M1- Mathématiques Appliquées - ENSTA / Paris-Sud
Modélisation statistique (MAP-STA1), 2018-2019
christine.keribin@math.u-psud.fr
Chargés de TD: Baptiste Broto, Timothée Mathieu, Solène Thépaud

Semaine 2: Estimateur du maximum de vraisemblance
Eléments de corrigé

1 Vraisemblance

Comparer le domaine de définition de la densité un n-échantillon iid d'une loi de Bernoulli $\mathcal{B}(1,\theta)$ et celui de sa vraisemblance; déterminer une statistique exhaustive; donner l'allure de la vraisemblance; déterminer l'estimateur du maximum de vraisemblance.

La densité de l'échantillon est une fonction de $\{0,1\}^{\otimes n}$ *dans* $[0,1]$, *tandis que la vraisemblance est une fonction de* $[0,1]$ *dans* $[0,1]$. *Les observation étant indépendantes, la vraisemblance est le produit des vraisemblances individuelles:*

$$L(\theta;x) = \prod_i f(x_i;\theta) = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i} = \theta^s(1-\theta)^{n-s}$$

avec $S=\sum_i X_i$ *qui est une statistique exhaustive par le théorème de factorisation* ($h(x)=1$ *ici). La forme de la vraisemblance dépend de la valeur de la statistique exhaustive qui prend ses valeurs dans* $s=0,\dots,n$. *La vraisemblance est représentée pour quelques valeurs de* s *figure 1. En annulant la dérivée de la log vraisemblance, on déduit* $\widehat{\theta}=S/n$, *et la dérivée seconde vaut* $-s(1-2\theta)/(\theta(1-\theta))^2<0$ *donc* $\widehat{\theta}$ *est bien l'EMV.*

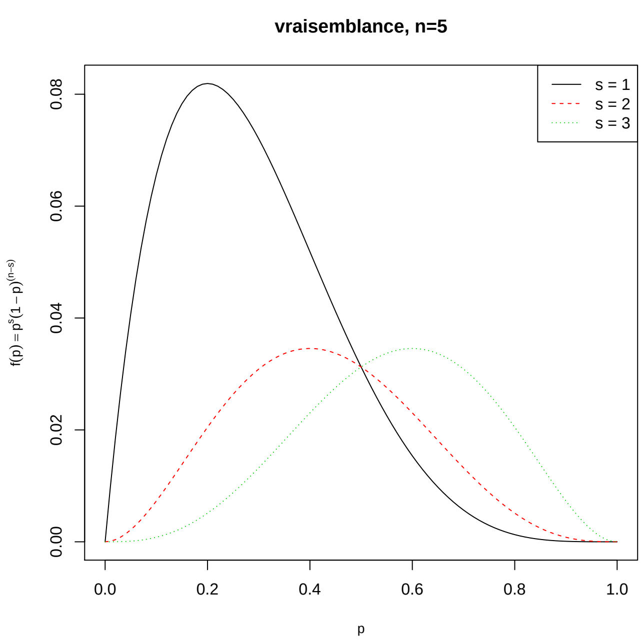

Figure 1: La fonction de vraisemblance réprésentée pour trois valeurs de la statistique exhaustive

Reprendre ces questions pour un n-échantillon de loi Uniforme $\mathcal{U}_{[0;\theta]}$ ($\theta>0$), puis de loi Gaussienne $\mathcal{N}(\mu,\sigma^2)$

Loi uniforme ($\theta > 0$): *La densité de l'échantillon par rapport à la mesure de Lebesgue est une fonction de* $\mathbb{R}^n$ *dans* $\{0, 1/\theta\}$*, tandis que la vraisemblance est une fonction de* θ*, de* $\mathbb{R}^{+*}$ *dans* $\mathbb{R}^+$*. Le support de la loi dépend du paramètre inconnu, le modèle n'est pas régulier. La densité d'un n-échantillon peut s'écrire*

$$f(x_1, \dots x_n; \theta) = \theta^{-n} \mathbb{1}_{0 \leq \min_i x_i \leq \max_i x_i \leq \theta} = \theta^{-n} \mathbb{1}_{0 \leq \max_i x_i \leq \theta}.$$

Donc $\max_i X_i$ *est une statistique exhaustive d'après le théorème de factorisation.* $\theta \mapsto L(\theta, x)$ *est une fonction nulle pour* $\theta < \max_i x_i$*. Elle vaut* $1/\theta^n$ *(et donc décroissante) pour* $\theta > max_i x_i$*. Ainsi, son maximum est en* $\max_i x_i$*, d'où l'EMV est* $\max_i X_i$*.*

