Partiel de Statistique

Mardi 7 décembre 2010, 10h00-12h00

Partiel sans documents (une feuille manuscrite A4 est autorisée)
Barème indicatif. Exo 1 : 5+3+4 = 12pts, Exo 2 : 2+2+4 = 8 pts, Exo 3 : 2+2=4pts

Exercice 1 : Estimation

On considère n variables aléatoires $X_1, ..., X_n$ indépendantes suivant la même loi de densité

$$f\left(x;\sigma^2\right) = \frac{x}{\sigma^2}\exp\left[-\frac{x^2}{2\sigma^2}\right] I_{\mathbb{R}^+}(x)$$

avec $\sigma > 0$ et où $I_{\mathbb{R}^+}(x)$ est la fonction indicatrice sur $\mathbb{R}^+$ ($I_{\mathbb{R}^+}(x) = 1$ si $x \in \mathbb{R}^+$ et $I_{\mathbb{R}^+}(x) = 0$ si $x \notin \mathbb{R}^+$). Cette loi est appelée loi de Rayleigh et on utilisera la notation habituelle $X_k \sim \mathcal{R}\left(\sigma^2\right)$. On admettra que les premiers moments d'une loi de Rayleigh sont donnés par les relations suivantes

$$E\left[X_k\right] = \sigma\sqrt{\frac{\pi}{2}}, E\left[X_k^2\right] = 2\sigma^2, E\left[X_k^3\right] = \left(3\sqrt{\frac{\pi}{2}}\right)\sigma^3 \text{ et } E\left[X_k^4\right] = 8\sigma^4.$$

Cet exercice étudie différents estimateurs du paramètre $\theta = \sigma^2$.

Estimateur du Maximum de Vraisemblance

1) Montrer que la vraisemblance de $(x_1, ..., x_n)$ admet un unique maximum global pour une valeur de θ que l'on déterminera. En déduire l'estimateur du maximum de vraisemblance du paramètre θ noté $\widehat{\theta}_{\mathrm{MV}}$.
2) Déterminer la moyenne et la variance de $\widehat{\theta}_{\mathrm{MV}}$. En déduire que l'estimateur $\widehat{\theta}_{\mathrm{MV}}$ est un estimateur sans biais et convergent du paramètre θ.
3) Déterminer la borne de Cramer-Rao pour les esimateurs non biaisés du paramètre θ. L'estimateur $\widehat{\theta}_{\mathrm{MV}}$ est-il l'estimateur efficace du paramètre θ ?

Estimation Bayésienne

On suppose désormais qu'on dispose d'une information a priori sur le paramètre θ résumée dans la loi inverse-gamma $IG(\alpha, \beta)$ de densité

$$f\left(\theta|\,\alpha,\beta\right) = \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{1}{\theta^{\alpha+1}}\exp\left(-\frac{\beta}{\theta}\right) I_{\mathbb{R}^+}(\theta)$$

1) Montrer que la loi a posteriori de $\theta|\,x_1, ..., x_n$ est aussi une loi inverse-gamma dont on précisera les paramètres.
2) Déterminer l'estimateur du maximum a posteriori du paramètre θ noté $\widehat{\theta}_{\mathrm{MAP}}$. Expliquer le comportement de $\widehat{\theta}_{\mathrm{MAP}}$ lorsque $n \to \infty$.

Méthode des Moments

1) Á l'aide de l'expression de la moyenne d'une loi de Rayleigh, donner un estimateur de $\theta = \sigma^2$ noté $\widehat{\theta}_{\mathrm{M}}$ en utilisant la méthode des moments.
2) Déterminer la moyenne de l'estimateur $\widehat{\theta}_{\mathrm{M}}$ et en déduire un estimateur non-biaisé du paramètre θ construit à partir de $\widehat{\theta}_{\mathrm{M}}$.

$\boxed{\textbf{Exercice } 2 : \textbf{Test de Neyman-Pearson}}$

On considère n variables aléatoires $X_1, ..., X_n$ indépendantes telles que $X_k \sim \mathcal{R}\left(\sigma^2\right)$. On désire effectuer le test d'hypothèses

$$\begin{aligned} H_0 &: \sigma^2 = \sigma_0^2 \\ H_1 &: \sigma^2 = \sigma_1^2 \end{aligned}$$

où σ_0^2 et σ_1^2 sont deux quantités connues.
1) Déterminer la statistique du test de Neyman-Pearson notée T_n (qui ne dépend que des variables aléatoires $X_1, ..., X_n$) et discuter la forme de la région critique de ce test en fonction des valeurs de σ_0^2 et σ_1^2. On supposera dans la suite de cet exercice que $\sigma_1^2 > \sigma_0^2$.
2) On rappelle qu'une variable aléatoire Y de loi du khi-deux à n degrés de liberté notée $Y \sim \chi_n^2$ admet la densité

$$g_n(y) = \frac{y^{\frac{n}{2}-1} \exp\left(-\frac{y}{2}\right)}{2^{n/2}\Gamma\left(n/2\right)} I_{\mathbb{R}^+}\left(y\right)$$

et la fonction caractéristique

$$\phi_n(u) = E\left[e^{iuY}\right] = \frac{1}{\left(1 - 2iu\right)^{n/2}}, \qquad u \in \mathbb{R}$$

