Workshop 4

August 3, 2024

Unless said otherwise, the random variables involved in each problem are supposed to be defined on a common sample space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathbb{E}X$ means $\int_\Omega X \, \mathrm{d}\mathbb{P}$.

1 Laws of large numbers

Exercise 1.

1. Let X and Y be real valued random variables with finite mean and variance. Show that if X and Y are uncorrelated,[1] then for all $a, b \in \mathbb{R}$,

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var} X + b^2 \operatorname{Var} Y. \tag{1}$$

2. Let $(X_n)_{n\in\mathbb{N}}$ be uncorrelated random variables, not necessarily identically distributed, with finite mean and variance. Set $\mu_i = \mathbb{E}(X_i)$ and $\sigma_i^2 = \operatorname{Var}(X_i)$ for each $i \in \mathbb{N}$. Prove that for any $a > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i\right| \geq a\right) \leq \frac{\sum_{i=1}^n \sigma_i^2}{a^2}. \tag{2}$$

Deduce that if all X_n are identically distributed, then for all $\epsilon > 0$, $\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^n X_i - \mu_1\right| \geq \epsilon\right) \to 0$ as $n \to \infty$.

Exercise 2 (Weierstrass' approximation theorem, cf. Williams 7.4)**.** Let f be a continuous function on $[0, 1]$. Prove that for any $\epsilon > 0$, there exists a polynomial B such that

$$\sup_{x\in[0,1]} |B(x) - f(x)| < \epsilon. \tag{3}$$

(Hint: Consider the polynomials $B_n(x) = \sum_{k=1}^n f(\frac{k}{n})\binom{n}{k}x^k(1-x)^{n-k}$. Write them as an expectation over a random variable S, and then express S as a sum of random variables. Use the weak law.)

2 Cramér-Rao inequality

The purpose of this section is to prove an important result in statistics: the Cramér-Rao inequality (4).

Exercise 3. Let $(\rho_\theta)_{\theta\in\Theta}$ be collection of probability measures, indexed by an open set $\Theta \subset \mathbb{R}$. We suppose that each ρ_θ has a density $f(x;\theta)$ with respect to (w.r.t.) the Lebesgue measure, that f is differentiable w.r.t. θ and moreover that $A = \{x : f(x;\theta) = 0\}$ does not depend on θ. We denote by $\mathbb{E}_\theta$ the integration w.r.t. $\rho_\theta^{\otimes n}$ (for some n that depends on the context).

A collection $\mathbf{X} = (X_1, X_2, ..., X_n)$ of random variables is called *a random sample of size n for the population ρ_θ* if $X_1, ..., X_n$ are independent and each has law ρ_θ. We denote by $f_n(\mathbf{x};\theta)$ their joint density w.r.t. the n-dimensional Lebesgue measure. An *estimator* (e.g. of the unknown parameter θ) is a function $W(\mathbf{X}) = W(X_1, ..., X_n)$.

Suppose that

[1] This means that $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

1. $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta W(\mathbf{X}) = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta}(W(\mathbf{x})f_n(\mathbf{x};\theta))\,\mathrm{d}\mathbf{x}$, and
2. $\operatorname{Var}_\theta W(\mathbf{X}) < \infty$.

Prove that

1. $\mathbb{E}_\theta(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f_n(\mathbf{X};\theta)) = 0$,
2. $\mathbb{E}_\theta\left((\frac{\mathrm{d}}{\mathrm{d}\theta}\log f_n(\mathbf{X};\theta))^2\right) = n\mathbb{E}_\theta\left((\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X;\theta))^2\right)$, and that
3. $$\operatorname{Var}_\theta W(\mathbf{X}) \geq \frac{(\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}_\theta W(\mathbf{X}))^2}{n\mathbb{E}_\theta\left((\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X;\theta))^2\right)} \tag{4}$$

In particular, if W is an unbiased estimator of θ i.e. $\mathbb{E}_\theta W(\mathbf{X}) = \theta$, then

$$\operatorname{Var}_\theta W(\mathbf{X}) \geq \frac{1}{n\mathbb{E}_\theta\left((\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(X;\theta))^2\right)}. \tag{5}$$

3 Some additional problems

Exercise 4 (Differentiation under the integral sign, cf. Cohn Ex 2.4.6)**.** Let $(X, \mathcal{F}, \mu)$ be a measure space, I an open subinterval of $\mathbb{R}$, and $f : X \times I \to \mathbb{R}$ a function. Suppose that

1. for each $t \in I$, the function $x \mapsto f(x,t)$ is integrable,
2. for each $x \in X$, the function $t \mapsto f(x,t)$ is differentiable, and
3. there is an integrable function $g : X \to [0,\infty)$ such that for all $t, t_0 \in I$,
$$\left|\frac{f(x,t) - f(x,t_0)}{t - t_0}\right| \leq g(x).$$

Prove that
$$\frac{\mathrm{d}}{\mathrm{d}t}\int_X f(x,t)\,\mathrm{d}\mu = \int_X \frac{\partial}{\partial t} f(x,t)\,\mathrm{d}\mu. \tag{6}$$

Exercise 5.

1. Let Z be a random variable. Choose X and Y appropriately in Hölder's inequality (see Section 6.13 in Williams) to show that
$$g(t) := \log \mathbb{E}|Z^t| \tag{7}$$
is a convex function on the interval[2] of values of $t \geq 1$ such that $\mathbb{E}|Z^t| < \infty$. *Pay attention to the conditions under which Hölder's inequality is valid.*

2. Deduce from the previous result that if $r_1, ..., r_n$ are real numbers such that $\mathbb{E}|X^{r_i}| < \infty$ for all $i \in \{1, ..., n\}$ and $s = n^{-1}\sum_{i=1}^n r_i$, then
$$\left(\prod_{i=1}^n \mathbb{E}|X^{r_i}|\right)^{1/n} \geq \mathbb{E}|X^s|. \tag{8}$$
(Hint: Use induction in n.)

[2] It is an interval because of the monotonicity of p-norms: if $\mathbb{E}|Z^t| < \infty$, then $\mathbb{E}|Z^r| < \infty$ for every r in $[1,t]$.