UNIVERSITETET I OSLO

Matematisk institutt

OBLIGATORY ASSIGNMENT: **STK4090/9090 – Statistical Large Sample Theory**

WITH: **Nils Lid Hjort and Emil Aas Stoltenberg**

HANDED OUT: **April 1, 2025**

DUE: **April 22, 2025**

FORMAT FOR YOUR ANSWER: **A .pdf-file.**

This oblig contains six exercises and comprises five pages, with Exercise 6 being optional.

Exercise 1 *Estimating the signal-to-noise ratio.* Consider independent observations $Y_1, \ldots, Y_n$ from a normal distribution with mean μ and variance σ^2.

(a) Write down the log-likelihood function $\ell_n(\mu, \sigma)$ and derive formulae for the maximum likelihood estimator, say $(\widehat{\mu}_n, \widehat{\sigma}_n)$ (notice that we are estimating the standard deviation here). Identify also the exact distributions of these estimators.

(b) Show that

$$\sqrt{n}\begin{pmatrix}\widehat{\mu}_n - \mu \\ \widehat{\sigma}_n - \sigma\end{pmatrix} \to_d \mathrm{N}_2\Big(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma^2 & 0 \\ 0 & \frac{1}{2}\sigma^2\end{pmatrix}\Big).$$

Briefly explain or investigate (via simulations, for example) how accurate the ensuing approximations are.

(c) Consider the parameter $\gamma = \mu/\sigma$, sometimes called the signal-to-noise ratio or the normalised mean. With the estimator $\widehat{\gamma}_n = \widehat{\mu}_n/\widehat{\sigma}_n$, show that

$$\sqrt{n}(\widehat{\gamma}_n - \gamma) \to_d \mathrm{N}(0, 1 + \tfrac{1}{2}\gamma^2).$$

(d) Let now $h(x) = \sqrt{2}\log(x/\sqrt{2} + \sqrt{1 + \frac{1}{2}x^2})$. Show that $\sqrt{n}(h(\widehat{\gamma}_n) - h(\gamma)) \to_d \mathrm{N}(0, 1)$. The function h is an example of what is called a variance stabilising transformation.

(e) Suppose you observe $\widehat{\mu}_n = 3.333$ and $\widehat{\sigma}_n = 0.222$ from $n = 50$ observations. Find an approximate 90 percent confidence interval for γ based on **(d)**.

Exercise 2 *Local asymptotics and contiguity.* The CLT and Lindeberg machineries yield normal limits and hence approximations in situations where independent observations come from given models. It is sometimes useful to extend such results to situations where observations stem from distributions close to, but not equal to, the postulated start models. The standard $\sqrt{n}$ speed of convergence for the CLT and relatives leads naturally to the notion of $O(1/\sqrt{n})$ neighbourhoods. Let $f_{a,b}(x)$ be the density of the Gamma distribution, that is

$$f_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1}\exp(-bx),\ x \geq 0,$$

for positive parameters a, b, such that $f_{1,b}$ is the density of the exponential distribution with expectation $1/b$.

(a) Show that the score function of the gamma distribution is

$$\begin{pmatrix}u_{a,b}(x)\\ v_{a,b}(x)\end{pmatrix} = \begin{pmatrix}\log(xb) - \psi(a)\\ a/b - x\end{pmatrix},$$

where $\psi(z)$ is the digamma function, $\psi(z) = \Gamma'(z)/\Gamma(z) = \int_0^\infty t^{z-1}\log(t)\exp(-t)\,\mathrm{d}t/\Gamma(z)$ for $z = a + ib$ with $a > 0$.

(b) Let $X_1, \ldots, X_n$ be i.i.d. exponentially distributed random variables with expectation $1/b$. For some h such that $1 + h/\sqrt{n} > 0$, introduce the distributions

$$Q_n(B_n) = \int_{B_n} \prod_{i=1}^n f_{1+h/\sqrt{n},b}(x_i)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n, \quad \text{and} \quad P_n(B_n) = \int_{B_n} \prod_{i=1}^n f_{1,b}(x_i)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n,$$

defined for events $B_n \subset \mathbb{R}^n$, and write $A_n(h) = \sum_{i=1}^n \{\log f_{1+h/\sqrt{n},b}(X_i) - \log f_{1,b}(X_i)\}$. Show that

$$A_n(h) = \frac{h}{\sqrt{n}} \sum_{i=1}^n u_{1,b}(X_i) - \tfrac{1}{2} h^2 \frac{\pi^2}{6} + o_{P_n}(1), \quad \text{as } n \to \infty,$$

and explain what this entails about the sequences of distributions Q_n and P_n.

