MML revision lecture, Dec 2, 2022

Yingzhen Li, yingzhen.li@imperial.ac.uk

Clarifications on exam settings

The exam will be **closed-book** and it will last for 90mins.

- This format has been approved by DoC.
- The rule is strict: you are NOT allowed to bring a piece of A4 paper.
- Instead we will provide a formula sheet containing all the "complicated" formula/identities/theorems that you might need to solve the questions in the exam.
- The formula sheet's length is about 1.5-page of A4 paper. (We are not "hiding" any important detail!)

General suggestions on revision and on taking exams

- Firm up your basic Math skills, in particular:
  - Vector/matrix calculus, e.g., how to derive derivatives, automatic differentiation, etc.
  - Linear algebra, e.g., matrix product, positive definite matrices, eigendecomposition, how to write objectives in matrix forms, etc.
  - Probability & Statistics, e.g., sum & product rules, Bayes' rule, (conditional) independence, deriving & estimating mean and variance, an estimator's bias & variance, consistency properties, common distributions (e.g., Bernoulli, Gaussian), etc.
  - Machine learning methods, e.g., maximum likelihood, posterior inference, deriving optimal solution for a given objective, etc.
- Relevance of previous exams:
  - MML course content this year contains quite some differences compared to previous years. This means quite some of the previous years' exam questions are not relevant (if the concepts are not taught in this year).
  - DoC policy does NOT allow us to provide example answersheets for previous exams.
- On exam day:
  - The exam contains 2 questions (20 points each, 40 points in total). Each of the question contains 3 small sub-parts.
  - You are suggested to have a quick overview of all the questions at the beginning, to prioritise your time on different sub-parts.
  - Try NOT to spend $\geq$20mins on each sub-part of the questions.
  - We will look into derivations for marking, not just the final answer. However, if there are explicit statements like "give 3-sentence justification for your answer", writing much longer justifications WON'T give you extra marks.

Good luck!

consistency: want to estimate θ_0

construct estimator $f(\theta_1, \theta_2, \cdots \theta_N)$, $\theta_1 \cdots \theta_N \overset{iid}{\sim} P_\theta$

$\lim_{N\to\infty} f(\theta_1, \cdots \theta_N) \xrightarrow{P} \theta_0$ ($\xrightarrow{P}$ refers to weak LLN)

Revision session Question 1 (estimator properties)

Consider the following estimator of the variance of a scalar random variable X with distribution P_X:

$$\hat{\sigma}^2 := \frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x}_N)^2, \quad \bar{x}_N := \frac{1}{N}\sum_{n=1}^{N} x_n, \quad x_1, ..., x_n \sim P_X.$$

Show your derivations in answering the following questions:

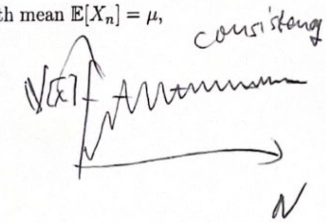
a. Is $\hat{\sigma}^2$ an *unbiased* estimator of the ground-truth variance $\mathbb{V}[X]$?

b. Is $\hat{\sigma}^2$ a *consistent* estimator of the ground-truth variance $\mathbb{V}[X]$?

You might find the weak LLN useful: for i.i.d. scalar random variables $X_1, ..., X_N$ with mean $\mathbb{E}[X_n] = \mu$, we have for any $\epsilon > 0$,

$$\lim_{N\to+\infty} P\left(|\frac{1}{N}\sum_{n=1}^{N} X_n - \mu| < \epsilon\right) = 1.$$

consistency



a).

$$\mathbb{E}[\hat{\sigma}^2] \overset{?}{=} \mathbb{V}[X]$$

$$\mathbb{E}[X_n] = \mu$$

$$\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}[\tfrac{1}{N}\textstyle\sum_{n=1}^N (X_n - \bar{X}_n)^2]$$

$$= \mathbb{E}[\tfrac{1}{N}\textstyle\sum_{n=1}^N (X_n - \mu + \mu - \bar{X}_n)^2]$$

$$= \mathbb{E}[\tfrac{1}{N}\textstyle\sum_{n=1}^N [(X_n-\mu)^2 + (\mu - \bar{X}_n)^2 + 2(X_n - \mu)(\mu - \bar{X}_n)]]$$

$$= \mathbb{E}[\tfrac{1}{N}\textstyle\sum_{n=1}^N (X_n-\mu)^2] + \mathbb{E}[(\mu-\bar{X}_n)^2] + 2\mathbb{E}[\tfrac{1}{N}\textstyle\sum_{n=1}^N (X_n-\mu)(\mu-\bar{X}_n)]$$

$$= \mathbb{V}[X] - \frac{1}{N^2}\underbrace{\mathbb{E}[(\textstyle\sum_{n=1}^N (X_n-\mu)^2)]}_{= \frac{1}{N}\mathbb{V}[X]}$$

$$= (1 - \tfrac{1}{N})\mathbb{V}[X] \neq \mathbb{V}[X]$$

biased estimator

b)

$$\hat{\sigma}^2 = \tfrac{1}{N}\textstyle\sum_{n=1}^N (X_n - \bar{X}_n)^2 = \tfrac{1}{N}\sum_{n=1}^N (X_n - \mu + \mu - \bar{X}_n)^2$$

$$= \tfrac{1}{N}\textstyle\sum_{n=1}^N (X_n-\mu)^2 + \tfrac{1}{N}\sum_{n=1}^N (\mu-\bar{X}_n)^2 + \tfrac{2}{N}\sum_{n=1}^N (X_n-\mu)(\mu-\bar{X}_n)$$

$$= \tfrac{1}{N}\textstyle\sum_{n=1}^N (X_n-\mu)^2 + (\mu-\bar{X}_n)^2 + 2(\underbrace{\tfrac{1}{N}\sum_{n=1}^N X_n}_{\bar{X}_n} - \mu)(\mu-\bar{X}_n)$$

$$= \tfrac{1}{N}\textstyle\sum_{n=1}^N (X_n-\mu)^2 - (\mu-\bar{X}_n)^2$$

by weak LLN: $\frac{1}{N}\sum_{n=1}^N (X_n-\mu)^2 \xrightarrow{P} \mathbb{V}[X]$, $\bar{X}_n \xrightarrow{P} \mu \Rightarrow (\mu-\bar{X}_n)^2 \xrightarrow{P} 0$

$$\Rightarrow \hat{\sigma}^2 \xrightarrow{P} \mathbb{V}[X]$$

Revision session Question 2 (sum rule, product rule and Bayes rule)

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?"

Show your derivations in answering the following question:

- What is the probability of wining a car if you switch your choice?

let's assume $P(C=i)$ be the prob. of the car behind door # i:

prior: $P(C=i) = \frac{1}{3}, \quad i=1,2,3.$

we want $P(C=2 \mid \underbrace{\text{"You pick door 1}}_{y=1}, \underbrace{\text{host pick door 3"}}_{h=3})$

Bayes' rule:

$$P(C=2 \mid y=1, h=3) = \frac{P(h=3 \mid y=1, C=2)\, P(C=2)}{P(h=3 \mid y=1)}$$

with $P(h=3 \mid y=1) = \sum_{i=1}^{3} P(h=3 \mid y=1, C=i)\, P(C=i)$

$$P(h=3 \mid y=1, C=i) = \begin{cases} \frac{1}{2}, & i=1 \\ 1, & i=2 \\ 0, & i=3 \end{cases}$$

$\Rightarrow$ Bayes' rule:

$$P(C=2 \mid y=1, h=3) = \frac{1 \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{2}{3}$$