

# STAT 3100 Lecture Note

Week Three (Sep 20 & 22, 2022)

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca, zhiyanggeezhou.github.io)

2022/Sep/21 22:31:49

---

## Multivariate normal (MVN) distribution (con'd)

### Marginals of MVN

- Suppose  $p$ -vector  $\mathbf{X} = [X_1, \dots, X_p]^\top$  and  $q$ -vector  $\mathbf{Y} = [Y_1, \dots, Y_q]^\top$  are jointly normally distributed. Then,  $\mathbf{X}$  and  $\mathbf{Y}$  are independent  $\Leftrightarrow \text{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}_{p \times q}$ .
- If  $\mathbf{X}$  is of MVN, then its all margins are still of MVN. The inverse proposition does NOT hold.
  - Cautionary example: Let  $Y = XZ$ , where  $X \sim \mathcal{N}(0, 1)$ ;  $Z$  is independent of  $X$  with  $\Pr(Z = 1) = \Pr(Z = -1) = .5$ .  $X$  and  $Y$  both turn out to be of standard normal, but they are not jointly normal. (Why?)

---

```
if (!("plot3D" %in% rownames(installed.packages())))  
  install.packages("plot3D")  
set.seed(1)  
xsize = 1e4L  
X = rnorm(xsize)  
Z = rbinom(n = xsize, 1, .5)  
Y = (2 * Z - 1) * X  
# 3d histogram of (X, Y)  
plot3D::hist3D(z=table(cut(X, 100), cut(Y, 100)), border = "black")  
# plot the support of joint pdf of (X, Y)  
plot3D::image2D(z=table(cut(X, 100), cut(Y, 100)), border = "black")
```

## Normal sampling theory (CB Sec. 5.3)

(Default) stochastic representations for  $\chi^2$ -,  $t$ -, and  $F$ -r.v. (HMC Chp. 3)

- $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$  if  $[X_1, \dots, X_n]^\top \sim \text{MVN}(\mathbf{0}, \mathbf{I}_n)$ ;
- $X/\sqrt{Y/n} \sim t(n)$  if  $X \sim \mathcal{N}(0, 1)$  and  $Y \sim \chi^2(n)$  are independent;
- $(X/m)/(Y/n) \sim F(m, n)$  if  $X \sim \chi^2(m)$  and  $Y \sim \chi^2(n)$  are independent.

### Important identities for iid normal samples

Let  $\mathbf{X} = [X_1, \dots, X_n]^\top \sim \text{MVN}(\mu \mathbf{1}_n, \sigma^2 \mathbf{I}_n)$ ,  $\bar{X} = n^{-1} \sum_{i=1}^n X_i = n^{-1} \mathbf{1}_n^\top \mathbf{X}$ , and  $S^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2 = (n-1)^{-1} \mathbf{X}^\top (\mathbf{I}_n - n^{-1} \mathbf{1}_n \mathbf{1}_n^\top) \mathbf{X}$ , where  $\mathbf{1}_n = [1, \dots, 1]^\top$ , i.e., a column  $n$ -vector whose entries are all

ones.

- $n^{1/2}(\bar{X} - \mu)/\sigma \sim \mathcal{N}(0, 1)$
- $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$
- $\bar{X}$  and  $S^2$  are independent of each other
- $n^{1/2}(\bar{X} - \mu)/S \sim t(n-1)$

## Taylor series (CB Def 5.5.20 & Thm 5.5.21)

### Taylor series about $x_0 \in \mathbb{R}$ for univariate functions

- Suppose  $f$  has derivative of order  $n+1$  within an open interval of  $x_0$ , say  $(x_0 - \varepsilon, x_0 + \varepsilon)$  with  $\varepsilon > 0$ . Then, for  $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$ ,

$$f(x) \approx \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k = f(x_0) + \sum_{k=1}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k,$$

where  $f^{(n)}(x_0) = \frac{d^n}{dx^n} f(x)|_{x=x_0}$ .