(c) Show that

$$\begin{pmatrix} \sqrt{n}(\bar{X}_n - 1/b) \\ A_n(h) \end{pmatrix} \to_d \mathrm{N}_2\Big(\begin{pmatrix} 0 \\ -\frac{1}{2} h^2 \pi^2/6 \end{pmatrix}, \begin{pmatrix} 1/b^2 & h/b \\ h/b & h^2\pi^2/6 \end{pmatrix} \Big), \quad \text{under } P_n.$$

where $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$.

(d) Let $\widehat{b}_n = 1/\bar{X}_n$ be the maximum likelihood estimator under the exponential model. Show that

$$\sqrt{n}(\widehat{b}_n - b) \to_d \mathrm{N}(-hb, b^2), \quad \text{under } Q_n.$$

In other words, this is the limit distribution when our exponential model has this form of local misspecification.

Exercise 3 *A continuity type theorem for m.g.f.s.* Lévy's continuity theorem says that if a sequence of characteristic functions has a limit that is continuous in zero, then this limit is must be the characteristic function of some random variable, and we have convergence in distribution. In this exercise we work out something analogous for moment generating functions. We take as given that the distribution of a random variable X is determined by the moment generating function $M(t) = \mathrm{E}\exp(tX)$ provided $M(t)$ is finite in a neighbourhood of zero. A full proof of this fact is included in the optional Exercise 6.

(a) Let X be a random variable with moment generating function $M(t)$ on $(-s, s)$, $s > 0$. Let r be a positive number smaller than s, and show that

$$\Pr(|X| > K) \leq \exp(-rK)\{M(r) + M(-r)\}.$$

(b) Suppose that $(X_n)_{n\geq 1}$ and X have moment generating functions $(M_n)_{n\geq 1}$ and M, that these are finite on some interval $(-s, s)$, and that $M_n(t) \to M(t)$ on this interval. Show that $X_n \to_d X$.

(c) Suppose that $X_1, X_2, \ldots$ have moment generating functions $M_1, M_2, \ldots$ that exist on a common interval $(-s, s)$, and that $M_n(t) \to L(t)$ for some function $L(t)$ that is finite on $(-s, s)$. Show that X_n converges in distribution to a random variable with moment generating function L.

Exercise 4 *A nonnormal limit of a sum.* Normally limits are normal, but not always. Here we shall indeed work with variables with mean zero and variance one, where the sample averages have nonnormal limits. The basic construction is as follows. Let $U_1, U_2, \ldots$ be i.i.d., with mean zero and variance one, and with moment-generating function $M_0(s) = \mathrm{E}\exp(sU_i)$ finite in a neighbourhood around zero; in particular, all moments for the U_i are finite. Let independently of these $J_1, J_2, \ldots$ be independent Bernoulli variables with $\Pr\{J_i = 1\} = 1/i$, $\Pr\{J_i = 0\} = 1 - 1/i$. Then form

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n J_i \sqrt{i}\, U_i = \sum_{i=1}^n J_i \sqrt{i/n}\, U_i.$$

A picture to have in mind is that most of the terms will be zero, with non-zero contributions becoming both more rare and more big as time proceeds.

(a) Show that there will with probability one be infinitely many $J_i = 1$, i.e. non-zero terms in the Z_n sum as n grows.

(b) Show that the terms $J_i \sqrt{i} U_i$ have mean zero and variance one; hence also the normalised sample average Z_n has mean zero and variance one. Find also an expression for the kurtosis

$$\kappa_n = \mathrm{E}\, Z_n^4 - 3$$

of Z_n, and show that $\kappa_n \to \frac{1}{2} a_4$, where $a_4 = \mathrm{E}\, U_i^4$. Compare this to what we are 'used to' from the Lindeberg central limit theorem.

