

Stats 100B – Intro to Statistics

University of California, Los Angeles

Duc Vu

Summer 2021

This is stats 100B taught by Professor Christou. The formal name of the class is **Introduction to Mathematical Statistics**. There is not an official textbook used for the course. Instead, handouts and reference materials are distributed and can be accessed through the class [website](#). You can find other math/stats lecture notes through my personal [blog](#). Let me know through my [email](#) if you notice something mathematically wrong/concerning. Thank you!

Contents

1	Lec 1: Aug 3, 2021	4
1.1	Review of Stats 100A	4
1.2	Exponential Families	5
1.3	Moment Generating Functions	7
2	Lec 2: Aug 4, 2021	10
2.1	Moment Generating Functions (Cont'd)	10
2.2	Joint MGF	13
3	Lec 3: Aug 10, 2021	16
3.1	Method of Transformation	16
3.2	Joint MGF (Cont'd)	18
3.3	Multivariate Normal Distribution	19
4	Lec 4: Aug 12, 2021	21
4.1	Multivariate Normal Distribution (Cont'd)	21
4.2	Statistical Independence	23
4.3	Conditional PDF of Normal Distribution	24
5	Lec 5: Aug 17, 2021	26
5.1	Multinomial Distribution	26
5.2	Chi-Squared Distribution	26
5.3	Non-central Chi-Squared Distribution	29
5.4	t-Distribution	30
6	Lec 6: Aug 19, 2021	32
6.1	t Distribution (Cont'd)	32
6.2	F Distribution	33
6.3	Properties of Estimators	34
7	Lec 7: Aug 24, 2021	37
7.1	Properties of Estimators (Cont'd)	37
7.2	Bias and Mean Square Error	40
7.3	Method of Maximum Likelihood	42

8 Lec 8: Aug 26, 2021	45
8.1 Method of Maximum Likelihood (Cont'd)	45
8.2 Order Statistics	46
8.3 MLE with Multi-Parameters	48
8.4 Method of Moments	49
8.5 Simple Linear Models	50

List of Theorems

2.9 Central Limit Theorem	13
-------------------------------------	----

List of Definitions

1.3 Exponential Family	5
1.8 Moment Generating Function	7
2.10 Joint MGF	13
5.2 Chi-Squared	26
5.4 Non-central χ^2	29
6.5 Non-Central t Distribution	33
6.6 F Distribution	33
6.7 Non-Central F Distribution	33
7.2 Consistency/Converge in Probability	38

§1 | Lec 1: Aug 3, 2021

§1.1 Review of Stats 100A

Let X be a random variable.

	Discrete RV	Continuous RV
Distribution Function	pmf	pdf
Expected Value	$EX = \sum_x xp(x)$	$EX = \int_x xf(x) dx$
Expectation Function	$Eg(x) = \sum_x g(x)p(x)$	$Eg(x) = \int_x g(x)f(x)dx$
Variance	$EX^2 - (EX)^2$	$EX^2 - (EX)^2$

Let X, Y be random variables with the joint pdf/pmf $f(x, y)$. If X, Y are independent, then

$$f(x, y) = f(x) \cdot f(y)$$

where $f(x)$ is the marginal pdf of x and $f(y)$ is the marginal pdf of y . Also,

$$f(x) = \int_y f(x, y) dy$$

$$f(y) = \int_x f(x, y) dx$$

Theorem 1.1

X, Y are independent if and only if

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

Remark 1.2. $g(x)$ and $h(y)$ are not necessarily the marginal pdf of x and y respectively.

Proof. Let $c = \int_x g(x) dx$ and $d = \int_y h(y) dy$. Notice that

$$c \cdot d = \int_x \int_y \underbrace{g(x)h(y)}_{f(x,y)} dx dy = 1$$

Now, we find $f(x)$ and $f(y)$

$$f(x) = \int_y f(x, y) dy = \int_y g(x)h(y) dy = g(x)d$$

$$f(y) = \int_x f(x, y) dx = \int_x g(x)h(y) dx = h(y)c$$

So,

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

Therefore, X, Y are independent. □

Let $X \sim \Gamma(\alpha, \beta)$. Then, for $x > 0, \alpha > 0, \beta > 0$,

$$f(x) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha) \beta^\alpha}$$

where

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

We have the following properties

$$\begin{aligned}\Gamma(\alpha + 1) &= \alpha \Gamma(\alpha) \\ \Gamma(\alpha + 2) &= (\alpha + 1) \Gamma(\alpha + 1) \\ &= (\alpha + 1) \Gamma(\alpha - 1)\end{aligned}$$

If α is an integer, then

$$\Gamma(\alpha) = (\alpha - 1)!$$

Kernel function of $\Gamma(\alpha, \beta)$ is

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

Let's make a substitution $y = \frac{x}{\beta}$. Then,

$$\begin{aligned}\int_0^\infty x^{\alpha-1} e^{-\frac{x}{\beta}} dx &= \int_0^\infty (\beta y)^{\alpha-1} e^{-y} \beta dy \\ &= \beta^\alpha \int_0^\infty y^{\alpha-1} e^{-y} dy \\ &= \beta^\alpha \Gamma(\alpha)\end{aligned}$$

So

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

§1.2 Exponential Families

Definition 1.3 (Exponential Family) — A random variable X belongs in the exponential family if its pdf/pmf can be expressed as follows

$$f(x|\theta) = h(x) \cdot c(\theta) \cdot e^{\sum_{i=1}^k w_i(\theta) \cdot t_i(x)}$$

Example 1.4

Let $X \sim b(n, p)$ with n fixed. Show that this belongs in an exponential family.

$$\begin{aligned} p(x) &= \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} (1-p)^n \left(\frac{p}{1-p} \right)^x \\ &= \binom{n}{x} (1-p)^n e^{\ln\left(\frac{p}{1-p}\right)x} \\ &= \binom{n}{x} (1-p)^n e^{(\ln \frac{p}{1-p})x} \end{aligned}$$

So, we have

$$\begin{aligned} h(x) &= \binom{n}{x} \\ c(\theta) &= (1-p)^n \\ w_1(\theta) &= \ln \frac{p}{1-p} \\ t_1(x) &= x \end{aligned}$$

Notice that in this case we have one parameter, and that is $\theta = p$.

Example 1.5

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

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots$$

Show that it is an exponential family.

$$p(x) = \frac{1}{x!} e^{-\lambda} e^{\ln \lambda^x} = \frac{1}{x!} e^{-\lambda} e^{(\ln \lambda)x}$$

where $h(x) = \frac{1}{x!}$, $c(\theta) = e^{-\lambda}$, $w_1(\theta) = \ln \lambda$, $t_1(x) = x$.

Theorem 1.6 a) $E \left[\sum \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(x) \right] = -\frac{\partial \ln c(\theta)}{\partial \theta_j}$

b) $\text{var} \left(\sum_{i=1}^k \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(x) \right) = -\frac{\partial^2 \ln c(\theta)}{\partial \theta_j^2} - E \left[\sum_{i=1}^k \frac{\partial^2 w_i(\theta)}{\partial \theta_j^2} t_i(x) \right]$

Example 1.7

If $X \sim \text{Poisson}(\lambda)$ then show that $EX = \lambda$. From the theorem above (a)

$$E \left[\frac{1}{\lambda} X \right] = -(-1) \implies EX = \lambda$$

Exercise 1.1. $X \sim N(\mu, \sigma)$. Show that $f(X)$ belongs to an exponential family.

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

§1.3 Moment Generating Functions

Definition 1.8 (Moment Generating Function) — Let X be a random variable. Then the mgf of X is

$$M_X(t) = Ee^{tX} = \begin{cases} \int_x e^{tx} f(x) dx, & \text{for continuous RV} \\ \sum_x e^{tx} p(x), & \text{for discrete RV} \end{cases}$$

Moments:

$$M_X(t) = \int_x e^{tx} f(x) dx$$

$$M'_X(t) = \int_x x e^{tx} f(x) dx$$

$$M'_X(0) = \int_x x f(x) dx = EX$$

$$M''_X(t) = \int_x x^2 e^{tx} f(x) dx$$

$$M''_X(0) = \int_x x^2 f(x) dx = EX^2$$

$$\text{var}(X) = EX^2 - (EX)^2$$

Theorem 1.9

Let $\phi(t) = \ln M_X(t)$. Then

$$\phi'(0) = EX$$

$$\phi''(0) = \text{var}(X)$$

Proof. We have

$$\phi'(t) = \frac{M'_X(t)}{M_X(t)}$$

$$\phi'(0) = \frac{M'_X(0)}{M_X(0)} = \frac{E(X)}{1} = EX$$

and

$$\phi''(t) = \frac{M''_X(t) \cdot M_X(t) - (M'_X(t))^2}{(M_X(t))^2}$$

$$= \dots$$

$$= EX^2 - (EX)^2$$

$$= \text{var}(X)$$

□

The MGF of

- Binomial – $X \sim b(n, p)$

$$\begin{aligned}
 p(x) &= \binom{n}{x} p^x (1-p)^{n-x} \\
 M_X(t) &= Ee^{tx} = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\
 &= \sum_{x=0}^n \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\
 &= (pe^t + 1 - p)^n
 \end{aligned}$$

- Poisson

$$\begin{aligned}
 p(x) &= \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots \\
 M_X(t) &= \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\
 &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\
 &= e^{-\lambda} e^{\lambda e^t} \\
 &= e^{\lambda(e^t - 1)}
 \end{aligned}$$

- Gamma – $X \sim \Gamma(\alpha, \beta)$, $x, \alpha, \beta > 0$. Note that if $\lambda = 1$ and $\beta = \frac{1}{\lambda}$, then $f(x) = \lambda e^{-\lambda x}$, i.e. exponential distribution.

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

Let $y = x \left(\frac{1}{\beta} - t \right)$. Then, after some “massage”, we obtain

$$M_X(t) = (1 - \beta t)^{-\alpha}$$

- Exponential – $X \sim \exp(\lambda)$. Then,

$$M_X(t) = \left(1 - \frac{t}{\lambda} \right)^{-1}$$

- Normal – $Z \sim N(0, 1)$

$$\begin{aligned}
 f(z) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty \\
 M_Z(t) &= Ee^{tz} = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
 &= e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz \\
 &= e^{\frac{1}{2}t^2}
 \end{aligned}$$

Properties of MGF:

Theorem 1.10

If X, Y are independent, then

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

Proof. We have

$$\begin{aligned} M_{X+Y}(t) &= Ee^{t(X+Y)} \\ &= E(e^{tX} \cdot e^{tY}) \\ &= (Ee^{tX})(Ee^{tY}) \\ &= M_X(t) \cdot M_Y(t) \end{aligned}$$

□

Example 1.11

Let X_1, X_2, \dots, X_n be i.i.d random variables with $X_i \sim \exp(\lambda)$. Find the distribution of $X_1 + X_2 + \dots + X_n$. From the theorem above, we have

$$\begin{aligned} M_{X_1+X_2+\dots+X_n}(t) &= M_{X_1}(t) \cdot M_{X_2}(t) \dots M_{X_n}(t) \\ &= \left(1 - \frac{t}{\lambda}\right)^{-1} \left(1 - \frac{t}{\lambda}\right)^{-1} \dots \left(1 - \frac{t}{\lambda}\right)^{-1} \\ &= \left(1 - \frac{t}{\lambda}\right)^{-n} \end{aligned}$$

Thus, the sum $X_1 + X_2 + \dots + X_n \sim \Gamma\left(n, \frac{1}{\lambda}\right)$.

§2 | Lec 2: Aug 4, 2021

§2.1 Moment Generating Functions (Cont'd)

Example 2.1 (Method of MGF)

$X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$, and X, Y are independent.

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^t-1)} \end{aligned}$$

Thus, $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$ (by uniqueness theorem, i.e., each distribution has its own unique generating function).

Example 2.2 (Method of MGF)

Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d}}{\sim} \text{Poisson}(\lambda)$ and $T = X_1 + X_2 + \dots + X_n$.

$$\begin{aligned} M_T(t) &= (M_{X_i}(t))^n \\ &= \left(e^{\lambda(e^t-1)} \right)^n \\ &= e^{n\lambda(e^t-1)} \end{aligned}$$

So, $T \sim \text{Poisson}(n\lambda)$.

Example 2.3 (Method of PMF)

From Example 2.1, we have

$$\begin{aligned} P(X + Y = k) &= \sum_{i=0}^k p(X = i, Y = k - i) \\ &= \sum_{i=0}^k p(X = i) \cdot p(Y = k - i) \\ &= \sum_{i=0}^k \frac{\lambda_1^i e^{-\lambda_1}}{i!} \cdot \frac{\lambda_2^{k-i} e^{-\lambda_2}}{(k-i)!} \\ &= e^{-(\lambda_1+\lambda_2)} \sum_{i=0}^k \frac{\lambda_1^i \lambda_2^{k-i}}{i!(k-i)!} \cdot \frac{k!}{k!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^k \binom{k}{i} \lambda_1^i \lambda_2^{k-i} \\ &= \frac{(\lambda_1 + \lambda_2)^k e^{-(\lambda_1+\lambda_2)}}{k!} \end{aligned}$$

Thus, $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Example 2.4

Suppose $X \sim b(n_1, p)$, $Y \sim b(n_2, p)$, and X, Y are independent. Find the distribution of $X + Y$.

