

# STAT 4100 Lecture Note

Week Four (Sep 26 & 28, 2022)

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

2022/Sep/29 16:15:30

---

## Generating functions (con'd)

### Moment generating function (con'd)

- Application
  - Characterizing distributions:  $M_{\mathbf{X}}(\mathbf{t})$  and  $M_{\mathbf{Y}}(\mathbf{t})$  are both well-defined and equal for all  $\mathbf{t}$  in a neighborhood of  $\mathbf{0} \Leftrightarrow \mathbf{X} \stackrel{d}{=} \mathbf{Y}$ 
    - \* Proofs for laws of large numbers and central limit theorems.
  - 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  $\mu_n = E(X - EX)^n = \sum_{k=0}^n \binom{n}{k} (-1)^{n-k} \mu'_k (\mu'_1)^{n-k}$

### Characteristic function

- For univariate  $X$ :  $\phi_X(t) = E \exp(itX)$  for all  $t \in \mathbb{R}$ 
  - Fourier transform of  $f_X$
  - Inverse:  $f_X(x) = (2\pi)^{-1} \int_{\mathbb{R}} \phi_X(t) \exp(-itx) dt$
  - $\mu'_n = EX^n = (-i)^n \phi_X^{(n)}(0)$
- 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  $\mathbf{t} \in \mathbb{R}^p$ 
  - Fourier transform of  $f_{\mathbf{X}}$
  - Inverse:  $f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-p} \int_{\mathbb{R}^p} \phi_{\mathbf{X}}(\mathbf{t}) \exp(-i\mathbf{t}^\top \mathbf{x}) d\mathbf{t}$
- $\phi_{\mathbf{X}}(\mathbf{t}) = \phi_{\mathbf{Y}}(\mathbf{t})$  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  $\theta \in \Theta \subset \mathbb{R}^p$  with small or moderate  $p$ 
  - Say  $\{f(\cdot | \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ , where  $f$  is either a pdf or a pmf and  $\Theta$  is the set of all the possible values of  $\theta$
- 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

## Exponential family (CB Sec 3.4)

- Original parameterization

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

- Normal (CB Example 3.4.4):
  - $h(x) = \mathbf{1}_{\mathbb{R}}(x)$
  - $c(\mu, \sigma) = (2\pi\sigma^2)^{-1/2} \exp\{-\mu^2/(2\sigma^2)\} \mathbf{1}_{\mathbb{R}}(\mu) \mathbf{1}_{\mathbb{R}^+}(\sigma)$
  - $w_1(\mu, \sigma) = \sigma^{-2} \mathbf{1}_{\mathbb{R}^+}(\sigma)$  &  $w_2(\mu, \sigma) = \mu \sigma^{-2} \mathbf{1}_{\mathbb{R}^+}(\sigma)$
  - $t_1(x) = -x^2/2$  &  $t_2(x) = x$
- Binomial (CB Example 3.4.1):
  - $h(x) = \binom{n}{x} \mathbf{1}_{\{0, \dots, n\}}(x)$
  - $c(p) = (1-p)^n \mathbf{1}_{(0,1)}(p)$
  - $w_1(p) = \ln\{p/(1-p)\} \mathbf{1}_{(0,1)}(p)$
  - $t_1(x) = x$
- Other special cases: gamma, beta, Poisson, negative binomial

## Method of moments (MOM, CB Sec 7.2.1)

- Procedure
  1. Equate raw moments to their empirical counterparts.
  2. Solve the resulting simultaneous equations for  $\theta = (\theta_1, \dots, \theta_p)$ .
- Features
  - Easy implementation
  - Start point for more complex methods
  - No constraint
  - Not uniquely defined
  - 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.  $N(\theta_1, \theta_2)$ ,  $(\theta_1, \theta_2) \in \mathbb{R} \times \mathbb{R}^+$ .
  - b.  $\text{Binom}(\theta_1, \theta_2)$  with pmf

$$p_X(x | \theta_1, \theta_2) = \binom{\theta_1}{x} \theta_2^x (1 - \theta_2)^{\theta_1 - x} \mathbf{1}_{\{0, \dots, \theta_1\}}(x), \quad (\theta_1, \theta_2) \in \mathbb{Z}^+ \times (0, 1).$$

## Exercise Lec7.2

- Let  $X_1, \dots, X_n$  iid follow pdf  $f(x | \theta) = \theta x^{\theta-1} \mathbf{1}_{[0,1]}(x)$ ,  $\theta > 0$ .

