STAT 330 - Mathematical Statistics

Cameron Roopnarine

Last updated: March 17, 2021

Contents

Chapter 2

Univariate Random Variable

LECTURE 1 | 2020-09-09

Review probability model, random variable (r.v.), expectation, and moment generating function.

2.1 Probability Model and Random Variable

DEFINITION 2.1.1: Probability model

A **probability model** is used for a random experiment. It has three important components:

(I) Sample space
(II) Event
(III) Probability function

DEFINITION 2.1.2: Sample space

A **sample space** S is the collection of all possible outcomes of one single random experiment.

DEFINITION 2.1.3: Event

An **event** $A, B, \ldots$ is a subset of S.

EXAMPLE 2.1.4

Toss a coin twice.

- $S = \{(H,H),(H,T),(T,H),(T,T)\}$
- A: First toss is a head (H).

Clearly, $A = \{(H,H),(H,T)\} \subseteq S$, so A is an event.

DEFINITION 2.1.5: Probability function

A **probability function** $\mathbb{P}(\,\cdot\,)$ is a function of events and satisfies:

(I) For any event A, $\mathbb{P}(A) \geq 0$

(II) $\mathbb{P}(S) = 1$

(III) *Additivity property*: If $A_1, A_2, A_3, \ldots$ are pairwise mutually exclusive events; that is, $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

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

EXAMPLE 2.1.6

Toss a coin twice, given one event A,

$$\mathbb{P}(A) = \frac{\#\text{ of outcomes in } A}{4}$$

where 4 is the total number of outcomes in S. $\mathbb{P}(\,\cdot\,)$ satisfies the three properties, therefore $\mathbb{P}(\,\cdot\,)$ is a probability function.

PROPOSITION 2.1.7: Additional Properties of the Probability Function

(1) $\mathbb{P}(\emptyset) = 0$ *where* $\emptyset = \bigcup_{i=1}^{\infty} \emptyset$.

(2) *Let* $\bar{A}$ *be the complementary event of* A.

  (i) $\bar{A} \cup A = S$

  (ii) $\bar{A} \cap A = \emptyset$

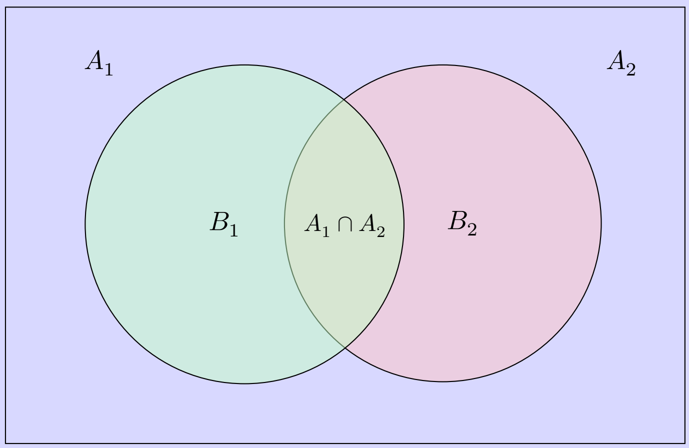
$$\mathbb{P}(A) + \mathbb{P}(\bar{A}) = 1$$

(3) *If* A_1 *and* A_2 *are mutually exclusive, then*

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

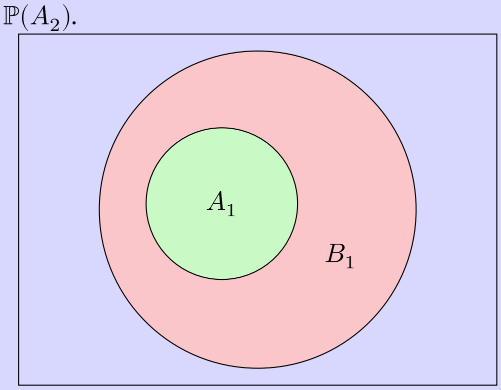
(4) *Generally,*

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



$$A_1 \cup A_2 = B_1 \cup (A_1 \cap A_2) \cup B_2$$

(5) *If* $A_1 \subseteq A_2$, *then* $\mathbb{P}(A_1) \leq \mathbb{P}(A_2)$.



$$A_2 = A_1 \cup B_1$$

Proof of Proposition 2.1.7

Proof of (1): Let $A_1 = \emptyset$, $A_2 = \emptyset$, $A_3 = \emptyset$, ..., then

$$\mathbb{P}(\emptyset) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) = \sum_{i=1}^{n} \mathbb{P}(\emptyset)$$

DEFINITION 2.1.8: Conditional probability

Suppose A and B are two events with $\mathbb{P}(B) > 0$. The **conditional probability** of A given that B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

DEFINITION 2.1.9: Independent events

Suppose A and B are two events. A and B are **independent events** if and only if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Clearly, $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ if and only if A and B are independent since

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

EXAMPLE 2.1.10

Toss a coin twice.

- A: First toss is H
- B: Second toss is T

$$\mathbb{P}(A) = \frac{\text{\# of outcomes in } A}{4} = \frac{2}{4} \quad \text{and} \quad \mathbb{P}(B) = \frac{2}{4}$$

$$\mathbb{P}(A \cap B) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B)$$

Therefore, A and B are independent.

DEFINITION 2.1.11: Random variable

A **random variable** (r.v.) X is a function from a sample space S to the real numbers **R**; that is,

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

satisfies for any given $x \in \mathbf{R}$ $\{X \leq x\}$ is an event.

$$\{X \leq x\} = \{w \in S : X(w) \leq x\} \subseteq S$$

EXAMPLE 2.1.12

Toss a coin twice. Let X be the number of heads (H) in two tosses. Verify that X is a random variable.
Solution. Possible values of X: $0, 1, 2$. Given $x \in \mathbf{R}$, $\{X \leq x\}$.

- $x < 0 \implies \{X \leq x\} = \emptyset$
- $x = 0 \implies \{X \leq x\} = \{X = 0\} = \{(T, T)\} \subseteq S$
- $x = 1 \implies \{X \leq x\} = \{X = 1\} = \{(H, T), (T, H)\} \subseteq S$
- $x = 2 \implies \{X \leq x\} = \{X = 2\} = \{(H, H)\} \subseteq S$

Thus, X is a random variable.

DEFINITION 2.1.13: Cumulative distribution function

The **cumulative distribution function** (c.d.f.) of a random variable X is defined by

$$F(x) = \mathbb{P}(X \leq x) \quad \text{for all } x \in \mathbf{R}$$

Note that the c.d.f. is defined for all **R**.

DEFINITION 2.1.14: Properties — Cumulative Distribution Function

(1) F is a non-decreasing function; that is, if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
By looking at:
- $\{X \leq x_1\} \subseteq \{X \leq x_2\}$ if $x_1 \leq x_2$.

(2) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.
By looking at:
- $\{X \leq x\} \to S$ as $x \to \infty$.
- $\{X \leq x\} \to \emptyset$ as $x \to -\infty$.

(3) $F(x)$ is a right continuous function; that is, for any $a \in \mathbf{R}$,

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

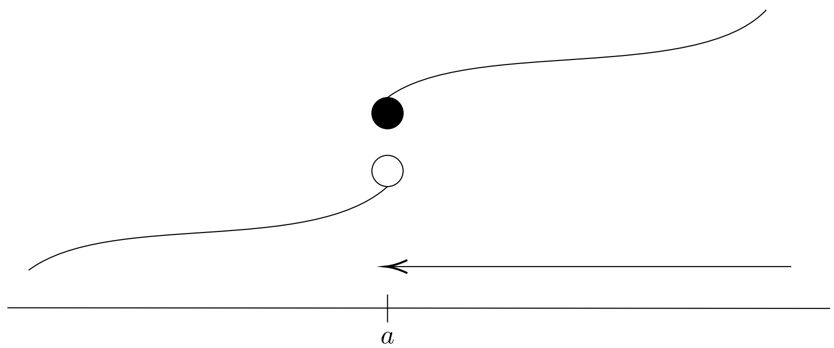


Figure 2.1: Right Continuous

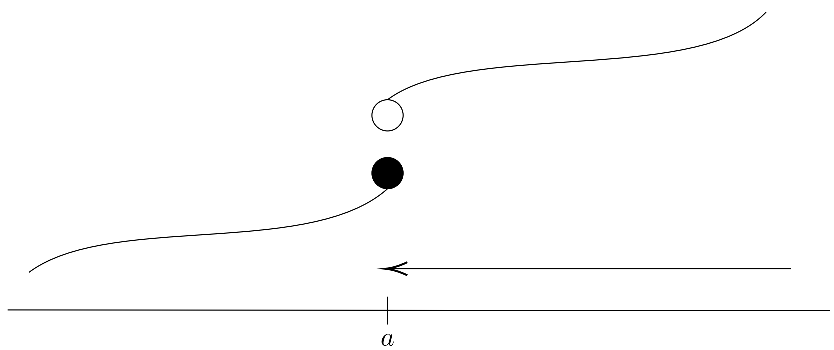


Figure 2.2: Not Right Continuous

PROPOSITION 2.1.15: Additional Properties of Cumulative Distribution Function

(1) $\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F(b) - F(a)$
(2) $\mathbb{P}(X = x) = \mathbb{P}(\textit{Jump at } x) = \lim_{t\to x^+} F(t) - \lim_{t\to x^-} F(t) = F(x) - \lim_{t\to x^-} F(t)$

LECTURE 2 | 2020-09-09

2.2 Discrete Random Variables

DEFINITION 2.2.1: Discrete random variable

If a random variable X can only take finite or countable values, X is a **discrete random variable**.

REMARK 2.2.2

When we say **countable**, we mean something you can enumerate such as **Z** or $\mathbf{N}^+$.

DEFINITION 2.2.3: Probability function

If X is a discrete random variable, then the **probability function** (p.f.) of X is given by

$$f(x) = \begin{cases} \mathbb{P}(X = x) & \text{if } X \text{ can take value } x \\ 0 & \text{if } X \text{ cannot take value } x \end{cases}$$

DEFINITION 2.2.4: Support

The set $A = \{x : f(x) > 0\}$ is called the **support** of X. These are all the possible values X can take.

PROPOSITION 2.2.5: Properties of the Probability Function

(1) $f(x) \geq 0$ *for all* $x \in \mathbf{R}$
(2) $\sum_{x \in A} f(x) = 1$

Review some commonly used discrete random variables:

- **Bernoulli.** $X \sim$ Bernoulli(p) where X can only take two possible values 0 (failure) or 1 (success). Let p be the probability of a success for a single trial. So,

$$\mathbb{P}(X = 1) = p \quad \text{and} \quad \mathbb{P}(X = 0) = 1 - p$$

  EXAMPLE 2.2.6: Bernoulli

  Toss a coin twice. Let X be the number of heads. Therefore, $X \sim$ Bernoulli(p).

- **Binomial.** $X \sim$ Binomial(n, p). Suppose we have **Bernoulli Trials**:
  - We run n trials
  - Each trial is independent of each other
  - Each trial has two possible outcomes: 0 (failure), 1 (success)

$$\mathbb{P}(X = 1) = p$$

  Let X be the number of success across these n trials and p be the success probability for a single trial.

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

  X_i is the outcome of the i^{th} trial.

$$\mathbb{P}(X_i = 1) = p$$

  where $X_i \sim$ Bernoulli(p).

- **Geometric.** $X \sim$ Geometric(p). Let X be the number of failures before the first success.

EXAMPLE 2.2.7: Geometric

X = number of tails before you get the first head. Therefore, $X \sim \text{Geometric}(p)$.

X can take values $0, 1, 2, \ldots$.

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

- **Negative Binomial.** $X \sim \text{Negative Binomial}(r, p)$. Let X be the number of failures before you get r success. X can take values $0, 1, 2, \ldots$

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

EXAMPLE 2.2.8: Negative Binomial

X = number of tails before you get the r^{th} head. Therefore, $X \sim \text{Negative Binomial}(r, p)$.

- **Poisson.** $X \sim \text{Poisson}(\mu)$

$$f(x) = \mathbb{P}(X = x) = \frac{\mu^x}{x!} e^{-\mu}$$

where $0 \leq x \in \mathbf{Z}$.

2.3 Continuous Random Variables

DEFINITION 2.3.1: Continuous random variable

If the possible values of X is an interval or real line, X is a **continuous random variable**. In this case, $F(x)$ is continuous and differentiable almost everywhere. (It's not differentiable for at most a countable set of points).

Note that this is not a rigorous definition, but it will be used in this course.

DEFINITION 2.3.2: Probability density function, Support

The **probability density function** (p.d.f.) of a continuous random variable is

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

The set $A = \{x : f(x) > 0\}$ is called the **support** of X.

Continuous case: $f(x) \neq \mathbb{P}(X = x)$

$$\mathbb{P}(x < X \leq x + \delta) \approx f(x)\delta$$

since

$$\lim_{\delta \to 0} \frac{F(x+\delta) - F(x)}{\delta} = F'(x) = f(x)$$

where $F(x + \delta) - F(x) = \mathbb{P}(x \leq X \leq x + \delta)$.

DEFINITION 2.3.3: Properties — Probability Density Function

(I) $f(x) \geq 0$ for all $x \in \mathbf{R}$

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

(III) $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\,dt$ with $F(-\infty) = 0$

(IV) $f(x) = F'(x)$

(V) $\mathbb{P}(X = x) = 0 \neq f(x)$

(VI) $\mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = F(b) - F(a) = \displaystyle\int_{a}^{b} f(x)\,dx$ since $\mathbb{P}(X = a) = \mathbb{P}(X = b) = 0$.

EXAMPLE 2.3.4

Suppose the c.d.f. of X is

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

Find the p.d.f. of X.

Solution.

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

We note that $X \sim \text{Uniform}(a, b)$.

EXAMPLE 2.3.5

Let the p.d.f. be defined as follows.

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

(i) For what values of θ is f a p.d.f.

(ii) Find $F(x)$.

(iii) Find $\mathbb{P}(-2 < X < 3)$.

Solution.

(i) Note that $\dfrac{\theta}{x^{\theta+1}} \geq 0$ for all $\theta \geq 0$.

Case 1: $\theta = 0$. $f(x) \equiv 0$, then f cannot be a p.d.f. since $\int_{-\infty}^{\infty} f(x)\,dx = 0 \neq 1$

Case 2: $\theta > 0$.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx = \int_{1}^{\infty} \frac{\theta}{x^{\theta+1}}\,dx = \Big[-x^{-\theta}\Big]_{1}^{\infty} = 1$$

Therefore, f is a p.d.f. when $\theta > 0$.

(ii) $F(x) = \mathbb{P}(X \leq x)$.

Case 1: $x < 1$.

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = 0$$

Case 2: $x \geq 1$.

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{1} f(t)\,dt + \int_{1}^{x} f(t)\,dt = \int_{1}^{x} \frac{\theta}{t^{\theta+1}}\,dt = \left[-t^{-\theta}\right]_{1}^{x} = 1 - x^{-\theta}$$

Therefore,

$$F(x) = \begin{cases} 1 - x^{-\theta} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

(iii) $\mathbb{P}(-2 < X < 3)$. Either use the c.d.f. we found or the p.d.f.
Using the c.d.f. we have

$$F(3) - F(-2) = (1 - 3^{-\theta}) - 0$$

Using the p.d.f. we have

$$\int_{-2}^{3} f(x)\,dx = \int_{-2}^{1} f(x)\,dx + \int_{1}^{3} f(x)\,dx = \int_{1}^{3} f(x)\,dx = \text{exercise}$$

Lecture 3 | 2020-09-13

We first introduce a function that will be used.

DEFINITION 2.3.6: Gamma function

The **gamma function**, denoted $\Gamma(\alpha)$ for all $\alpha > 0$, is given by

$$\Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1} e^{-x}\,dx$$

PROPOSITION 2.3.7: Properties of the Gamma Function

(1) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ *for* $\alpha > 1$
(2) $\Gamma(n) = (n-1)!$ *when* $n \in \mathbf{Z}^{+}$*, where* $\Gamma(1) = 1$.
(3) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

EXAMPLE 2.3.8

The p.d.f. is given by

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

where $\alpha > 0$ and $\beta > 0$. $X \sim \text{Gamma}(\alpha, \beta)$.
We also say that α is the scale parameter and β is the shape parameter for this distribution.
Verify that $f(x)$ is a p.d.f.
Solution. Showing $f(x) \geq 0$ is trivial. Now,

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{\infty} \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}\,dx$$

Let $y = x/\beta \implies x = y\beta$ and $dx = \beta\, dy$. Therefore,

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

EXAMPLE 2.3.9

Suppose the p.d.f. is given by

$$f(x) = \begin{cases} \dfrac{\beta}{\theta^{\beta}}x^{\beta-1}\exp\left\{-\left(\dfrac{x}{\theta}\right)^{\beta}\right\} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

with $\alpha > 0$ and $\beta > 0$. $X \sim$ Weibull(θ, β).
Verify that $f(x)$ is a p.d.f.
Solution. $f(x) \geq 0$ for every $x \in \mathbf{R}$. Now,

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

Let $y = (x/\theta)^{\beta} \implies x = \theta y^{1/\beta}$ and $dx = (\theta/\beta)y^{(1/\beta)-1}\, dy$. Therefore,

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

EXAMPLE 2.3.10: Normal

The p.d.f. is given by

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

for $x \in \mathbf{R}$, $-\infty < \mu < \infty$, $\sigma^2 > 0$. Verify that $f(x)$ is a p.d.f.
Solution. $f(x) \geq 0$ for every $x \in \mathbf{R}$.
Case 1: $\mu = 0$ and $\sigma^2 = 1$, then we say X follows a **standard normal** distribution. We want to show that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\} dx = 1$$

Since the function is symmetrical around 0, we have the following equivalent integral.

$$2\int_0^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\} dx$$

Let $y = x^2/2 \implies x = \sqrt{2y}$ and $dx = \dfrac{\sqrt{2}}{2}y^{-1/2}\, dy$. Therefore,

$$= \frac{2}{\sqrt{2\pi}}\int_0^{\infty} e^{-y}\frac{\sqrt{2}}{2}y^{-1/2}\, dy = \frac{1}{\sqrt{\pi}}\int_0^{\infty} y^{1/2-1}e^{-y}\, dy = \left(\frac{1}{\sqrt{\pi}}\right)\Gamma\left(\frac{1}{2}\right) = 1$$

Case 2: For general μ and σ^2,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx$$

Let $z = \dfrac{x-\mu}{\sigma} \implies x = \mu + \sigma z$ and $dx = \sigma\,dz$. Therefore,

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{z^2}{2}\right\} \sigma\,dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} dz = 1$$

using Case 1.

2.4 Expectation

DEFINITION 2.4.1: Expectation (Discrete)

Suppose X is a discrete random variable with support A and p.f. $f(x)$. Then,

$$\mathbb{E}[X] = \sum_{x\in A} x f(x)$$

if $\sum_{x\in A} |x| f(x) < \infty$ (finite). If $\sum_{x\in A} |x| f(x) = \infty$ (infinite), then $\mathbb{E}[X]$ does not exist.

DEFINITION 2.4.2: Expectation (Continuous)

Suppose X is a continuous random variable with support A and p.d.f. $f(x)$. Then,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,dx$$

if $\int_{-\infty}^{\infty} |x| f(x)\,dx < \infty$ (finite). Similarly, if $\int_{-\infty}^{\infty} |x| f(x)\,dx = \infty$ (infinite), then $\mathbb{E}[X]$ does not exist.

EXAMPLE 2.4.3: Discrete

Suppose

$$f(x) = \frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{x+1}$$

for $x = 1, 2, \ldots$. The support set is $A = \{1, 2, \ldots\}$. We note that $f(x)$ is a p.f. since $f(x) \geq 0$ and

$$\sum_{x\in A} f(x) = \sum_{x=1}^{\infty} \left(\frac{1}{x} - \frac{1}{x+1}\right) = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \cdots = 1$$

Find $\mathbb{E}[X]$.
Solution.

$$\sum_{x\in A} |x| f(x) = \sum_{x=1}^{\infty} x\left(\frac{1}{x} - \frac{1}{x+1}\right) = \sum_{x=1}^{\infty} \frac{1}{x+1} = \infty$$

Therefore, $\mathbb{E}[X]$ does not exist!

EXAMPLE 2.4.4: Continuous

Let the p.d.f. be defined as $f(x) = \dfrac{1}{x^2+1}$ for $x \in \mathbf{R}$. This is known as the Cauchy distribution (or Student's T-distribution with 1 degree of freedom). Find $\mathbb{E}[X]$.

Solution.

$$\int_{-\infty}^{\infty} |x| f(x)\, dx = \int_{-\infty}^{\infty} |x| \frac{1}{x^2+1}\, dx = 2\int_{0}^{\infty} \frac{x}{x^2+1}\, dx = \left[\ln|x^2+1|\right]_0^{\infty} = \infty$$

$\mathbb{E}[X]$ does not exist! The following is **wrong**:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{-\infty}^{\infty} \frac{x}{1+x^2}\, dx = 0$$

since the integral above with $|x|$ is infinite. You must always remember to check that the $\mathbb{E}[X]$ is finite (using $|x|$) for both the discrete and continuous case whenever the support is **negative**.

EXAMPLE 2.4.5: Bernoulli and Binomial Random Variable

Suppose $X \sim \text{Bernoulli}(p)$.

$$\mathbb{P}(X=1) = p \quad \text{and} \quad \mathbb{P}(X=0) = 1-p$$

We know $\mathbb{E}[X] = (1)\mathbb{P}(X=1) + (0)\mathbb{P}(X=0) = p$.
Now suppose $X \sim \text{Binomial}(n,p)$. Find $\mathbb{E}[X]$.
Solution.

$$\mathbb{E}[X] = \sum_{x\in A} x f(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

This is hard to do. But, we know we can use the relationship between the Binomial and Bernoulli random variable so,

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

Therefore,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np$$

EXAMPLE 2.4.6

Suppose for a random variable X the p.d.f. is given by $f(x) = \dfrac{\theta}{x^{\theta+1}}$ for $x \geq 1$ and 0 when $x < 1$. Assume $\theta > 0$. Find $\mathbb{E}[X]$, and determine the values of θ for which $\mathbb{E}[X]$ exists.
Solution.

$$\int_{-\infty}^{\infty} |x| f(x)\, dx = \int_{1}^{\infty} (x) \frac{\theta}{x^{\theta+1}}\, dx = \theta \int_{1}^{\infty} \frac{1}{x^{\theta}}\, dx < \infty \iff \theta > 1$$

from MATH 138. Therefore, if $\theta > 1$ then $\mathbb{E}[X]$ exists. Also,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{1}^{\infty} \frac{\theta x}{x^{\theta+1}}\, dx = \theta \int_{1}^{\infty} \frac{1}{x^{\theta}}\, dx = \frac{\theta}{\theta-1}$$

DEFINITION 2.4.7: Expectation (Discrete)

If X is a discrete random variable with probability function $f(x)$ and support set A, then the **expectation** of the random variable $g(X)$ is defined by

$$\mathbb{E}[g(X)] = \sum_{x \in A} g(x) f(x)$$

provided the sum converges absolutely; that is, provided

$$\sum_{x \in A} |g(x)| f(x) < \infty$$

DEFINITION 2.4.8: Expectation (Continuous)

If X is a continuous random variable with p.d.f. $f(x)$ and support set A, then the **expectation** of the random variable $g(X)$ is defined by

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

provided the integral converges absolutely; that is, provided

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

THEOREM 2.4.9: Expectation is a Linear Operator

Suppose X is a random variable with probability (density) function $f(x)$, and a and b are real constants, and $g(x)$ and $h(x)$ are real-valued functions. Then,

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

$$\mathbb{E}[ag(X) + bh(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)]$$

DEFINITION 2.4.10: Variance

The variance of a random variable is defined as

$$\sigma^2 = \mathbb{V}(X) = \mathbb{E}[(X - \mu)^2]$$

where $\mu = \mathbb{E}[X]$.

