Examen du 21 janvier 2003, 13h05-16h05

SE 203: ESTIMATION ET INTRODUCTION AUX TESTS

ENSAE, *Paul Doukhan*

Les documents distribués en cours et les notes manuscrites ansi que les calculettes sont autorisés.
On traitera les exercices dans un ordre arbitraire, à la condition de le préciser clairement.
On pose $\Phi(x) = \int_{-\infty}^{\infty} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}$, et $\overline{\Phi}(x) = 1 - \Phi(x)$ est la fonction de queue gaussienne.
Barême indicatif: *1 point par question, sauf les questions A-7 et D-5 qui en compteront 2.*

Exercice A *(Modèle exponentiel)*

Soit $X_1, X_2, \ldots$, une suite iid de loi exponentielle de paramètre λ. On rappelle que leur densité s'écrit $\lambda e^{-\lambda x}$ pour $x \geq 0$.

1. Calculez la vraisemblance de l'échantillon $X^{(n)} = (X_1, \ldots, X_n)$. En déduire une statistique S exhaustive; on prouvera que cette statistique est complète.
2. Déterminez l'estimateur du maximum de vraisemblance fondé sur cet échantillon,
3. l'information de Fischer correspondante.
4. Ce modèle est-il régulier, si $\lambda \geq 0$, si $\lambda > 0$?
5. L'estimateur du maximum de vraisemblance précédent est-il efficace?
6. Asymptotiquement efficace?
7. Calculez la densité f_n de la loi de $S = X_1 + \cdots + X_n$. Indiquez l'intérêt de ce calcul: en particulier, expliquez pourquoi et de quelle manière il pourrait permettre d'obtenir des estimateurs de λ de bonne qualité (on ne demande pas de calculer explicitement de tels estimateurs).

Exercice B *(Paramètre de translation)*

On s'intéresse au problème d'estimation du paramètre de translation $\theta \in \mathbb{R}$ dans le modèle de densité (sur $\mathbb{R}$) $f_\theta(x) = f_0(x - \theta)$ donné par une densité f_0 (fixée, une fois pour toutes). On observe ainsi un $n-$échantillon iid $X = (X_1, \ldots, X_n)$ de loi f_θ, où $f_0(x) \neq 0$ p.s., $\int x f_0(x)\, dx = 0$ et $\int x^2 f_0(x)\, dx = 1$.

1. Déterminez la densité p_θ et la vraisemblance $V_\theta = p_\theta(X)$ du $n-$échantillon $X = (X_1, \ldots, X_n)$.
2. Etablir des conditions sur f_0 pour que ce modèle soit régulier. Ecrire l'information de Fisher de ce modèle.
3. Montrez que la borne de Cramer Rao s'écrit simplement à partir de l'inégalité de Cauchy Schwartz dans ce cas.
4. Prouvez que l'estimateur $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ est efficace si et seulement si $f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ est la densité gaussienne standard.
5. Dans le cas général, prouvez que cet estimateur est sans biais et la normalité asymptotique de $\sqrt{n}(\overline{X} - \theta)$ sous la loi P_θ.

Exercice C *(Détection d'un signal)*

Une antenne est réglée pour recevoir des signaux à une fréquence donnée. A chaque instant, elle reçoit une onde S d'amplitude $f(t)$. Cette amplitude est générée, soit par le signal, soit pour le bruit électromagnétique ambiant, les données sont collectées aux temps $T/N, 2T/N, \ldots, N$. Le modèle stochastique choisi est le suivant. Si S n'a pas émis de signal, on reçoit une amplitude $X_i = \xi_i \sim \mathcal{N}(0,\sigma_\xi^2)$ au temps iT/N. Si S a émis un signal, l'amplitude de ce signal est modélisée par une variable aléatoire $\eta_i \sim \mathcal{N}(0,\sigma_\eta^2)$; on reçoit alors un signal d'amplitude $X_i = \xi_i + \eta_i$ au temps iT/N. Les suites ξ_i est η_i sont mutuellement indépendantes. On suppose que les paramètres σ_ξ^2 et σ_η^2 sont connus.