- a. Find an MOM estimator of  $\theta$ .
- b. Can we employ the second (raw) moment instead of the first one?

## Maximum Likelihood Estimator (MLE, CB Sec 7.2.2)

- Likelihood function:  $L : \Theta \rightarrow \mathbb{R}$  such that, given  $\mathbf{x}$  (a realization of  $\mathbf{X}$ ),

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; \mathbf{x}) = f_{\mathbf{X}}(\mathbf{x} \mid \boldsymbol{\theta}),$$

where  $f_{\mathbf{X}}$  is the joint pdf or pmf.

- For each  $\mathbf{x}$ , let  $\hat{\boldsymbol{\theta}}(\mathbf{x})$  be the maximizer of  $L(\boldsymbol{\theta}; \mathbf{x})$  (or log-likelihood  $\ell(\boldsymbol{\theta}; \mathbf{x}) = \ln L(\boldsymbol{\theta}; \mathbf{x})$ ) with respect to  $\boldsymbol{\theta}$  constrained in  $\Theta$ , i.e.,

$$\hat{\boldsymbol{\theta}}(\mathbf{x}) = \arg \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; \mathbf{x}) = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}; \mathbf{x}).$$

Then the statistic  $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{X})$  is the MLE for  $\boldsymbol{\theta} \in \Theta$ .

- Invariance property of MLE (CB Thm 7.2.10): As long as  $\hat{\boldsymbol{\theta}}$  is the MLE of  $\boldsymbol{\theta}$ , for ANY function  $g$ , the  $g(\hat{\boldsymbol{\theta}})$  is the MLE of  $g(\boldsymbol{\theta})$ .

- If  $\ell$  is differentiable, the score function  $\mathbf{S}$  is defined as its gradient

$$\mathbf{S}(\boldsymbol{\theta}) = \mathbf{S}(\boldsymbol{\theta}; \mathbf{x}) = \left[ \frac{\partial}{\partial \theta_1} \ell(\boldsymbol{\theta}; \mathbf{x}), \dots, \frac{\partial}{\partial \theta_p} \ell(\boldsymbol{\theta}; \mathbf{x}) \right]^\top.$$

- If  $\ell$  is twice differentiable, we have hessian of  $\ell(\boldsymbol{\theta}; \mathbf{x})$

$$\mathbf{H}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta}; \mathbf{x}) = \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\boldsymbol{\theta}; \mathbf{x}) \right]_{p \times p}.$$

- Maximizing differentiable (but non-monotonic)  $\ell(\boldsymbol{\theta})$  with  $\boldsymbol{\theta} \in \Theta$ 
  1. Find out stationary points, i.e., solutions to simultaneous equations  $\mathbf{S}(\boldsymbol{\theta}) = \mathbf{0}$
  2. Determine the global maximizer within  $\Theta$ : by comparing values of likelihood (or log-likelihood) evaluated at stationary points and boundary points of  $\Theta$

## Exercise Lec7.3

- Suppose  $X_1, \dots, X_n$  are iid as the following distributions. Find MLEs for corresponding parameters.
  - a.  $N(\mu, \sigma^2)$ ,  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$ .
  - b. Bernoulli( $\theta$ ):  $p(x \mid \theta) = \theta^x (1 - \theta)^{1-x} \mathbf{1}_{\{0,1\}}(x)$ ,  $\theta \in [0, 1/2]$ .
  - c. Two-parameter exponential:  $f(x \mid \alpha, \beta) = \beta^{-1} \exp\{-(x - \alpha)/\beta\} \mathbf{1}_{(\alpha, \infty)}(x)$ ,  $(\alpha, \beta) \in \mathbb{R} \times \mathbb{R}^+$ .

## Other examples of estimating equations

- Least-squares estimator
- Generalized estimating equations (GEE)
- M-estimator