DEFINITION 2.4.11: Special Expectations

(I) The k^{th} moment (about 0): $\mathbb{E}[X^k]$
(II) The k^{th} moment about the mean $\mathbb{E}[(X - \mu)^k]$

THEOREM 2.4.12: Properties of Variance

If X is a random variable, then

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2$$

where $\mu = \mathbb{E}[X]$. Note that the variance of X exists if $\mathbb{E}[X^2] < \infty$.

EXAMPLE 2.4.13

Suppose $X \sim \text{Poisson}(\theta)$, the p.f. is defined as $f(x) = \frac{\theta^x}{x!}e^{-\theta}$ for $0 \leq x \in \mathbf{Z}$. Find $\mathbb{E}[X]$ and $\mathbb{V}(X)$.
Solution. The support is non-negative, so $|x| = x$. Therefore,

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

Let $y = x - 1$, then

$$\mathbb{E}[X] = \theta \sum_{y=0}^{\infty} \frac{\theta^y}{y!}e^{-\theta} = \theta(e^{-\theta})e^{\theta}$$

since we know $e^{\theta} = \sum_{y=0}^{\infty} \frac{\theta^y}{y!}$. Therefore, $\mathbb{E}[X] = \theta$.

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2$$

Let's find $\mathbb{E}[X^2]$:

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{x=0}^{\infty} x^2\frac{\theta^x}{x!}e^{-\theta} \\
&= \sum_{x=1}^{\infty} \frac{x}{(x-1)!}\theta^x e^{-\theta} \\
&= \sum_{x=1}^{\infty} \frac{(x-1)+1}{(x-1)!}\theta^x e^{-\theta} \\
&= \sum_{x=1}^{\infty} \frac{x-1}{(x-1)!}\theta^x e^{-\theta} + \sum_{x=1}^{\infty} \frac{1}{(x-1)!}\theta^x e^{-\theta}
\end{aligned}$$

Looking at the first sum (since the second sum was computed before and is θ)

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

Let $y = x - 2$:

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

Therefore,

$$\mathbb{E}[X^2] = \theta^2 + \theta$$

Therefore,

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2 = (\theta^2 + \theta) - \theta^2 = \theta$$

LECTURE 4 | 2020-09-13

EXAMPLE 2.4.14

If $X \sim \text{Gamma}(\alpha, \beta)$, prove that

$$\mathbb{E}[X^p] = \frac{\beta^p \Gamma(\alpha + p)}{\Gamma(\alpha)}$$

for $p > -\alpha$.

Solution. Recall that

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

So,

$$\mathbb{E}[X^p] = \int_{-\infty}^{\infty} x^p f(x)\,dx = \int_0^\infty x^p \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}\,dx$$

There are two methods to solve this integral:
Method 1: Rewrite the function as the p.d.f. of a gamma distribution.

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

which is close to the p.d.f. of Gamma$(p+\alpha, \beta)$.

$$= \int_0^\infty \frac{x^{p+\alpha-1}e^{-x/\beta}}{\Gamma(\alpha+p)\beta^{\alpha+p}} \times \underbrace{\frac{\Gamma(\alpha+p)\beta^{\alpha+p}}{\Gamma(\alpha)\beta^\alpha}}_{\text{constant}}\,dx = \frac{\Gamma(\alpha+p)\beta^p}{\Gamma(\alpha)} \times 1$$

Method 2: Rewrite the function as a gamma function.

$$\mathbb{E}[X^p] = \int_0^\infty \frac{x^{(p+\alpha)-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}\,dx$$

Let $y = x/\beta \implies x = \beta y$ and $dx = \beta\,dy$. Therefore,

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

Additionally,

- $\mathbb{E}[X] = \dfrac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha\beta$
- $\mathbb{E}[X^2] = \dfrac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)} = \dfrac{\beta^2(\alpha+1)\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha(\alpha+1)\beta^2$
- $\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2 = \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$

2.5 Moment Generating Functions

DEFINITION 2.5.1: Moment generating function

Suppose X is a random variable, then

$$M(t) = \mathbb{E}[e^{tX}]$$

is called the **moment generating function** (m.g.f.) of X if $M(t)$ exists for $t \in (-h, h)$ with some $h > 0$.

REMARK 2.5.2

If we are able to find some $h > 0$ such that for any $t \in (-h, h)$, $\mathbb{E}[e^{tX}] < \infty$, then we say $M(t)$ is the m.g.f. of X.

EXAMPLE 2.5.3

Suppose $X \sim \text{Gamma}(\alpha, \beta)$. Find $M(t)$. Recall the p.d.f. is

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

Solution.

$$\begin{aligned} M(t) &= \mathbb{E}[e^{tX}] \\ &= \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \\ &= \int_{0}^{\infty} e^{tx} \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}\, dx \\ &= \int_{0}^{\infty} \frac{x^{\alpha-1}\text{exp}\left\{-\frac{x}{\frac{1}{(1/\beta)-t}}\right\}}{\Gamma(\alpha)\beta^{\alpha}}\, dx \\ &= \int_{0}^{\infty} \frac{x^{\alpha-1}e^{-x/\tilde{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}\, dx \end{aligned}$$

where

$$\tilde{\beta} = \frac{1}{\left(\dfrac{1}{\beta} - t\right)}$$

Continuing,

$$\begin{aligned} &= \int_{0}^{\infty} \frac{x^{\alpha-1}e^{-x/\tilde{\beta}}}{\Gamma(\alpha)\tilde{\beta}^{\alpha}} \left(\frac{\tilde{\beta}^{\alpha}}{\beta^{\alpha}}\right) dx \\ &= \frac{\tilde{\beta}^{\alpha}}{\beta^{\alpha}}(1) \\ &= (1-\beta t)^{-\alpha} \end{aligned}$$

The moment generating function must be non-negative since $1 - \beta t > 0$ and therefore, $t < 1/\beta$. Take $h = 1/\beta$.

EXAMPLE 2.5.4

If $X \sim \text{Poisson}(\theta)$, the p.f. is given by $f(x) = \dfrac{\theta^x e^{-\theta}}{x!}$ for $0 \leq x \in \mathbf{Z}$. Find $M(t)$.

Solution.

$$\begin{aligned}
M(t) &= \mathbb{E}[e^{tX}] \\
&= \sum_{x=0}^{\infty} e^{tx} \frac{\theta^x e^{-\theta}}{x!} \\
&= \sum_{x=0}^{\infty} \frac{(e^t\theta)^x e^{-\theta}}{x!} \\
&= e^{-\theta} \sum_{x=0}^{\infty} \frac{(e^t\theta)^x}{x!} \\
&= e^{-\theta} \exp\{e^t\theta\} \\
&= \exp\{\theta(e^t - 1)\}
\end{aligned}$$

for all $t \in \mathbf{R}$.

Three important properties of $M(t)$.

THEOREM 2.5.5: Moment Generating Function of a Linear Function

Suppose that the moment generating function of X is $M_X(t)$. Then $Y = aX + b$ has moment generating function

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

Proof of Theorem 2.5.5

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(aX+b)}] = e^{bt}\mathbb{E}[e^{atX}] = e^{bt} M_X(at)$$

EXAMPLE 2.5.6

(i) If $Z \sim \mathcal{N}(0,1)$, find $M_Z(t)$.
(ii) If $X \sim \mathcal{N}(\mu, \sigma^2)$, find $M_X(t)$.

Solution.

(i)

$$\begin{aligned}
M_Z(t) &= \mathbb{E}[e^{tZ}] \\
&= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2 - 2tx}{2}\right\} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-t)^2 - t^2}{2}\right\} dx && \text{complete the square} \\
&= \exp\left\{\frac{t^2}{2}\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-t)^2}{2}\right\} dx
\end{aligned}$$

where the integral is the p.d.f. of $\mathcal{N}(\mu = t, \sigma^2 = 1)$. Therefore,

$$\mathbb{E}[e^{tZ}] = \exp\left\{\frac{t^2}{2}\right\}$$

(ii) $X = \sigma Z + \mu$ where $Z \sim \mathcal{N}(0,1)$.

$$M_X(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} \exp\left\{\frac{(\sigma t)^2}{2}\right\} = \exp\left\{\frac{(\sigma t)^2}{2} + \mu t\right\}$$

THEOREM 2.5.7: Moments from Moment Generating Function

Suppose X has moment generating function $M(t)$.

$$M^{(k)}(0) = \mathbb{E}[X^k]$$

EXAMPLE 2.5.8

Gamma(α, β) has m.g.f. $M(t) = (1 - \beta t)^{-\alpha}$ for $t < 1/\beta$. What is $\mathbb{E}[X]$ and $\mathbb{V}(X)$?
Solution. For $\mathbb{E}[X]$ we find $M'(t)$.

$$M'(t) = (-\alpha)(1 - \beta t)^{-\alpha-1}(-\beta) = (\alpha\beta)(1 - \beta t)^{-\alpha-1}$$

We know,

$$\mathbb{E}[X] = M'(0) = \alpha\beta$$

For $\mathbb{V}(X)$ we find $M''(t)$.

$$M''(t) = (\alpha\beta)(-\alpha - 1)(-\beta)(1 - \beta t)^{-\alpha-2}$$

Now, $M''(0) = \alpha\beta^2(\alpha + 1) = \mathbb{E}[X^2]$. Therefore,

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2 = \alpha\beta^2(\alpha + 1) - (\alpha\beta)^2 = \alpha\beta^2$$

EXAMPLE 2.5.9

The m.g.f. of Poisson(θ) is $M(t) = \exp\{\theta(e^t - 1)\}$. Find $\mathbb{E}[X]$ and $\mathbb{V}(X)$.
Solution.

$$M'(t) = \exp\{\theta(e^t - 1)\}\theta e^t$$

Therefore,

$$\mathbb{E}[X] = M'(0) = \theta$$

Now,

$$M''(t) = \exp\{\theta(e^t - 1)\}\theta^2 e^{2t} + \theta e^t \exp\{\theta(e^t - 1)\}$$

Therefore,

$$M''(0) = \mathbb{E}[X^2] = \theta^2 + \theta$$

So,

$$\mathbb{V}(X) = \mathbb{E}[X^2] - \mu^2 = \theta^2 + \theta - (\theta)^2 = \theta$$

THEOREM 2.5.10: Uniqueness of Moment Generating Functions

X and Y have the same moment generating function if and only if X and Y have the same distribution.

EXAMPLE 2.5.11

Suppose X has m.g.f. $M_X(t) = \exp\left\{\frac{t^2}{2}\right\}$.

(i) Find m.g.f. of $Y = 2X - 1$
(ii) Find $\mathbb{E}[Y]$ and $\mathbb{V}(Y)$
(iii) What is the distribution of Y.

Solution.

(i) $M_Y(t) = e^{-t}\exp\left\{\frac{(2t)^2}{2}\right\} = \exp\{2t^2 - t\}$.
(ii)

$$M'_Y(t) = \exp\{2t^2 - t\}(4t - 1)$$

Therefore,

$$\mathbb{E}[Y] = M_Y'(0) = -1$$

Also,

$$M_Y''(t) = \text{exp}\{2t^2 - t\}(4t-1)^2 + 4\text{exp}\{2t^2 - t\}$$

and

$$\mathbb{E}[Y^2] = M_Y''(0) = 1 + 4 = 5$$

Therefore,

$$\mathbb{V}(Y) = \mathbb{E}[Y^2] - \mu^2 = 5 - 1 = 4$$

(iii) $M_Y(t) = \text{exp}\{2t^2 - t\}$ is the m.g.f. of $\mathcal{N}(-1, 4)$ since if $X \sim \mathcal{N}(\mu, \sigma^2)$, then (by previous example)

$$M_X(t) = e^{\mu t}\text{exp}\left\{\frac{\sigma^2 t^2}{2}\right\}$$

Lecture 5 | 2020-09-20

EXAMPLE 2.5.12: Uniqueness Theorem

Suppose $M_X(t) = (1-2t)^{-1}$. What is the distribution of X?
Solution. $X \sim \text{Gamma}(\alpha = 1, \beta = 2)$.

Chapter 3

Multivariate Random Variables

3.1 Joint and Marginal Cumulative Distribution Functions

Purpose: to characterize a joint distribution of two random variables.

DEFINITION 3.1.1: Joint cumulative distribution function

Suppose X and Y are two random variables. The **joint cumulative distribution function** of X and Y is given by

$$F(x,y) = \mathbb{P}(X \leq x, Y \leq y)$$

for $(x,y) \in \mathbf{R}^2$.

$\mathbb{P}(X \leq x, Y \leq y)$: "What is the probability these two events occur simultaneously"

REMARK 3.1.2

Since $\{X \leq x\}$ and $\{Y \leq y\}$ are both events, $F(x,y)$ is well-defined as we consider $\{X \leq x\} \cap \{Y \leq y\}$.

REMARK 3.1.3

If we have more than two random variables, say $X_1, X_2, \dots, X_n$ We can similarly define the cumulative distribution function as

$$F(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$

However, in this course we will only focus on two events X and Y.

DEFINITION 3.1.4: Joint cumulative distribution function

(I) F is non-decreasing in x for fixed y

(II) F is non-decreasing in y for fixed x

(III) $\lim\limits_{x \to -\infty} F(x,y) = 0$ and $\lim\limits_{y \to -\infty} F(x,y) = 0$

By looking at

$$\underbrace{\{X \leq x\}}_{\substack{\to 0 \\ \text{as } x \to -\infty}} \cap \underbrace{\{Y \leq y\}}_{\substack{\to 0 \\ \text{as } y \to -\infty}}$$

(IV)

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

DEFINITION 3.1.5: Marginal distribution function

The **marginal distribution function** of X is given by

$$F_1(x) = \lim_{y\to\infty} F(x,y) = \mathbb{P}(X \leq x)$$

for $x \in \mathbf{R}$.
The **marginal distribution function** if Y is given by

$$F_2(y) = \lim_{x\to\infty} F(x,y) = \mathbb{P}(Y \leq y)$$

for $y \in \mathbf{R}$.

REMARK 3.1.6

The definition of marginal distribution function tells us that we can know all information about marginal c.d.f. from the joint c.d.f. but the marginal c.d.f. cannot give full information about joint c.d.f.

3.2 Bivariate Discrete Distributions

DEFINITION 3.2.1: Joint discrete random variables

Suppose X and Y are both discrete random variables, then (X, Y) are **joint discrete random variables** X and Y.

DEFINITION 3.2.2: Joint probability function, Support

Suppose X and Y are discrete random variables. The **joint probability function** of X and Y is given by

$$f(x,y) = \mathbb{P}(X = x, Y = y)$$

for $(x,y) \in \mathbf{R}^2$.
The set $A = \{(x,y) : f(x,y) > 0\}$ is called the **joint support** of (X, Y).

DEFINITION 3.2.3: Properties — Joint Probability Function

(I) $f(x,y) \geq 0$ for $(x,y) \in \mathbf{R}^2$

(II) $\displaystyle\sum_{(x,y)\in A} f(x,y) = 1$

(III) For any set $R \subseteq \mathbf{R}^2$

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

EXAMPLE 3.2.4

Suppose we want to find $\mathbb{P}(X \leq Y)$. What is the corresponding set R?
Solution. $R = \{(x,y) : x \leq y\}$
Suppose we want to find $\mathbb{P}(X + Y \leq 1)$. What is the corresponding set R?
Solution. $R = \{(x,y) : x + y \leq 1\}$

DEFINITION 3.2.5: Marginal probability function

Suppose X and Y are discrete random variables with joint probability function $f(x, y)$.
The **marginal probability function** of X is given by

$$f_1(x) = \mathbb{P}(X = x) = \mathbb{P}(X = x, Y < \infty) = \sum_y f(x, y)$$

for $x \in \mathbf{R}$.
The **marginal probability function** of Y is given by

$$f_2(y) = \mathbb{P}(Y = y) = \mathbb{P}(X < \infty, Y = y) = \sum_x f(x, y)$$

for $y \in \mathbf{R}$.

EXAMPLE 3.2.6

Suppose that X and Y are discrete random variables with joint p.f. $f(x, y) = kq^2p^{x+y}$ where

- $0 \leq x \in \mathbf{Z}$
- $0 \leq y \in \mathbf{Z}$
- $0 < p < 1$
- $q = 1 - p$

(i) Determine k.
(ii) Find marginal p.f. of X and find marginal p.f. of Y.
(iii) Find $\mathbb{P}(X \leq Y)$.

Solution.

(i) $k > 0$ since if $k = 0$ then the summation of the joint p.f. will be 0 (but needs to be 1).

$$\sum_{x=0}^{\infty}\sum_{y=0}^{\infty} f(x, y) = 1$$

Therefore,

$$k\left(\sum_{x=0}^{\infty}\sum_{y=0}^{\infty} p^{x+y}q^2\right) = kq^2\left(\sum_{x=0}^{\infty} p^x\right)\left(\sum_{y=0}^{\infty} p^y\right) = kq^2\left(\frac{1}{1-p}\right)\left(\frac{1}{1-p}\right) = k$$

Thus, $k = 1$.

(ii) Marginal p.f. of X:

$$f_1(x) = \mathbb{P}(X = x) = \sum_{y=0}^{\infty} q^2p^{x+y} = q^2p^x\left(\sum_{y=0}^{\infty} p^y\right) = q^2p^x\left(\frac{1}{1-p}\right) = p^x(1-p)$$

Support of X: $[0, \infty)$.
By symmetry,

$$f_2(y) = \mathbb{P}(Y = y) = qp^y$$

Support of Y: $[0, \infty)$.

(iii) Find $\mathbb{P}(X \leq Y)$.

$$\begin{aligned}
\mathbb{P}(X \leq Y) &= \sum_{x=0}^{\infty} \sum_{y=x}^{\infty} \left(q^2 p^{x+y} \right) \\
&= \sum_{x=0}^{\infty} q^2 p^x \sum_{y=x}^{\infty} p^y \\
&= \sum_{x=0}^{\infty} q^2 p^x \left(\frac{p^x}{1-p} \right) \\
&= q \sum_{x=0}^{\infty} p^{2x} \\
&= q \left(\frac{1}{1-p^2} \right) \\
&= \frac{1}{1+p}
\end{aligned}$$

REMARK 3.2.7: Interesting Fact

If X and Y are *continuous* random variables and have the same distribution and ***independent***,

$$\mathbb{P}(X \leq Y) = \frac{1}{2}$$

3.3 Bivariate Continuous Distributions

DEFINITION 3.3.1: Joint probability density function, Support

If the joint c.d.f. of (X, Y) can be written as

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

for all $(x, y) \in \mathbf{R}^2$, then X and Y are joint continuous random variables with **joint probability density function** $f(x, y)$ where

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

The set $A = \{(x, y) : f(x, y) > 0\}$ is called the **support** of (X, Y).

REMARK 3.3.2

We will arbitrarily define $f(x, y)$ to be equal to 0 when $\dfrac{\partial^2}{\partial x \partial y}[F(x, y)]$ does not exist, although we can define it to be any real number.

DEFINITION 3.3.3: Properties — Joint Probability Density Function

(I) $f(x,y) \geq 0$ for all $(x,y) \in \mathbf{R}^2$
(II) For any set $R \subseteq \mathbf{R}^2$:

$$\mathbb{P}((X,Y) \in R) = \iint\limits_{(x,y)\in R} f(x,y)dx\,dy$$

EXAMPLE 3.3.4

To find $\mathbb{P}(X \leq Y)$, the region is $R = \{(x,y) : x \leq y\}$. Therefore,

$$\mathbb{P}(X \leq y) = \iint\limits_{x \leq y} f(x,y)dx\,dy$$

DEFINITION 3.3.5: Marginal probability density function

Suppose X and Y are continuous random variables with p.d.f. $f(x,y)$. The **marginal probability density function** of X is given by

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

for $x \in \mathbf{R}$ and the **marginal probability density function** of Y is given by

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

for $y \in \mathbf{R}$.

$$\mathbb{P}((X,Y) \in \mathbf{R}) = \iint\limits_{R} f(x,y)dx\,dy = \int_x \int_y f(x,y)\,dx\,dy$$

Helpful theorem from MATH 237 that some of you may have forgotten:

THEOREM 3.3.6: †

y first, then x
Let $R \subset \mathbf{R}^2$ be defined by

$$y_\ell(x) \leq y \leq y_u(x) \quad \text{and} \quad x_\ell \leq x \leq x_u$$

where $y_\ell(x)$ and $y_u(x)$ are continuous for $x_\ell \leq x \leq x_u$. If $f(x,y)$ is continuous on R, then

$$\iint\limits_{R} f(x,y)dA = \int_{x_\ell}^{x_u} \int_{y_\ell(x)}^{y_u(x)} f(x,y)\,dy\,dx$$

x first, then y
Let $R \subset \mathbf{R}^2$ be defined by

$$x_\ell(y) \leq x \leq x_u(y) \quad \text{and} \quad y_\ell \leq y \leq y_u$$

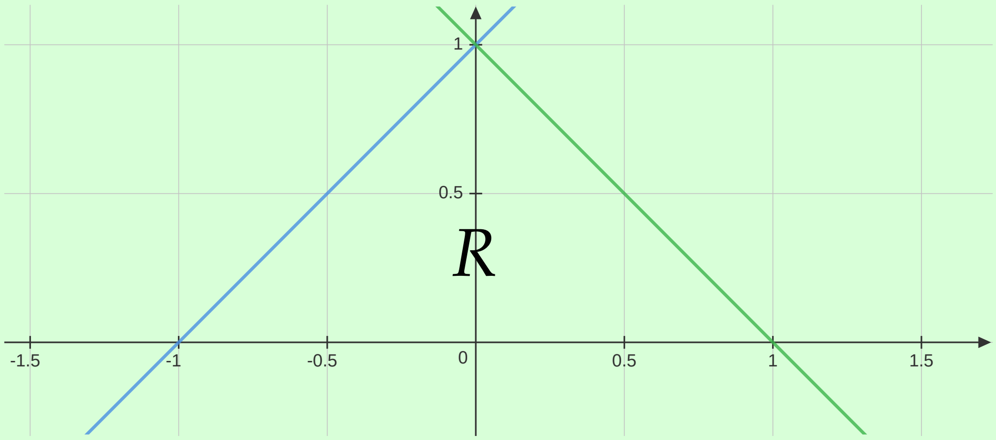
where $x_\ell(y)$ and $x_u(y)$ are continuous for $y_\ell \leq y \leq y_u$. If $f(x,y)$ is continuous on R, then

$$\iint\limits_{R} f(x,y)dA = \int_{y_\ell}^{y_u} \int_{x_\ell(y)}^{x_u(y)} f(x,y)\,dx\,dy$$

We use ℓ for "lower" and u for "upper."

EXAMPLE 3.3.7

Describe the region R above the x-axis.



Solution. R can be described by the set of two inequalities (you can actually verify this in Desmos if you *really* forgot how this works):

$$0 \leq y \leq 1$$

$$y-1 \leq x \leq 1-y$$

Using the theorem above,

$$\int_0^1 \int_{y-1}^{1-y} f(x,y)\,dx\,dy$$

LECTURE 6 | 2020-09-20

<u>Author's note</u>: Diagrams will be omitted for most of the text, unless the example is not trivial. Students are encouraged to draw the diagrams when following the examples.

