Final examen in *Asymptotic Statistics*

Thursday 15th of December 2016

Duration 4 hours

Manuscript notes of the lectures are allowed

1 Be Gaussian or not that is the question

Let $x \in \mathbb{R}$ be a fixed given point and $\theta \in \mathbb{R}$ be an unknown parameter. Let $X_1, \ldots, X_n$ be an i.i.d. sample with common law $\mathcal{N}(\theta, 1)$. Set

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt.$$

To estimate $p := \mathbb{P}(X_1 \leq x)$, we propose the two following estimators :

$$\widehat{p}_n := \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{X_i \leq x},$$
$$\tilde{p}_n := \Phi(x - \overline{X}_n).$$

Here, as usual, $\overline{X}_n$ denotes the empirical mean associated with the sample $X_1, \ldots, X_n$.

1. Show that both estimators converge almost surely to p.
2. Show that both estimators are asymptotically Gaussian when they are properly normalized. Compare the asymptotic variances.
3. What estimator would you choose ?

2 Do you speak contiguity ?

On a given measurable metric space, let p_1 and p_2 be two probability densities with respect to some $\sigma-$finite measure μ. We define the the total variation distance between p_1 and p_2 by

$$\|p_1 - p_2\|_1 := \int |p_1 - p_2| d\mu.$$

1. Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures. Set

$$|||\mathbb{P} - \mathbb{Q}|||_1 := \sup_A |\mathbb{P}(A) - \mathbb{Q}(A)|.$$

   Here, the supremum runs over all measurable sets. Show that if $\mathbb{P} = p_1\mu$ and $\mathbb{Q} = p_2\mu$ then

$$|||\mathbb{P} - \mathbb{Q}|||_1 = \|p_1 - p_2\|_1.$$

2. Let $(\mathbb{P}_n)_n$ and $(\mathbb{Q}_n)_n$ be two sequence of probability measures that are both dominated by μ. Assume that $|||\mathbb{P}_n - \mathbb{Q}_n|||_1 \to 0$, show that $\mathbb{P}_n$ and $\mathbb{Q}_n$ are mutually contiguous.
3. Let for $\theta > 0$, $\mathbb{P}_\theta$ be the uniform distribution on $[0, \theta]$. Let $\mathbb{P}_\theta^n$ denote the distribution of n i.i.d. draws from $\mathbb{P}_\theta$. Let $h \in \mathbb{R}$, discuss the contiguity of $\mathbb{P}_1^n$ and $\mathbb{P}_{1+h/\sqrt{n}}^n$.

3 Strange whistle : la, la, LAN for AR(1)

Let $\theta \in \mathbb{R}$ with $|\theta| < 1$. Let X_0 be a centred Gaussian random variable with variance $(1-\theta^2)^{-1}$ and (ε_n) be an i.i.d. sequence of standard Gaussian random variables. We assume that the sequence (ε_n) and X_0 are independent. For $n \in \mathbb{N}$, we set

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}.$$

1. For $n \geq 1$, write the joint density of $X_0, \varepsilon_1, \ldots, \varepsilon_n$. Compute the joint density of $X_0, X_1, \ldots, X_n$.
2. Let $\mathbb{P}_\theta^n$ be the distribution of $X_0, X_1, \ldots, X_n$. For $h \in \mathbb{R}$ small enough, compute the likelihood ratio $L_\theta^n(h) := \dfrac{\mathrm{d}\mathbb{P}^n_{\theta+\frac{h}{\sqrt{n}}}}{\mathrm{d}\mathbb{P}^n_\theta}(X_0, X_1, \ldots, X_n)$.
3. Set
$$S_n := \frac{1}{n}\sum_{i=1}^{n} X_i^2 \text{ and } T_n := \frac{1}{n}\sum_{i=0}^{n-1} X_i X_{i+1}.$$
Show that S_n and T_n are respectively unbiased estimator of $r(0) := \mathbb{E}(X_i^2) = \dfrac{1}{1-\theta^2}$ and $r(1) := \mathbb{E}(X_i X_{i+1}) = \dfrac{\theta}{1-\theta^2}$, $(i \in \mathbb{N})$.
4. For what follows, we will admit the following Theorem

   Theorem 1

   (a) The covariance matrix of the random vector $\sqrt{n}(S_n, T_n)^T$ converges to a positive matrix Γ.

   (b) $\sqrt{n}[(S_n, T_n)^T - (r(0), r(1))^T]$ converges in distribution to a two dimensional centred Gaussian vector with covariance matrix Γ.

   Show that $\log L_\theta^n(h)$ converges in distribution towards a Gaussian random variable with mean $m(h)$ and variance $V(h)$. Express $m(h)$ and $V(h)$ as functions of h, Γ, $r(0)$ and $r(1)$.
5. What is your intuition on the relationship between $m(h)$ and $V(h)$?

4 No life without Sobol

Let f be a measurable function from $\mathbb{R}^d$ to $\mathbb{R}$ and $X = (X_1, X_2, \ldots, X_d)$ be a random vector with independent components.

Set $I_d := \{1, 2, \ldots d\}$ and let $\mathbf{u}$ be a subset of I_d. Further, set

$$S^{\mathbf{u}} := \frac{\text{Var}\left(\mathbb{E}\left[Y|X_k, k \in \mathbf{u}\right]\right)}{\text{Var}(Y)}.$$

Let $X^{\mathbf{u}}$ be the random vector in $\mathbb{R}^d$ such that $X_k^{\mathbf{u}} = X_k$ if $k \in \mathbf{u}$ and $X_k^{\mathbf{u}} = X_k'$ if $k \notin \mathbf{u}$ where X_k' has the same law as X_k and is independent of all the other random variables (for example if $d = 5$ and $\mathbf{u} = \{2, 4\}$, $X = (X_1, X_2, X_3, X_4, X_5)$ and $X^{\mathbf{u}} = (X_1', X_2, X_3', X_4, X_5'))$. Further, let

$$Y := f(X) \text{ and } Y^{\mathbf{u}} := f(X^{\mathbf{u}}).$$

1. Show that

$$\text{Var}\left(\mathbb{E}\left[Y|X_k, k \in \mathbf{u}\right]\right) = \text{Cov}\left(Y, Y^{\mathbf{u}}\right).$$

2. Now let $(Y_j, Y_j^{\mathbf{u}})_{1 \le j \le N}$ be a N i.i.d. sample with the same distribution as $(Y, Y^{\mathbf{u}})$. We set further

$$S_N = \frac{\frac{1}{N}\sum_{j=1}^N Y_j Y_j^{\mathbf{u}} - \left(\frac{1}{N}\sum_{j=1}^N Y_j\right)\left(\frac{1}{N}\sum_{j=1}^N Y_j^{\mathbf{u}}\right)}{\frac{1}{N}\sum_{j=1}^N Y_j^2 - \left(\frac{1}{N}\sum_{j=1}^N Y_j\right)^2}.$$

   Show that

$$\sqrt{N}\left(S_N - S^{\mathbf{u}}\right) \underset{N \to \infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, \Sigma_S\right).$$

   Compute explicitly Σ_S.

3. Consider now the particular case where $d = 3$ and the inputs X_1, X_2, X_3 are i.i.d and uniformly distributed on $[-\pi, \pi]$. Further, let Y and f be defined by

$$Y = f(X_1, X_2, X_3) := \sin(X_1) + 7\sin(X_2)^2 + 0.1 X_3^4 \sin(X_1).$$

   Take successively $\mathbf{u} = \{1\}, \{2\}, \{3\}$. In each case, compute explicitly the exact values of $S^{\mathbf{u}}$ and Σ_S.