2$

Let $X \sim BVN(\mu, \Sigma)$. Then

$$\boxed{(X-\mu)^T \Sigma^{-1} (X-\mu) \sim \chi_2^2.}$$

Proof. To show this, we need to show

$$(X-\mu)^T \Sigma^{-1} (X-\mu) = \sum_{i=1}^{2} Z_i^2,$$

where $Z_1, Z_2 \sim N(0,1)$ are iid.

Sketch: We can write

$$\Sigma^{-1} = \Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}},$$

where we define $\Sigma^{-\frac{1}{2}}$ as follows:

If A is positive definite, then define

$$A^{\frac{1}{2}} = \sum_{j=1}^{d} \lambda_j^{\frac{1}{2}} x_j x_j^T, \qquad \begin{aligned} \lambda_j &= \text{eigenvalues of } A, \\ x_j &= \text{corresponding eigenvectors.} \end{aligned}$$

Then we show $\Sigma^{-\frac{1}{2}}(X-\mu) \sim BVN(0, I_{2\times 2})$.

Chapter 4: Functions of Random Variables

💡₁ Given the crvs $X_1, \dots, X_n$ with known joint distribution, we are interested to find the distribution of

$Y = h(X_1, \dots, X_n)$.

💡₂ Three methods:

① cdf technique
② 1-1 bivariate transformation
③ mgf technique.

CDF TECHNIQUE (4.1)

💡 Method:

① Find the cdf of $Y = h(X_1, \dots, X_n)$,

$$\boxed{P(Y \le y) = P(h(X_1, \dots, X_n) \le y)}$$

② Get the pdf of Y:

$$\boxed{f_Y(y) = F_Y'(y).}$$

EXAMPLE 1 (Y IS UNIVARIATE)

💡 If $X \sim N(0,1)$, find the pdf of $Y = X^2$.

Sol$^{\underline{n}}$. If $y \le 0$, $P(Y \le y) = 0$.

If $y > 0$, $P(Y \le y) = P(X^2 \le y)$

$$= P(-\sqrt{y} \le X \le \sqrt{y})$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$
$$= \int_{-\infty}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx - \int_{-\infty}^{-\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx.$$

Therefore, the pdf of Y is 0 if $y \le 0$, & for $y > 0$, it is

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{\sqrt{2\pi}} e^{\frac{-y}{2}} \cdot \frac{1}{2} y^{-\frac{1}{2}} + \frac{1}{\sqrt{2\pi}} e^{\frac{-y}{2}} \cdot \frac{1}{2} y^{-\frac{1}{2}}$$
$$= \frac{1}{\sqrt{2\pi}} e^{\frac{-y}{2}} y^{-\frac{1}{2}}.$$

In particular, also notice that $\chi_1^2 \sim \text{Gamma}(\frac{1}{2}, 2)$

EXAMPLE 2 (Y IS UNIVARIATE)

💡 The pdf of X is

$$f(x) = \frac{\theta}{x^{\theta+1}}, \quad x \ge 1, \ \theta > 0.$$

Find the pdf of $Y = \log X$.

Sol$^{\underline{n}}$. Support of Y is $[0, \infty)$.

If $y \le 0$, $F_Y(y) = P(Y \le y) = 0$.

If $y > 0$,

$$F_Y(y) = P(Y \le y)$$
$$= P(\log X \le y)$$
$$= P(X \le e^y)$$
$$= \int_1^{e^y} \frac{\theta}{x^{\theta+1}} dx \quad \text{(since support of } X \text{ is } 1 \le x < \infty)$$
$$= \left[-x^{-\theta}\right]_1^{e^y}$$
$$= 1 - e^{-y\theta}.$$

Thus, the pdf of Y is

$$f_Y(y) = \theta e^{-y\theta}, \quad \text{for } y > 0.$$

EXAMPLE 3 (Y IS A FUNC OF ≥1 RV)

💡 The joint pdf of X, Y is

$$f(x,y) = 3y, \quad 0 \le x \le y \le 1.$$

Find the marginal pdf of $T = XY$ & $S = \frac{Y}{X}$.

Sol$^{\underline{n}}$. The support of T is $[0,1]$.

If $t < 0$, $P(T \le t) = 0$, & if $t > 1$, $P(T \le t) = 1$.

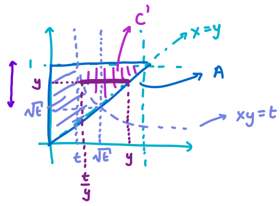
If $0 \le t \le 1$, then

$$P(T \le t) = P(XY \le t).$$

Consider $P((X,Y) \in C')$.

Given a particular y, x can range from $[\frac{t}{y}, y]$.

y can range from $\sqrt{t}$ to 1.



Then

$$P((X,Y) \in C') = \int_{\sqrt{t}}^{1} dy \int_{\frac{t}{y}}^{y} 3y \, dx$$
$$= \int_{\sqrt{t}}^{1} \left[3yx\right]_{x=\frac{t}{y}}^{x=y} dy$$
$$= \int_{\sqrt{t}}^{1} 3y^2 - 3t \, dy$$
$$= \left[y^3 - 3ty\right]_{y=\sqrt{t}}^{y=1}$$
$$= (1 - 3t) - (t^{3/2} - 3t^{3/2})$$
$$= 1 + 2t^{3/2} - 3t.$$

$$\therefore P((X,Y) \in C) = 1 - P((X,Y) \in C')$$

$$P(XY \le t). \quad = 1 - (1 + 2t^{3/2} - 3t)$$
$$= 3t - 2t^{3/2}.$$

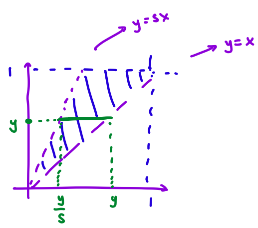
So the pdf of T is

$$f_T(t) = \frac{d}{dt}(3t - 2t^{3/2}) = 3 - 3t^{\frac{1}{2}}, \quad 0 < t < 1.$$

The support of S is $[1, \infty)$.

When $s < 1$, $P(S \le s) = 0$.

When $s \ge 1$,



$$P(S \le s) = P(\frac{Y}{X} \le s)$$
$$= P(Y \le sX).$$
$$= \int_0^1 dy \int_{y/s}^{y} 3y \, dx$$
$$= \int_0^1 3y(y - \frac{y}{s}) \, dy$$
$$= \int_0^1 3y^2 - \frac{3y^2}{s} \, dy$$
$$= \left[y^3 - \frac{y^3}{s}\right]_0^1 = 1 - \frac{1}{s}.$$

Thus

pdf :: $f(s) = \frac{1}{s^2}$, $s \ge 1$, & 0 otherwise.

EXAMPLE 4 (FIND DISTN OF MAX/MIN)

Let $X_1, \dots, X_n$ are iid Unif $[0, \theta]$. Find the pdf of

① $X_{(n)} = \max_{1 \le i \le n} X_i$

② $X_{(1)} = \min_{1 \le i \le n} X_i$.

① $X_{(n)}$: support of $X_{(n)}$ is $[0, \theta]$.

i) when $x \le 0$, $P(X_{(n)} \le x) = 0$.

ii) when $x \ge \theta$, $P(X_{(n)} \le x) = 1$.

iii) when $0 < x < \theta$,

$$
\begin{aligned}
P(X_{(n)} \le x) &= P\left(\bigcap_{i=1}^{n} (X_i \le x)\right) \\
&= \prod_{i=1}^{n} P(X_i \le x) \quad \text{(since } X_i \text{ are iid)} \\
&= \prod_{i=1}^{n} \frac{x}{\theta} = \frac{x^n}{\theta^n}.
\end{aligned}
$$

$\therefore$ pdf of $X_{(n)}$ is $f_{X_{(n)}} = \dfrac{n x^{n-1}}{\theta^n}$, $0 \le x \le \theta$

② Support of $X_{(1)}$ is $[0, \theta]$.

i) If $x \le 0$, $P(X_{(1)} \le x) = 0$

ii) If $x \ge \theta$, $P(X_{(1)} \le x) = 1$.

iii) If $0 < x < \theta$,

$$
\begin{aligned}
P(X_{(1)} \le x) &= 1 - P(X_{(1)} > x) \\
&= 1 - P\left(\bigcap_{i=1}^{n} (X_i > x)\right) \\
&= 1 - \prod_{i=1}^{n} P(X_i > x) \quad \text{(since } X_i \text{ are iid)} \\
&= 1 - \prod_{i=1}^{n} \left(1 - \frac{x}{\theta}\right) \\
&= 1 - \left(1 - \frac{x}{\theta}\right)^n \\
&= 1 - \left(\frac{\theta - x}{\theta}\right)^n.
\end{aligned}
$$

$\therefore$ pdf of $X_{(1)}$ is $f_{X_{(1)}} = \dfrac{n(\theta - x)^{n-1}}{\theta^n}$, $0 \le x \le \theta$.

1-1 BIVARIATE TRANSFORMATION (4.2)

1. Given the joint pdf $f(x,y)$ of X,Y, we want to find the joint pdf of

$$\boxed{U = h_1(X,Y), \quad V = h_2(X,Y).}$$

2. Then, a "1-1 bivariate transformation" is

$$\boxed{u = h_1(x,y), \quad v = h_2(x,y).}$$

These are 1-1 if there exist another 2 functions such that

$$\boxed{x = w_1(u,v), \quad y = w_2(u,v).}$$

JACOBIAN

1. The "Jacobian" of (u,v), where $u = h_1(x,y)$ & $v = h_2(x,y)$, is

$$\boxed{\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}}$$

2. If we can write $x = w_1(u,v)$ & $y = w_2(u,v)$, then we define

$$\boxed{\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}}$$

3. Then, the pdf of (U,V) is

$$\boxed{\begin{aligned} g(u,v) &= f(x,y) \cdot \left| \frac{\partial(x,y)}{\partial(u,v)} \right| \\ &= f(w_1(u,v), w_2(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right|. \end{aligned}}$$

EXAMPLE 1

Let $X, Y \sim N(0,1)$ be independent.
Let $U = X+Y$, $V = X-Y$. Find the joint pdf of (U,V).

