Exercise 4

Bent Nielsen
University of Oxford

Slightly revised version of material written by Anders Kock 2022

30 October 2023

In solving the problems, you may freely use and refer to results in the slides. Please hand in for marking problems 1–2. Problem 3 is optional and only to be covered in the classes in case you have time.

1. (Binomial distribution: Sufficiency, completeness and UMVU estimation) Fix $n \in \mathbb{N}$ and consider the measurable space $(\{0, \ldots, n\}, \mathcal{P}(\{0, \ldots, n\}))$ on which, for known $n \in \mathbb{N}$, the binomial distribution with density

$$p_\pi(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x} \qquad \pi \in (0,1),\ x \in \{0, \ldots, n\}$$

is defined. Here x interpreted as the number of successes and π as the success probability in each of the n draws.

   (a) Show that the binomial distribution forms an exponential family. What is the sufficient statistic?

   (b) Show that the sufficient statistic from (a) is also complete.

   (c) Show that $\frac{x}{n}$ is a UMVUE for π.

   (d) Is $\frac{x}{n}$ the unique UMVUE?

   (e) Is it true in general that the existence of a UMVUE for π guarantees the existence for a UMVUE for $g(\pi)$ for any function $g : (0,1) \to \mathbb{R}$?

2. (Estimating a probability or a quantile) Assume that you have observed n observation $X_1, \dots, X_n$ from $\mathsf{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ and $\sigma^2 \in (0, \infty)$. (That is, $X_1, \dots, X_n$ are independent. Therefore they have a joint normal distribution.) Write $\theta = (\mu, \sigma^2)$.

   (a) Let $p \in (0,1)$ be given and assume that you wish to estimate the corresponding quantile $z_p \in \mathbb{R}$, i.e. z_p satisfies $\mathbb{P}_\theta(X_1 \geq z_p) = p$.

      i. Show that $z_p = \mu + \sigma\Phi^{-1}(1-p)$, where Φ is the cdf of the standard normal distribution.

      ii. Show that $\bar{x}_n + sK_{n,1}\Phi^{-1}(1-p)$ is a UMVUE for z_p, where $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $s^2 = \sum_{i=1}^n (X_i - \bar{x}_n)^2 = \sum_{i=1}^n X_i^2 - n\bar{x}_n^2$ and $K_{n,1} = \frac{\Gamma([n-1]/2)}{2^{1/2}\Gamma(n/2)}$.
      Hint: Recall (and use freely) from the lectures that $\mathbb{E}_\theta sK_{n,1} = \sigma$.

   (b) Assume that $\sigma^2 = 1$ is known. Let $u \in \mathbb{R}$ be given and assume that you wish to estimate the probability of a drawing a smaller value than u, i.e. $p_u := \mathbb{P}_\mu(X_1 \leq u)$.

      i. Show that $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is a complete sufficient statistic for μ.

      ii. Show that $\mathbb{1}_{\{X_1 \leq u\}}$ is an unbiased estimator for p_u.

      iii. Show that $\bar{x}_n$ is independent of $X_1 - \bar{x}_n$.
      Hint: Use that $\bar{x}_n$ and $X_1 - \bar{x}_n$ are jointly normally distributed such that it suffices to show that they have a covariance of zero.

      iv. Argue that $\delta_n := \mathbb{E}_\mu(\mathbb{1}_{\{X_1 \leq u\}} | \bar{x}_n)$ is a UMVUE for p_u.

      v. Show that $\delta_n = \mathbb{E}_\mu(\mathbb{1}_{\{X_1 \leq u\}} | \bar{x}_n) = \Phi\left(\sqrt{\frac{n}{n-1}}(u - \bar{x}_n)\right)$.
      Hint: Use that $\mathbb{1}_{\{X_1 \leq u\}} = \mathbb{1}_{\{X_1 - \bar{x}_n \leq u - \bar{x}_n\}}$ and that $\bar{x}_n$ and $X_1 - \bar{x}_n$ are independent such that the conditional expectation reduces to an unconditional one.

      vi. Is $\delta_n = \Phi\left(\sqrt{\frac{n}{n-1}}(u - \bar{x}_n)\right)$ the unique UMVUE for p_u?

3. (Uniqueness of uniform minimum risk unbiased estimators and loss functions) Consider the experiment $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\otimes_{i=1}^n \mathsf{N}(\mu, 1) : \mu \in \mathbb{R}\}\right)$ for $n \in \mathbb{N}$ known and with μ to be estimated. That is, we wish to estimate μ based on n independent observations. Consider two estimators $\delta_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2$ defined via

$$\delta_1(x) = x_1 \qquad \text{and} \qquad \delta_2(x) = \frac{1}{n}\sum_{i=1}^n x_i,$$

where we write $x = (x_1, \ldots, x_n)$.

(a) Are δ_1 and δ_2 unbiased for μ?

(b) Consider the quadratic loss function $L(u, v) = (u - v)^2$. Which of δ_1 and δ_2 has the lowest estimation risk? Is the risk these estimators equal to their variance for this loss function?

(c) Consider the loss function $L_{\text{exp}}(u, v) = e^{(u-v)^4}$ [note that we could subtract 1 if we insist on the loss being zero when $u = v$]. What is the risk of δ_2? and of δ_1? Does δ_2 still have lower risk than δ_1? Does there exist any unbiased estimator with finite risk at any $\mu \in \mathbb{R}$?

(d) Does there exist a biased estimator with finite risk at all $\mu \in \mathbb{R}$?