EXAMPLE 3.3.8

Let X and Y be continuous random variables with joint p.d.f.

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

(i) Show $f(x,y)$ is a joint p.d.f.

(ii) Find

 (a) $\mathbb{P}(X \leq 1/3, Y \leq 1/2)$

 (b) $\mathbb{P}(X \leq Y)$

 (c) $\mathbb{P}(X + Y \leq 1/2)$

 (d) $\mathbb{P}(XY \leq 1/2)$

(iii) Find marginal p.d.f. of X and Y.

Solution.

(i) Note that $f(x,y) \geq 0$. We need to show

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

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

(ii) (a) $R = \{(x,y) : 0 \le x \le 1/3, 0 \le y \le 1/2\}$.

$$\int_0^{1/3} \int_0^{1/2} (x+y)\,dy\,dx = \frac{5}{72}$$

(b) $R = \{(x,y) : 0 \le x \le 1, x \le y \le 1\}$.

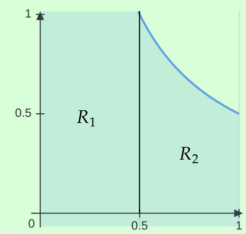
$$\int_0^1 \int_x^1 (x+y)\,dy\,dx = \frac{1}{2}$$

(c) $R = \{(x,y) : 0 \le x \le 1/2, 0 \le y \le (1/2) - x\}$

$$\int_0^{1/2} \int_0^{(1/2)-x} (x+y)\,dy\,dx = \frac{1}{24}$$

(d) $R_1 = \{(x,y) : 0 \le x \le 1/2, 0 \le y \le 1\}$ and $R_2 = \{(x,y) : 1/2 \le x \le 1, 0 \le y \le (1/2)/x\}$. Therefore, we need to evaluate two double integrals.

$$\int_0^{1/2} \int_0^1 (x+y)\,dy\,dx + \int_{1/2}^1 \int_0^{(1/2)/x} (x+y)\,dy\,dx = \frac{3}{4}$$



(iii) The support of X is $[0,1]$.

$$f_1(x) = 0 \iff x < 0 \text{ or } x > 1$$

Therefore, we focus on $0 \le x \le 1$.

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

Thus,

$$f_1(x) = \begin{cases} x + \frac{1}{2} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$f_2(y)$ is similar by symmetry.

EXAMPLE 3.3.9

Suppose

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

is the joint p.d.f. of (X,Y).

(i) Find k.

(ii) Find

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

(b) $\mathbb{P}(X \le Y)$

(c) $\mathbb{P}(X + Y \ge 1)$

(iii) Marginal p.d.f. of X and Y.
(iv) Suppose $T = X + Y$, find the p.d.f. of T.

Solution.

(i) We know $f(x, y) \geq 0 \iff k \geq 0$. Actually, $k > 0$ since if $k = 0$, then $f(x, y) \equiv 0$. We solve k by solving the following:

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

Therefore,

$$\int_{0}^{\infty} \int_{x}^{\infty} ke^{-x-y}\, dy\, dx = \frac{k}{2}$$

Thus, $k/2 = 1 \implies k = 2$.

(ii) (a) $\mathbb{P}(X \leq 1/3, Y \leq 1/2)$.

$$R = \{(x, y) : 0 \leq x \leq 1/3, x \leq y \leq 1/2\}$$

Therefore,

$$\begin{aligned}\mathbb{P}(X \leq 1/3, Y \leq 1/2) &= \int_{0}^{1/3} \int_{x}^{1/2} 2e^{-x-y}\, dy\, dx \\ &= 1 - e^{-2/3} + 2(e^{-5/6} - e^{-1/2}) \\ &\approx 0.1427\end{aligned}$$

(b) $\mathbb{P}(X \leq Y)$. Note that the region is the same as the support. Therefore,

$$\mathbb{P}(X \leq Y) = \iint_{x \leq y} f(x, y) dx\, dy = 1$$

(c) $\mathbb{P}(X + Y \geq 1)$. Note that this region is a bit complicated, so we will consider $1 - \mathbb{P}(X + Y < 1) = 1 - \mathbb{P}(X + Y \leq 1)$. The equal sign does not account for any area (it's continuous, but not required to know in this course).

$$R = \{(x, y) : 0 \leq x \leq 1/2, x \leq y \leq 1 - x\}$$

$$\begin{aligned}\mathbb{P}(X + Y \leq 1) &= \int_{0}^{1/2} \int_{x}^{1-x} 2e^{-x}e^{-y}\, dy\, dx \\ &= 1 - 2e^{-1}\end{aligned}$$

Thus, $\mathbb{P}(X + Y \geq 1) = 1 - \mathbb{P}(X + Y \leq 1) = 1 - (1 - 2e^{-1}) = 2e^{-1}$.

(iii) Marginal p.d.f. of X. The support of X is $(0, \infty)$. We know $x > 0$, so

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_{x}^{\infty} 2e^{-x-y}\, dy = 2e^{-x}\left[-e^{-y}\right]_{x}^{\infty} = 2e^{-2x}$$

The marginal p.d.f. of Y. The support of Y is $(0, \infty)$. We know $y > 0$, so

$$f_2(y) = \int_{0}^{y} 2e^{-x-y}\, dx = 2e^{-y}\left[-e^{-x}\right]_{0}^{y} = 2e^{-y}\left(1 - e^{-y}\right) = 2e^{-y} - 2e^{-2y}$$

(iv) Suppose $T = X + Y$, find the p.d.f. of T. We first find the c.d.f. of T, then we take the derivative of T.
Support of T is $(0, \infty)$.

When $t \leq 0$, $F_T(t) = \mathbb{P}(T \leq t) = 0$, so we only focus on $t > 0$, so $F_T(t) = \mathbb{P}(T \leq t)$.

$$R = \{(x, y) : 0 \leq x \leq t/2, x \leq y \leq t - x\}$$

Therefore,

$$\begin{aligned} F_T(t) &= \int_0^{t/2} \int_x^{t-x} 2e^{-x-y}\, dy\, dx \\ &= 1 - e^{-t} - te^{-t} \end{aligned}$$

So,

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

Therefore, by computing $\dfrac{d}{dt}[F_T(t)]$, the p.d.f. of T is

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

LECTURE 7 | 2020-09-27

3.4 Independence

DEFINITION 3.4.1: Independent

For any two random variables, we say X and Y are **independent** if and only if

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

for any two sets A and B of real numbers.

THEOREM 3.4.2: Independent Random Variables

Suppose X and Y are random variables. X and Y are independent if and only if

(1) $F(x, y) = F_1(x)F_2(y)$, or

(2) $f(x, y) = f_1(x)f_2(y)$.

THEOREM 3.4.3

Let g and h be real-valued functions. If X and Y are independent, then $g(X)$ and $h(Y)$ are independent.

EXAMPLE 3.4.4

If X and Y are independent, then X^2 and Y^2 are independent. However, if X^2 and Y^2 are independent, then X and Y may not be independent. Can you find an example here? Choose X where

$$\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$$

EXAMPLE 3.4.5

Consider the joint discrete random variable $f(x,y) = q^2 p^{x+y}$, where $0 \leq x \in \mathbf{Z}$ and $0 \leq y \in \mathbf{Z}$. Then $f_1(x) = qp^x$ and $f_2(y) = qp^y$. Therefore, $f(x,y) = f_1(x)f_2(y)$ shows that X and Y are independent.

Consider $f(x,y) = \begin{cases} x+y & 0 \leq x \leq 1, \quad 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$ We've shown that

$$f_1(x) = \begin{cases} x + \dfrac{1}{2} & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_2(y) = \begin{cases} y + \dfrac{1}{2} & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We see that $f(x,y) \neq f_1(x)f_2(y)$ therefore, X and Y are not independent.

THEOREM 3.4.6: Factorization Theorem for Independence

Suppose X and Y are random variables with joint probability (density) function $f(x,y)$. Suppose also that A is the support set of (X,Y), A_1 is the support set of X, and A_2 is the support set of Y. Then X and Y are independent random variables if and only if there exist non-negative functions $g(x)$ and $h(y)$ such that

$$f(x,y) = g(x)h(y) \quad \forall (x,y) \in A_1 \times A_2$$

where $A_1 \times A_2 = \{(x,y) : x \in A_1, y \in A_2\}$.

REMARK 3.4.7

Equivalently, we can check that both conditions are met:

- The support of A is a square or rectangle.
- The range of X does not depend on the values of y and the range of Y does not depend on the values of x.

EXAMPLE 3.4.8

$f(x,y) = \dfrac{\theta^{x+y}e^{-2\theta}}{x!y!}$ where $0 \leq x, y \in \mathbf{Z}$. Are X and Y independent or not? Find the marginal p.f. of X and Y.

Solution.

$$f(x,y) = \underbrace{\frac{\theta^x}{x!}e^{-\theta}}_{g(x)} \underbrace{\frac{\theta^y}{y!}e^{-\theta}}_{h(y)}$$

The range of X does not depend on the value of y. Therefore, X and Y are independent.

$$f_1(x) = \sum_{y=0}^{\infty} f(x,y) = \frac{\theta^x e^{-\theta}}{x!} \quad 0 \leq x \in \mathbf{Z}$$

$$f_2(y) = \sum_{x=0}^{\infty} f(x,y) = \frac{\theta^y e^{-\theta}}{y!} \quad 0 \leq y \in \mathbf{Z}$$

If we've shown that X and Y are independent, then we can verify

$$f(x,y) = g(x)h(y)$$

With $f_1(x) = C_1 g(x)$ and $f_2(y) = C_2 h(y)$ where $C_1, C_2 \in \mathbf{R}$ is a constant. We know that $C_1 C_2 = 1$.

EXAMPLE 3.4.9

If X and Y have joint p.d.f. $f(x,y) = \frac{3}{2}y(1-x^2)$ where $-1 \le x \le 1$ and $0 \le y \le 1$. Are X and Y independent? Find $f_1(x)$ and $f_2(y)$.
Solution. $f(x,y) = \underbrace{(1-x^2)}_{h(x)}\underbrace{\tfrac{3}{2}y}_{g(y)}$ and $A = \{(x,y) : -1 \le x \le 1, 0 \le y \le 1\}$ is a rectangle. Therefore, X and Y are independent. So,

$$f_1(x) = C_1 h(x) = C_1(1-x^2) \quad \text{for } -1 \le x \le 1$$

So, let's consider the integral:

$$\int_{-1}^{1} f_1(x)\,dx = C_1 \int_{-1}^{1} (1-x^2)\,dx = 1 \implies C_1 = \frac{3}{4}$$

Using our previous result, we know that

$$f_2(y) = \frac{1}{C_1}h(y) = \frac{4}{3}\cdot\frac{3}{2}y = 2y \quad 0 \le y \le 1$$

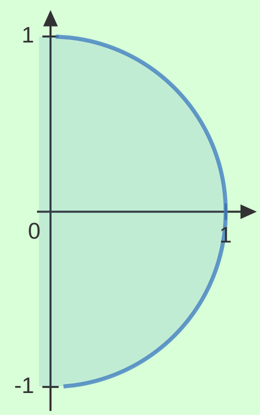
EXAMPLE 3.4.10: Uniform Distribution on a Semicircle

$f(x,y) = \frac{2}{\pi}$ where $0 \le x \le \sqrt{1-y^2}$ and $-1 \le y \le 1$. The area of the semicircle is given by $\pi/2$. Are X and Y independent? Find $f_1(x)$ and $f_2(y)$.
Solution. $f(x,y) = 2/\pi$. Take $g(x) = 1$ and $h(y) = 2/\pi$. Also, this is not a rectangle, so X and Y are not independent. Similarly, for a particular value of x we can easily see that y depends on x.
The support of X is $[0,1]$.

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

The support of Y is $[-1,1]$.

$$f_2(y) = \int_{0}^{\sqrt{1-y^2}} \frac{2}{\pi}\,dx = \frac{2}{\pi}\sqrt{1-y^2}$$



Neither of these marginal distributions are uniform.

Lecture 8 | 2020-09-27

3.5 Joint Expectation

This section: extend the definition of expectation from univariate to bivariate cases.

DEFINITION 3.5.1: Joint exepectation

Suppose $h(x, y)$ is a real-valued function.
If X and Y are discrete random variables with joint probability function $f(x, y)$ and support set A then

$$\mathbb{E}[h(X,Y)] = \sum_{(x,y)\in A} h(x,y)f(x,y)$$

provided the joint sum converges absolutely.
If X and Y are continuous random variables with joint probability density function $f(x, y)$ then

$$\mathbb{E}[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)f(x,y)\,dx\,dy$$

provided the joint integral converges absolutely.

EXAMPLE 3.5.2

$$\mathbb{E}[XY] = \begin{cases} \displaystyle\sum_x\sum_y xyf(x,y) & X, Y \text{ are joint discrete} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dx\,dy & X, Y \text{ are joint continuous} \end{cases}$$

$$\mathbb{E}[X] = \begin{cases} \displaystyle\sum_x\sum_y xf(x,y) & X, Y \text{ are joint discrete} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dx\,dy & X, Y \text{ are joint continuous} \end{cases}$$

Alternatively,

$$\mathbb{E}[X] = \sum_x xf_1(x) = \sum_x x\left[\sum_y f(x,y)\right]$$

PROPOSITION 3.5.3: Linearity Property

Suppose X and Y are random variables with joint probability (density) function $f(x, y)$, a and b are constants, and $g(x, y)$ and $h(x, y)$ are real-valued functions. Then

$$\mathbb{E}[ag(X,Y) + bh(X,Y)] = a\mathbb{E}[g(X,Y)] + b\mathbb{E}[h(X,Y)]$$

COROLLARY 3.5.4

If $X_1, \ldots, X_n$ are random variables and $a_1, \ldots, a_n$ are real constants then

$$\mathbb{E}\left[\sum_{i=1}^n a_iX_i\right] = \sum_{i=1}^n a_i\mathbb{E}[X_i]$$

THEOREM 3.5.5: Expectation and Independence

(1) If X and Y are independent random variables and $g(x)$ and $h(y)$ are real-valued functions then

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$$

(2) More generally, if $X_1, \dots, X_n$ are independent random variables and h is a real-valued function then

$$\mathbb{E}\left[\prod_{i=1}^{n} h(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}[h(X_i)]$$

DEFINITION 3.5.6: Covariance

The **covariance** of random variables X and Y is defined by

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

where $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$.

THEOREM 3.5.7: Covariance and Independence

If X and Y are random variables then

$$\textit{Cov}(X, Y) = \mathbb{E}[XY] - \mu_X\mu_Y$$

If X and Y are independent then Cov$(X, Y) = 0$.

Proof of Theorem 3.5.7

$$\begin{aligned}
\text{Cov}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\
&= \mathbb{E}[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y] \\
&= \mathbb{E}[XY] - \mu_X\mathbb{E}[Y] - \mu_Y\mathbb{E}[X] + \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

Now, if X and Y are independent, then by Theorem 3.5.5, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 0$. Thus, Cov$(X, Y) = 0$.

THEOREM 3.5.8: Results for Covariance

(1) Cov$(X, X) = \mathbb{E}[(X - \mu_X)(X - \mu_X)] = \mathbb{E}[(X - \mu_X)^2] = \mathbb{V}(X)$
(2) Cov$(X + Y, Z) =$ Cov$(X, Z) +$ Cov(Y, Z)

THEOREM 3.5.9: Variance of a Linear Combination

(1) Suppose X and Y are random variables and a and b are real constants then

$$\mathbb{V}(aX+bY) = a^2\mathbb{V}(X) = b^2\mathbb{V}(Y) + 2ab\text{Cov}(X,Y)$$

(2) Suppose $X_1,\ldots,X_n$ are random variables and $a_1,\ldots,a_n$ are real constants then

$$\mathbb{V}\left(\sum_{i=1}^{n} a_iX_i\right) = \sum_{i=1}^{n} a_i^2\mathbb{V}(X_i) + \underbrace{\sum_{i\neq j} a_ia_j\text{Cov}(X_i,X_j)}_{\binom{n}{2}\text{ terms}} = \sum_{i=1}^{n} a_i^2\mathbb{V}(X_i) + 2\underbrace{\sum_{i<j} a_ia_j\text{Cov}(X_i,X_j)}_{\binom{n}{2}\text{ terms}}$$

(3) If $X_1,\ldots,X_n$ are random variables and $a_1,\ldots,a_n$ are real constants then

$$\mathbb{V}\left(\sum_{i=1}^{n} a_iX_i\right) = \sum_{i=1}^{n} a_i^2\mathbb{V}(X_i)$$

EXAMPLE 3.5.10

Suppose the joint p.f. of X and Y is $f(x,y) = \dfrac{\theta^{x+y}e^{-2\theta}}{x!y!}$, where $0 \leq x,y \in \mathbf{Z}$. Find $\mathbb{V}(2X+3Y)$.

Solution.

$$f(x,y) = \underbrace{\left(\frac{\theta^x e^{-\theta}}{x!}\right)}_{g(x)}\underbrace{\left(\frac{\theta^y e^{-\theta}}{y!}\right)}_{h(y)}$$

Thus, the range of X does not depend on Y. Therefore, X and Y are independent. In other words, we can write

$$f_1(x) = C\frac{\theta^x e^{-\theta}}{x!} \quad 0 \leq x \in \mathbf{Z}$$

Since $\displaystyle\sum_{x=0}^{\infty}\frac{\theta^x e^{-\theta}}{x!} = 1$ as it is Poisson we get that $C=1$. Also,

$$f_2(y) = \frac{\theta^y e^{-\theta}}{y!} \quad 0 \leq y \in \mathbf{Z}$$

Thus, $\mathbb{V}(X) = \theta$ and $\mathbb{V}(Y) = \theta$. Finally,

$$\mathbb{V}(2X+3Y) = 4\mathbb{V}(X) + 9\mathbb{V}(Y) = 13\theta$$

EXAMPLE 3.5.11

The joint p.d.f. of X and Y is $f(x,y) = \begin{cases} x+y & x,y\in[0,1] \\ 0 & \text{otherwise}\end{cases}$.

Find $\mathbb{V}(X+Y)$.

Solution. We know $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{Cov}(X,Y)$. Recall that

$$f_1(x) = \begin{cases} x+\frac{1}{2} & x\in[0,1] \\ 0 & \text{otherwise}\end{cases} \qquad f_2(y) = \begin{cases} y+\frac{1}{2} & y\in[0,1] \\ 0 & \text{otherwise}\end{cases}$$

$$\mathbb{E}[X] = \int_0^1 x\left(x+\frac{1}{2}\right)dx = \frac{7}{12}$$

$$\mathbb{E}[X^2] = \int_0^1 x^2\left(x+\frac{1}{2}\right)dx = \frac{5}{12}$$

$$\implies \mathbb{V}(X) = \mathbb{E}[X^2] - \mu_X^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$$

We know that $\mathbb{E}[Y] = 7/12$, $\mathbb{V}(Y) = 11/144$. Now,

$$\mathbb{E}[XY] = \int_0^1\int_0^1 xy(x+y)\,dy\,dx = \frac{1}{3}$$

$$\implies \text{Cov}(X,Y) = \mathbb{E}[XY] - \mu_X\mu_Y = \frac{1}{3} - \left(\frac{7}{12}\right)\left(\frac{7}{12}\right) = -\frac{1}{144}$$

Hence,

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{Cov}(X,Y) = \frac{11}{144} + \frac{11}{144} - \frac{2}{144} = \frac{20}{144} = \frac{5}{36}$$

DEFINITION 3.5.12: Correlation coefficient

The **correlation coefficient** of random variables X and Y is defined by

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

REMARK 3.5.13

$\rho(X,Y)$ can only be used to characterize **linear** association between X and Y. For example, there might exist some quadratic relationship between X and Y but $\rho(X,Y) \to 0$.

EXAMPLE 3.5.14

$Y = X^2$ and $X \sim \mathcal{N}(0,1)$. Note that $\rho(X,Y) = 0$, but obviously there is some relationship between X and Y.

THEOREM 3.5.15

If $\rho(X,Y)$ is the correlation coefficient of random variables X and Y, then $-1 \le \rho(X,Y) \le 1$

(1) $\rho(X,Y) = 1 \iff Y = aX + b$ *with* $a > 0$.

(2) $\rho(X,Y) = -1 \iff Y = aX + b$ *with* $a < 0$.

EXAMPLE 3.5.16

Let $f(x,y) = \begin{cases} x+y & x,y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$. Find $\rho(X,Y)$.

Solution. Recall that $\mathbb{V}(X) = \mathbb{V}(Y) = 11/144$ and $\text{Cov}(X,Y) = -1/144$. So,

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

3.6 Conditional Distributions

DEFINITION 3.6.1: Conditional probability (density) function

Suppose that X and Y have joint probability (density) function $f(x, y)$, and marginal probability (density) functions $f_1(x)$ and $f_2(y)$ respectively. Suppose also that the support set of (X, Y) is $A = \{(x, y) : f(x, y) > 0\}$.
The **conditional probability (density) function** of X given $Y = y$ is

$$f_1(x \mid y) = \frac{f(x, y)}{f_2(y)} \quad \text{provided } f_2(y) > 0 \quad (x, y) \in A$$

The **conditional probability (density) function** of Y given $X = x$

$$f_2(y \mid x) = \frac{f(x, y)}{f_1(x)} \quad \text{provided } f_1(x) > 0 \quad (x, y) \in A$$

PROPOSITION 3.6.2: Properties — Conditional Probability Function

$f_1(x \mid y)$ and $f_2(y \mid x)$ are both probability functions; that is,

$$f_1(x \mid y) \geq 0 \quad \textit{and} \quad \sum_x f_1(x \mid y) = 1 \implies f_1(x \mid y) \textit{ is a p.f.}$$

$$f_2(y \mid x) \geq 0 \quad \textit{and} \quad \sum_y f_2(y \mid x) = 1 \implies f_2(y \mid x) \textit{ is a p.f.}$$

PROPOSITION 3.6.3: Properties — Conditional Probability Function

$f_1(x \mid y)$ and $f_2(y \mid x)$ are both probability density functions; that is,

$$f_1(x \mid y) \geq 0 \quad \textit{and} \quad \int_{-\infty}^{\infty} f_1(x \mid y)\, dx = 1 \implies f_1(x \mid y) \textit{ is a p.d.f.}$$

$$f_2(y \mid x) \geq 0 \quad \textit{and} \quad \int_{-\infty}^{\infty} f_2(y \mid x)\, dy = 1 \implies f_2(y \mid x) \textit{ is a p.d.f.}$$

EXAMPLE 3.6.4

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

Find

(i) $f_1(x \mid y)$

(ii) $f_2(y \mid x)$

Solution.

(i) To find $f_1(x \mid y)$, we need to calculate $f_2(y)$.

$$f_2(y) = \int_y^1 8xy\, dx = -4y^3 + 4y \quad 0 < y < 1$$

By definition,

$$f_1(x \mid y) = \frac{f(x, y)}{f_2(y)} = \frac{8xy}{4y - 4y^3} = \frac{2x}{1 - y^2} \quad 0 < y < 1$$

Given $0 < y < 1$, the support of X is $y < x < 1$.

(ii) To find $f_2(y \mid x)$, we need to calculate $f_1(x)$.

$$f_1(x) = \int_0^x 8xy\,dy = 4x^3 \quad 0 < x < 1$$

By definition,

$$f_2(y \mid x) = \frac{f(x,y)}{f_1(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2} \quad 0 < x < 1$$

Given $0 < x < 1$, the support of Y is $0 < y < x$.

