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

List of Theorems

List of Definitions

§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. $\square$

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.

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

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

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

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:

$$\begin{aligned}
M_X(t) &= \int_x e^{tx} f(x)\, dx \\
M_X^{'}(t) &= \int_x x e^{tx} f(x)\, dx \\
M_X^{'}(0) &= \int_x x f(x)\, dx = EX \\
M_X^{''}(t) &= \int_x x^2 e^{tx} f(x)\, dx \\
M_X^{''}(0) &= \int_x x^2 f(x)\, dx = EX^2 \\
\text{var}(X) &= EX^2 - (EX)^2
\end{aligned}$$

Theorem 1.9
Let $\phi(t) = \ln M_X(t)$. Then

$$\begin{aligned}
\phi'(0) &= EX \\
\phi''(0) &= \text{var}(X)
\end{aligned}$$

Proof. We have

$$\begin{aligned}
\phi'(t) &= \frac{M_X^{'}(t)}{M_X(t)} \\
\phi'(0) &= \frac{M_X^{'}(0)}{M_X(0)} = \frac{E(X)}{1} = EX
\end{aligned}$$

and

$$\begin{aligned}
\phi''(t) &= \frac{M_X^{''}(t) \cdot M_X(t) - \left(M_X^{'}(t)\right)^2}{(M_X(t))^2} \\
&= \dots \\
&= EX^2 - (EX)^2 \\
&= \text{var}(X)
\end{aligned}$$

□

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

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

- 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.

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

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 – $Z \sim N(0,1)$

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

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

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

□

Example 1.11

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

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

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

§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 $X_1, X_2, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Poisson}(\lambda)$ and $T = X_1 + X_2 + \ldots + X_n$.

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

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 $X + Y$.

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

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:

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

Then, $F_Y(y) = F_X\left(\frac{y}{c}\right)$. 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 $X_1, X_2, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. Let $T = X_1 + X_2 + \ldots + X_n$. 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 $\overline{X} = \frac{\sum X_i}{n} = \frac{T}{n}$. Find $M_{\overline{X}}(t)$.

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

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

Recall

Theorem 2.9 (Central Limit Theorem)

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

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

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

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

□

§2.2 Joint MGF

Let $X = \begin{bmatrix} X_1 & X_2 & \ldots & X_n \end{bmatrix}^\top$ be a random vector and $t = \begin{bmatrix} t_1 & t_2 & \ldots & t_n \end{bmatrix}^\top$.

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 $\mu = \begin{bmatrix} \mu_1 & \mu_2 & \ldots & \mu_n \end{bmatrix}^\top$, 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

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

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 $a = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}^\top$ be a vector of constants. Find the mean and variance of $a^\top X$.

$$\begin{aligned} Ea^\top X &= a^\top EX = a^\top \mu \\ \operatorname{var}(a^\top X) &= E(a^\top X - a^\top \mu)^2 \\ &= a^\top \left[E(X-\mu)(X-\mu)^\top \right] a \\ &= a^\top \Sigma a \end{aligned}$$

or using summation, we have

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

Example 2.11

For $n = 3$,

$$\begin{aligned} \operatorname{var}(a_1X_1 + a_2X_2 + a_3X_3) &= \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \\ &= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + a_3^2\sigma_3^2 + 2a_1a_2\sigma_{12} + 2a_1a_3\sigma_{13} + 2a_2a_3\sigma_{23} \end{aligned}$$

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.

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

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

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

Note that $\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB) \neq \mathrm{tr}(BAC)$
Let $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, $t = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$. Then,

$$\begin{aligned} M_X(t) &= E\left(e^{t_1X_1 + t_2X_2}\right) \\ &= \int_{x_1}\int_{x_2} e^{t_1x_1+t_2x_2} f(x_1,x_2)\,dx_1\,dx_2 \\ M_1(t) = \frac{\partial M_X(t)}{\partial t_1} &= \int_{x_1}\int_{x_2} x_1 e^{t_1x_1+t_2x_2} f(x_1,x_2)\,dx_1\,dx_2 \end{aligned}$$

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,

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

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

- 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.

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

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

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

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}\begin{pmatrix} x_1 = h_1^{-1}(y_1, y_2) \\ x_2 = h_2^{-1}(y_1, y_2) \end{pmatrix} \cdot |J|$$

or by using the original function

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}\begin{pmatrix} x_1 = h_1^{-1}(y_1, y_2) \\ x_2 = h_2^{-1}(y_1, y_2) \end{pmatrix} \cdot |J|^{-1}$$

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

$$f_{X_1X_2}(x_1, x_2) = f(x_1) \cdot f(x_2) = \lambda_1\lambda_2 e^{-(\lambda_1 x_1 + \lambda_2 x_2)}$$

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

$$x_1 = \frac{u+v}{2}$$
$$x_2 = \frac{u-v}{2}$$

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.

$$\begin{aligned} x &= uv \\ y &= u - uv \end{aligned}$$

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

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

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 $v_i = 0$,

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

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

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

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

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

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 $Z_1, Z_2, \ldots, Z_n$ be i.i.d and $\sim N(0,1)$. Show that $\mathbf{Z} \sim N(\mathbf{0}, I)$.

$$\begin{aligned} f(\mathbf{z}) &= f(\mathbf{z_1}) \cdot \ldots f(\mathbf{z_n}) \\ f(z_i) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2} \end{aligned}$$

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.

$$\begin{aligned} M_Z(\mathbf{t}) &= Ee^{\mathbf{t}^\top \mathbf{z}} = Ee^{t_1z_1 + \ldots + t_nz_n} \\ &= Ee^{t_1z} \ldots Ee^{t_nz_n} \\ &= e^{\frac{1}{2}t_1^2} \ldots e^{\frac{1}{2}t_n^2} \\ &= e^{\frac{1}{2}\sum t_i^2} \\ &= e^{\frac{1}{2}\mathbf{t}^\top \mathbf{t}} \end{aligned}$$

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 $\lambda_1, \ldots, \lambda_n$ 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 $\mathbf{e_1}, \ldots, \mathbf{e_n}$. They are orthogonal, i.e., $\mathbf{P}\mathbf{P}^\top = I$ in which $P = \begin{pmatrix} \mathbf{e_1} & \mathbf{e_2} & \ldots & \mathbf{e_n} \end{pmatrix}$. 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 $Z_1, \ldots, Z_n \overset{\text{i.i.d}}{\sim} N(0,1)$. 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

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

Then

$$\begin{aligned} Z_1 &= v_1 y_1 + v_{12} y_2 + \ldots + v_{1n} y_n - \text{const1} \\ Z_2 &= v_{21} y_1 + v_{22} y_2 + \ldots + v_{2n} y_n - \text{const2} \\ &\vdots \\ Z_n &= v_{n1} y_1 + v_{n2} y_2 + \ldots + v_{nn} y_n - \text{const}n \end{aligned}$$

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

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

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

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

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

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

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

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

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 $\mathbf{AY} + \mathbf{C}$ is multivariate normal.

Proof. Consider the MGF

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

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

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

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

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

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

In addition, we have

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

Theorem 4.2

Let

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

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

where $\mathbf{A} = \begin{pmatrix} I & \mathbf{0} \end{pmatrix}$. Then

$$\begin{aligned} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} &\sim N\left(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top\right) \\ &\sim N\left(\begin{pmatrix} I & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} I & \mathbf{0} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I \\ \mathbf{0} \end{pmatrix}\right) \\ &\sim N\left(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}\right) \end{aligned}$$

Also, the <u>linear combination</u> follows the normal distribution in which

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

§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

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

or

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

So if $\operatorname{cov}(\mathbf{Q}_1, \mathbf{Q}_2) = \mathbf{0}$, 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

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

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

§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

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

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

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

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

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

Variance

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

Notice that $\text{cov}(\mathbf{U}, \mathbf{V}) = 0$, 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 $\text{cov}(\mathbf{AY}, \mathbf{BY}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\top$

We have

$$\begin{aligned} \text{cov}(\mathbf{U}, \mathbf{V}) &= \text{cov}(\mathbf{Q}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Q_2}, \mathbf{Q_2}) \\ &= \text{cov}(\mathbf{Q}_1, \mathbf{Q}_2) - \text{cov}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Q_2}, \mathbf{Q}_2) \\ &= \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{22} \\ &= \mathbf{0}\end{aligned}$$

Observe that

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

Then

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

but $\mathbf{Q_2} = \mathbf{V}$

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

§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 $p_1, p_2, \ldots, p_r$ and $\sum_{i=1}^{r} p_i = 1$. Let X_i be the number of the n experiments that result in outcome i where $i = 1, 2, \ldots, r$. Then

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

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

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

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

$$M_{\mathbf{X}}(\mathbf{t}) = \sum_{X_1} \ldots \sum_{X_r} \frac{n!}{x_1! \ldots x_r!} (p_1 e^{t_1})^{X_1} \ldots (p_r e^{t_2})^{X_r}$$

Using the multinomial theorem, we have

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

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

By setting every other X_i to 0, we have

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

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

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

§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:

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

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.,

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

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

For $\mathcal{X}_n^2$, let $Z_1, \ldots, Z_n \overset{\text{i.i.d}}{\sim} N(0,1)$. 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 $\Gamma\left(\frac{n}{2}, 2\right)$. Let $X_1, X_2, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. 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 $X \sim \Gamma\left(\frac{n}{2}, 2\right)$, we can easily deduce that the mean and variance are

$$\begin{aligned} EX &= n \\ \text{var}(X) &= 2n \end{aligned}$$

<u>Distribution of Quadratic Forms of Normally Distributed RV</u>:

a) Let $Z_1, Z_2, \ldots, Z_n \overset{\text{i.i.d}}{\sim} N(0,1)$. 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 $\mathbf{V} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y} - \boldsymbol{\mu})$.

$$\begin{aligned} E\mathbf{V} &= \boldsymbol{\Sigma}^{-\frac{1}{2}} E\left[\mathbf{Y} - \boldsymbol{\mu}\right] = \mathbf{0} \\ \text{var}(\mathbf{V}) &= \text{var}\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y}\right) \\ &= \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-\frac{1}{2}} = I \end{aligned}$$

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

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

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

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

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum (X_i - \overline{X})^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$$

We have

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

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

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

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

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

Then,

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

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

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

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

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

$\overline{X}$ is independent of $\begin{pmatrix} X_1 - \overline{X} & \ldots & X_n - \overline{X} \end{pmatrix}^\top$ and therefore is independent of S^2.

