Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4295 Statistical inference
Fall 2023

Exercise set 5

Read the questions carefully and make your own assumptions if needed.

1 Casella-Berger

⑦ 2, 7, 9

2 Empirical distribution and empirical statistics

Let $x_1, \dots, x_n$ be points in a sample space X.

a) What are the properties of the family $\mathcal{F}$ of events in X? Is $\{x_1\}$ an event?

b) Fix $x \in \mathrm{X}$ and define the Dirac measure δ_x by

$$\mathcal{F} \ni A \mapsto \delta_x(A) = A(x) = 1_A(x) = \chi_A(x) = (x \in A) \tag{1}$$

Explain what this means. Show that δ_x is a probability measure. Let $\Omega \ni \omega \mapsto Y_1(\omega) = x_1$. Prove that Y_1 is a random point and determine P_{Y_1}.

c) The empirical distribution P_n is defined by

$$\mathrm{P}_n = (\delta_{x_1} + \cdots + \delta_{x_n})/n \tag{2}$$

Is P_n a probability measure? Let $M \sim \mathrm{U}\{1, \dots, n\}$. Determine the distribution of $Y = x_M$ and the distribution of $Y \mid M = 1$. Is $(Y = y)$ an event?

d) Assume $Y_n \sim \mathrm{P}_n$. Explain what this means. Prove

$$\mathrm{E}\,\phi(Y_n) = \mathrm{E}_n\,\phi = \int \phi(x)\,\mathrm{P}_n(dx) \tag{3}$$

Is P_n and $\mathrm{E}_n\,\phi$ statistics? What assumptions are needed? Determine $\mathrm{E}_n\,\phi$ using $\mathrm{E}(\phi(Y)) = \mathrm{E}(\mathrm{E}(\phi(Y) \mid M))$. Is $(Y_n = y)$ an event?

e) Find formulas for $\mathrm{E}\,Y_n$ and $\mathrm{Std}\,Y_n$ for the case where $\mathrm{X} = \mathbb{R}$. Assume that all intervals $(-\infty, x]$ are events. Is $\{x_1\}$ an event? Are the level sets of Y_n events?

f) Let $(W \mid M = m) \sim \mathrm{N}(x_m, \sigma_m^2)$. Determine the density g_n of W. Determine the CDFs F_n and G_n of Y and W. What happens when $\sigma_m \downarrow 0$? What about $\mathrm{E}(\phi(W))$ given that ϕ is continuous? Is $\mathrm{X} = \mathbb{R}^k$ meaningful here?

g) Let $(W \mid M = m) \sim (x_m + \mathrm{U}(-\epsilon_m, \epsilon_m))$. Determine the density g_n of W. Determine the CDFs F_n and G_n of Y and W. What happens when $\epsilon_m \downarrow 0$? What about $\mathrm{E}(\phi(W))$ given that ϕ is continuous? Is $\mathrm{X} = \mathbb{R}^k$ meaningful here?

3 Completeness and sufficiency in non-parametric statistics

Let $x_1, \ldots, x_n$ be a random sample from a distribution on a sample space X.

a) Is the empirical distribution P_n a complete minimal sufficient statistic?

b) The empirical statistic $\mathrm{E}_n\,\phi$ is a uniformly minimal variance unbiased (UMVU) estimate of $\mathrm{E}\,\phi(X_1)$. Why?

c) Is the loss $l = (t-\tau)^2$ convex as a function of t. What about $l = (t-\tau)^p$? Is $\mathrm{E}_n\,\phi$ is a uniformly minimal risk unbiased (UMRU) estimate of $\mathrm{E}\,\phi(X_1)$ given a convex loss? Hint: Jensen's inequality.

d) Assume $\mathrm{X} = \mathbb{R}$ and that every interval $(-\infty, x]$ is an event in X. Prove that the sample mean and variance are UMVU estimates of $\mathrm{E}\,X_1$ and $\mathrm{Var}\,X_1$.

e) Assume $\mathrm{X} = \mathbb{R}$ and that every interval $(-\infty, x]$ is an event in X. Assume additionally that $X_1 \sim f(x)\,dx$ where f is unknown. Are the sample mean and variance UMVU estimates? What additional assumptions are needed? What if the additional assumption is replaced by assuming that f is a density with respect to a measure $\mu = w_1\delta_{y_1} + w_2\delta_{y_2} + \cdots$.

f) Reconsider task **e)** with the added assumption that $X_1 = \mu + Z_1$ where both μ and the distribution of $Z_1 \sim -Z_1$ is unknown. Prove that the statistical model is invariant with respect to shift transformations. Are the sample mean and variance equivariant?

g) Reconsider task **f)** with the added assumption that $Z_1 \sim \mathrm{U}(-\frac{1}{2}, \frac{1}{2})$.

4 Estimation in parametric families

Let $x_1, \ldots, x_n$ be a random sample from $f(x) = (x > 0)\beta^{-1}e^{-x/\beta}$ with unknown $\beta > 0$ and a prior $\pi(\beta) = \beta^{-1}$. Let P_n be the empirical distribution.

a) Determine the moment- and Bayes-estimates of β.

b) Explain that $\mathrm{E}\,X_1$, $\mathrm{Var}\,X_1$, $f(x)$, f, $F(x) = \int_0^x f(x)\,dx$, F, and $F^{-1}(q)$ are parameters. Interpretation? What are the corresponding parameter spaces and range of possible parameter values? Find the corresponding maximum likelihood estimates (MLEs).

c) Show that the statistical model is an exponential family. Determine a natural parameter, a natural statistic T, and its distribution.

d) Let $Y \sim \mathrm{P}_n$. Is Y a statistic? Explain that P_n, $\mathrm{E}\,Y$, $\mathrm{Var}\,Y$, $F_Y(x)$, F_Y, and $F_Y^{-1}(q)$ are estimates. Interpretation? Are they unbiased or equivariant?

e) Discuss properties of all of the above estimators. Sufficiency principle obeyed? Are they optimal, or can they be used to obtain optimal estimators?

5 Casella-Berger

(7) 18, 24, 34, 35, 45