- Called the Maclaurin series if  $x_0 = 0$

### Taylor series about $\mathbf{x}_0 \in \mathbb{R}^p$ for multivariate functions

Under regularity conditions,

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^\top \nabla f(\mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^\top \mathbf{H}(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0),$$

where the gradient  $\nabla f(\mathbf{x}_0) = [\frac{\partial}{\partial x_1} f(\mathbf{x}_0), \dots, \frac{\partial}{\partial x_p} f(\mathbf{x}_0)]^\top$  and the Hessian  $\mathbf{H}(\mathbf{x}_0) = [\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}_0)]_{p \times p}$ .

### Application

- Approximate unknown or complex  $f$  with a polynomial
  - $\Delta$ -method
  - Asymptotic theory for maximum likelihood estimators
- Moment generating function (mgf):  $M_X(t) = E\{\exp(tX)\} = \sum_{n=0}^{\infty} t^n E(X^n)/n!$ 
  - Maclaurin series of  $\exp(tX)$ :  $\exp(tX) = \sum_{n=0}^{\infty} (tX)^n/n! \Rightarrow E(X^n) = (\partial^n / \partial t^n) M_X(t) |_{t=0}$

## Generating functions

### Moment generating function (mgf, CB Sec. 2.3)

- Univariate r.v.  $X$ 
  - mgf  $M_X(t) = E\{\exp(tX)\}$  if  $E\{\exp(tX)\} < \infty$  for  $t$  in a neighborhood of 0; otherwise we say that the mgf does not exist or is undefined.
    - \* Continuous  $X$ :  $M_X(t) = \int_{-\infty}^{\infty} \exp(tx) f_X(x) dx$
    - \* Discrete  $X$ :  $M_X(t) = \sum_{\{x: x \in \text{supp}(X)\}} \exp(tx) p_X(x)$
  - $M_{aX+b}(t) = \exp(bt) M_X(at)$

- 
- Multivariate r.v.  $\mathbf{X} = (X_1, \dots, X_p)^\top \in \mathbb{R}^p$

- mgf  $M_{\mathbf{X}}(\mathbf{t})$  is defined as

$$M_{\mathbf{X}}(\mathbf{t}) = E\{\exp(\mathbf{t}^\top \mathbf{X})\} = \begin{cases} \int_{\mathbb{R}^p} \exp(\mathbf{t}^\top \mathbf{X}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} & \text{continuous } \mathbf{X} \\ \sum_{\{\mathbf{x}: \mathbf{x} \in \text{supp}(\mathbf{X})\}} \exp(\mathbf{t}^\top \mathbf{X}) p_{\mathbf{X}}(\mathbf{x}) & \text{discrete } \mathbf{X} \end{cases}$$

provided that  $E\{\exp(\mathbf{t}^\top \mathbf{X})\} < \infty$  for  $\mathbf{t} = [t_1, \dots, t_p]^\top$  in some neighborhood of  $\mathbf{0}$ ; otherwise we say that the mgf does not exist or is undefined.

- $M_{\mathbf{A}\mathbf{X}+\mathbf{b}}(\mathbf{t}) = \exp(\mathbf{b}^\top \mathbf{t}) M_{\mathbf{X}}(\mathbf{A}^\top \mathbf{t}) = \exp(\mathbf{b}^\top \mathbf{t}) E\{\exp(\mathbf{t}^\top \mathbf{A}\mathbf{X})\}$ 
  - \* Specifically, independent  $X_1, \dots, X_p \Rightarrow M_{\mathbf{X}}(\mathbf{t}) = \prod_{i=1}^p M_{X_i}(t_i)$

- Application
  - Computing moments
    - \*  $n$ th raw moment  $\mu'_n = EX^n = \sum_{k=0}^n \binom{n}{k} \mu_k (\mu'_1)^{n-k}$
    - \*  $n$ th central moment  $\mu_n = E(X - EX)^n = \sum_{k=0}^n \binom{n}{k} (-1)^{n-k} \mu'_k (\mu'_1)^{n-k}$
  - Proving laws of large numbers and central limit theorems
    - \* A distribution is uniquely determined by its mgf if the mgf is well-defined

## Example Lec6.1

- Find the mgfs of following distributions.
  - $\mathcal{N}(\mu, \sigma^2)$ .
  - $\text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .
  - Cauchy distribution:  $f_X(x) = \{\pi(1+x^2)\}^{-1}$ ,  $x \in \mathbb{R}$ .

