Mathematical Statistics - Section 1 - NYU Spring 2019 - Midterm

NAME:

Warm-up First, we consider the usual family of exponential distributions. For any $\theta > 0$, we consider the pdf $x \mapsto f(x;\theta)$, defined for $x \in [0, +\infty)$ by $f(x;\theta) := \theta e^{-\theta x}$. We let $\theta_\star$ be the true, unknown value of the parameter, and $X_1, \dots X_n$ be n independent random variables, identically distributed with pdf $x \mapsto f(x;\theta_\star)$.

1. **Compute** the mean and the variance of a random variable with pdf $x \mapsto f(x;\theta_\star)$.

2. **Define** the empirical mean $\hat{m}_n$ of $X_1, \dots, X_n$.

3. **State** an asymptotic normality result for $\hat{m}_n$. *Make sure that your statement is complete, and includes an explicit expression for the expectation/variance that appear; the mode of convergence etc.*

4. **Compute** the max-likelihood estimator for $\theta_\star$.

5. **Compute** the Fisher information $I(\theta_\star)$.

The "multinoulli" distribution For any choice of two parameters p, q in the interval $(0,1)$ such that $0 < p+q < 1$, we consider a random variable X distributed according to the "multinoulli distribution" $\mathbb{P}_{p,q}$. It has three possible outcomes: $0, 1$, or 2 and we let

$$\mathbb{P}_{p,q}(X=0) = p, \quad \mathbb{P}_{p,q}(X=1) = q, \quad \mathbb{P}_{p,q}(X=2) = 1-(p+q).$$

Let $p_\star, q_\star$ be fixed, unknown parameters (the "real ones"), that we now want to estimate.

1. **Compute** the first moment $\mathsf{FM} := \mathbb{E}_{p_\star,q_\star}[X]$ and the second moment $\mathsf{SM} := \mathbb{E}_{p_\star,q_\star}[X^2]$.

2. **Show** that we have the relations

$$p_\star = 1 + \frac{\mathsf{SM} - 3\mathsf{FM}}{2}, \quad q_\star = 2\mathsf{FM} - \mathsf{SM}.$$

3. Let $X_1, \dots, X_n$ be i.i.d. random variables distributed according to $\mathbb{P}_{p_\star,q_\star}$. We let

$$\widehat{\mathsf{FM}}_n := \frac{1}{n}\sum_{i=1}^n X_i, \quad \widehat{\mathsf{SM}}_n := \frac{1}{n}\sum_{i=1}^n X_i^2.$$

Justify, briefly but precisely, why the following convergences hold:

$$\widehat{\mathsf{FM}}_n \longrightarrow_{n\to\infty} \mathsf{FM} \text{ in probability}, \quad \widehat{\mathsf{SM}}_n \longrightarrow_{n\to\infty} \mathsf{SM} \text{ in probability.}$$

4. The method of moments suggests to consider the following estimators for $p_\star, q_\star$:

$$\hat{p}_n := 1 + \frac{\widehat{\mathsf{SM}}_n - 3\widehat{\mathsf{FM}}_n}{2}, \quad \hat{q}_n := 2\widehat{\mathsf{FM}}_n - \widehat{\mathsf{SM}}_n.$$

Prove that $\hat{p}_n$ is an unbiased, consistent estimator of $p_\star$ and that $\hat{q}_n$ is an unbiased, consistent estimator of $q_\star$.
Hint: unbiasedness can be proven directly. For consistency, you may want to use the result of the previous question.

Do it yourself: a fallible, fair coin. We consider the following situation: a rudimentary computer is programmed as a "random number generator". Each time that we run the program, it answers 0 or 1 with equal probability. However, once in a while, something goes wrong, in which case the output of the program is $\frac{1}{2}$.

We let $R_1, \dots, R_n$ be the results obtained in the course of n runs, we assume that the results are independent. We let $p_\star$ be the (unknown) probability that the program "fails" and answers $\frac{1}{2}$.

- **Frame** the problem as a parametric statistical model.
- **Design** an estimator for $p_\star$ (there is more than one possible answer)
- **Justify**, concisely but precisely, its properties: (asymptotic) unbiasedness, consistency, asymptotic normality... ?