(c) We already know from point (b) that if Z_n has a limit distribution, it can't be normal. Working with the moment-generating function, show that

$$M_n(t) = \mathrm{E}\, \exp(tZ_n) = \prod_{i=1}^n \Big[1 + \frac{1}{i}\{M_0(t\sqrt{i/n}) - 1\}\Big],$$

for all t around zero for which $M_0(t)$ is finite.

(d) Here one may show that

$$\prod_{i=1}^n \Big[1 + \frac{1}{i}\{M_0(t\sqrt{i/n}) - 1\}\Big] \to \exp\Big\{\int_0^1 \frac{M_0(t\sqrt{x}) - 1}{x}\,\mathrm{d}x\Big\}. \qquad (1)$$

Work first with Special Case One, where we let U_i have the simple symmetric two-point distribution $\Pr\{U_i = 1\} = \Pr\{U_i = -1\} = \frac{1}{2}$. Find the limiting kurtosis for Z_n in this case. Show that

$$M_0(s) = \tfrac{1}{2}e^s + \tfrac{1}{2}e^{-s} = 1 + (1/2!)s^2 + (1/4!)s^4 + \cdots,$$

and use this to find an infinite-sum expression for the limit of $M_n(t)$. Have you now proved that Z_n has a limit distribution?

(e) Then work with Special Case Two, where the U_i have a double exponential distribution, of the form

$$f(u) = \tfrac{1}{2}\sqrt{2}\exp(-\sqrt{2}|u|)$$

on the real line (the $\sqrt{2}$ factor is there to ensure variance one). Find the moment-generating function $M_0(s)$ for the U_i, and then the moment-generating function $M(t)$ for the limit distribution of Z_n.

(f) For most cases, regarding the distribution for the U_i, it is hard to learn the explicit distribution for Z_n (even in cases where there might be a clear distribution for its limit). For Special Case Two, however, attempt to find the explicit distribution for Z_n, for any given n.

(g) We do not define this question as the most important one, on this occasion, but please attempt to prove the limit result of (1).

Exercise 5 *A Markov coin.* First a fair coin is flipped, then the second flip depends on the outcome of the first flip, the third flip on the outcome of the second, and so on. More formally, the random variable X_0 is Bernoulli(1/2), while the zero-one random variables $X_1, X_2, \ldots$ form a Markov chain evolving according to the transition probability matrix

$$\begin{pmatrix} 1-p_0 & p_0 \\ 1-p_1 & p_1 \end{pmatrix},$$

where $0 < p_j < 1$ for $j = 0, 1$. Thus $\Pr(X_j = 1 \,|\, X_0, \ldots, X_{j-1}) = \Pr(X_j = 1 \,|\, X_{j-1})$ almost surely for each j, and $\Pr(X_j = 1 \,|\, X_{j-1} = 0) = p_0$ and $\Pr(X_j = 1 \,|\, X_{j-1} = 1) = p_1$. To solve this exercise we need some results from Markov chain theory: Since $X_0, X_1, X_2, \ldots$ is an irreducible ergodic Markov chain, the limiting probabilities $\pi_0 = \lim_{n\to\infty} \Pr(X_n = 0 \,|\, X_0 = a)$ and $\pi_1 = \lim_{n\to\infty} \Pr(X_n = 1 \,|\, X_0 = a)$ exist and are independent of whether $a = 0$ or $a = 1$. They are

$$\pi_0 = \lim_{n\to\infty} \Pr(X_n = 0 \,|\, X_0 = 0) = \lim_{n\to\infty} \Pr(X_n = 0 \,|\, X_0 = 1) = \frac{1-p_1}{1+p_0-p_1},$$

and

$$\pi_1 = \lim_{n\to\infty} \Pr(X_n = 1 \,|\, X_0 = 0) = \lim_{n\to\infty} \Pr(X_n = 1 \,|\, X_0 = 1) = \frac{p_0}{1+p_0-p_1},$$

a result you can find in any book on Markov chain theory, for example Ross (2010, Theorem 4.1, p. 215). We also state without proof that

$$\frac{1}{n}\sum_{i=0}^{n-1}(1-X_i) \to_p \pi_0, \quad \text{and} \quad \frac{1}{n}\sum_{i=0}^{n-1} X_i \to_p \pi_1,$$

see for example Norris (1998, Theorem 1.10.2, p. 53).