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

We have

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

Also,

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

and

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

Proof. (of the above theorem) We have

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

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, $\frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$. □

<u>Summary</u>:

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

This is known as central chi-squared distribution.

§5.3 Non-central Chi-Squared Distribution

Definition 5.4 (Non-central $\mathcal{X}^2$) — If $Y \sim N(\mu, 1)$, then $Y^2 \sim \mathcal{X}^2_1$ (NCP $= \mu^2$) 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, $\frac{Y^2}{\sigma^2} \sim \mathcal{X}_1^2$ (NCP $= \frac{\mu^2}{\sigma^2}$).
Let's find the MGF of $\mathcal{X}_1^2$ (NCP $= \theta$). Let $Q \sim \mathcal{X}_1^2$ (NCP $= \theta$).

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

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

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

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

Example 5.5

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

a) Find the distribution of W where

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

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

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

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, $X + Y \sim \mathcal{X}_{n+m}^2$. So,

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

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 $U \sim \mathcal{X}_{df}$. If Z, U are independent, then

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

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

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

Summary:

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

§6 Lec 6: Aug 19, 2021

§6.1 t Distribution (Cont'd)

Example 6.1

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

$$P\left[\frac{\frac{\overline{X}-0}{\sigma/\sqrt{10}}}{\sqrt{\frac{9S_X^2}{\sigma^2}/9}} < \sqrt{90}c\right] = 0.95$$
$$P(t_9 < \sqrt{90}c) = 0.95$$

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

Example 6.2

$X_1, X_2, X_3, X_4, X_5 \overset{\text{i.i.d}}{\sim} N(0,\sigma)$. 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

$$X_1 - X_2 \sim N(0, \sigma\sqrt{2})$$
$$\frac{X_3^2}{\sigma^2} + \frac{X_4^2}{\sigma^2} + \frac{X_5^2}{\sigma^2} \sim \mathcal{X}_3^2$$

So

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

Example 6.3

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

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

We have

$$E\left[\overline{X}+\overline{Y}-2\overline{W}\right] = \mu + \mu - 2\mu = 0$$
$$\text{var}(\overline{X}+\overline{Y}-2\overline{W}) = \frac{\sigma^2}{10} + \frac{\sigma^2}{10} + \frac{4\sigma^2}{10} = \frac{6\sigma^2}{10}$$

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

Example 6.4 (Cont'd from above)

$$P\left(\frac{\frac{\overline{X}+\overline{Y}-2\overline{W}-0}{\sigma\sqrt{\frac{6}{10}}}}{\sqrt{\frac{9S_X^2+9S_Y^2+9S_W^2}{\sigma^2}/27}} < \sqrt{\frac{270}{6}}c\right) = 0.95$$
$$P\left(t_{27} < \sqrt{\frac{270}{6}}c\right) = 0.95$$
$$\implies t_{0.95;27} = \sqrt{\frac{270}{6}}c = 1.703$$

Definition 6.5 (Non-Central t Distribution) — Let $U \sim N(\delta, 1)$ and $V \sim \mathcal{X}_n^2$ 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 $U \sim \mathcal{X}_n^2$ and $V \sim \mathcal{X}_m^2$. If U, V are independent, then $\frac{\frac{U}{n}}{\frac{V}{m}} \sim F_{n,m}$.

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

$$\frac{(n-1)S_1^2}{\sigma_1^2} \sim \mathcal{X}_{n-1}^2 \quad \text{and} \quad \frac{(m-1)S_2^2}{\sigma_2^2} \sim \mathcal{X}_{m-1}^2$$
$$\frac{\frac{(n-1)S_1^2}{\sigma_1^2}/(n-1)}{\frac{(m-1)S_2^2}{\sigma_2^2}/(m-1)} = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F_{n-1,m-1}$$

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

$$t = \frac{Z}{\sqrt{\frac{U}{n}}}$$
$$Z \sim N(0,1) \quad \text{and} \quad U \sim \mathcal{X}_n^2$$

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

Definition 6.7 (Non-Central F Distribution) — Let $U \sim \mathcal{X}_n^2(\text{NCP} = \theta)$ and $V \sim \mathcal{X}_m^2$. 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

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

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

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

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

Example 6.9

Consider $\hat{p} = \frac{X}{n}$ where $X_1, X_2, \ldots, X_n$ 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

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

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

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

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

<u>*Note*</u>: Have

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

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 $\overline{X}$ an efficient estimator of μ?

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

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,

$$
\begin{aligned}
I(\theta) &= E\left[\frac{\partial \ln f(x)}{\partial \theta}\right]^2 \\
\frac{\partial \ln f(x)}{\partial \mu} &= \frac{2}{2\sigma^2}(x-\mu) = \frac{x-\mu}{\sigma^2} \\
I(\theta) &= E\left[\frac{X-\mu}{\sigma^2}\right]^2 = \frac{E\left[X-\mu\right]^2}{\sigma^2} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}
\end{aligned}
$$

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

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

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

Example 6.13

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

$$
\begin{aligned}
ES^2 &= \sigma^2 \\
\text{var}(S^2) &= \frac{2\sigma^4}{n-1}
\end{aligned}
$$

Cramer-Rao lower bound:

$$
\begin{aligned}
\ln f(x) &= -\frac{1}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}(x-\mu)^2 \\
\frac{\partial \ln f(x)}{\partial \sigma^2} &= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(x-\mu)^2 \\
\frac{\partial^2 \ln f(x)}{\partial \sigma^2} &= \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x-\mu)^2 \\
I(\theta) &= -E\left[\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x-\mu)^2\right] \\
&= -\frac{1}{2\sigma^4} + \frac{1}{\sigma^4} = \frac{1}{2\sigma^4}
\end{aligned}
$$

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

$$L = f(x_1, x_2, \ldots, x_n; \theta) = f(x_1;\theta)\cdot f(x_2;\theta)\ldots f(x_n;\theta)$$
$$\ln L = \ln f(x_1;\theta) + \ln f(x_2;\theta) + \ldots + \ln f(x_n;\theta)$$
$$\frac{\partial \ln L}{\partial \theta} = \frac{\partial \ln f(x_1;\theta)}{\partial \theta} + \ldots + \frac{\partial \ln f(x_n;\theta)}{\partial \theta}$$

Information in the sample is

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

Recall the Cramer-Rao Inequality:

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

Let's prove this inequality.

Proof. Let $X_1, X_2, \ldots, X_n$ i.i.d R.V. with pdf $f(x;\theta)$. Let $\hat{\theta} = g(x_1, x_2, \ldots, x_n)$ 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

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

Generalize this to arbitrary n and we get

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

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

$$-1 \leq \rho \leq 1$$
$$\rho^2_{\hat{\theta}Q} \leq 1$$
$$\frac{\operatorname{cov}^2(\hat{\theta}, Q)}{\operatorname{var}(\hat{\theta})\operatorname{var}(Q)} \leq 1$$
$$\frac{\left[E\hat{\theta}Q - (E\hat{\theta})(EQ)\right]^2}{\operatorname{var}(\hat{\theta})\left[EQ^2 - (EQ)^2\right]} \leq 1$$

Note that

$$E\left[\hat{\theta}Q\right] = 1$$
$$EQ = 0$$
$$EQ^2 = nI(\theta)$$

Plug these into the inequality to obtain

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

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

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

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

Example 7.1

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

$$\hat{\theta}_1 = \frac{X_1+X_2}{2}$$
$$\hat{\theta}_2 = \overline{X}$$

Then,

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

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 $\hat{\theta}_1$.

<u>Consistency</u>:

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

Theorem 7.3

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

Example 7.4

Consider $\overline{X}$

$$E\overline{X} = \mu$$
$$\mathrm{var}(\overline{X}) = \frac{\sigma^2}{n} \overset{n\to\infty}{\longrightarrow} 0$$

$\implies$ consistent.

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

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

and

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

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

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

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

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

Back to the proof, we have

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

Let $\varepsilon = k\sqrt{\mathrm{var}\,\hat{\theta}}$. Then,

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

If $\mathrm{var}(\hat{\theta}) \overset{n\to\infty}{\longrightarrow} 0$, we conclude that $\hat{\theta}$ is consistent. $\square$

Example 7.5

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

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

As $\mathrm{var}(S^2) \overset{n\to\infty}{\longrightarrow} 0$, then S^2 is consistent.

Example 7.6

$X_1, \ldots, X_n$ i.i.d Bernoulli R.V.

$$\hat{p} = \frac{X}{n}$$
$$E\frac{X}{n} = p$$
$$\text{var}(\frac{X}{n}) = \frac{p(1-p)}{n} \overset{n\to\infty}{\longrightarrow} 0$$

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

§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 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

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

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

Consider $\overline{X}$ as an estimator of θ.