Sol^{n}. $\begin{cases} U = X+Y \\ V = X-Y \end{cases} \Rightarrow \begin{cases} X = \frac{U+V}{2} \\ Y = \frac{U-V}{2} \end{cases}.$

Jacobian is

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = -\frac{1}{2}.$$

$\therefore$ The joint pdf of U & V is

$$\begin{aligned} g(u,v) &= f(x,y)\,|J| \\ &= f\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \cdot \left|-\frac{1}{2}\right| \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{u+v}{2}\right)^2\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{u-v}{2}\right)^2\right\} \cdot \left|-\frac{1}{2}\right| \\ &= \frac{1}{4\pi} \exp\left\{-\frac{1}{8}\left[(u+v)^2 + (u-v)^2\right]\right\} \\ &= \frac{1}{4\pi} \exp\left\{-\frac{1}{4}(u^2+v^2)\right\}, \quad -\infty < u,v < \infty. \end{aligned}$$

EXAMPLE 2

The joint pdf of X & Y is
$f(x,y) = e^{-x-y}$, $0 < x, y < \infty$.
Find the pdf of $U = X+Y$.

Sol^{n}. Let $V = X$. Then

$\begin{cases} u = x+y \\ v = x \end{cases} \Rightarrow \begin{cases} x = v \\ y = u-v. \end{cases}$

Since $x,y>0$, the joint support of (U,V) is $0<v<u<\infty$.

$$\therefore J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1.$$

Thus the joint pdf of U & V is

$$\begin{aligned} g(u,v) &= |J|\, f(x,y) = |-1|\, e^{-x-y} \\ &= e^{-(v)-(u-v)} = e^{-u}, \quad 0<v<u<\infty. \end{aligned}$$

Hence the marginal pdf of U is

$$\begin{aligned} g_1(u) &= \int_0^u g(u,v)\,dv, \quad 0<u<\infty \\ &= \int_0^u e^{-u}\,dv \end{aligned}$$

$\therefore g_1(u) = ue^{-u}$, $0<u<\infty$.

EXAMPLE 3 (SUPPORT)

Let the support of X & Y is $0<x<y<1$.
Find the support of (U,V), where
$\begin{cases} U = X \\ V = XY. \end{cases}$

Sol^{n}. $\begin{cases} u = x \\ v = xy \end{cases} \Rightarrow \begin{cases} x = u \\ y = \frac{v}{u}. \end{cases}$

Since $0<x<y<1$,
$\therefore 0<u<\frac{v}{u}<1$.
$\therefore 0<u<\frac{v}{u}$ & $\frac{v}{u}<1$
$\Rightarrow 0<u^2<v<u<1$.

EXAMPLE 4 (SUPPORT)

Suppose the joint support of X & Y is $0<x,y<1$.
Find the joint support of (U,V), where
$\begin{cases} U = \frac{X}{Y} \\ V = XY. \end{cases}$

Sol^{n}. $\begin{cases} u = \frac{x}{y} \\ v = xy \end{cases} \Rightarrow \begin{cases} x = \sqrt{uv} \\ y = \sqrt{\frac{v}{u}}. \end{cases}$

Then $0<x,y<1 \Rightarrow 0<\sqrt{uv}<1$, $0<\sqrt{\frac{v}{u}}<1$.
$\therefore u,v>0$, &
$uv<1$ & $\frac{v}{u}<1$.
$\therefore u<\frac{1}{v}$ & $v<u$.

$\therefore$ the joint support of (U,V) is $0<v<u<\frac{1}{v}$.

MGF TECHNIQUE (4.3)

Idea:

① Find the mgf of a rv.

② By the uniqueness property of mgfs, we can identify its distribution & the pdf of this rv.

2. Suppose $X_1, \dots, X_n$ are independent. Then the mgf of $T = \sum_{i=1}^n X_i$ is

$$\boxed{M_T(t) = \prod_{i=1}^n M_{X_i}(t).}$$

Proof.
$$\begin{aligned} M_T(t) &= E(e^{t\sum_{i=1}^n X_i}) \\ &= E(e^{\sum_{i=1}^n tX_i}) \\ &= E(\prod_{i=1}^n e^{tX_i}) \\ &= \prod_{i=1}^n E(e^{tX_i}) \quad \text{(since } X_i \text{ are independent)} \\ &= \prod_{i=1}^n M_{X_i}(t). \end{aligned}$$

3. If $X_1, \dots, X_n$ are iid, then

$$\boxed{M_T(t) = [M_{X_1}(t)]^n.}$$

NORMAL DISTRIBUTION

1. If $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

Proof. Let $Y = aX + b$. Then
$$\begin{aligned} M_Y(t) &= e^{bt} M_X(at) \\ &= \cdots \\ &= e^{(a\mu+b)t + \frac{a^2\sigma^2}{2}t^2} \end{aligned}$$
so $Y \sim N(a\mu + b, a^2\sigma^2)$.

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

Proof. This follows from the previous result.

3. If $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$ are independent, then

$$\boxed{\sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2\right).}$$

Proof. Let $T = \sum_{i=1}^n a_i X_i$. Then
$$\begin{aligned} M_T(t) &= \prod_{i=1}^n M_{a_iX_i}(t) \quad \text{(since } X_1, \dots, X_n \text{ are independent)} \\ &= \prod_{i=1}^n M_{X_i}(a_i t) \\ &= \prod_{i=1}^n e^{(a_i t)\mu_i + \frac{\sigma_i^2}{2}(a_i^2 t^2)} \\ &= \exp\left\{ t \cdot \sum_{i=1}^n a_i\mu_i + \frac{t^2}{2}\sum_{i=1}^n a_i^2\sigma_i^2 \right\}, \end{aligned}$$
so that $T \sim N\left(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2\right)$. #

4. If $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then

$$\boxed{\begin{aligned} &\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2) \\ &\frac{1}{n}\sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right). \end{aligned}}$$

We call $\frac{1}{n}\sum_{i=1}^n X_i$ the "sample mean".

χ^2-DISTRIBUTION

1. Recall $\chi_1^2 = [N(0,1)]^2$.
Then if $X \sim N(\mu, \sigma^2)$, thus $\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi_1^2$.

2. Additivity of χ^2:
If $Y_i \sim \chi_{k_i}^2$, $1 \le i \le n$, & $Y_1, \dots, Y_n$ are independent, then

$$\boxed{\sum_{i=1}^n Y_i \sim \chi_d^2, \quad d = \sum_{i=1}^n k_i.}$$

Proof. We know
$\chi_1^2 = \text{Gamma}(\frac{1}{2}, 2)$.
We can then show
$\chi_\nu^2 = \text{Gamma}(\frac{\nu}{2}, 2)$.
Then it follows $Y_i \sim \text{Gamma}(\frac{k_i}{2}, 2)$, and so
$M_{Y_i}(t) = (1-2t)^{-\frac{k_i}{2}}$.
Let $T = \sum_{i=1}^n Y_i$. Then
$$\begin{aligned} M_T(t) &= \prod_{i=1}^n M_{Y_i}(t) \quad \text{(as } Y_i \text{ are independent)} \\ &= \prod_{i=1}^n (1-2t)^{-\frac{k_i}{2}} \\ &= (1-2t)^{-\frac{d}{2}}, \quad d = \sum_{i=1}^n k_i. \end{aligned}$$
This is exactly the mgf of $\text{Gamma}(\frac{d}{2}, 2) = \chi_d^2$, so it follows $T \sim \chi_d^2$ by uniqueness.

3. We can similarly show

$$\boxed{\chi_\nu^2 = \sum_{i=1}^\nu Z_i^2, \quad Z_i \overset{iid}{\sim} N(0,1).}$$

4. In particular, if $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then

$$\boxed{\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_n^2.}$$

t DISTRIBUTION

If $X \sim N(0,1)$, $Y \sim \chi_n^2$ are independent, then we say

$$\boxed{\frac{X}{\sqrt{Y/n}} \sim t_n.}$$

The support of t_n is $(-\infty, \infty)$.

F DISTRIBUTION

If $X \sim \chi_n^2$, $Y \sim \chi_m^2$ are independent, then

$$\boxed{\frac{(X/n)}{(Y/m)} \sim F_{n,m}.}$$

n, m are the two degrees of freedom parameters.

The support of $F_{n,m}$ is $(0, \infty)$.

EXAMPLE 1

Let $X \sim \chi_n^2$, $Y \sim \chi_m^2$ are independent.
Then we know $X + Y \sim \chi_{n+m}^2$.
Does $\frac{X/n}{(X+Y)/n} \sim F_{n,n+m}$?

Sol^{n}. No, because X & $X+Y$ are not independent.
To see this, see that
$$\begin{aligned} \text{Cov}(X, X+Y) &= E(X(X+Y)) - E(X)E(X+Y) \\ &> 0. \end{aligned}$$

EXAMPLE 2 (χ^2)

Let $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.
Then we know

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_n^2.$$

When μ is unknown, we replace μ by

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

Then we can show

$$\boxed{\frac{\sum_{i=1}^{n}(X_i - \bar{X})}{\sigma^2} \sim \chi_{n-1}^2.}$$

Proof. First, we show $\bar{X}$ is independent of

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

To do this, we observe that if X is independent of (Y, Z), then X is independent of $g(Y, Z)$.

Then, consider

$$(\bar{X}, \underbrace{X_1 - \bar{X}, \dots, X_n - \bar{X}}_{Y}).$$

We claim $\text{Cov}(\bar{X}, X_i - \bar{X}) = 0$. To do this, see that

$$\begin{aligned}
\text{Cov}\left(\frac{1}{n}\sum_{j=1}^{n} X_j, X_i\right) &= E\left(\frac{1}{n}\sum_{j=1}^{n} X_j X_i\right) - E\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right)E(X_i) \\
&= \frac{1}{n}\Big[\sum_{j\neq i} E(X_j)E(X_i) + E(X_i^2) \\
&\quad - \sum_{j \neq i} E(X_j)E(X_i) - E(X_i)^2\Big] \\
&= \frac{1}{n}\text{Var}(X_i).
\end{aligned}$$

Thus $\bar{X}$ is independent of S^2.