EXAMPLE 3.6.5

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

Recall that $f_1(x) = x + 1/2$ for $0 \le x \le 1$ and $f_2(y) = y + 1/2$ for $0 \le y \le 1$. Therefore,

$$f_1(x \mid y) = \frac{f(x,y)}{f_2(y)} = \frac{x+y}{y+1/2}$$

Given $0 \le y \le 1$, the support of X is $0 \le x \le 1$.

$$f_2(y \mid x) = \frac{x+y}{x+1/2}$$

Given $0 \le x \le 1$, the support of Y is $0 \le y \le 1$.

EXAMPLE 3.6.6

$f(x,y) = q^2 p^{x+y}$ where $0 \le x, y \in \mathbf{Z}$. Note we derived that $f_1(x) = qp^x$ and $f_2(y) = qp^y$. Therefore,

$$f_1(x \mid y) = \frac{f(x,y)}{f_2(y)} = qp^x = f_1(x)$$

$$f_2(y \mid x) = \frac{f(x,y)}{f_1(x)} = qp^y = f_2(y)$$

This is another way to show independence of X and Y.

THEOREM 3.6.7

X and Y are independent if and only if
(1) $f_1(x \mid y) = f_1(x)$, and
(2) $f_2(y \mid x) = f_2(y)$.

THEOREM 3.6.8: Product Rule

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

EXAMPLE 3.6.9: Product rule

Suppose $Y \sim \text{Poisson}(\theta)$ and $X \mid Y = y \sim \text{Binomial}(y, p)$. Find the marginal p.f. of X.
Before we get to the solution of this problem, let's consider a physical setup.

- Y: number of students who go to Tim Hortons in one day. Note that $Y \sim \text{Poisson}(\theta)$.
- $X \mid Y = y$: number of students among these y visitors

What is the distribution of X? We guess that $X \sim \text{Poisson}(\theta p)$.
Solution.

$$f_1(x \mid y) = \binom{y}{x} p^x (1-p)^{y-x} \quad x = 0, 1, \dots, y$$

$$f_2(y) = \frac{\theta^y}{y!} e^{-\theta} \quad 0 \le y \in \mathbf{Z}$$

$$\begin{aligned}
f(x, y) &= f_1(x \mid y) f_2(y) \\
&= \left(\frac{y!}{x!(y-x)!} p^x (1-p)^{y-x} \right) \frac{\theta^y}{y!} e^{-\theta} \\
&= \left(\frac{\theta^x p^x}{x!} \right) \frac{\theta^{y-x}(1-p)^{y-x}}{(y-x)!} e^{-\theta}
\end{aligned}$$

(X, Y) support is $x = 0, 1, \dots, y$ and $0 \le y \in \mathbf{Z}$. Therefore,

$$\begin{aligned}
f_1(x) &= \sum_y f(x, y) \\
&= \sum_{y=x}^{\infty} \left(\frac{(\theta p)^x}{x!} \right) \left(\frac{(\theta(1-p))^{y-x}}{(y-x)!} e^{-\theta} \right) \\
&= \frac{e^{-\theta}(\theta p)^x}{x!} \sum_{h=0}^{\infty} \frac{\left[\theta(1-p)\right]^h}{h!} && h = y - x \\
&= \frac{e^{-\theta}(\theta p)^x}{x!} e^{\theta(1-p)} \\
&= \frac{(\theta p)^x}{x!} e^{-\theta p}
\end{aligned}$$

Therefore, $0 \le x \in \mathbf{Z}$ and so $X \sim \text{Poisson}(\theta p)$.

EXAMPLE 3.6.10

Suppose Y has p.d.f. $f_2(y) = \dfrac{y^{\alpha-1}}{\Gamma(\alpha)} e^{-y}$ for $y > 0$; that is, $Y \sim \text{Gamma}(\alpha, \beta = 1)$. The conditional p.d.f. of X given $Y = y$ is

$$f_1(x \mid y) = y e^{-xy} \quad \text{for } x > 0, y > 0$$

Find the marginal p.d.f. of X.
Solution. Firstly, find the joint p.d.f. of (X, Y) is

$$f(x, y) = f_1(x \mid y) f_2(y) = y e^{-xy} \frac{y^{\alpha-1}}{\Gamma(\alpha)} e^{-y} = \frac{y^\alpha}{\Gamma(\alpha)} e^{-(x+1)y}$$

The support of X is $(0, \infty)$. Recall that the gamma function is $\Gamma(\alpha) = \displaystyle\int_0^\infty x^{\alpha-1} e^{-x}\, dx$.
The marginal p.d.f. of X is

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

Let $t = (x+1)y$, therefore $y = t/(x+1)$ and $dy = dt/(x+1)$.

$$\int_0^\infty \frac{t^\alpha}{(x+1)^\alpha \Gamma(\alpha)} e^{-t} \frac{1}{x+1}\, dt = \frac{1}{(x+1)^{\alpha+1}\Gamma(\alpha)} \int_0^\infty t^\alpha e^{-t}\, dt = \frac{1}{(x+1)^{\alpha+1}\Gamma(\alpha)} \Gamma(\alpha+1)$$

By Proposition 2.3.7, we know that $\Gamma(\alpha+1) = (\alpha)\Gamma(\alpha)$. Therefore,

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

That is, $f_1(x) = \dfrac{\alpha}{(x+1)^{\alpha+1}}$ and the support of X is positive.

LECTURE 10 | 2020-10-04

3.7 Conditional Expectation

DEFINITION 3.7.1: Conditional expectation

The **conditional expectation** of $g(Y)$ given $X = x$ is defined as

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

REMARK 3.7.2

- Supplementary notes: $\mathbb{E}[g(Y) \mid X = x]$ is denoted by $\mathbb{E}[g(Y) \mid x]$.

We're interested in

1. The conditional mean of Y given $X = x$ is denoted $\mathbb{E}[Y \mid X = x]$ since $g(Y) = Y$.
2. The conditional variance of Y given $X = x$ is denoted by $\mathbb{V}(Y \mid X = x)$ and is given by

$$\mathbb{V}(Y \mid X = x) = \mathbb{E}[Y^2 \mid X = x] - \left(\mathbb{E}[Y \mid X = x]\right)^2$$

3. $\mathbb{E}[e^{tY} \mid X = x]$, that is, $g(Y) = e^{tY}$.

THEOREM 3.7.3: Independence

If X and Y are independent random variables then

$$\mathbb{E}[g(Y) \mid X = x] = \mathbb{E}[g(Y)] \quad \textit{and} \quad \mathbb{E}[h(X) \mid Y = y] = \mathbb{E}[h(X)]$$

In other words, the conditional expression becomes an unconditional one.

EXAMPLE 3.7.4

If X and Y are independent, then

$$\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y] \quad \text{and} \quad \mathbb{V}(Y \mid X = x) = \mathbb{V}(Y)$$

Also, $\mathbb{V}(Y \mid X = x) = \mathbb{E}[Y^2 \mid X = x] - \left(\mathbb{E}[Y \mid X = x]\right)^2 = \mathbb{E}[Y^2] - \left(\mathbb{E}[Y]\right)^2$

THEOREM 3.7.5: Substitution Rule

If X and Y be random variables and $h : \mathbf{R}^2 \to \mathbf{R}$ then

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

EXAMPLE 3.7.6

- $\mathbb{E}[X + Y \mid X = x] = \mathbb{E}[x + Y \mid X = x] = x + \mathbb{E}[Y \mid X = x]$
- $\mathbb{E}[XY \mid X = x] = \mathbb{E}[xY \mid X = x] = x\mathbb{E}[Y \mid X = x]$

THEOREM 3.7.7

The conditional expectation has all properties of expectation like linearity.

EXAMPLE 3.7.8

$f(x, y) = \begin{cases} 8xy & 0 < y < x < 1 \\ 0 & \text{otherwise} \end{cases}$ We've found that $f_1(x \mid y) = (2x)/(1 - y^2)$ for $0 < y < 1$ and $y < x < 1$.

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f_1(x \mid y)\, dx = \int_y^1 (x)\frac{2x}{1 - y^2}\, dx = \left(\frac{2}{3}\right)\frac{1 - y^3}{1 - y^2} = \left(\frac{2}{3}\right)\frac{y^2 + y + 1}{y + 1}$$

$$\mathbb{E}[X^2 \mid Y = y] = \int_y^1 (x^2)\frac{2x}{1 - y^2}\, dy = \left(\frac{2}{4}\right)\frac{1 - y^4}{1 - y^2} = \left(\frac{1}{2}\right)(y^2 + 1) \quad 0 < y < 1$$

$$\mathbb{V}(X \mid Y = y) = \left(\frac{1}{2}\right)(1 + y^2) - \left(\frac{4}{9}\right)\frac{(1 + y + y^2)^2}{(1 + y)^2} \quad 0 < y < 1$$

EXAMPLE 3.7.9

Suppose $Y \sim \text{Poisson}(\theta)$ and $X \mid Y = y \sim \text{Binomial}(y, p)$. Then,

$$\mathbb{E}[X \mid Y = y] = yp \quad \text{and} \quad \mathbb{V}(X \mid Y = y) = yp(1 - p)$$

REMARK 3.7.10

Note that $\mathbb{E}[g(Y) \mid X] \neq \mathbb{E}[g(Y) \mid X = x]$.
$\mathbb{E}[g(Y) \mid X]$ is a random variable because it's a function of X, denoted by $h(X)$. Its value is given by $h(x) = \mathbb{E}[g(Y) \mid X = x]$ for $X = x$.
How to get it? Two steps.

- Step 1: Find $\mathbb{E}[g(Y) \mid X = x] = h(x)$
- Step 2: Replace x with X to get the random variable $\mathbb{E}[g(Y) \mid X] = h(X)$.

EXAMPLE 3.7.11

Suppose $Y \sim \text{Poisson}(\theta)$ and $X \mid Y = y \sim \text{Binomial}(y, p)$. Then,

$$\mathbb{E}[X \mid Y = y] = yp \implies \mathbb{E}[X \mid Y] = Yp$$

These concepts lead to the Double Expectation Theorem or more commonly known as the Law of Total Expectation.

THEOREM 3.7.12: Double Expectation (Law of Total Expectation)

Suppose X and Y are random variables then

$$\mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}[g(Y) \mid X]]$$

In particular, $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$.

EXAMPLE 3.7.13

Suppose $Y \sim \text{Poisson}(\theta)$ and $X \mid Y = y \sim \text{Binomial}(y, p)$. Find $\mathbb{E}[X]$.
Solution. By Theorem 3.7.12 we have

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[Yp] = p\mathbb{E}[Y] = p\theta$$

Recall that we've shown that $X \sim \text{Poisson}(p\theta) \implies \mathbb{E}[X] = p\theta$.

LECTURE 11 | 2020-10-18

THEOREM 3.7.14: Law of Total Variance

Suppose X and Y are random variables then

$$\mathbb{V}(Y) = \mathbb{E}[\mathbb{V}(Y \mid X)] + \mathbb{V}(\mathbb{E}[Y \mid X])$$

REMARK 3.7.15

$\mathbb{V}(Y \mid X)$ is a random variable and function of X.
How to get it? Two steps:

1. $\mathbb{V}(Y \mid X = x) = \mathbb{E}[Y^2 \mid X = x] - (\mathbb{E}[Y \mid X = x])^2$.
2. Replace x with X to get the random variable $\mathbb{V}(Y \mid X)$.

EXAMPLE 3.7.16

$Y \sim \text{Poisson}(\theta)$, $X \mid Y = y \sim \text{Binomial}(y, p)$. Find $\mathbb{V}(X)$.
Solution. We know that $X \sim \text{Poisson}(p\theta)$, then $\mathbb{V}(X) = p\theta$. But we can alternatively use the Double Expectation Theorem.

$$\mathbb{V}(X) = \mathbb{E}[\mathbb{V}(X \mid Y)] + \mathbb{V}(\mathbb{E}[X \mid Y])$$

To find $\mathbb{V}(X \mid Y)$,

$$\mathbb{V}(X \mid Y = y) = yp(1-p) \implies \mathbb{V}(X \mid Y) = Yp(1-p)$$

To find $\mathbb{E}[X \mid Y]$,

$$\mathbb{E}[X \mid Y = y] = yp \implies \mathbb{E}[X \mid Y] = Yp$$

Therefore,

$$\mathbb{V}(X) = \mathbb{E}[Yp(1-p)] + \mathbb{V}(pY) = p(1-p)\mathbb{E}[Y] + p^2\mathbb{V}(Y) = p(1-p)\theta + p^2\theta = p\theta$$

EXAMPLE 3.7.17

Suppose $X \sim \text{Uniform}(0, 1)$ and $Y \mid X = x \sim \text{Binomial}(10, x)$. Find $\mathbb{E}[Y]$ and $\mathbb{V}(Y)$.

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$$

Two steps to find $\mathbb{E}[Y \mid X]$.

$$\mathbb{E}[Y \mid X = x] = 10x \implies \mathbb{E}[Y \mid X] = 10X$$

$$\mathbb{E}[Y] = \mathbb{E}[10X] = 10\mathbb{E}[X] = 10\left(\frac{1+0}{2}\right) = 5$$

$$\mathbb{V}(Y) = \mathbb{E}[\mathbb{V}(Y \mid X)] = \mathbb{V}(\mathbb{E}[Y \mid X])$$

Two steps to find $\mathbb{V}(Y \mid X)$.

$$\mathbb{V}(Y \mid X = x) = 10x(1-p) \implies \mathbb{V}(Y \mid X) = 10X(1-X)$$

$$\begin{aligned}
\mathbb{V}(Y) &= \mathbb{E}[10X(1-X)] + \mathbb{V}(10X) \\
&= 10\mathbb{E}[X] - 10\mathbb{E}[X^2] + 100\mathbb{V}(X) \\
&= 10\left(\frac{1+0}{2}\right) - 10\big[\mathbb{V}(X) + (\mathbb{E}[X])^2\big] + 100\mathbb{V}(X) \\
&= 5 - 10\left[\frac{(0-1)^2}{12} + \left(\frac{1+0}{2}\right)^2\right] + 100\left[\frac{(0-1)^2}{12}\right] \\
&= 5 - 10\left(\frac{1}{12} + \frac{1}{4}\right) + 100\left(\frac{1}{12}\right) \\
&= 5 - 10\left(\frac{1}{3}\right) + \frac{100}{12} \\
&= 10
\end{aligned}$$

EXAMPLE 3.7.18

Suppose $Y \sim \text{Poisson}(\theta)$ and $X \mid Y = y \sim \text{Binomial}(y, p)$. Find the m.g.f. of X using the Double Expectation Theorem. [We *could* use the formula sheet to find $M_X(t)$ since we already know $X \sim \text{Poisson}(p\theta)$]

Solution. By definition, the m.g.f. of X is

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}[\mathbb{E}[e^{tX} \mid Y]]$$

Given $Y = y$,

$$\begin{aligned}
\mathbb{E}[e^{tX} \mid Y = y] &= \sum_{x=0}^{y} e^{tx}\binom{y}{x}p^x(1-p)^{y-x} \\
&= \sum_{x=0}^{y}\binom{y}{x}(pe^t)^x(1-p)^{y-x} \\
&= (1-p+pe^t)^y
\end{aligned}$$

Therefore, $\mathbb{E}[e^{tX} \mid Y] = (1-p+pe^t)^Y$. Therefore,

$$\begin{aligned}
M_X(t) &= \mathbb{E}\big[(1-p+pe^t)^Y\big] \\
&= \sum_{y=0}^{\infty}(1-p+pe^t)^y\frac{\theta^y e^{-\theta}}{y!} \\
&= e^{-\theta}\sum_{y=0}^{\infty}\frac{\big[\theta(1-p+pe^t)\big]^y}{y!} \\
&= e^{-\theta}\text{exp}\{\theta(1-p+pe^t)\} \\
&= \text{exp}\{\theta p(e^t-1)\}
\end{aligned}$$

Actually, this is the m.g.f. of $\text{Poisson}(\theta p)$.

3.8 Joint Moment Generating Functions

DEFINITION 3.8.1: Joint moment generating function

If X and Y are random variables, then

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

is called the **joint moment generating function** of X and Y if $M(t_1, t_2)$ exists for $|t_1| < h_1$ and $|t_2| < h_2$ for some $h_1, h_2 > 0$.

REMARK 3.8.2

In general, suppose $X_1, \dots, X_n$ are random variables, then

$$M(t_1, \dots, t_n) = \mathbb{E}\left[\exp\left\{\sum_{i=1}^{n} t_i X_i\right\}\right]$$

is the **joint moment generating function** if it exists for $|t_i| < h_i$ for some $h_i > 0$ where $i = 1, \dots, n$.

REMARK 3.8.3: Applications of Joint Moment Generating Functions

(1) From joint m.g.f. to marginal m.g.f. Given $M(t_1, t_2)$ for $|t_1| < h_1$ and $|t_2| < h_2$ with $h_1, h_2 > 0$,

$$M_X(t_1) = M(t_1, t_2 = 0) = \mathbb{E}[e^{t_1 X}]$$

$$M_Y(t_2) = M(0, t_2) = \mathbb{E}[e^{t_2 Y}]$$

(2) Independence Property. X and Y are independent if and only if

$$M(t_1, t_2) = M_X(t_1) M_Y(t_2)$$

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

$$M(t_1, \dots, t_n) = \prod_{i=1}^{n} M_{X_i}(t_i)$$

EXAMPLE 3.8.4

Suppose $f(x, y) = e^{-y}$ for $0 < x < y$ is the joint p.d.f. of (X, Y). Find the joint m.g.f. of X and Y. Are they independent? Find the marginal p.d.f. of X and Y.

Solution.

$$
\begin{aligned}
M(t_1,t_2) &= \mathbb{E}[e^{t_1X+t_2Y}] \\
&= \int_0^\infty \left[\int_0^y e^{t_1x+t_2y}e^{-y}\,dx\right] dy \\
&= \int_0^\infty e^{(t_2-1)y}\left[\frac{1}{t_1}e^{t_1x}\right]_0^y dy \\
&= \frac{1}{t_1}\int_0^\infty e^{(t_2-1)y}(e^{t_1y}-1)\,dy \\
&= \frac{1}{t_1}\int_0^\infty e^{(t_1+t_2-1)y}-e^{(t_2-1)y}\,dy \\
&= \frac{1}{t_1}\left(\frac{1}{1-t_1-t_2}-\frac{1}{1-t_2}\right) \\
&= \frac{1}{(1-t_1-t_2)(1-t_2)}
\end{aligned}
$$

with $t_2-1<0$ and $t_1+t_2-1<0$. Therefore, $t_2<1$ and $t_1+t_2<1$.

$$M_X(t_1) = M(t_1,t_2=0) = \frac{1}{1-t_1}$$

which is the m.g.f. of Exponential(1).

$$M_Y(t_2) = M(t_1=0,t_2) = \frac{1}{(1-t_2)^2}$$

which is the m.g.f. of Gamma$(\alpha=2,\beta=1)$. Note that the joint support is a triangle (not a rectangle), so obviously $M(t_1,t_2)\neq M_X(t_1)M_Y(t_2)$. Thus, X and Y are not independent.

EXAMPLE 3.8.5: Additivity of Poisson Random Variables

Suppose $X\sim$ Poisson(θ_1) and $Y\sim$ Poisson(θ_2) with X and Y independent. Prove that $X+Y\sim$ Poisson$(\theta_1+\theta_2)$.
Solution. We can try to find the p.d.f. of $X+Y$ (direct method). Alternatively, find $M_{X+Y}(t)$.

$$
\begin{aligned}
M_{X+Y}(t) &= \mathbb{E}[e^{tX+tY}] \\
&= \mathbb{E}[e^{tX}e^{tY}] && X \text{ and } Y \text{ independent} \\
&= \mathbb{E}[e^{tX}]\mathbb{E}[e^{tY}] \\
&= \exp\{\theta_1(e^t-1)\}\exp\{\theta_2(e^t-1)\} \\
&= \exp\{(\theta_1+\theta_2)(e^t-1)\}
\end{aligned}
$$

which is the m.g.f. of Poisson$(\theta_1+\theta_2)$.

LECTURE 12 | 2020-10-18

3.9 Multinomial Distribution

DEFINITION 3.9.1: Multinomial distribution

$(X_1, \dots, X_k)$ are joint discrete random variables with joint p.f. given by

$$f(x_1, \dots, x_k) = \mathbb{P}(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1!x_2! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

where $x_i = 0, 1, \dots, n$ $(i = 1, 2, \dots, k)$. Furthermore, $\sum_{i=1}^k x_i = n$, $\sum_{i=1}^k p_i = 1$, for $0 < p_i < 1$ $i = 1, \dots, k$. Then, $(X_1, \dots, X_k)$ follows a **multinomial distribution.**

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

EXAMPLE 3.9.2: Possible Application

- There are k boxes and each box has same balls.
- The probability of choosing a ball from the i^{th} box is p_i for $i = 1, 2, \dots, k$.
- We randomly choose n balls from k boxes.

Let $X_i :=$ number of boxes from the i^{th} box for $i = 1, 2, \dots, k$. Then,

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

Note: if there are only two boxes, then $X_1 \sim \text{Binomial}(n, p_1)$.

PROPOSITION 3.9.3: Properties — Multinomial Distribution

If $(X_1, \dots, X_k) \sim$ Multinomial$(n; p_1, \dots, p_k)$, then

(1) $M(t_1, \dots, t_k) = \mathbb{E}[e^{t_1 X_1 + \cdots + t_k X_k}] = (p_1 e^{t_1} + \cdots + p_k e^{t_k})^n$ where $|t_i| < \infty$ for $i = 1, \dots, k$.
(2) $X_i \sim$ Binomial(n, p_i) for $i = 1, \dots, k$.
(3) If $T = X_i + X_j$ for $i \neq j$, then $T \sim$ Binomial$(n, p_i + p_j)$
(4) Cov$(X_i, X_j) = -np_i p_j$ for $i \neq j$
(5) The conditional probability function of X_i given $X_j = x_j$ for $i \neq j$ is

$$X_i \mid X_j = x_j \sim \textit{Binomial}\left(n - x_j, \frac{p_i}{1 - p_j}\right)$$

(6) The conditional distribution of X_i given $T = X_i + X_j$ for $i \neq j$ is

$$X_i \mid X_i + X_j \sim \textit{Binomial}\left(t, \frac{p_i}{p_i + p_j}\right)$$

Proof of Proposition 3.9.3

Proof of (1): Too long for my poor soul to type. Proof requires the Multinomial Theorem.
Proof of (2): The moment generating function of X_i for $i = 1, \dots, k$ is

$$M(0, \dots, 0, t, 0, \dots, 0) = \left[p_i e^{t_i} + (1 - p_i)\right]^n \quad t_i \in \mathbf{R}$$

which is the moment generating function of a Binomial(n, p_i) random variable. By Theorem 2.5.10 we have $X_i \sim$ Binomial(n, p_i) for $i = 1, \dots, k$.