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= (pe^t + 1 - p)^{n_1} (pe^t + 1 - p)^{n_2} \\ &= (pe^t + 1 - p)^{n_1+n_2} \end{aligned}$$

Thus, $X + Y \sim b(n_1 + n_2, p)$.

Properties of MGF:

a) MGF of $X + a$ is

$$\begin{aligned} M_{X+a}(t) &= Ee^{t(X+a)} \\ &= e^{ta} \cdot Ee^{tX} = e^{ta} M_X(t) \end{aligned}$$

b) MGF of bX is

$$\begin{aligned} M_{bX}(t) &= Ee^{tbX} \\ &= Ee^{t^*X} \\ &= M_X(t^*) = M_X(bt) \end{aligned}$$

Example 2.5

$X \sim \Gamma(\alpha, \beta)$. Then,

$$M_X(t) = (1 - \beta t)^{-\alpha}$$

Let $Y = cX$ where $c > 0$. We want to find the distribution of Y .

(a) Method of MGF:

$$\begin{aligned} M_Y(t) &= M_{cX}(t) = M_X(ct) \\ &= (1 - c\beta t)^{-\alpha} \end{aligned}$$

Therefore, $Y \sim \Gamma(\alpha, c\beta)$.

(b) Method of CDF:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(cX \leq y) \\ &= P(X \leq \frac{y}{c}) \end{aligned}$$

Then, $F_Y(y) = F_X(\frac{y}{c})$. Take derivative w.r.t. y

$$\begin{aligned} f_Y(y) &= \frac{1}{c} f_X\left(\frac{y}{c}\right) \\ f(x) &= \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha} \end{aligned}$$

Lastly, replace X with $\frac{Y}{c}$.

c) MGF of $\frac{X+a}{b}$ is

$$\begin{aligned} M_{\frac{X+a}{b}}(t) &= Ee^{t \cdot \frac{X+a}{b}} \\ &= e^{t \frac{a}{b}} Ee^{t \frac{X}{b}} \\ &= e^{t \frac{a}{b}} \cdot M_X\left(\frac{t}{b}\right) \end{aligned}$$

Use these properties to find the MGF of $X \sim N(\mu, \sigma)$. Recall that if $Z \sim N(0, 1)$, then

$$M_Z(t) = e^{\frac{1}{2}t^2}$$

So, standardizing x to obtain

$$Z = \frac{X - \mu}{\sigma} \implies X = \mu + \sigma Z$$

Then,

$$\begin{aligned} M_X(t) &= M_{\mu + \sigma Z}(t) \\ &= Ee^{t(\mu + \sigma z)} \\ &= e^{t\mu} M_Z(\sigma t) \\ &= e^{t\mu} e^{\frac{1}{2}t^2\sigma^2} \end{aligned}$$

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

Example 2.6

Let $X \sim N(\mu_1, \sigma_1)$ and $Y \sim N(\mu_2, \sigma_2)$ and X, Y are independent. We want to find the distribution of $X + Y$.

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= e^{t\mu_1 + \frac{1}{2}t^2\sigma_1^2} \cdot e^{t\mu_2 + \frac{1}{2}t^2\sigma_2^2} \\ &= e^{t(\mu_1 + \mu_2) + \frac{1}{2}t^2(\sigma_1^2 + \sigma_2^2)} \end{aligned}$$

Thus, $X + Y \sim N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$.

Example 2.7

Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$. Let $T = X_1 + X_2 + \dots + X_n$. Then

$$\begin{aligned} M_T(t) &= (M_{X_i}(t))^n \\ &= \left(e^{t\mu + \frac{1}{2}t^2\sigma^2}\right)^n \\ &= e^{tn\mu + \frac{1}{2}t^2n\sigma^2} \end{aligned}$$

Thus, $T \sim N(n\mu, \sigma\sqrt{n})$.

Example 2.8

Let $\bar{X} = \frac{\sum X_i}{n} = \frac{T}{n}$. Find $M_{\bar{X}}(t)$.

$$\begin{aligned} M_{\bar{X}}(t) &= M_T\left(\frac{t}{n}\right) \\ &= e^{t\mu + \frac{1}{2}t^2 \frac{\sigma^2}{n}} \end{aligned}$$

Therefore, $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$.

Recall

Theorem 2.9 (Central Limit Theorem)

Let $T = X_1 + \dots + X_n$ with mean μ and variance σ^2 (can follow any distribution other than normal). As $n \rightarrow \infty$,

$$\frac{T - n\mu}{\sigma\sqrt{n}} \rightarrow N(0, 1)$$

Proof. Start with the MGF and as $n \rightarrow \infty$ we obtain

$$M_{\frac{T - n\mu}{\sigma\sqrt{n}}}(t) \rightarrow e^{\frac{1}{2}t^2}$$

□

§2.2 Joint MGF

Let $X = [X_1 \ X_2 \ \dots \ X_n]^\top$ be a random vector and $t = [t_1 \ t_2 \ \dots \ t_n]^\top$.

Definition 2.10 (Joint MGF) — Joint MGF of X is defined as

$$M_X(t) = Ee^{t^\top X} = Ee^{\sum t_i X_i}$$

Let X be a random vector (as above) with mean vector $\mu = [\mu_1 \ \mu_2 \ \dots \ \mu_n]^\top$, i.e.,

$$\mu = EX = E \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} EX_1 \\ EX_2 \\ \vdots \\ EX_n \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

and variance covariance matrix is defined as

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \sigma_n^2 \end{bmatrix} = E[(X - \mu)(X - \mu)^\top]$$

Special Case: For i.i.d random variables,

$$\mu = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} = \mu \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \mu \mathbf{1}$$

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix} = \sigma^2 I$$

Now, let's discuss two results.

1. Let $a = [a_1 \ a_2 \ \dots \ a_n]^\top$ be a vector of constants. Find the mean and variance of $a^\top X$.

$$\begin{aligned} E a^\top X &= a^\top E X = a^\top \mu \\ \text{var}(a^\top X) &= E(a^\top X - a^\top \mu)^2 \\ &= a^\top [E(X - \mu)(X - \mu)^\top] a \\ &= a^\top \Sigma a \end{aligned}$$

or using summation, we have

$$\text{var}(a^\top X) = \sum_{i=1}^n a_i^2 \text{var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j>i}^n a_i a_j \text{cov}(X_i, X_j)$$

Example 2.11

For $n = 3$,

$$\begin{aligned} \text{var}(a_1 X_1 + a_2 X_2 + a_3 X_3) &= [a_1 \ a_2 \ a_3] \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \\ &= a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + a_3^2 \sigma_3^2 + 2a_1 a_2 \sigma_{12} + 2a_1 a_3 \sigma_{13} + 2a_2 a_3 \sigma_{23} \end{aligned}$$

2. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{bmatrix}$$

be a $p \times n$ matrix of constants. Find mean and variance of the vector AX .

$$\begin{aligned} E(AX) &= AEX = A\mu \\ \text{var}(AX) &= E[(AX - A\mu)(AX - A\mu)^\top] \\ &= AE(X - \mu)(X - \mu)^\top A^\top \\ &= A\Sigma A^\top \end{aligned}$$

Consider $X^\top A X$ where $X : n \times 1$, $A : n \times n$ symmetric. For example, $n = 2$,

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$$

Then $X^\top AX = 5X_1^2 + 3X_2^2 + 4X_1X_2$.

$$\begin{aligned}
 E \left[\underbrace{X^\top AX}_{\text{scalar}} \right] &= E \operatorname{tr}(X^\top AX) \\
 &= E (\operatorname{tr} AXX^\top) \\
 &= \operatorname{tr} (EAXX^\top) \\
 &= \operatorname{tr} (AEXX^\top) \\
 &= \operatorname{tr} (A(\Sigma + \mu\mu^\top)) \\
 &= \operatorname{tr}(A\Sigma) + \operatorname{tr}(A\mu\mu^\top) \\
 &= \operatorname{tr}(A\Sigma) + \mu^\top A\mu
 \end{aligned}$$

Note that $\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB) \neq \operatorname{tr}(BAC)$

Let $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, $t = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$. Then,

$$\begin{aligned}
 M_X(t) &= E(e^{t_1 X_1 + t_2 X_2}) \\
 &= \int_{x_1} \int_{x_2} e^{t_1 x_1 + t_2 x_2} f(x_1, x_2) dx_1 dx_2 \\
 M_1(t) &= \frac{\partial M_X(t)}{\partial t_1} = \int_{x_1} \int_{x_2} x_1 e^{t_1 x_1 + t_2 x_2} f(x_1, x_2) dx_1 dx_2
 \end{aligned}$$

Set $t = 0$, we obtain

$$\begin{aligned}
 M_1(0) &= \int \int x_1 f(x_1, x_2) dx_1 dx_2 \\
 &= \int_{x_1} x_1 \left[\int_{x_2} f(x_1, x_2) dx_2 \right] dx_1 \\
 &= \int_{x_1} x_1 f(x_1) dx_1 \\
 &= EX_1
 \end{aligned}$$

So,

$$\begin{aligned}
 \operatorname{var}(X_1) &= EX_1^2 - (EX_1)^2 \\
 \operatorname{cov}(X_1, X_2) &= E(X_1, X_2) - (EX_1)(EX_2)
 \end{aligned}$$

§3 | Lec 3: Aug 10, 2021

§3.1 Method of Transformation

Let X be a random variable and $Y = g(X)$ be a function of X . If $g(X)$ is increasing or decreasing function of X , then the pdf of Y is given by

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

This is known as the **method of transformation**.

Example 3.1 (Increasing Function Case)

Let $Y = 3X - 1$.

- Method of CDF:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(3X - 1 \leq y) \\ &= P(X \leq \frac{y+1}{3}) \\ &= F_X\left(\frac{y+1}{3}\right) \end{aligned}$$

Thus, $f_Y(y) = \frac{1}{3}f_X\left(\frac{y+1}{3}\right)$

- Method of transformation

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{y+1}{3}\right) \left| \frac{d}{dy} \left(\frac{y+1}{3}\right) \right| \\ &= \frac{1}{3}f_X\left(\frac{y+1}{3}\right) \end{aligned}$$

Example 3.2

$X \sim \Gamma(\alpha, \beta)$. Let $Y = cX$ for some $c > 0$. Find the pdf of Y using the method of transformation.

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{y}{c}\right) \frac{d}{dy} \left(\frac{y}{c}\right) \\ &= \frac{y^{\alpha-1} \exp\left(\frac{-y}{\beta c}\right) \frac{1}{c}}{\beta^\alpha \Gamma(\alpha) c^{\alpha-1}} \\ &= \frac{y^{\alpha-1} \exp\left(-\frac{y}{c\beta}\right)}{\Gamma(\alpha)(c\beta)^\alpha} \end{aligned}$$

$$\Rightarrow Y \sim \Gamma(\alpha, c\beta).$$

Let X_1, X_2 be random variables with joint pdf $f_{x_1 x_2}(x_1, x_2)$. Now, suppose that $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$. We want to find the joint pdf of Y_1, Y_2 .

Let $x_1 = h^{-1}(y_1, y_2)$ and $x_2 = h_2^{-1}(y_1, y_2)$. Now, let's find the Jacobian of the transformation.

$$J = \begin{vmatrix} \frac{\partial h_1^{-1}(y_1, y_2)}{\partial y_1} & \frac{\partial h_1^{-1}(y_1, y_2)}{\partial y_2} \\ \frac{\partial h_2^{-1}(y_1, y_2)}{\partial y_1} & \frac{\partial h_2^{-1}(y_1, y_2)}{\partial y_2} \end{vmatrix}$$

or

$$J = \begin{vmatrix} \frac{\partial g_1(x_1, x_2)}{\partial x_1} & \frac{\partial g_1(x_1, x_2)}{\partial x_2} \\ \frac{\partial g_2(x_1, x_2)}{\partial x_1} & \frac{\partial g_2(x_1, x_2)}{\partial x_2} \end{vmatrix}$$

Finally, we find the joint pdf of Y_1 and Y_2 by using the inverse function

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2} \left(\begin{matrix} x_1 = h_1^{-1}(y_1, y_2) \\ x_2 = h_2^{-1}(y_1, y_2) \end{matrix} \right) \cdot |J|$$

or by using the original function

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2} \left(\begin{matrix} x_1 = h_1^{-1}(y_1, y_2) \\ x_2 = h_2^{-1}(y_1, y_2) \end{matrix} \right) \cdot |J|^{-1}$$

Example 3.3

Let $X_1 \sim \exp(\lambda_1)$ and $X_2 \sim \exp(\lambda_2)$. Suppose $U = X_1 + X_2$ and $V = X_1 - X_2$. Find the joint pdf of U and V if X_1, X_2 are independent.