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

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

It can be easily deduced that $\hat{\theta_1} = \overline{X} - \frac{1}{2}$ is unbiased as $E\hat{\theta}_1 = \theta$.

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

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

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

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

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 $X_1, X_2, \ldots, X_n$ be i.i.d R.V. with pdf $f(x;\theta)$. The joint pdf is also called the likelihood function is

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

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

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

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

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

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

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

$$\frac{\partial \ln L}{\partial \lambda} = \frac{\sum x_i}{\lambda} - n = 0$$
$$\hat{\lambda} = \frac{\sum x_i}{n} = \overline{x}$$

Then, we can check for unbiased and consistent through

$$E\overline{X} = \lambda$$
$$\text{var}(\overline{X}) = \frac{\lambda}{n}$$

Example 7.11

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

$$\begin{aligned} L &= \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \dots \lambda e^{-\lambda x_n} \\ L &= \lambda^n e^{\lambda \sum x_i} \\ \ln L &= n \ln \lambda - \lambda \sum x_i \end{aligned}$$

Maximize $\ln L$,

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

Example 7.12

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

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

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

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

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

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

Note that

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

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

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

§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

$X_1, X_2, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \theta)$. 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 $n \to \infty$,

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

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

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

$\implies \hat{\mu_1} = \overline{x}$. Similarly,

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

$\implies \hat{\mu_2} = \overline{y}$. Now, for σ^2,

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

$\implies \hat{\sigma^2} = \frac{\sum(x_i-\overline{x})^2 + \sum(y_i-\overline{y})^2}{n+m} = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m}$. 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, $\frac{(n+m-2)S_p^2}{\sigma^2} \sim \mathcal{X}^2_{n+m-2}$ and also $\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\right)$. Then,

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

which is from the definition of t distribution.

§8.2 Order Statistics

Example 8.2

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} U(0,\theta)$. 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 $\hat{\theta} = \max(X_1, \ldots, X_n)$ or $\hat{\theta} = X_{(n)}$ – n^{th} order statistic. Similarly, $X_{(1)} = \min(X_1, \ldots, X_n)$ – first order statistic.

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

$$\begin{aligned} F_{X_{(n)}}(x) &= P\left(X_{(n)} \le x\right) \\ &= P\left(X_1 \le x, X_2 \le x, \ldots, X_n \le x\right) \\ &= P(X_1 \le x) \cdot P(X_2 \le x) \ldots P(X_n \le x) \\ &= F(x) \cdot F(x) \ldots F(x) \end{aligned}$$

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

$$\hat{\theta_1} = \frac{n}{n+1}X_{(n)}$$

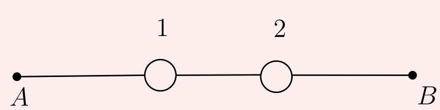
So, $E\hat{\theta_1} = \theta$.

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

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

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 $F(x) = 1 - e^{-\frac{1}{100}x}$.

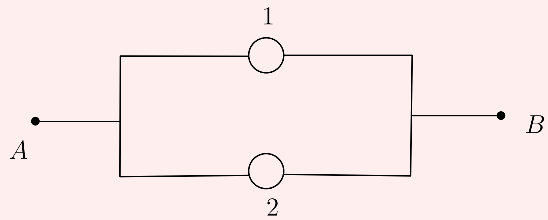


If either component fails, then the system fails $\implies$ minimum of the two.

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

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, $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \exp(\lambda)$

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

Then,

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

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

§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

$X_1, \dots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. 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 $n \to \infty$.

§8.4 Method of Moments

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

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

Example 8.7

Find the method of moments estimator of θ.

1. $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Poisson}(\lambda)$.
$$\hat{\lambda} = \overline{x}$$

2. $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} U(0, \theta)$
$$\frac{\hat{\theta}}{2} = \overline{x} \to \hat{\theta} = 2\overline{x}$$
Here
$$E\hat{\theta} = E2\overline{X} = 2E\overline{X} = 2\mu = 2\frac{\theta}{2} = \theta$$
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. $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$.
$$\begin{aligned} \hat{\mu} &= \overline{x} \\ \sigma^2 &= EX^2 - (EX)^2 \\ \hat{\sigma^2} &= \frac{\sum X_i^2}{n} - \overline{X}^2 = \frac{1}{n}\left[\sum X_i^2 - n\overline{X}^2\right] \\ &= \frac{\sum (X_i - \overline{X})^2}{n} \end{aligned}$$
which is the same as the MLE.

4. $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$.
$$\begin{aligned} \sigma^2 &= EX^2 - (EX)^2 \\ &= EX^2 \implies \hat{\sigma^2} = \frac{\sum X_i^2}{n} \end{aligned}$$

5. $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim}$ R.V. with pdf
$$f(x) = (\theta+1)x^\theta, \quad 0 < x < 1$$
First, find we find EX
$$\begin{aligned} EX &= \int_0^1 x f(x)\, dx = (\theta+1)\int_0^1 x^{\theta+1}\, dx \\ &= (\theta+1)\frac{x^{\theta+2}}{\theta+2}\Big|_0^1 = \frac{\theta+1}{\theta+2} \end{aligned}$$
$\implies \overline{X} = \frac{\hat{\theta}+1}{\hat{\theta}+2} \to \hat{\theta} = \frac{2\overline{X}-1}{1-\overline{X}}$.

§8.5 Simple Linear Models

Consider $Y_1, \ldots, Y_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. Write this statement as a model
$$y_i = \mu + \varepsilon_i$$
with $\varepsilon_i \sim N(0, \sigma)$ and $\varepsilon_1, \ldots, \varepsilon_n$ are independent.

Suppose now that $y_i = b_0 + b_1 x_i + \varepsilon_i$ in which $\varepsilon_1, \dots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$ 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 $n\sum(x_i - \overline{x})^2 \geq 0$ which confirms the existence of a solution to the system of equations. Finally, MLEs of $\hat{b_0}$ and $\hat{b_1}$ are

$$\begin{aligned}
\hat{b_0} &= \overline{y} - \hat{b_1}\overline{x} \\
\hat{b_1} &= \frac{\sum(x_i - \overline{x})y_i}{\sum(x_i - \overline{x})^2}
\end{aligned}$$

§9 Lec 9: Aug 31, 2021

§9.1 Simple Regression

Consider $Y_i = b_0 + b_1 X_i + \varepsilon_i$ in which we assume $E[\varepsilon_i] = 0$ and $\text{var}(\varepsilon_i) = \sigma^2$ and $\varepsilon_1, \ldots, \varepsilon_n$ are independent. This is known as Gauss-Markov Conditions. Also, assume $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$. So, $Y_i \sim N(b_0 + b_1 X_i, \sigma)$. Then, the pdf is

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

Since Y_i are independent, we can find the likelihood function as follows

$$\begin{aligned} L &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(y_i - b_0 - b_1 x_i)^2} \\ \ln L &= -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum(y_i - b_0 - b_1 x_i)^2 \end{aligned}$$

Recall from last lecture,

$$\begin{aligned} \hat{b_0} &= \overline{y} - \hat{b_1}\overline{x} \\ \hat{b_1} &= \frac{\sum(x_i - \overline{x})y_i}{\sum(x_i - \overline{x})^2} \end{aligned}$$

Properties:

$$\begin{aligned} E\hat{b_1} &= E\left[\frac{\sum(x_i - \overline{x})y_i}{\sum(x_i - \overline{x})^2}\right] \\ &= \frac{\sum(x_i - \overline{x})EY_i}{\sum(x_i - \overline{x})^2} \\ &= \frac{\sum(x_i - \overline{x})(b_0 + b_1 x_i)}{\sum(x_i - \overline{x})^2} \\ &= \frac{b_0\sum(x_i - \overline{x})}{\sum(x_i - \overline{x})^2} + b_1\frac{\sum(x_i - \overline{x})x_i}{\sum(x_i - \overline{x})^2} \\ &= b_1 \end{aligned}$$

Note that $\sum(x_i - \overline{x}) = 0$ and $\sum(x_i - \overline{x})x_i = \sum(x_i - \overline{x})^2$. Now, for $\hat{b_0}$,

$$\begin{aligned} E\left[\hat{b_0}\right] &= E\left[\overline{y} - \hat{b_1}\overline{x}\right] \\ &= E\overline{y} - \overline{x}E\hat{b_1} \\ &= b_0 \end{aligned}$$

Note that

$$\begin{aligned} \overline{y} &= \frac{\sum y_i}{n} = \frac{\sum(b_0 + b_1 x_i + \varepsilon_i)}{n} \\ &= \frac{nb_0 + b_1\sum x_i + \sum\varepsilon_i}{n} \\ &= b_0 + b_1\overline{x} + \frac{\sum\varepsilon_i}{n} \end{aligned}$$

Thus, $\hat{b_0}, \hat{b_1}$ are unbiased estimator of b_0, b_1.
Let's denote the residual as $e_i = Y_i - \hat{Y_i}$. Also,

$$\hat{Y_i} = \overline{Y} + \hat{b_1}(X_i - \overline{X})$$

Therefore, the residual can be expressed as

$$e_i = Y_i - \overline{Y} - \hat{b_1}(X_i - \overline{X})$$

<u>Estimation of σ^2</u>

$$\begin{aligned}
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum (y_i - b_0 - b_1 x_i)^2 \\
\hat{\sigma^2} &= \frac{\sum (y_i - \hat{b_0} - \hat{b_1} x_i)^2}{n} \\
&= \frac{\sum (y_i - \hat{y}_i)^2}{n} \\
&= \frac{\sum e_i^2}{n}
\end{aligned}$$