We then show $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$. See that

$$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2.$$

Then

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \\
&\quad + 2\underbrace{\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \mu)}_{0}.
\end{aligned}$$

So

$$\underbrace{\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2}}_{\text{mgf: } (1-2t)^{-\frac{n}{2}}} = \underbrace{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}}_{M(t)} + \underbrace{\frac{n(\bar{X} - \mu)^2}{\sigma^2}}.$$

Since $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$,

$\therefore \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1)$.

$\therefore \left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}\right)^2 \sim \chi_1^2 = \text{Gamma}(\frac{1}{2}, 2)$.

$\therefore$ MGF $= (1-2t)^{-\frac{1}{2}}$.

Taking MGFs of both sides, we see that

$$(1-2t)^{-\frac{n}{2}} = M(t)(1-2t)^{-\frac{1}{2}}$$

and so

$$M(t) = (1-2t)^{\frac{-(n-1)}{2}},$$

which is the mgf of χ_{n-1}^2. Thus

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 = \text{Gamma}\left(\frac{n-1}{2}, 2\right).$$

Let

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.}$$

Then

$$\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.}$$

EXAMPLE 3 (t)

Let $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Then

$$\boxed{\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.}$$

Proof. $\frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \sim N(0,1)$ & $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ are independent.

Therefore

$$\frac{\left[\frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}}\right]}{\sqrt{\frac{(n-1)S^2}{\sigma^2} / n-1}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{S} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

by definition of t-distribution.

EXAMPLE 4 (F)

If $X_1, \dots, X_n \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$, & $Y_1, \dots, Y_m \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$, & these 2 samples are independent, define

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$
$$S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2.$$

Then

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1, m-1}.$$

Proof. We know $\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi_{n-1}^2$ & $\frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi_{m-1}^2$, and these are independent.

Thus, by def$^{\underline{n}}$ of F dist$^{\underline{n}}$,

$$\frac{\frac{(n-1)S_1^2}{\sigma_1^2} / (n-1)}{\frac{(m-1)S_2^2}{\sigma_2^2} / (m-1)} \sim F_{n-1, m-1}.$$

$$\Rightarrow \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1, m-1},$$ as needed.

Chapter 5: Limiting Distributions

1. The main problem to be solved:
   we are interested in the distribution of $\sqrt{n}(\bar{X}-\mu)$,
   where $X_1, \dots, X_n$ are iid, & $E(X_i)=\mu$, $Var(X_i)=\sigma^2$,
   & $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
2. We don't know the distribution of X_i;
   thus, it is impossible to know the exact distribution of $\sqrt{n}(\bar{X}-\mu)$.
3. Instead, we find an approximate distribution for $\sqrt{n}(\bar{X}-\mu)$.

LIMITING / ASYMPTOTIC DISTRIBUTION

1. Let $F_n(x)$ be the true cdf of $\sqrt{n}(\bar{X}-\mu)$.
   By definition,

$$\boxed{F_n(x) = P(\sqrt{n}(\bar{X}-\mu) \le x).}$$

2. Consider $\lim_{n\to\infty} F_n(x) = F(x)$, where F is a known cdf.
   Then we can use $F(x)$ to approximate $F_n(x)$.

CONVERGENCE IN DISTRIBUTION: $X_n \xrightarrow{d} X$ (5.1)

1. Let $X_1, \dots, X_n$ be a sequence of rvs such that X_n has cdf $F_n(x)$.

   Then, let X be another rv with cdf $F(x)$.

   If

$$\boxed{\lim_{n\to\infty} F_n(x) = F(x) \text{ for all } x \text{ such that which } F(x) \text{ is continuous}}$$

   then we say X_n "converges in distribution" to X,
   and write $X_n \xrightarrow{d} X$.
2. Note:
   1. $F(x)$ is called the "limiting distribution" or "asymptotic distribution" of X_n.
   2. The cdf of X_n converges, rather than just the rv X_n.
      ↳ we don't actually know that the actual rv X_n converges to X! We only know the cdf converges.
   3. $\lim_{n\to\infty} F_n(x) = F(x)$ only holds for the continuous points of $F(x)$.
   4. This definition applies to both discrete & cts rvs.

EXAMPLE 1

1. If

$$\boxed{F(x) = \begin{cases} 1, & x \ge a \\ 0, & x < a \end{cases}}$$

   We note F is not cts at $x=a$.
   If you want to show $X_n \xrightarrow{d} X$, we are not interested of $\lim_{n\to\infty} F_n(a)$.
2. We only need to show

$$\boxed{\lim_{n\to\infty} F_n(a) = \begin{cases} 0, & x < a \\ 1, & x > a. \end{cases}}$$

EXAMPLE 2

Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Unif}[0,1]$.
Let

$$X_{(1)} = \min_{1\le i\le n} X_i \quad , \quad X_{(n)} = \max_{1\le i\le n} X_i .$$

Find the limiting distribution of

a) $nX_{(1)}$ & $n(1-X_{(n)})$; &

b) $X_{(1)}$ & $X_{(n)}$.

Soln. a) $nX_{(1)}$: support is $[0,n]$.

Then the cdf of $nX_{(1)}$ is

$$P(nX_{(1)} \le x) = \begin{cases} 0 & , \quad x \le 0 \\ P(X_{(1)} \le \frac{x}{n}) & , \quad 0<x<n . \\ 1 & , \quad x \ge n \end{cases}$$

For $0<x<n$,

$$\begin{aligned} P(X_{(1)} \le \tfrac{x}{n}) &= 1 - P(X_{(1)} > \tfrac{x}{n}) \\ &= 1 - P(X_1 > \tfrac{x}{n}, \dots, X_n > \tfrac{x}{n}) \\ &= 1 - \prod_{i=1}^{n} P(X_1 > \tfrac{x}{n}) \qquad \text{(by independence of } X_i) \\ &= 1 - (1-\tfrac{x}{n})^n . \end{aligned}$$

Therefore,

$$P(nX_{(1)} \le x) = \begin{cases} 0 & , \quad x \le 0 \\ 1-(1-\frac{x}{n})^n & , \quad 0<x<n \\ 1 & , \quad x \ge n . \end{cases}$$

Hence

$$\lim_{n\to\infty} P(nX_{(1)} \le x) = \begin{cases} 0 & , \quad x \le 0 \\ 1-e^{-x} & , \quad 0<x<n \\ 1 & , \quad x \ge n . \end{cases}$$

This is the cdf of Exp(1).

$n(1-X_{(n)})$: support is $[0,n]$.

when $x \le 0$, $P(n(1-X_{(n)}) \le x) = 0$, & when $x \ge n$, $P(n(1-X_{(n)}) \le x) = 1$.

Then, when $0<x<n$,

$$\begin{aligned} P(n(1-X_{(n)}) \le x) &= P(1-X_{(n)} \le \tfrac{x}{n}) \\ &= P(X_{(n)} \ge 1-\tfrac{x}{n}) \\ &= 1 - P(X_{(n)} \le 1-\tfrac{x}{n}) \\ &= 1 - P(X_1 \le 1-\tfrac{x}{n}, \dots, X_n \le 1-\tfrac{x}{n}) \\ &= 1 - \prod_{i=1}^{n} P(X_1 \le 1-\tfrac{x}{n}) \qquad \text{(since } X_i \text{ are iid)} \\ &= 1 - (1-\tfrac{x}{n})^n . \end{aligned}$$

$$\therefore \; F_n(x) = \begin{cases} 0 & , \quad x \le 0 \\ 1-(1-\frac{x}{n})^n & , \quad 0<x<n \\ 1 & , \quad x \ge n . \end{cases}$$

Hence

$$F(x) = \lim_{n\to\infty} F_n(x) = \begin{cases} 0 & , \quad x<0 \\ 1-e^{-x} & , \quad x \ge 0 . \end{cases}$$

In particular, $F(x)$ is the cdf of Exp(1).

b) $X_{(1)}$: support is $[0,1]$.

when $x \le 0$, $P(X_{(1)} \le x) = 0$, & when $x \ge 1$, $P(X_{(1)} \le x) = 1$.

when $0<x<1$,

$$\begin{aligned} P(X_{(1)} \le x) &= 1 - P(X_{(1)} \ge x) \\ &= 1 - P(X_1 \ge x, \dots, X_n \ge x) \\ &= 1 - \prod_{i=1}^{n} P(X_1 \ge x) \qquad \text{(by iid of } X_i\text{'s)} \\ &= 1 - \prod_{i=1}^{n} (1 - P(X_1 \le x)) \\ &= 1 - (1-x)^n . \end{aligned}$$

Thus

$$F_n(x) = \begin{cases} 0 & , \quad x \le 0 \\ 1-(1-x)^n & , \quad 0<x<1 \\ 1 & , \quad x \ge 1 . \end{cases}$$

Then

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & , \quad x \le 0 \\ 1 & , \quad 0<x<1 \\ 1 & , \quad x \ge 1 . \end{cases} \Big\} \text{ ie } x>0 .$$

This is not a cdf since this is not right continuous at $x=0$!

Instead, the limiting cdf is

$$F(x) = \begin{cases} 0 , & x<0 \\ 1 , & x \ge 0 . \end{cases}$$

This is right continuous, but is not continuous at $x=0$.

$X_{(n)}$: support is $[0,1]$. Then

$P(X_{(n)} \le x) = 0$ if $x \le 0$, $P(X_{(n)} \le x) = 1$ if $x \ge 1$.