The joint pdf of X_1, X_2

$$f_{X_1 X_2}(x_1, x_2) = f(x_1) \cdot f(x_2) = \lambda_1 \lambda_2 e^{-(\lambda_1 x_1 + \lambda_2 x_2)}$$

First, let's find x_1 and x_2 in terms of u and v .

$$\begin{aligned} x_1 &= \frac{u + v}{2} \\ x_2 &= \frac{u - v}{2} \end{aligned}$$

Then, we can calculate the Jacobian as follows

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

or if we want to use the original function then

$$J = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

So, the pdf is

$$f_{UV}(u, v) = \frac{\lambda_1 \lambda_2}{2} \exp \left(-\lambda_1 \frac{u + v}{2} - \lambda_2 \frac{u - v}{2} \right)$$

Example 3.4

Let $X \sim \Gamma(\alpha_1, \beta)$ and $Y \sim \Gamma(\alpha_2, \beta)$, X, Y are independent. Let $U = X + Y$ and $V = \frac{X}{X+Y}$. Find the joint pdf of U, V .

$$x = uv$$

$$y = u - uv$$

The Jacobian is

$$J = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -u$$

So the pdf is

$$\begin{aligned} f_{UV}(u, v) &= \frac{(uv)^{\alpha_1-1} (u(1-v))^{\alpha_2-1} \exp\left(-\frac{u}{\beta}\right)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} \cdot u \\ &= \frac{u^{\alpha_1+\alpha_2-1} \exp\left(-\frac{u}{\beta}\right) v^{\alpha_1-1} (1-v)^{\alpha_2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} \end{aligned}$$

From the example above notice that if we multiply $\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$, then we obtain

$$f_{UV}(u, v) = \frac{u^{\alpha_1+\alpha_2-1} \exp\left(-\frac{u}{\beta}\right)}{\Gamma(\alpha_1 + \alpha_2)\beta^{\alpha_1+\alpha_2}} \cdot \frac{v^{\alpha_1-1} (1-v)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)}$$

We can observe that $U \sim \Gamma(\alpha_1 + \alpha_2, \beta)$ and $V \sim \text{beta}(\alpha_1, \alpha_2)$ where $B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$. Also, we can observe that U and V are independent.

§3.2 Joint MGF (Cont'd)

Consider

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}$$

$$M_X(t) = Ee^{t^\top X} = Ee^{\sum t_i X_i}$$

Suppose

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{pmatrix} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}$$

and similarly,

$$\mathbf{t} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$$

Apply what we assume,

$$\begin{aligned} M_X(t) &= Ee^{t^\top X} = E \exp\left(\begin{pmatrix} \mathbf{u}^\top & \mathbf{v}^\top \end{pmatrix} \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}\right) \\ &= E \exp\left(\sum u_i y_i + \sum v_i z_i\right) \end{aligned}$$

Now, we let all $v_i = 0$,

$$M_X(t) = E \exp \left(\sum u_i y_i \right) = E \exp (\mathbf{u}^\top \mathbf{Y}) = M_Y(\mathbf{u})$$

In general,

$$M_Y(u) = M_X(u, 0)$$

$$M_Z(v) = M_X(0, v)$$

Example 3.5

For $n = 3$,

$$M_X(t_1, t_2, t_3) = (1 - t_1 + 2t_2)^{-4} (1 - t_1 + 3t_3)^{-3} (1 - t_1)^{-2}$$

Then, say we want to find $M_{X_1}(t_1)$ – set $t_2 = t_3 = 0$,

$$M_X(t_1, 0, 0) = (1 - t_1)^{-9}$$

or for t_1, t_3

$$M_{X_1 X_3}(t_1, t_3) = M_X(t_1, 0, t_3) = (1 - t_1)^{-6} (1 - t_1 + 3t_3)^{-3}$$

Note (on independence): Use the same notation as above \mathbf{X}, \mathbf{t} . \mathbf{Y} and \mathbf{Z} are independent if and only if

$$M_X(t) = E \exp (\mathbf{u}^\top \mathbf{Y} + \mathbf{v}^\top \mathbf{Z}) = E e^{\mathbf{u}^\top \mathbf{Y}} \cdot E e^{\mathbf{v}^\top \mathbf{Z}} = M_Y(\mathbf{u}) \cdot M_Z(\mathbf{v})$$

Example 3.6

Consider:

$$M_X(t_1, t_2, t_3) = (1 - t_1 + 2t_2)^{-4} (1 - t_1 + 3t_3)^{-3} (1 - t_1)^{-2}$$

1. Find MGF of $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$

$$M_{X_1 X_3}(t_1, 0, t_3) = (1 - t_1)^{-6} (1 - t_1 + 3t_3)^{-3}$$

2. Find MGF of X_1

$$M_{X_1}(t_1) = (1 - t_1)^{-9}$$

3. Find MGF of X_3

$$M_{X_3}(t_3) = (1 + 3t_3)^{-3}$$

4. Are X_1, X_3 independent?

Notice that $M_{X_1 X_3}(t_1, t_3) \neq M_{X_1}(t_1) \cdot M_{X_3}(t_3)$. Thus, X_1, X_3 are not independent.

§3.3 Multivariate Normal Distribution

Suppose \mathbf{Y} is a random vector ($n \times 1$) with mean vector $\boldsymbol{\mu}$ and variance covariance matrix $\boldsymbol{\Sigma}$. Then, we say that $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if its joint pdf is given by the following

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right)$$

If $n = 2$, then we have the bivariate normal distribution with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$$

or

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & p\sigma_1\sigma_2 \\ p\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where $p = \frac{\sigma_{12}}{\sigma_1\sigma_2}$. We now want to find the joint MGF of $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let Z_1, Z_2, \dots, Z_n be i.i.d and $\sim N(0, 1)$. Show that $\mathbf{Z} \sim N(\mathbf{0}, I)$.

$$f(\mathbf{z}) = f(\mathbf{z}_1) \cdot \dots \cdot f(\mathbf{z}_n)$$

$$f(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2}$$

So,

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\mathbf{z}^\top \mathbf{z}\right)$$

Thus, $\mathbf{Z} \sim N(\mathbf{0}, I)$.

Now, let's find the joint MGF.

$$\begin{aligned} M_Z(\mathbf{t}) &= Ee^{\mathbf{t}^\top \mathbf{z}} = Ee^{t_1 z_1 + \dots + t_n z_n} \\ &= Ee^{t_1 z_1} \dots Ee^{t_n z_n} \\ &= e^{\frac{1}{2}t_1^2} \dots e^{\frac{1}{2}t_n^2} \\ &= e^{\frac{1}{2}\sum t_i^2} \\ &= e^{\frac{1}{2}\mathbf{t}^\top \mathbf{t}} \end{aligned}$$

Suppose now $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Show that $\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{y} - \boldsymbol{\mu})$ follows $N_n(\mathbf{0}, I)$.

Notice that $\boldsymbol{\Sigma}$ is a symmetric matrix and its spectral decomposition is given by

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^\top$$

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\boldsymbol{\Sigma}$ using $|\boldsymbol{\Sigma} - \lambda I| = 0$. We also have the corresponding eigenvectors in which $\boldsymbol{\Sigma}\mathbf{x} = \lambda\mathbf{x}$. The normalized eigenvectors are denoted with $\mathbf{e}_1, \dots, \mathbf{e}_n$. They are orthogonal, i.e., $\mathbf{P}\mathbf{P}^\top = I$ in which $\mathbf{P} = (\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_n)$. In addition, observe that $\mathbf{e}_1^\top \mathbf{e}_1 = 1$, $\mathbf{e}_1^\top \mathbf{e}_2 = 0$ (for example).

Remark 3.7. Using spectral decomposition, we can compute $\boldsymbol{\Sigma}^{-1}$, $\boldsymbol{\Sigma}^{-\frac{1}{2}}$, $\boldsymbol{\Sigma}^{\frac{1}{2}}$ more conveniently by

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} &= \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}^\top \\ \boldsymbol{\Sigma}^{\frac{1}{2}} &= \mathbf{P}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{P}^\top \\ \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}} &= \boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}^{-\frac{1}{2}} &= \mathbf{P}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{P}^\top \\ \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}} &= \boldsymbol{\Sigma}^{-1} \end{aligned}$$

§4 | Lec 4: Aug 12, 2021

§4.1 Multivariate Normal Distribution (Cont'd)

If $Z_1, \dots, Z_n \stackrel{\text{i.i.d}}{\sim} N(0, 1)$. Then $\mathbf{Z} \sim N(\mathbf{0}, I)$ and

$$M_{\mathbf{Z}}(\mathbf{t}) = Ee^{\mathbf{t}^\top \mathbf{z}} = e^{\frac{1}{2}\mathbf{t}^\top \mathbf{t}}$$

If $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then let's show that $\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y} - \boldsymbol{\mu})$ follows $N(\mathbf{0}, I)$.

Note: From univariate normal, if $Y \sim N(\mu, \sigma)$, then $Z = \frac{Y - \mu}{\sigma} = (\sigma^2)^{-\frac{1}{2}}(Y - \mu) \sim N(0, 1)$.

Proof. We have

$$\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y} - \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}$$

Let

$$\boldsymbol{\Sigma}^{-\frac{1}{2}} = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ \vdots & & & \vdots \\ v_{n1} & v_{n2} & \dots & v_{nn} \end{pmatrix}$$

Then

$$\begin{aligned} Z_1 &= v_{11}y_1 + v_{12}y_2 + \dots + v_{1n}y_n - \text{const1} \\ Z_2 &= v_{21}y_1 + v_{22}y_2 + \dots + v_{2n}y_n - \text{const2} \\ &\vdots \\ Z_n &= v_{n1}y_1 + v_{n2}y_2 + \dots + v_{nn}y_n - \text{constn} \end{aligned}$$

a) Pdf of \mathbf{Y}

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})}$$

and

$$\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y} - \boldsymbol{\mu})$$

So

$$\mathbf{Y} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}$$

b) Jacobian

$$J = \begin{vmatrix} \frac{\partial Z_1}{\partial y_1} & \dots & \frac{\partial Z_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial Z_n}{\partial y_{n1}} & \dots & \frac{\partial Z_n}{\partial y_{nn}} \end{vmatrix} = |\boldsymbol{\Sigma}^{-\frac{1}{2}}| = |\boldsymbol{\Sigma}|^{-\frac{1}{2}}$$

Finally, we can find the pdf of Z as follows

$$\begin{aligned} f(\mathbf{z}) &= \frac{1}{(2\pi)^{\frac{n}{2}}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{z} + \boldsymbol{\mu} - \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{z} + \boldsymbol{\mu} - \boldsymbol{\mu}\right)\right) \cdot |\boldsymbol{\Sigma}|^{\frac{1}{2}} \\ f(\mathbf{z}) &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}} \end{aligned}$$

Thus, $\mathbf{Z} \sim N(\mathbf{0}, I)$. □

Now, we use this result to find the joint MGF of $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $Z \sim N(0, 1)$, then

$$M_Z(t) = e^{-\frac{1}{2}t^2}$$

For the MGF of $Y \sim N(\mu, \sigma)$,

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

Then, for multivariate normal, $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\mathbf{Z} = \boldsymbol{\Sigma}^{\frac{1}{2}}(\mathbf{Y} - \boldsymbol{\mu})$$

Solve for \mathbf{Y}

$$\mathbf{Y} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}$$

So, the MGF is

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= M_{\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}}(\mathbf{t}) \\ &= Ee^{\mathbf{t}^\top (\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu})} \\ &= e^{\mathbf{t}^\top \boldsymbol{\mu}} \cdot Ee^{(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{t})^\top \mathbf{Z}} \end{aligned}$$

Let $\mathbf{t}^* = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{t}$.

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= e^{\mathbf{t}^\top \boldsymbol{\mu}} \cdot Ee^{\mathbf{t}^{*\top} \mathbf{Z}} \\ &= e^{\mathbf{t}^\top \boldsymbol{\mu}} \cdot e^{\frac{1}{2}\mathbf{t}^{*\top} \mathbf{t}^*} \end{aligned}$$

Replace $\mathbf{t}^* = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{t}$ to obtain

$$M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}}$$

Theorem 4.1

Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Suppose \mathbf{A} is a $m \times n$ matrix of constants and \mathbf{C} is a $m \times 1$ vector of constants. The distribution of $\mathbf{A}\mathbf{Y} + \mathbf{C}$ is multivariate normal.

