Afdeling Wiskunde	Exam Mathematische Statistiek/Asymptotic Statistics
Vrije Universiteit	20 December 2005

Give clear but brief explications of your answers!
You may write your answers in Dutch, English or French.
This exam has 5 problems on 2 pages.

1. The random variables $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ are independent, with every X_i Poisson distributed with mean μ and every Y_i Poisson distributed with mean ν. To test the null hypothesis $H_0: \mu = \nu$ we consider the test statistics

$$T_n = 2n(\sqrt{\bar{X}_n} - \sqrt{\bar{Y}_n})^2.$$

   a. Find critical values such that the asymptotic level of the sequence of tests is α (as $n \to \infty$).
   b. Show that the test rejects alternatives (with $\mu \neq \nu$) with probability tending to 1 (as $n \to \infty$).

2. Assume that $(Y_1, X_1), \ldots, (Y_n, X_n)$ are independent and identically distributed stochastic vectors distributed according to the non-linear regression model

$$Y_i = e^{\theta_0 X_i} + e_i,$$

for $\theta_0 \in (0, 1)$ and independent random variables X_i and e_i with the uniform distribution on $(0, 1)$ and a distribution with $\mathrm{E}e_i = 0$ and $\mathrm{E}e_i^2 < \infty$, respectively ($i = 1, 2, \ldots, n$). Let $\hat{\theta}_n$ be the point of minimum of

$$\theta \mapsto \sum_{i=1}^{n} \left(Y_i - e^{\theta X_i}\right)^2, \qquad \theta \in [0, 1].$$

   a. Give a heuristic argument showing that the sequence $\hat{\theta}_n$ converges in probability to θ_0.
   b. Which theorems and/or lemmas can be used to make this argument mathematically rigorous?
   c. Which limit distribution do you expect for the sequence $\sqrt{n}(\hat{\theta}_n - \theta_0)$?
   d. Assume that each e_i possesses the standard normal distribution. Find an expression for the Fisher information for θ.
   e. Relate the answers to questions b) and c) by reference to a theorem on the asymptotic properties of certain M-estimators.

3. Assume that X_n possesses a multinomial distribution with parameters n and $p = (p_1, \ldots, p_k)$.
   a. Formulate a theorem concerning the limit in distribution of the sequence of variables $\sum_{i=1}^{k} (X_{ni} - np_i)^2/(np_i)$ as $n \to \infty$.
   b. Prove this theorem.

4. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with density f.
   a. Give the formula for a kernel estimator $\hat{f}_n$ for f.
   b. Give the definition of the mean integrated square error (MISE) of $\hat{f}_n$ and its decomposition in square bias and variance.
   c. If it is known that f is once differentiable with derivative satisfying $\int f'(x)^2\, ds < \infty$, then it can be shown that the square bias is bounded above by a multiple of $C_f h^2$ for h the bandwidth of the kernel estimator and $C_f := \int f'(x)^2\, ds$. Which (asymptotic) choice of bandwidth can be recommended as $n \to \infty$ if no more is a-priori known about f?

5. The random variables $X_1, \ldots, X_n$ are i.i.d. with a uniform distribution on $[-\theta, \theta]$ for some unknown parameter $\theta > 0$. To estimate θ we consider the estimators $S_n = (X_{(n)} - X_{(1)})/2$ and $T_n = \sqrt{(3/n)\sum_{i=1}^n X_i^2}$.

a. Show that the sequence $n(X_{(1)} + \theta, \theta - X_{(n)})$ converges in distribution to a random vector $2\theta(Z_1, Z_2)$ for a pair of independent standard exponential variables Z_1, Z_2.

b. Determine the limit distribution of the sequence $n(S_n - \theta)$.

c. Indicate how an asymptotic confidence interval for θ based on S_n can be constructed.

d. Determine the limit distribution of the sequence $\sqrt{n}(T_n - \theta)$.

e. Which of the two estimator sequences S_n or T_n is preferable?

Credits:

1a: 6	2a: 4	3a: 8	4a: 2	5a: 3
1b: 3	2b: 2		4b: 3	5b: 2
	2c: 4		4c: 3	5c: 2
	2d: 3			5d: 3
	2e: 3			5e: 2

Mark = total/53*9+1.

Graded exams can be inspected at the student administration of the Vrije Universiteit from three weeks after the exam onwards.