Let's check whether $\hat{\sigma^2}$ is unbiased.

$$E\hat{\sigma^2} = E\left[\frac{\sum e_i^2}{n}\right] = \frac{\sum Ee_i^2}{n} = \frac{1}{n}\sum \left[\operatorname{var}(e_i) + (Ee_i)^2\right]$$

Note that

$$\begin{aligned}
Ee_i &= E(y_i - \hat{y}_i) = E\left[y_i - (\hat{b_0} + \hat{b_1} x_i)\right] \\
&= Ey_i - E(\hat{b_0} + \hat{b_1} x_i) \\
&= b_0 + b_1 x_i - (b_0 + b_1 x_i) = 0 \\
\operatorname{var}(e_i) &= \operatorname{var}\left(y_i - \overline{y} - \hat{b_1}(x_i - \overline{x})\right) \\
&= \operatorname{var}(y_i) + \operatorname{var}(\overline{y}) + (x_i - \overline{x})\operatorname{var}(\hat{b_1}) - 2\operatorname{cov}(y_i, \overline{y}) - 2(x_i - \overline{x})\operatorname{cov}(y_i, \hat{b_1}) \\
&\quad + 2(x_i - \overline{x})\operatorname{cov}(\overline{y}, \hat{b_1})
\end{aligned}$$

First, let's find the variance of $\hat{b_1}$.

$$\begin{aligned}
\operatorname{var}(\hat{b_1}) &= \operatorname{var}\left(\frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2}\right) \\
&= \frac{1}{\left(\sum (x_i - \overline{x})^2\right)^2}\sum (x_i - \overline{x})^2 \operatorname{var}(y_i) \\
&= \frac{\sigma^2}{\sum (x_i - \overline{x})^2}
\end{aligned}$$

$\implies \hat{b_1} \sim N\left(b_1, \frac{\sigma}{\sqrt{\sum (x_i - \overline{x})^2}}\right)$. Now, let's find $\operatorname{cov}(y_i, \hat{b_1})$.

$$\begin{aligned}
\operatorname{cov}(y_i, \hat{b_1}) &= \operatorname{cov}\left(y_i, \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2}\right) \\
&= \operatorname{cov}\left(y_i, \frac{(x_1 - \overline{x}) y_1}{\sum (x_i - \overline{x})^2} + \ldots + \frac{(x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2} + \ldots + \frac{(x_n - \overline{x}) y_n}{\sum (x_i - \overline{x})^2}\right) \\
&= \frac{(x_i - \overline{x})}{\sum (x_i - \overline{x})^2}\sigma^2
\end{aligned}$$

Note that $\text{cov}(y_i, y_j) = 0$ for all i, j. Similarly, for $\text{cov}(\overline{y}, \hat{b_1})$,

$$\begin{aligned}
\text{cov}(\overline{y}, \hat{b_1}) &= \text{cov}\left(\frac{y_1 + \ldots + y_n}{n}, \frac{\sum(x_i - \overline{x})y_i}{\sum(x_i - \overline{x})^2}\right) \\
&= \text{cov}\left(\frac{y_1}{n} + \ldots + \frac{y_n}{n}, \frac{(x_i - \overline{x})y_1}{\sum(x_i - \overline{x})^2} + \ldots + \frac{(x_n - \overline{x})y_n}{\sum(x_i - \overline{x})^2}\right) \\
&= \frac{\sigma^2(x_1 - \overline{x})}{n\sum(x_i - \overline{x})^2} + \ldots + \frac{\sigma^2(x_n - \overline{x})}{n\sum(x_i - \overline{x})^2} \\
&= \frac{\sigma^2 \sum(x_i - \overline{x})}{n\sum(x_i - \overline{x})^2} = 0
\end{aligned}$$

Now that we have everything, let's get back to $\text{var}(e_i)$.

$$\begin{aligned}
\text{var}(e_i) &= \text{var}\left(y_i - \overline{y} - \hat{b_1}(x_i - \overline{x})\right) \\
&= \text{var}(y_i) + \text{var}(\overline{y}) + (x_i - \overline{x})\,\text{var}(\hat{b_1}) - 2\,\text{cov}(y_i, \overline{y}) - 2(x_i - \overline{x})\,\text{cov}(y_i, \hat{b_1}) \\
&\quad + 2(x_i - \overline{x})\,\text{cov}(\overline{y}, \hat{b_1}) \\
&= \sigma^2 + \frac{\sigma^2}{n} - \frac{\sigma^2(x_i - \overline{x})}{\sum(x_i - \overline{x})^2} - 2\frac{\sigma^2}{n} - \frac{2\sigma^2(x_i - \overline{x})^2}{\sum(x_i - \overline{x})^2} \\
&= \sigma^2\left(1 - \frac{1}{n} - \frac{(x_i - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right)
\end{aligned}$$

Finally, we can compute $E\hat{\sigma^2}$.

$$\begin{aligned}
E\hat{\sigma^2} &= \frac{\sum \text{var}(e_i)}{n} \\
&= \frac{\sigma^2 \sum\left(1 - \frac{1}{n} - \frac{(x_i - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right)}{n} \\
&= \sigma^2 \frac{n-2}{n}
\end{aligned}$$

We can adjust it to be unbiased as

$$S_e^2 = \frac{n}{n-2}\hat{\sigma^2} \quad \text{or} \quad S_e^2 = \frac{\sum e_i^2}{n-2}$$

Problem 9.1. Show that $\frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2$.

Since $Y_i \sim N(b_0 + b_1 X_i, \sigma)$, it follows that

$$\sum\left(\frac{Y_i - b_0 - b_1 X_i}{\sigma}\right)^2 \sim \mathcal{X}_n^2$$

Consider

$$\begin{aligned}
\sum\left(\frac{y_i - b_0 - b_1 x_i}{\sigma}\right)^2 &= \frac{\sum(y_i - \hat{b_0} - \hat{b_1}x_i + (\hat{b_0} - b_0) + (\hat{b_1} - b_1)x_i)^2}{\sigma^2} \\
&= \frac{\sum e_i^2}{\sigma^2} + \frac{n(\hat{b_0} - b_0)^2}{\sigma^2} + \frac{(\hat{b_1} - b_1)^2 \sum x_i^2}{\sigma^2} + \frac{2(\hat{b_0} - b_0)\sum e_i}{\sigma^2} \\
&\quad + \frac{2(\hat{b_1} - b_1)\sum e_i x_i}{\sigma^2} + \frac{2(\hat{b_0} - b_0)(\hat{b_1} - b_1)\sum x_i}{\sigma^2}
\end{aligned}$$

Note that

$$\begin{cases} \sum e_i = 0 \\ \sum e_i x_i = 0 \end{cases}$$

Then,

$$\frac{\sum(y_i - b_0 - b_1 x_i)^2}{\sigma^2} = \frac{(n-2)S_e^2}{\sigma^2} + \frac{n(\hat{b_0} - b_0)^2}{\sigma^2} + \frac{(\hat{b_1} - b_1)^2 \sum x_i^2}{\sigma^2} + \frac{2(\hat{b_0} - b_0)(\hat{b_1} - b_1)\sum x_i}{\sigma^2} \tag{**}$$

Let $D = \hat{b_0} + \hat{b_1}\overline{X} = \overline{Y}$ and consider

$$\frac{(\hat{b_1} - b_1)^2}{\mathrm{var}(\hat{b_1})} + \frac{[D - (b_0 + b_1\overline{x})]^2}{\mathrm{var}(D)} = \frac{(\hat{b_1} - b_1)^2}{\sigma^2}\sum(x_i - \overline{x})^2 + \frac{(\hat{b_0} - b_0 + (\hat{b_1} - b_1)\overline{x})^2}{\sigma^2/n} \tag{*}$$

Notice that lHS of (*) $\sim \mathcal{X}_2^2$ and we can manipulate its RHS to be equal to the last three terms of the RHS of (**). Thus, we can conclude that $\frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}_{n-2}^2$.

§9.2 Order Statistics (Cont'd)

Let $X_1, \ldots, x_n$ i.i.d R.V. with pdf $f(x)$ and cdf $F(x)$ and $x_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be ordered ran variables. Recall that

$$\begin{aligned} g_{X_{(n)}}(x) &= nF^{n-1}(x)f(x) \\ g_{X_{(1)}}(x) &= n\left(1 - F(x)\right)^{n-1} f(x) \end{aligned}$$

Now, let's find the pdf of $X_{(j)}$. Begin with cdf of $X_{(j)}$

$$\begin{aligned} F_{X_{(j)}}(x) &= P\left(X_{(j)} \le x\right) = P\left(Y \ge j\right) \\ &= \sum_{k=j}^{n} \binom{n}{k} F(x)^k \left(1 - F(x)\right)^{n-k} \end{aligned}$$

where $Y \sim b\left(n, F(x)\right)$. Note that

$$P\left(x \le X_{(j)} \le x + dx\right) \approx g_{X_{(j)}} dx = \begin{pmatrix} & n & \\ j-1 & 1 & n-j \end{pmatrix} F(x)^{j-1} f(x) dx \left(1 - F(x)\right)^{n-j}$$

Thus, the pdf of $X_{(j)}$ is

$$g_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} F(x)^{j-1} \left(1 - F(x)\right)^{n-j} f(x)$$

Using the same approach we can find the joint pdf of $X_{(i)}$ and $X_{(j)}$

$$\begin{aligned} P\left(q \le Q \le q + dq,\ w \le W \le w + dw\right) &\approx f(q, w) dq dw \\ P\left(u \le X_{(i)} \le u + du,\ v \le X_{(j)} \le v + dv\right) &\approx g_{X_{(i)}X_{(j)}}(u, v) du dv \end{aligned}$$

Lastly, we define the range and midrange as follows

$$\begin{aligned} R &= X_{(n)} - X_{(1)} \\ Q &= \frac{X_{(1)} + X_{(n)}}{2} \end{aligned}$$

Exercise 9.1. Find the joint pdf of R and Q.