Proof. Consider the MGF

$$\begin{aligned} M_{\mathbf{A}\mathbf{Y} + \mathbf{C}}(\mathbf{t}) &= Ee^{\mathbf{t}^\top (\mathbf{A}\mathbf{Y} + \mathbf{C})} \\ &= e^{\mathbf{t}^\top \mathbf{C}} Ee^{(\mathbf{A}^\top \mathbf{t})^\top \mathbf{Y}} \end{aligned}$$

Let $\mathbf{t}^* = \mathbf{A}^\top \mathbf{t}$.

$$\begin{aligned} M_{\mathbf{A}\mathbf{Y} + \mathbf{C}} &= e^{\mathbf{t}^\top \mathbf{C}} Ee^{\mathbf{t}^{*\top} \mathbf{Y}} \\ &= e^{\mathbf{t}^\top \mathbf{C}} \cdot M_{\mathbf{Y}}(\mathbf{t}^*) \\ &= e^{\mathbf{t}^\top \mathbf{C}} e^{\mathbf{t}^{*\top} \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^{*\top} \boldsymbol{\Sigma} \mathbf{t}^*} \end{aligned}$$

Substitute $\mathbf{t}^* = \mathbf{A}^\top \mathbf{t}$ to get

$$M_{\mathbf{A}\mathbf{Y} + \mathbf{C}}(\mathbf{t}) = e^{\mathbf{t}^\top (\mathbf{A}\boldsymbol{\mu} + \mathbf{C}) + \frac{1}{2}\mathbf{t}^\top \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top \mathbf{t}}$$

Thus, $\mathbf{A}\mathbf{Y} + \mathbf{C} \sim N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{C}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$. □

In addition, we have

$$\begin{aligned} E(\mathbf{A}\mathbf{Y} + \mathbf{C}) &= \mathbf{A}\boldsymbol{\mu} + \mathbf{C} \\ \text{var}(\mathbf{A}\mathbf{Y} + \mathbf{C}) &= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top \end{aligned}$$

Theorem 4.2

Let

$$\mathbf{Q}_1 = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & | & 0 & 0 & 0 \\ 0 & 1 & | & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix} = \mathbf{A}\mathbf{Y}$$

where $\mathbf{A} = (I \ \mathbf{0})$. Then

$$\begin{aligned} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} &\sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top) \\ &\sim N\left((I \ \mathbf{0}) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, (I \ \mathbf{0}) \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I \\ \mathbf{0} \end{pmatrix}\right) \\ &\sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \end{aligned}$$

Also, the linear combination follows the normal distribution in which

$$a_1 Y_1 + a_2 Y_2 + \dots a_n Y_n = \mathbf{a}^\top \mathbf{Y} \sim N(\mathbf{a}^\top \boldsymbol{\mu}, \sqrt{\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}})$$

§4.2 Statistical Independence

Suppose

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{pmatrix}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

Then

$$M_{\mathbf{Y}}(\mathbf{t}) = \exp\left(\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}\right) = \exp\left(\mathbf{t}_1^\top \boldsymbol{\mu}_1 + \mathbf{t}_2^\top \boldsymbol{\mu}_2 + \frac{1}{2} \mathbf{t}_1^\top \boldsymbol{\Sigma}_{11} \mathbf{t}_1 + \frac{1}{2} \mathbf{t}_2^\top \boldsymbol{\Sigma}_{22} \mathbf{t}_2 + \mathbf{t}_1^\top \boldsymbol{\Sigma}_{12} \mathbf{t}_2\right)$$

If $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, then

$$M_{\mathbf{Y}}(\mathbf{t}) = \exp\left(\mathbf{t}_1^\top \boldsymbol{\mu}_1 + \frac{1}{2} \mathbf{t}_1^\top \boldsymbol{\Sigma}_{11} \mathbf{t}_1\right) \cdot \exp\left(\mathbf{t}_2^\top \boldsymbol{\mu}_2 + \frac{1}{2} \mathbf{t}_2^\top \boldsymbol{\Sigma}_{22} \mathbf{t}_2\right)$$

or

$$M_{\mathbf{Y}}(\mathbf{t}) = M_{\mathbf{Q}_1}(\mathbf{t}_1) \cdot M_{\mathbf{Q}_2}(\mathbf{t}_2)$$

So if $\text{cov}(\mathbf{Q}_1, \mathbf{Q}_2) = \mathbf{0}$, then $\mathbf{Q}_1, \mathbf{Q}_2$ are independent.

Theorem 4.3

Let $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and consider \mathbf{AY} and \mathbf{BY} .

$$\begin{pmatrix} \mathbf{AY} \\ \mathbf{BY} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{Y} = \mathbf{LY}$$

Then

$$\begin{aligned} \text{var}(\mathbf{LY}) &= \mathbf{L}\boldsymbol{\Sigma}\mathbf{L}^\top \\ &= \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \boldsymbol{\Sigma} \begin{pmatrix} \mathbf{A}^\top & \mathbf{B}^\top \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top & \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\top \\ \mathbf{B}\boldsymbol{\Sigma}\mathbf{A}^\top & \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top \end{pmatrix} \end{aligned}$$

\mathbf{AY} and \mathbf{BY} are independent if $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\top = \mathbf{0}$ or check $\text{cov}(\mathbf{AY}, \mathbf{BY}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\top$.

§4.3 Conditional PDF of Normal Distribution

Consider the bivariate case ($n = 2$).

$$\begin{aligned} \mathbf{Y} &= \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \\ f(y_2|y_1) &= \frac{f(y_1, y_2)}{f(y_1)} \end{aligned}$$

Notice that $f(y_1, y_2)$ is bivariate normal. Thus, $f(y_1)$ is univariate normal, $Y_1 \sim N(\mu_1, \sigma_1)$. So

$$f(y_1) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2}(y_1 - \mu_1)^2}$$

The conditional pdf then is

$$f_{Y_1|Y_2}(y_1|y_2) = \frac{1}{\sqrt{\sigma_1^2(1-\rho^2)}\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{[y_1 - \mu_1 - \rho \frac{\sigma_1}{\sigma_2}(y_2 - \mu_2)]^2}{\sigma_1^2(1-\rho^2)} \right) \right]$$

In general, suppose

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

and $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then, the conditional distribution of \mathbf{Q}_1 given \mathbf{Q}_2 is also multivariate normal, i.e., $\mathbf{Q}_1|\mathbf{Q}_2 \sim N(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$ where

$$\begin{aligned} \boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{Q}_2 - \boldsymbol{\mu}_2) \\ \boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \end{aligned}$$

Proof. Let

$$\begin{aligned} \mathbf{U} &= \mathbf{Q}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Q}_2 \\ \mathbf{V} &= \mathbf{Q}_2 \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{pmatrix} \\ &= \mathbf{A} \cdot \mathbf{Y} \end{aligned}$$

Let's find the mean and variance covariance matrix of $\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix}$.

$$\begin{aligned} E \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} &= \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \\ &= \begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix} \end{aligned}$$

Variance

$$\begin{aligned} \text{var} \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} &= \mathbf{A}\Sigma\mathbf{A}^\top \\ &= \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I & \mathbf{0}^\top \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix} \end{aligned}$$

Notice that $\text{cov}(\mathbf{U}, \mathbf{V}) = \mathbf{0}$, so \mathbf{U}, \mathbf{V} are independent because jointly they follow multivariate normal. \square

Question 4.1. Find $\text{cov}(\mathbf{U}, \mathbf{V})$ using $\text{cov}(\mathbf{A}\mathbf{Y}, \mathbf{B}\mathbf{Y}) = \mathbf{A}\Sigma\mathbf{B}^\top$

We have

$$\begin{aligned} \text{cov}(\mathbf{U}, \mathbf{V}) &= \text{cov}(\mathbf{Q}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{Q}_2, \mathbf{Q}_2) \\ &= \text{cov}(\mathbf{Q}_1, \mathbf{Q}_2) - \text{cov}(\Sigma_{12}\Sigma_{22}^{-1}\mathbf{Q}_2, \mathbf{Q}_2) \\ &= \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} \\ &= \mathbf{0} \end{aligned}$$

Observe that

$$\mathbf{Q}_1 = \mathbf{U} + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{Q}_2$$

Then

$$\mathbf{Q}_1|\mathbf{Q}_2 = \mathbf{U}|\mathbf{Q}_2 + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{Q}_2$$

but $\mathbf{Q}_2 = \mathbf{V}$

$$\begin{aligned} \mathbf{Q}_1|\mathbf{Q}_2 &= \mathbf{U}|\mathbf{V} + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{Q}_2 \\ &= \mathbf{U} + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{Q}_2 \\ E(\mathbf{Q}_1|\mathbf{Q}_2) &= \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{Q}_2 \\ &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{Q}_2 - \mu_2) \\ \text{var}(\mathbf{Q}_1|\mathbf{Q}_2) &= \text{var}(\mathbf{U}) \\ &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

§5 | Lec 5: Aug 17, 2021

§5.1 Multinomial Distribution

Suppose a sequence of n independent experiments is performed, and each one results in one of r possible outcomes with probabilities p_1, p_2, \dots, p_r and $\sum_{i=1}^r p_i = 1$. Let X_i be the number of the n experiments that result in outcome i where $i = 1, 2, \dots, r$. Then

$$P(X_1 = x_1, X_2 = x_2, \dots, X_r = x_r) = \frac{n!}{\underbrace{x_1! x_2! \dots x_r!}_{{}^n_{(x_1 \ x_2 \dots x_r)}}} p_1^{x_1} p_2^{x_2} \dots p_r^{x_r}$$

Also, notice that $X_1 + X_2 + \dots + X_r = n$. We have $\mathbf{X} \sim M(n, \mathbf{p})$.

Example 5.1

Roll a die 20 times. We want to find the probability of $P(X_1 = 3, X_2 = 2, X_3 = 4, X_4 = 5, X_5 = 1, X_6 = 5)$

$$P(X_1 = 3, X_2 = 2, X_3 = 4, X_4 = 5, X_5 = 1, X_6 = 5) = \frac{20!}{3!2!4!5!1!5!} \frac{1}{6^3} \frac{1}{6^2} \frac{1}{6^4} \frac{1}{6^5} \frac{1}{6^1} \frac{1}{6^5}$$

Now, let's examine the MGF of $\mathbf{X} \sim M(n, \mathbf{p})$.

$$M_{\mathbf{X}}(\mathbf{t}) = E e^{\mathbf{t}^\top \mathbf{X}} = E e^{t_1 X_1 + \dots + t_r X_r} = \sum_{X_1} \dots \sum_{X_r} e^{t_1 X_1 + \dots + t_r X_r} \frac{n!}{x_1! \dots x_r!} p_1^{x_1} \dots p_r^{x_r}$$

in which $X_1 + \dots + X_r = n$. Then, rearranging the expression, we obtain

$$M_{\mathbf{X}}(\mathbf{t}) = \sum_{X_1} \dots \sum_{X_r} \frac{n!}{x_1! \dots x_r!} (p_1 e^{t_1})^{X_1} \dots (p_r e^{t_r})^{X_r}$$

Using the multinomial theorem, we have

$$M_{\mathbf{X}}(\mathbf{t}) = (p_1 e^{t_1} + \dots + p_r e^{t_r})^n$$

Question 5.1. Find $M_{X_1}(t_1)$.

By setting every other X_i to 0, we have

$$M_{X_1}(t_1) = M_{\mathbf{X}}(t_1, 0, 0, \dots, 0) = (p_1 e^{t_1} + p_2 + \dots + p_r)^n = (p_1 e^{t_1} + 1 - p_1)^n$$

$$\implies X_1 \sim b(n, p_1).$$

Exercise 5.1. Find the mean vector and variance covariance matrix of X . Using Handout #10 to find $\text{var}(X_i), \text{cov}(X_i, X_j)$.

§5.2 Chi-Squared Distribution

Definition 5.2 (Chi-Squared) — Let $Z \sim N(0, 1)$ and $X = Z^2$. Then we say that $X \sim \mathcal{X}_1^2$. Notice that the subscript denotes the degree of freedom.

Now, let's find the pdf of \mathcal{X}_1^2 .

- Method of CDF:

$$\begin{aligned}
 F_X(x) &= P(X \leq x) = P(Z^2 \leq x) \\
 &= P(-\sqrt{x} \leq Z \leq \sqrt{x}) \\
 &= P(Z \leq \sqrt{x}) - P(Z \leq -\sqrt{x}) \\
 &= F_Z(\sqrt{x}) - F_Z(-\sqrt{x})
 \end{aligned}$$

So, $f_X(x) = \frac{1}{2}x^{-\frac{1}{2}}f_Z(\sqrt{x}) + \frac{1}{2}x^{-\frac{1}{2}}f_Z(-\sqrt{x})$, i.e.,

$$f_X(x) = \frac{x^{-\frac{1}{2}}e^{-\frac{x}{2}}}{\sqrt{\pi}\sqrt{2}} = \frac{x^{-\frac{1}{2}}e^{-\frac{x}{2}}}{\Gamma(\frac{1}{2})\sqrt{2}}$$

We can observe that \mathcal{X}_1^2 follows the same distribution as $\Gamma(\frac{1}{2}, 2)$.

For \mathcal{X}_n^2 , let $Z_1, \dots, Z_n \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$. We want to find the pdf of $\sum Z_i^2$. First, the MGF of Z_i^2 is

$$M_{Z_i^2}(t) = (1 - 2t)^{-\frac{1}{2}}$$

As Z_i are independent, we have

$$M_{\sum Z_i^2}(t) = \left(M_{Z_i^2}(t)\right)^n = (1 - 2t)^{-\frac{n}{2}}$$

Similar to the case of degree of freedom equals to 1, we deduce that \mathcal{X}_n^2 is the same as $\Gamma(\frac{n}{2}, 2)$.

Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$. Let's form a \mathcal{X}^2 distribution. Since $\sum Z_i^2 \sim \mathcal{X}_n^2$, it follows that

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

For $X \sim \mathcal{X}_n^2$ or $X \sim \Gamma(\frac{n}{2}, 2)$, we can easily deduce that the mean and variance are

$$EX = n$$

$$\text{var}(X) = 2n$$

Distribution of Quadratic Forms of Normally Distributed RV:

a) Let $Z_1, Z_2, \dots, Z_n \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$. Then $\mathbf{Z} \sim N(\mathbf{0}, I)$.

$$\sum Z_i^2 \sim \mathcal{X}_n^2 \quad \text{or} \quad \mathbf{Z}^\top \mathbf{Z} \sim \mathcal{X}_n^2$$

b) $\mathbf{Z} \sim N(\mathbf{0}, \sigma^2 I)$.

$$\sum \frac{Z_i^2}{\sigma^2} \sim \mathcal{X}_n^2 \quad \text{or} \quad \frac{\mathbf{Z}^\top \mathbf{Z}}{\sigma^2} \sim \mathcal{X}_n^2$$

c) $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 I)$

$$\sum \left(\frac{Y_i - \mu}{\sigma}\right)^2 \sim \mathcal{X}_n^2 \quad \text{or} \quad \frac{(\mathbf{Y} - \boldsymbol{\mu})^\top (\mathbf{Y} - \boldsymbol{\mu})}{\sigma^2} \sim \mathcal{X}_n^2$$

d) $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\mathbf{V} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y} - \boldsymbol{\mu})$.

$$E\mathbf{V} = \boldsymbol{\Sigma}^{-\frac{1}{2}}E[\mathbf{Y} - \boldsymbol{\mu}] = \mathbf{0}$$

$$\text{var}(\mathbf{V}) = \text{var}\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y}\right)$$

$$= \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-\frac{1}{2}} = I$$

So, $\mathbf{V} \sim N(\mathbf{0}, I)$. From a), we have $\mathbf{V}^\top \mathbf{V} \sim \mathcal{X}_n^2$. Thus,

$$(\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \sim \mathcal{X}_n^2$$

Theorem 5.3

Let $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$. We define sample variance as follows

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

Then, $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$.

From the above result, we know that $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2$. We want to show

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$$

We have

$$\begin{aligned} \sum \left(\frac{X_i - \mu \pm \bar{X}}{\sigma} \right)^2 &= \frac{\sum (X_i - \bar{X} + \bar{X} - \mu)^2}{\sigma^2} \\ &= \frac{\sum (X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} + \frac{2(\bar{X} - \mu) \sum (X_i - \bar{X})}{\sigma^2} \end{aligned}$$

Note that $\sum (X_i - \bar{X}) = \sum X_i - n\bar{X} = 0$. Then,

$$\sum \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2$$

Notice that $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ and thus $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$. In addition, \bar{X} and S^2 are independent. Consider

$$\begin{pmatrix} X_1 - \bar{X} \\ X_2 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix} = \left(I - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \right) \mathbf{X}$$

Then,

$$\begin{pmatrix} X_1 - \bar{X} \\ X_2 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix} = \begin{pmatrix} \frac{1}{n} \mathbf{1} \\ I - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \end{pmatrix} \mathbf{X} = \mathbf{A}\mathbf{X}$$

But $\mathbf{X} \sim N(\boldsymbol{\mu}, \sigma^2 I)$ as $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$. Then,

$$\begin{aligned} \text{var}(\mathbf{A}\mathbf{X}) &= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top \\ &= \begin{pmatrix} \frac{1}{n} \mathbf{1}^\top \\ I - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \end{pmatrix} \sigma^2 I \begin{pmatrix} \frac{1}{n} \mathbf{1} & I - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \end{pmatrix} \end{aligned}$$

Note that $I - \frac{1}{n} \mathbf{1}\mathbf{1}^\top$ is symmetric and idempotent.

$$\text{var}(\mathbf{A}\mathbf{X}) = \sigma^2 \begin{pmatrix} \frac{1}{n} & \mathbf{0}^\top \\ \mathbf{0} & I - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \end{pmatrix}$$

\bar{X} is independent of $(X_1 - \bar{X} \dots X_n - \bar{X})^\top$ and therefore is independent of S^2 .

Exercise 5.2. Verify $\text{cov}(\bar{X}, X_1 - \bar{X}) = 0$.

We have

$$\begin{aligned}\text{cov}(\bar{X}, X_1 - \bar{X}) &= \text{cov}(\bar{X}, X_1) - \text{cov}(\bar{X}, \bar{X}) \\ &= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0\end{aligned}$$

Also,

$$\begin{aligned}\text{var}(X_1 - \bar{X}) &= \text{var}(X_1) + \text{var}(\bar{X}) - 2\text{cov}(X_1, \bar{X}) \\ &= \sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n} \\ &= \sigma^2 \left(1 - \frac{1}{n}\right)\end{aligned}$$

and

$$\begin{aligned}\text{cov}(X_1 - \bar{X}, X_2 - \bar{X}) &= \text{cov}(X_1, X_2) - \text{cov}(X_1, \bar{X}) - \text{cov}(\bar{X}, X_2) + \text{cov}(\bar{X}, \bar{X}) \\ &= 0 - \frac{\sigma^2}{n} - \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = -\frac{\sigma^2}{n}\end{aligned}$$

Proof. (of the above theorem) We have

$$\underbrace{\sum \left(\frac{X_i - \mu}{\sigma} \right)^2}_Q = \underbrace{\frac{(n-1)S^2}{\sigma^2}}_{Q_1} + \underbrace{\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2}_{Q_2}$$

Using MGF,

$$\begin{aligned}M_Q(t) &= M_{Q_1}(t) \cdot M_{Q_2}(t) \\ M_{Q_1}(t) &= \frac{M_Q(t)}{M_{Q_2}(t)} \\ &= \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}} \\ &= (1-2t)^{-\frac{n-1}{2}}\end{aligned}$$

Therefore, $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$. □

Summary:

1. $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2$
2. $\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi_{n-1}^2$.

This is known as central chi-squared distribution.

§5.3 Non-central Chi-Squared Distribution

Definition 5.4 (Non-central χ^2) — If $Y \sim N(\mu, 1)$, then $Y^2 \sim \chi_1^2$ (NCP = μ^2) where NCP means non-centrality parameter.

Let $Y \sim N(\mu, \sigma)$. Then $\frac{Y}{\sigma} \sim N\left(\frac{\mu}{\sigma}, 1\right)$. Thus, $\frac{Y^2}{\sigma^2} \sim \chi_1^2$ (NCP = $\frac{\mu^2}{\sigma^2}$).
 Let's find the MGF of χ_1^2 (NCP = θ). Let $Q \sim \chi_1^2$ (NCP = θ).

$$M_Q(t) = (1 - 2t)^{-\frac{1}{2}} e^{\theta \frac{t}{1-2t}}$$

Let $Y_1, Y_2, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$. Then

$$M_{\frac{Y_1^2}{\sigma^2} + \frac{Y_2^2}{\sigma^2} + \dots + \frac{Y_n^2}{\sigma^2}}(t) = \left((1 - 2t)^{-\frac{1}{2}} e^{\frac{\mu^2}{\sigma^2} \frac{t}{1-2t}} \right)^n = (1 - 2t)^{-\frac{n}{2}} e^{\frac{n\mu^2}{\sigma^2} \frac{t}{1-2t}}$$

Thus, $\sum \frac{Y_i^2}{\sigma^2} \sim \chi_n^2$ (NCP = $\frac{n\mu^2}{\sigma^2}$).

Example 5.5

$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu_1, 1)$ and $Y_1, \dots, Y_m \stackrel{\text{i.i.d.}}{\sim} N(\mu_2, 1)$. The two samples X and Y are independent.

a) Find the distribution of W where

$$W = \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2$$

We have $\frac{(n-1)S_X^2}{1} \sim \chi_{n-1}^2$ and $\frac{(m-1)S_Y^2}{1} \sim \chi_{m-1}^2$. Note that

$$\left. \begin{array}{l} X \sim \chi_n^2 \\ Y \sim \chi_m^2 \\ X, Y \text{ are independent} \end{array} \right\}$$

Then

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= (1 - 2t)^{-\frac{n}{2}} (1 - 2t)^{-\frac{m}{2}} \\ &= (1 - 2t)^{-\frac{n+m}{2}} \end{aligned}$$

Thus, $X + Y \sim \chi_{n+m}^2$. So,

$$\frac{(n-1)S_X^2}{1} + \frac{(m-1)S_Y^2}{1} \sim \chi_{n+m-2}^2$$

b) The mean of W is $n + m - 2$.

c) The variance of W is $2(n + m - 2)$.

§5.4 t-Distribution

Let $Z \sim N(0, 1)$ and $U \sim \chi_{df}$. If Z, U are independent, then

$$\frac{Z}{\sqrt{\frac{U}{df}}} \sim t_{df}$$

Application: Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$. Then

$$\left. \begin{array}{l} \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \\ \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 \end{array} \right\} \Rightarrow \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/n-1}} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

Summary:

$$\begin{aligned} \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} &\sim N(0, 1) \\ \frac{\bar{X}-\mu}{S/\sqrt{n}} &\sim t_{n-1} \end{aligned}$$

§6 | Lec 6: Aug 19, 2021

§6.1 t Distribution (Cont'd)

Example 6.1

$X_1, \dots, X_{10} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma)$. Find c s.t. $P(\frac{\bar{X}}{\sqrt{9S_X^2}} < c) = 0.95$.

$$P\left[\frac{\frac{\bar{X}-0}{\sigma/\sqrt{10}}}{\sqrt{\frac{9S_X^2}{\sigma^2}}/9} < \sqrt{90}c\right] = 0.95$$

$$P(t_9 < \sqrt{90}c) = 0.95$$

As $t_{0.95;9} = \sqrt{90}c = 1.833 \implies c = 0.19$.

Example 6.2

$X_1, X_2, X_3, X_4, X_5 \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma)$. Find c so that $\frac{c(X_1 - X_2)}{\sqrt{X_3^2 + X_4^2 + X_5^2}}$ follows t distribution.

We have

$$X_1 - X_2 \sim N(0, \sigma\sqrt{2})$$

$$\frac{X_3^2}{\sigma^2} + \frac{X_4^2}{\sigma^2} + \frac{X_5^2}{\sigma^2} \sim \chi_3^2$$

So

$$\frac{\frac{X_1 - X_2 - 0}{\sigma\sqrt{2}}}{\sqrt{\frac{X_3^2 + X_4^2 + X_5^2}{\sigma^2}}/3} = \sqrt{3} \frac{X_1 - X_2}{\sqrt{X_3^2 + X_4^2 + X_5^2}} \sim t_3$$

Example 6.3

Let $X_1, \dots, X_{10} \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$ and $Y_1, \dots, Y_{10} \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$ and $W_1, \dots, W_{10} \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$ where X, Y, W are independent. Find c s.t.

$$P\left(\frac{\bar{X} + \bar{Y} - 2\bar{W}}{\sqrt{9S_X^2 + 9S_Y^2 + 9S_W^2}} < c\right) = 0.95$$

We have

$$E[\bar{X} + \bar{Y} - 2\bar{W}] = \mu + \mu - 2\mu = 0$$

$$\text{var}(\bar{X} + \bar{Y} - 2\bar{W}) = \frac{\sigma^2}{10} + \frac{\sigma^2}{10} + \frac{4\sigma^2}{10} = \frac{6\sigma^2}{10}$$

$\implies \bar{X} + \bar{Y} - 2\bar{W} \sim N(0, \sigma\sqrt{\frac{6}{10}})$. Also, $\frac{9S_X^2}{\sigma^2} + \frac{9S_Y^2}{\sigma^2} + \frac{9S_W^2}{\sigma^2} \sim \chi_{27}^2$.

Example 6.4 (Cont'd from above)

$$P\left(\frac{\frac{\bar{X} + \bar{Y} - 2\bar{W} - 0}{\sigma\sqrt{\frac{6}{10}}}}{\sqrt{\frac{9S_X^2 + 9S_Y^2 + 9S_W^2}{\sigma^2}}/27} < \sqrt{\frac{270}{6}}c\right) = 0.95$$

$$P\left(t_{27} < \sqrt{\frac{270}{6}}c\right) = 0.95$$

$$\Rightarrow t_{0.95;27} = \sqrt{\frac{270}{6}}c = 1.703$$

Definition 6.5 (Non-Central t Distribution) — Let $U \sim N(\delta, 1)$ and $V \sim \chi_n^2$ where U, V are independent. Then $\frac{U}{\sqrt{\frac{V}{n}}} \sim t_n(\text{NCP} = \delta)$.

§6.2 F Distribution

Definition 6.6 (F Distribution) — Let $U \sim \chi_n^2$ and $V \sim \chi_m^2$. If U, V are independent, then $\frac{\frac{U}{n}}{\frac{V}{m}} \sim F_{n,m}$.

Application: Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu_1, \sigma_1)$ and $Y_1, \dots, Y_m \stackrel{\text{i.i.d.}}{\sim} N(\mu_2, \sigma_2)$. We have

$$\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi_{n-1}^2 \quad \text{and} \quad \frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi_{m-1}^2$$

$$\frac{\frac{(n-1)S_1^2}{\sigma_1^2}/(n-1)}{\frac{(m-1)S_2^2}{\sigma_2^2}/(m-1)} = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F_{n-1, m-1}$$

Special Case: Suppose $\sigma_1^2 = \sigma_2^2$. Then $\frac{S_1^2}{S_2^2} \sim F_{n-1, m-1}$.

