Final examen in *Asymptotic Statistics*

Thursday 21th of December 2017

Duration 3 hours-13h30-16h30

Manuscript notes and handout of the lectures are allowed

1 Cheap Fish

Let Z be a random variable with Poisson distribution of mean $\mu > 0$ ($Z \sim \mathcal{P}(\mu)$). We recall the following formulas

$$\begin{aligned} \mathbb{E}(Z) &= \mu, \\ \mathbb{E}(Z^2) &= \mu + \mu^2, \\ \mathbb{E}(Z^3) &= \mu + 3\mu^2 + \mu^3, \\ \mathbb{E}(Z^4) &= \mu + 7\mu^2 + 6\mu^3 + \mu^4. \end{aligned}$$

Recall further that the sum of two independent Poisson distributed random variables is also Poisson distributed. For $\lambda^* > 0$, let $X_1, \ldots, X_n$ be an i.i.d. sample with common law $\mathcal{P}(\sqrt{\lambda^*})$. To estimate the parameter λ^*, we consider the maximum likelihood estimator $\widehat{\lambda}$.

1. Show that $\widehat{\lambda} = \overline{X_n}^2$. Here, as usual, $\overline{X_n}$ denotes the empirical mean built on the sample $X_1, \ldots, X_n$. Compute the two first moments of $\widehat{\lambda}$. Compute its mean square error :
$$R_{\widehat{\lambda}}(\lambda^*) := \mathbb{E}[(\widehat{\lambda} - \lambda^*)^2].$$
2. Modify $\widehat{\lambda}$ to build an unbiased estimator $\widehat{\widehat{\lambda}}$. Compute $R_{\widehat{\widehat{\lambda}}}(\lambda^*)$.
3. Show that $\sqrt{n}(\widehat{\lambda} - \lambda^*)$ and $\sqrt{n}(\widehat{\widehat{\lambda}} - \lambda^*)$ converge both in distribution towards the same law. What is the limit law ?
4. What estimator should we use ? Why ?
5. Show that the statistical model is LAN. Is $\widehat{\widehat{\lambda}}$ optimal ?

2 Sobol indices in $\mathbb{R}^p$

For $i = 1, 2$ let $\left(E_i, \mathcal{A}_\rangle, \mu_i\right)$ be two probability spaces and f be a measurable application from $E_1 \times E_2$ to $\mathbb{R}^p$. Let X be a random variable taking its value in E_1 and Z arandom variable taking its value in $\mathbb{E}_2$. We assume that X and Z are independent. Set

$$Y := f(X, Z),$$

and assume that

$$\mathbb{E}\left(\|Y\|^2\right) < +\infty.$$

Here, $\|\cdot\|$ denotes the euclidean norm on $\mathbb{R}^p$.
We also set $\mathbb{E}[Y] := (E[Y_1], \ldots, \mathbb{E}[Y_p])^t$ (the mean vector of Y), and $\Sigma := (\mathrm{Cov}(Y_i, Y_j))_{1\le i,j\le p}$ (its covariance matrix). Recall that Y satisfies the following decomposition

$$\begin{aligned} Y =& \mathbb{E}[Y] + (\mathbb{E}[Y|X] - \mathbb{E}[Y] + \mathbb{E}[Y|Z] - \mathbb{E}[Y]) \\ &+ Y - (\mathbb{E}[Y] + \mathbb{E}[Y|X] - \mathbb{E}[Y] + \mathbb{E}[Y|Z] - \mathbb{E}[Y]) . \end{aligned} \tag{1}$$

The aim is to define some generalization of the classical Sobol indices defined in the lectures in the case $p = 1$ to a more general p. This generalized index is

$$S(Y) = \left(S^1(Y), S^2(Y)\right) = \left(\frac{\mathrm{Var}\,(\mathbb{E}[Y|X])}{\mathrm{Var}(Y)}, \frac{\mathrm{Var}\,(\mathbb{E}[Y|Z])}{\mathrm{Var}(Y)}\right)$$

Our index should verify the following statements

1. They are invariant by the action of translation, multiplication by a scalar and left-composition by any orthogonal matrix. In order words

$$S(Y + c) = S(Y), \text{and} S(\lambda Y) \qquad = |\lambda| S(Y) \forall (c, \lambda) \in \mathbb{R}^p \times \mathbb{R} \tag{2}$$

$$S(OY) = S(Y), \forall O \in O_p(\mathbb{R}) \tag{3}$$

   where $O_p(\mathbb{R})$ is the set of all orthogonal matrices of order p.

