École Normale Supérieure
FIMFA — Statistique 2017–2018

Examen

1. Tests de rapports de vraisemblance (généralisés).

On considère un modèle exponentiel en forme canonique sur $(\mathcal{X}, \mathcal{F})$ dominé par la loi de probabité ν, paramétré par l'ouvert $\Theta \subset \mathbb{R}^p$. On note $Z(\theta) = \int_{\mathcal{X}} \exp(\langle \theta, T(x) \rangle) \nu(\mathrm{d}x)$. On choisit comme densité de P_θ par rapport à ν,

$$p_\theta(x) = \exp(\langle \theta, T(x) \rangle - \log Z(\theta)) \,.$$

On suppose le modèle identifiable.

On veut développer un test d'adéquation général dans ce modèle.

(a) (1½ points) Proposer une statistique S_n, paramétrée par $\theta_0 \in \Theta$ ($S_n(\theta_0)$ est un raccourci pour $S_n(\theta_0, X_1, \ldots, X_n)$) telle que sous $P_{\theta_0}^{\mathbb{N}}$, $(S_n(\theta_0))_n$ converge en loi vers une distribution non dégénérée qui ne dépend pas de θ_0. Préciser la loi limite.

(b) (1½ points) Proposer un test d'adéquation générique, de niveau asymptotique $\alpha \in]0, 1[$ pour distinguer l'hypothèse nulle $H_0 : X_1, \ldots, X_n \sim_{i.i.d.} P_{\theta_0}$ de l'alternative $H_1 : X_1, \ldots, X_n \sim_{i.i.d.} P_\theta, \theta \neq \theta_0$.

(c) (1 point) Étudier la limite de l'erreur de seconde espèce, lorsque vous appliquez les tests d'adéquation générique de niveau asymptotique $\alpha \in]0, 1[$, lorsque les données sont engendrées i.i.d. selon P_θ avec $\theta \neq \theta_0$.

(d) (3 points) On considère une suite d'alternatives contigues $(\theta_n)_n$ définie par $\theta_n = \theta_0 + h/\sqrt{n}$ avec $h \in \mathbb{R}^p$.

   i. Quelle est la loi limite de $\ell_n(\theta_n) - \ell_n(\theta_0)$ sous $P_{\theta_0}^{\otimes \mathbb{N}}$?

   ii. sous $P_{\theta_n}^{\otimes n}$ lorsque $n \to \infty$?

   iii. Si on considère la suite des tests de rapport de vraisemblance de niveau $\alpha \in]0, 1[$, quelle est la limite des seuils à utiliser ?

   iv. Quelle est la puissance limite de ces tests ?

(e) (3 points) Essayez de caractériser le comportement asymptotique de $S_n(\theta_0)$ sous $P_{\theta_n}^{\otimes n}$.

(f) (2 points) Calculer la puissance limite des tests définis à partir de $S_n(\theta_0)$ lorsque $\theta_n = \theta_0 + h/\sqrt{n}$. Les tests sont-ils asymptotiquement sans biais ?

Dans le cas unidimensionnel ($p = 1$), comparer les puissances asymptotiques du test de niveau α construit à partir de S_n et du test de rapport de vraisemblance de niveau α pour distinguer θ_0 de θ_n.

2. Estimation de n à partir de données binomiales.

Dans cet exercice, on observe $X_1, \ldots, X_N$ identiquement et indépendamment distribuées selon une loi binomiale $B(n, p)$ avec $n \in \{1, 2, \ldots\}$ inconnu et $p \in]0, 1[$ inconnu.

Attention : dans cet exercice, n est un paramètre, la taille de l'échantillon est notée N.

(a) (1 point) Majorer la distance en variation $\mathrm{d}_{\mathrm{TV}}\left(P^{\otimes n}, Q^{\otimes n}\right)$ entre $P^{\otimes n}$ et $Q^{\otimes n}$ par une quantité qui dépend de $\mathrm{d}_{\mathrm{TV}}(P, Q)$ et de n.

(b) (1 point) Pour $\lambda < n$, construisez une majoration de la distance en variation entre $B(n, \lambda/n)$ et la loi de Poisson d'espérance λ.

(c) (2 points) Pour $\lambda < n$, construisez une majoration de la distance en variation entre $B(n, \lambda/n)$ et $B(2n, \lambda/(2n))$.