Properties:

1. If $X \sim F_{n,m}$, then $\frac{1}{X} \sim F_{m,n}$.
2. $F_{\alpha; m, n} = \frac{1}{F_{1-\alpha; n, m}}$.
3. Relationship to t distribution

$$t = \frac{Z}{\sqrt{\frac{U}{n}}}$$

$$Z \sim N(0, 1) \quad \text{and} \quad U \sim \chi_n^2$$

Then, $\frac{Z^2}{\frac{U}{n}} \sim F_{1,n}$; thus, $t_n^2 = F_{1,n}$.

Definition 6.7 (Non-Central F Distribution) — Let $U \sim \chi_n^2(\text{NCP} = \theta)$ and $V \sim \chi_m^2$. If U, V are independent, then $\frac{\frac{U}{n}}{\frac{V}{m}} \sim F_{n,m}(\text{NCP} = \theta)$.

§6.3 Properties of Estimators

Unbiased Estimators: Let θ be a parameter of a distribution and let $\hat{\theta}$ be an estimator of θ . We say that $\hat{\theta}$ is unbiased if $E\hat{\theta} = \theta$.

Example 6.8

X_1, \dots, X_n i.i.d random variables. Since $E\bar{X} = \mu$ and $EX_i = \mu$, \bar{X} has the unbiased properties.

Let $S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$ be the sample variance. Let's find ES^2 .

$$\begin{aligned}
 E \sum (X_i - \bar{X})^2 &= E \left[\sum (X_i - \mu - (\bar{X} - \mu))^2 \right] \\
 &= E \left[\sum (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum (X_i - \mu) \right] \\
 &= E \left[\sum (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2n(\bar{X} - \mu)^2 \right] \\
 &= E \left[\sum (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \right] \\
 &= \sum E(X_i - \mu)^2 - nE(\bar{X} - \mu)^2 \\
 &= n\sigma^2 - n\frac{\sigma^2}{n} \\
 &= (n-1)\sigma^2
 \end{aligned}$$

Thus, $ES^2 = E \frac{\sum (X_i - \bar{X})^2}{n-1} = \frac{(n-1)\sigma^2}{n-1} = \sigma^2$.

Example 6.9

Consider $\hat{p} = \frac{X}{n}$ where X_1, X_2, \dots, X_n i.i.d Bernoulli R.V. and $X = \#$ of successes among n trials, $X \sim b(n, p)$.

$$E\hat{p} = E \frac{X}{n} = \frac{np}{n} = p$$

Example 6.10

X_1, \dots, X_n i.i.d $N(\mu_1, \sigma_1)$ and Y_1, \dots, Y_m i.i.d $N(\mu_2, \sigma_2)$.

$$E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2$$

Efficient Estimators: Let θ be a parameter of a distribution and let $\hat{\theta}$ be an (unbiased) estimator of θ . Then,

$$\begin{aligned}
 \text{var}(\hat{\theta}) &\geq \frac{1}{nE \left[\frac{\partial \ln f(x; \theta)}{\partial \theta} \right]^2} \\
 &\geq \frac{1}{-nE \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2}}
 \end{aligned}$$

Note: Have

$$\begin{aligned}
 \int_x f(x; \theta) dx &= 1 \\
 \int_x \frac{\partial f(x; \theta)}{\partial \theta} dx &= 0 \\
 \int_x \frac{\partial f(x; \theta)}{\partial \theta} \frac{f(x; \theta)}{f(x; \theta)} dx &= 0 \\
 \int_x \frac{\partial \ln f(x; \theta)}{\partial \theta} f(x; \theta) dx &= 0 \\
 \text{or } E \frac{\partial \ln f(x; \theta)}{\partial \theta} &= 0
 \end{aligned} \tag{*}$$

We define the score function as follows

$$S = \frac{\partial \ln f(x; \theta)}{\partial \theta}$$

where $ES = 0$. Take derivative w.r.t. θ for (*) and we obtain

$$\begin{aligned}
 \int \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} f(x; \theta) dx + \int \frac{\partial \ln f(x; \theta)}{\partial \theta} \cdot \frac{\partial f(x; \theta)}{\partial \theta} dx &= 0 \\
 \int \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} f(x; \theta) dx + \int \frac{\partial \ln f(x; \theta)}{\partial \theta} \frac{\partial f(x; \theta)}{\partial \theta} \frac{f(x; \theta)}{f(x; \theta)} dx &= 0 \\
 - \int \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} f(x; \theta) dx &= \int \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 f(x; \theta) dx
 \end{aligned}$$

Thus, $-E \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} = E \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 = I(\theta)$ – information in one of the observation and $nI(\theta)$ is the information in the sample. Another way to find $I(\theta)$ is $\text{var}(S) = I(\theta)$.

$$\text{var}(S) = ES^2 - (ES)^2$$

But $ES = 0$, so $\text{var}(S) = ES^2 = E \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2$.

$\hat{\theta}$ is an efficient estimator if $\text{var}(\hat{\theta}) = \frac{1}{nI(\theta)}$.

Example 6.11

Let X_1, \dots, X_n i.i.d follows $N(\mu, \sigma)$. Is \bar{X} an efficient estimator of μ ?

$$\begin{aligned}
 E\bar{X} &= \mu \quad \text{unbiased} \\
 \text{var}(\bar{X}) &= \frac{\sigma^2}{n}
 \end{aligned}$$

1. First method:

$$\begin{aligned}
 f(x) &= (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \\
 \ln f(x) &= -\frac{1}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}(x-\mu)^2
 \end{aligned}$$

Example 6.12 (Cont'd from above)

Then,

$$I(\theta) = E \left[\frac{\partial \ln f(x)}{\partial \theta} \right]^2$$

$$\frac{\partial \ln f(x)}{\partial \mu} = \frac{2}{2\sigma^2}(x - \mu) = \frac{x - \mu}{\sigma^2}$$

$$I(\theta) = E \left[\frac{X - \mu}{\sigma^2} \right]^2 = \frac{E[X - \mu]^2}{\sigma^2} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

2. Second method:

$$I(\theta) = -E \left[\frac{\partial^2 \ln f(x)}{\partial \mu^2} \right] = - \left(-\frac{1}{\sigma^2} \right) = \frac{1}{\sigma^2}$$

3. Third method

$$I(\theta) = \text{var}(S) = \text{var} \left(\frac{X - \mu}{\sigma^2} \right) = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

Now, the Cramer-Rao lower bound is

$$\text{var}(\hat{\mu}) \geq \frac{1}{nI(\theta)} = \frac{1}{n \frac{1}{\sigma^2}} = \frac{\sigma^2}{n}$$

Our estimator is \bar{X} which has $\text{var}(\bar{X}) = \frac{\sigma^2}{n}$ (same as the Cramer-Rao lower bound). Thus, \bar{X} is an efficient estimator of μ .

Example 6.13

Is $S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$ an efficient estimator of σ^2 ?

$$ES^2 = \sigma^2$$

$$\text{var}(S^2) = \frac{2\sigma^4}{n-1}$$

Cramer-Rao lower bound:

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

$$\frac{\partial \ln f(x)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(x - \mu)^2$$

$$\frac{\partial^2 \ln f(x)}{\partial \sigma^2} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x - \mu)^2$$

$$I(\theta) = -E \left[\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x - \mu)^2 \right]$$

$$= -\frac{1}{2\sigma^4} + \frac{1}{\sigma^4} = \frac{1}{2\sigma^4}$$

Thus, $\text{var}(\sigma^2) \geq \frac{1}{nI(\theta)} = \frac{2\sigma^4}{n}$. Our estimator S^2 has variance $\frac{2\sigma^4}{n-1}$. So S^2 is not an efficient estimator of σ^2 (asymptotically efficient estimator for large enough n).

§7 | Lec 7: Aug 24, 2021

§7.1 Properties of Estimators (Cont'd)

Let X_1, X_2, \dots, X_n i.i.d R.V. with pdf $f(x; \theta)$. Then the joint pdf is

$$\begin{aligned} L &= f(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \dots f(x_n; \theta) \\ \ln L &= \ln f(x_1; \theta) + \ln f(x_2; \theta) + \dots + \ln f(x_n; \theta) \\ \frac{\partial \ln L}{\partial \theta} &= \frac{\partial \ln f(x_1; \theta)}{\partial \theta} + \dots + \frac{\partial \ln f(x_n; \theta)}{\partial \theta} \end{aligned}$$

Information in the sample is

$$\begin{aligned} E \left(\frac{\partial \ln L}{\partial \theta} \right)^2 &= E \left(\frac{\partial \ln f(x_1; \theta)}{\partial \theta} + \dots + \frac{\partial \ln f(x_n; \theta)}{\partial \theta} \right)^2 \\ &= E \left\{ \left(\frac{\partial \ln f(x_1; \theta)}{\partial \theta} \right)^2 + \dots + \left(\frac{\partial \ln f(x_n; \theta)}{\partial \theta} \right)^2 + 2 \frac{\partial \ln f(x_1; \theta)}{\partial \theta} \frac{\partial \ln f(x_2; \theta)}{\partial \theta} + \dots \right\} \\ &= I(\theta) + \dots + I(\theta) + 0 = nI(\theta) \end{aligned}$$

Recall the Cramer-Rao Inequality:

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(\theta)}$$

Let's prove this inequality.

Proof. Let X_1, X_2, \dots, X_n i.i.d R.V. with pdf $f(x; \theta)$. Let $\hat{\theta} = g(x_1, x_2, \dots, x_n)$ be an unbiased estimator of θ . Let assume $n = 2$. Then $\hat{\theta} = g(x_1, x_2)$ and

$$\begin{aligned} E\hat{\theta} &= \iint g(x_1, x_2) f(x_1, x_2; \theta) dx_1 dx_2 = \theta \\ &= \iint g(x_1, x_2) f(x_1; \theta) f(x_2; \theta) dx_1 dx_2 \end{aligned}$$

Now, we take the derivative w.r.t. θ on both sides

$$\begin{aligned} \iint g(x_1, x_2) \frac{\partial f(x_1; \theta)}{\partial \theta} f(x_2; \theta) \frac{f(x_1; \theta)}{f(x_1; \theta)} dx_1 dx_2 + \iint g(x_1, x_2) f(x_1; \theta) \frac{\partial f(x_2; \theta)}{\partial \theta} \frac{f(x_2; \theta)}{f(x_2; \theta)} dx_1 dx_2 &= 1 \\ \iint g(x_1, x_2) \frac{\partial \ln f(x_1; \theta)}{\partial \theta} f(x_1; \theta) f(x_2; \theta) dx_1 dx_2 + \iint g(x_1, x_2) \frac{\partial \ln f(x_2; \theta)}{\partial \theta} f(x_1; \theta) f(x_2; \theta) dx_1 dx_2 &= 1 \\ \iint g(x_1, x_2) \left[\frac{\partial \ln f(x_1; \theta)}{\partial \theta} + \frac{\partial \ln f(x_2; \theta)}{\partial \theta} \right] f(x_1; \theta) f(x_2; \theta) dx_1 dx_2 &= 1 \end{aligned}$$

Generalize this to arbitrary n and we get

$$\int \dots \int g(x_1, \dots, x_n) \sum_{i=1}^n \frac{\partial \ln f(x_i; \theta)}{\partial \theta} f(x_1; \theta) \dots f(x_n; \theta) dx_1 \dots dx_n = 1$$

Let $Q = \sum_{i=1}^n \frac{\partial \ln f(x_i; \theta)}{\partial \theta}$. Then, $E[\hat{\theta}Q] = 1$. Now, consider $\rho_{\hat{\theta}Q}$

$$\begin{aligned} -1 &\leq \rho \leq 1 \\ \rho_{\hat{\theta}Q}^2 &\leq 1 \\ \frac{\text{cov}^2(\hat{\theta}, Q)}{\text{var}(\hat{\theta}) \text{var}(Q)} &\leq 1 \\ \frac{[E\hat{\theta}Q - (E\hat{\theta})(EQ)]^2}{\text{var}(\hat{\theta}) [EQ^2 - (EQ)^2]} &\leq 1 \end{aligned}$$

Note that

$$\begin{aligned} E[\hat{\theta}Q] &= 1 \\ EQ &= 0 \\ EQ^2 &= nI(\theta) \end{aligned}$$

Plug these into the inequality to obtain

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(\theta)} \quad \square$$

Relative Efficiency: Suppose X_1, X_2, \dots, X_n i.i.d R.V. with pdf $f(x; \theta)$. Then,

$$\begin{aligned} EX_1 = \theta &\implies \text{var}(X_1) = \sigma^2 \\ E\left[\frac{X_1 + X_2}{2}\right] = \theta &\implies \text{var}\left(\frac{X_1 + X_2}{2}\right) = \frac{\sigma^2}{2} \\ &\vdots \\ E\bar{X} = \theta &\implies \text{var}(\bar{X}) = \frac{\sigma^2}{n} \end{aligned}$$

So unbiased is not an unique property, we choose the one with the smallest variance. In this case, we choose \bar{X} .

Example 7.1

$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d}}{\sim} \text{Poisson}(\lambda)$ and consider

$$\begin{aligned} \hat{\theta}_1 &= \frac{X_1 + X_2}{2} \\ \hat{\theta}_2 &= \bar{X} \end{aligned}$$

Then,

$$\begin{aligned} E\hat{\theta}_1 &= \lambda; \quad \text{var}(\hat{\theta}_1) = \frac{\lambda}{2} \\ E\hat{\theta}_2 &= \lambda; \quad \text{var}(\hat{\theta}_2) = \frac{\lambda}{n} \end{aligned}$$

The ratio between the variances is

$$\frac{\text{var}(\hat{\theta}_2)}{\text{var}(\hat{\theta}_1)} = \frac{\lambda/n}{\lambda/2} = \frac{2}{n}$$

Suppose $n = 10$. Then the variability associated with $\hat{\theta}_2$ is 20% of the variability associated with $\hat{\theta}_1$.