1. Définir un modèle statistique paramétré par $\theta \in \{0,1\}$ rendant compte de ces observations; calculez sa vraisemblance.
2. Construire le test de Neyman-Pearson de niveau α pour décider de l'hypothèse H_0 (l'observation est un bruit blanc pur), contre $H_1 = H_0^c$ (l'hypothèse contraire). Précisez sa puissance (on notera que ce test au niveau α est donné par un quantile de loi du χ^2_N à N degrés de liberté).
3. Lorsque N est grand, donnez l'approximation $\widetilde{\beta}$ de la puissance fondée sur le théorème de limite centrale.
4. Alors, comment choisir N pour que ce test soit asymptotiquement non biaisé? On pourra utiliser le comportement asymptotique d'une loi du χ^2_N lorsque $N \uparrow \infty$.
5. Lorsque N est grand, montrez qu'il existe une constante $C > 0$ telle que la différence d'un $\alpha-$quantile de loi du χ^2_N et du quantile gaussien correspondant est de valeur absolue $\leq CN^{-1/8}$ (uniformément par rapport à α).

Exercice D *(Une variante du modèle Tobit)*

Soit $(\xi_n)_{n\in\mathbb{N}}$ une suite iid réelle de fonction de répartition $F : \mathbb{R} \to]0,1[$. On suppose la fonction F bijective. On pose $Y_n = \sigma\xi_n + m$ pour $n \in \mathbb{N}$. On considère le modèle statistique $(P_\theta^{\otimes n})_{\theta\in\Theta}$, loi du $n-$uplet $(X_1,\ldots,X_n)$ où X_i vaut 0 lorsque $Y_i < 0$, 1 si $Y_i \in [0,1]$ et 2, sinon, de plus Θ est l'ensemble des couples $\theta = (m,\sigma) \in \mathbb{R} \times \mathbb{R}^{+*}$.

1. Traduisez la propriété de F en termes de la loi de ξ_1.
2. Prouvez que le modèle est identifiable. Soient $S = \{(p_A,p_B,p_C) \in (\mathbb{R}^+)^3 | \ p_A + p_B + p_C = 1\}$, le simplexe de dimension 3 et sa tranche $T = \{(p_A,p_C) \in (\mathbb{R}^+)^2 | \ p_A + p_C \leq 1\}$. Alors S est en bijection avec T en vertu de la relation $p_B = 1 - p_A - p_C$. Posons $p_A = \mathbb{P}(Z_1 < 0)$, $p_B = \mathbb{P}(Z_1 \in [0,1])$ et $p_C = \mathbb{P}(Z_1 > 1)$. Pour prouver que le modèle est identifiable, on pourrait, par exemple, montrer que la fonction $K : S \to \Theta$, définie par $(p_A,p_C) \mapsto \theta = K(p_A,p_C)$ est bijective.
3. Montrez que l'estimateur du maximum de vraisemblance θ_n de θ s'écrit $\widehat{\theta} = K(\widehat{p}_A,\widehat{p}_C)$ pour des estimations par maximum de vraisemblance $\widehat{p}_A$, $\widehat{p}_C$ que l'on précisera.
4. Déterminez les propriétés asymptotiques de $(\widehat{p}_A,\widehat{p}_C)$ lorsque $n \to \infty$. Comment en déduiriez-vous un test de l'hypothèse $\theta = \theta_0$ contre $\theta \neq \theta_0$.
5. Que se passe-t-il lorsque la variable X_i prend 2 valeurs (0, ou 1) selon que $Y_i < 0$ ou $Y_i \geq 0$? Et lorsqu'elle prend 4 valeurs ($X_i = 2$ si $1 < Y_i \leq 2$ et $X_i = 3$ pour $Y_i > 2$)? Ces modèles sont-il identifiables, que dire de leur estimateur du maximum de vraisemblance?

Corrigé de l'examen SE203, session 2002-2003

Exercice A *(Modèle exponentiel)*

1. $\log p_\lambda(x) = n \log \lambda - \lambda(x_1 + \cdots + x_n)$, donc $S = X_1 + \cdots + X_n$ est une statistique exhaustive, on voit facilement qu'elle est complète.
2. De plus, $\dot{L}_\lambda(x) = n/\lambda - (x_1 + \cdots + x_n)$, s'annule lorsque $\lambda = 1/\overline{x}$ et $\widehat{\lambda} = 1/\overline{X}$
3. Pour $n = 1$, on obtient $I_1(\lambda) = \mathrm{Var}_\lambda X_1 = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$, donc $I_n(\lambda) = \frac{n}{\lambda^2}$.
4. Le modèle est régulier, si $\lambda > 0$ mais pas si $\lambda \geq 0$.
5. L'estimateur du maximum de vraisemblance est biaisé dans ce cas: l'inégalité de Cauchy Schwartz implique en effet que $\mathbb{E}_\lambda \widehat{\lambda} \mathbb{E}_\lambda \overline{X} > 1$, (car $1 = \sqrt{\widehat{\lambda}} \cdot \sqrt{\overline{X}}$), le fait que $(\overline{X})^2$ n'est pas constant implique que l'inégalité est stricte. Par suite la borne de Cramer Rao ne s'applique pas ici, un calcul direct prouve qu'il n'est pas efficace (la totalité des points sera accordée avec l'une des ces deux réponses).