(a) For each j, let $\mathcal{F}_j$ be the σ-algebra generated by $X_0, \ldots, X_j$. Show that

$$\mathrm{E}\,(X_j \,|\, \mathcal{F}_{j-1}) = (1 - X_{j-1})p_0 + X_{j-1}p_1,$$

almost surely, i.e., that the right hand side is a version of the conditional expectation.

(b) Define $Y_{n,k} = n^{-1/2}(1-X_{k-1})(X_k - p_0)$ and $Z_{n,k} = n^{-1/2}X_{k-1}(X_k - p_1)$ for $k = 1, \ldots, n$, and show that

$$\sum_{k=1}^{n}\begin{pmatrix} Y_{n,k} \\ Z_{n,k}\end{pmatrix} \to_d \mathrm{N}_2\Big(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\pi_0 p_0(1-p_0) & 0 \\ 0 & \pi_1 p_1(1-p_1)\end{pmatrix}\Big).$$

(c) Write $N_{a,b}(k) = \sum_{j=1}^{k} I\{(X_{j-1}, X_j) = (a,b)\}$ for the number of transitions from a to b occurring up to and including time k, and set $N_{a,b}(0) = 0$. Show that the likelihood function based on observing $X_0, \ldots, X_n$ is

$$L(p_0, p_1) = \tfrac{1}{2}\prod_{j=1}^{n} p_{X_{j-1}}^{X_j}(1 - p_{X_{j-1}})^{1-X_j} = \tfrac{1}{2}\, p_0^{N_{0,1}(n)} p_1^{N_{1,1}(n)}(1-p_0)^{N_{0,0}(n)}(1-p_1)^{N_{1,0}(n)},$$

and find expressions for the maximum likelihood estimators, say $\widehat{p}_{0,n}$ and $\widehat{p}_{1,n}$, of p_0 and p_1.

(d) We would like to say something about the uncertainty in our estimates, derive confidence intervals, etc. Deduce from your efforts so far that

$$\sqrt{n}\begin{pmatrix}\widehat{p}_{0,n} - p_0 \\ \widehat{p}_{1,n} - p_1\end{pmatrix} \to_d \mathrm{N}_2\Big(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}p_0(1-p_0)/\pi_0 & 0 \\ 0 & p_1(1-p_1)/\pi_1\end{pmatrix}\Big).$$

(e) Consider the estimand $\gamma = p_0/p_1$ and the estimator $\widehat{\gamma}_n = \widehat{p}_{0,n}/\widehat{p}_{1,n}$. Find the limiting distribution of $\sqrt{n}(\widehat{\gamma}_n - \gamma)$.

Exercise 6 *Uniqueness of the moment generating function.* We know (from the lectures) that characteristic functions characterise distributions, i.e., $\mathrm{E}\exp(itX) = \mathrm{E}\exp(itY)$ for all $t \in \mathbb{R}$ if and only if $X \sim Y$. This exercise is, in part, about showing that moment generating functions also characterise distributions. The moment generating function of a random variable X is

$$M(t) = \mathrm{E}\exp(tX),$$

provided the expectation is finite for all t in some open interval around zero.

(a) Suppose that the m.g.f. of X exists for $t \in (-t_0, t_0)$. For some $t \in (-t_0, t_0)$, the sum $M(-t) + M(t)$ is then clearly also finite. Use this to show that all the even absolute moments of X exists, that is $\mathrm{E}\,|X|^{2k} < \infty$; then fill in the odd gaps to conclude that all absolute moments of X are finite.

(b) The converse is not true, though. A random variable may have moments of all orders, but no m.g.f. Show that the log-normal is a case in point.

(c) For X as in (a) show that

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \operatorname{E} X^k, \quad t \in (-t_0, t_0).$$

From this we see that if Y is a random variable such that $\operatorname{E} Y^k = \operatorname{E} X^k$ for all $k \geq 0$, then their moment generating functions coincide. Show that the converse also holds, that is, if Y has m.g.f. M_Y and $M_X(t) = M_Y(t)$ for all $t \in (-t_0, t_0)$, then $\operatorname{E} Y^k = \operatorname{E} X^k$ for all $k \geq 0$.