For $0<x<1$,

$$\begin{aligned} P(X_{(n)} \le x) &= P(X_1 \le x, \dots, X_n \le x) \\ &= \prod_{i=1}^{n} P(X_1 \le x) \qquad \text{(by iid of } X_i) \\ &= x^n . \end{aligned}$$

Thus

$$F_n(x) = \begin{cases} 0 , & x \le 0 \\ x^n , & 0<x<1 \\ 1 , & x \ge 1 . \end{cases}$$

Hence

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 , & x \le 0 \\ 0 , & 0<x<1 \\ 1 , & x \ge 1 . \end{cases}$$

Thus, the limiting cdf is

$$F(x) = \begin{cases} 0 & , \quad x<1 \\ 1 & , \quad x \ge 1 \end{cases}$$

CONVERGENCE IN PROBABILITY: $X_n \xrightarrow{P} X$ (5.2)

1. Let $X_1, \dots, X_n$ be a sequence of rvs such that X_n has cdf $F_n(x)$. Let X be another rv with cdf $F(x)$.

   If
   $$\boxed{\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0}$$
   or equivalently if
   $$\boxed{\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1}$$
   for any given $\varepsilon > 0$,

   then we say X_n "converges in probability" to X and write $X_n \xrightarrow{P} X$.

2. Note that here, it is the convergence / limit for a probability rather than a cdf (like in convergence in distribution).

3. In particular, as $n \to \infty$, X_n cannot be "ε" away from X. That is to say, X_n gets pretty "close" to X as $n \to \infty$.

4. So, we expect that $F_n(x)$ becomes very close to $F(x)$.

5. If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$;
   ie convergence in probability implies convergence in distribution.

CONVERGENCE IN PROBABILITY TO A CONSTANT

1. Let $X_1, \dots, X_n$ be a sequence of rvs, & let c be a constant.

   If
   $$\boxed{\lim_{n\to\infty} P(|X_n - c| > \varepsilon) = 0 \quad \forall \varepsilon > 0,}$$
   then we say X_n converges in probability to c, and write $X_n \xrightarrow{P} c$.

2. Let $X_1, \dots, X_n$ be a sequence of rvs with cdf $F_n(x)$.

   Then, if
   $$\boxed{\lim_{n\to\infty} F_n(x) = \begin{cases} 0, & x < c \\ 1, & x > c \end{cases}}$$
   *note: we don't need to consider when $x = c$.

   or in other words if the limiting distribution of X_n is
   $$\boxed{F(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}}$$
   ie if $X_n \xrightarrow{d} c$,

   then necessarily $X_n \xrightarrow{P} c$.

3. In other words, $X_n \xrightarrow{P} c \iff X_n \xrightarrow{d} c$, where c is a constant.

Proof. We need to show for any $\varepsilon > 0$,
$$\lim_{n\to\infty} P(|X_n - c| > \varepsilon) = 0.$$
Then, note $P(|X_n - c| > \varepsilon) \geq 0$.

Next, for any $\varepsilon > 0$, note
$$\begin{aligned} P(|X_n - c| > \varepsilon) &= P(\{X_n - c > \varepsilon\} \cup \{X_n - c < -\varepsilon\}) \\ &= P(X_n - c > \varepsilon) + P(X_n - c < -\varepsilon) \quad \text{(since events are mutually exclusive)} \\ &= P(X_n > c + \varepsilon) + P(X_n < c - \varepsilon) \\ &= 1 - P(X_n < c + \varepsilon) + P(X_n < c - \varepsilon) \\ &= 1 - F_n(c + \varepsilon) + F_n(c - \varepsilon). \end{aligned}$$
Intuitively, see that
$$\lim_{n\to\infty} F_n(c + \varepsilon) = F(c + \varepsilon) = 1$$
&
$$\lim_{n\to\infty} F_n(c - \varepsilon) = F(c - \varepsilon) = 0.$$
So
$$\lim_{n\to\infty} P(|X_n - c| > \varepsilon) = 1 - 1 + 0 = 0,$$
which suffices to prove the claim. ∎

EXAMPLE 1

Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Unif}[0,1]$.

We showed
$$\lim_{n\to\infty} P(X_{(1)} \leq x) = \begin{cases} 0, & x < 0 \\ 1, & x > 0 \end{cases}$$
and
$$\lim_{n\to\infty} P(X_{(n)} \leq x) = \begin{cases} 0, & x < 1 \\ 1, & x > 1; \end{cases}$$
ie $X_{(1)} \xrightarrow{d} 0$ & $X_{(n)} \xrightarrow{d} 1$.

Thus by the theorem, $X_{(1)} \xrightarrow{P} 0$ & $X_{(n)} \xrightarrow{P} 1$.

EXAMPLE 2

Let $X_1, \dots, X_n$ be iid with pdf
$$f(x) = e^{-(x-\theta)}, \quad x \geq \theta.$$
Let $X_{(1)} = \min_{1 \leq i \leq n} X_i$. Show $X_{(1)} \xrightarrow{P} \theta$.

Proof. We could show this by def^n of convergence in probability.

Alternatively, we show $X_{(1)} \xrightarrow{d} \theta$, as that implies $X_{(1)} \xrightarrow{P} \theta$.

In other words, we want to show
$$\lim_{n\to\infty} P(X_{(1)} \leq x) = \begin{cases} 0, & x < \theta \\ 1, & x > \theta. \end{cases}$$
The support of $X_{(1)}$ is $[\theta, \infty)$.

So, if $x \leq \theta$, $P(X_{(1)} \leq x) = 0$.

If $x > \theta$,
$$\begin{aligned} P(X_{(1)} \leq x) &= 1 - P(X_{(1)} > x) \\ &= 1 - P(X_1 > x, \dots, X_n > x) \\ &= 1 - \prod_{i=1}^{n} P(X_i > x) \qquad \text{(by iid of } X_i\text{'s)} \\ &= 1 - \left[\int_x^\infty e^{-(t-\theta)}\, dt \right]^n \\ &= 1 - \left[e^{-(x-\theta)} \right]^n \\ &= 1 - e^{-n(x-\theta)} \end{aligned}$$
when $x > \theta$, $x - \theta > 0$, so $\lim_{n\to\infty} e^{-n(x-\theta)} = 0$.

Thus $P(X_{(1)} \leq x) \to 1$ as $n \to \infty$.

Hence
$$\lim_{n\to\infty} P(X_{(1)} \leq x) = \begin{cases} 0, & x < \theta \\ 1, & x > \theta, \end{cases}$$
as required. (So $X_{(1)} \xrightarrow{d} \theta$, $\Rightarrow X_{(1)} \xrightarrow{P} \theta$).

MARKOV'S INEQUALITY

Let X be a rv. Then, for $k, c > 0$, we have

$$\boxed{P(|X| \geq c) \leq \frac{E(|X|^k)}{c^k}.}$$

* this bounds probability by a moment of X.

Proof. Idea: Use the fact that

$$P(X>c) = \int_c^\infty f(x)\,dx \leq \int_c^\infty \frac{|x|^k}{c^k} f(x)\,dx$$
$$\leq \int_{-\infty}^\infty \frac{|x|^k}{c^k} f(x)\,dx$$
$$= \frac{E(|X|^k)}{c^k}.$$

Often, we will take $k=2$, and

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

where $\mu = E(X)$.

* This is known as Chebyshev's inequality.

WEAK LAW OF LARGE NUMBERS (WLLN)

Let $X_1, \dots, X_n$ be independent with a common mean μ & common variance σ^2.

Then

$$\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} \mu.}$$

Proof. By definition, for any $\varepsilon > 0$, we want to show

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| > \varepsilon\right) \to 0$$

as $n \to \infty$.

i) We know $P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| > \varepsilon\right) \geq 0$.

ii) So, if we can show

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| > \varepsilon\right) \leq a_n$$

& $a_n \to 0$ as $n \to \infty$, by squeeze theorem we would be done.

So, by Chebyshev's inequality,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| > \varepsilon\right) \leq \frac{E\left[\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right|^2\right]}{\varepsilon^2} = \frac{Var(\bar{X})}{\varepsilon^2}$$

We know $E(\bar{X}) = \mu$ & $Var(\bar{X}) = \frac{\sigma^2}{n}$. Thus

$$P(|\bar{X} - \mu| > \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2} = \frac{1}{n}\left(\frac{\sigma^2}{\varepsilon^2}\right).$$

This converges to 0 as $n \to \infty$ since $\frac{\sigma^2}{\varepsilon^2}$ is fixed.

So we are done. ∎

EXAMPLE 1

Let $X_1, \dots, X_n \overset{iid}{\sim} \chi^2_1$.

Then we can show

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

Sol^{n}. Note $E(X_i) = E(Z^2) = Var(Z) + E(Z)^2 = 1$,

where $Z \sim N(0,1)$.

We can also show

$$Var(X_i) = Var(\chi_1^2)$$
$$= Var(Gamma(\tfrac{1}{2}, 2))$$
$$= \tfrac{1}{2}(2)^2 = 2.$$

Since $E(X_i), Var(X_i) < \infty$, we can apply WLLN.

Thus $\bar{X}_n \xrightarrow{P} E(X_i) = 1$.

EXAMPLE 2

Suppose $Y_n \sim \chi^2_n$. Then

$$\frac{Y_n}{n} \xrightarrow{P} 1.$$

Sol^{n}. Note $Y_n = \sum_{i=1}^n X_i$, $X_1, \dots, X_n \sim \chi^2_1$.

Thus $\frac{Y_n}{n} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$; the result follows from the previous example.

EXAMPLE 3

Let $X_1, \dots, X_n \overset{iid}{\sim} Poi(\mu)$. Then

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} E(X_i) = \mu.$$

Sol^{n}. We need to show the conditions for WLLN is satisfied; after that the conclusion follows.

We note $E(\bar{X}_n) = \mu$ & $Var(\bar{X}_n) = \frac{\mu}{n}$, which are both finite. The conclusion thus follows.

SOME USEFUL LIMIT THEOREMS (5.3)

CONVERGENCE OF MGFs

Let $X_1, \ldots, X_n$ be a sequence of rvs such that X_n has mgf $M_n(t)$.

Let X be a rv with mgf $M(t)$.

If there exists an $h>0$ such that

$$\boxed{\lim_{n\to\infty} M_n(t) = M(t) \quad \forall t \in (-h,h),}$$

then $X_n \xrightarrow{d} X$.