   De fait, le calcul (qui utilise la loi déterminée en question 7) prouve que $\mathrm{Var}_\lambda \widehat{\lambda} = \frac{n^2\lambda^2}{(n-1)^2(n-2)} > \frac{\lambda^2}{n} = I_n^{-1}(\lambda)$. Notons enfin que $\lim_n n \mathrm{Var}_\lambda \widehat{\lambda} = I_1^{-1}(\lambda)$ (bien que cette propriété ne soit pas l'efficacité asymptotique).
6. C'est un résultat du cours, mais pour le prouver, on peut écrire,

   $$\sqrt{n}\left(\frac{1}{\overline{X}} - \lambda\right) = \sqrt{n}\left(\frac{\overline{X}}{(\overline{X})^2} - \frac{\lambda^2}{\lambda}\right)$$

   le lemme de Slutsky permet de conclure avec la relation $\mathbb{E}_\lambda X_1 = \frac{1}{\lambda}$.
7. On calcule par récurrence $f_n(x) = \frac{\lambda^n}{(n-1)!} x^n e^{-\lambda x}$ pour $x \geq 0$ (on peut aussi noter que $2\lambda X_1 \sim \gamma(1,1/2)$ et utiliser l'additivité de ces lois). La statistique S est exhaustive (question 1) donc un bon estimateur de λ s'écrit sous la forme $T = g(S)$, pour obtenir un estimateur optimal, on pourrait minimiser $R_\lambda(T) = \mathbb{E}_\lambda(g(S) - \lambda)^2$ ou son risque bayésien $\int R_\lambda(T) d\nu(\lambda)$. Dans le cas d'un mesure ν, exponentielle, elle aussi, les calculs sont possibles lorsque l'on a des éléments de calcul des variations. De tels calculs n'ont d'intérêt que pour des estimateurs biaisés ou pas réguliers en vertu de la borne FDCR.

Exercice B *(Paramètre de translation)*

1. $V_\theta = p_\theta(X) = \prod_i f_0(X_i - \theta)$.
2. Il suffit que f_0 soit dérivable et $I = \int f_0'^2/f_0 < \infty$ (en réalité, l'absolue-continuité de f_0 qui implique la dérivabilité p.s., et équivaut au fait que la dérivée définie p.s. soit intégrable, suffit avec cette condition). L'information de Fisher de ce modèle vaut nI.
3. Comme $\int x f_0(x) dx = 0$, on a

   $$\mathrm{Var}_\theta X_1 = \int x^2 f_\theta(x) dx = \int \left((x+\theta)^2 - \theta^2\right) f_0(x) dx = \int x^2 f_0(x) dx$$

et, puisque $\int x f_0'(x)dx = -\int f_0(x)dx = -1$ (intégration par parties), l'inégalité FDCR s'écrit ainsi

$$1 = \left(\int \left(\frac{f_0'(x)}{f_0(x)} \cdot x\right) f_0(x)dx\right)^2 \leq \int \left(\frac{f_0'(x)}{f_0(x)}\right)^2 f_0(x)dx \int x^2 f_0(x)dx$$

4. L'efficacité dans l'inégalité de Schwartz a lieu lorsque les fonctions $\frac{f_0'}{f_0}$ et $x \mapsto x$ sont proportionnelles.
5. C'est le TLC.

Exercice C *(Détection d'un signal)*

1. La densité du modèle s'écrit avec $\theta = 0$ ou $\theta = 1$ selon qu'on est dans l'hypothèse nulle H_0 d'un bruit pur ou l'hypothèse H_1 d'un signal bruité.

$$\log p_\theta(x) = -\frac{N}{2}\log(2\pi(\sigma_\xi^2 + \theta\sigma_\eta^2)) - \frac{1}{\sigma_\xi^2 + \theta\sigma_\eta^2}\sum_{i=1}^N X_i^2$$

