Afdeling Wiskunde	Exam Mathematische Statistiek/Asymptotic Statistics
Vrije Universiteit	10 February 2005

Exam Mathematische Statistiek (6 ECTS): problems 1ab, 2, 3, 4, 5, 6.
Exam Asymptotic Statistics (8 ECTS): problems 1abc, 2, 3, 4, 6a, 7.
Give clear but brief explications of your answers!
You may write your answers in Dutch, English or French.
Graded exams can be inspected at the student administration of the Vrije Universiteit from three weeks after the exam onwards.

1. The random variables X_n and Y_n are independent and binomially distributed with parameters (n,p) and (n,q), respectively. To test the null hypothesis $H_0: p = q$ we consider the test statistics
$$T_n = \frac{4n(X_n - Y_n)^2}{(X_n + Y_n)(2n - X_n - Y_n)}.$$
   a. Find critical values such that the asymptotic level of the test is α.
   b. Show that the test rejects alternatives (with $p \neq q$) with probability tending to 1.
   c. (*Only Asymptotic Statistics.*) Find the limiting power at $(p,q) = (p_0, p_0 + h/\sqrt{n})$ as a function of h.

2. The random variables $X_1, \ldots, X_n$ are independent and $N(\mu_0, \sigma_0^2)$-distributed. To estimate the parameters (μ_0, σ_0) we consider the minimizer $(\hat\mu_n, \hat\sigma_n)$ of the function
$$(\mu, \sigma) \mapsto \frac{1}{n}\sum_{i=1}^n \Bigl(\frac{X_i - \mu}{\sigma}\Bigr)^4 + 12 \log \sigma.$$
   a. Give a heuristic argument showing that the estimators $(\hat\mu_n, \hat\sigma_n)$ are asymptotically consistent. (You may use that $\mathrm{E}(X_1 - \mu_0)^4 = 3\sigma_0^4$.)
   b. Which limit distribution do you expect for the sequence $\sqrt{n}(\hat\mu_n - \mu_0, \hat\sigma_n - \sigma_0)$?
   c. Determine the relative efficiency of $\hat\mu_n$ and $\bar X_n$ as estimators of μ_0. (You may use that $\mathrm{E}(X_1 - \mu_0)^6 = 15\sigma_0^6$.)
   d. Give a precise proof of the consistency of $\hat\mu_n$ as in a). (It is not necessary to consider $\hat\sigma_n$ in your arguments.)

3. The random vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. and bivariate-normally distributed with mean $\mu \neq 0$ and $\operatorname{var} X_i = \operatorname{var} Y_i = 1$ and $\operatorname{cov}(X_i, Y_i) = 1/2$. Find constants a_n and b_n such $b_n(\bar X_n \bar Y_n - a_n)$ converges in distribution to a non-degenerate limit distribution. Which limit distribution?

4. 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.

5. (*Only Mathematische Statistiek.*) Suppose that (X, Y) possesses a bivariate-normal distribution with expectation 0 and covariance matrix $\begin{pmatrix} 1 & 2 \\ 2 & 9 \end{pmatrix}$.
   a. The variable $X^2 + Y^2$ is distributed as a linear combination of χ_1^2-variables. Which linear combination?
   b. Determine constants A and B such that the variable $AX^2 + BXY + CY^2$ is χ^2-distributed with two degrees of freedom.

6. 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. *(Only Mathematische Statistiek.)* Give the definition of the mean integrated square error (MISE) of $\hat{f}_n$.
   c. *(Only Mathematische Statistiek.)* Express the bias of $\hat{f}_n(x)$ in f (and the kernel).

7. *(Only Asymptotic Statistics.)* Consider the following statistical experiments (i) and (ii):

(i) We observe an i.i.d. sample $X_1, \ldots, X_n$ from a density $x \mapsto p_\theta(x)$ for which the corresponding model $\{p_\theta : \theta \in \mathbb{R}\}$ satisfies the local asymptotic normality assumption: for any $h \in \mathbb{R}$:

$$\log \prod_{i=1}^{n} \frac{p_{\theta+h/\sqrt{n}}(X_i)}{p_\theta(X_i)} = h\Delta_{n,\theta} - \tfrac{1}{2}h^2 I_\theta + o_P(1),$$

under θ, where $I_\theta \in (0,\infty)$ and $\Delta_{n,\theta}$ is a sequence of random variables that tends in distribution under θ to the normal distribution with mean zero and variance I_θ.

(ii) We observe a single observation X from the $N(h, 1/I_\theta)$-distribution, where h is the unknown parameter.

   a. Formulate a theorem that relates these two experiments.
   b. Show that $S_n \rightsquigarrow S$ implies that $\liminf \mathrm{E} S_n^2 \geq \mathrm{E} S^2$, for any sequence of random variables S_n and S.
   c. It is known that in experiment (ii) any (randomized) estimator T of h satisfies $\sup_{h\in\mathbb{R}} \mathrm{E}_h (T-h)^2 \geq 1/I_\theta$. What does this imply for

$$\sup_{h\in\mathbb{R}} \mathrm{E}_{\theta+h/\sqrt{n}} n(T_n - \theta - h/\sqrt{n})^2$$

for estimators T_n in the experiments (i), as $n \to \infty$? (You may (but do not have to) assume that the sequence $\sqrt{n}(T_n - \theta)$ tends in distribution to a limit.)

Credits:

1a: 5	2a: 5	3a:10	4a: 5	5a: 3	6a: 3	7a: 6
1b: 5	2b: 5		4b: 5	5b: 3	6b: 3	7b: 1
1c: 3	2c: 5				6c: 4	7c: 3
	2d: 5					

Mark = total/66*9+1.