Vrije Universiteit Amsterdam
Department of Mathematics

Midterm Mathematical Statistics (XB_0049)
October 23, 2023, 15.30–17.45

You are not allowed to use calculators, phones, laptops or other tools. Clearly put your name and student number on all sheets that you submit. Your answer may contain quantiles from a distribution. Expressions for densities, expectation, variance and quantiles for some distributions are given on page 3 of the exam. You can use these expressions if needed for answering the questions below.

Unless stated differently, always add an explanation to your answer.

1. Suppose $X_1, \ldots, X_n$ are independent and identically distributed. Let $\sigma^2 = \mathrm{Var}(X_1)$.

   (a) [1 PT]. Show that
   $$\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \overline{X_n^2} - \left(\bar{X}_n\right)^2.$$
   Here $\overline{X_n^2} = n^{-1}\sum_{i=1}^n X_i^2$.

   (b) [2 PT]. Show that
   $$S^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$
   is unbiased for estimating σ^2.

2. Consider a population with three kinds of individuals, say 1, 2 and 3 occurring in the following proportions: $p_1 = \theta^2$, $p_2 = 2\theta(1-\theta)$, $p_3 = (1-\theta)^2$ $(0 < \theta < 1)$. In a random sample of size n chosen from the population, n_1 individuals have been labeled 1, n_2 labelled 2, n_3 labelled 3, respectively $(n_1 + n_2 + n_3 = n)$.

   (a) [2 PT]. Write down the likelihood function for θ.

   (b) [2 PT]. Derive an expression for the maximum likelihood estimator for θ.

   (c) [3 PT]. Does the distribution of the data belong to the exponential family? If so, what are the natural parameter and the sufficient statistic?

3. [2 PT]. Suppose X has a discrete distribution parametrised by θ. Suppose $T = \psi(X)$ is a sufficient statistic for θ. Prove that there exist functions g and h such that
   $$\mathrm{P}_\theta(X = x) = g(\psi(x); \theta)h(x).$$

4. Suppose $X_1, \ldots, X_n$ are independent and identically distributed. Each $X_i \sim Unif\,(0, \theta)$ (uniform distribution on $[0, \theta]$).

   (a) [3 PT]. Derive the maximum likelihood estimator (mle) $\hat{\Theta}$ and show that

$$\mathrm{P}(\hat{\Theta} \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ (x/\theta)^n & \text{if } 0 \leq x \leq \theta \\ 1 & \text{if } x > 0 \end{cases}.$$

   (b) [3 PT]. It can be shown that

$$\operatorname{var} \hat{\Theta} = \frac{n}{(n+1)^2(n+2)}\theta^2.$$

   Use this result to compute the Mean Squared Error of $\hat{\Theta}$ for estimating θ.

   (c) [2 PT]. Another estimator for θ is given by $T := 2\bar{X}_n$. Compute the Mean Squared Error of T for estimating θ.

   (d) [1 PT]. Which of the two estimators do you prefer? Explain.

5. Suppose $X_1, \ldots, X_n$ are independent and identically distributed with the Exponential distribution with parameter θ.

   (a) [HARD, 1 PT]. Explain why $2\theta \sum_{i=1}^n X_i \sim \chi^2_{2n}$.
   Hint: for an integer k, the density of the χ^2_k distribution is given by

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}, \qquad x \geq 0.$$

   For integers m, $\Gamma(m) = (m-1)!$. First show that $2\theta X_1 \sim \chi^2_2$.

   (b) [3 PT]. Use this result to derive an expression for a 90% confidence interval for θ.

Distributions and notation

- If $X \sim Unif\,(0,\theta)$, then $\mathbb{E}X = \theta/2$ and $\mathrm{Var}(X) = \theta^2/12$.
- If $X \sim N(\mu,\sigma^2)$, then its density is given by $p(x) = (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{(x-\mu)^2)}{2\sigma^2}\right)$ and $\Phi(x) := \mathbb{P}(X \le x)$. The (lower) quantile ξ_α is defined by $\Phi(\xi_\alpha) = \alpha$, for $\alpha \in (0,1)$.
- If $X \sim \chi^2_k$, then $\chi^2_{k,\alpha}$ is defined by $\mathbb{P}(X \le \chi^2_{k,\alpha}) = \alpha$.
- If $X \sim t_k$, then $t_{k,\alpha}$ is defined by $\mathbb{P}(X \le t_{k,\alpha}) = \alpha$.
- If $X \sim Ga(\alpha,\beta)$, then its density is given by $p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\mathbf{1}_{(0,\infty)}(x)$. $\mathbb{E}X = \alpha/\beta$, $\mathrm{Var}(X) = \alpha/\beta^2$. The special case of $\alpha = 1$ corresponds to the exponential distribution with parameter β.
- If $X \sim Bin(n,p)$, then $\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, for $x \in \{0,1,\ldots,n\}$. $\mathbb{E}X = np$, $\mathrm{Var}(X) = np(1-p)$.
- If $X \sim Beta(\alpha,\beta)$, then its density is given by $p(x) = B(\alpha,\beta)^{-1} x^{\alpha-1}(1-x)^{\beta-1}\mathbf{1}_{[0,1]}(x)$. $\mathbb{E}X = \alpha/(\alpha+\beta)$, $\mathrm{Var}(X) = \alpha\beta/\left((\alpha+\beta)^2(\alpha+\beta+1)\right)$