References [...] available on the website
of this course

http://www.math.nagoya-u.ac.jp/~richard/SMLspring2019.html

Appendix 2

A few exercises from [CB]

Example 2.2.6 (minimizing distance)
Example 2.3.13 (Poisson approximation)
Section 2.4 (about exchange of differentiation and integration)
Exercise 2.17 about the median m

$$P(X \leq m) \geq \frac{1}{2} \quad \text{and} \quad P(X \geq m) \geq \frac{1}{2}$$

Exercise 2.18 (also about the median)
Exercise 2.28 (about skewness and kurtosis)
Thm 3.6.1 (related to Markov inequality)
Thm 2.1.8 (function of a r.v. X, make a summary of the result)

About common distributions:

make a 1 page (maximum) summary about one of them, with an example of application.
Use [CB], [FEHP], or internet (Wikipedia)

Inequalities

Appendix 3

see [CB], sec. 4.7

Lemma (Markov inequality) If $X \geq 0$, then
$$P(X > t) \leq \frac{E(X)}{t} \quad \forall t > 0.$$

$\Rightarrow$ if $\mu := E(X)$ and $\sigma := Var(X)$, $P\left(\frac{(X-\mu)^2}{\sigma^2} \geq t^2\right) \leq \frac{1}{t^2}$.
(Chebychev inequality)

Lemma (key of many inequalities)
If $a, b \geq 0$ and $p, q > 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$

then $\frac{1}{p} a^p + \frac{1}{q} b^q \geq ab$

this is a generalization of $\frac{1}{2} a^2 + \frac{1}{2} b^2 \geq ab$
$\Leftrightarrow$ $2ab \leq a^2 + b^2$ (obtained from $(a-b)^2 \geq 0$).

Thm (Hölder's inequality)
Let X, Y be 2 random variables, and $p, q > 1$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$|E(XY)| \leq E(|XY|) \leq \left(E(|X|^p)\right)^{1/p} \left(E(|Y|^q)\right)^{1/q}. \quad (*)$$

When $p = q = 2$, one gets the Cauchy-Schwarz inequality:

$$|E(XY)| \leq E(|XY|) \leq \left(E(|X|^2)\right)^{1/2} \left(E(|Y|^2)\right)^{1/2},$$

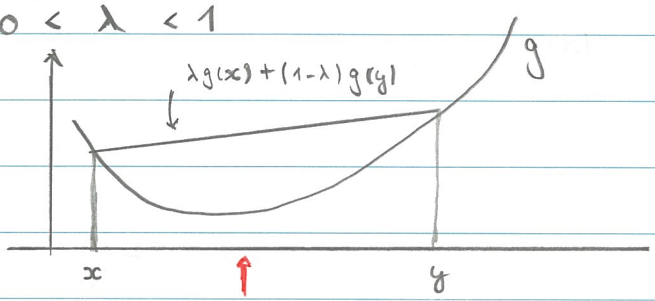
From $(*)$ one can also deduce the Lyapounov's inequality:
$$\left(E(|X|^r)\right)^{1/r} \leq \left(E(|X|^s)\right)^{1/s}$$

for $1 < r < s < \infty$.

<u>Thm</u> (Minkowski's inequality)
Let X, Y be random variables, and $1 \leq p < \infty$.
Then

$$\left(E(|X+Y|^p)\right)^{1/p} \leq \left(E(|X|^p)\right)^{1/p} + \left(E(|Y|^p)\right)^{1/p}.$$

Recall that a function $g : \mathbb{R} \to \mathbb{R}$ is <u>convex</u>
if $g(\underbrace{\lambda x + (1-\lambda) y}_{*}) \leq \lambda g(x) + (1-\lambda) g(y) \quad \forall x, y \in \mathbb{R}$
and $0 < \lambda < 1$



* any point between x and y

<u>Thm</u> (Jensen's inequality)
Let X be a random variable, and $g : \mathbb{R} \to \mathbb{R}$, Borel measurable and convex. Then

$$E(g(X)) \geq g(E(X)).$$

One directly deduces : $E(X^2) \geq (E(X))^2$
and $E(1/X) \geq 1/E(X)$.

If $a_1, \dots, a_N \in \mathbb{R}_+$, one sets
arithmetic mean : $a_A := \frac{1}{N}(a_1 + a_2 + \dots + a_N)$
geometric mean : $a_G := (a_1 a_2 \dots a_N)^{1/N}$
harmonic mean : $a_H = n\left(\frac{1}{a_1} + \frac{1}{a_2} + \dots + \frac{1}{a_N}\right)^{-1}$
and one has $a_H \leq a_G \leq a_A$.

Convergence

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. There exist several types of convergence of this sequence to a random variable X_∞.

__Def__: $(X_n)_{n\in\mathbb{N}}$ converges to X_∞ __in probability__ if $\forall \varepsilon > 0$, $P(|X_n - X_\infty| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0$.

