PRINCIPLES OF STATISTICS Part II
Example Sheet 1 (of 4) RDS/Michaelmas 2024

1. Find the Fisher information for $\theta \in (0,1)$ in the model $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Bern}(\theta)$ where $\theta \in [0,1]$. Show that the MLE is unbiased and achieves the Cramér–Rao lower bound.

2. Find the Fisher Information matrix $I(\beta, \sigma^2)$ in the normal linear model $Y = X\beta + \varepsilon$ where $X \in \mathbb{R}^{n\times p}$ is a deterministic matrix of predictors with full column rank, $\beta \in \mathbb{R}^p$ and $\sigma^2 > 0$. Show that the MLE $\hat{\beta}$ for β is unbiased and achieves the Cramér–Rao lower bound.

3. Suppose we wish to estimate the mean μ of a random variable $X \sim N(\mu, 1)$ and we can either do this using data formed of (i) n i.i.d. copies $X_1, \ldots, X_n$ of X; or (ii) $N > n$ i.i.d. observations $W_1, \ldots, W_N$ each having distribution equal to that of $\text{sgn}(X)$. Suppose that it is expected that $|\mu| \leq M$. By considering the Fisher information, explain why we might choose option (ii) over option (i) when

$$N > \frac{\Phi(M)\Phi(-M)}{\phi^2(M)} n,$$

where ϕ and Φ are the standard normal density and distribution functions respectively.

4. Prove that an unbiased estimator $\hat{\theta}(X) \in \mathbb{R}$ achieves the Cramér–Rao lower bound if and only if (almost surely)

$$\hat{\theta} = \theta + I(\theta)^{-1} S(\theta).$$

[*Hint: Recall that for random variables* U, V *with* $\mathbb{E}(U^2)$, $\mathbb{E}(V^2) < \infty$, *we have* $(\mathbb{E}|UV|)^2 \leq \mathbb{E}(U^2)\mathbb{E}(V^2)$, *with equality if and only if* $U = cV$ *almost surely, for some* $c \in \mathbb{R}$.]

5. Suppose we have pairs

$$(Y_1, X_1), \ldots, (Y_n, X_n) \stackrel{\text{i.i.d.}}{\sim} N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \underbrace{\begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}}_{=:\Sigma} \right),$$

where Σ is positive definite, and we are interested in estimating μ_1.

(a) Consider first the setting where (only) Σ is known. Find the MLE of μ_1 in this case and show that it is unbiased and achieves the Cramér–Rao lower bound v_1 for estimating μ_1.

(b) Now suppose that both Σ and μ_2 are known. Find the Cramér–Rao lower bound v_2 in this case and show that $v_2 \leq v_1$ with equality if and only if $\rho = 0$. Show that the MLE is given by

$$\bar{Y} - \frac{\rho}{\sigma_2^2}(\bar{X} - \mu_2)$$

and that it is unbiased and achieves the bound v_2.

[*Hint: It may help to use the fact that for* $\nabla_x(x^\top Ax) = (A + A^\top)x$ *for a matrix* $A \in \mathbb{R}^{d\times d}$ *and vector* $x \in \mathbb{R}^d$.]

6. Suppose we have data i.i.d. copies of $X_1, \ldots, X_n$ of a random variable $X \in \mathbb{R}$ assumed to follow the model $X = \mu + \varepsilon$, where $\varepsilon \sim t_\nu$; we wish to estimate the unknown parameter $\mu \in \mathbb{R}$ and the degrees of freedom $\nu > 2$ is known to us. Show that

$$\frac{\mathrm{Var}_\mu(\bar{X})}{I_n(\mu)} = \frac{\nu(\nu+3)}{(\nu-2)(\nu+1)}.$$

[*Hint: The following facts may be of use. If $A \sim \chi^2_k$, then $\mathbb{E}(A) = (k-2)^{-1}$ provided $k > 2$. Now if $B \sim \chi^2_l$ and A and B are independent, then*

$$\frac{A}{A+B} \sim \textit{Beta}(k/2, l/2),$$

a Beta distribution with parameters $k/2$ and $l/2$, provided $k, l > 0$. If $Z \sim \textit{Beta}(a, b)$ then

$$\mathbb{E}(Z) = \frac{a}{a+b} \qquad \mathrm{Var}(Z) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Also the t_ν distribution has density proportional to

$$f(x) = (1 + x^2/\nu)^{-(\nu+1)/2}.$$

]

7. (a) Suppose that random vectors $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$. Show that $(X_n, Y_n) \xrightarrow{p} (X, Y)$.

   (b) Give an example to show that we can have $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$, but (X_n, Y_n) does not converge in distribution.

   (c) Show that if random vectors $X_n \xrightarrow{d} c$ for some deterministic constant $c \in \mathbb{R}^d$, then $X_n \xrightarrow{p} c$.

   (d) Show that for a sequence of real-valued random variables $(X_n)_{n \in \mathbb{N}}$, we have $X_n \xrightarrow{p} 0$ if and only if $\mathbb{E}(\min(|X_n|, M)) \to 0$ for some $M \geq 0$. Give an example to show that we can have $X_n \xrightarrow{p} 0$ but $\mathbb{E}|X_n| \to \infty$.