Proof of (3): The moment generating function of $T = X_i + X_j$ for $i \neq j$ is

$$\begin{aligned} M_T(t) &= \mathbb{E}[e^{tT}] \\ &= \mathbb{E}\big[e^{t(X_i+X_j)}\big] \\ &= \mathbb{E}\big[e^{tX_i+tX_j}\big] \\ &= M(0,\dots,0,t,0,\dots,0,t,0,\dots,0) \\ &= (p_1+\cdots+p_ie^t+\cdots+p_je^t+\cdots+p_{k-1}+p_k)^n && t \in \mathbf{R} \\ &= \big[(p_i+p_j)e^t+(1-p_i-p_j)\big]^n && t \in \mathbf{R} \end{aligned}$$

which is the moment generating function of a Binomial$(n, p_i + p_j)$ random variable. By Theorem 2.5.10 we have $T \sim$ Binomial$(n, p_i + p_j)$ for $i \neq j$.
Proof of (4): By (2) we have $\mathbb{E}[X_i] = np_i$, $\mathbb{V}(X_i) = np_i(1-p_i)$, and $\mathbb{V}(X_j) = np_j(1-p_j)$. By (3) we have $X_i + X_j \sim$ Binomial$(n, p_i + p_j)$, so $\mathbb{V}(X_i + X_j) = n(p_i + p_j)(1 - p_i - p_j)$. Thus,

$$\text{Cov}(X_i + X_j, X_i + X_j) = \mathbb{V}(X_i) + \mathbb{V}(X_j) + 2\text{Cov}(X_i, X_j)$$

$$\implies n(p_i+p_j)(1-p_i-p_j) = np_i(1-p_i) + np_j(1-p_j) + 2\text{Cov}(X_i, X_j)$$

Therefore, $\text{Cov}(X_i, X_j) = -np_ip_j$.
Proof of (5): There are x_j outcomes from the j^{th} category. Therefore, there are $(n - x_j)$ balls chosen from the remaining $(k-1)$ boxes. We are not allowed to choose from the j^{th} box, we are only allowed to choose from the remaining $(k-1)$ boxes. Therefore, proportionally we get the success probability as $p_i/(1-p_j)$.

EXERCISE 3.9.4

Prove property (6) from Proposition 3.9.3.

3.10 Bivariate Normal Distribution

DEFINITION 3.10.1: Bivariate normal distribution

Suppose that X_1 and X_2 are continuous random variables with joint probability density function

$$f(x_1, x_2) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \quad (x_1, x_2) \in \mathbf{R}^2$$

Also,

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{2\times 1}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}_{2\times 1}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}_{k\times k}$$

and Σ is positive semi-definite. Also, $|\Sigma|$ is the determinant of Σ. Then, $\boldsymbol{X} = (X_1, X_2)^\top$ follows a **bivariate normal distribution**, and we write

$$\boldsymbol{X} \sim \text{BVN}(\boldsymbol{\mu}, \Sigma)$$

REMARK 3.10.2: †

Alternatively, we could write

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

PROPOSITION 3.10.3: Properties — Bivariate Normal Distribution

(1) X_1, X_2 *has joint moment generating function*

$$M(t_1, t_2) = \mathbb{E}[e^{t_1X_1+t_2X_2}] = exp\left\{\boldsymbol{t}^\top\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top\Sigma\boldsymbol{t}\right\} \quad \forall \boldsymbol{t} \in \mathbf{R}^2$$

(2) *Marginally,*

$$M_{X_1}(t_1) = M(t_1, 0) = exp\left\{t_1\mu_1 + \frac{1}{2}t_1^2\sigma_1^2\right\}$$

which is the m.g.f. of $\mathcal{N}(\mu_1, \sigma_1^2)$*; that is,* $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$*. Also,* $\mathbb{E}[X_1] = \mu_1$ *and* $\mathbb{V}(X_1) = \sigma_1^2$.

$$M_{X_2}(t_2) = M(0, t_2) = exp\left\{t_2\mu_2 + \frac{1}{2}t_2^2\sigma_2^2\right\}$$

which is the m.g.f. of $\mathcal{N}(\mu_2, \sigma_2^2)$*; that is,* $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$*. Also,* $\mathbb{E}[X_2] = \mu_2$ *and* $\mathbb{V}(X_2) = \sigma_2^2$.

(3) *Conditional distribution.*

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

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

$$f_2(x_2 \mid x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}$$

$$f_1(x_1 \mid x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}$$

(4) *Cov*$(X_1, X_2) = \rho\sigma_1\sigma_2$

(5) $\rho = 0 \iff X_1$ *and* X_2 *are independent.*

(6) *Linear transformations of bivariate normal are still normal.*

(7) $(\boldsymbol{X} - \boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2(2)$

Proof of Proposition 3.10.3

Proof of (4): We want to find $\mathbb{E}[X_1X_2] = \mathbb{E}[\mathbb{E}[X_1X_2 \mid X_1]]$.
Step 1:

$$\mathbb{E}[X_1X_2 \mid X_1 = x_1] = \mathbb{E}[x_1X_2 \mid X_1 = x_1] = x_1\mathbb{E}[X_2 \mid X_1 = x_1] = x_1\left(\mu_2 + \frac{\rho\sigma_2(x_1-\mu_1)}{\sigma_1}\right)$$

Step 2:

$$\mathbb{E}[X_1X_2 \mid X_1] = X_1\left(\mu_2 + \frac{\rho\sigma_2(X_1-\mu_1)}{\sigma_1}\right)$$

$$\begin{aligned}\mathbb{E}[X_1X_2] &= \mathbb{E}\left[X_1\mu_2 + \frac{X_1\rho\sigma_2(X_1-\mu_1)}{\sigma_1}\right] \\ &= \mu_2\mathbb{E}[X_1] + \frac{\rho\sigma_2}{\sigma_1}(\mathbb{E}[X_1^2] - \mu_1\mathbb{E}[X_1]) \\ &= \mu_2\mu_1 + \frac{\rho\sigma_2}{\sigma_1}(\mu_1^2+\sigma_1^2-\mu_1^2) \\ &= \mu_1\mu_2 + \rho\sigma_1\sigma_2\end{aligned}$$

Thus,

$$\text{Cov}(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \rho\sigma_1\sigma_2$$

$$\text{Corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\mathbb{V}(X_1)\mathbb{V}(X_2)}} = \rho$$

Proof of (5): We know if X_1 and X_2 are independent, then $\rho = 0$. If $\rho = 0$, e.g., $X_2X_1 = x_1 \sim \mathcal{N}(\mu_2, \sigma_1^2)$ and $X_1X_2 = x_2 \sim \mathcal{N}(\mu_1, \sigma_1^2)$. In summary: If joint bivariate normal then uncorrelated = independence.

Proof of (6): Let $\boldsymbol{c} = (c_1, c_2)^\top$, then $\boldsymbol{c}^\top X = c_1X_2 + c_2X_2 \sim \mathcal{N}(c_1\mu_1 + c_2\mu_2, \boldsymbol{c}^\top \Sigma \boldsymbol{c})$. Furthermore, if $A \in \mathbf{R}^{2\times 2}$, and $\boldsymbol{b} = (b_1, b_2)^\top$, then

$$AX + \boldsymbol{b} \sim \text{BVN}(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A^\top)$$

Two linear combinations of BVN is joint BVN.

REMARK 3.10.4

Remark of (7): Note $\chi^2(1) := Z^2$ where $Z \sim \mathcal{N}(0, 1)$.

$$\chi^2(n) = \sum_{i=1}^{n} Z_i^2$$

where $Z_1, \ldots, Z_n$ are independent $\mathcal{N}(0, 1)$.

Chapter 4

Function of Random Variables

LECTURE 13 | 2020-10-25

Let $(X_1, \dots, X_n)$ be continuous random variables. We want to find the distribution of $Y = h(X_1, \dots, X_n)$.

Three methods here:

(1) Cumulative Distribution Function Technique

(2) One-to-One Transformation

(3) Moment Generating Function Technique

- (1) and (3) are useful to find marginal distribution $Y = h(X_1, \dots, X_n)$.
- (3) is useful to find both univariate and multivariate functions. For example,

$$Y_1 = h_1(X_1, \dots, X_n) \quad \text{and} \quad Y_2 = h_2(X_1, \dots, X_n)$$

If we want to find the joint distribution of more than one function, we can use this method.

4.1 Cumulative Distribution Function Technique

Tutorial 5: $T = \mathbb{E}[X \mid Y] = \frac{3}{4}Y$.

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

Step 1: Find the c.d.f. of Y by definition.

$$F_Y(y) = \mathbb{P}(Y \le y) = 1 - \mathbb{P}(Y > y)$$

Step 2: Find the p.d.f. of Y by

$$f(y) = F_Y'(y)$$

EXAMPLE 4.1.1: Cumulative Distribution Function Technique

Suppose the joint p.d.f. of (X, Y) is $f(x, y) = 3y$ for $0 \le x \le y \le 1$. Find the p.d.f. of $T = XY$ and p.d.f. of $S = Y/X$.

Solution. $T = XY$. Support of T is $(0, 1)$.

- If $t \ge 1$, then $F_T(t) = \mathbb{P}(T \le t) = 1$.
- If $t \le 0$, then $F_T(t) = 0$.

- If $0 < t < 1$, then

$$\begin{aligned}F_T(t) &= \mathbb{P}(T \le t)\\ &= \mathbb{P}(XY \le t)\\ &= 1 - \mathbb{P}(XY > t)\\ &= 1 - \left(\int_{\sqrt{t}}^{1}\int_{t/y}^{y} 3y\,dx\,dy\right)\\ &= 1 - (2t^{3/2} - 3t + 1)\\ &= 3t - 2t^{3/2}\end{aligned}$$

Therefore, the p.d.f. of T for $0 < t < 1$ is

$$f_T(t) = 3 - 3\sqrt{t}$$

$S = Y/X$. Support of S is $(1, \infty)$.

- If $s < 1$, then $F_S(s) = 0$
- If $s \ge 1$, then

$$\begin{aligned}F_S(s) &= \mathbb{P}(S \le s)\\ &= \mathbb{P}\left(\frac{Y}{X} \le s\right)\\ &= \mathbb{P}(Y \le sX)\\ &= \int_0^1\int_{y/s}^{y} 3y\,dx\,dy\\ &= 1 - \frac{1}{s}\end{aligned}$$

Therefore, the p.d.f. of S for $s \ge 1$ is

$$f_S(s) = \frac{1}{s^2}$$

EXAMPLE 4.1.2: Distribution of maximum and minimum

Suppose $(X_1, \dots, X_n) \overset{\text{iid}}{\sim} \text{Uniform}(0, \theta)$. Find the p.d.f. of the largest order statistic; that is,

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

and the smallest order statistic; that is,

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

Solution. $F_{X_{(n)}}(y) = \mathbb{P}(X_{(n)} \le y)$.

- If $y \le 0$, then $F_{X_{(n)}}(y) = 0$.
- If $y \ge \theta$, then $F_{X_{(n)}}(y) = 1$.
- If $0 < y < \theta$, then

$$\begin{aligned}F_{X_{(n)}}(y) &= \mathbb{P}(X_{(n)} \le y)\\ &= \mathbb{P}(X_1 \le y, \dots, X_n \le y)\\ &= \mathbb{P}(X_1 \le y)\cdots\mathbb{P}(X_n \le y)\\ &= \left(\frac{y}{\theta}\right)^n\end{aligned}$$

The p.d.f. of $X_{(n)}$ for $0 < y < \theta$ is

$$f_{X_{(n)}}(y) = \frac{n}{\theta^n} y^{n-1}$$

For $X_{(1)}$ the support is $[0, \theta]$. If $0 < y < \theta$,

$$\begin{aligned} F_{X_{(1)}}(y) &= \mathbb{P}(X_{(1)} \leq y) \\ &= 1 - \mathbb{P}(X_{(1)} > y) \\ &= 1 - \left[\mathbb{P}(X_1 > y) \cdots \mathbb{P}(X_n > y)\right] \\ &= 1 - \left(\frac{\theta - y}{\theta}\right)^n \end{aligned}$$

The p.d.f. of $X_{(1)}$ for $0 < y < \theta$ is

$$f_{X_{(1)}}(y) = \frac{n}{\theta}\left(1 - \frac{y}{\theta}\right)^{n-1}$$

EXERCISE 4.1.3

If $X_1, \dots, X_n \stackrel{\text{iid}}{\sim}$ Exponential(1), find $X_{(n)}$ and $X_{(1)}$.

LECTURE 14 | 2020-10-25

4.2 One-to-One Transformations (Univariate)

EXAMPLE 4.2.1: Cumulative Distribution Function Technique

If $X \sim \mathcal{N}(0,1)$, find the p.d.f. of $Y = X^2$.
Solution. Support of Y is $[0,\infty)$. The c.d.f. of Y for $y > 0$ is

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(X^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

The p.d.f. of Y is

$$\begin{aligned} f_Y(y) &= F_Y'(y) \\ &= F_X'(\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right) - F_X'(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right) \\ &= \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})] \\ &= \frac{1}{2\sqrt{y}}\left[\frac{1}{\sqrt{2\pi}}\text{exp}\left\{-\frac{(\sqrt{y})^2}{2}\right\} + \frac{1}{\sqrt{2\pi}}\text{exp}\left\{-\frac{(-\sqrt{y})^2}{2}\right\}\right] \\ &= \frac{1}{2\sqrt{y}}\left[\frac{2}{\sqrt{2\pi}}\text{exp}\left\{-\frac{y}{2}\right\}\right] \\ &= \frac{1}{\sqrt{2\pi}}y^{-1/2}\text{exp}\left\{-\frac{y}{2}\right\} \end{aligned}$$

The p.d.f. of Y is also $\chi^2(1)$ or Gamma$(\alpha = 1/2, \beta = 2)$.

EXAMPLE 4.2.2: Cumulative Distribution Function Technique

Suppose the p.d.f. of X is $f(x) = \dfrac{\theta}{x^{\theta+1}}$ for $x \geq 1$ and $\theta > 0$. Find the p.d.f. of $Y = \ln(X)$.
Solution. Support of Y is $[0,\infty)$. The c.d.f. of Y for $y > 0$ is

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(\ln(X) \leq y) \\ &= \mathbb{P}(X \leq e^y) \\ &= \int_1^{e^y} \frac{\theta}{x^{\theta+1}}\,dx \\ &= 1 - e^{-y\theta} \end{aligned}$$

The p.d.f. of Y is

$$F_Y'(y) = f_Y(y) = \begin{cases} \theta e^{-y\theta} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

Special case: If $h(x)$ is a one-to-one transformation on the support of X, then we have a formula to find p.d.f. of $Y = h(X)$.

THEOREM 4.2.3: One-to-One Univariate Transformations

If $h(x)$ is one-to-one transformation on the support of X, then the probability density function of Y is given by

$$g_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

REMARK 4.2.4

Replace x in the right-hand side by function of y; that is, $x = h^{-1}(y)$ (inverse of h).

EXAMPLE 4.2.5: One-to-One Transformation (Univariate)

Suppose the p.d.f. of X is $f(x) = \dfrac{\theta}{x^{\theta+1}}$ for $x \geq 1$ and $\theta > 0$. Find the p.d.f. of $Y = \ln(X)$.
Solution. Support of Y is $[0, \infty)$. $h(x) = \ln(x)$ is a one-to-one transformation. For $y > 0$ we have

$$\begin{aligned} g_Y(y) &= f_X(x)\left|\frac{dx}{dy}\right| & \qquad y = \ln(x) \implies x = e^y \\ &= f_X(e^y)|e^y| \\ &= \frac{\theta}{(e^y)^{\theta+1}}(e^y) \\ &= \theta e^{-\theta y} \end{aligned}$$

Note that $\dfrac{dx}{dy} = \dfrac{1}{dy/dx} = \dfrac{1}{1/x} = x$. So we could've done

$$\begin{aligned} g_Y(y) &= f_X(x)\left|\frac{dx}{dy}\right| \\ &= f_X(e^y)|x| \\ &= \frac{\theta}{(e^y)^{\theta+1}}(e^y) \\ &= \theta e^{-y\theta} \end{aligned}$$

EXAMPLE 4.2.6: One-to-One Transformation (Univariate)

Suppose $X \sim \mathcal{N}(0, 1)$ and the c.d.f. of X is $\Phi(x)$. Find the p.d.f. of $Y = \Phi(X)$.
Solution. Support of Y is $[0, 1]$. The p.d.f. of Y for $0 \leq y \leq 1$ is

$$\begin{aligned} g_Y(y) &= f_X(x)\left|\frac{dx}{dy}\right| \\ &= f_X(x)\left|\frac{1}{dy/dx}\right| & \qquad y = \Phi(x) \implies \frac{dy}{dx} = \Phi'(x) = f_X(x) \\ &= f_X(x)\left|\frac{1}{f_X(x)}\right| \\ &= 1 \end{aligned}$$

Thus, $Y \sim \text{Uniform}(0, 1)$.

EXAMPLE 4.2.7: One-to-One Transformation (Univariate)

Suppose $X \sim \text{Uniform}(0,1)$. Find the p.d.f. of $Y = -\ln(X)$.
Solution. Support of Y is $[0,\infty)$. Note that $y = -\ln(x) \implies dy/dx = -1/x$. The p.d.f. of Y for $y > 0$ is

$$\begin{aligned} g_Y(y) &= f_X(x)\left|\frac{dx}{dy}\right| \\ &= 1\left|\frac{1}{dy/dx}\right| \\ &= x \\ &= e^{-y} \end{aligned}$$

where the last equality follows since $y = -\ln(x) \implies x = e^{-y}$ for $y > 0$.

REMARK 4.2.8

The c.d.f. technique is always useful, but the one-to-one transformation is less useful, and you are more likely to make a mistake. It is not recommended using the formula.

LECTURE 15 | 2020-11-01

Find the p.d.f. of $Y = h(X)$. Two possible ways:

- Method 1: CDF Technique
- Method 2: If $h(x)$ is a one-to-one function, then

$$g_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

4.3 One-to-One Transformations (Bivariate)

Given X and Y, the joint p.d.f. of (X,Y) is $f(x,y)$. We would like to find the joint p.d.f. of

$$U = h_1(X,Y) \quad \text{and} \quad V = h_2(X,Y)$$

One-to-one bivariate transformation

$$u = h_1(x,y) \quad \text{and} \quad v = h_2(x,y)$$

The two functions are a one-to-one transformation if there exist another two unique functions such that

$$x = w_1(u,v) \quad \text{and} \quad y = w_2(u,v)$$

for (x,y) in support of (X,Y).

THEOREM 4.3.1: One-to-One Bivariate Transformations

The p.d.f. of $U = h_1(X,Y)$ and $V = h_2(X,Y)$ is given by

$$g(u,v) = f(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|$$

where the Jacobian matrix is

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

Step 1: Find support of (U, V) by making use of h_1, h_2, and support of (X, Y).

Step 2: $u = h_1(x, y)$ and $v = h_2(x, y)$ implies $x = w_1(u, v)$ and $y = w_2(u, v)$, compute Jacobian:

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix}$$

Step 3:

$$g(u,v) = f(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|$$

EXAMPLE 4.3.2: One-to-One Transformation (Bivariate)

Suppose $X \sim \mathcal{N}(0,1)$ and $\mathcal{N}(0,1)$ independent. Find the joint p.d.f. of $U = X + Y$ and $V = X - Y$.
Solution. Since X and Y are independent, the joint p.d.f. of X and Y is given by

$$f(x,y) = f_X(x)f_Y(y) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\} = \frac{1}{2\pi}\exp\left\{-\frac{x^2+y^2}{2}\right\}$$

Step 1: $u = x + y$ and $v = x - y$ implies $x = (u + v)/2$ and $y = (u - v)/2$. Support of U and V is $(-\infty, \infty)$.
Step 2: Jacobian is given by

$$\begin{aligned}
\frac{\partial(x,y)}{\partial(u,v)} &= \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix} \\
&= \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} \\
&= \left(\frac{1}{2}\right)\left(-\frac{1}{2}\right) - \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) \\
&= -\frac{2}{4} \\
&= -\frac{1}{2}
\end{aligned}$$

Step 3:

$$\begin{aligned}
g(u,v) &= f(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right| \\
&= \frac{1}{2\pi}\exp\left\{-\frac{x^2+y^2}{2}\right\}\left|-\frac{1}{2}\right| \\
&= \frac{1}{4\pi}\exp\left\{-\frac{[(u+v)/2]^2+[(u-v)/2]^2}{2}\right\} \\
&= \frac{1}{4\pi}\exp\left\{-\frac{u^2+v^2}{4}\right\}
\end{aligned}$$

EXAMPLE 4.3.3: One-to-One Transformation (Bivariate)

Suppose that X and Y are continuous random variables with joint p.d.f. $f(x,y) = e^{-x-y}$ for $0 < x < \infty$ and $0 < y < \infty$. Find the joint p.d.f. of $U = X + Y$ and $V = X$. Find the marginal p.d.f. of U.
Solution. $u = x + y$ and $v = x$ implies $x = v$ and $y = u - v$. Therefore, $0 < v < \infty$ and $0 < u - v < \infty$.

In other words, the joint support of (U, V) is $0 < v < u < \infty$. Jacobian is

$$\begin{aligned}\frac{\partial(x,y)}{\partial(u,v)} &= \begin{vmatrix}\partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v\end{vmatrix} \\ &= \begin{vmatrix}0 & 1 \\ 1 & -1\end{vmatrix} \\ &= -1\end{aligned}$$

Therefore, the joint p.d.f. of (U, V) for $0 < v < u < \infty$ is

$$\begin{aligned}g(u,v) &= f(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right| \\ &= e^{-x-y}|-1| \\ &= e^{-(x+y)} \\ &= e^{-u}\end{aligned}$$

Support of U is $(0, \infty)$. The marginal p.d.f. of U for $u > 0$ is

$$f_1(u) = \int_{-\infty}^{\infty} g(u,v)\,dv = \int_0^u e^{-u}\,dv = ue^{-u}$$

Find the p.d.f. of $U = X + Y$.

1. CDF Technique
2. Define $V = X$ (or $V = Y$), find (U, V) with the Theorem.

EXAMPLE 4.3.4: Support of One-to-One Transformation (Bivariate)

Suppose that the support of (X, Y) is $0 < x < y < 1$. Find the support of (U, V) where $U = X$ and $V = XY$.
Solution. $u = x$ and $v = xy$ implies $x = u$ and $y = v/u$.

$$0 < u < \frac{v}{u} < 1 \implies 0 < u^2 < v < u < 1$$

(multiply by u)

EXAMPLE 4.3.5: Support of One-to-One Transformation (Bivariate)

Suppose the support of (X, Y) is $0 < x < 1$ and $0 < y < 1$. Find the support of (U, V) where $U = X/Y$ and $V = XY$.
Solution. $u = x/y$ and $v = xy$.

$$uv = x^2 \implies x = \sqrt{uv}$$

$$y = \frac{v}{x} \implies y = \frac{v}{\sqrt{uv}} = \frac{v^{1/2}}{u^{1/2}v^{1/2}} = \sqrt{\frac{v}{u}}$$

So,

$$0 < \sqrt{uv} < 1 \implies 0 < uv < 1 \implies 0 < u < \frac{1}{v} \quad (v > 0)$$

$$0 < \sqrt{\frac{v}{u}} < 1 \implies 0 < \frac{v}{u} < 1 \implies 0 < v < u \quad (u > 0)$$

Combining, we get $0 < v < u < 1/v$.

LECTURE 16 | 2020-11-01

4.4 Moment Generating Function Technique

Idea:

(1) Find the moment generating function of a random variable

(2) Use the uniqueness theorem of moment generating function to find the distribution of the random variable and then the p.d.f. of a random variable.

THEOREM 4.4.1

Suppose $X_1, \dots, X_n$ are independent, then $T = \sum_{i=1}^n X_i$ has moment generating function

$$M_T(t) = \mathbb{E}\left[exp\left\{t\sum_{i=1}^n X_i\right\}\right] = \mathbb{E}\left[\prod_{i=1}^n exp\{tX_i\}\right] = \prod_{i=1}^n \mathbb{E}[exp\{tX_i\}] = \prod_{i=1}^n M_{X_i}(t)$$

In particular, if $X_1, \dots, X_n$ are independently and identically distributed, then they have the exact same moment generating function $M(t)$; that is,

$$M_T(t) = [M(t)]^n$$

Next, we use the m.g.f. technique to find properties of normal, χ^2, t-distribution, and F-distributions.

