

Groupe 8

S. Gadat : sebastien.gadat @ tse-fr.eu.

Exercice 1 (Un café pour commencer (*)). Une cafétéria souhaite remplacer son service par un ou plusieurs distributeurs automatiques de boisson. Afin de calibrer au plus juste le nombre de distributeurs nécessaires, elle mène donc une enquête pour connaître au mieux les besoins de ses utilisateurs.

1. Dans un premier temps, elle observe les temps *successifs* d'arrivée de clients dans son local, enregistrant ainsi un vecteur $T = (T_1, \dots, T_n)$. En considérant les intervalles de temps ainsi formés, et en admettant que les clients arrivent les uns après les autres selon un flux identique toute la journée, proposer un modèle statistique. On notera pour cela

$$\forall i \geq 1 : \quad \Delta_i = T_i - T_{i-1},$$

$$T_0 = 0 \leq T_1 \leq \dots \leq T_n \leq \dots$$

$$\Delta_i = T_i - T_{i-1} \quad \text{Avantage de manipuler les } \Delta_i : \text{ ce sont des quantités } \geq 0$$

Proposer 1 modèle stat ? C'est quoi déjà ??

triplet : - espace mesurable + tribu
- famille de lois de probas.

$$\mathcal{L}_i : \mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \{E(\lambda), \lambda > 0\}$$

On peut faire 1 par + :

$$\mathbb{R}_+^m, \mathcal{B}(\mathbb{R}_+)^m, \{E(\lambda)^{\otimes m}, \lambda > 0\} \quad \triangle!$$

Nous avons ajouté ici l'hypothèse non innocente que les m obs. sont II ! Ça c'est ok. Mais il y a aussi l'hyp probablement assez éloignée de la réalité que le λ reste le même tout le temps...

2. Proposer un estimateur du temps moyen d'arrivée d'un nouveau client, noté $\hat{\Delta}_n$ ainsi qu'une approximation du nombre de clients arrivant sur un intervalle de temps donné de longueur $h > 0$ fixé.

$$\hat{\Delta}_n = \frac{1}{n} \sum_{i=1}^n \Delta_i \quad \text{est un estimateur du temps}$$

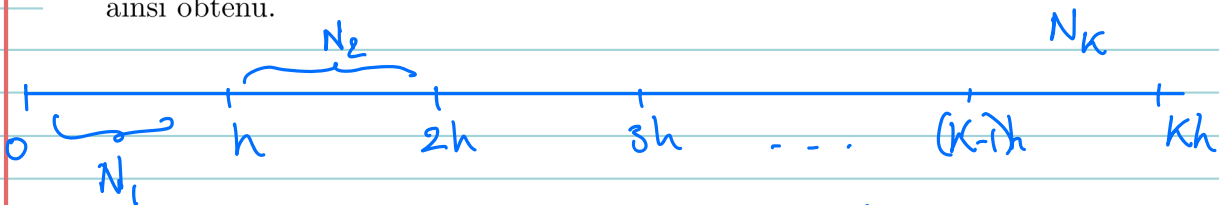
moyen d'attente.

! Dans le param $\lambda \rightarrow \lambda e^{-\lambda a}$, on a
 $E[A] = \frac{1}{\lambda}$.

Approximation du nombre de clients sur $[0; h]$?

$$\frac{h}{\hat{\Delta}_n}$$

3. Dans un second temps, pour une durée $h > 0$ fixée, elle compte sur une fenêtre de temps entre 0 et Kh le nombre de clients arrivant tous les intervalles de temps de taille h . Proposer alors un second modèle statistique pour le vecteur $N = (N_1, N_2, \dots, N_K)$ ainsi obtenu.



Modèle Poissonien $\{P(\theta)^{\otimes K}, \theta > 0\}$

Espace mesurable $\mathbb{N}$

Tribu: $\mathcal{P}(\mathbb{N}^K)$

$$4] \quad \hat{N}_{K,h} = \frac{N_1 + \dots + N_K}{K}$$

5] Il existe un lien très fort entre les 2 modèles:

si les intertemps sont $E(\lambda)$, alors les N_k sont $P(\lambda h)$
et si les arrivées suivent 1 processus de Poisson de param λ
alors on a la réciproque.

Objectif : Modèle translation/échelle. Cas Gaussien (questions 3/4) ()**

Exercice 3 (Modèle de translation et d'échelle). Soit g une densité par rapport à la mesure Leb. On considère le modèle statistique

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{p_{n,\theta} \cdot \text{Leb}^{\otimes n} : \theta \in \Theta := \mathbb{R} \times \mathbb{R}_*^+\})$$

où

$$p_{n,\theta}(x_1, \dots, x_n) := \sigma^{-n} \prod_{i=1}^n g\left(\frac{x_i - \mu}{\sigma}\right), \quad \theta := (\mu, \sigma).$$

On note $(X_1, \dots, X_n)$ les variables canoniques : pour tout $i \in \{1, \dots, n\}$ et $(x_1, \dots, x_n) \in \mathbb{R}^n$, on a $X_i(x_1, \dots, x_n) = x_i$.