2. Donc

$$\log\frac{p_1(x)}{p_0(x)} = -\frac{N}{2}\log\left(\frac{\sigma_\xi^2}{\sigma_\eta^2} + 1\right) + \frac{\sigma_\eta^2}{\sigma_\xi^2(\sigma_\xi^2 + \sigma_\eta^2)}\sum_{i=1}^N X_i^2$$

ainsi le test rejette l'hypothèse nulle lorsque $\sum_{i=1}^N X_i^2 > \sigma_\xi^2 \chi^2_{N,1-\alpha}$. La puissance du test vaut $\beta = \mathbb{P}\left((\sigma_\xi^2 + \sigma_\eta^2)\chi_N^2 \geq \sigma_\xi^2\chi^2_{N,1-\alpha}\right)$

3. La fonction $\Phi(x)$ est croissante et $(\Phi(z) \geq \alpha \Leftrightarrow z \geq \varphi_\alpha)$.
On utilise l'approximation $\chi^2_{N,1-\alpha} \simeq \widetilde{\chi}^2_{N,1-\alpha} = N + \sqrt{2N}\varphi_{1-\alpha}$ obtenue *via* le TLC; en effet les valeurs $\mathbb{E}\chi_N^2 = 1$,Var $\chi_N^2 = 2$ s'obtiennent rapidement grâce à des intégrations par parties.
Ainsi, en posant $\rho = \frac{\sigma_\xi^2}{\sigma_\xi^2 + \sigma_\eta^2}$, on réécrit

$$\beta = \mathbb{P}\left(\chi_N^2 \geq \rho\chi^2_{N,1-\alpha}\right) = \mathbb{P}\left(\frac{\chi_N^2 - N}{\sqrt{2N}} \geq \frac{\rho\chi^2_{N,1-\alpha} - N}{\sqrt{2N}}\right)$$

et l'approximation de la puissance est donc

$$\widetilde{\beta} = \overline{\Phi}\left(\frac{\rho\widetilde{\chi}^2_{N,1-\alpha} - N}{\sqrt{2N}}\right) = \overline{\Phi}\left(\frac{\sigma_\xi^2\varphi_{1-\alpha} - \sigma_\eta^2\sqrt{\frac{N}{2}}}{\sigma_\xi^2 + \sigma_\eta^2}\right)$$

4. Donc le test est asymptotiquement sans biais lorsque

$$\frac{\sigma_\xi^2\varphi_{1-\alpha} - \sigma_\eta^2\sqrt{\frac{N}{2}}}{\sigma_\xi^2 + \sigma_\eta^2} \leq \varphi_\alpha = -\varphi_{1-\alpha}$$

c'est-à-dire $\sqrt{\frac{N}{2}} \geq \varphi_{1-\alpha}\left(1+2\frac{\sigma_\xi^2}{\sigma_\eta^2}\right)$.
On obtient donc $\widetilde{\beta} \geq \alpha$ lorsque

$$N > 2\left(\varphi_{1-\alpha}\left(1+2\frac{\sigma_\xi^2}{\sigma_\eta^2}\right)\right)^2,$$

5. est une variante du théorème 1.2 du cours polycopié, on y remplace les variables de Bernoulli par des carrés de gaussiennes indépendantes .

Exercice D *(Une variante du modèle Tobit)*

1. La fonction $F : \mathbb{R} \to]0,1[$ est alors continue et d'inverse continue. De plus, la loi de ξ_1 est sans atome ($\mathbb{P}(\xi_1 = x) = 0$ pour tout $x \in \mathbb{R}$) et charge tout intervalle ($\mathbb{P}(\xi_1 \in]a,b[) > 0$ pour $a < b$).
En termes de loi, ceci signifie (de manière peu parlante) que la loi de ξ_1 est une mesure équivalente à la mesure de Lebesgue. Pour finir, rappelons que l'existence d'une densit n'est là que pour simplifier les conditions: il existe des loi continues et non absolument continues comme celle de Cantor
Seule l'hypothèse F bijective sera utilisée par la suite.
2. On a $p_A = F(-m/\sigma)$, $p_C = 1 - F((1-m)/\sigma)$. La loi de $(X_1,\ldots,X_n)$ s'écrit

$$\mathbb{P}_\theta((X_1,\ldots,X_n) = (x_1,\ldots,x_n)) = P_\theta(x_1,\ldots,x_n) = p_A^{n_A} p_B^{n_B} p_C^{n_C},$$