§9.3 Sufficiency

Let $X_1, X_2, \ldots, X_n \overset{\text{i.i.d}}{\sim} \exp(\lambda)$. Then, $2\lambda \sum X_i \sim \mathcal{X}^2_{2n}$. Then,

$$P\left(\mathcal{X}^2_{\frac{\alpha}{2};2n} \leq 2\lambda \sum X_i \leq \mathcal{X}^2_{1-\frac{\alpha}{2};2n}\right) = 1-\alpha$$

This is known as the confidence interval. Also, we can manipulate the above expression to obtain

$$P\left(\frac{\mathcal{X}^2_{\frac{\alpha}{2};2n}}{\sum X_i} \leq \lambda \leq \frac{\mathcal{X}^2_{1-\frac{\alpha}{2};2n}}{\sum X_i}\right) = 1-\alpha$$
$$P\left(L \leq \lambda \leq U\right) = 1-\alpha$$

The MLE of λ is $\hat{\lambda} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$
Sufficiency Principle
Let $X_1, \ldots, X_n$ be a random sample and $T(\mathbf{X}) = T(\mathbf{x})$ is a sufficient statistic. Let $Y_1, \ldots, Y_n$ be another sample and $T(\mathbf{Y}) = T(\mathbf{y})$ be the sufficient statistic. If $T(\mathbf{x}) = T(\mathbf{y})$ then the inference we make on the parameter θ will be the same.

§10 Lec 10: Sep 2, 2021

§10.1 Sufficiency (Cont'd)

Definition 10.1 (Sufficient Statistic) — Let $X_1, \dots, X_n$ be a random sample and $T(\mathbf{X})$ be a function of $X_1, \dots, X_n$. We say that $T(\mathbf{X})$ is a sufficient statistic if the conditional distribution of $\mathbf{X} = \mathbf{x}$ $(X_1 = x_1, \dots, X_n = x_n)$ given $T(\mathbf{X}) = T(\mathbf{x})$ is free of θ and we check $\frac{L(X,\theta)}{q(T(\mathbf{x}),\theta)} = H(x_1, \dots, x_n)$ where L is the likelihood function and q is the pdf/pmf of $T(\mathbf{x})$.

Example 10.2

Let $X_1, \dots, X_n \overset{\text{i.i.d}}{\sim} \text{Bernoulli}(p)$. Is $Y = \sum X_i$ a sufficient statistic of p?
We know $Y \sim b(n,p)$. Then,

$$\begin{aligned}
\frac{L(\mathbf{x};p)}{q\left(T(\mathbf{x});p\right)} &= \frac{\prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}}{\binom{n}{y}p^y(1-p)^{n-y}} \\
&= \frac{p^{\sum x_i}(1-p)^{n-\sum x_i}}{\binom{n}{y}p^y(1-p)^{n-y}} \\
&= \frac{1}{\binom{n}{y}}
\end{aligned}$$

which is free of p. Therefore, $\sum X_i$ is a sufficient statistic of p.

Example 10.3

Let $X_1, \dots, X_n \overset{\text{i.i.d}}{\sim} \Gamma(\alpha, \beta)$. Suppose that α is known. Is $\sum X_i$ is a sufficient statistic?
First, the pdf of X_i is

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

and $Y = \sum X_i \sim \Gamma(n\alpha, \beta)$. So,

$$\begin{aligned}
\frac{L(\mathbf{x},\beta)}{q\left(T(\mathbf{x}),\beta\right)} &= \frac{\frac{(\prod x_i)^{\alpha-1}e^{-\sum x_i/\beta}}{\Gamma^n(\alpha)\beta^{n\alpha}}}{\frac{y^{n\alpha-1}e^{-\frac{y}{\beta}}}{\Gamma(n\alpha)b^{n\alpha}}} \\
&= \frac{\Gamma(n\alpha)(\prod x_i)^{\alpha-1}}{y^{n\alpha-1}\Gamma^n(\alpha)}
\end{aligned}$$

Yes, $\sum X_i$ is a sufficient statistic of β.

Example 10.4

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. Suppose σ^2 is known. Is $\overline{X}$ a sufficient statistic for μ?
First, $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. Therefore,

$$\begin{aligned}\frac{L(\mathbf{x}; \mu, \sigma^2)}{q\left(T(\mathbf{x}), \mu\right)} &= \frac{(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2}}{(2\pi\frac{\sigma^2}{n})^{-\frac{n}{2}} e^{-\frac{n}{2\sigma^2}(\overline{x}-\mu)^2}} \\ &= \frac{(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\overline{x})^2} e^{-\frac{n}{2\sigma^2}(\overline{x}-\mu)^2}}{(2\pi\frac{\sigma^2}{n})^{-\frac{n}{2}} \cdot e^{-\frac{n}{2\sigma^2}(\overline{x}-\mu)^2}} \\ &= \frac{(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\overline{x})^2}}{\left(2\pi\frac{\sigma^2}{n}\right)^{-\frac{n}{2}}}\end{aligned}$$

Thus we have a function free of μ. Note that

$$\begin{aligned}\sum(x_i - \mu \pm \overline{x})^2 &= \sum(x_i - \overline{x} + \overline{x} - \mu)^2 \\ &= \sum(x_i - \overline{x})^2 + n(\overline{x} - \mu)^2\end{aligned}$$

Theorem 10.5 (Factorization)

Let $X_1, \ldots, X_n$ be a random sample and $L(\mathbf{X}; \theta)$ be the likelihood function. We say that $T(\mathbf{X})$ is a sufficient statistic if

$$L(\mathbf{X}; \theta) = g\left(T(\mathbf{x}); \theta\right) \cdot h(x)$$

where $h(x) = 1$ if necessary.

Example 10.6

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \exp\left(\frac{1}{\lambda}\right)$. Find a sufficient statistic using the factorization theorem.

$$\begin{aligned}f(x_i) &= \frac{1}{\theta} e^{-\frac{1}{\theta}x_i} \\ L(\mathbf{x}; \theta) &= \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum x_i}\end{aligned}$$

Let $h(x) = 1$. Thus, $\sum x_i$ is a sufficient statistic.

Example 10.7

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. Suppose σ^2 is known.

$$\begin{aligned}L &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\overline{x})^2} e^{-\frac{n}{2\sigma^2}(\overline{x}-\mu)^2}\end{aligned}$$

We can see that $\overline{X}$ is sufficient statistic as there is no μ in the first two terms.

Example 10.8

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. Find sufficient statistic for μ and σ^2.

$$L = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\underbrace{\sum (x_i - \overline{x})^2}_{(n-1)S^2} - \frac{n}{2\sigma^2}(\overline{x} - \mu)^2\right)$$

Let $h(x) = 1$. Thus, $(\overline{X}, S^2)$ are sufficient statistic for (μ, σ^2).

Properties of Sufficient Statistic:

1. Let $T(\mathbf{x})$ be a sufficient statistic and $U^* = V[T(\mathbf{x})]$. Then, $T(\mathbf{x}) = V^{-1}(U^*)$. Show that U^* is a sufficient statistic.

   Proof. Using the factorization theorem, we get

$$\begin{aligned} L(\mathbf{x}; \theta) &= g(T(x); \theta) \cdot h(\mathbf{x}) \\ &= g\left(V^{-1}(U^*); \theta\right) \cdot h(\mathbf{x}) \\ &= g^*(U^*; \theta) \cdot h(\mathbf{x}) \end{aligned}$$

□

Example 10.9

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$.

$$L = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum x_i^2}$$

Using the factorization theorem, it follows that $\sum X_i^2$ is a sufficient statistic. Now, let $U^* = \frac{\sum X_i^2}{n}$. Then,

$$L = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}nu^*}$$

Thus, U^* is also a sufficient statistic.

2. Let $X_1, \ldots, X_n$ be a random sample and $L(\mathbf{X}; \theta)$ be the likelihood function. From the definition of sufficiency,

$$\frac{L(\mathbf{X}; \theta)}{q(T(\mathbf{X}); \theta)} = H(x_1, \ldots, x_n)$$

   Then,

$$L(\mathbf{X}; \theta) = q(T(\mathbf{X}); \theta) \cdot H(x_1, \ldots, x_n)$$

   The MLEs are function of sufficient statistics.

3. Exponential Families and Sufficient Statistics: Recall that if $f(x; \theta)$ belong in an exponential family then

$$f(x; \theta) = h(x)c(\theta)e^{\sum w_i(\theta)t_i(x)}$$

   Suppose $X_1, \ldots, X_n$ is a random sample from this distribution. Then,

$$L = \prod h(x_i)c^n(\theta)\exp\left(\sum_{i=1}^k\left(w_i(\theta)\sum_{j=1}^n t_i(x_j)\right)\right)$$

   From the factorization theorem, we can deduce that $(\sum_{i=1}^n t_1(x_i), \ldots, \sum_{i=1}^n t_k(x_i))$ are sufficient statistic for $(w_1(\theta), \ldots, w_k(\theta))$.