1. Montrer que sous $p_{n,\theta} \cdot \text{Leb}^{\otimes n}$, les statistiques $(X_1, \dots, X_n)$ sont i.i.d. et identifier leur loi.

Vocabulaire : transl. échelle (μ, σ) , c'est regarder

$$\mu + \sigma \varepsilon$$

On obtient que sous $p_{n,\theta}$, la densité est sous forme produit, donc les X_i sont i.i.d. De +, la densité pour chaque X_i est $\sigma^{-1} g(\frac{x-\mu}{\sigma})$ donc c'est la même : les (X_i) sont i.i.d.

$$x \mapsto \sigma^{-1} g\left(\frac{x-\mu}{\sigma}\right) \text{ est la densité de chaque } X_i$$

On va mettre $X_i = \mu + \sigma \varepsilon_i$, où ε_i i.i.d. $\sim g$.

Pour montrer cela, on va utiliser la méthode de la fonction nœtte.

Soit φ continue bornée, on calcule : $Z = \mu + \sigma \varepsilon$

$$\mathbb{E}[\varphi(Z)] = \mathbb{E}[\varphi(\mu + \sigma \varepsilon)]$$

$$= \int \varphi(\mu + \sigma s) g(s) ds$$

$$z = \mu + \sigma s$$

$$= \int \varphi(z) \frac{g(\frac{z-\mu}{\sigma})}{\sigma} dz$$

$$s = \frac{z-\mu}{\sigma}$$

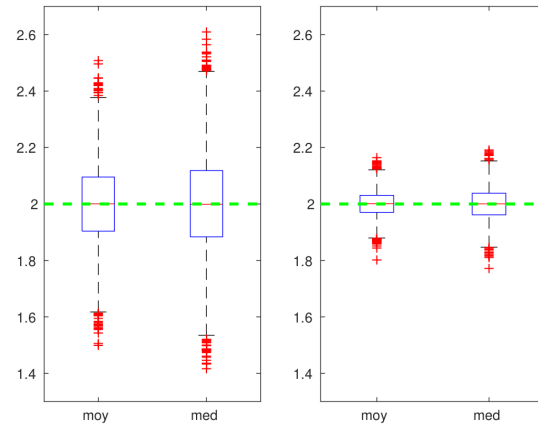
Z a donc la même densité que X .

Conclusion : (X_i) i.i.d. et $X_i = \mu + \sigma \varepsilon_i$ avec $\varepsilon_i \sim g$.

Supposons que g est une densité gaussienne centrée réduite. On définit les statistiques

$$S_n := \sum_{i=1}^n X_i, \quad K_n := \sum_{i=1}^n (X_i - n^{-1} S_n)^2.$$

3. Proposer un estimateur de μ puis de σ^2 , utilisant ces deux statistiques. *Commentaires : il n'y a pas unicité d'un estimateur de θ . Dans le cas de l'estimateur de la moyenne, la figure 1 permet de comparer deux stratégies d'estimation : l'une repose sur l'estimateur de la moyenne empirique, et l'autre sur l'estimateur de la médiane empirique. Cette figure affiche le boxplot de 5000 réalisations indépendantes de chacun des estimateurs dans le cas $n = 50$ (gauche) et le cas $n = 500$ (droite). Qu'en concluez-vous ? dans le corrigé rédigé, on trouvera les mêmes analyses graphiques pour le cas où g est une densité de Laplace ; pensez-vous que la conclusion sera la-même ?*



- $\hat{\mu}_n = \frac{S_n}{n}$ moyenne empirique. C'est 1 bon candidat pour μ !

$\mu = \mathbb{E}[X]$ " $\hat{\mu}_n$ va converger vers μ " p.s.
 LFGN - p.s. Borel Cantelli
 (même si on l'ess des exts où il y a conv.)
 ! Yukor Tchekyap → $\forall \varepsilon > 0: \lim P(|\hat{\mu}_n - \mu| > \varepsilon) = 0$
 Transf $\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{D} Z_0 \quad \forall \varphi \in \mathcal{C}_b^0: \mathbb{E} \varphi(\frac{S_n - n\mu}{\sqrt{n}}) \rightarrow \mathbb{E} \varphi(Z_0)$
 TCL

- $\frac{1}{n} \sum_{i=1}^n \left(X_i - \frac{S_n}{n} \right)^2$ variance empirique.

4. Déterminer le modèle statistique induit par les statistiques (S_n, K_n) . La démonstration du théorème de Gosset est donnée dans le polycopié de cours (Théorème IV-5-24).

