EXAMEN STATISTIQUE 1MF2E

Lundi 3 avril 2023 (8h00-9h45)

Partiel sans document (Une feuille A4 recto-verso autorisée)

Exercice 1 : Estimation (10 points)

On considère n observations $x_1, ..., x_n$ issues d'un échantillon $(X_1, ..., X_n)$ distribué suivant la même loi beta $B(1, \theta)$ de densité

$$f(x_i; \theta) = \begin{cases} \theta(1 - x_i)^{\theta - 1} \text{ si } x_i \in]0, 1[\\ 0 \quad \text{sinon} \end{cases}$$

avec $\theta > 0$.

1. (2pts) Montrer que l'estimateur du maximum de vraisemblance du paramètre θ noté $\widehat{\theta}_{\text{MV}}$ en fonction des variables X_i est
$$\widehat{\theta}_{\text{MV}} = -\frac{n}{\sum_{i=1}^{n} \ln(1 - X_i)}.$$
On prendra soin de vérifier que cette vraisemblance admet un unique maximum global.

2. (2pts) Déterminer la loi de la variable aléatoire $Y_i = -\ln(1 - X_i)$ et en déduire sa fonction caractéristique (en s'aidant des tables). En déduire la loi de $T_n = -\sum_{i=1}^{n} \ln(1 - X_i)$. En utilisant le fait que si une variable Y suit une loi gamma $\mathcal{G}(\nu, \theta)$, son inverse $Z = \frac{1}{Y}$ suit une loi inverse gamma $\mathcal{IG}(\nu, \theta)$, montrer que $\widehat{\theta}_{\text{MV}} = nZ_n$, où Z_n suit une loi inverse gamma $\mathcal{IG}(n, \theta)$.

3. (2pts) Déterminer la moyenne de l'estimateur $\widehat{\theta}_{\text{MV}}$. En déduire un estimateur sans biais et convergent du paramètre θ noté θ_n^*.

4. (1pt) L'estimateur θ_n^* est-il l'estimateur efficace du paramètre θ ?

5. (3pts) On suppose maintenant qu'on dispose d'une information a priori sur le paramètre θ résumée dans la loi a priori de densité de probabilité $p(\theta)$ définie par
$$p(\theta) = \lambda \exp(-\lambda\theta) \mathcal{I}_{\mathbb{R}^+}(\theta),$$
où $\mathcal{I}_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$ (on remarquera que cette loi est une loi gamma $\mathcal{G}(1, \lambda)$).
   - Déterminer l'estimateur MAP de θ noté $\widehat{\theta}_{\text{MAP}}$.
   - Montrer que la loi a posteriori de $\theta | x_1, ..., x_n$ est une loi gamma dont on déterminera les paramètres. En déduire l'estimateur de la moyenne a posteriori du paramètre θ noté $\widehat{\theta}_{\text{MMSE}}$.

Exercice 2 : Test Statistique (10 **points)**

On considère un ensemble de variables aléatoires $(X_1, ..., X_n)$ indépendantes et de même loi beta $B(1, \theta)$ de densité

$$f(x_i; \theta) = \begin{cases} \theta(1 - x_i)^{\theta - 1} \text{ si } x_i \in]0, 1[\\ 0 \quad \text{sinon} \end{cases}$$

avec $\theta > 0$. On désire tester les deux hypothèses

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1 > \theta_0.$$

1. (2pts) À l'aide du théorème de Neyman-Pearson, montrer que la statistique du test le plus puissant est $T_n = -\sum_{i=1}^{n} \ln(1 - X_i)$ et indiquer la région critique de ce test.
2. (1pt) On admet que si X_i suit une loi de densité $f(x_i; \theta)$, alors $Y_i = -\ln(1 - X_i)$ suit une loi gamma $\mathcal{G}(1, \theta)$. En utilisant le théorème central limite, déterminer la loi asymptotique de T_n sous les deux hypothèses.
3. (2pts) En utilisant la loi asymptotique de la question précédente, déterminer les risques de première et seconde espèce α et β du test en fonction des paramètres θ_0 et θ_1 et de la fonction de répartition d'une loi normale $\mathcal{N}(0, 1)$ notée F.
4. (2pts) Déterminer les courbes COR associées à ce test montrer qu'elles ne dépendent que de n et de $\frac{\theta_1}{\theta_0}$. Tracer l'allure de ces courbes COR pour différentes valeurs de n dans une première figure et pour différentes valeurs de $\frac{\theta_1}{\theta_0}$ dans une seconde figure.
5. (3pts) On désire vérifier que l'ensemble des observations $(x_1, ..., x_n)$ suit une loi beta $B(1, \theta)$ avec le paramètre $\theta = \frac{1}{2}$ à l'aide d'un test du χ^2. Déterminer la fonction de répartition de cette loi et en déduire que l'intervalle $]0, 1[$ est la réunion de trois intervalles équiprobables pour la loi $B\left(1, \frac{1}{2}\right)$ que l'on précisera. On compte le nombre d'observations x_i appartenant à ces trois intervalles et on trouve $K_1 = 13$, $K_2 = 8$ et $K_3 = 9$. Quelle est la valeur de la statistique du test du $\chi 2$? Exprimer le seuil de ce test noté S_α en fonction du risque α et de l'inverse de la fonction de répartition d'une loi du $\chi 2$ dont on précisera le nombre de degrés de liberté. On donne $S_{0.05} = 5.991$. Qu'en conclut-on ?