CENTRAL LIMIT THEOREM (CLT)

Let $X_1, \ldots, X_n$ be iid with common mean μ & common variance $\sigma^2 < \infty$.

* note they need to be iid !!

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

Then the limiting distribution of $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ is $N(0,1)$, ie

$$\boxed{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1).}$$

Proof. We will use the above theorem to prove CLT.

① First, we find mgf of $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$, $M_n(t)$

mgf of $N(0,1)$, $M(t) = e^{t^2/2}$.

② Show $\lim_{n\to\infty} M_n(t) = e^{t^2/2}$ for $|t| < h$, where $h > 0$.

Step 1: Find mgf of $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$.

First, see that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)}{\sigma}$$

$$= \sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}.$$

Let $Y_i = \frac{X_i - \mu}{\sigma}$. Then $E(Y_i) = 0$ & $Var(Y_i) = 1$.

So

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^{n} Y_i.$$

Assume Y_i has mgf $M(t)$, & $M^{(k)}(t)$ exists for any k.

The mgf of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}$ is

$$M_n(t) = \prod_{i=1}^{n} M_{\frac{Y_i}{\sqrt{n}}}(t) \qquad \text{(as } Y_i \text{ are independent)}$$

$$= \prod_{i=1}^{n} M_{Y_i}\left(\frac{t}{\sqrt{n}}\right)$$

$$= \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Then we know

$M(0) = 1$ (by definition of $M(t)$)

$M'(0) = E(Y_i) = 0$

$M''(0) = E(Y_i^2)$

$= Var(Y_i) + E(Y_i)^2$

$= 1 + 0^2 = 1.$

We want to show $\lim_{n\to\infty}\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = e^{\frac{t^2}{2}}$.

Consider the Taylor expansion

$$M\left(\frac{t}{\sqrt{n}}\right) = \frac{M(0)}{0!} + \frac{t}{\sqrt{n}}\frac{M'(0)}{1!} + \frac{t^2}{n}\frac{M''(0)}{2!} + O\left(\frac{1}{n}\right)$$

$$= 1 + \frac{t}{\sqrt{n}}(0) + \frac{t^2}{2n}(1) + O\left(\frac{1}{n}\right)$$

$$= 1 + \frac{t^2}{2n} + O\left(\frac{1}{n}\right).$$

So

$$\lim_{n\to\infty}\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + O\left(\frac{1}{n}\right)\right]^n = \lim_{n\to\infty}\left[1 + \left(\frac{t^2}{2}\right)\frac{1}{n} + O\left(\frac{1}{n}\right)\right]^n$$

$$= e^{t^2/2}. \qquad \text{(by def}^n \text{ of } e^x\text{)}$$

Since the mgf of $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ converges to that of $N(0,1)$ for any $t \in \mathbb{R}$,

it follows that $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$.

EXAMPLE 1

Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \chi^2_1$, $Y_n = \sum_{i=1}^{n} X_i$.

Show that

$$\frac{Y_n - n}{\sqrt{2n}} \xrightarrow{d} Z \sim N(0,1).$$

Sol^n. We know

$$\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_i)\right)}{\sqrt{Var(X_i)}} \xrightarrow{d} N(0,1).$$

Previously, we showed $E(X_i) = E(\chi^2_1) = 1$ & $Var(X_i) = 2$.

Then

$$\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_i)\right)}{\sqrt{Var(X_i)}} = \frac{1}{\sqrt{n}} \cdot \frac{Y_n - n}{\sqrt{2}} = \frac{Y_n - n}{\sqrt{2n}},$$

and so

$$\frac{Y_n - n}{\sqrt{2n}} \xrightarrow{d} N(0,1).$$

EXAMPLE 2

Let $X_1, \ldots, X_n \overset{iid}{\sim} Poi(\mu)$, & $Y_n = \sum_{i=1}^{n} X_i$.

Find the limiting distribution of $\frac{Y_n - n\mu}{\sqrt{n\mu}}$.

Sol^n. We can directly apply CLT to solve this.

Note

$$\frac{Y_n - n\mu}{\sqrt{n\mu}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\mu}} = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)}{\sqrt{\mu}}$$

$$= \frac{\sqrt{n}(\bar{X}_n - E(X_i))}{\sqrt{Var(X_i)}}$$

which $\xrightarrow{d} N(0,1)$ by CLT.

CONTINUOUS MAPPING THEOREM

Let g be a continuous function.

Then

① If $X_n \xrightarrow{P} X$, then $g(X_n) \xrightarrow{P} g(X)$.

② If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$.

SLUTSKY'S THEOREM

Let $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{P} a$ (or $Y_n \xrightarrow{d} a$), where "a" is a constant.

Then

① $X_n + Y_n \xrightarrow{d} X + a$; &

② $X_n Y_n \xrightarrow{d} aX$

③ $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{a}$ (if $a \neq 0$).

* If $Y_n \xrightarrow{d} Y$, we cannot say $X_n + Y_n \xrightarrow{d} X + Y$!

eg Take $X_1 = \cdots = X_n = \cdots = Z \sim N(0,1)$, & $Y_n = X_n$.

Take $X = Z$, $Y = -Z$.

Then $X_n \xrightarrow{d} X$ & $Y_n \xrightarrow{d} Y$.

But $X + Y = 0$ & $X_n - Y_n \sim N(0,2)$. So $X_n - Y_n \not\xrightarrow{d} X + Y$.

EXAMPLE 4

Let $X_1, \dots, X_n \overset{iid}{\sim} Poi(\mu)$. Find the limiting distribution of $U_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\bar{X}}}$ & $V_n = \sqrt{n}(\bar{X}_n - \mu)$.

Sol^n. We showed

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}} \xrightarrow{d} N(0,1).$$

By WLLN, $\bar{X}_n \xrightarrow{P} \mu$.

Thus, by cts mapping theorem, it follows that

$$\sqrt{\bar{X}_n} \xrightarrow{P} \sqrt{\mu}.$$

So, by Slutsky's theorem,

$$U_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\bar{X}_n}} = \underbrace{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}}}_{Z_n} \cdot \frac{\sqrt{\mu}}{\sqrt{\bar{X}_n}}$$

Then $Z_n \xrightarrow{d} N(0,1)$, & by CLT, $\frac{\sqrt{\mu}}{\sqrt{\bar{X}_n}} \xrightarrow{P} 1$ as $\sqrt{\bar{X}_n} \xrightarrow{P} \sqrt{\mu}$.

Thus, by Slutsky's theorem,

$$U_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\bar{X}_n}} \xrightarrow{d} N(0,1).$$

Next, consider

$$V_n = \sqrt{n}(\bar{X}_n - \mu) = \underbrace{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}}}_{Z_n} \cdot \sqrt{\mu}.$$

Then $Z_n \xrightarrow{d} N(0,1)$. So, by Slutsky's theorem,

$$V_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}} \cdot \sqrt{\mu} = \sqrt{\mu} \cdot Z_n \xrightarrow{d} \sqrt{\mu} \cdot Z.$$

Then $Z \sim N(0,1) \Rightarrow \sqrt{\mu}\, Z \sim N(0, \mu)$.

EXAMPLE 5

Note

① $X_n \xrightarrow{P} a \Rightarrow X_n^2 \xrightarrow{P} a^2$ & $\sqrt{X_n} \xrightarrow{P} \sqrt{a}$ (if $a, X_n \geq 0$)

② If $X_n \xrightarrow{d} Z \sim N(0,1)$, then

$$2X_n \xrightarrow{d} 2Z \sim N(0,4) \quad \&$$
$$X_n^2 \xrightarrow{d} Z^2 \sim \chi_1^2.$$

③ If $X_n \xrightarrow{d} X \sim N(0,1)$, $Y_n \xrightarrow{P} b \neq 0$, then

$$X_n + Y_n \xrightarrow{d} X + b \sim N(b, 1)$$
$$X_n Y_n \xrightarrow{d} bX \sim N(0, b^2)$$
$$\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{b} \sim N(0, \tfrac{1}{b^2}).$$

EXAMPLE 6

Let $X_1, \dots, X_n \overset{iid}{\sim} Unif[0,1]$, & $U_n = \max_{1 \le i \le n} X_i$.

Find the limiting distribution of

① e^{U_n}

② $\sin(1 - U_n)$

③ $e^{-n(1-U_n)}$

④ $(U_n + 1)^2 [n(1 - U_n)]$.

① We know $U_n \xrightarrow{P} 1$ (from earlier); ie $U_n \xrightarrow{d} 1$.

So, by cts mapping thm, $e^{U_n} \xrightarrow{d} e^1 = e$.

② Since $U_n \xrightarrow{d} 1$, so by cts mapping thm

$$\sin(1 - U_n) \xrightarrow{d} \sin(1-1) = 0.$$

③ We have shown

$$n(1 - U_n) \xrightarrow{d} X, \quad X \sim Exp(1).$$

So, by cts mapping thm,

$$e^{-n(1-U_n)} \xrightarrow{d} e^{-X}, \quad X \sim Exp(1).$$

Let $Y = e^{-X}$. The support of Y is $(0,1)$, as $X \in (0, \infty)$.

So, when $y \le 0$, $P(Y \le y) = 0$, & $y \ge 1$, $P(Y \le y) = 1$.

When $0 < y < 1$,

$$\begin{aligned} P(Y \le y) &= P(e^{-X} \le y) \\ &= P(X \geq -\ln(y)) \\ &= \int_{-\ln(y)}^{\infty} e^{-x}\, dx \\ &= [-e^{-x}]_{-\ln(y)}^{\infty} \\ &= e^{-\ln(y)} - 0 \\ &= y. \end{aligned}$$

Thus the pdf of Y is

$$g(y) = \begin{cases} 1 & , 0 \le y \le 1 \\ 0 & , \text{otherwise}; \end{cases}$$

ie $Y \sim Unif[0,1]$.

④ Since $U_n \xrightarrow{d} 1$, thus by cts mapping thm

$$(U_n + 1)^2 \xrightarrow{d} (1+1)^2 = 4.$$

Thus $(U_n + 1)^2 \xrightarrow{P} 4$ as well.