avec $n_A = \sum_i \mathbb{1}_{(x_i=0)}$, $n_B = \sum_i \mathbb{1}_{(x_i=1)}$ et $n_C = \sum_i \mathbb{1}_{(x_i=2)}$. Posons $a = F^{-1}(p_A)$, $c = F^{-1}(1-p_C)$, alors $m + a\sigma = 0, m + c\sigma = 1$ donc $\sigma = 1/(c-a)$, $m = -a/(c-a)$. L'application $\theta \mapsto (p_A,p_C)$ est donc bijective et le modèle est identifiable.
3. On trouve $\widehat{p}_A = N_A/n$, $\widehat{p}_B = N_B/n$ et $\widehat{p}_C = N_C/n$ avec des notations évidentes donc

$$\widehat{\sigma} = \frac{1}{F^{-1}\left(\frac{n-N_C}{n}\right) - F^{-1}\left(\frac{N_A}{n}\right)},\ \widehat{m} = \frac{-F^{-1}\left(\frac{N_A}{n}\right)}{F^{-1}\left(\frac{n-N_C}{n}\right) - F^{-1}\left(\frac{N_A}{n}\right)}$$

4. On a $\sqrt{n}(\widehat{p}_A - p_A, \widehat{p}_C - p_C) \to \mathcal{N}_2(0,\Sigma)$ pour Σ, la matrice de covariance (2×2) du couple $(\mathbb{1}_{(Z_1<0)}, \mathbb{1}_{(Z_1>1)})$. Un test asymptotique de l'hypothèse $\theta = \theta_0$ contre $\theta \neq \theta_0$ suit du TLC $\sqrt{n}(\widehat{p}_A - p_A^0, \widehat{p}_C - p_C^0) \to \mathcal{N}_2(0,\Sigma)$ obtenu lorsque $(p_A^0,p_C^0) = K^{-1}(\theta_0)$.
La normalité asymptotique de $\sqrt{n}(\widehat{\theta} - \theta)$ suit aussi du TLC précédent pour $(\widehat{p}_A,\widehat{p}_C)$ et du lemme de Slutsky car K est différentiable; elle permet de déterminer un test analogue.
5. – Dans le premier cas, l'application $\theta \mapsto p_A$ n'est plus injective lorsque l'on pose $p_A = \mathbb{P}(Z_1 < 0)$. De plus tous les couples tels que $F\left(-\frac{m}{\sigma}\right) = \frac{N_A}{n}$ sont des estimateurs du maximum de vraisemblance, qui existe mais n'est pas unique. On peut même paramétrer l'ensemble EMV de ces estimateurs

$$EMV = \left\{ \left(-tF^{-1}\left(\frac{N_A}{n}\right), t\right) \middle|\ t \in \mathbb{R} \right\}$$

– Dans le second cas, avec des notations analogues aux précédentes, l'application $\theta \mapsto H(\theta) = (p_A,p_B,p_C,p_D)$ n'est plus surjective et l'estimateur du maximum de vraisemblance n'existe plus forcément: Il existe si et seulement si

$$\left(\frac{N_A}{n},\frac{N_B}{n},\frac{N_C}{n},\frac{N_D}{n}\right) \in H(\Theta)$$

De plus, lorsque F est dérivable (la loi de ξ_1 a une densité), l'image $H(\Theta)$ est une sous-variété différentiable de dimension 2 dans $\mathbb{R}^4$, elle est donc d'intérieur vide dans le simplexe de dimension 4, $S_4 = \{(p_A,p_B,p_C,p_D) \in (\mathbb{R}^+)^4 |\ p_A + p_B + p_C + p_D = 1\}$, qui est une variété (à bords) de dimension 3 (car il est homéomorphe à sa tranche $T_3 = \{(p_A,p_B,p_C) \in (\mathbb{R}^+)^4 |\ p_A + p_B + p_C \leq 1\}$ qui est d'intérieur non vide, en vertu de la relation $p_D = 1 - (p_A + p_B + p_C)$.
Il est alors toujours unique: un tel estimateur est aussi estimateur du maximum de vraisemblance dans le modèle déja étudié; en effet, en posant $\widetilde{X}_i = \min\{X_i,2\}$, on retombe précisément sur la suite envisagée dans les questions précédentes.

Remarque. Notez que les questions sont de difficulté essentiellement croissante; la dernière question trouve des réponses diverses dont on ne s'étonnera pas si elles manipulent des notions qui ne vous sont pas familières. Le barême autorise la note maximale sans la traiter.

Barême indicatif
Exercice A sur (9) points
chaque question est notée 1 point, sauf les questions 5 et 7.
Exercice B sur 5 points
chaque question est notée 1 point.
Exercice C sur 5 points
chaque question est notée 1 point.
Exercice D sur (5+1) points
chaque question est notée 1 point, sauf la dernière pour laquelle 1 point additionnel peut être envisagé.