LOIS DE PROBABILITÉ CONTINUES

$\mathbf{m}$: moyenne σ^2 : variance **F. C. :** fonction caractéristique

LOI	Densité de probabilité	m	σ^2	F. C.
Uniforme	$f(x)=\frac{1}{b-a}$ $x\in\,]a,b[$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\dfrac{e^{itb}-e^{ita}}{it(b-a)}$
Gamma $\mathcal{G}(\nu,\theta)$	$f(x)=\frac{\theta^\nu}{\Gamma(\nu)}e^{-\theta x}x^{\nu-1}$ $\theta>0,\ \nu>0$ $x\geq 0$ avec $\Gamma(n+1)=n!\ \forall n\in\mathbb{N}$	$\dfrac{\nu}{\theta}$	$\dfrac{\nu}{\theta^2}$	$\dfrac{1}{\left(1-i\frac{t}{\theta}\right)^\nu}$
Inverse gamma $\mathcal{IG}(\nu,\theta)$	$f(x)=\frac{\theta^\nu}{\Gamma(\nu)}e^{-\frac{\theta}{x}}\frac{1}{x^{\nu+1}}$ $\theta>0,\ \nu>0$ $x\geq 0$ avec $\Gamma(n+1)=n!\ \forall n\in\mathbb{N}$	$\frac{\theta}{\nu-1}$ si $\nu>1$	$\frac{\theta^2}{(\nu-1)^2(\nu-2)}$ si $\nu>2$	(*)
Première loi de Laplace	$f(x)=\frac{1}{2}e^{-\lvert x\rvert},\quad x\in\mathbb{R}$	0	2	$\dfrac{1}{1+t^2}$
Normale univariée $\mathcal{N}\left(m,\sigma^2\right)$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}},\quad x\in\mathbb{R}$	m	σ^2	$e^{imt-\frac{\sigma^2t^2}{2}}$
Normale multivariée $\mathcal{N}_p(\boldsymbol{m},\boldsymbol{\Sigma})$	$f(x)=Ke^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{m})}$ $K=\frac{1}{\sqrt{(2\pi)^p\det(\boldsymbol{\Sigma})}}$ $\boldsymbol{x}=(x_1,...,x_p)^T\in\mathbb{R}^p$	$\boldsymbol{m}$	$\boldsymbol{\Sigma}$	$e^{i\boldsymbol{u}^T\boldsymbol{m}-\frac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}\boldsymbol{u}}$
Khi$_2$ χ^2_ν $\Gamma\left(\frac{1}{2},\frac{\nu}{2}\right)$	$f(x)=ke^{-\frac{x}{2}}x^{\frac{\nu}{2}-1}$ $k=\dfrac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}$ $\nu\in\mathbb{N}^*,\ x\geq 0$	ν	2ν	$\dfrac{1}{(1-2it)^{\frac{\nu}{2}}}$
Cauchy $c_{\lambda,\alpha}$	$f(x)=\dfrac{1}{\pi\lambda\left(1+\left(\frac{x-\alpha}{\lambda}\right)^2\right)}$ $\lambda>0,\ \alpha\in\mathbb{R}$	(-)	(-)	$e^{i\alpha t-\lambda\lvert t\rvert}$
Beta $B(a,b)$	$f(x)=kx^{a-1}(1-x)^{b-1}$ $k=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ $a>0,\ b>0$ $x\in\,]0,1[$ avec $\Gamma(n+1)=n!\ \forall n\in\mathbb{N}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	(*)

LOIS DE PROBABILITÉ DISCRÈTES

$\mathbf{m}$: moyenne σ^2 : variance **F. C. :** fonction caractéristique

$p_k = P[X = k]$ $p_{1,\ldots,m} = P[X_1 = k_1, \ldots, X_m = k_m]$

LOI	**Probabilités**	$\mathbf{m}$	σ^2	**F. C.**
Uniforme	$p_k = \frac{1}{n}$ $k \in \{1, \ldots, n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\dfrac{e^{it}\left(1-e^{itn}\right)}{n\left(1-e^{it}\right)}$
Bernoulli	$p_1 = P[X = 1] = p$ $p_0 = P[X = 0] = q$ $p \in [0,1] \quad q = 1 - p$	p	pq	$pe^{it} + q$
Binomiale $B(n,p)$	$p_k = C_n^k p^k q^{n-k}$ $p \in [0,1] \quad q = 1 - p$ $k \in \{0, 1, \ldots, n\}$	np	npq	$\left(pe^{it} + q\right)^n$
Binomiale négative	$p_k = C_{n+k-1}^{n-1} p^n q^k$ $p \in [0,1] \quad q = 1 - p$ $k \in \mathbb{N}$	$n\dfrac{q}{p}$	$n\dfrac{q}{p^2}$	$\left(\dfrac{p}{1-qe^{it}}\right)^n$
Multinomiale	$p_{1,\ldots,m} = \frac{n!}{k_1!\ldots k_m!} p_1^{k_1} \ldots p_m^{k_m}$ $p_j \in [0,1] \quad q_j = 1 - p_j$ $k_j \in \{0, 1, \ldots, n\}$ $\sum_{j=1}^m k_j = n \quad \sum_{j=1}^m p_j = 1$	np_j	Variance : $np_j q_j$ Covariance : $-np_j p_k$	$\left(\sum_{j=1}^m p_j e^{it}\right)^n$
Poisson $P(\lambda)$	$p_k = e^{-\lambda} \dfrac{\lambda^k}{k!}$ $\lambda > 0 \quad k \in \mathbb{N}$	λ	λ	$\exp\left[\lambda\left(e^{it} - 1\right)\right]$
Géométrique	$p_k = pq^{k-1}$ $p \in [0,1] \quad q = 1 - p$ $k \in \mathbb{N}^*$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{pe^{it}}{1-qe^{it}}$