Mathematical Statistics: Midterm Exam

Name:

Net ID:

You **may** use your notes and the course lecture notes, but no other sources.

The test is out of **50 points**. There are two sections: a true or false section with five questions (worth 15 points total), and a free response section with four questions (worth 35 points total).

True or false [3 pt each: 1 pt for correct answer, 2 pts for justification]

1. All unbiased estimators are consistent.

2. All consistent estimators are unbiased.

3. Fix a model $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$. If C is a $1-\alpha$ confidence set for θ, then the test

$$\psi(\omega) = \begin{cases} 0 & \exists \theta_0 \in \Theta_0 : \theta_0 \in C(\omega) \\ 1 & \text{otherwise} \end{cases}$$

has level α.

4. If $\sqrt{n}(T_n - \mu) \xrightarrow{d} Z$ for some $\mu \in \mathbb{R}$ and $Z \sim \mathcal{N}(0,1)$, then $\sqrt{n}(T_n^2 - \mu^2) \xrightarrow{d} Z^2$.

5. The mean squared error of an estimator never depends on the true value of the parameter.

ANOVA model [35 pts total] Given positive integers $n_1, \ldots, n_k$, consider independent observations Y_{ij}, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$ satisfying

$$Y_{ij} = \theta_i + \varepsilon_{ij} \,,$$

where $\theta_1, \ldots, \theta_k \in \mathbb{R}$ are unknown and the random variables ε_{ij} are all independent and satisfy

$$\mathbb{E}\varepsilon_{ij} = 0 \,, \mathrm{Var}\, \varepsilon_{ij} = \sigma^2 \,,$$

where we assume for simplicity that σ is known. This is called the ANOVA model, and it is used to understand the differences between different groups of test subjects.

1. **[5 pts]** Fix $t = (t_1, \ldots, t_k)$. Propose an estimator for $\sum_{i=1}^k t_i \theta_i$, and calculate its bias and variance.

2. **[5 pts]** Assume that the random variables ε_{ij} are σ^2-subgaussian. Given $\alpha \in (0,1)$, construct a $1-\alpha$ confidence set for $\sum_{i=1}^{k} t_i \theta_i$.

3. **[10 pts]** Assuming as above that the random variables ε_{ij} are σ^2-subgaussian, construct a confidence set $C \subseteq \mathbb{R}^k$ such that

$$\mathbb{P}_\theta \left\{(\theta_1, \ldots, \theta_k) \in C\right\} \geq 1 - \alpha \,.$$

Comment on the difference between the confidence set constructed in this set and the confidence set constructed in Part 2.

4. **[15 pts]** Assuming now that the random variables ε_{ij} are i.i.d. $\mathcal{N}(0, \sigma^2)$. Write *explicitly* the test statistic for a likelihood ratio test between

$$\Theta_0 := \{(\theta_1, \dots, \theta_k) : \theta_1 = \cdots = \theta_k\}$$
$$\Theta_1 := \mathbb{R}^k \setminus \Theta_0 \,.$$

Give an intuitive interpretation of this statistic.

This page intentionally left blank

This page intentionally left blank

This page intentionally left blank