(d) Suppose that X and Y have m.g.f.s M_X and M_Y such that $M_X(t) = M_Y(t)$ for $t \in (-t_0, t_0)$ and let φ_X and φ_Y be their characteristic functions. Let $\varphi_X^{(k)}(t) = \operatorname{E}(iX)^k \exp(itX)$ be the kth derivative of φ_X, with $\varphi_Y^{(k)}$ similarly defined. Show that

$$\varphi_X(t+h) = \operatorname{E} \exp(itX) \sum_{k=0}^{\infty} \frac{(ih)^k}{k!} X^k = \sum_{k=0}^{\infty} \frac{\varphi_X^{(k)}(t)}{k!} h^k, \quad \text{for } h \in (-t_0, t_0),$$

and similarly

$$\varphi_Y(t+h) = \sum_{k=0}^{\infty} \frac{\varphi_Y^{(k)}(t)}{k!} h^k, \quad \text{for } h \in (-t_0, t_0).$$

Setting $t = 0$ in these two expressions we get

$$\varphi_X(h) = \sum_{k=0}^{\infty} \frac{(ih)^k \operatorname{E} X^k}{k!} = \sum_{k=0}^{\infty} \frac{(ih)^k \operatorname{E} Y^k}{k!} = \varphi_Y(h),$$

showing that $\varphi_X(t) = \varphi_Y(t)$ for all $t \in (-t_0, t_0)$, and consequently $\varphi_X^{(k)}(t) = \varphi_Y^{(k)}(t)$ for all $t \in (-t_0, t_0)$ and each $k \geq 1$. But then, for $\varepsilon > 0$

$$\varphi_X(t_0 - \varepsilon + h) = \sum_{k=0}^{\infty} \frac{\varphi_X^{(k)}(t_0 - \varepsilon)}{k!} h^k = \sum_{k=0}^{\infty} \frac{\varphi_Y^{(k)}(t_0 - \varepsilon)}{k!} h^k = \varphi_Y(t_0 - \varepsilon + h), \quad \text{for } h \in (-t_0, t_0),$$

and similarly, $\varphi_X(-t_0 + \varepsilon + h) = \varphi_Y(-t_0 + \varepsilon + h)$ for $h \in (-t_0, t_0)$; from which we conclude that $\varphi_X(t) = \varphi_Y(t)$ for $t \in (-2t_0 + \varepsilon, 2t_0 - \varepsilon)$, and since $\varepsilon > 0$ was arbitrary and the c.f.s are continuous

$$\varphi_X(t) = \varphi_Y(t), \quad \text{for } t \in (-2t_0, 2t_0).$$

This implies that $\varphi_X^{(k)}(t) = \varphi_Y^{(k)}(t)$ for all $k \geq 0$ and $t \in (-2t_0, 2t_0)$, and therefore

$$\varphi_X(2t_0 - \varepsilon + h) = \sum_{k=0}^{\infty} \frac{\varphi_X^{(k)}(2t_0 - \varepsilon)}{k!} h^k = \sum_{k=0}^{\infty} \frac{\varphi_Y^{(k)}(2t_0 - \varepsilon)}{k!} h^k = \varphi_Y(2t_0 - \varepsilon + h), \quad \text{for } h \in (-t_0, t_0),$$

and similarly $\varphi_X(-2t_0 + \varepsilon + h) = \varphi_Y(-2t_0 + \varepsilon + h)$ for $h \in (-t_0, t_0)$; from which we conclude that $\varphi_X(t) = \varphi_Y(t)$ for $t \in (-3t_0, 3t_0)$ since $\varepsilon > 0$ was arbitrary and c.f.s are continuous. Run this argument once over and deduce that $\varphi_X(t) = \varphi_Y(t)$ for $t \in (-4t_0, 4t_0)$, and so on until you have covered the entire real line, at which point we have that $\varphi_X = \varphi_Y$.

References

Norris, J. R. (1998). *Markov Chains.* Cambridge University Press, Cambridge.

Ross, S. M. (2010). *Introduction to Probability Models. Tenth Edition.* Academic Press.