2. The index should be easy to estimate.

2.1 Case $p = 1$

2.1.1

Show that when $p = 1$ properties (2) and (3) are satisfied by Sobol indices.

2.2 Case $p \geq 1$

2.2.1 Building an index

1. For any square matrix M of size p, from equation (1) show that

   $$\mathrm{Tr}(M\Sigma) = \mathrm{Tr}(M\mathrm{Var}(E[Y|X])) + \mathrm{Tr}(M\mathrm{Var}(E[Y|Z])) + \mathrm{Tr}(M\mathrm{Var}(Y - E[Y|X] - E[Y|Z])).$$

2. Let us consider the sensitivity index associated with M

   $$S(M;Y) = \left(\frac{\mathrm{Tr}(M\mathrm{Var}(E[Y|X]))}{\mathrm{Tr}(M\Sigma)}, \frac{\mathrm{Tr}(M\mathrm{Var}(E[Y|Z]))}{\mathrm{Tr}(M\Sigma)}\right).$$

   and show that

   (a) for any orthogonal matrix O we have $S(M; OY) = S(O^t M O, Y)$

   (b) show that if $p = 1$ then the previous index does not depends on M. Show that, in this case, we recover the classical Sobol index.

3. Show that for $M = I_p$ (the identy matrix) properties (2) and (3) hold for $S(M;Y)$
4. Let
   Assume further that M has full rank ; and that $S(M;Y)$ is invariant by left-composition of f by any isometry of $\mathbb{R}^k$. We aim to prove
$$S(M;Y) = S(I_p, Y). \tag{4}$$
   (a) Show that for any symetric matrix V if M satisfies $M^t = -M$ then $\mathrm{Tr}(MV) = 0$. Deduce that we can then assume that M is a symetric matrix.
   (b) Show that if $M = PDP$ where D is a diagonal matrix and $P^tP = I_p$ we have
$$S(M, Y) = S(D, P^tY).$$
   Deduce that, we can then assume that M is a diagonal matrix.
   (c) Show that $M = \lambda I_p$ is then the only possible choice.
   (d) Conclude
5. Let X_1 and X_2 be i.i.d. standard Gaussian random variables and $(a, b) \in \mathbb{R}^2$. Compute $S(Y)$ in the following two cases
   (a) $Y = \begin{pmatrix} aX_1 \\ X_2 \end{pmatrix}$,
   (b) $Y = \begin{pmatrix} X_1 + X_1X_2 + X_2 \\ aX_1 + bX_1X_2 + X_2 \end{pmatrix}$.

2.3 Estimation of $S^1 = \frac{\mathrm{Tr}(\mathrm{Var}(E[Y|X_1]))}{\mathrm{Tr}(M\Sigma)}$

Let $(X_1, \ldots, X_N)$ and $(X'_1, \ldots, X'_N)$ be two independent i.i.d sample of law μ_1. Let further $(Z_1, \ldots, Z_N)$ be an i.i.d sample of law μ_2. Set $Y^i := f(X_i, Z_i)$ and $Y^{1,i} := f(X_i, Z'_i)$.

We define

$$S_N := \frac{\sum_{l=1}^k \left(\frac{1}{N} \sum_{i=1}^N Y_l^i Y_l^{1,i} - \left(\frac{1}{N} \sum_{i=1}^N \frac{Y_l^i + Y_l^{1,i}}{2} \right)^2 \right)}{\sum_{l=1}^k \left(\frac{1}{N} \sum_{i=1}^N \frac{(Y_l^i)^2 + (Y_l^{1,i})^2}{2} - \left(\frac{1}{N} \sum_{i=1}^N \frac{Y_l^i + Y_l^{1,i}}{2} \right)^2 \right)}. \tag{5}$$

1. Show that S_N converges almost surely to S^1
2. Show that $\sqrt{N}\,(S_N - S^1)$ converges in law to a centered Gaussian distribution.
3. Compute the variance of this limit distribution.