LEMMA 4.4.2

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

Proof of Lemma 4.4.2

Recall that the m.g.f. of $X \sim \mathcal{N}(\mu, \sigma^2)$ is

$$M_X(t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$$

Therefore,

$$\begin{aligned} M_{aX+b}(t) &= \mathbb{E}\left[e^{t(aX+b)}\right] \\ &= e^{bt}\mathbb{E}[e^{taX}] \\ &= e^{bt}M_X(ta) \\ &= e^{bt}\exp\left\{\mu(ta) + \frac{\sigma^2(at)^2}{2}\right\} \\ &= \exp\left\{(a\mu + b)t + \frac{a^2\sigma^2 t^2}{2}\right\} \end{aligned}$$

which is the m.g.f. $\mathcal{N}(a\mu + b, a^2\sigma^2)$.

THEOREM 4.4.3

If $X \sim \mathcal{N}(\mu, \sigma^2)$*, then*

$$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

THEOREM 4.4.4: Linear Combination of Independent Normal Random Variables

If $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$*,* $i = 1, \dots, n$ *independently, then*

$$\sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Proof of Theorem 4.4.4

By Lemma 4.4.2, we have $a_i X_i \sim \mathcal{N}(a_i \mu_i, a_i^2 \sigma_i^2)$ for $i = 1, \dots, n$ and the m.g.f.

$$M_{a_i X_i}(t) = \exp\left\{(a_i \mu_i)t + \frac{a_i^2 \sigma_i^2}{2} t^2\right\}$$

Therefore,

$$\begin{aligned}
M_{\sum_{i=1}^n a_i X_i}(t) &= \mathbb{E}\left[\exp\left\{t \sum_{i=1}^{n} a_i X_i\right\}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} e^{(a_i X_i)t}\right] \\
&= \prod_{i=1}^{n} \mathbb{E}\left[e^{(a_i X_i)t}\right] \\
&= \prod_{i=1}^{n} M_{a_i X_i}(t) \\
&= \prod_{i=1}^{n} \exp\left\{(a_i \mu_i)t + \frac{\sigma_i^2 a_i^2}{2} t^2\right\} \\
&= \exp\left\{\left(\sum_{i=1}^{n} a_i \mu_i\right)t + \frac{(\sum_{i=1}^{n} a_i^2 \sigma_i^2)t^2}{2}\right\}
\end{aligned}$$

which is the m.g.f. of $\mathcal{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$.

COROLLARY 4.4.5

If $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$*, then*

(1) $\displaystyle\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$

(2) $\displaystyle\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

Proof of Corollary 4.4.5

(1) Let $a_i = 1$, $\mu_i = \mu$, $\sigma_i^2 = \sigma$ in Theorem 4.4.4.
(2) Let $a_i = \frac{1}{n}$, $\mu_i = \mu$, $\sigma_i^2 = \sigma$ in Theorem 4.4.4.

DEFINITION 4.4.6: Chi-Squared Distribution

If $Z_1, \dots, Z_k \sim \mathcal{N}(0,1)$ are independent and $0 < k \in \mathbf{Z}$, then

$$Q = \sum_{i=1}^{k} Z_i^2$$

follows a **chi-squared distribution** with k degrees of freedom and write $Q \sim \chi^2(k)$.

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2(1)$$

If $Y_i \sim \chi^2(k_i)$ are independent, then

$$\sum_{i=1}^{n} Y_i \sim \chi^2\left(\sum_{i=1}^{n} k_i\right)$$

The m.g.f. of $\chi^2(1)$ is $(1-2t)^{-1/2}$. Derive the m.g.f. $\chi^2(n)$: $(1-2t)^{-n/2}$.

$$\chi^2(n) = \sum_{i=1}^{n} X_i^2 \quad X_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$

Let $T = \sum_{i=1}^{n} Y_i$, then

$$M_T(t) = \prod_{i=1}^{n} M_{Y_i}(t) = \prod_{i=1}^{n} (1-2t)^{-k/2} = (1-2t)^{-\sum_{i=1}^{n} k_i/2}$$

If $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$

$$\sum_{i=1}^{n} \left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi(n)$$

DEFINITION 4.4.7: Student's t-distribution

Let $Z \sim \mathcal{N}(0,1)$ and $Q \sim \chi^2(\nu)$ be independent, then

$$T = \frac{Z}{\sqrt{Q/\nu}}$$

follows a **student's t-distribution** with k degrees of freedom and write $T \sim t(\nu)$ where $\nu > 0$.
Support of T: $(-\infty, \infty)$.

DEFINITION 4.4.8: F-distribution

If $X \sim \chi^2(n)$ and $Y \sim \chi^2(m)$ are independent, then

$$\frac{X/n}{Y/m} \sim F(n, m)$$

follows a **F-distribution**.
Support of $F(n, m)$:

- If $n = 1$: $[0, \infty)$.
- If $n \neq 1$: $(0, \infty)$.

If $X \sim \chi^2(n)$ and $Y \sim \chi^2(m)$ are independent, then

$$X + Y \sim \chi^2(n + m)$$

EXERCISE 4.4.9

Prove or disprove.

$$\frac{X/n}{(X+Y)/(n+m)} \sim F(n, n+m)$$

Solution. False. Define $Z = \dfrac{(X+Y)/(m+n)}{X/n}$, we have

$$Z = \frac{n}{m+n}\left(\frac{X+Y}{X}\right) = \frac{n}{m+n} + \frac{n}{m+n}\left(\frac{Y}{X}\right) = \frac{n}{m+n} + \frac{Y/m}{X/n}\left(\frac{m}{m+n}\right)$$

Assume $n > 2$, then

$$\mathbb{E}[Z] = \frac{n}{m+n} + \mathbb{E}\left[\frac{Y/m}{X/n}\right]\left(\frac{m}{m+n}\right) = \frac{n}{m+n} + \frac{n}{n-2}\left(\frac{m}{m+n}\right) \neq \frac{n}{n-2}$$

Thus, Z does not follow $F(m+n, n)$, hence $\dfrac{1}{Z}$ does not follow $F(n, n+m)$.

LEMMA 4.4.10: Useful Identity

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

Proof of Lemma 4.4.10

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \bar{X}) + \sum_{i=1}^{n}(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2
\end{aligned}$$

since

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\bar{X}) &= \sum_{i=1}^{n}X_i-\sum_{i=1}^{n}\bar{X}\\ &= \sum_{i=1}^{n}X_i-n\bar{X}\\ &= \sum_{i=1}^{n}X_i-n\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)\\ &= \sum_{i=1}^{n}X_i-\sum_{i=1}^{n}X_i\\ &= 0\end{aligned}$$

THEOREM 4.4.11

If $X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \dots, n$ independently, then

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

Proof of Theorem 4.4.11

By Lemma 4.4.10 we have

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

Dividing both sides by σ^2 gives

$$\underbrace{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}}_{Y} = \underbrace{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}}_{U} + \underbrace{\frac{n(\bar{X}-\mu)^2}{\sigma^2}}_{V}$$

Note that $\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1)$, thus

$$V = \frac{n(\bar{X}-\mu)^2}{\sigma^2} = \left[\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right]^2 \sim \chi^2(1)$$

Previously, we derived $Y = \dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$.
Since U and V are independent and $Y = U + V$, then

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(U+V)}] = \mathbb{E}[e^{tU}]\mathbb{E}[e^{tV}] = M_U(t)M_V(t)$$

Thus,

$$(1-2t)^{-n/2} = M_U(t)(1-2t)^{-1/2} \quad t<\frac{1}{2}$$

$$\implies M_U(t) = (1-2t)^{-(n-1)/2} \quad t<\frac{1}{2}$$

which is the m.g.f. of $\chi^2(n-1)$.

Why $\bar{X}$ is independent of $\sum_{i=1}^{n}(X_i - \bar{X})^2$?

$$(\underbrace{\bar{X}}_{0}, X_1 - \bar{X}, \dots, X_n - \bar{X}) \sim \text{MVN}(\cdot)$$

Verify that $\bar{X}$ independent of $(X_1 - \bar{X}, \dots, X_n - \bar{X})$ by calculating the correlation.

EXAMPLE 4.4.12: t-distribution

If $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, then

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

where

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

is defined as the sample variance ($\mathbb{E}[S^2] = \sigma^2$).
Solution.

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

are independent, then

$$\frac{\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

EXAMPLE 4.4.13: F-distribution

If $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y_1, \dots, Y_m \overset{\text{iid}}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$ are independent. Define

$$S_1^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S_2^2 = \frac{\sum_{i=1}^{m}(Y_i - \bar{Y})^2}{m-1}, \quad \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$$

Then,

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1)$$

Reasoning:

$$\frac{S_1^2}{\sigma_1^2} = \frac{\dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma_1^2}}{n-1} \sim \frac{\chi^2(n-1)}{n-1}$$

$$\frac{S_2^2}{\sigma_2^2} \sim \frac{\chi^2(m-1)}{m-1}$$

are independent, therefore,

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \frac{\chi^2(n-1)/(n-1)}{\chi^2(m-1)/(m-1)} = F(n-1, m-1)$$

Chapter 5

Limiting/Asymptotic Distribution

LECTURE 17 | 2020-11-08

Motivation: We're very interested in the distribution $\sqrt{n}(\bar{X}-\mu)$, here $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ with c.d.f. F with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}(X_i) = \sigma^2$,

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

In practice, we don't know the distribution of X_i.

REMARK 5.0.1

(i) It is impossible to find the exact distribution of $\sqrt{n}(\bar{X} - \mu)$.
(ii) Main idea: are we able to find an approximate distribution for $\sqrt{n}(\bar{X} - \mu)$? Concept of limiting/asymptotic distribution is introduced for this purpose.

Let $F_n(x)$ be the c.d.f. of $\sqrt{n}(\bar{X} - \mu)$; that is, $F_n(x) = \mathbb{P}\big(\sqrt{n}(\bar{X} - \mu) \leq x\big)$. Consider: $\lim\limits_{n\to\infty} F_n(x)$ (pointwise limit) and find that $\lim\limits_{n\to\infty} F_n(x) = F(x)$ where $F(x)$ is a known distribution, e.g., normal c.d.f. then we can use $F(x)$ to approximate $F_n(x)$ for a sufficiently large n.

To continue, we need some formal definition of this limit mathematically.

5.1 Convergence in Distribution

DEFINITION 5.1.1: Convergence in Distribution

Let $X_1, \ldots, X_n$ be a sequence of random variables such that X_n has c.d.f. $F_n(x)$. Let X be another random variable with c.d.f. $F(x)$. If

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all x at which $F(x)$ is continuous, then we say X_n **converges in distribution** to X, and write $X_n \xrightarrow{\text{d}} X$.

REMARK 5.1.2

(i) $F(x)$ is called the limiting distribution (or asymptotic distribution) of X_n.
(ii) It's the c.d.f. to which X_n converges to, not the random variables. This means, $F_n(x) \approx F(x)$ for n sufficiently large, however X_n is not approximately, X.
(iii) $\lim_{n\to\infty} F_n(x) = F(x)$ only for continuous points of $F(x)$, e.g.,

$$F(x) = \begin{cases} 1 & x \geq a \\ 0 & x < a \end{cases}$$

which is the c.d.f. of constant $X = a$; that is, $\mathbb{P}(X = a) = 1$. It's easy to tell that the c.d.f. of X is not continuous. $X_n \to X$ with c.d.f. $F(x)$ if $\lim_{n\to\infty} F_n(x) = F(x)$ for $x \neq a$; that is,

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 1 & x > a \\ 0 & x < a \end{cases}$$

we don't care what's the limit of $F_n(x)$ as $n \to \infty$.
(iv) This definition holds for both discrete and continuous random variables.

THEOREM 5.1.3: e Limit

Let $b, c \in \mathbf{R}$, $\lim_{n\to\infty} \psi(n) = 0$.

$$\lim_{n\to\infty} \left[1 + \frac{b}{n} + \frac{\psi(n)}{n}\right]^{cn} = e^{bc}$$

COROLLARY 5.1.4

Let $b, c \in \mathbf{R}$.

$$\lim_{n\to\infty} \left[1 + \frac{b}{n}\right]^{cn} = e^{bc}$$

EXAMPLE 5.1.5

Suppose that $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Uniform}(0, 1)$. Let $X_{(1)} = \min(X_1, \ldots, X_n)$ and $X_{(n)} = \max(X_1, \ldots, X_n)$. Find the limiting distribution of
(i) $nX_{(1)}$ and $n(1 - X_{(n)})$
(ii) $X_{(1)}$ and $X_{(n)}$

Solution.
(i) $nX_{(1)}$. Support is $[0, n]$, so the c.d.f. of $nX_{(1)}$ is:
- $x \geq n$, $F_n(x) = \mathbb{P}(nX_{(1)} \leq x) = 1$
- $x \leq 0$, $F_n(x) = \mathbb{P}(nX_{(1)} \leq x) = 0$
- $0 < x < n$,

$$\begin{aligned} F_n(x) &= \mathbb{P}(nX_{(1)} \leq x) \\ &= \mathbb{P}\Big(X_{(1)} \leq \frac{x}{n}\Big) \\ &= 1 - \mathbb{P}\Big(X_1 > \frac{x}{n}, \ldots, X_n > \frac{x}{n}\Big) \\ &= 1 - \left[\mathbb{P}\Big(X_1 > \frac{x}{n}\Big)\right]^n \\ &= 1 - \left(1 - \frac{x}{n}\right)^n \end{aligned}$$

Therefore,

$$\mathbb{P}(nX_{(1)} \le x) := F_n(x) = \begin{cases} 0 & x \le 0 \\ 1 - \left(1 - \frac{x}{n}\right)^n & 0 < x < n \\ 1 & x \ge n \end{cases}$$

$$\implies \lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-x} & x > 0 \end{cases}$$

Aside: $\lim_{n\to\infty}\left(1+\frac{x}{n}\right) = e^x$ which is the c.d.f. of Exponential(1).

$n(1-X_{(n)})$. Support is $[0,n]$, so the c.d.f. of $n(1-X_{(n)})$ is

- $x \ge n$, $F_n(x) = \mathbb{P}(n(1-X_{(n)}) \le x) = 1$
- $x \le 0$, $F_n(x) = \mathbb{P}(n(1-X_{(n)}) \le x) = 0$
- $0 < x < n$,

$$\begin{aligned} F_n(x) &= \mathbb{P}(n(1-X_{(n)}) \le x) \\ &= \mathbb{P}\Big(1 - X_{(n)} \le \frac{x}{n}\Big) \\ &= 1 - \mathbb{P}\Big(X_{(n)} < 1 - \frac{x}{n}\Big) \\ &= 1 - \mathbb{P}\Big(X_1 < 1 - \frac{x}{n}, \dots, X_n < 1 - \frac{x}{n}\Big) \\ &= 1 - \Big[\mathbb{P}\Big(X_1 < 1 - \frac{x}{n}\Big)\Big]^n \\ &= 1 - \left(1 - \frac{x}{n}\right)^n \end{aligned}$$

Therefore,

$$F_n(x) = \begin{cases} 0 & x \le 0 \\ 1 - \left(1 - \frac{x}{n}\right)^n & 0 < x < n \\ 1 & x \ge n \end{cases}$$

$$\implies \lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-x} & x > 0 \end{cases}$$

which is the c.d.f. of Exponential(1).

(ii) $X_{(1)}$. Support $(0,1)$.

$$F_n(x) = \mathbb{P}(X_{(1)} \le x) = \begin{cases} 0 & x \le 0 \\ 1 - (1-x)^n & 0 < x < 1 \\ 1 & x \ge 1 \end{cases}$$

$$\implies \lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x \le 0 \\ 1 & 0 < x < 1 \\ 1 & x \ge 1 \end{cases} = \begin{cases} 0 & x \le 0 \\ 1 & x > 0 \end{cases}$$

Question: What is $F(x)$?

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 & x \ge 0 \end{cases}$$

will make $F(x)$ right-continuous. $F(x)$ is not continuous at $x = 0$. Here, we don't require that $F_n(x)$ converges to $F(x)$ at $x = 0$. $F(x)$ is actually the c.d.f. of X which satisfies $\mathbb{P}(X = 0) = 1$.

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x \le 0 \\ 0 & 0 < x < 1 \\ 1 & x \ge 1 \end{cases} = \begin{cases} 0 & x < 1 \\ 1 & x \ge 1 \end{cases}$$

which is right-continuous.
Therefore, $\lim_{n\to\infty} F_n(x) = F(x)$ is the limiting distribution in this case only.

LECTURE 18 | 2020-11-08

5.2 Convergence in Probability

DEFINITION 5.2.1: Converges in Probability

Let $X_1, \dots, X_n$ be a sequence of random variables such that X_n has c.d.f. $F_n(x)$. Let X be a random variable with c.d.f. $F(x)$. If for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \ge \varepsilon) = 0$$

or

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1$$

then we say X_n **converges in probability** to X, and write $X_n \xrightarrow{\mathbb{P}} X$.

REMARK 5.2.2

(i) Here it's the convergence or limit for a probability, therefore it's called convergence in probability.

(ii) "Meaning" of $X_n \xrightarrow{\mathbb{P}} X$. As $n \to \infty$, X_n cannot be "ε" away from X. That is, X_n becomes very close to X as $n \to \infty$. Because of that, we expect that $F_n(x)$ becomes very close to $F(x)$.

THEOREM 5.2.3: Convergnece in Probability Implies Convergence in Distribution

If $X_n \xrightarrow{\mathbb{P}} X$, then $X_n \xrightarrow{d} X$.

We consider a special case.

DEFINITION 5.2.4: Convergence in Probability to a Constant

Let $X_1, \dots, X_n$ be a sequence of random variables, and b be a constant. If $\lim_{n\to\infty} \mathbb{P}(|X_n - b| \ge \varepsilon) = 0$ for any $\varepsilon > 0$. We say X_n **converges in probability** to b, and write $X_n \xrightarrow{\mathbb{P}} b$.

THEOREM 5.2.5

Let $X_1, \dots, X_n$ be a sequence of random variables such that X_n has c.d.f. $F_n(x)$. If

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x < b \\ 1 & x > b \end{cases}$$

or limiting distribution of X_n is

$$F(x) = \begin{cases} 0 & x < b \\ 1 & x \geq b \end{cases}$$

(c.d.f. of X, which satisfies $\mathbb{P}(X = b) = 1$), then $X_n \xrightarrow{\mathbb{P}} b$ and write $X_n \xrightarrow{d} b$.
In other words, $X_n \xrightarrow{d} b$ implies $X_n \xrightarrow{\mathbb{P}} b$. Therefore,

$$X_n \xrightarrow{d} b \iff X_n \xrightarrow{\mathbb{P}} b$$

Proof of Theorem 5.2.5

For any $\varepsilon > 0$, $\mathbb{P}(|X_n - b| \geq \varepsilon) \to 0$ as $n \to \infty$.

(i) Lower bound: $\mathbb{P}(|X_n - b| \geq \varepsilon) \geq 0$

(ii) Upper bound:

$$\begin{aligned} \mathbb{P}(|X_n - b| \geq \varepsilon) &= \mathbb{P}((X_n \geq b + \varepsilon) \cup (X_n \leq b - \varepsilon)) \\ &= 1 - \mathbb{P}(X_n < b + \varepsilon) + \underbrace{\mathbb{P}(X_n \leq b - \varepsilon)}_{F_n(b-\varepsilon)} \\ &\leq 1 - \mathbb{P}\left(X_n \leq b + \frac{\varepsilon}{2}\right) + F_n(b - \varepsilon) \\ &= 1 - F_n\left(b + \frac{\varepsilon}{2}\right) + F_n(b - \varepsilon) \end{aligned}$$

as $n \to \infty$, $F_n\left(b + \frac{\varepsilon}{2}\right) \geq 1$ and $\lim_{n\to\infty} F_n(b - \varepsilon) = 0$, so the upper bound will be $1 - 1 + 0 = 0$, and hence

$$0 \leq \lim_{n\to\infty} \mathbb{P}(|X_n - b| \geq \varepsilon) \leq 0$$

and hence

$$X_n \xrightarrow{d} b \iff X_n \xrightarrow{\mathbb{P}} b$$

EXAMPLE 5.2.6

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Uniform}(0, 1)$. In Example 5.1.5, we showed that

$$\lim_{n\to\infty} \mathbb{P}(X_{(1)} \leq x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases} \implies X_{(1)} \xrightarrow{d} 0 \implies X_{(1)} \xrightarrow{\mathbb{P}} 0$$

$$\lim_{n\to\infty} \mathbb{P}(X_{(n)} \leq x) = \begin{cases} 0 & x < 1 \\ 1 & x > 1 \end{cases} \implies X_{(n)} \xrightarrow{d} 1 \implies X_{(n)} \xrightarrow{\mathbb{P}} 1$$

EXAMPLE 5.2.7

$X_1, \dots, X_n$ are i.i.d with p.d.f. $f(x) = e^{-(x-\theta)}$, $x > \theta$. Show that $X_{(1)} \xrightarrow{\mathbb{P}} \theta$.

Solution 1. Only need to show that $X_{(1)} \xrightarrow{\mathrm{d}} \theta$; that is,

$$\lim_{n\to\infty} \mathbb{P}(X_{(1)} \le x) = \begin{cases} 0 & x < \theta \\ 1 & x > \theta \end{cases}$$

or limiting distribution of $X_{(1)}$ is

$$F(x) = \begin{cases} 0 & x < \theta \\ 1 & x \ge \theta \end{cases}$$

Support $X_{(1)}$ is (θ, ∞). $\mathbb{P}(X_{(1)} \le x) = F_n(x)$.

- $x \le \theta$, $F_n(x) = 0$
- $x > \theta$,

$$\begin{aligned} \mathbb{P}(X_{(1)} \le x) &= 1 - [\mathbb{P}(X_1 > x)]^n \\ &= 1 - e^{-n(x-\theta)} \end{aligned}$$

since $\mathbb{P}(X_1 > x) = \displaystyle\int_x^\infty e^{-(t-\theta)}\,dt = e^{-(x-\theta)}$. Therefore,

$$F_n(x) = \begin{cases} 0 & x \le \theta \\ 1 - e^{-n(x-\theta)} & x > \theta \end{cases} \implies \lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x < \theta \\ 1 & x > \theta \end{cases}$$

So, $X_{(1)} \xrightarrow{\mathrm{d}} \theta \implies X_{(1)} \xrightarrow{\mathbb{P}} \theta$.

Solution 2. By definition, for any $\varepsilon > 0$,

- Lower bound: $\mathbb{P}(|X_{(1)} - \theta| \ge \varepsilon) \ge 0$
- Upper bound:

$$\begin{aligned} \mathbb{P}(|X_{(1)} - \theta| \ge \varepsilon) &= \mathbb{P}((X_{(1)} \ge \theta + \varepsilon) \cup (X_{(1)} \le \theta - \varepsilon)) \\ &= \mathbb{P}(X_{(1)} \ge \theta + \varepsilon) + \mathbb{P}(X_{(1)} \le \theta - \varepsilon) \\ &= [\mathbb{P}(X_1 > \theta + \varepsilon)]^n \\ &= e^{-n(\theta+\varepsilon-\theta)} \\ &= e^{-n\varepsilon} \to 0 \text{ as } n \to \infty \end{aligned}$$

Therefore, $\mathbb{P}(|X_{(1)} - \theta| \ge \varepsilon) = 0$ as $n \to \infty$ which implies $X_{(1)} \xrightarrow{\mathbb{P}} \theta$ by definition.