Consistency:

Definition 7.2 (Consistency/Converge in Probability) — $P\left(\left|\hat{\theta} - \theta\right| < \varepsilon\right) \rightarrow 1$ as $n \rightarrow \infty$ or $P\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right) \rightarrow 0$ as $n \rightarrow \infty$.

Theorem 7.3

Let $\hat{\theta}$ be an unbiased estimator of θ . Then, $\hat{\theta}$ is consistent if $\text{var}(\hat{\theta}) \rightarrow 0$ as $n \rightarrow \infty$.

Example 7.4

Consider \bar{X}

$$E\bar{X} = \mu$$

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$$

\Rightarrow consistent.

Proof. Let X be a random variable with mean μ and variance σ^2 . For $k > 0$,

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

and

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

by Chebyshev's Inequality. Suppose $X \sim N(\mu, \sigma)$. Then,

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95\%$$

Using Chebyshev's Inequality with $k = 2$ we get

$$P(|X - \mu| \leq 2\sigma) \geq 1 - \frac{1}{2^2} = 75\%$$

Back to the proof, we have

$$P\left(|\hat{\theta} - \theta| > k\sqrt{\text{var}(\hat{\theta})}\right) \leq \frac{1}{k^2}$$

Let $\varepsilon = k\sqrt{\text{var}(\hat{\theta})}$. Then,

$$P\left(|\hat{\theta} - \theta| > \varepsilon\right) \leq \frac{\text{var}(\hat{\theta})}{\varepsilon^2}$$

If $\text{var}(\hat{\theta}) \xrightarrow{n \rightarrow \infty} 0$, we conclude that $\hat{\theta}$ is consistent. \square

Example 7.5

$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$.

$$ES^2 = \sigma^2$$

$$\text{var}(S^2) = \frac{2\sigma^4}{n-1}$$

As $\text{var}(S^2) \xrightarrow{n \rightarrow \infty} 0$, then S^2 is consistent.

Example 7.6

X_1, \dots, X_n i.i.d Bernoulli R.V.

$$\begin{aligned}\hat{p} &= \frac{X}{n} \\ E \frac{X}{n} &= p \\ \text{var}(\frac{X}{n}) &= \frac{p(1-p)}{n} \xrightarrow{n \rightarrow \infty} 0\end{aligned}$$

$$X = X_1 + \dots + X_n \implies X \sim b(n, p).$$

§7.2 Bias and Mean Square Error

First, let's define the bias term as $B = E(\hat{\theta}) - \theta$. If $B = 0$ then $\hat{\theta}$ is unbiased. Let's define the mean square error (MSE)

$$\text{MSE} = E \left(\hat{\theta} - \theta \right)^2 = \text{var}(\hat{\theta}) + B^2$$

as a measure of the quality of an estimator. If $\hat{\theta}$ is unbiased, then $\text{MSE} = \text{var}(\hat{\theta})$. Let's show that $E(\hat{\theta} - \theta)^2 = \text{var}(\hat{\theta}) + B^2$.

$$\begin{aligned}E \left(\hat{\theta} - \theta \right)^2 &= E \left(\hat{\theta} - \theta \pm E\hat{\theta} \right)^2 \\ &= E \left(\hat{\theta} - E\hat{\theta} + \underbrace{E\hat{\theta} - \theta}_B \right)^2 \\ &= E \left(\hat{\theta} - E\hat{\theta} \right)^2 + B^2 + 2B \underbrace{E \left(\hat{\theta} - E\hat{\theta} \right)}_0 \\ E \left(\hat{\theta} - \theta \right)^2 &= \text{var}(\hat{\theta}) + B^2\end{aligned}$$

Example 7.7

$X_1, \dots, X_n \stackrel{\text{i.i.d}}{\sim} U(\theta, \theta + 1)$.

- a) Show that \bar{X} is a biased estimator of θ and compute the bias.

Consider \bar{X} as an estimator of θ .

$$\begin{aligned}E\bar{X} &= \mu = \frac{\theta + \theta + 1}{2} = \theta + \frac{1}{2} \\ \implies B &= \frac{1}{2}\end{aligned}$$

- b) Find a function of \bar{X} that is unbiased estimator of θ .

It can be easily deduced that $\hat{\theta}_1 = \bar{X} - \frac{1}{2}$ is unbiased as $E\hat{\theta}_1 = \theta$.

Example 7.8 c) Find the MSE when \bar{X} is used as an estimator of θ .

$$\text{MSE}(\bar{X}) = \text{var}(\bar{X}) + B^2 = \frac{\sigma^2}{n} + \frac{1}{4} = \frac{(\theta + 1 - \theta)^2}{12n} + \frac{1}{4} = \frac{1}{12n} + \frac{1}{4}$$

d) Find the MSE of $\hat{\theta}_1$.

$$\begin{aligned} \text{MSE}(\hat{\theta}_1) &= \text{var}(\hat{\theta}_1) \\ &= \text{var}\left(\bar{X} - \frac{1}{2}\right) = \text{var}(\bar{X}) \\ &= \frac{\sigma^2}{n} = \frac{1}{12n} \end{aligned}$$

Example 7.9

$X \sim b(100, p)$. Consider

$$\hat{p}_1 = \frac{X}{100}, \quad \hat{p}_2 = \frac{X+3}{100}, \quad \hat{p}_3 = \frac{X+3}{106}$$

Find the MSE of each \hat{p}_i .

\hat{p}_1 is unbiased so

$$\begin{aligned} \text{MSE}(\hat{p}_1) &= \text{var}(\hat{p}_1) = \text{var}\left(\frac{X}{100}\right) \\ &= \frac{p(1-p)}{100} \end{aligned}$$

For \hat{p}_2 , we have

$$\begin{aligned} B &= E\hat{p}_2 - p \\ &= E\frac{X+3}{100} - p = \frac{3}{100} \end{aligned}$$

and

$$\begin{aligned} \text{MSE}(\hat{p}_2) &= \text{var}(\hat{p}_2) + B^2 \\ &= \text{var}\left(\frac{X+3}{100}\right) + \frac{3^2}{100^2} \\ &= \frac{p(1-p)}{100} + \left(\frac{3}{100}\right)^2 \end{aligned}$$

Similarly for \hat{p}_3

$$B = E\frac{X+3}{106} - p = \frac{100p+3}{106} - p$$

and

$$\begin{aligned} \text{MSE}(\hat{p}_3) &= \text{var}\left(\frac{X+3}{106}\right) - \left(\frac{100p+3}{106} - p\right)^2 \\ &= \frac{100p(1-p)}{106^2} - \left(\frac{100p+3}{106} - p\right)^2 \end{aligned}$$

§7.3 Method of Maximum Likelihood

This method requires a distribution assumption. Let X_1, X_2, \dots, X_n be i.i.d R.V. with pdf $f(x; \theta)$. The joint pdf is also called the likelihood function is

$$L = f(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \dots f(x_n; \theta)$$

Our goal here is to maximize the likelihood function L w.r.t. θ to find $\hat{\theta}$. To make the computation easier, we maximize the log likelihood by solving

$$\frac{\partial \ln L}{\partial \theta} = 0$$

Example 7.10

$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d}}{\sim} \text{Poisson}(\lambda)$. Find the MLE of λ .

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Then, $L = P(x_1; \lambda) \cdot P(x_2; \lambda) \dots P(x_n; \lambda)$

$$L = \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \dots \frac{\lambda^{x_n} e^{-\lambda}}{x_n!} = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod x_i!}$$

So, taking log-likelihood we get

$$\ln L = \sum x_i \ln \lambda - n\lambda - \ln \prod x_i!$$

Maximize $\ln L$ w.r.t. λ we obtain

$$\begin{aligned} \frac{\partial \ln L}{\partial \lambda} &= \frac{\sum x_i}{\lambda} - n = 0 \\ \hat{\lambda} &= \frac{\sum x_i}{n} = \bar{x} \end{aligned}$$

Then, we can check for unbiased and consistent through

$$\begin{aligned} E\bar{X} &= \lambda \\ \text{var}(\bar{X}) &= \frac{\lambda}{n} \end{aligned}$$

Example 7.11

Let $X_1, \dots, X_n \stackrel{\text{i.i.d}}{\sim} \exp(\lambda)$. Find the MLE of λ .

$$L = \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \dots \lambda e^{-\lambda x_n}$$

$$L = \lambda^n e^{\lambda \sum x_i}$$

$$\ln L = n \ln \lambda - \lambda \sum x_i$$

Maximize $\ln L$,

$$\begin{aligned} \frac{\partial \ln L}{\partial \lambda} &= \frac{n}{\lambda} - \sum x_i = 0 \\ \hat{\lambda} &= \frac{n}{\sum x_i} = \frac{1}{\bar{x}} \end{aligned}$$

Example 7.12

$X_1, \dots, X_n \stackrel{\text{i.i.d}}{\sim} N(\mu, \sigma^2)$. Find the MLE of μ and σ^2 .

$$\begin{aligned} f(x_i; \mu, \sigma^2) &= \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ L &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2\right) \\ \ln L &= -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \end{aligned}$$

So we maximize $\ln L$ w.r.t. μ as follows

$$\begin{aligned} \frac{\partial \ln L}{\partial \mu} &= \frac{2}{2\sigma^2} \sum (x_i - \mu) = 0 \\ \sum x_i - n\mu &= 0 \\ \hat{\mu} &= \frac{\sum x_i}{n} = \bar{x} \end{aligned}$$

We maximize $\ln L$ w.r.t. σ^2 as follows

$$\begin{aligned} \frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4} \sum (x_i - \mu)^2 = 0 \\ \hat{\sigma}^2 &= \frac{\sum (x_i - \hat{\mu})^2}{n} \\ \text{or } \hat{\sigma}^2 &= \frac{\sum (x_i - \bar{x})^2}{n} \end{aligned}$$

which is biased. So we adjust the last expression to be unbiased by setting

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Note that

$$\begin{aligned} E\hat{\sigma}^2 &= E \frac{\sum (x_i - \bar{x})^2}{n} \\ &= \frac{1}{n} E \sum (x_i - \bar{x})^2 \\ &= \frac{(n-1)\sigma^2}{n} \end{aligned}$$

Adjust $\hat{\sigma}^2$ as follows. We need to find c s.t.

$$\begin{aligned} Ec\hat{\sigma}^2 &= \sigma^2 \\ cE\hat{\sigma}^2 &= \sigma^2 \\ c &= \frac{\sigma^2}{E\hat{\sigma}^2} = \frac{n}{n-1} \end{aligned}$$

Therefore, unbiased estimator of σ^2 is

$$S^2 = c\hat{\sigma}^2 = \frac{n}{n-1} \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

§8 | Lec 8: Aug 26, 2021

§8.1 Method of Maximum Likelihood (Cont'd)

Recall that we previously compute $I(\theta)$ through

$$I(\theta) = E \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 = -E \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2}$$

Using the log-likelihood function we find the information in the sample through $-E \frac{\partial^2 \ln L}{\partial \theta^2}$.

Example 8.1

$X_1, X_2, \dots, X_n \stackrel{\text{i.i.d}}{\sim} N(\mu, \theta)$. From lecture 7, we found that

$$I(\theta) = \frac{1}{\sigma^2} \implies nI(\theta) = \frac{n}{\sigma^2} - \text{information in the sample}$$

Let's find the information in the sample using the log likelihood function.

$$\begin{aligned} \ln L &= -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \\ \frac{\partial \ln L}{\partial \mu} &= \frac{1}{\sigma^2} \sum (x_i - \mu) \\ \frac{\partial^2 \ln L}{\partial \theta^2} &= -\frac{n}{\sigma^2} \\ -E \left(-\frac{n}{\sigma^2} \right) &= \frac{n}{\sigma^2} \end{aligned}$$

Asymptotic Properties of MLE

Let $\hat{\theta}$ be the MLE of θ . Then as $n \rightarrow \infty$,

$$\hat{\theta} \stackrel{\text{apprx}}{\sim} N \left(\theta, \sqrt{\frac{1}{nI(\theta)}} \right)$$

Let $X_1, \dots, X_n \stackrel{\text{i.i.d}}{\sim} N(\mu_1, \sigma)$ and $Y_1, \dots, Y_m \stackrel{\text{i.i.d}}{\sim} N(\mu_2, \sigma)$. The two samples are independent. We want to find the MLEs of μ_1, μ_2, σ^2 .

$$\begin{aligned} L &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu_1)^2} \cdot (2\pi\sigma^2)^{-\frac{m}{2}} e^{-\frac{1}{2\sigma^2} \sum (y_i - \mu_2)^2} \\ \ln L &= -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum (x_i - \mu_1)^2 - \frac{m}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum (y_i - \mu_2)^2 \\ \frac{\partial \ln L}{\partial \mu_1} &= \frac{2}{2\sigma^2} \sum (x_i - \mu_1) = 0 \end{aligned}$$

$\implies \hat{\mu}_1 = \bar{x}$. Similarly,

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{2}{2\sigma^2} \sum (y_i - \mu_2) = 0$$

$\implies \hat{\mu}_2 = \bar{y}$. Now, for σ^2 ,

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu_1)^2 - \frac{m}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (y_i - \mu_2)^2 = 0$$

$\implies \hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n+m} = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m}$. Let's check whether $\hat{\sigma}^2$ is unbiased.

$$E\hat{\sigma}^2 = \frac{(n-1)ES_X^2 + (m-1)ES_Y^2}{n+m} = \frac{(n+m-2)\sigma^2}{n+m}$$

which is biased. So the adjusted unbiased expression is

$$S_p^2 = \frac{n+m}{n+m-2} \hat{\sigma}^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

Also, observe that

$$\frac{(n+m-2)S_p^2}{\sigma^2} = \frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2}$$

Thus, $\frac{(n+m-2)S_p^2}{\sigma^2} \sim \chi_{n+m-2}^2$ and also $\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\right)$. Then,

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

which is from the definition of t distribution.

