Department of Mathematics	final exam Asymptotic Statistics
VU	January 15, 2020

- **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.**

1. Let $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(n, q)$ be independent observations, with $n \in \mathbb{N}$ known and $p, q \in (0, 1)$ unknown parameters.

   (a) Define $\hat{p} = X/n$ and $\hat{q} = Y/n$. Prove that as $n \to \infty$, the random variable

   $$T = \frac{\hat{p} - \hat{q} - (p - q)}{\sqrt{\frac{\hat{p}(1-\hat{p}) + \hat{q}(1-\hat{q})}{n}}}$$

   converges in distribution to a standard normal.

   (b) Give an approximate confidence interval for the difference $p - q$ with asymptotic confidence level $1 - \alpha$. Motivate your answer.

2. Let $X_1, \ldots, X_n$ be i.i.d. observations with density

   $$f_{\theta_0}(x) = \frac{1}{2} e^{-|x - \theta_0|}, \qquad x \in \mathbb{R},$$

   where $\theta_0 \in \mathbb{R}$ is an unknown parameter. Let p be an odd natural number and consider the Z-estimator $\hat{\theta}_n^{(p)}$ defined as zero of the criterion function $\Psi_n^{(p)}$ defined by

   $$\Psi_n^{(p)}(\theta) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \theta)^p, \qquad \theta \in \mathbb{R}.$$

   (a) Show that for every odd natural number p, the estimator $\hat{\theta}_n^{(p)}$ is well defined and unique (i.e. that $\Psi_n^{(p)}$ has a unique zero for every odd p).

   In the following you may use the fact that for every even natural number k, it holds that

   $$\int_{\mathbb{R}} x^k f_0(x)\, dx = k!.$$

   You don't have to prove this.

   (b) Show that $\hat{\theta}_n^{(p)}$ is a consistent estimator for θ_0 for every odd natural number p.

   (c) Determine the limiting distribution of $\sqrt{n}(\hat{\theta}_n^{(p)} - \theta_0)$ for every odd natural number p.

   (d) Among all the estimators $\hat{\theta}_n^{(p)}$ considered, which one should we prefer?

3. Let $X_1, \ldots, X_n$ be independent normal random variables with unknown mean μ and variance 1. Clearly, the distribution function of X_i at the fixed point x is given by $F(x) = \Phi(x - \mu)$, where Φ is distribution function of the standard normal.

   We want to compare two estimators for $F(x)$, namely $\mathbb{F}_n(x)$ and $\Phi(x - \bar{X}_n)$, where $\mathbb{F}_n$ is the empirical distribution function.

   (a) Intuitively, without a formal computation, which estimator do you expect to be the better of the two? Motivate your answer.

   (b) Give the definition of $\mathbb{F}_n(x)$. Show it is a consistent estimator for $F(x)$ and determine its asymptotic variance.

   (c) Show that $\Phi(x - \bar{X}_n)$ is a consistent estimator for $F(x)$ and determine its asymptotic variance.

   (d) Use your answers of parts (a)–(c) to conjecture an inequality involving the density φ and the distribution function Φ of the standard normal distribution.

4. (a) Give the statement of the theorem that gives a minimax lower bound for estimators in smooth parametric models.

   (b) Give a proof of the theorem.

5. Consider the normal means model where we observe $Y \sim N_n(\theta_0, I)$, with $\theta_0 \in \mathbb{R}^n$ an unknown parameter vector. Let $\hat{\theta}$ be the lasso estimator defined by

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^n}{\operatorname{argmin}} \Big(\|Y - \theta\|^2 + \lambda \|\theta\|_1 \Big),$$

   where $\lambda > 0$ is a tuning parameter and $\|\theta\|_1 = \sum |\theta_i|$ is the ℓ^1-norm of the vector θ.

   (a) Show that $\hat{\theta}$ is given by $\hat{\theta}_i = h_\lambda(Y_i)$ for $i = 1, \ldots, n$, where

$$h_\lambda(y) = \begin{cases} y + \lambda/2 & \text{if } y < -\lambda/2, \\ 0 & \text{if } -\lambda/2 \le y \le \lambda/2, \\ y - \lambda/2 & \text{if } y > \lambda/2. \end{cases}$$

   (b) **This is a BONUS question.** Show that if θ_0 is s-sparse, i.e. if θ_0 has at most s non-zero coordinates, then for an appropriate choice of λ the risk $\mathbb{E}_{\theta_0} \|\hat{\theta} - \theta_0\|^2$ is bounded by a constant times $s \log(n/s)$ if $s = o(n)$. You may use the fact that for Z standard normal and $p > 0$, there exists a constant $C_p > 0$ such that for all $a \ge 1$,

$$\mathbb{E} Z^p 1_{Z > a} \le C_p a^{p-1} e^{-\frac{1}{2}a^2}.$$

Norming:

1(a): 4	**2(a): 3**	**3(a): 1**	**4(a): 2**	**5(a): 3**
1(b): 2	**2(b): 3**	**3(b): 4**	**4(b): 4**	**5(b): 4**
	2(c): 4	**3(c): 4**		
	2(d): 2	**3(d): 2**		