Brief Summary:

- Convergence in distribution.
- Convergence in probability.
- Special case. Convergence in probability to a <u>constant</u> if and only if convergence to distribution.
- $X_{(1)} = \min_{1\le i\le n} X_i$ and $X_{(n)} = \max_{1\le i\le n} X_i$.

Next, our main job is to study convergence in distribution and probability $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Sequence of results:

- Convergence in probability of $\bar{X}_n$, WLLN.

- Convergence in distribution of $\sqrt{n}(\bar{X}_n - \mu)$. CLT.
- Combine them together: Slutsky's Theorem, Delta Method.

THEOREM 5.2.8: Markov's Inequality

Suppose that X is a random variable. For any $k > 0$, $c > 0$, we have

$$\mathbb{P}(|X| \geq c) \leq \frac{\mathbb{E}[|X|^k]}{c^k}$$

Markov's Inequality relates probability to moments.

In most situations, we take $k = 2$; that is, we consider

$$\mathbb{P}(|X| \geq c) \leq \frac{\mathbb{E}[X^2]}{c^2}$$

EXAMPLE 5.2.9: Weak Law of Large Numbers

Suppose $X_1, \dots, X_n$ are i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}(X_i) = \sigma^2 < \infty$, then

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mu$$

Solution. By definition, we only need to show that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\big(|\bar{X}_n - \mu| \geq \varepsilon\big) = 0$$

Lower bound: $\mathbb{P}\big(|\bar{X}_n - \mu| \geq \varepsilon\big) \geq 0$.
Upper bound:

$$\mathbb{P}\big(|\bar{X}_n - \mu| \geq \varepsilon\big) \leq \frac{\mathbb{E}[(\bar{X}_n - \mu)^2]}{\varepsilon^2}$$

Aside: $\mathbb{E}[\bar{X}_n] = \mu$, $\mathbb{V}(X_n) = \frac{\sigma^2}{n}$, so

$$\frac{\mathbb{E}[(\bar{X}_n - \mu)^2]}{\varepsilon^2} = \frac{\mathbb{V}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0 \text{ as } n \to \infty$$

By Squeeze Theorem, $\mathbb{P}\big(|\bar{X}_n - \mu| \geq \varepsilon\big) = 0$.

EXERCISE 5.2.10: Markov's Inequality

If $X_1, \dots, X_n$ are independent. $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}(X_i) = \sigma_i^2$ for $i = 1, \dots, n$. $\max_{1\leq i\leq n} \sigma_i^2 \leq c$. Show that

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mu$$

EXAMPLE 5.2.11

If $X_1, \dots, X_n \overset{\text{iid}}{\sim} \chi^2(1)$, then $\bar{X}_n \xrightarrow{\mathbb{P}} 1$.
Solution. $\mathbb{E}[X_1] = \cdots = \mathbb{E}[X_n] = 1$.
From term test 1, $\chi^2(1)$ m.g.f. is $(1-2t)^{-1/2}$ which is Gamma$(\alpha = 1/2, \beta = 2)$, so $\mathbb{V}(\chi^2(1)) = \alpha\beta^2 = (1/2)(2)^2 = 2$. By WLLN, $\bar{X}_n \xrightarrow{\mathbb{P}} \mu = 1$.

EXAMPLE 5.2.12

If $Y_n \sim \chi^2(n)$, then $\dfrac{Y_n}{n} \xrightarrow{\mathbb{P}} 1$.

Solution. We can write $Y_n = \sum_{i=1}^n X_i$ where $X_i \overset{\text{iid}}{\sim} \chi^2(1)$, then

$$\frac{Y_n}{n} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}_n \xrightarrow{\mathbb{P}} 1$$

EXAMPLE 5.2.13

If $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ Poisson(μ), then $\bar{X}_n \xrightarrow{\mathbb{P}} \mu$.

Solution. $\mathbb{E}[X_i] = \mu < \infty$ and $\mathbb{V}(X_i) = \mu < \infty$, so by WLLN, $\bar{X}_n \xrightarrow{\mathbb{P}} \mu$.

EXERCISE 5.2.14

If $Y_n \sim$ Poisson(n), then

$$\frac{Y_n}{n} \xrightarrow{\mathbb{P}} 1$$

Solution. $Y_n = \sum_{i=1}^n X_i$ where $X_i \overset{\text{iid}}{\sim}$ Poisson(1), so by WLLN,

$$\frac{Y_n}{n} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}_n \xrightarrow{\mathbb{P}} 1$$

LECTURE 19 | 2020-11-15

5.3 Some Useful Limit Theorems

In this section, we'll discuss some theorems regarding convergence in distribution of $\bar{X}_n$ or function of $\bar{X}_n$.

THEOREM 5.3.1: Central Limit Theorem (CLT)

Let $X_1, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}(X_i) = \sigma^2 < \infty$. 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 the c.d.f. of $\mathcal{N}(0, 1)$.

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$$

The proof is not hard, we use a standard method, but we need to put several pieces together. We need the following theorem.

THEOREM 5.3.2

Let $X_1, \dots, X_n$ be a sequence of random variables such that X_n has m.g.f. $M_n(t)$. Let X be another random variable with m.g.f. $M(t)$. If there exists some $h > 0$, such that

$$\lim_{n\to\infty} M_n(t) = M(t)$$

for all $t \in (-h, h)$, then

$$X_n \xrightarrow{d} X$$

Therefore, our next steps:

(1) Find the m.g.f. of $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$, denoted by $M_n(t)$.

Find the m.g.f. of $\mathcal{N}(0,1)$, denoted by $M(t) = \exp\left\{\dfrac{t^2}{2}\right\}$.

(2) We try to show that for $t \in (-h, h)$ where $h > 0$ that

$$\lim_{n\to\infty} M_n(t) = \exp\left\{\frac{t^2}{2}\right\}$$

Step 1: Find m.g.f. of $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$.

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n (X_i - \mu)}{\sigma} = \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n (X_i - \mu)}{\sigma}$$

Let $Y_i = \dfrac{X_i - \mu}{\sigma}$, then $Y_1, \dots, Y_n$ are i.i.d. with $\mathbb{E}[Y_i] = 0$ and $\mathbb{V}(Y_i) = 1$. Then,

$$\begin{aligned}
M_n(t) &= \mathbb{E}\left[\exp\left\{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{\frac{t}{\sqrt{n}}\sum_{i=1}^n Y_i\right\}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^n \exp\left\{\frac{t}{\sqrt{n}} Y_i\right\}\right] \\
&= \prod_{i=1}^n \mathbb{E}\left[\exp\left\{\frac{t}{\sqrt{n}} Y_i\right\}\right]
\end{aligned}$$

Suppose Y_i has m.g.f. $M_Y(t)$, then

- $M_Y(0) = 1$
- $M_Y'(0) = 0$
- $M_Y''(0) = \mathbb{E}[Y^2] = \mathbb{V}(Y) + (\mathbb{E}[Y])^2 = 1$

$$M_n(t) = \left[M_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

since $\mathbb{E}\left[\exp\left\{\dfrac{t}{\sqrt{n}} Y_i\right\}\right] = M_Y\left(\dfrac{t}{\sqrt{n}}\right)$.

Step 2: We want to show that

$$\lim_{n\to\infty}\left[M_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n = \exp\left\{\frac{t^2}{2}\right\}$$

$$\begin{aligned} M_Y\left(\frac{t}{\sqrt{n}}\right) &= M_Y(0) + M_Y'(0)\left(\frac{t}{\sqrt{n}}\right) + \frac{M_Y''(0)}{2!}\left(\frac{t}{\sqrt{n}}\right)^2 + o\left[\left(\frac{t}{\sqrt{n}}\right)^2\right] \\ &= 1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \end{aligned}$$

Then,

$$\lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n = \lim_{n\to\infty}\left(1 + \frac{t^2}{2n}\right)^n = \exp\left\{\frac{t^2}{2}\right\}$$

EXAMPLE 5.3.3

Suppose that $X_1, \dots, X_n \overset{\text{iid}}{\sim} \chi^2(1)$ and $Y_n = \sum_{i=1}^n X_i$. Show that

$$\frac{Y_n - n}{\sqrt{2n}} \xrightarrow{\text{d}} \mathcal{N}(0,1)$$

Solution. $\mathbb{E}[X_i] = 1 = \mu$ and $\mathbb{V}(X_i) = 2 = \sigma^2 < \infty$. CLT tells us

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sqrt{n}(\bar{X}_n - 1)}{\sqrt{2}} \xrightarrow{\text{d}} Z \sim \mathcal{N}(0,1)$$

$$\frac{Y_n - n}{\sqrt{2n}} = \frac{\sum_{i=1}^n X_i - n}{\sqrt{2n}} = \frac{n(\bar{X}_n - 1)}{\sqrt{2n}} = \frac{\sqrt{n}(\bar{X}_n - 1)}{\sqrt{2}} \xrightarrow{\text{d}} Z \sim \mathcal{N}(0,1)$$

Suppose that $Y_n \sim \chi^2(n)$, we might ask you to prove

$$\frac{Y_n - n}{\sqrt{2n}} \xrightarrow{\text{d}} Z \sim \mathcal{N}(0,1)$$

and you might have to figure out $Y_n = \sum_{i=1}^n X_i$ where $X_1, \dots, X_n \overset{\text{iid}}{\sim} \chi^2(1)$.

EXAMPLE 5.3.4

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\mu)$. Let $Y_n = \sum_{i=1}^n X_i$. Find the limiting distribution of $\dfrac{Y_n - n\mu}{\sqrt{n\mu}}$.

Solution. CLT tells us that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}} \xrightarrow{\text{d}} Z \sim \mathcal{N}(0,1)$$

Now,

$$\frac{Y_n - n\mu}{\sqrt{n\mu}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\mu}} = \frac{n(\bar{X}_n - \mu)}{\sqrt{n\mu}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}} \xrightarrow{\text{d}} Z \sim \mathcal{N}(0,1)$$

Alternatively: If $Y_n \sim \text{Poisson}(n\mu)$, what is the limiting distribution of $\dfrac{Y_n - n\mu}{\sqrt{n\mu}}$?

THEOREM 5.3.5: Continuous Mapping Theorem

Suppose that $g(\cdot)$ is a continuous function.

(1) *If $X_n \xrightarrow{\mathbb{P}} a$, then $g(X_n) \xrightarrow{\mathbb{P}} g(a)$.*

(2) *If $X_n \xrightarrow{\text{d}} X$, then $g(X_n) \xrightarrow{\text{d}} g(X)$.*

THEOREM 5.3.6: Slutsky's Theorem

If $X_n \xrightarrow{d} X$, and $Y_n \xrightarrow{\mathbb{P}} b$, then

(a) *$X_n + Y_n \xrightarrow{d} X + b$. If we replace b by Y it is still true.*

(b) *$X_n Y_n \xrightarrow{d} bX$.*

(c) *$\dfrac{X_n}{Y_n} \xrightarrow{d} \dfrac{X}{b}$ for $b \neq 0$.*

EXERCISE 5.3.7

Find a counter-example to the following statement. If $X_n \xrightarrow{\text{d}} X$ and $Y_n \xrightarrow{\text{d}} Y$, then

$$X_n + Y_n \xrightarrow{\text{d}} X + Y$$

EXAMPLE 5.3.8

(i) If $X_n \geq 0$, $a \geq 0$, then $\sqrt{X_n} \xrightarrow{\mathbb{P}} \sqrt{a}$.

(ii) If $X_n \xrightarrow{\mathbb{P}} a$, then $X_n^2 \xrightarrow{\mathbb{P}} a^2$.

(iii) If $X_n \xrightarrow{\text{d}} X \sim \mathcal{N}(0,1)$, then

- $2X_n \xrightarrow{\text{d}} 2X \sim \mathcal{N}(0,4)$.
- $2X_n + 1 \xrightarrow{\text{d}} 2X + 1 \sim \mathcal{N}(1,4)$.
- $X_n^2 \xrightarrow{\text{d}} X^2 \sim \chi^2(1)$.

(iv) If $X_n \xrightarrow{\text{d}} X \sim \mathcal{N}(0,1)$ and $Y_n \xrightarrow{\mathbb{P}} b$ for $b \neq 0$, then

- $X_n + Y_n \xrightarrow{\text{d}} X + b \sim \mathcal{N}(b,1)$.
- $X_n Y_n \xrightarrow{\text{d}} bX \sim \mathcal{N}(0,b^2)$.
- $\dfrac{X_n}{Yn} \xrightarrow{\text{d}} \dfrac{X}{b} \sim \mathcal{N}\left(0, \dfrac{1}{b^2}\right)$.

EXAMPLE 5.3.9

Suppose that $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\mu)$. Find the limiting distribution of

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\bar{X}_n}}$$

$$U_n = \sqrt{n}(\bar{X}_n - \mu)$$

Solution. For Z_n,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}} \xrightarrow{\text{d}} Z \sim \mathcal{N}(0,1)$$

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\bar{X}_n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}} \cdot \frac{\sqrt{\mu}}{\sqrt{\bar{X}_n}}$$

By continuous mapping theorem,

$$\frac{\sqrt{\mu}}{\sqrt{\bar{X}_n}} \xrightarrow{\mathbb{P}} \frac{\sqrt{\mu}}{\sqrt{\mu}} = 1 \text{ since } \bar{X}_n \xrightarrow{\mathbb{P}} \mu \text{ by WLLN}$$

For U_n, by Slutsky's theorem, $Z_n \xrightarrow{\text{d}} Z \sim \mathcal{N}(0,1)$.

$$U_n = \sqrt{n}(\bar{X}_n - \mu) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}}\sqrt{\mu} \xrightarrow{\text{d}} \sqrt{\mu}Z \sim \mathcal{N}(0,\mu) \text{ by continuous mapping theorem}$$

Lecture 20 | 2020-11-15

EXAMPLE 5.3.10

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Uniform}(0,1)$ and $U_n = \max_{1\le i\le n} X_i$. In the first two examples of this chapter, we've shown that

$$U_n = X_{(n)} \xrightarrow{\mathbb{P}} 1$$

$$U_n \xrightarrow{\text{d}} 1$$

and

$$n(1 - X_{(n)}) = n(1 - U_n) \xrightarrow{\text{d}} X \sim \text{Exponential}(1)$$

Then,

(i) e^{U_n}
(ii) $\sin(1 - U_n)$
(iii) $e^{-n(1-U_n)}$
(iv) $(U_n + 1)^2[n(1 - U_n)]$

Solution.

(i) e^{U_n} Take $g(x) = e^x$. Continuous mapping theorem:

$$U_n \xrightarrow{\mathbb{P}} 1 \implies e^{U_n} \xrightarrow{\mathbb{P}} e^1$$

(ii) $\sin(1 - U_n)$. Take $g(x) = \sin(1 - x)$.

$$\sin(1 - U_n) \xrightarrow{\mathbb{P}} \sin(1 - 1) = 0$$

(iii) $e^{-n(1-U_n)}$.

$$n(1 - U_n) \xrightarrow{\text{d}} X \sim \text{Exponential}(1)$$

Continuous mapping theorem. Take $g(x) = e^{-x}$,

$$e^{-n(1-U_n)} \xrightarrow{\text{d}} e^{-X} \quad X \sim \text{Exponential}(1)$$

How to find c.d.f. of e^{-X}? Let $Y = e^{-X}$. Support of Y is $(0, 1)$. For any $0 < y < 1$,

$$\begin{aligned}
F_y(y) &= \mathbb{P}(e^{-X} \leq y) \\
&= \mathbb{P}(-X \leq \ln(y)) \\
&= \mathbb{P}(X \geq -\ln(y)) \\
&= \int_{-\ln(y)}^{\infty} e^{-x}\, dx \\
&= y
\end{aligned}$$

Therefore,

$$e^{-n(1-U_n)} \xrightarrow{\text{d}} Y \sim \text{Exponential}(1)$$

(iv) $(U_n + 1)^2[n(1 - U_n)]$. Since $U_n \xrightarrow{\mathbb{P}} 1$, Take $g(x) = (1 + x)^2$. Continuous mapping theorem:

$$(U_n + 1)^2 \xrightarrow{\mathbb{P}} (1 + 1)^2 = 4$$

$$n(1 - U_n) \xrightarrow{\text{d}} X \sim \text{Exponential}(1)$$

Slutsky's Theorem:

$$(U_n + 1)^2[n(1 - U_n)] \xrightarrow{\text{d}} 4X$$

Let $Y = 4X$. Support of Y is $(0, \infty)$. For $0 < y < \infty$,

$$\begin{aligned}
\mathbb{P}(Y \leq y) &= \mathbb{P}(4X \leq y) \\
&= \mathbb{P}\Big(X \leq \frac{y}{4}\Big) \\
&= \int_0^{y/4} e^{-x}\, dx \\
&= 1 - e^{-y/4}
\end{aligned}$$

Hence, the p.d.f. of Y is

$$f_Y(y) = \frac{1}{4} e^{-y/4} \quad (y > 0)$$

$Y \sim \text{Exponential}(4)$.

THEOREM 5.3.11: Delta Method

Let $X_1, \ldots, X_n$ be a sequence of random variables such that

$$\sqrt{n}(X_n - \theta) \xrightarrow{\text{d}} X \sim \mathcal{N}(0, \sigma^2)$$

and $g(x)$ is differentiable at $x = \theta$ and $g'(\theta) \neq 0$. Then,

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{\text{d}} W \sim \mathcal{N}(0, [g'(\theta)]^2 \sigma^2)$$

Background: $\sqrt{n}(X_n - \theta)\sqrt{n}(X_n - \theta) \xrightarrow{\text{d}} X \sim \mathcal{N}(0, \sigma^2)$. This implies that

$$\sqrt{n}(X_n - \theta) \overset{\text{d}}{\approx} \mathcal{N}(0, \sigma^2)$$

equivalently,

$$X_n \overset{\mathrm{d}}{\approx} \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$$

Question: What's the approximate distribution of $g(X_n)$? Delta method tells us that

$$\sqrt{n}[g(X_n) - g(\theta)] \approx \mathcal{N}(0, [g'(\theta)]^2\sigma^2)$$

$$\implies g(X_n) \overset{\mathrm{d}}{\approx} \mathcal{N}\left(g(\theta), \frac{[g'(\theta)]\sigma^2}{n}\right)$$

Not rigorous derivation. By 1st order Taylor expansion:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) \quad (x \approx x_0)$$

$$g(X_n) \approx g(\theta) + g'(\theta)(X_n - \theta) \implies \sqrt{n}[g(X_n) - g(\theta)] \approx \underbrace{\sqrt{n}(X_n - \theta)}_{\mathcal{N}(0,\sigma^2)} g'(\theta)$$

By continuous mapping theorem,

$$\sqrt{n}(X_n - \theta)g'(\theta) \xrightarrow{\mathrm{d}} g'(\theta)X \sim \mathcal{N}(0, [g'(\theta)]^2\sigma^2)$$

Not rigorous since we only considered the 1st Taylor expansion, "why can we drop other terms?"

EXAMPLE 5.3.12

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\mu)$. Find limiting distribution of

$$Z_n = \sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\mu})$$

Recall in Example 5.3.9:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mu)$$

since $\mathbb{E}[X_i] = \mu$, $\mathbb{V}(X_i) = \mu$. Take $g(x) = \sqrt{x}$, $g'(x) = \frac{1}{2}x^{-1/2}$.

$$Z_n = \sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\mu}) \xrightarrow{\mathrm{d}} \mathcal{N}(0, [g'(\mu)]\sigma^2) = \mathcal{N}\left(0, \frac{1}{4}\mu^{-1}\mu = \frac{1}{4}\right) = \mathcal{N}\left(0, \frac{1}{4}\right)$$

EXAMPLE 5.3.13

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Exponential}(\theta)$. 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. $V_n = \sqrt{n}(\ln(\bar{X}_n) - \ln(\theta))$

Solution.

1. $\bar{X}_n$. By WLLN, $\mathbb{E}[X_i] = \theta$, $\mathbb{V}(X_i) = \theta^2$ (also available on cheat sheet), so $\bar{X}_n \xrightarrow{\mathbb{P}} \theta$.
2. $Z_n = \dfrac{\sqrt{n}(\bar{X}_n - \theta)}{\bar{X}_n}$.

$$\frac{\sqrt{n}(\bar{X}_n - \theta)}{\theta} \xrightarrow{\mathrm{d}} \mathcal{N}(0, 1) \quad \text{CLT}$$

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \theta)}{\bar{X}_n} = \frac{\sqrt{n}(\bar{X}_n - \theta)}{\theta} \frac{\theta}{\bar{X}_n}$$

by continuous mapping theorem, take $g(x) = \frac{\theta}{x}$,

$$\frac{\theta}{\bar{X}_n} \xrightarrow{\mathbb{P}} 1$$

By Slutsky's Theorem,

$$Z_n \xrightarrow{\mathrm{d}} Z \sim \mathcal{N}(0,1)(1)$$

3. $U_n = \sqrt{n}(\bar{X}_n - \theta)$.

$$U_n = \frac{\sqrt{n}(\bar{X}_n - \theta)}{\theta}(\theta) \xrightarrow{\mathrm{d}} Z \sim \mathcal{N}(0,1)$$

   $g(x) = \theta x$, continuous mapping theorem

$$U_n \xrightarrow{\mathrm{d}} \theta Z \sim \mathcal{N}(0,\theta^2)$$

4. $V_n = \sqrt{n}(\ln(\bar{X}_n) - \ln(\theta))$. $g(x) = \ln(x)$. $g'(x) = 1/x$. By Delta Method,

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\mathrm{d}} \mathcal{N}(0,\sigma^2)$$

   Delta method,

$$\sqrt{n}(\ln(\bar{X}_n)\ln(\theta)) \xrightarrow{\mathrm{d}} \mathcal{N}(0,[g'(\sigma)]^2\theta^2) = \mathcal{N}(0,1)$$

Chapter 6

Point Estimation

Lecture 21 | 2020-11-22

6.1 Introduction

Background: Suppose $X_1, \ldots, X_n$ are i.i.d. random variables from $f(x; \theta)$. Here, θ is unknown, but fixed. It can be a scalar or a vector since

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$$

Clearly, if $k = 1$, θ is a scalar, and if $k > 1$, θ is a vector.

Purpose: Given $X_1, \ldots, X_n$ we'd like to estimate θ.

EXAMPLE 6.1.1

- If $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$, then $\theta = \mu$ is a scalar.
- If $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, then $\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$ is a vector.

Notation:

- Θ: parameter space, it contains all possible values of θ.

  EXAMPLE 6.1.2

  - If $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$, then
  $$\Theta = \{\mu : -\infty < \mu < \infty\}$$
  - If $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, then
  $$\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$$

- Data: $(X_1, \ldots, X_n)$ data are random variables.
- Observed data/observations: $(x_1, \ldots, x_n)$, they're observed values of $(X_1, \ldots, X_n)$. Note that $x_1, \ldots, x_n$ are not random variables.
- Statistic: function of data, does not depend on any unknown parameter. Denoted by $T = T(X_1, \ldots, X_n)$.

EXAMPLE 6.1.3

– If $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$, then
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$
  is a statistic.

– $\sqrt{n}(\bar{X}_n - \mu)$ is <u>not</u> a statistic, since it depends on $\theta = \mu$.

- Estimator & estimate
  1. If a statistic $T = T(X_1, \dots, X_n)$ is used to estimate θ, then $T = T(X_1, \dots, X_n)$ is an estimator (which must be a statistic, and also a random variable) of θ.
  2. The observed value of T, denote it by $t = T(x_1, \dots, x_n)$ is called an estimate (which is an observed value, therefore not a random variable) of θ.

