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. □

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

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

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

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

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

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

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

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

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,

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

§3 Lec 3: Aug 10, 2021

§3.1 Method of Transformation

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

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

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

Example 3.1 (Increasing Function Case)
Let $Y = 3X - 1$.

- Method of CDF:

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

  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 $f_{x_1x_2}(x_1, x_2)$. 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

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

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

$$M_{X_1X_3}(t_1, t_3) = M_X(t_1, 0, t_3) = (1 - t_1)^{-6}(1 - t_1 + 3t_3)^{-3}$$

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 $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix}$

$$M_{X_1X_3}(t_1, 0, t_3) = (1 - t_1)^{-6}(1 - t_1 + 3t_3)^{-3}$$

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 $M_{X_1X_3}(t_1, t_3) \neq M_{X_1}(t_1) \cdot M_{X_3}(t_3)$. 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.

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

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 $N(\mathbf{0}, I)$.
<u>*Note*</u>: 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

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

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \left(\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}\right) \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix} = \mathbf{AY}$$

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 linear combination 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

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

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

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

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

Question 4.1. Find $\operatorname{cov}(\mathbf{U}, \mathbf{V})$ using $\operatorname{cov}(\mathbf{AY}, \mathbf{BY}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\top$

We have

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

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

$$\begin{aligned} \mathbf{Q}_1 | \mathbf{Q}_2 &= \mathbf{U} | \mathbf{V} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Q}_2 \\ &= \mathbf{U} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Q_2} \\ E(\mathbf{Q}_1 | \mathbf{Q}_2) &= \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Q_2} \\ &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{Q}_2 - \boldsymbol{\mu}_2) \\ \operatorname{var}(\mathbf{Q_1} | \mathbf{Q_2}) &= \operatorname{var}(\mathbf{U}) \\ &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \end{aligned}$$