Then $n(1 - U_n) \xrightarrow{d} X$, where $X \sim Exp(1)$.

So, by Slutsky's theorem,

$$(U_n + 1)^2 [n(1 - U_n)] \xrightarrow{d} 4X.$$

Let $Y = 4X$. The support of Y is $[0, \infty)$.

So when $y \le 0$, $P(Y \le y) = 0$.

When $y > 0$,

$$\begin{aligned} P(Y \le y) &= P(4X \le y) \\ &= P(X \le \tfrac{y}{4}) \\ &= \int_0^{y/4} e^{-x}\, dx \\ &= [-e^{-x}]_0^{y/4} \\ &= 1 - e^{-\frac{y}{4}}. \end{aligned}$$

Thus the pdf of Y is $\frac{1}{4} e^{-\frac{1}{4}y}$, $y \ge 0$.

In particular $Y \sim Exp(4)$.

DELTA METHOD

We use this to find the limiting distribution of $g(\bar{X})$, $\sqrt{n}(g(\bar{X}) - g(\mu))$

Suppose that $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} X \sim N(0, \sigma^2)$, and g is differentiable at $x = \mu$, $g'(\mu) \neq 0$.

Then

$$\boxed{\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{d} W \sim N(0, (g'(\mu))^2 \sigma^2).}$$

How to understand this result?

Using first-order Taylor expansion,

$$g(\bar{X}) = g(\mu) + g'(\mu)(\bar{X} - \mu) + \text{high order}$$

$$\therefore \sqrt{n}[g(\bar{X}) - g(\mu)] = \sqrt{n}[g'(\mu)(\bar{X} - \mu)] + \underbrace{\sqrt{n}[\text{high order}]}_{\text{negligible}}$$

So

$$\sqrt{n}[g(\bar{X}) - g(\mu)] \approx \underbrace{\sqrt{n}(\bar{X} - \mu)}_{\xrightarrow{d} N(0,\sigma^2)} \cdot \underbrace{g'(\mu)}_{\text{constant}}.$$

Thus, by cts mapping thm,

$$g'(\mu) N(0, \sigma^2) = N(0, [g'(\mu)]^2 \sigma^2),$$

which shows the result.

Thus

$$\boxed{g(\bar{X}) \sim N\left(g(\mu), \frac{g'(\mu)^2 \sigma^2}{n}\right) \text{ approximately}}$$

Idea. Since $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$,

equivalently, $\bar{X} \approx N(\underset{\text{approx mean}}{\mu}, \underset{\text{approx variance}}{\frac{\sigma^2}{n}})$ approximately.

What is the approximate distribution of $g(\bar{X})$?

↳ Delta Method tells us that

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{d} N(0, (g'(\mu))^2 \sigma^2).$$

In other words,

$$g(\bar{X}) \approx N\left(g(\mu), \frac{g'(\mu)^2 \sigma^2}{n}\right).$$

EXAMPLE 1

Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Poi}(\mu)$. Find the limiting distribution of $Z_n = \sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\mu})$.

Sol^n. From prev results, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \mu)$.

Take $g(x) = \sqrt{x}$. By Delta method,

$$\sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\mu}) \xrightarrow{d} N(0, [g'(\mu)]^2 \sigma^2),$$

where $g'(\mu) = \frac{1}{2} x^{-\frac{1}{2}} \Big|_{x=\mu} = \frac{1}{2\sqrt{\mu}}$.

Since $\sigma^2 = \mu$, thus $[g'(\mu)]^2 \cdot \sigma^2 = \frac{1}{4}$.

$$\therefore \sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\mu}) \xrightarrow{d} N(0, \tfrac{1}{4}).$$

EXAMPLE 2

Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Exp}(\text{mean} = \theta)$.

Find the limiting distribution of

① $\bar{X}_n$;

② $Z_n = \dfrac{\sqrt{n}(\bar{X}_n - \theta)}{\bar{X}_n}$;

③ $U_n = \sqrt{n}(\bar{X}_n - \theta)$; &

④ $V_n = \sqrt{n}(\log \bar{X}_n - \log \theta)$.

* pdf $= \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $x \geq 0$

① $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. $E(X_i) = \theta$, $Var(X_i) = \theta^2 < \infty$.

$\therefore$ By WLLN, $\bar{X}_n \xrightarrow{P} \theta$.

② $$Z_n = \underbrace{\frac{\sqrt{n}(\bar{X}_n - \theta)}{\theta}}_{\substack{\xrightarrow{d} N(0,1) \\ \text{by CLT}}} \cdot \underbrace{\frac{\theta}{\bar{X}_n}}_{\xrightarrow{P} 1 \text{ (by below)}}$$

If you take $g(x) = \frac{\theta}{x}$, by cts mapping thm

$g(\bar{X}_n) \xrightarrow{P} g(\theta) = 1$.

So, by Slutsky's thm,

$$Z_n \xrightarrow{d} Z \cdot 1 = Z \sim N(0,1).$$

③ $$U_n = \underbrace{\frac{\sqrt{n}(\bar{X}_n - \theta)}{\theta}}_{\substack{\xrightarrow{d} N(0,1) \\ \text{by CLT}}} \cdot \theta.$$

Thus, by cts mapping theorem, if we take $g(x) = \theta x$,

$U_n \xrightarrow{d} \theta Z \sim N(0, \theta^2)$, where $Z \sim N(0,1)$.

④ $V_n = \sqrt{n}(\log \bar{X}_n - \log \theta)$.

We know

$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{d} N(0, \theta^2)$

Take $g(x) = \log x$. $\Rightarrow$ $g'(\theta) = \frac{1}{\theta}$.

So, by Delta method,

$$\sqrt{n}(\log \bar{X}_n - \log \theta) \xrightarrow{d} N(0, (g'(\theta))^2 \theta^2)$$
$$= N(0, \tfrac{1}{\theta^2} \theta^2)$$
$$= N(0,1).$$

Chapter 6: Point Estimation

1. Suppose $X_1, \dots, X_n$ are iid rv from $f(x;\theta)$.
   (either a pf for discrete rvs, or pdf for continuous rvs).
2. Here θ is unknown & consists of a finite number of unknown parameters, ie $\theta = (\theta_1, \dots, \theta_k)^T$.
   θ could be a scalar $(k=1)$ or vector $(k>1)$.

eg[1] $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, 1)$, then $\theta = \mu$, a scalar.

eg[2] $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then $\underbrace{\theta = (\mu, \sigma^2)^T}$, a vector.

we use column vector in statistics (convention).

PARAMETER SPACE: Θ

1. Θ is the parameter space; it contains all possible values of θ.

eg[1] if $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, 1)$: $\Theta = \{\mu : -\infty < \mu < \infty\}$

eg[2] if $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$: $\Theta = \{(\mu, \sigma^2)^T : -\infty < \mu < \infty,\ \sigma^2 > 0\}$

DATA & OBSERVATION

1. If $X_1, \dots, X_n \overset{iid}{\sim} f(x;\theta)$, these are the data.
2. Let $x_1, \dots, x_n$ be the observed values of $X_1, \dots, X_n$; these are not random.

STATISTIC

1. A "statistic" is a function of data & does not depend on θ.
2. We denote it by $T = T(X_1, \dots, X_n)$.

eg $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, 1)$.

$\Rightarrow \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a statistic.

But

$\sqrt{n}(\bar{X}_n - \mu)$ is <u>not</u> a statistic.

ESTIMATOR & ESTIMATE

1. If a statistic $T = T(X_1, \dots, X_n)$ is used to estimate an unknown parameter θ, then T is called an "estimator" of θ.
2. An "estimate" is the observed value of T;
   ie $t = T(x_1, \dots, x_n)$ is an estimate of θ.

eg $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, 1)$ & observed data $(x_1, \dots, x_n)$.

Then

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ (an estimator)

$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ (an estimate; an observed value of $\bar{X}_n$).

3. We use $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ to denote an <u>estimator</u> for θ.

eg $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ & $E(X_1)$ are estimators of μ.

4. We may also write $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ to be an <u>estimate</u> for θ.

METHOD OF MOMENTS (6.2)

Let $X_1, \dots, X_n$ be iid with pf $f(x;\theta)$, or pdf $f(x;\theta)$.

We want to estimate $\theta = (\theta_1, \dots, \theta_k)^T$.

This will be a "method of moments" (MM) estimator.

Method:

① Let $\boxed{\mu_j = E(X_i^j), \quad j = 1, \dots, k}$ be the j^{th} population moment.

Then μ_j is a function of $\theta = (\theta_1, \dots, \theta_k)^T$.

② Let $\boxed{\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j, \quad j = 1, \dots, k.}$

This is the "j^{th} sample moment".

- note $\hat{\mu}_j$ is an unbiased estimator of μ_j.
- $\hat{\theta}$ is an unbiased estimator of θ if $E(\hat{\theta}) = \theta$.

③ Then, we want to choose estimators $\hat{\theta}$ of θ such that

$$\boxed{\mu_j(\hat{\theta}) = \mu_j((\hat{\theta}_1, \dots, \hat{\theta}_k)^T) = \hat{\mu}_j, \quad j = 1, \dots, k.}$$

$\hat{\theta}$ will be the MM estimator of θ.

EXAMPLE 1 (1-D CASE)

Let $X_1, \dots, X_n$ be iid from

a) $Poi(\theta)$

b) $Unif[0, \theta]$; &

c) $f(x;\theta) = \theta x^{\theta - 1}$, $0 < x < 1$, $\theta > 0$.

Find the MM estimator of θ for each of them.

a) $\mu_1 = E(X_i) = \theta$. $\quad \hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Then, the MM estimator $\hat{\theta}$ satisfies

$$\mu_1(\hat{\theta}) = \hat{\mu}_1.$$

$$\hat{\theta} \; " \quad " \; \frac{1}{n}\sum_{i=1}^{n} X_i. \qquad \therefore \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

b) $X_i \sim Unif[0, \theta]$.