EXAMPLE 6.1.4

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$ with observed data is $(x_1, \dots, x_n)$.

– $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ is an estimator.

– $\bar{x}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ is an estimate.

REMARK 6.1.5

We prefer using
- $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ to denote an estimator of θ.
- (Slight abuse of notation) $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ to denote an estimate of θ. That is: $\hat{\theta}$ is used for both estimator and estimate.
  – If $\hat{\theta}$ is a random variable, then $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ is regarded as an estimator.
  – If $\hat{\theta}$ is an observed value, then $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ is regarded as an estimate.

6.2 Method of Moments

Problem setup: Suppose $X_1, \dots, X_n$ are i.i.d. with p.f. $f(x;\theta)$ or p.d.f. $f(x;\theta)$. We need to estimate $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^\top$.

Method: Method of moments estimator (MM estimator).

1. Population moment. Let $\mu_j = \mathbb{E}[X_i^j] = \mathbb{E}[X^j]$ for $j = 1, \dots, k$.
   - μ_j is called j^{th} population moment.
   - μ_j is a function of $\boldsymbol{\theta}$, and we write it as $\mu_j(\boldsymbol{\theta})$.
2. Sample moments. Let $\hat{\mu}_j = \sum_{i=1}^{n} X_i^j$ for $j = 1, \dots, k$.
   - $\hat{\mu}_j$: j^{th} sample moment.
   - $\mathbb{E}[\hat{\mu}_j] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i^j] = \mu_j$.
3. Idea of method of moments. Choose estimators $\hat{\boldsymbol{\theta}}$ such that $\mu_j(\hat{\boldsymbol{\theta}}) = \hat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$ for $j = 1, \dots, k$. There are k unknown parameters and k equations.

The estimator $\hat{\boldsymbol{\theta}}$ is called the method of moment estimator of $\boldsymbol{\theta}$.

EXAMPLE 6.2.1

$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$

- First population moment is $\mathbb{E}[X_1] = \mu_1 = \theta \to \mu_1(\theta) = \theta$
- First sample moment is $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^n X_i$
- MM estimator satisfies $\mu_1(\hat{\theta}) = \hat{\mu}_1$ and $\hat{\theta} = \hat{\mu}_1 = \frac{1}{n} X_i$ which is MM estimator of θ.

Lecture 22 | 2020-11-22

EXAMPLE 6.2.2

$X_1, \ldots, X_n$ are i.i.d.

1. Exponential(θ)
2. Uniform$(0, \theta)$
3. $f(x; \theta) = \theta x^{\theta-1}$ with $0 < x < 1$ and $\theta > 0$

Solution.

1. Exponential(θ). $\mu_1 = \mathbb{E}[X_1] = \theta$. $\mu_1(\theta) = \theta$

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^n X_i$$

$\mu_1(\hat{\theta}) = \hat{\mu}_1$. Since μ_1 is the identity map,

$$\hat{\theta} = \hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^n X_i$$

2. Uniform$(0, \theta)$.

$$\mu_1 = \mathbb{E}[X_1] = \int_0^\theta x\left(\frac{1}{\theta}\right) dx = \theta/2$$

$\mu_1(\theta) = \frac{\theta}{2}$

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^n X_i$$

$$\mu_1(\hat{\theta}) = \frac{\hat{\theta}}{2} = \hat{\mu}_1$$

Therefore,

$$\hat{\theta}_{\text{MM}} = 2\hat{\mu}_1 = \frac{2}{n}\sum_{i=1}^n X_i$$

3. $f(x; \theta) = \theta x^{\theta-1}$ with $0 < x < 1$ and $\theta > 0$

$$\mu_1 = \mathbb{E}[X_1] = \int_0^1 x\theta x^{\theta-1}\, dx = \frac{\theta}{1+\theta}$$

$\mu_1(\theta) = \dfrac{\theta}{1+\theta}$

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^n X_i$$

$$\mu_1(\hat{\theta}) = \frac{\hat{\theta}}{1+\hat{\theta}} = \hat{\mu}_1$$

Therefore,

$$\hat{\theta}_{\text{MM}} = \frac{\hat{\mu}_1}{1-\hat{\mu}_1} = \frac{\bar{X}}{1-\bar{X}}$$

4. $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$.

$$\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$$

$$\begin{aligned} \mu_1 &= \mathbb{E}[X_1] = \mu \\ \mu_2 &= \mathbb{E}[X_1^2] = \mathbb{V}(X) + [\mathbb{E}[X_1]]^2 = \mu^2 + \sigma^2 \end{aligned}$$

$$\begin{aligned} \mu_1(\mu, \sigma^2) &= \mu \\ \mu_2(\mu, \sigma^2) &= \mu^2 + \sigma^2 \end{aligned}$$

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^n X_i$$

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^n X_i^2$$

$$\mu_1(\hat{\mu}, \hat{\sigma}^2) = \hat{\mu} = \hat{\mu}_1 = \bar{X}$$

$$\mu_2(\hat{\mu}, \hat{\sigma}^2) = (\hat{\mu})^2 + \hat{\sigma}^2 = \hat{\mu}_2$$

Therefore,

$$\begin{aligned} \hat{\mu}_{\text{MM}} &= \bar{X}_n \\ \hat{\sigma}^2_{\text{MM}} &= \hat{\mu}_2 - (\bar{X}_n)^2 \\ &= \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n \\ &= \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 \end{aligned}$$

6.3 Maximum Likelihood Method

This section: introduce the most commonly used method for estimating unknown parameter θ referred to as maximum likelihood method.

- Likelihood function
  1. Suppose $X_1, \dots, X_n$ are i.i.d. from $f(x; \theta)$
  2. Given $(x_1, \dots, x_n)$, the observed value of $(X_1, \dots, X_n)$. We calculate the joint p.f. of $(X_1, \dots, X_n)$ at observed data $(x_1, \dots, x_n)$ or joint p.d.f. of $(X_1, \dots, X_n)$ at observed data $(x_1, \dots, x_n)$.

     Discrete random variables joint p.d.f. of $(X_1, \dots, X_n)$ at $(x_1, \dots, x_n)$:

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i) = \prod_{i=1}^n f(x_i; \theta)$$

Continuous random variables joint p.d.f. of $(X_1, \dots, X_n)$ at $(x_1, \dots, x_n)$:

$$f_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)$$

3. We use $L(\theta; x_1, \dots, x_n)$ or simply $L(\theta)$ to denote it. That is to say,

$$L(\theta; x_1, \dots, x_n) = \begin{cases} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) & \text{discrete} \\ f_{X_1,\dots,X_n}(x_1, \dots, x_n) & \text{continuous} \end{cases} = \prod_{i=1}^{n} f(x_i; \theta)$$

Here, $L(\theta; x_1, \dots, x_n)$ is called the likelihood function of θ.

Comments:

1. Likelihood function measures how likely we get the observed data for a given θ.
2. Smaller $L(\theta)$ means θ is less likely to generate the observed data.
3. Larger $L(\theta)$ means θ is more likely to generate the observed data.

Idea of Maximum Likelihood Method

Choose θ to maximize $L(\theta)$ or choose θ such that it most likely generates the observed data.

Maximum likelihood estimator/estimate (MLE)

1. ML estimate maximizes $L(\theta)$, and we use $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ to denote it.

$$\hat{\theta} = \hat{\theta}(x_1, \dots, x_n) = \arg\max_{\theta \in \Theta} L(\theta)$$

2. ML estimator: $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$
3. Log-likelihood function: log of likelihood function:

$$\ell(\theta) = \ln[L(\theta)]$$

Then: an immediate result is:

$$\hat{\theta} = \hat{\theta}(x_1, \dots, x_n) = \arg\max_{\theta \in \Theta} \ell(\theta) \arg\max_{\theta \in \Theta} L(\theta)$$

4. Invariance principal of ML estimator $\tau(\theta)$ is a function of θ. $\tau(\hat{\theta})$ is the ML estimator of $\tau(\theta)$ if $\hat{\theta}$ is the ML estimator of θ.

EXAMPLE 6.3.1

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$. Find ML estimator of θ.

Solution.

$$f(x; \theta) = \frac{\theta^x}{x!} e^{-\theta}$$

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta} = \frac{\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} (x_i!)} e^{-n\theta}$$

$$\ell(\theta) = \left(\sum_{i=1}^{n} x_i \right) \ln(\theta) - n\theta - \sum_{i=1}^{n} \ln(x_i!)$$

$$\frac{d\ell(\theta)}{d\theta} = \frac{\sum_{i=1}^n x_i}{\theta} - n$$

ML estimator of θ satisfies

$$\left[\frac{d\ell}{d\theta}\right]_{\theta=\hat{\theta}} = 0 \implies \frac{\sum_{i=1}^n x_i}{\hat{\theta}} - n = 0 \implies \hat{\theta} = \frac{\sum_{i=1}^n x_i}{n}$$

ML estimator of θ is

$$\hat{\theta} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{(same as the MM estimator)}$$

REMARK 6.3.2

- ML estimator of θ^2 is $(\hat{\theta})^2$
- ML estimator of $e^{-\theta}$ is $e^{-\hat{\theta}}$

EXAMPLE 6.3.3

$X_1, \dots, X_n$ are i.i.d. from $f(x;\theta) = \theta x^{\theta-1}$ with $0 < x < 1$, $\theta > 0$. Find ML estimator of θ.
Solution.

$$L(\theta) = \prod_{i=1}^n f(x_i;\theta) = \prod_{i=1}^n \theta x_i^{\theta-1} = \theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1}$$

$$\ell(\theta) = n\ln(\theta) + (\theta-1)\sum_{i=1}^n \ln(x_i)$$

$$\frac{d\ell(\theta)}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^n \ln(x_i)$$

ML estimate $\hat{\theta}$ satisfies

$$\left[\frac{d\ell}{d\theta}\right]_{\theta=\hat{\theta}} = 0 \implies \frac{n}{\hat{\theta}} + \sum_{i=1}^n \ln(x_i) = 0 \implies \hat{\theta} = -\frac{n}{\sum_{i=1}^n \ln(x_i)}$$

ML estimator:

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^n \ln(X_i)} \quad \text{(is different from MM estimator)}$$

LECTURE 23 | 2020-11-29

EXAMPLE 6.3.4

Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Find ML estimator of θ.
Solution.

$$L(\mu,\sigma^2) = \prod_{i=1}^n f(x_i;\mu,\sigma^2)$$

$$
\begin{aligned}
\ell(\mu,\sigma^2) &= \sum_{i=1}^{n} \ln[f(x_i;\mu,\sigma^2)] \\
&= \sum_{i=1}^{n} \ln\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}\right] \\
&= \sum_{i=1}^{n} \left[-\frac{(x_i-\mu)^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \ell}{\partial \mu} &= \frac{\sum_{i=1}^{n}(\mu - x_i)}{\sigma^2} \\
\frac{\partial \ell}{\partial \sigma^2} &= \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} - \left(\frac{n}{2}\right)\left(\frac{1}{\sigma^2}\right)
\end{aligned}
$$

ML estimate satisfies

$$
\begin{aligned}
\frac{\sum_{i=1}^{n}(X_i-\hat{\mu})^2}{\hat{\sigma}^2} &= 0 \\
\frac{\sum_{i=1}^{n}(X_i-\hat{\mu})}{2(\hat{\sigma}^2)^2} - \left(\frac{n}{2}\right)\left(\frac{1}{\sigma^2}\right) &= 0
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\mu})^2
\end{aligned}
$$

ML estimator of (μ,σ^2) is

$$
\begin{aligned}
\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n \\
\hat{\sigma} &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2
\end{aligned}
$$

which is the same as MM estimator.

EXAMPLE 6.3.5

$$
f(x,\theta) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}
$$

Note that the support of X depends on θ. Find ML estimator of θ.
Solution.

$$
L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} \left(\frac{1}{\theta}\right)^n & 0 \le x_1,\dots,x_n \le \theta \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{1}{\theta^n} & 0 \le x_{(1)}, x_{(n)} \le \theta \\ 0 & \text{otherwise} \end{cases}
$$

- $\theta < x_{(n)}$, $L(\theta) = 0$
- $\theta \ge x_{(n)}$, $L(\theta)$ is a strictly monotone decreasing function of θ.

This implies that the ML estimate of θ is

$$x_{(n)} = \max(x_1, \dots, x_n)$$

ML estimator of θ is

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

is different from the MM estimator $\hat{\theta}_{\text{MM}} = 2\bar{X}_n$.
Which estimator is better? $\hat{\theta}_{\text{MM}}$ or $\hat{\theta}_{\text{ML}}$? STAT 450 covers this.

- Biased or unbiased estimator. Let $\hat{\theta}$ denote one estimator of θ. If $\mathbb{E}[\hat{\theta}] = \theta$, then $\hat{\theta}$ is an unbiased estimator of θ. Otherwise, $\hat{\theta}$ is a biased estimator of θ.

6.4 Properties of ML Estimator

In this section:

1. We only consider the case that the support of $X_1, \dots, X_n$ does not depend on θ.
2. We talk about random variables, only concerned about ML estimator.
3. We only consider θ is 1-dimensional or θ is a scalar.

We define some notation first.

DEFINITION 6.4.1: Score Function

The **score function** is defined as

$$S(\theta) = S(\theta; \boldsymbol{x}) = \frac{d}{d\theta}\ell(\theta) = \frac{d}{d\theta} \ln[L(\theta)]$$

where $\boldsymbol{x}$ are the observed data. When the support of $X_1, \dots, X_n$ does not depend on θ, then $S(\hat{\theta}) = 0$.

DEFINITION 6.4.2: Information Function

The **information function** is defined as

$$I(\theta) = I(\theta; \boldsymbol{x}) = -\frac{d^2}{d\theta^2}\ell(\theta) = -\frac{d^2}{d\theta^2} \ln[L(\theta)]$$

where $\boldsymbol{x}$ are the observed data. $I(\hat{\theta})$ is called the **observed information**.

DEFINITION 6.4.3: Fisher Information/Expected Information

The **fisher information (expected information)** is defined as

$$J(\theta) = \mathbb{E}[I(\theta; \boldsymbol{X})] = -\mathbb{E}\left[\frac{d^2}{d\theta^2}\ell(\theta; \boldsymbol{X})\right]$$

where $\boldsymbol{X}$ is the potential data.
In particular, when $\boldsymbol{X} = (X_1, \dots, X_n)$ is i.i.d. from $f(x, \theta)$, then

$$\ell(\theta; \boldsymbol{x}) = \sum_{i=1}^{n} \ln[f(x_i; \theta)]$$

$$I(\theta; \boldsymbol{X}) = -\frac{d^2}{d\theta^2}\sum_{i=1}^{n} \ln[f(X_i; \theta)] = -\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\ln[f(X_i; \theta)]$$

Therefore,

$$J(\theta) = \mathbb{E}\left[-\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\ln[f(X_i; \theta)]\right] = -\mathbb{E}\left[\frac{d^2}{d\theta^2}\ln[f(X_1; \theta)]\right]$$

DEFINITION 6.4.4: Fisher Information of One Observation

The **fisher information of one observation** is

$$J_1(\theta) = -n\mathbb{E}\left[\frac{d^2}{d\theta^2}\ln[f(X_1; \theta)]\right]$$

The **fisher information in n observations** is

$$J(\theta) = nJ_1(\theta)$$

EXAMPLE 6.4.5

Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$.

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \theta)$$

$$\ell(\theta; \boldsymbol{x}) = \sum_{i=1}^{n} \ln[f(x_i; \theta)] = \sum_{i=1}^{n} \ln\left[\frac{\theta^{x_i}e^{-\theta}}{x!}\right] = \left(\sum_{i=1}^{n} x_i\right)\ln(\theta) - n\ln(\theta) - \sum_{i=1}^{n}\ln(x_i!)$$

Score function:

$$S(\theta; \boldsymbol{x}) = \frac{\partial}{\partial\theta}\ell(\theta; \boldsymbol{x}) = \frac{\sum_{i=1}^{n} x_i}{\theta} - n$$

Observed information function:

$$I(\theta; \boldsymbol{x}) = -\frac{\partial S}{\partial\theta}S(\theta; \boldsymbol{x}) = \frac{\sum_{i=1}^{n} x_i}{\theta^2}$$

Fisher information:

$$J(\theta) = \mathbb{E}[I(\theta; \boldsymbol{X})] = \mathbb{E}\left[\frac{\sum_{i=1}^{n} X_i}{\theta^2}\right] = \frac{n\mathbb{E}[X_1]}{\theta^2} = \frac{n\theta}{\theta^2} = \frac{n}{\theta}$$

Recall that: $\hat{\theta}_{\text{ML}} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

$$\implies \mathbb{V}(\hat{\theta}_{\text{ML}}) = \frac{\mathbb{V}(X_i)}{n} = \frac{\theta}{n}$$

Is there any relationship between $J(\theta)$ and $\mathbb{V}(\hat{\theta}_{\text{ML}})$?

THEOREM 6.4.6: Cramér–Rao Bound

The variance of any unbiased estimator $\hat{\theta}$ of θ is bounded by the reciprocal of the Fisher information $J(\theta)$:

$$\mathbb{V}(\theta) \geq \frac{1}{J(\theta)}$$

COROLLARY 6.4.7

If T is an unbiased estimator of $g(\theta)$, then

$$\mathbb{V}(T) \geq \frac{[g'(\theta)]^2}{J(\theta)}$$

THEOREM 6.4.8

ML estimator satisfies (when support of $X_1, \dots, X_n$ does not depend on θ)

(1) $\hat{\theta} \xrightarrow{\mathbb{P}} \theta$ as $n \to \infty$.

(2) $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \dfrac{1}{J_1(\theta)}\right)$

(3) By delta-method, $\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} \mathcal{N}\left(0, \dfrac{[g'(\theta)]^2}{J_1(\theta)}\right)$

REMARK 6.4.9

(1) Tells us that $\hat{\theta}$ is close to θ as $n \to \infty$.

(2) Tells us that $\sqrt{n}(\hat{\theta} - \theta) \approx \mathcal{N}\left(0, \dfrac{1}{J_1(\theta)}\right) \implies \hat{\theta} \approx \mathcal{N}\left(\theta, \dfrac{1}{nJ_1(\theta)}\right) = \mathcal{N}\left(\theta, \dfrac{1}{J(\theta)}\right)$

$$\mathbb{V}(\hat{\theta}) \approx \frac{1}{J(\theta)}$$

which is the CR lower-bound. $\mathbb{E}[\hat{\theta}] \approx \theta$.

- $\hat{\theta}$ is asymptotically unbiased.
- $\hat{\theta}$ is asymptotically efficient.

(3) Tells us that $g(\hat{\theta}) \approx \mathcal{N}\left(g(\theta), \dfrac{[g'(\theta)]^2}{J(\theta)}\right)$.

- $g(\hat{\theta})$ is asymptotically unbiased.
- $\mathbb{V}(g(\hat{\theta})) \approx \dfrac{[g'(\theta)]^2}{J(\theta)}$ which is the CR lower-bound.

Conclusion: ML estimator is asymptotically optimal.

LECTURE 24 | 2020-11-29

EXAMPLE 6.4.10

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$.

(i) Find ML estimator of θ, $\hat{\theta}$.

(ii) Find ML estimator of $\mathbb{P}(X_1 = 0) := g(\theta)$.

(iii) Find limiting distribution of

$$\sqrt{n}(\hat{\theta} - \theta)$$

(iv) Find limiting distribution of

$$\sqrt{n}(g(\hat{\theta}) - g(\theta))$$

(v) Is $\hat{\theta}$ (or $g(\hat{\theta})$) unbiased?

Solution.

(i) $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$

(ii) $\mathbb{P}(X_1 = 0) = e^{-\theta} := g(\theta)$, therefore, ML estimator of $g(\theta)$ is $g(\hat{\theta}) = e^{-\bar{X}_n}$ by the invariance property.

(iii) $\sqrt{n}(\hat{\theta} - \theta) = \sqrt{\bar{X}_n - \theta}$.

- Method 1 (ML estimator): First, note that the support does not depend on θ. If it does, then see Method 2.

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\text{d}} \mathcal{N}\left(0, \frac{1}{J_1(\theta)}\right)$$

  Since we've shown $J(\theta) = \frac{n}{\theta}$, find

$$J_1(\theta) = \frac{J(\theta)}{n} = \frac{1}{\theta}$$

  Therefore,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\text{d}} \mathcal{N}(0, \theta)$$

- Method 2 (CLT): $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$ so $\mathbb{E}[X_i] = \theta$ and $\mathbb{V}(X_i) = \theta$.

$$\implies \frac{\sqrt{n}(\bar{X}_n - \theta)}{\sqrt{\theta}} \xrightarrow{\text{d}} Z \sim \mathcal{N}(0, 1)$$

  Continuous mapping theorem,

$$\sqrt{\theta}\frac{\sqrt{n}(\bar{X}_n - \theta)}{\sqrt{\theta}} \xrightarrow{\text{d}} \sqrt{\theta}Z \sim \mathcal{N}(0, \theta)$$

(iv)

$$\sqrt{n}(g(\hat{\theta}) - g(\theta))$$

- Method 1 (ML estimator of $g(\theta)$): First, note that the support does not depend on θ. If it does, then see Method 2.

$$\sqrt{n}(e^{-\bar{X}_n} - e^{\theta}) \xrightarrow{\text{d}} \mathcal{N}(0, \frac{[g'(\theta)]^2}{J_1(\theta)})$$

  Here, $g(x) = e^{-x}$.

$$\frac{[g'(\theta)]^2}{J_1(\theta)} = e^{-2\theta}\theta$$

  So, $\sqrt{n}(e^{-\bar{X}_n} - e^{\theta}) \xrightarrow{\text{d}} \mathcal{N}(0, e^{-2\theta}\theta)$

- By using delta method

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\text{d}} \mathcal{N}(0, \theta)$$

Take $g(x) = e^{-x}$. Therefore,

$$\sqrt{n}(g(\bar{X}_n) - g(\theta)) \xrightarrow{\text{d}} \mathcal{N}(0, [g'(\theta)]^2\theta)$$

(v) Approximate mean of $\hat{\theta}$ is θ. Approximate mean of $e^{-\hat{\theta}}$ is $e^{-\theta}$. However, we want the exact expectation.

- Part 1

$$\mathbb{E}[e^{-\bar{X}_n}] = \mathbb{E}\left[\exp\left\{-\frac{1}{n}\sum_{i=1}^{n} X_i\right\}\right]$$

$X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta) \implies \sum_{i=1}^{n} X_i \sim \text{Poisson}(n\theta)$

$$\mathbb{E}[e^{-T/n}] = M_T\left(-\frac{1}{n}\right) = \exp\{n\theta(e^{-1/n} - 1)\}$$

Therefore,

$$\mathbb{E}[g(\hat{\theta})] \neq e^{-\theta}$$

$$\lim_{n\to\infty} e^{\theta[ne^{-1/n} - 1]}$$

Consider

$$e^x = 1 + x + x^2 + o(x^2)$$

$$e^{-1/n} = 1 - \frac{1}{n} + \frac{1}{n^2} + o\left(\frac{1}{n^2}\right)$$

Therefore,

$$n(e^{-1/n} - 1) = -1 + \frac{1}{n} + o\left(\frac{1}{n}\right)$$

Asymptotic mean of $\hat{\theta}$ is θ since

$$\lim_{n\to\infty} \mathbb{E}[\hat{\theta}] = \theta$$

- Part 2

$$\mathbb{E}[\hat{\theta}] = \theta$$

$$\mathbb{E}[\theta] = \mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mathbb{E}[X_1] = \theta$$

$\hat{\theta}$ is an unbiased estimator of θ.