Sorbonne Université
Master Mathématiques et Applications
MU4MA015 : Statistique

9 décembre 2024
Aucun document autorisé
Durée : 1 heure

Contrôle Continu

Rappels

- Il sera tenu compte du soin apporté à la présentation et à la rédaction.
- Si vous n'arrivez pas à démontrer un résultat, vous pouvez l'admettre pour la suite de l'exercice.
- Si $T \sim \Gamma(a,b)$ avec $a > 2$ et $b > 0$, alors $\mathbb{E}[1/T] = \frac{b}{a-1}$ et $\mathbb{E}[1/T^2] = \frac{b^2}{(a-1)(a-2)}$.

Exercice 1 (6 points)

Soit $n \geq 3$. On considère le cadre bayésien

$$\boldsymbol{\theta} \sim \Pi = \mathcal{E}(1)$$
$$\mathbf{X}|\boldsymbol{\theta} \sim P_{\boldsymbol{\theta}}^{\otimes n} = \mathcal{E}(\boldsymbol{\theta})^{\otimes n}.$$

1. Donner la loi a posteriori.
2. On considère la perte
$$\ell(\theta, t) = e^{\theta}(t-\theta)^2.$$
Donner un estimateur de Bayes $T^\star = T^\star(\mathbf{X})$ pour Π et cette fonction de perte.
3. Montrer que, sous $P_\theta^{\otimes n}$, la variable $n\overline{X}_n$ suit une loi Gamma dont on précisera les paramètres. Grâce au rappel, en déduire $\mathbb{E}_\theta\left[\frac{1}{n\overline{X}_n}\right]$ et $\mathbb{E}_\theta\left[\frac{1}{(n\overline{X}_n)^2}\right]$.
4. Montrer que le risque ponctuel $R(\theta, T^\star) = \mathbb{E}_\theta[\ell(\theta, T^\star)]$ peut se mettre sous la forme
$$R(\theta, T^\star) = \frac{an+b}{(n-1)(n-2)}\theta^2 e^{\theta},$$
où a et b sont deux constantes positives que l'on précisera.
5. Que vaut le risque de Bayes $R_B(\Pi)$? En déduire le risque minimax R_M.
6. Pouvez-vous proposer un autre estimateur de Bayes ?
7. On souhaite tester
$$H_0 : \boldsymbol{\theta} = 1 \quad \text{contre} \quad H_1 : \boldsymbol{\theta} = 2\,.$$
Pour cela, on considère maintenant la loi a priori $\Pi = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_2$ et la fonction de perte équilibrée. Montrer que le test de Bayes peut se mettre sous la forme $\varphi^*(\mathbf{X}) = \mathbb{1}_{\{\overline{X}_n \leq c\}}$, pour une certaine constante $c \in \mathbb{R}$ que l'on déterminera.

Exercice 2 (4 points)

On considère, dans cet exercice, un modèle linéaire gaussien

$$Y = X\beta + \varepsilon = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} \beta + \varepsilon,$$

où, comme d'habitude,

— Y est un vecteur aléatoire de n observations $Y_1, \ldots, Y_n$;
— X est une matrice de mesures connues et déterministes, de taille $n \times 2$ et de rang 2 ;
— β est un vecteur déterministe inconnu de taille 2 ;
— ε est un vecteur aléatoire d'erreurs inconnues avec $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ où $\sigma^2 > 0$ est inconnu.

1. Par quoi faut-il remplacer `???` pour que le script suivant construise une réalisation `y` du vecteur aléatoire Y.

[3]:
```
n = 100
noise = stats.norm(scale=true_sigma2())
t = measure_t(n)
beta = true_beta()
X = np.column_stack((np.ones(n), t))
y = ???
```

2. Étant donné les informations suivantes sur la matrice X et sur la réalisation $y = Y(\omega)$, calculer l'estimateur des moindres carrés de β, noté $\widehat{\beta}$, et donner sa réalisation $\widehat{\beta}(\omega)$.

$$\sum_{i=1}^{n} t_i = 0, \qquad \sum_{i=1}^{n} t_i^2 = 400, \qquad \sum_{i=1}^{n} t_i y_i = 100, \qquad \sum_{i=1}^{n} y_i = 100, \qquad \widehat{\sigma}^2 = 1.$$

3. Que produit le script suivant (justifier par des éléments théoriques) ? Le résultat peut-il être lu dans le résumé des résultats donné ensuite via la commande `res.summary()` ?

[10]:
```
model = sm.OLS(y, X)
res = model.fit()
```

[11]:
```
level = .05
stu = stats.t(df=n-2)
ic_bound = np.sqrt(res.scale / np.diag(X.T @ X)) * stu.ppf(1 - level/2)
for c, s in zip(res.params, ic_bound):
    print(f"[{c-s:0.3f}, {c+s:0.3f}]")
```

```
[0.802, 1.198]
[0.151, 0.349]
```

[12]:
```
res.summary()
```

[12]:

Dep. Variable :	y	**R-squared :**	0.203
Model :	OLS	**Adj. R-squared :**	0.195
Method :	Least Squares	**F-statistic :**	25.00
Date :	Fri, 29 Nov 2024	**Prob (F-statistic) :**	2.51e-06
Time :	11 :20 :50	**Log-Likelihood :**	-140.88
No. Observations :	100	**AIC :**	285.8
Df Residuals :	98	**BIC :**	291.0
Df Model :	1		
Covariance Type :	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	1.0000	0.100	10.000	0.000	0.802	1.198
x1	0.2500	0.050	5.000	0.000	0.151	0.349

Omnibus :	4.728	**Durbin-Watson :**	2.206
Prob(Omnibus) :	0.094	**Jarque-Bera (JB) :**	5.178
Skew :	-0.239	**Prob(JB) :**	0.0751
Kurtosis :	4.007	**Cond. No.**	2.00

4. On rappelle que l'ellipse centrée en $(x_0, y_0) \in \mathbb{R}^2$ et définie par :

$$\left\{ (x', y') \in \mathbb{R}^2, \frac{(x' - x_0)^2}{a^2} + \frac{(y' - y_0)^2}{b^2} \leq 1 \right\},$$

(où $a > 0$ et $b > 0$) a pour aire πab.

Commenter le script suivant (utiliser des éléments théoriques) et son résultat.

[13]:
```
fish = stats.f(dfn=2, dfd=n-2)

box_bound = np.sqrt(res.scale / np.diag(X.T @ X)) * stu.ppf(1 - level/4)

print(f"Aire 1 : {np.prod(2 * box_bound):.3f}")
print(f"Aire 2 : {np.pi / np.prod( np.sqrt(np.diag(X.T @ X) / (2 * res.scale * fish.ppf(1 - level))) ):.3f}")
```

```
Aire 1 : 0.104
Aire 2 : 0.097
```