Then $\mu_1 = E(X_i) = \frac{\theta}{2}$ & $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Hence, the MM estimator of θ, $\hat{\theta}$, satisfies

$$\mu_1(\hat{\theta}) = \hat{\mu}_1.$$

$$\frac{\hat{\theta}}{2} \; " \quad " \; \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Thus $\hat{\theta} = \frac{2}{n}\sum_{i=1}^{n} X_i = 2\bar{X}$.

c) The pdf of X_i is $\theta x^{\theta - 1}$, $0 < x < 1$.

Then

$$\mu_1 = E(X_i) = \int_0^1 x \cdot \theta x^{\theta - 1}\, dx = \int_0^1 \theta x^{\theta}\, dx = \left[\frac{\theta}{\theta + 1} x^{\theta + 1}\right]_0^1 = \frac{\theta}{\theta + 1},$$

& $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Hence, the MM estimator of θ, $\hat{\theta}$, satisfies

$$\mu_1(\hat{\theta}) = \hat{\mu}_1.$$

In other words,

$$\frac{\hat{\theta}}{\hat{\theta} + 1} = \frac{1}{n}\sum_{i=1}^{n} X_i. \Rightarrow \hat{\theta} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i}{1 - \frac{1}{n}\sum_{i=1}^{n} X_i} = \frac{\bar{X}}{1 - \bar{X}}.$$

EXAMPLE 2 (2-D CASE)

Let $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, ie $\theta = (\mu, \sigma^2)^T$, $k = 2$.

Find the MM estimator of θ.

Soln. $\mu_1 = E(X_i) = \mu$. $\quad \hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\mu_2 = E(X_i^2) = \mu^2 + \sigma^2$. $\quad \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

The MM estimator of μ & σ^2 satisfies

$$\mu_1(\hat{\mu}, \hat{\sigma}^2) = \hat{\mu}_1 \qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

$$\mu_2(\hat{\mu}, \hat{\sigma}^2) = \hat{\mu}_2. \Rightarrow (\hat{\mu})^2 + \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Thus

$$\hat{\mu} = \bar{X}, \; \&$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2.$$

MAXIMUM LIKELIHOOD METHOD (6.3)

This is the most commonly used method for estimating the unknown parameter θ.

LIKELIHOOD FUNCTION

1 Let $X_1, \dots, X_n$ be iid with pf or pdf $f(x;\theta)$.
Let $(x_1, \dots, x_n)$ be the observed values of $(X_1, \dots, X_n)$.
We calculate the joint pf or pdf of $(X_1, \dots, X_n)$ at observed value $(x_1, \dots, x_n)$.

2 Discrete case:
We want the joint pdf of $(X_1, \dots, X_n)$ at $(x_1, \dots, x_n)$:

$$\boxed{P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} P(X = x_i) = \prod_{i=1}^{n} f(x_i;\theta)}$$

(since X_i are iid)

3 Continuous case:
We want the joint pdf of $(X_1, \dots, X_n)$ at $(x_1, \dots, x_n)$:

$$\boxed{f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} f(x_i;\theta)}$$

since X_i's are iid

4 We use $L(\theta; x_1, \dots, x_n)$, or simply $L(\theta)$, to denote the likelihood function.

That is,

$$\boxed{L(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i;\theta).}$$

- the derivation is different in discrete vs cts
- but result is the same.

L is called the "likelihood function of θ".

5 Remark:

① "Likelihood" measures the possibility we get the observed value for a given θ.

② Smaller $L(\theta)$ indicates θ is less likely to generate the observed data;

③ Larger $L(\theta)$ indicates θ is more likely to generate the observed data.

METHOD

Idea: Pick θ to maximize $L(\theta)$; ie pick θ such that it is most likely to generate the observed data.

MAXIMUM LIKELIHOOD (ML) ESTIMATOR/ ESTIMATE

1 The ML estimate maximizes $L(\theta)$;
we use $\hat{\theta} = \theta(x_1, \dots, x_n)$ to denote the estimate.

2 In particular,

$$\boxed{\hat{\theta} = \hat{\theta}(x_1, \dots, x_n) = \underset{\theta \in \Theta}{\operatorname{argmax}}\ L(\theta).}$$

3 The ML estimator is

$$\boxed{\hat{\theta} = \hat{\theta}(X_1, \dots, X_n).}$$

LOG LIKELIHOOD FUNCTION

1 The "log-likelihood function" is

$$\boxed{\ell(\theta) = \log L(\theta).}$$

2 Then,

$$\boxed{\hat{\theta} = \hat{\theta}(x_1, \dots, x_n) = \underset{\theta \in \Theta}{\operatorname{argmax}}\ \ell(\theta) = \underset{\theta \in \Theta}{\operatorname{argmax}}\ L(\theta).}$$

INVARIANCE PRINCIPLE OF ML ESTIMATOR

Let $\tau = g(\theta)$. Then the ML estimator of τ is

$$\boxed{\hat{\tau} = g(\hat{\theta}),}$$

where $\hat{\theta}$ is the ML estimator of θ.

EXAMPLE 1

Let $X_1, \dots, X_n$ be iid from

a) $Poi(\theta)$;

b) $Unif[0, \theta]$;

c) $f(x;\theta) = \theta x^{\theta - 1}$, $0<x<1$, $\theta>0$

d) $N(\mu, \sigma^2)$.

Find the ML estimator of θ.

Soln. a) $L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i;\theta)$

$$= \prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{x_i!}$$

$$= \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} e^{-n\theta}.$$

Thus, the log likelihood function is

$$\ell(\theta) = \left(\sum_{i=1}^n x_i\right)\log\theta - \sum_{i=1}^n \log(x_i!) - n\theta.$$

Then

$$\ell'(\theta) = \left(\sum_{i=1}^n x_i\right)\frac{1}{\theta} - n.$$

Set

$$\ell'(\hat\theta) = \frac{1}{\hat\theta}\sum_{i=1}^n x_i - n = 0.$$

Thus

$$\hat\theta = \frac{1}{n}\sum_{i=1}^n x_i$$

Then, note

$$\ell''(\theta) = -\frac{\sum_{i=1}^n x_i}{\theta^2} < 0;$$

which shows ℓ is concave, and so

$$\hat\theta = \frac{1}{n}\sum_{i=1}^n X_i$$

is the ML estimator (MLE) of θ.

b) $X_i \overset{iid}{\sim} Unif[0,\theta]$, $i=1,\dots,n$.

The likelihood function is

$$L(\theta; x_1,\dots,x_n) = \prod_{i=1}^n f(x_i;\theta)$$

$$= \prod_{i=1}^n \frac{1}{\theta}\underbrace{\mathbb{1}(0\le x_i\le\theta)}$$

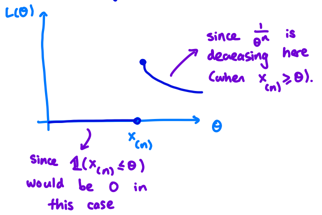
This is 1 if $0 \le x_i \le \theta$, & 0 otherwise

$$= \frac{1}{\theta^n}\prod_{i=1}^n \mathbb{1}(0\le x_i\le\theta).$$

$$= \frac{1}{\theta^n}\prod_{i=1}^n \mathbb{1}(x_{(1)}\ge 0)\,\mathbb{1}(x_{(n)}\le\theta)$$

$x_{(1)} = \min_{1\le i\le n} x_i$ $\quad$ $x_{(n)} = \max_{1\le i\le n} x_i$

Draw a figure:



Thus, $L(\theta)$ is maximized at $\hat\theta = x_{(n)}$.

$\therefore$ The ML estimator of θ is

$$\hat\theta = X_{(n)} = \max_{1\le i\le n} X_i.$$

* this is different from the MM estimator of θ.

c) $X_i \overset{iid}{\sim} f(x;\theta) = \theta x^{\theta-1}$, $0<x<1$.

Thus, the likelihood function is

$$L(\theta; x_1,\dots,x_n) = \prod_{i=1}^n f(x;\theta)$$

$$= \prod_{i=1}^n \theta x_i^{\theta-1}$$

$$= \theta^n\left(\prod_{i=1}^n x_i\right)^{\theta-1}.$$

The log-likelihood function is

$$\ell(\theta) = n\log\theta + (\theta-1)\sum_{i=1}^n \log(x_i).$$

Then

$$\ell'(\theta) = \frac{n}{\theta} + \sum_{i=1}^n \log(x_i).$$

Find $\hat\theta$ that satisfies

$$\ell'(\hat\theta) = \frac{n}{\hat\theta} + \sum_{i=1}^n \log(x_i) = 0.$$

ie

$$\hat\theta = -\frac{n}{\sum_{i=1}^n \log(x_i)}.$$

Then $\ell''(\theta) = -\frac{n}{\theta^2} < 0$, so ℓ is concave.

Thus $\hat\theta = -\frac{n}{\sum_{i=1}^n \log X_i}$ is the MLE of θ.

d) Likelihood function is

$$L(\theta; x_1,\dots,x_n) = \prod_{i=1}^n f(x_i;\mu,\sigma^2)$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}}.$$

Then, the log-likelihood function is

$$\ell(\theta) = \log L(\theta; x_1,\dots,x_n)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}.$$

Then

$$\frac{\partial\ell(\theta)}{\partial\mu} = \frac{\sum_{i=1}^n (x_i-\mu)}{\sigma^2}, \quad \frac{\partial\ell(\theta)}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i-\mu)^2}{2(\sigma^2)^2}$$

$$= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^4}.$$

We want to find $\hat\mu$ & $\hat{\sigma^2}$ s.t.

$$\begin{cases} \left.\frac{d\ell}{d\mu}\right|_{\mu=\hat\mu,\ \sigma^2=\hat{\sigma^2}} = \frac{\sum_{i=1}^n (x_i-\hat\mu)}{\hat{\sigma}^2} = 0 \\ \left.\frac{d\ell}{d(\sigma^2)}\right|_{\mu=\hat\mu,\ \sigma^2=\hat{\sigma^2}} = -\frac{n}{2\hat{\sigma}^2} + \frac{\sum_{i=1}^n (x_i-\hat\mu)^2}{2(\hat{\sigma}^2)^2} = 0. \end{cases}$$

This evaluates to

$$\hat\mu = \frac{1}{n}\sum_{i=1}^n x_i$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (x_i-\hat\mu)^2.$$

So, the MLE of μ, σ^2 is

$$\hat\mu = \frac{1}{n}\sum_{i=1}^n X_i, \quad \hat{\sigma^2} = \frac{\sum_{i=1}^n (X_i-\hat\mu)^2}{n}.$$

PROPERTIES OF ML ESTIMATOR (6·4)

Here,

1. We consider θ to be a scalar $(k=1)$
2. We consider the ML estimator (a r.v.)
3. The support of X_i does not depend on θ.

eg we cannot apply the thms here to $X_1, \dots, X_n \overset{iid}{\sim} \text{Unif}[0,\theta]$.