__Def__: $(X_n)_{n\in\mathbb{N}}$ converges to X_∞ __almost surely__ if $P(\lim X_n = X_\infty) = 1$

$$\Leftrightarrow P(\{s \in S \mid \lim_{n\to\infty} X_n(s) = X_\infty(s)\}) = 1.$$

__Example__: Let $(X_i)_{i\in\mathbb{N}}$ be iid with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Set $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Weak law of large number: $\overline{X}_n \to \mu$ in probability

Strong law of large number: $\overline{X}_n \to \mu$ almost surely

discrete random variable taking only the value μ.

__Remarks__:

1) If $X_n \to X_\infty$ in probability and $g: \mathbb{R} \to \mathbb{R}$ is continuous, then $g(X_n) \to g(X_\infty)$ in probability.

2) Almost sure convergence $\Rightarrow$ convergence in probability.

<u>Def</u> : $(X_n)_{n \in \mathbb{N}}$ converges in distribution to X_∞ if $\lim_{n \to \infty} F_{X_n}(x) = F_{X_\infty}(x)$ at all $x \in \mathbb{R}$ with F_{X_∞} continuous at x.

cdf = cumulative distribution function

<u>Remark</u> : Convergence in probability $\Rightarrow$ convergence in distribution

$\Leftarrow$ if X_∞ is a constant random variable.

<u>Central limit theorem</u> :

Let $(X_i)_{i \in \mathbb{N}}$ be iid with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 \in (0, \infty)$. Set $\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$, and $Z_n := \sqrt{n}\, \frac{\bar{X}_n - \mu}{\sigma}$. Then

$$F_{Z_n}(x) \xrightarrow[n \to \infty]{} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy ,$$

that is Z_n converges in distribution to the random variable having the standard normal distribution.

<u>Lemma</u> (Slutsky's theorem)

Let $X_n \to X_\infty$ in distribution, and $Y_n \to a$ in probability, with a a constant r. v. Then

1) $X_n Y_n \to a X_\infty$ in distribution

2) $X_n + Y_n \to X_\infty + a$ in distribution.

A few exercises

- Study and summerized the hierarchical models, as presented in Section 4.4 and in its examples.
- Study Examples 4.2.2 and 4.2.4 about conditional pdf or pmf.
- Study the proof of thm 4.5.7 about the corelation ρ_{XY} of X and Y.
- Study the proof of Thm 5.2.6 about sample mean, variance and standard deviation.
- Study the proof of thm 5.3.1 about a family of narmal distribution.
- Study the student's t distribution as presented in Section 5.3.2.

Apendix 5

- Example 6.2.17 on ancillary statistic
  + Example 6.2.15 .
- Exercises 6.2 or 6.3 on sufficient statistics.
- Example 7.2.7 on log-likelihood.
- Example 7.2.9 on maximum with differentiation.
- Study example 7.2.13 on large variability.
- Exercise 7.22 with Example 7.2.16.

Likelihood ratio test

Appendix 7

Aim : testing hypothesis about mean of a normal random sample with unknown variance.

unknown

Let $X = (X_1, \dots, X_N)$ with $X_j \sim n(\mu, \sigma^2)$ be the random sample, and let $H_0 : \mu = \mu_0$ for a given $\mu_0 \in \mathbb{R}$. On the other hand, $\sigma^2 > 0$ is arbitrary. Set $\Theta := (\mu, \sigma^2)$.

Recall that $L(\Theta | \underline{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left(-\sum_{j=1}^{N} (x_j - \mu)^2 / 2\sigma^2\right)$.

The maximum of $\Theta \mapsto L(\Theta | \underline{x})$ is realized for $\mu = \bar{x}$ and $\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} (x_j - \bar{x})^2$, see Example 7.2.11 in [CB]. By the same argument one gets that if $\mu = \mu_0$, the maximum of $\sigma^2 \mapsto L(\mu_0, \sigma^2 | \underline{x})$ is realized for $\sigma^2 = \frac{1}{N} \sum_j (x_j - \mu_0)^2$. These results are obtained by looking at the critical points of $\Theta \mapsto L(\Theta | \underline{x})$.

Then, set $\sigma_0^2 := \frac{1}{N}\sum_j (x_j - \mu_0)^2$ and $\sigma^2 := \frac{1}{N}\sum_j (x_j - \bar{x})^2$.

$$\lambda(\underline{x}) := \frac{\sup_{\sigma^2>0} L(\mu_0, \sigma^2 | \underline{x})}{\sup_{\mu, \sigma^2} L(\mu, \sigma^2 | \underline{x})}$$

$$= \frac{(2\pi\sigma_0^2)^{-N/2} \exp\left(-\sum_j (x_j - \mu_0)^2 / 2\sigma_0^2\right)}{(2\pi\sigma^2)^{-N/2} \exp\left(-\sum_j (x_j - \bar{x})^2 / 2\sigma^2\right)}$$

$$= \left(\frac{\sigma^2}{\sigma_0^2}\right)^{N/2} \frac{e^{-N/2}}{e^{-N/2}}$$

$$= \left(\frac{\sigma^2}{\sigma_0^2}\right)^{N/2}$$

$$= \left(\frac{\sum_j (x_j - \bar{x})^2}{\sum_j (x_j - \mu_0)^2}\right)^{N/2} .$$

not known but fixed

Thus, one would reject H_0 if $\lambda(\underline{x}) \leq c$ for a given $c \in (0, 1)$. But let us go one more step. Observe that $\sum_j (x_j - \mu_0)^2 = \sum_j (x_j - \bar{x})^2 + N(\bar{x} - \mu_0)^2$.