§8.2 Order Statistics

Example 8.2

Let $X_1, \dots, X_n \stackrel{\text{i.i.d}}{\sim} U(0, \theta)$. Find the MLE of θ .

We have

$$f(x) = \frac{1}{\theta}$$

$$L = \frac{1}{\theta^n}$$

Then,

$$\ln L = -n \ln \theta$$

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta}$$

We say that $\hat{\theta} = \max(X_1, \dots, X_n)$ or $\hat{\theta} = X_{(n)}$ - n^{th} order statistic. Similarly, $X_{(1)} = \min(X_1, \dots, X_n)$ - first order statistic.

Let X_1, \dots, X_n i.i.d R.V. with pdf $f(x)$ and cdf $F(x)$. Denote the ordered random variables with $X_{(1)} < X_{(2)} < \dots < X_{(n)}$. We want to find

- Pdf of $X_{(n)}$
- Pdf of $X_{(i)}$
- Pdf of $X_{(j)}$
- Joint pdf of $X_{(i)}$ and $X_{(j)}$

Let's first find the pdf of $X_{(n)}$. Begin with cdf of $X_{(n)}$

$$\begin{aligned} F_{X_{(n)}}(x) &= P(X_{(n)} \leq x) \\ &= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= P(X_1 \leq x) \cdot P(X_2 \leq x) \dots P(X_n \leq x) \\ &= F(x) \cdot F(x) \dots F(x) \end{aligned}$$

So, $F_{X_{(n)}}(x) = [F(x)]^n$. Thus, $g_{X_{(n)}}(x) = n[F(x)]^{n-1} f(x)$ (maximum).

Example 8.3 (The maximum)

For $U(0, \theta)$, the MLE of θ is $\hat{\theta} = X_{(n)}$. Find the pdf of $X_{(n)}$.
We have

$$f(x) = \frac{1}{\theta}$$

$$F(x) = \frac{x}{\theta}$$

Thus, $g_{X_{(n)}}(x) = n \left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta}$.

Is $\hat{\theta} = X_{(n)}$ an unbiased estimator of θ ?

$$\begin{aligned} E\hat{\theta} &= EX_{(n)} = \int_0^\theta xn \left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta} dx \\ &= \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{n}{\theta^n} \frac{x^{n+1}}{n+1} \Big|_0^\theta \\ &= \frac{n}{n+1} \theta \end{aligned}$$

It's therefore biased, but we can adjust it so that it's unbiased

$$\hat{\theta}_1 = \frac{n}{n+1} X_{(n)}$$

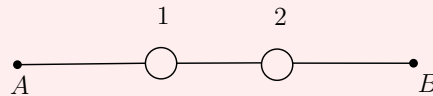
So, $E\hat{\theta}_1 = \theta$.

Now, let's find the pdf of $X_{(1)} = \min(X_1, \dots, X_n)$.

$$\begin{aligned} P(X_{(1)} > x) &= P(X_1 > x, X_2 > x, \dots, X_n > x) \\ &= P(X_1 > x) \cdot P(X_2 > x) \dots P(X_n > x) \\ 1 - P(X_{(1)} \leq x) &= [1 - P(X_1 \leq x)] \cdot [1 - P(X_2 \leq x)] \dots [1 - P(X_n \leq x)] \\ 1 - F_{X_{(1)}}(x) &= [1 - F(x)]^n \end{aligned}$$

Therefore, $g_{X_{(1)}}(x) = n [1 - F(x)]^{n-1} f(x)$ (minimum).

Example 8.4 a) $f(x) = \frac{1}{100} e^{-\frac{1}{100}x}$, $\lambda = \frac{1}{100}$ - exponential distribution and the cdf is $F(x) = 1 - e^{-\frac{1}{100}x}$.

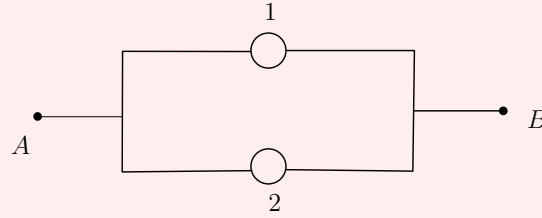


If either component fails, then the system fails \implies minimum of the two.

$$\begin{aligned} g_{X_{(1)}}(x) &= 2 \left[1 - \left(1 - e^{-\frac{1}{100}x} \right) \right]^{2-1} \frac{1}{100} e^{-\frac{1}{100}x} \\ &= \frac{1}{50} e^{-\frac{1}{50}x} \end{aligned}$$

Thus, $X_{(1)} \sim \exp\left(\frac{1}{50}\right)$ and $EX_{(1)} = 50$.

Example 8.5 b) In the case that the system does not fail until both component fails



Thus, in this case we want to look at the maximum.

$$g_{X_{(n)}}(x) = 2 \left[1 - e^{-\frac{1}{100}x} \right]^{2-1} \frac{1}{100} e^{-\frac{1}{100}x}$$

$$g_{X_{(n)}}(x) = \frac{1}{50} \left[1 - e^{-\frac{1}{100}x} \right] e^{-\frac{1}{100}x}$$

In general, $X_1, \dots, X_n \stackrel{\text{i.i.d}}{\sim} \exp(\lambda)$

$$f(x) = \lambda e^{-\lambda x}, \quad F(x) = 1 - e^{-\lambda x}$$

Then,

$$g_{X_{(1)}}(x) = n \left[1 - (1 - e^{-\lambda x}) \right]^{n-1} \lambda e^{-\lambda x}$$

$$= n \lambda e^{-n\lambda x}$$

$$\Rightarrow X_{(1)} \sim \exp(n\lambda).$$

§8.3 MLE with Multi-Parameters

For efficiency, we need the information matrix denoted $I(\theta)$. Suppose

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}$$

Then

$$I(\theta) = -E \begin{pmatrix} \frac{\partial^2 \ln L}{\partial \theta_1^2} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_p} \\ \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \ln L}{\partial \theta_2 \partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L}{\partial \theta_p \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_p \partial \theta_2} & \cdots & \frac{\partial^2 \ln L}{\partial \theta_p^2} \end{pmatrix}$$

in which $I(\theta)$ is $p \times p$ and symmetric matrix. Then we just need to find $I^{-1}(\theta)$.

Example 8.6

$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$. Find $I(\theta)$.

$$\begin{aligned}\ln L &= -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \\ \frac{\partial \ln L}{\partial \mu} &= \frac{1}{\sigma^2} \sum (x_i - \mu) = \frac{1}{\sigma^2} (\sum x_i - n\mu) \\ \frac{\partial^2 \ln L}{\partial \mu^2} &= -\frac{n}{\sigma^2} \\ \frac{\partial \ln L}{\partial \mu \partial \sigma^2} &= -\frac{1}{\sigma^4} \sum (x_i - \mu) \\ \frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu)^2 \\ \frac{\partial \ln L}{\partial \sigma^{2(2)}} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum (x_i - \mu)^2\end{aligned}$$

Thus,

$$\begin{aligned}I(\theta) &= -E \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4} \sum (x_i - \mu) \\ -\frac{1}{\sigma^4} \sum (x_i - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum (x_i - \mu)^2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} n\sigma^2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix} \\ I^{-1}(\theta) &= \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}\end{aligned}$$

If $\hat{\theta}$ is the MLE of θ , then approximately

$$\hat{\theta} \sim N(\theta, I^{-1}(\theta))$$

as $n \rightarrow \infty$.

§8.4 Method of Moments

The moments of the population are estimated by the sample moments.

$$\begin{aligned}EX &= \mu \approx \bar{X} = \frac{1}{n} \sum X_i \\ EX^2 &\approx \frac{1}{n} \sum X_i^2 \\ &\vdots \\ EX^k &\approx \frac{1}{n} \sum X_i^k\end{aligned}$$

Example 8.7

Find the method of moments estimator of θ .

1. $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$.

$$\hat{\lambda} = \bar{x}$$

2. $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} U(0, \theta)$

$$\frac{\hat{\theta}}{2} = \bar{x} \rightarrow \hat{\theta} = 2\bar{x}$$

Here

$$E\hat{\theta} = E2\bar{X} = 2E\bar{X} = 2\mu = 2\frac{\theta}{2} = \theta$$

It's unbiased. Compare $\hat{\theta}$ with the MLE of θ , $\hat{\theta}_1 = \frac{n}{n+1}X_{(n)}$. We finally choose the one with smaller variance between $\hat{\theta}$ and $\hat{\theta}_1$.

3. $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$.

$$\begin{aligned}\hat{\mu} &= \bar{x} \\ \sigma^2 &= EX^2 - (EX)^2 \\ \hat{\sigma}^2 &= \frac{\sum X_i^2}{n} - \bar{X}^2 = \frac{1}{n} \left[\sum X_i^2 - n\bar{X}^2 \right] \\ &= \frac{\sum (X_i - \bar{X})^2}{n}\end{aligned}$$

which is the same as the MLE.

4. $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma)$.

$$\begin{aligned}\sigma^2 &= EX^2 - (EX)^2 \\ &= EX^2 \implies \hat{\sigma}^2 = \frac{\sum X_i^2}{n}\end{aligned}$$

5. $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{R.V. with pdf}$

$$f(x) = (\theta + 1)x^\theta, \quad 0 < x < 1$$

First, find we find EX

$$\begin{aligned}EX &= \int_0^1 xf(x) dx = (\theta + 1) \int_0^1 x^{\theta+1} dx \\ &= (\theta + 1) \frac{x^{\theta+2}}{\theta + 2} \Big|_0^1 = \frac{\theta + 1}{\theta + 2}\end{aligned}$$

$$\implies \bar{X} = \frac{\hat{\theta}+1}{\hat{\theta}+2} \rightarrow \hat{\theta} = \frac{2\bar{X}-1}{1-\bar{X}}.$$

§8.5 Simple Linear Models

Consider $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma)$. Write this statement as a model

$$y_i = \mu + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma)$ and $\varepsilon_1, \dots, \varepsilon_n$ are independent.

Suppose now that $y_i = b_0 + b_1 x_i + \varepsilon_i$ in which $\varepsilon_1, \dots, \varepsilon_n \stackrel{\text{i.i.d}}{\sim} N(0, \sigma)$ and x_1, \dots, x_n non-random. Note that y is referred as the response variable and x is the predictor variable. Let's estimate b_0, b_1, σ^2 using method of maximum likelihood. Because $\varepsilon_i \sim N(0, \sigma)$, it follows that $Y_i \sim N(b_0 + b_1 x_i, \sigma)$. Then,

$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i - b_0 - b_1 x_i)^2}$$

So,

$$\begin{aligned} L &= \prod_{i=1}^n f(y_i) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum (y_i - b_0 - b_1 x_i)^2} \\ \ln L &= -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum (y_i - b_0 - b_1 x_i)^2 \\ \frac{\partial \ln L}{\partial b_0} &= \frac{1}{\sigma^2} \sum (y_i - b_0 - b_1 x_i) = 0 \\ \frac{\partial \ln L}{\partial b_1} &= \frac{1}{\sigma^2} \sum (y_i - b_0 - b_1 x_i) x_i = 0 \end{aligned}$$

Then,

$$\begin{aligned} \sum y_i - nb_0 - b_1 \sum x_i &= 0 \\ \sum x_i y_i - b_0 \sum x_i - b_1 \sum x_i^2 &= 0 \end{aligned}$$

Massage these equations to get

$$\begin{aligned} nb_0 + b_1 \sum x_i &= \sum y_i \\ b_0 \sum x_i + b_1 \sum x_i^2 &= \sum x_i y_i \end{aligned}$$

This is known as the normal equations. Then, we can solve it as follows

$$\begin{aligned} \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} &= \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} \\ \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \end{pmatrix} &= \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix} \end{aligned}$$

Notice that the determinant of the inverse matrix is $n \sum (x_i - \bar{x})^2 \geq 0$ which confirms the existence of a solution to the system of equations. Finally, MLEs of \hat{b}_0 and \hat{b}_1 are

$$\begin{aligned} \hat{b}_0 &= \bar{y} - \hat{b}_1 \bar{x} \\ \hat{b}_1 &= \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} \end{aligned}$$