Gm va chercher la loi de (S_n, K_n) . Et attention ce n'est pas facile !

$$S_n = \sum_{i=1}^n X_i \quad K_n = \sum_{i=1}^n \left(X_i - \frac{S_n}{n} \right)^2$$

X_i iid $N(\mu, \sigma^2)$

Prop : $S_n \sim N(n\mu, n\sigma^2)$

Démo : $X = (X_1, \dots, X_m)$ est $\perp \vec{VG}$

Donc $\sum_{i=1}^m X_i$ est $\perp$ v.a. Gaussienne car C. Lin de $\vec{VG}$

On identifie la loi de S_m par $\mathbb{E} S_m$ et $\text{Var}(S_m)$:

$$\mathbb{E} S_m = m \mu$$

$$\text{Var}(S_m) = \sum_{i=1}^m \text{Var}(X_i) = m \sigma^2$$



par LL

□

Au passage: cov. nulle pour les $\vec{VG} \Leftrightarrow \perp$.

$$\mathbb{E}[e^{-tX}] = e^{-t\mu + \frac{t^2}{2}\sigma^2}$$

$$X \sim N(\mu, \sigma^2)$$

$$\mathbb{E}[e^{-\langle t, X \rangle}] = e^{-\langle t, \mu \rangle + \frac{T^T \Sigma^2 t}{2}}$$

$$X \sim N(\mu, \Sigma^2)$$

Pourquoi K_m a-t-il un rapport avec un χ^2 ?

$$S_m = m\mu + \sigma \sum_{i=1}^m \varepsilon_i$$

$$X_i = \mu + \sigma \varepsilon_i$$

$$K_m = \sum_{i=1}^m \left(\mu + \sigma \varepsilon_i - \frac{S_m}{m} \right)^2$$

On se ramène au

cas des $\varepsilon_i \text{ i.i.d } N(0,1)$

$$= \sum_{i=1}^m \left(\cancel{\mu} + \sigma \varepsilon_i - \cancel{\mu} - \sigma \frac{\sum \varepsilon_i}{m} \right)^2$$

$$= \sigma^2 \sum_{i=1}^m \left(\varepsilon_i - \frac{\sum \varepsilon_i}{m} \right)^2$$

$$\varepsilon = (\varepsilon_1, \dots, \varepsilon_m) \xrightarrow{\vec{VG}} N(0, I_m)$$

$$\sum_{i=1}^m \varepsilon_i = \langle \varepsilon, \mathbb{1} \rangle$$

$$\sum_{i=1}^m \left(\varepsilon_i - \frac{\langle \varepsilon, \mathbb{1} \rangle}{m} \right)^2 \quad \text{c'est 1 norme au carré!}$$

$$\left\| \varepsilon - \frac{\langle \varepsilon, \mathbb{1} \rangle}{m} \mathbb{1} \right\|^2 \text{ très précisément!}$$

Qu'est-ce que $\varepsilon - \frac{\langle \varepsilon, \mathbb{1} \rangle}{m} \mathbb{1}$? (c'est la proj $\perp$ de ε sur $\mathbb{1}^\perp$)

$$\frac{\langle \varepsilon, \mathbb{1} \rangle}{m} \mathbb{1} ? \quad \frac{\langle \mathbb{1}, \varepsilon \rangle}{m} \mathbb{1}$$

$$\varepsilon = \frac{\langle \varepsilon, \mathbb{1} \rangle}{m} \mathbb{1} + \left(\varepsilon - \frac{\langle \varepsilon, \mathbb{1} \rangle}{m} \mathbb{1} \right)$$

$$= \frac{\mathbb{1} \langle \mathbb{1}, \varepsilon \rangle}{m} + \underline{\hspace{2cm}}$$

$$= \frac{\mathbb{1}}{m} \cdot \mathbb{1}^T \varepsilon + \underline{\hspace{2cm}}$$

$$= \underbrace{\left(\frac{\mathbb{1} \mathbb{1}^T}{m} \right)}_P \varepsilon + \underbrace{\left(\text{Id} - \frac{\mathbb{1} \mathbb{1}^T}{m} \right)}_{\text{Id} - P} \varepsilon$$