Then, $\lambda(\underline{x}) \leq c \iff \left(\frac{\sum_j (x_j - \bar{x})^2}{\sum_j (x_j - \mu_0)^2}\right)^{N/2} \leq c$

$$\iff \frac{\sum_j (x_j - \bar{x})^2}{\sum_j (x_j - \bar{x})^2 + N(\bar{x} - \mu_0)^2} \leq c^{2/N}$$

$$\iff \frac{1}{1 + \frac{N(\bar{x} - \mu_0)^2}{\sum_j (x_j - \bar{x})^2}} \leq c^{2/N}$$

$$\Longleftrightarrow \quad \frac{N(\bar{x}-\mu_0)^2}{\sum_j (x_j-\bar{x})^2} \geq \frac{1}{c^{2/N}} - 1$$

sample mean

$$\Longleftrightarrow \quad \frac{(\sqrt{N}(\bar{x}-\mu_0))^2}{\underbrace{\frac{1}{N-1}\sum_j (x_j-\bar{x})^2}_{\text{sample variance } S^2}} \geq (N-1)(c^{-2/N}-1)$$

$$\Longleftrightarrow \quad \frac{|\bar{x}-\mu_0|}{S/\sqrt{N}} \geq \underbrace{\left((N-1)(c^{-2/N}-1)\right)^{1/2}}_{=: c'}$$

follow a student t-distribution with parameter $N-1$.

Then, H_0 would be rejected if $\frac{|\bar{x}-\mu_0|}{S/\sqrt{N}} \geq c'$,

Conclusion: The likelihood ratio test is equivalent to a t-test (in this setting).

Some useful exercises:

Example 7.2.11 + 7.2.12
Example 8.2.3
Thm 8.2.4 about sufficient statistics
Exercise 8.3
Example 8.2.9

Summary Chap. IV and V

Framework: $\underset{\sim}{X} = (X_1, \ldots, X_N)$ a random sample, with $X_j \sim f(\cdot \mid \theta)$, pmf or pdf, $f_{\underset{\sim}{X}}(\cdot \mid \theta)$ is the joint pmf or pdf.

IV : Point estimation :

Aim: infer an estimator (= expression) for θ based on $\underset{\sim}{X}$, or an estimate (= value) for θ based on $\underset{\sim}{x}$ (= values obtained for $\underset{\sim}{X}$ after the experiment).

Estimators or estimates are statistics $W(\underset{\sim}{X})$ or $W(\underset{\sim}{x})$.

Different methods (for finding $W(\underset{\sim}{X})$ or $W(\underset{\sim}{x})$):

- Method of moments
- Maximum likelihood estimator (MLE)
- Bayes estimator.

there are other methods ...

Evaluate the estimator $W = W(\underline{X})$:

- Mean square error (MSE)

⇝ Compute $E_\theta((W-\theta)^2) = Var_\theta(W) + Bias_\theta(W)^2$

$$Bias_\theta(W) = (E_\theta(W-\theta))^2$$

and look for W minimizing this expression.

⇝ Best estimator or best unbiased estimator.

when $Bias_\theta(W) = 0$

- Loss function optimality :

⇝ Replace $s \mapsto (s-\theta)^2$ by other functions.

$\underline{V}$: Hypothesis testing :

Aim : Define 2 complementary hypotheses H_0 and H_1, and test them.

Example : $\theta \in \Theta_0$

Different tests (for accepting or rejecting H_0)

- Likelihood ration tests (LRT)
- Baysian tests
- Union-intersection and intersection-union tests.

Evaluate the test (which tests $H_0 : \theta \in \Theta_0$) :

Decision : ← based on experiments and the test.

Truth :	Accept H_0	Reject H_0
H_0	OK	Type I error
H_1	Type II error	OK

We would like to minimize Type I and Type II errors.

⇝ Power function β (for estimating Type I error)

⇝ Minimize β on Θ_0 and maximize β on Θ_0^c.

⇝ Size α-test and level α-test.

⇝ Uniformly most powerful (UMP) level α-test

A more continuous evaluation : p-value.

Additional work :

- p-value and conditioning p 399 of [CB]
  + Example 8.3.30
- Section 8.3.5 on loss function optimality

About robustness (from page 481 – 482 of [CB])

... any statistical procedure should possess the following desirable features:

(1) It should have a reasonably good (optimal or nearly optimal) efficiency at the assumed model.
(2) It should be robust in the sense that small deviations from the model assumptions should impair the performance only slightly....
(3) Somewhat larger deviations from the model should not cause a catastrophe.

Additional exercises

- look at the proof of Thm 10.1.12
- review the content of Apendix 3
- study Example 10.2.1
- study the notion of breakdown value, of Definition 10.2.2
- study Examples 10.2.3 and 10.2.4 on median and mean.