(d) (3 points) On considère maintenant un estimateur $\widehat{n}_N$ de n à partir de N observations indépendantes. Montrer que

$$\sup_{P=B(n,p):(n,p)\in\mathbb{N}\times]0,1[} P^{\otimes N}\left\{\left|\frac{\widehat{n}_N - n}{n}\right| > \frac{1}{4}\right\} \geq \frac{1}{8}\,.$$

(e) (1 point) On suppose maintenant $p > 0$ connu. Montrez que l'estimateur au maximum de vraisemblance existe et donnez une méthode pour le calculer.

3. Régression logistique

La régression logistique est une technique statistique utilisée pour modéliser des problèmes de classification. On aborde ici le problème de la classification binaire. Dans ce problème, les données sont des couples $(x_i, y_i)_{i\leq n}$ avec $x_i \in \mathcal{X} = \mathbb{R}^d$ et $Y_i \in \mathcal{Y} = \{-1, 1\}$. Elles sont supposées obtenues par un échantillonnage i.i.d. selon une loi P inconnue sur $\mathcal{X} \times \mathcal{Y}$.

Dans un premier temps nous allons supposer que la loi conditionnelle de Y sachant X a une forme particulière. Pour $y \in \{-1, 1\}$

$$P_\theta\{Y = y|X = x\} = \frac{1}{1 + \mathrm{e}^{-y\langle\theta,x\rangle}} \,.$$

On note $p_{\theta,x} = \frac{1}{1+\mathrm{e}^{-\langle\theta,x\rangle}} = P_\theta\{Y = 1|X = x\}$.

Ceci définit un modèle dominé. La dominante est formée par le produit de la loi marginale de X (inconnue mais qu'on ne cherche pas à estimer) et de la loi uniforme sur $\{-1, 1\}$. La loi marginale de X est un *paramètre de nuisance*. Le vecteur θ est le paramètre d'intérêt. On notera $\theta^0 \in \mathbb{R}^d$ le paramètre inconnu qui définit la loi sous laquelle on échantillonne.

(a) (1 point) Montrer que la log-vraisemblance en θ peut se réduire à

$$\ell_n(\theta) = \sum_{i=1}^{n} -\log\left(1 + \mathrm{e}^{-y_i\langle\theta,x_i\rangle}\right) \,.$$

(b) (1 point) On note $\nabla\ell_n(\theta)$ la fonction score en θ. Montrer que

$$\mathbb{E}_{\theta^0}\left[\nabla\ell_n(\theta^0)\right] = 0 \,.$$

(c) (2 points) Calculer l'information de Fisher $I(\theta^0)$. Donner une condition suffisante sur la loi conditionnelle de Y sachant X et sur la covariance de X pour que l'information de Fisher soit définie positive.

(d) (1 point) Montrer que l'opposé du Hessien $-\nabla^2\ell_n(\theta)$ est semi-défini positif.

(f) (1 point) Une méthode du premier ordre : descente de gradient

$$\theta^{t+1} = \theta^t - \gamma\nabla\ell_n(\theta^t)$$

Choisir le pas du gradient γ.

(g) (1 point) Méthode de Newton-Raphson

$$\theta^{t+1} = \theta^t - (\nabla^2\ell_n(\theta^t))^{-1}\nabla\ell_n(\theta^t)$$

$$\theta^{t+1} = \theta^t - \left(\mathbf{X}^t\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^t(y - p_{\theta^t,\cdot}) \,.$$

IWLS (*iterated weighted least squares*)

(h) (1 point) Loi limite de la descente de gradient stochastique.

(i) (1 point) Région de confiance.

Le point de vue prédictif. L'hypothèse sur la loi conditionnelle de Y sachant X suit la forme mentionnée plus haut peut être inacceptable. En apprentissage, on ne formule pas d'hypothèse sur la loi conditionnelle de Y sachant X. On ne constente de chercher à determiner le vecteur θ^* qui minimise le risque de classification :

$$R(\theta) = \mathbb{E}\left[\mathbb{I}_{-Y\langle\theta,X\rangle>0}\right]$$

Pour estimer θ^*, on cherche $\widetilde{\theta}$ qui minimise le risque de classification empirique

$$R_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{-Y_i\langle\theta,X_i\rangle>0}$$

Dans le cas de la recherche du meilleur séparateur linéaire, la minimisation du risque empirique est un problème NP-difficile. Les praticiens cherchent plutôt à minimiser un risque qui se prête à un traitement algorithmique efficace.

(a) (1 point)

Question :	1	2	3	Total
Points :	12	8	10	30
Score :				