## Characteristic function

- For univariate  $X$ :  $\phi_X(t) = E \exp(itX)$  for all  $t \in \mathbb{R}$
- For Multivariate  $\mathbf{X} = [X_1, \dots, X_p]^\top$ :  $\phi_{\mathbf{X}}(\mathbf{t}) = E\{\exp(i\mathbf{t}^\top \mathbf{X})\}$  for all  $\mathbf{t} \in \mathbb{R}^p$
- $\phi_{\mathbf{X}}(\mathbf{t}) = \phi_{\mathbf{Y}}(\mathbf{t})$  for all  $\mathbf{t} \in \mathbb{R}^p \Leftrightarrow \mathbf{X} \stackrel{d}{=} \mathbf{Y}$

## Example Lec6.2

- Find the characteristic functions of following distributions.
  - $\mathcal{N}(\mu, \sigma^2)$ .
  - $\text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .
  - Cauchy distribution:  $f_X(x) = \{\pi(1+x^2)\}^{-1}$ ,  $x \in \mathbb{R}$ .

## Other generating functions

- Cumulant generating function
  - $K_X(t) = \ln M_X(t) = \sum_{n=0}^{\infty} \kappa_n t^n / n!$
  - $\kappa_n = K_X^{(n)}(0)$
- Probability-generating function
  - For discrete r.v.  $X$  taking values from  $\{0, 1, \dots\}$ ,  $G(z) = Et^X = \sum_{x=0}^{\infty} t^x p_X(x)$ .
  - $p_X(n) = \Pr(X = n) = G^{(n)}(1)/n!$

## Estimating equations

### Parametric models

- A parametric model is a set of distributions indexed by unknown  $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$  with small or moderate  $p$ 
  - Say  $\{f(\cdot | \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p\}$ , where  $f$  is either a pdf or a pmf

- Believed that the true parameter (vector)  $\theta_0$  ( $\in \Theta \subset \mathbb{R}^p$ ) is fixed
  - Rather than making  $\theta_0$  random in the Bayesian philosophy

### Method of moments (MOM, a.k.a. moment matching, CB Sec 7.2.1)

- Procedure
  1. Equate raw moments to their empirical counterparts.
  2. Solve the resulting simultaneous equations for  $\theta = [\theta_1, \dots, \theta_p]^\top$ .
- Features
  - Easy implementation
  - Start point for more complex methods
  - Ad hoc
  - No constraint
  - Not unique
  - No guarantee on optimality

### Exercise Lec7.1

- Let  $X_1, \dots, X_n$  iid follow the following distributions. Find MOM estimators for  $(\theta_1, \theta_2)$ .
  - a.  $N(\theta_1, \theta_2)$ .
  - b.  $\text{Binom}(\theta_1, \theta_2)$  with pmf

$$p_X(x \mid \theta_1, \theta_2) = \binom{\theta_1}{x} \theta_2^x (1 - \theta_2)^{\theta_1 - x} \mathbf{1}_{\{0, \dots, \theta_1\}}(x).$$

### Exercise Lec7.2

- Let  $X_1, \dots, X_n$  iid follow pdf  $f(x \mid \theta) = \theta x^{\theta-1} I_{\mathbb{R}^+ \times [0,1]}(\theta, x)$ .
  - a. Find an MOM estimator of  $\theta$ .
  - b. Can we employ the second (raw) moment instead of the first one?