Example 10.10

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

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

Then,

$$L(\mathbf{x}; \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{n\mu^2}{2\sigma^2}} e^{-\frac{1}{2\sigma^2}\sum x_i^2 + \frac{\mu}{\sigma^2}\sum x_i}$$

Next, let's discuss a theorem that provides the minimum variance unbiased estimators (MVUE).

Theorem 10.11 (Rao-Blackwell)

Let $X_1, \dots, X_n$ be a random sample from a distribution with parameter θ. Let $\hat{\theta}$ be an unbiased estimator of θ and U be a sufficient statistic for θ. Now define a new estimator $\hat{\theta^*} = E\left[\hat{\theta}|U\right] = h(u)$. Then $\hat{\theta^*}$ is the MVUE of θ.

Proof. Let X, Y be random variables with joint pdf $f(x, y)$. Then

$$\begin{aligned} EX &= E\left[E[X|Y]\right] \\ \text{var}(X) &= \text{var}\left(E[X|Y]\right) + E\left[\text{var}\left(X|Y\right)\right] \end{aligned}$$

So,

$$E\hat{\theta^*} = E\left[E[\hat{\theta}|U]\right] = E\hat{\theta} = \theta$$

So the variance is

$$\begin{aligned} \text{var}(\hat{\theta}) &= \text{var}\left(E\left[\hat{\theta}|U\right]\right) + E\left[\underbrace{\text{var}\left(\hat{\theta}|U\right)}_{\geq 0}\right] \\ &= \text{var}(\hat{\theta^*}) + C(\geq 0) \end{aligned}$$

$\implies \text{var}(\hat{\theta^*}) \leq \text{var}(\hat{\theta})$. □

Example 10.12

Let $X_1, \dots, X_n \overset{\text{i.i.d}}{\sim} \exp\left(\frac{1}{\theta}\right)$. Find the MVUE of θ.
First, we need to find a sufficient statistic. Let's use factorization theorem.

$$L = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum x_i}$$

So, $\sum x_i$ is a sufficient statistic for θ with $E\sum x_i = n\theta$. Then, $\hat{\theta^*} = \frac{\sum X_i}{n}$ is an unbiased estimator of θ and therefore it is MVUE of θ.

Example 10.13

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \exp\left(\frac{1}{\theta}\right)$. Find the MVUE of θ.
Let's start with $\overline{X}^2$.

$$\begin{aligned} E\overline{X}^2 &= \operatorname{var}(\overline{X}) + (E\overline{X})^2 \\ &= \frac{\sigma^2}{n} + \mu^2 = \frac{\theta^2}{n} + \theta^2 = \frac{n+1}{n}\theta^2 \end{aligned}$$

Thus, the MVUE of θ^2 is $\hat{\theta^*} = \frac{n}{n+1}\overline{X}^2$.

<u>Minimal Sufficient Statistics</u>:

Theorem 10.14 (Lehmann-Scheffe)

Let $X_1, \ldots, X_n$ be a random sample from a distribution with parameter θ and let $Y_1, \ldots, Y_n$ be another random sample from the same distribution. Then we can find a minimal sufficient statistics iff

$$\frac{L(x_1, \ldots, x_n; \theta)}{L(y_1, \ldots, y_n; \theta)} \text{ is free of } \theta$$

iff $T(\mathbf{x}) = T(\mathbf{y})$.

Example 10.15

$X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Bernoulli}(p)$ and $Y_1, \ldots, Y_n \overset{\text{i.i.d}}{\sim} \text{Bernoulli}(p)$. Then,

$$\frac{L(x_1, \ldots, x_n; p)}{L(y_1, \ldots, y_n; p)} = \frac{p^{\sum x_i}(1-p)^{n-\sum x_i}}{p^{\sum y_i}(1-p)^{n-\sum y_i}} = \left(\frac{p}{1-p}\right)^{\sum x_i - \sum y_i}$$

The above expression is free of p iff $\sum x_i = \sum y_i$, and then $\sum x_i$ is a minimal sufficient statistic.

§10.2 Confidence Intervals

Recall that

$$P\left(L \leq \theta \leq U\right) = \underbrace{1-\alpha}_{\text{confidence level}}$$

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. Suppose σ^2 is known. Let's find a confident interval for μ.

$$\begin{gathered} \overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \\ \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \\ P\left(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha \\ P\left(-Z_{\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha \\ P\left(\overline{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \end{gathered}$$

So $\mu \in \overline{X} \pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$.

§11 Lec 11: Sep 7, 2021

§11.1 Confidence Intervals (Cont'd)

Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. We want to find a $1-\alpha$ confidence interval for μ (σ is known). We know $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. So the pivot is $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ and thus $\mu \in \overline{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$. Recall that

$$P\left(\overline{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Sample size:

$$\overline{X} \pm \underbrace{Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}_{E}$$

and so $n = \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{E}\right)^2$
Let's consider now the case σ is not known. Recall that

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

So $\mu \in \overline{X} \pm t_{\frac{\alpha}{2};n-1} \cdot \frac{S}{\sqrt{n}}$.

Example 11.1
Two dice are rolled and the sum X of the two numbers is recorded. The distribution has mean $\mu = 7$ and standard deviation $\sigma = 2.42$ in which $\mu = \sum xp(x)$ and $\sigma = \sqrt{\sum x^2 p(x) - \mu^2}$. By Central Limit Theorem, $\overline{X} \sim N(7, \frac{2.42}{\sqrt{50}})$.

$$\pm 1.96 = \frac{\overline{X} - 7}{2.42/\sqrt{50}}$$

$\implies$ lower mean $= 6.33$ and upper mean $= 7.67$, and we are 95% confident that the mean lies in this interval or the empirical mean falls into this interval 95% of the time.

With regard to confidence interval for variance σ^2 of normal distribution, let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma)$. Then, $\frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$. Thus,

$$P\left(\mathcal{X}^2_{\frac{\alpha}{2};n-1} \le \frac{(n-1)S^2}{\sigma^2} \le \mathcal{X}^2_{1-\frac{\alpha}{2};n-1}\right) = 1-\alpha$$

Through some manipulation we get

$$P\left(\frac{(n-1)s^2}{\mathcal{X}_{1-\frac{\alpha}{2};n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\mathcal{X}^2_{\frac{\alpha}{2};n-1}}\right) = 1-\alpha$$

Remark 11.2. When the sample size n is large, the $\mathcal{X}^2_{n-1}$ distribution can be approximated by $N\left(n-1, \sqrt{2(n-1)}\right)$. So the confidence interval can be computed as follows

$$\frac{s^2}{1 + z_{\frac{\alpha}{2}}\sqrt{\frac{2}{n-1}}} \le \sigma^2 \le \frac{s^2}{1 - z_{\frac{\alpha}{2}}\sqrt{\frac{2}{n-1}}}$$

Confidence interval for the difference between two population means $\mu_1 - \mu_2$:
Let $X_1, \dots, X_n \overset{\text{i.i.d}}{\sim} N(\mu_1, \sigma)$ and $Y_1, \dots, Y_m \overset{\text{i.i.d}}{\sim} N(\mu_2, \sigma)$ where σ^2 is unknown. We want to find a $1-\alpha$ confidence interval for $\mu_1 - \mu_2$.
We know

$$\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

Aside: Suppose σ is known. Then,

$$\mu_1 - \mu_2 \in \overline{X} - \overline{Y} \pm z_{\frac{\alpha}{2}} \sigma \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Recall from previous lecture that

$$\frac{(n+m-2)S_p^2}{\sigma^2} \sim \mathcal{X}^2_{n+m-2}$$

where $S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$. Using the two distribution we get

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

Finally, we have

$$\mu_1 - \mu_2 \in \overline{X} - \overline{Y} \pm t_{\frac{\alpha}{2};n+m-2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Simple Regression:
Have $Y_i = b_0 + b_1 X_i + \varepsilon_i$ where $\varepsilon_1, \dots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma)$. Construct a $1-\alpha$ confidence interval for b_1.

$$\left.\begin{array}{l} \hat{b_1} \sim N\left(b_1, \frac{\sigma}{\sqrt{\sum(x_i - \overline{x})^2}}\right) \\ \frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}^2_{n-2} \end{array}\right\} \implies \frac{\hat{b_1} - b_1}{S_e/\sqrt{\sum(x_i - \overline{x})^2}} \sim t_{n-2}$$

Thus,

$$b_1 \in \hat{b_1} \pm t_{\frac{\alpha}{2};n-2} \cdot \frac{S_e}{\sqrt{\sum(x_i - \overline{x})^2}}$$

§11.2 Prediction Intervals

Let $Y_1, \dots, Y_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$. We want to predict Y_0. It turns out that $\hat{Y_0} = \overline{Y}$. In order to show that, let $\hat{Y_0} = \sum a_i Y_i$. Then, we want it to be unbiased, i.e., $E\hat{Y_0} = \mu$ or $E\sum a_i Y_i = \mu \implies \sum a_i = 1$. Lastly, we need to minimize $E(Y_0 - \hat{Y_0})$ – mean square error prediction. Since $\hat{Y_0}$ is unbiased,

$$E\left(Y_0 - \hat{Y_0}\right)^2 = \operatorname{var}\left(Y_0 - \hat{Y_0}\right)$$

Use Lagrange multiplier the optimization problem becomes

$$\begin{aligned} \min Q &= \operatorname{var}(y_0 - \hat{y_0}) - 2\lambda(\sum a_i - 1) \\ &= \sigma^2 + \sigma^2 \sum a_i^2 - 2\lambda(\sum a_i - 1) \end{aligned}$$

Then,

$$\frac{\partial Q}{\partial a_i} = 2\sigma^2 a_i - 2\lambda = 0$$
$$\implies a_i = \frac{\lambda}{\sigma^2}$$

Let's find λ.

$$\underbrace{\sum a_i}_{1} = \frac{n\lambda}{\sigma^2} \implies \lambda = \frac{\sigma^2}{n}$$

$\implies a_i = \frac{1}{n}$ and thus $\hat{Y_0} = \sum a_i Y_i = \overline{Y}$.