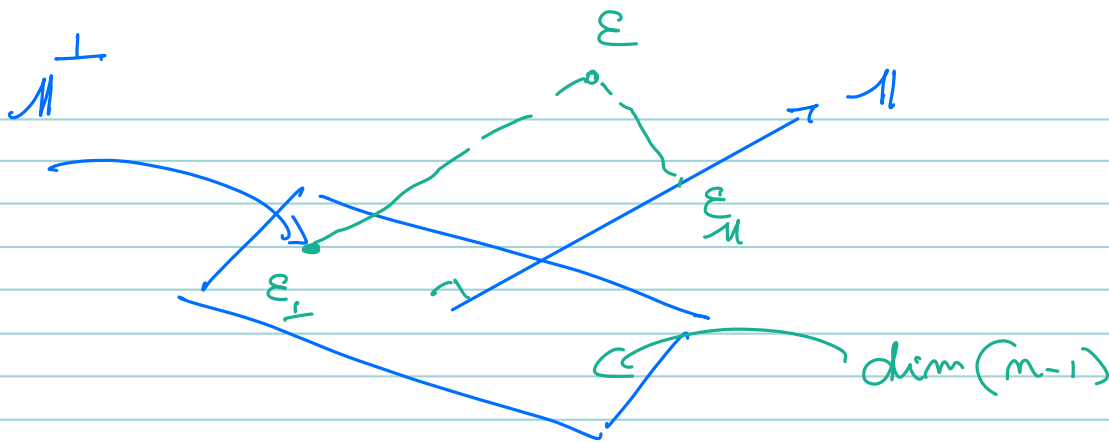
On va calculer $\text{Cov}(P\varepsilon, (\text{Id}-P)\varepsilon)$:

$$\begin{aligned} \mathbb{E} \left[P \varepsilon [(\text{Id}-P)\varepsilon]^T \right] &= \mathbb{E} \left[P \varepsilon \varepsilon^T (\text{Id}-P)^T \right] \\ &= P \underbrace{\mathbb{E}[\varepsilon \varepsilon^T]}_{\text{Id}_m} (\text{Id}-P)^T \\ &= P (\text{Id}-P) = 0 \end{aligned}$$

$P\varepsilon$ et $(\text{Id}-P)\varepsilon$ sont donc $\perp$.

On en déduit que S_m et K_m sont $\perp$. C'est 1 résultat super important pour les v.a. Gaussiennes.

Loi de K_m ?



$$K_m = \sigma^2 \| \varepsilon_{\perp} \|^2 = \sigma^2 (\tilde{\varepsilon}_1^2 + \dots + \tilde{\varepsilon}_{n-1}^2)$$

où les $\tilde{\varepsilon}_i$ sont les coord. de ε dans une base $\perp$ de $M^{\perp}$

Conclusion: $K_m \sim \chi^2(m-1) \times \sigma^2$

$$(S_m, K_m) \sim N(m\mu, m\sigma^2) \otimes \chi^2(m-1) \cdot \sigma^2$$

Cochran // Gosset

Objectif : Modèle de régression (linéaire) Gaussienne

Exercice 4 (()).** Nous cherchons à modéliser la dépendance d'une réponse par rapport à k variables explicatives, sur une population de n individus. On collecte donc

$$(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n),$$

où $y_i \in \mathbb{R}$ et $\mathbf{x}_i \in \mathbb{R}^k$ sont respectivement la *réponse* et les *variables explicatives* pour le i -ème individu. Considérons le modèle statistique suivant

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{p_{n,\theta} \cdot \text{Leb}^{\otimes n} : \theta = (\beta, \sigma) \in \Theta := \mathbb{R}^k \times \mathbb{R}_*^+\}),$$

où

$$p_{n,\theta}(y_1, \dots, y_n) := \sigma^{-n} \prod_{i=1}^n g\left(\frac{y_i - f(\mathbf{x}_i' \beta)}{\sigma}\right);$$

Le modèle de translation échelle est ici

$$Y_i = f(\langle X_i, \beta \rangle) + \sigma \varepsilon_i$$

Le modèle le + commun est $f = \text{Id}$, $\varepsilon_i \sim N(0, 1)$ et on retrouve

$$Y_i = \langle X_i, \beta \rangle + \sigma \varepsilon_i$$

3] $k=2$: la dimension des (x_i) est 2 . $X_i = \begin{pmatrix} 1 \\ x_i \end{pmatrix}$

$$\langle X_i, \beta \rangle = \beta_1 + \beta_2 x_i$$

$$Y_i = \beta_1 + \beta_2 x_i + \sigma \varepsilon_i, \quad \varepsilon_i \text{ iid } N(0,1)$$

1^{ère} idée : $F(\beta) = \sum_{i=1}^n (Y_i - (\beta_1 + \beta_2 x_i))^2$

$$\text{et } (\hat{\beta}_1, \hat{\beta}_2) = \underset{\beta \in \mathbb{R}^2}{\text{argmin}} F(\beta)$$

2^{ème} idée : β va maximiser la "ressemblance" des obs.

$$-\frac{1}{2} (X - \mu)^T \Sigma^{-1} (X - \mu)$$

$$\frac{e^{-\frac{1}{2} (X - \mu)^T \Sigma^{-1} (X - \mu)}}{(2\pi)^{d/2} \sqrt{\det(\Sigma^2)}}$$

$$L(\mu, \Sigma^2) = \dots - \frac{1}{2} \log(\det \Sigma^2)$$