TD3 - Asymptotic properties of estimators

Exercise 1  We are given two asymptotically normal estimators of the same parameter \( \theta \in \Theta \subseteq \mathbb{R} \), denoted by \( \hat{\theta}_n^1 \) and \( \hat{\theta}_n^2 \), which satisfy, for all \( \theta \in \Theta \),

\[
\frac{\sqrt{n}}{\sqrt{\hat{\theta}_n^1 - \theta}} \sim \mathcal{N}(0, \sigma_1^2) \\
\frac{\sqrt{n}}{\sqrt{\hat{\theta}_n^2 - \theta}} \sim \mathcal{N}(0, \sigma_2^2)
\]

where \( \sigma_1^2 < \sigma_2^2 \).

1. Using each estimator, derive an asymptotic confidence interval on \( \theta \) of level 0.95.

2. For any \( \epsilon > 0 \), how many samples \( n \) are needed to get a precision \( \epsilon \) on \( \theta \) with the estimator \( \hat{\theta}_n^2 \), with probability \( \approx 0.95 \)?

3. What fraction of this sample size is needed if we choose to use instead the estimator \( \hat{\theta}_n^1 \)?

Exercise 2  We are given an i.i.d sample \( X_1, \ldots, X_n \sim f_\theta \) where \( \theta > 0 \) is an unknown parameter and

\[
f_\theta(x) = \frac{1}{2} \theta^3 x^2 \exp(-\theta x); \quad x > 0.
\]

Denote \( \mu, \sigma^2 \) to be the mean and variance of \( f_\theta \) respectively.

1. Show that \( \mu = 3/\theta \) and \( \sigma^2 = 3/\theta^2 \).

2. Find the Cramer Rao bound for \( \mu \) and show that the estimator \( \hat{\mu}_n = \frac{1}{n} \sum_i X_i \) achieves this bound.

3. Now we consider the estimator \( \hat{\sigma}_n^2 = \frac{\hat{\sigma}_n^2}{3} \) of \( \sigma^2 \).
   
   (a) Compute the bias of \( \hat{\sigma}_n^2 \) and conclude that it is asymptotically unbiased.
   
   (b) Show that \( \hat{\sigma}_n^2 \) is a consistent estimator of \( \sigma^2 \).
   
   (c) Show that \( \hat{\sigma}_n^2 \) is asymptotically efficient.

Exercise 3  In a clinical trials involving two treatments, we observe the outcome of treatment 1 (a placebo) on a pool of \( n_1 \) patients. For \( i \in \{1, \ldots, n_1\} \), we record \( X_i = 1 \) if the treatment is a success for patient \( i \), \( X_i = 0 \) if it is a failure. The outcome of treatment 2 (the new drug) is observed on a pool of \( n_2 \) patients, with \( Y_j \in \{0, 1\} \) indicating success of failure for patient \( j \in \{1, \ldots, n_2\} \).

We assume that all the \( X_i \) and \( Y_j \) are independent and that \( X_i \sim B(p_1) \) and \( Y_j \sim B(p_2) \) where \( p_1 \) and \( p_2 \) are the probability of efficacy of treatment 1 and 2, respectively. We are interested in estimating the treatment effect \( \phi := p_2 - p_1 \).

1. Find the MLE estimator \( \hat{p} \) of the parameter \( p = (p_1, p_2) \) in \( [0, 1]^2 \).
2. Deduce a consistent estimator of \( \phi \), denoted by \( \hat{\phi} \). Compute its variance.

3. Compute the Fisher information of the observation, \( I^Z((p_1, p_2)) \), where \( Z = (X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}) \).

In the rest of the exercise, we assume that \( n_1 = n_2 = n \).

4. Give the limiting distribution of \( \hat{p} \) when \( n \) goes to infinity.

5. Using the multi-variate Delta method, find the asymptotic distribution of \( \hat{\phi} \).

6. Using this estimator, we want to construct a test of the hypothesis

\[
H_0 : (\phi = 0) \quad \text{against} \quad H_1 : (\phi > 0).
\]

We propose to use the test statistic

\[
T_n = \frac{\hat{p}_{2,n} - \hat{p}_{1,n}}{\sqrt{\frac{\hat{p}_{1,n}(1-i_{n,1})}{n} + \frac{\hat{p}_{2,n}(1-i_{n,2})}{n}}}
\]

(a) Find the (asymptotic) type I error \( \alpha \) of this test.

(b) Give an (asymptotic) upper bound on its type II error \( \beta \) for an alternative with an effect size \( \phi = \varepsilon \), for \( \varepsilon > 0 \) (that is when \( p_2 = p_1 + \varepsilon \)).

(c) How should we choose \( n \) to guarantee \( \alpha = 0.05 \), and power \( 1 - \beta = 0.80 \) when \( \varepsilon = 0.2 \).