§11.3 Hypothesis Testing

Definition 11.3 (Hypothesis Test) — A hypothesis test is a claim about a parameter of a population. The two hypotheses are called "null" and "alternative" hypotheses which are denoted with H_0 and H_a respectively. Note that

$$H_0: \ \theta \in \Theta_0$$
$$Ha: \ \theta \in \Theta_0^c$$

Example 11.4 1. Consider the simple regression model: $y_i = b_0 + b_1 x + \varepsilon_i$. We want to test $H_0: \ b_1 = 0$, i.e., there is no association between the respond and predictor variable. As a result, $H_a: \ b_1 \neq 0$, i.e., there is a linear association between y and x.

2. Test for proportion of defective items at a certain production line:

$$H_0: \ p = p_0$$
$$H_a: \ p > p_0$$

Assume $X_1, \dots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$ and $H_0: \ \mu = 10$, $H_a: \ \mu > 10$. We know $\overline{X} \sim N(10, \frac{\sigma}{\sqrt{n}})$. Suppose $\alpha = 5\%$.

$$P\left(T(x) \in RR | H_0\right) = 5\% - \text{Type I Error}$$

where RR refers to rejection region. Notice that

- Type I error $\alpha = P\left(\text{falsely rejecting } H_0\right)$
- Type II error $\beta = P\left(\text{falsely accepting } H_0\right)$ where $1 - \beta =$ power of the test.

Theorem 11.5 (Neyman-Pearson)

Suppose X is a random variable and we need to decide whether the probability distribution is either $f_0(x)$ or $f_1(x)$. Let k be some positive number, and define the following two sets

$$A = \left\{x \middle| \frac{f_0(x)}{f_1(x)} > k\right\}$$
$$R = \left\{x \middle| \frac{f_0(x)}{f_1(x)} < k\right\}$$

If data x is in set A, then we accept H_0. Similarly, if data x is in set R then we accept H_a.

Example 11.6

Let X be a single observation from the probability density function $f(x) = \theta x^{\theta-1}$ where $0 < x < 1$. Find the most powerful test using significance level $\alpha = 0.05$ for testing $H_0 : \ \theta = 1$ and $H_a : \ \theta = 2$.
We reject H_0 if

$$\frac{f_0(x)}{f_1(x)} < k$$
$$\frac{1}{2x} < k$$
$$x > \frac{1}{2k} = k'$$

which is the best critical region.

Question 11.1. How do we find k'?

Using $\alpha = 0.05$

$$P(X > k'|H_0) = 0.05$$
$$\int_{k'}^{1} dx = 0.05$$
$$1 - k' = 0.05 \implies k' = 0.95$$

Notice that the power of the test is $1 - \beta = P\left(X > 0.95|H_a\right)$.

Example 11.7

$X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$ where σ^2 is known. We test $H_0 : \ \mu = \mu_0$ and $H_a : \ \mu > \mu_0$.
We reject H_0 if

$$\frac{L(\theta_0)}{L(\theta_a)} < k$$
$$\frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i - \mu_0)^2\right)}{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i - \mu_a)^2\right)} < k$$
$$\sum x_i^2 + n\mu_a - 2\mu_a \sum x_i - \sum x_i^2 - n\mu_0^2 + 2\mu_0 \sum x_i < 2\sigma^2 \ln k$$
$$2\underbrace{(\mu_0 - \mu_a)}_{<0} \sum x_i < 2\sigma^2 \ln k - n\mu_a^2 + n\mu_0^2$$
$$\sum x_i/n > \frac{2\sigma^2 \ln k - n\mu_a^2 + n\mu_0^2}{2(\mu_0 - \mu_a)n}$$
$$\overline{X} > k'$$

which is the best critical region of size α. So

$$P\left(\overline{X} > k' \middle| H_0\right) = \alpha$$
$$\frac{k' - \mu_0}{\sigma/\sqrt{n}} = Z_\alpha$$
$$k' = \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}}$$

Example 11.8 a) $H_0: \ \lambda = 3$, $H_a: \ \lambda = 5$, and $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \exp\left(\frac{1}{\lambda}\right)$

$$\frac{L(\theta_0)}{L(\theta_a)} < k$$

$$\frac{\left(\frac{1}{3}\right)^n e^{-\frac{1}{3}\sum x_i}}{\left(\frac{1}{5}\right)^n e^{-\frac{1}{5}\sum x_i}} < k$$

$$e^{\left(\frac{1}{5}-\frac{1}{3}\right)\sum x_i} < \left(\frac{3}{5}\right)^n k$$

$$\sum x_i > \frac{\ln\left(\frac{3}{5}\right)^n k}{\left(\frac{1}{5}-\frac{1}{3}\right)} = k'$$

b) If $n = 12$ and using $\frac{2}{\lambda}\sum X_i \sim \mathcal{X}^2_{24}$, find the best critical region when the significance level $\alpha = 0.05$.

$$P\left(\sum X_i > k' \middle| H_0\right) = 0.05$$

$$P\left(2\frac{1}{3}\sum X_i > \frac{2}{3}k'\right) = 0.05$$

$$P\left(\mathcal{X}^2_{24} > \frac{2}{3}k'\right) = 0.05$$

$$\mathcal{X}^2_{0.95;24} = 36.42$$

$$\frac{2}{3}k' = 36.42 \implies k' \approx 54$$

§12 Lec 12: Sep 9, 2021

§12.1 Hypothesis Testing (Cont'd)

Recall the Neyman-Pearson Lemma: $X_1, \ldots, X_n$ random sample. We reject H_0 if

$$\frac{L(\theta_0)}{L(\theta_a)} < k$$

Example 12.1

$H_0: \ \mu = \mu_0$, $H_a: \ \mu > \mu_0$ and we have normal distribution (i.i.d) with n random variables and σ is known. The best critical rejection region of size α is $\overline{X} > k'$.

$$P\left(\overline{X} > k' | H_0\right) = \alpha$$

Under H_0, $\overline{X} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$. Thus,

$$k' = \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}}$$

Power of the test:

$$1 - \beta = P\left(\text{rejecting } H_0 \text{when } H_0 \text{ is false}\right) = P\left(\overline{X} > k' | H_0 \text{ is false then } \mu = \mu_a\right)$$

So

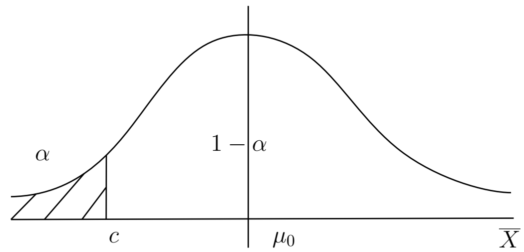
$$\begin{aligned} 1 - \beta &= P\left(\overline{X} > \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}}\right) \\ &= P\left(Z > \frac{\mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}} - \mu_a}{\sigma/\sqrt{n}}\right) \end{aligned}$$

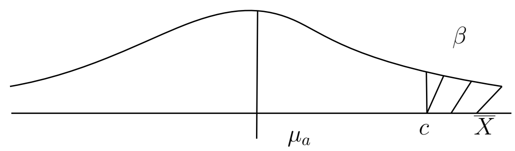
Observe how the following parameters affect the power

Parameters	$1-\beta$
$\alpha \downarrow$	$\downarrow$
$n \uparrow$	$\uparrow$
$\sigma \downarrow$	$\uparrow$
$\lvert\mu_a - \mu_0\rvert \uparrow$	$\uparrow$

<u>One-sided Test</u>:

$$H_0: \ \mu = \mu_0$$
$$H_a: \ \mu < \mu_0$$

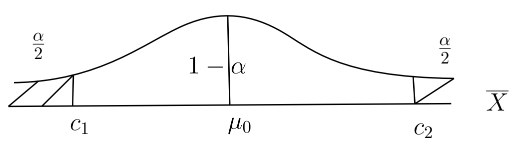


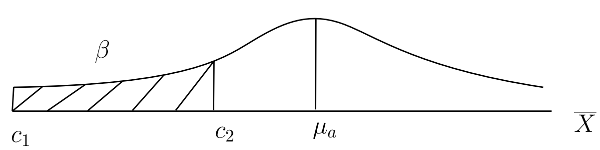


The power then is

$$\begin{aligned} 1-\beta &= P\left(\overline{X} < c | H_a\right) \\ &= P\left(\overline{X} < \mu_0 - Z_\alpha \frac{\sigma}{\sqrt{n}}\right) \\ &= P\left(Z < \frac{\mu_0 - Z_\alpha \frac{\sigma}{\sqrt{n}} - \mu_a}{\sigma/\sqrt{n}}\right) \end{aligned}$$

<u>Two-sided Test</u>: $H_0: \ \mu = \mu_0$, $H_a: \ \mu \neq \mu_0$





The power then is

$$\begin{aligned} 1-\beta &= P(\overline{X} < c_1) + P(\overline{X} > c_2) \\ &= P\left(\overline{X} < \mu_0 - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) + P\left(\overline{X} > \mu_0 + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \\ &= P\left(Z < \frac{\mu_0 - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} - \mu_a}{\sigma/\sqrt{n}}\right) + P\left(Z > \frac{\mu_0 + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} - \mu_a}{\sigma/\sqrt{n}}\right) \end{aligned}$$

Sample Size Determination:
Given α, β and $H_0: \ \mu = \mu_0$, $H_a: \ \mu > \mu_0$. Find n in order to detect a shift from μ_0 to μ_a (assume σ is known).

$$\left.\begin{aligned} Z_\alpha &= \tfrac{c-\mu_0}{\sigma/\sqrt{n}} \\ -Z_\beta &= \tfrac{c-\mu_a}{\sigma/\sqrt{n}} \end{aligned}\right\} \implies \left.\begin{aligned} c-\mu_0 &= Z_\alpha \tfrac{\sigma}{\sqrt{n}} \\ c-\mu_a &= -Z_\beta \tfrac{\sigma}{\sqrt{n}} \end{aligned}\right\}$$

Thus,

$$n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2}$$

Note that we always use positive values for Z_α, Z_β.