Montrer que si X_k suit une loi de rayleigh de paramètre σ^2, alors $Y_k = \frac{X_k^2}{\sigma^2}$ suit une loi du χ_2^2 à deux degrés de liberté. Calculer la fonction caractéristique de $\frac{T_n}{\sigma^2}$ et en déduire la loi de $\frac{T_n}{\sigma^2}$ sous les deux hypothèses H_0 et H_1.
3) Déterminer les risques α et β en fonction du seuil du test noté s_α, des paramètres σ_0^2 et σ_1^2, et de

$$G_{2n}(x) = \int_x^\infty g_{2n}(u)du$$

où g_{2n} est la densité d'une loi du khi-deux à $2n$ degrés de libertés.
4) Déterminer les caractéristiques opérationelles du récepteur (courbes COR) pour le test étudié dans cet exercice. Représenter les courbes COR pour différentes valeurs des paramètres σ_0^2 et σ_1^2.

$\boxed{\textbf{Exercice } 3 : \textbf{Test d'ajustement}}$

On jette $n = 120$ fois un dé et on observe $x_1, ..., x_{120}$ les 120 numéros relevés sur la face du dé regroupés dans ce tableau

face du dé	1	2	3	4	5	6
nombre de résultats	18	22	18	19	20	23

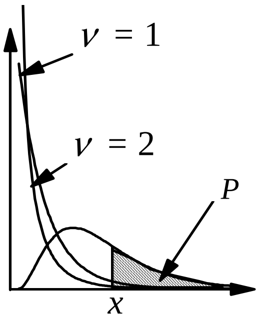
On désire tester si ce dé est truqué ou non.
1) Effectuer un test du χ^2 avec un risque $\alpha = 0.05$ et conclure (pour l'application numérique, on utilisera les tables jointes à cet énoncé).
2) Un non-statisticien propose le test suivant

$$\text{Rejet de } H_0 \text{ si } \left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{7}{2}\right)^2 > s_\alpha$$

où s_α est un seuil dépendant du risque de première espèce α. Rappeler la loi approchée de $\frac{1}{n}\sum_{i=1}^{n} X_i$ résultant de l'application du théorème de la limite centrale. En déduire une expression du seuil s_α en fonction de α et de l'inverse de la fonction G_1 introduite dans l'exercice précédent.

Distribution du χ^2_ν

$$\boxed{P\left[\chi^2_\nu \geq x\right] = P}$$



	P						
ν	**0.99**	**0.975**	**0.95**	**0.90**	**0.80**	**0.70**	**0.60**
1	0.000	0.001	0.004	0.016	0.064	0.148	0.275
2	0.020	0.051	0.103	0.211	0.446	0.713	1.022
3	0.115	0.216	0.352	0.584	1.005	1.424	1.869
4	0.297	0.484	0.711	1.064	1.649	2.195	2.753
5	0.554	0.831	1.145	1.610	2.343	3.000	3.655
6	0.872	1.237	1.635	2.204	3.070	3.828	4.570
7	1.239	1.690	2.167	2.833	3.822	4.671	5.493
8	1.646	2.180	2.733	3.490	4.594	5.527	6.423
9	2.088	2.700	3.325	4.168	5.380	6.393	7.357
10	2.558	3.247	3.940	4.865	6.179	7.267	8.295

Distribution du χ^2(suite)

	P							
ν	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.025**	**0.01**
1	0.455	0.708	1.074	1.642	2.706	3.841	5.024	6.635
2	1.386	1.833	2.408	3.219	4.605	5.991	7.378	9.210
3	2.366	2.946	3.665	4.642	6.251	7.815	9.348	11.345
4	3.357	4.045	4.878	5.989	7.779	9.488	11.143	13.277
5	4.351	5.132	6.064	7.289	9.236	11.071	12.833	15.086
6	5.348	6.211	7.231	8.558	10.645	12.592	14.449	16.812
7	6.346	7.283	8.383	9.803	12.017	14.067	16.013	18.475
8	7.344	8.351	9.524	11.030	13.362	15.507	17.535	20.090
9	8.343	9.414	10.656	12.242	14.684	16.919	19.023	21.666
10	9.342	10.473	11.781	13.442	15.987	18.307	20.483	23.209