8. Show the following, where $(X_n)_{n \in \mathbb{N}}$ is a sequence of random vectors taking values in $\mathbb{R}^d$.

   (a) If for real-valued random vectors, we have $X_n \xrightarrow{d} X$ and Ω_n is a sequence of events with $\mathbb{P}(\Omega_n) \to 1$, then $X_n \mathbb{1}_{\Omega_n} \xrightarrow{d} X$.

   (b) If $r_n(X_n - \theta)$ converges in distribution for some $\theta \in \mathbb{R}^d$ and $r_n \to \infty$, then $X_n \xrightarrow{p} \theta$.

9. Consider the setting of Question 5 but where we do not make assumptions on the distribution of each of the i.i.d. pairs (Y_i, X_i) beyond the existence of their covariance matrix. Show that the sample covariance

$$\hat{\rho} := \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X})$$

satisfies $\hat{\rho} \xrightarrow{p} \rho$.

10. We continue with the setting in Question 9, but with our target of interest being $\mu_1 = \mathbb{E}(X_1)$ as in Question 5.

(i) Write down an estimator $\hat{\mu}_1^{(1)}$ that satisfies $\sqrt{n}(\hat{\mu}_1^{(1)} - \mu_1) \xrightarrow{d} N(0, \sigma_1^2)$.

(ii) Now suppose that Σ and μ_2 are known. Find an estimator satisfying

$$\sqrt{n}(\hat{\mu}_1^{(2)} - \mu_1) \xrightarrow{d} N\left(0, \sigma_1^2 - \frac{\rho^2}{\sigma_2^2}\right).$$

(iii) Now suppose that only μ_2 is known. Find an estimator $\hat{\mu}_1^{(3)}$ satisfying the same distributional convergence result as in part (ii).

(iv) Finally, consider the setting where neither μ_2 nor Σ are known exactly, but we have an additional N i.i.d. copies of X_1. Find an estimator $\hat{\mu}_1^{(4)}$ that in the asymptotic regime where $n = o(N)$, satisfies the same distributional convergence result as in part (ii).

11. In this question, we consider a *random design* regression setting where we have available data i.i.d. $(Y_1, X_1), \dots, (Y_n, X_n) \in \mathbb{R} \times \mathbb{R}^p$, and study the asymptotic behaviour of the OLS estimator $\hat{\beta} := (X^\top X)^{-1} X^\top Y$, where $X \in \mathbb{R}^{n \times p}$ is the matrix with ith row $X_i \in \mathbb{R}^p$ and $Y := (Y_1, \dots, Y_n)^\top$; writing $\Omega_n := \{\frac{1}{n} X^\top X \text{ is invertible}\}$, on the event Ω_n^c we (arbitrarily) define $\hat{\beta} = 0$.

(i) Let $\Sigma := \mathbb{E}(X_1 X_1^\top)$ be finite and suppose that Σ is invertible. Show that $\frac{1}{n} X^\top X \xrightarrow{p} \Sigma$ and explain why $\mathbb{P}(\Omega_n) \to 1$.

(ii) Now suppose $\mathbb{E}(Y_1 \,|\, X_1) = \beta^\top X_1$ and let $\varepsilon_i := Y_i - \beta^\top X_i$ so $\mathbb{E}(\varepsilon_i \,|\, X_i) = 0$. Let $\Gamma := \mathrm{Cov}(\varepsilon_1 X_1) \in \mathbb{R}^{p \times p}$ be finite. Show that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N_p(0, \Sigma^{-1} \Gamma \Sigma^{-1}).$$

What happens when ε_i and X_i are in fact independent?

(iii) We now make no assumption on the conditional expectation of Y_1 given X_1, but define $\rho = \mathbb{E}(X_1 Y_1) \in \mathbb{R}^p$, $\beta := \Sigma^{-1} \rho$ (and retain the definition of ε_i and the assumption on Γ from above). Show that with our new β, we have the same distributional result as above.

(iv) Finally, writing $X_i = (W_i, Z_i) \in \mathbb{R} \times \mathbb{R}^{p-1}$, in the setting of the previous part, suppose we have a *partially linear model* where

$$\mathbb{E}(Y_i \,|\, W_i, Z_i) = W_i \theta + f(Z_i)$$

and $\mathbb{E}(f(Z_i)^2) < \infty$. Suppose additionally that $\mathbb{E}(W_i \,|\, Z_i) = Z_i^\top \gamma$. Show that writing $\hat{\theta}$ for the first component of $\hat{\beta}$, we have

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, (\Sigma^{-1} \Gamma \Sigma^{-1})_{11}).$$

[*Hint: Aim to compute relevant parts of* Σ *and* ρ *and use the matrix identity that for* $M \in \mathbb{R}^{p \times p}$, $b \in \mathbb{R}^p$ *and* $a \in \mathbb{R}$,

$$\begin{pmatrix} a & b^\top \\ b & M \end{pmatrix}^{-1} = \begin{pmatrix} s^{-1} & -s^{-1} b^\top M^{-1} \\ -s^{-1} M^{-1} b & M^{-1} + s^{-1} M^{-1} b b^\top M^{-1} \end{pmatrix},$$

where $s := a - b^\top M^{-1} b > 0$ *provided the matrix on the left is indeed invertible.*]