Example 12.2

For a certain candidate's political poll $n = 15$ voters are sampled. Assume that this sample is taken from an infinite population voters. We wish to test $H_0: \ p = 0.5$ against the alternative $H_a: \ p < 0.5$. The test statistic is X, which is the number of voter among the 15 sampled favoring this candidate.

a) Calculate the probability of type I error α for $RR = \{x \le 2\}$

$$\begin{aligned} \alpha &= P\left(X \le 2 | p = 0.5\right) \\ &= \sum_{x=0}^{2} \binom{15}{x} 0.5^x 0.5^{15-x} \end{aligned}$$

b) How about type II error where $p = 0.3$

$$\beta = P(X > 2) = \sum_{x=3}^{15} \binom{15}{x} 0.3^x 0.7^{15-x}$$

§12.2 Likelihood Ratio Test

We reject H_0 if

$$\Lambda = \frac{L(\hat{u})}{L(\hat{\Omega})} < k$$

where $L(\hat{u})$ is the maximized likelihood function under H_0 and $L(\hat{\Omega})$ is the maximized likelihood function under no restrictions.

Example 12.3

$X_1, \dots, X_n \overset{\text{i.i.d}}{\sim} N(\mu, \sigma)$ where σ^2 is unknown. $H_0 : \mu = \mu_0$, $H_a : \mu \neq \mu_0$.

$$\left.\begin{array}{l} \overline{X} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right) \\ \frac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2_{n-1} \end{array}\right\} \implies \frac{\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1} - \text{test statistic}$$

We reject H_0 if $t > t_{\frac{\alpha}{2};n-1}$ or $t < -t_{\frac{\alpha}{2};n-1}$.
Another approach is to use likelihood ratio test. Under H_0,

$$L = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right)$$
$$\ln L = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum(x_i-\mu_0)^2$$

MLE of σ^2 is

$$\hat{\sigma_0}^2 = \frac{\sum(x_i-\mu_0)^2}{n}$$

Under no restrictions the MLEs of μ and σ^2 are $\overline{X}$ and $\hat{\sigma_1}^2 = \frac{\sum(x_i-\overline{x})^2}{n}$. Then,

$$\Lambda = \frac{(2\pi\hat{\sigma_0}^2)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\hat{\sigma_0}^2}\sum(x_i-\mu_0)^2\right)}{(2\pi\hat{\sigma_1}^2)^{-\frac{n}{2}}\exp\left(-\frac{1}{2\hat{\sigma_1}^2}\sum(x_i-\overline{x})^2\right)} < k$$

Note: $\sum(x_i-\mu_0)^2 = n\hat{\sigma_0}^2$ and $\sum(x_i-\overline{x})^2 = n\hat{\sigma_1}^2$.

$$\left(\frac{\hat{\sigma_1}^2}{\hat{\sigma_0}^2}\right)^{\frac{n}{2}} \frac{e^{-\frac{n}{2}}}{e^{-\frac{n}{2}}} < k$$
$$\frac{\hat{\sigma_1}^2}{\hat{\sigma_0}^2} < k^{\frac{2}{n}}$$
$$\frac{\frac{\sum(x_i-\overline{x})^2}{n}}{\frac{\sum(x_i-\mu_0)^2}{n}} < k^{\frac{2}{n}}$$
$$\frac{\sum(x_i-\overline{x})^2}{\sum(x_i-\mu_0)^2} < k^{\frac{2}{n}}$$
$$\frac{\sum(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2 + n(\overline{x}-\mu_0)^2} < k^{\frac{2}{n}}$$
$$\frac{1}{1+\frac{n(\overline{x}-\mu_0)^2}{\sum(x_i-\overline{x})^2}} < k^{\frac{2}{n}}$$

So,

$$\left(\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right)^2 > (n-1)\left(k^{-\frac{2}{n}}-1\right) = k'$$

or $F_{1,n-1} = t^2_{n-1} > k'$.

$$P\left(t^2_{n-1} > k'\right) = \alpha$$
$$P\left(-\sqrt{k'} \le t_{n-1} \le \sqrt{k'}\right) = 1-\alpha$$

Asymptotic Result:

$$-2\ln\Lambda \sim \mathcal{X}^2_{\gamma_0-\gamma}$$

where γ_0 is number of free parameters under H_0 and γ is number of free parameters under no restrictions.

§12.3 Power Analysis

For unknown σ, if H_0 is true then

$$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Then,

$$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

If H_0 is not true

$$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(\frac{\mu_a-\mu_0}{\sigma/\sqrt{n}},1)$$

Then

$$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}\left(NCP=\frac{\mu_a-\mu_0}{\sigma/\sqrt{n}}\right)$$

Now, the power is

$$1-\beta = P\left(t_{n-1}(NCP=\delta) > t_{\frac{\alpha}{2};n-1}\right) + P\left(t_{n-1}(NCP=\delta) < -t_{\frac{\alpha}{2};n-1}\right)$$

To compute δ we need μ_a and σ.
Hypothesis Testing in Regression:
Have

$$\begin{aligned} Y_i &= b_0 + b_1X_i + \varepsilon_i \\ H_0 &: \ b_1 = 0 \\ H_a &: \ b_1 \neq 0 \end{aligned}$$

Recall that

$$\hat{b_1} \sim N\left(0, \frac{\sigma}{\sqrt{\sum(x_i-\overline{x})^2}}\right)$$

and

$$\frac{(n-2)S_e^2}{\sigma^2} \sim \mathcal{X}^2_{n-2}$$

Then,

$$\frac{\hat{b_1}}{S_e/\sqrt{\sum(x_i-\overline{x})^2}} \sim t_{n-2}$$

§12.4 Two Sample t Test

$X_1,\ldots,X_n \overset{\text{i.i.d}}{\sim} N(\mu_1,\sigma)$ and $Y_1,\ldots,Y_n \overset{\text{i.i.d}}{\sim} N(\mu_2,\sigma)$ where σ^2 is common but unknown. We want to test

$$\begin{aligned} H_0 &: \ \mu_1 = \mu_2 \quad (\text{or } \mu_1-\mu_2 = 0) \\ H_a &: \ \mu_1 \neq \mu_2 \quad (\text{or } \mu_1-\mu_2 \neq 0) \end{aligned}$$

Under H_0, we have

$$\left.\begin{aligned} &\overline{X}-\overline{Y} \sim N\left(0, \sigma\sqrt{\tfrac{1}{n}+\tfrac{1}{m}}\right) \\ &\tfrac{(n+m-2)S_p^2}{\sigma^2} \sim \mathcal{X}^2_{n+m-2} \end{aligned}\right\} \implies \frac{\overline{X}-\overline{Y}}{\sqrt{S_p^2\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim t_{n+m-2}$$

Using likelihood ratio test: Under H_0,

$$\begin{aligned} L &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2} (2\pi\sigma^2)^{-\frac{m}{2}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2} \\ \hat{\mu} &= \frac{n\overline{x}+m\overline{y}}{n+m} \\ \hat{\sigma_0^2} &= \frac{\sum(x_i-\hat{\mu})^2+\sum(y_i-\hat{\mu})^2}{n+n} \end{aligned}$$

Under no restriction, there are three free parameters

$$\begin{aligned} \hat{\mu_1} &= \overline{x} \\ \hat{\mu_2} &= \overline{y} \\ \hat{\sigma_1^2} &= \frac{\sum(x_i-\overline{x})^2+\sum(y_i-\overline{y})^2}{n+m} \end{aligned}$$

Then,

$$\begin{gathered} \frac{L(\hat{u})}{L(\hat{\Omega})} < k \\ \frac{\hat{\sigma_1}^2}{\hat{\sigma_0}^2} < k^{\frac{2}{n+m}} \\ \frac{\sum(x_i-\overline{x})^2+\sum(y_i-\overline{y})^2}{\sum(x_i-\hat{\mu})^2+\sum(y_i-\hat{\mu})^2} < k^{\frac{2}{n+m}} \end{gathered}$$

Note that

$$\begin{aligned} \sum(x_i-\hat{\mu})^2 &= \sum(x_i \pm \overline{x} - \frac{n\overline{x}+m\overline{y}}{n+m}) \\ &= \sum(x_i-\overline{x})^2 + n\left(\overline{x}-\frac{n\overline{x}+m\overline{y}}{n+m}\right)^2 \end{aligned}$$

Apply the same trick to the y term and we can simplify the inequality.