SCORE FUNCTION : $S(\theta)$

1. The "score function" is

$$\boxed{S(\theta) = S(\theta; x_1, \dots, x_n) = \frac{d\ell(\theta)}{d\theta} = \frac{d \log L(\theta)}{d\theta}}$$

2. If the support of X_i does not depend on θ, then

$$\boxed{S(\hat{\theta}) = 0,}$$

where $\hat{\theta}$ is the ML estimate.

INFORMATION FUNCTION : $I(\theta)$

The "information function" is

$$\boxed{I(\theta) = -\frac{d^2\ell(\theta)}{d\theta} = -\frac{dS(\theta)}{d\theta}.}$$

* I depends on $x_1, \dots, x_n$.

FISHER INFORMATION : $J(\theta)$

1. The "fisher information" is

$$\boxed{J(\theta) = E[I(\theta; x_1, \dots, x_n)].}$$

2. When $X_1, \dots, X_n \overset{iid}{\sim} f(x;\theta)$, then

$$\boxed{\ell(\theta; x_1, \dots, x_n) = \sum_{i=1}^{n} \log f(x_i;\theta),}$$

and so

$$\boxed{\begin{aligned} S(\theta) &= \sum_{i=1}^{n} \frac{d \log f(x_i;\theta)}{d\theta} \\ I(\theta) &= -\sum_{i=1}^{n} \frac{d^2 \log f(x_i;\theta)}{d\theta} \\ J(\theta) &= E[I(\theta; \underbrace{X_1, \dots, X_n}_{\text{these are rvs!}})] = -nE\left[\frac{d^2 \log f(X_i;\theta)}{d\theta^2}\right]. \end{aligned}}$$

3. let

$$\boxed{J_1(\theta) = -E\left[\frac{d^2 \log f(X_i;\theta)}{d\theta^2}\right];}$$

this is the information contained in one observation.

Then $J(\theta) = nJ_1(\theta)$; the information in n observations.

EXAMPLE 1

> Recall: $X_1, \dots, X_n \overset{iid}{\sim} \text{Poi}(\theta)$.
> We showed the MLE of θ is $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
> So $\text{Var}(\hat{\theta}) = \frac{\theta}{n}$.
> Find $J(\hat{\theta})$.

Sol^n. Note

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta}.$$

$$\therefore \ \ell(\theta) = -\sum_{i=1}^{n} \log(x_i!) - n\theta + \left(\sum_{i=1}^{n} x_i\right)\log\theta.$$

Then,

$$\Rightarrow S(\theta) = \frac{d\ell(\theta)}{d\theta} = \frac{1}{\theta}\sum_{i=1}^{n} x_i - n.$$

$$\Rightarrow I(\theta) = -\frac{dS(\theta)}{d\theta} = \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$$

So,

$$\begin{aligned} J(\theta) &= E[I(\theta; X_1, \dots, X_n)] \\ &= E\left(\frac{1}{\theta^2}\sum_{i=1}^{n} X_i\right) \\ &= n \cdot E\left(\frac{X_1}{\theta^2}\right) = \frac{n}{\theta^2}E(X_1) = \frac{n}{\theta}. \end{aligned}$$

Since $\text{Var}(\hat{\theta}) = \frac{\theta}{n}$, thus $\text{Var}(\hat{\theta}) = \frac{1}{J(\theta)}$.

UNBIASED ESTIMATOR

1. An estimator T is said to be an "unbiased" estimator of $\tau = g(\theta)$ if

$$\boxed{E(T) = \tau \quad \forall \theta \in \Theta.}$$

2. If $E(T) \neq \tau$, then T is a "biased" estimator of τ.
3. The "bias" of T is

$$\boxed{\text{Bias}(T) = E(T) - \tau.}$$

CRAMER-RAO (CR) LOWER BOUND

If T is an unbiased estimator of τ, then necessarily

$$\boxed{\text{Var}(T) \geq \frac{[g'(\theta)]^2}{J(\theta)}, \qquad T = T(X_1, \dots, X_n).}$$

MEAN SQUARED ERROR

The "mean squared error" of $\hat{\theta}$ is

$$\boxed{\begin{aligned} MSE(\hat{\theta}) &= E[(\hat{\theta}-\theta)^2] \\ &= [\text{Bias}(\hat{\theta})]^2 + \text{Var}(\hat{\theta}). \end{aligned}}$$

ASYMPTOTIC NORMALITY OF MLE

Under some regularity conditions (one of these conditions is the support of X_i does not depend on θ): if $\hat{\theta}$ is the MLE of θ, then

1. $\boxed{\hat{\theta} \xrightarrow{P} \theta \quad \text{as } n \to \infty.}$] consistency

   ↳ this shows $\hat{\theta}$ is close to θ when n is sufficiently large.

2. $\boxed{\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N\left(0, \frac{1}{J_1(\theta)}\right)}$

3. By Delta method,

$$\boxed{\sqrt{n}(\hat{\tau}-\tau) \xrightarrow{d} N\left(0, \frac{(g'(\theta))^2}{J_1(\theta)}\right),}$$

   where $\tau = g(\theta)$, & $\hat{\tau}$ is the MLE of τ.

4. In particular, ② implies

$$\boxed{\hat{\theta} \approx N\left(\theta, \frac{1}{nJ_1(\theta)}\right),}$$

   so that $E(\hat{\theta}) \approx \theta$ & $\text{Var}(\hat{\theta}) \approx \frac{1}{nJ_1(\theta)} = \frac{1}{J(\theta)}$.

   ↳ this is the CR lower bound!

5. Thus,

$$\boxed{\hat{\tau} = g(\hat{\theta}) \approx N\left(\tau, \frac{[g'(\theta)]^2}{nJ_1(\theta)}\right),}$$

   so that $E(\hat{\tau}) \approx \tau$ & $\text{Var}(\hat{\tau}) \approx \frac{[g'(\theta)]^2}{nJ_1(\theta)} = \frac{[g'(\theta)]^2}{J(\theta)}$.

EXAMPLE 1

Let $X_1, \dots, X_n \overset{iid}{\sim} Poi(\theta)$. Find

① the MLE of θ, $\hat{\theta}$;

② the ML estimator of $\tau = P(X_1 = 0)$, $\hat{\tau}$;

③ the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta)$;

④ the limiting distribution of $\sqrt{n}(\hat{\tau} - \tau)$; &

⑤ $E(\hat{\tau})$ & $E(\hat{\theta})$.

Sol<u>n</u>. ① From previous results,

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

② $\tau = P(X_1 = 0) = e^{-\theta}$.

So, by the invariance property of ML estimator,

$$\hat{\tau} = e^{-\hat{\theta}}.$$

③ $\sqrt{n}(\hat{\theta} - \theta)$.

Method 1: Under regularity conditions (which $Poi(\theta)$ satisfies):

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \frac{1}{J_1(\theta)}).$$

As $J(\theta) = \frac{n}{\theta} = n J_1(\theta) \Rightarrow J_1(\theta) = \frac{1}{\theta}$.

So

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \theta).$$

Method 2: Use CLT:

$$\frac{\sqrt{n}(\hat{\theta} - \theta)}{\sqrt{\theta}} \xrightarrow{d} N(0,1) \Rightarrow \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \sqrt{\theta} \cdot N(0,1)$$
$$= N(0, \theta).$$

④ $\sqrt{n}(\hat{\tau} - \tau)$.

Method 1: Under regularity conditions,

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, \frac{[g'(\theta)]^2}{n J_1(\theta)}).$$

Then $\tau = g(\theta) = e^{-\theta} \Rightarrow [g'(\theta)]^2 = (-e^{-\theta})^2 = e^{-2\theta}$.

Also $J_1(\theta) = \frac{1}{\theta}$ from earlier. Thus

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, \theta e^{-2\theta}).$$

Method 2: Delta method.

Take $g(\theta) = e^{-\theta}$. Then

$$\sqrt{n}(\hat{\tau} - \tau) = \sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} N(0, [g'(\theta)]^2 \cdot \theta)$$
$$= N(0, \theta \cdot e^{-2\theta}).$$

⑤ $\hat{\theta} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$\Rightarrow E(\hat{\theta}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \theta.$$

$$E(\hat{\tau}) = E(e^{-\bar{X}_n})$$
$$= E(e^{-\frac{1}{n}\sum_{i=1}^{n} X_i}).$$

Let $T = \sum_{i=1}^{n} X_i$. $\Rightarrow T \sim Poi(n\theta)$. So

$$E(\hat{\tau}) = E(e^{-T/n}).$$

We could then calculate $E(e^{-T/n}) = M_T(\frac{1}{n})$, or alternatively

$$E(e^{-T/n}) = \sum_{t=0}^{\infty} e^{-\frac{t}{n}} P(T=t)$$
$$= \sum_{t=0}^{\infty} e^{-\frac{t}{n}} \frac{(n\theta)^t}{t!} e^{-n\theta}$$
$$= e^{-n\theta} \sum_{t=0}^{\infty} \frac{(e^{-\frac{1}{n}} n\theta)^t}{t!}$$
$$= e^{e^{-\frac{1}{n}} n\theta - n\theta}.$$

good luck!