Loi gaussienne: *La densité de l'échantillon est une fonction de* $\mathbb{R}^n$ *dans* $\mathbb{R}^{+*}$*, tandis que la vraisemblance est une fonction de* $\mathbb{R} \times \mathbb{R}^{+*}$ *dans* $\mathbb{R}^{+*}$*. On écrit la vraisemblance pour* $\theta = (\mu, \sigma^2)$

$$L(\theta; x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp - \left(\frac{\sum_i (x_i - \mu)^2}{2\sigma^2} \right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp - \left(\frac{\sum_i x_i^2 - 2\mu \sum_i x_i + n\mu^2}{2\sigma^2} \right)$$

et $(\sum_i X_i, \sum_i X_i^2)$ *est une statistique exhaustive par le théorème de factorisation. La fonction de vraisemblance est une gaussienne (à sigma fixé) et la forme d'une inverse Gamma à* μ *fixé. La dérivée de la vraisemblance est le vecteur du score*

$$\left(\frac{\sum_i x_i - n\mu}{\sigma^2} \;\; ; \;\; -\frac{n}{2\sigma^2} + \frac{\sum_i (x_i - \mu)^2)}{2\sigma^4} \right)$$

qui possède un unique point critique $\widehat{\theta} = \left(\bar{x}, \sum_i (x_i - \bar{x})^2/n\right)$*. La hessienne au point critique est la matrice diagonale* $diag(-n/\widehat{\sigma}^4, -1/(2\widehat{\sigma}^4))$ *qui est bien définie négative; d'où l'EMV* $\widehat{\theta} = \left(\bar{X}, \sum_i (X_i - \bar{X})^2/n\right)$*.*

2 Loi exponentielle

Soit un n-échantillon de loi exponentielle dépendant du paramètre $\mu > 0$ et définie pour $x \geq 0$ par $f(x) = \frac{1}{\mu} \exp(-\frac{x}{\mu})$.

1. Calculer l'estimateur du maximum de vraisemblance $\widehat{\mu}$. Celui-ci est-t-il sans biais? efficace ? consistant ? uniformément de variance minimum parmi les estimateurs sans biais? *Soit*

   $$L_n(\mu) = -n \log \mu - \sum_i X_i/\mu$$

   la log-vraisemblance. L'EMV annule l'équation de score $dL_n/d\mu = 0$*, soit* $\widehat{\mu} = \bar{X}$ *et on vérifie que la dérivée seconde est négative autour du maximum. L'information de Fisher vaut* $I_n(\mu) = n/\mu^2$*. L'estimateur est sans biais (*$\mathbb{E}(\widehat{\mu}) = \mu$*), sa variance* $Var(\widehat{\mu}) = \mu^2/n$ *atteint la borne de Cramer-Rao* $I_n(\mu)^{-1}$*: il est donc efficace et donc optimal (UVMB). Il est consistant car sans biais et de variance tendant vers* 0*; ou LGN.*

2. Déterminer la loi de $\widehat{\mu}$ à distance finie, puis sa loi asymptotique. Déterminer un intervalle de confiance de μ de niveau $1-\alpha$. *L'échantillon étant iid,* $T_n = \sum_i X_i/\mu \sim \mathcal{G}a(n,1)$ *(passer par les fonctions caractéristiques) qui est une statistique pivotale (sa loi ne dépend pas de* μ*). On a l'intervalle de fluctuation pour* T *de probabilité* $1-\alpha$

$$\mathbb{P}_{\mu_0}\Big(g_n(\alpha/2) \leq T_n \leq g_n(1-\alpha/2)\Big) = 1-\alpha$$

avec $g_n(\alpha/2)$ *et* $g_n(1-\alpha/2)$ *les quantiles d'ordre* $\alpha/2$ *et* $1-\alpha/2$ *d'une loi* $\mathcal{G}a(n,1)$*. D'où l'IC de niveau* $1-\alpha$

$$IC = \left[\frac{\sum_i X_i}{g_n(1-\alpha/2)}; \frac{\sum_i X_i}{g_n(\alpha/2)}\right] \text{ avec } \mathbb{P}(IC \ni \mu) = 1-\alpha$$

D'après le TLC (moyenne empirique d'un n-échantillon iid de variance finie), $T_n^* = \sqrt{n}(\bar{X}/\mu - 1) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$*. D'où*

$$\mathbb{P}_{\mu_0}\Big(|T_n^*| < q^*_{1-\alpha/2}\Big) \simeq 1-\alpha$$

où $q^*_{1-\alpha/2}$ *est le quantile d'ordre* $1-\alpha/2$ *de la loi gaussienne centrée réduite. On en déduit l'IC de niveau asymptotique* $1-\alpha$,

$$IC_{asym} = \left[\frac{\sum_i X_i}{1+q^*_{1-\alpha/2}/\sqrt{n}}; \frac{\sum_i X_i}{1-q^*_{1-\alpha/2}/\sqrt{n}}\right] \text{ avec } \mathbb{P}(IC \ni \mu) \simeq 1-\alpha$$

A distance finie, le niveau vaut approximativement $1-\alpha$.

3. On estime maintenant $\lambda = 1/\mu$.

   (a) Quelle est la borne de Cramer-Rao pour un estimateur sans biais de λ?
   Soit $\lambda = h(\mu) = 1/\mu$*. La borne de Cramér-Rao pour* λ *est*

   $$BCR(\lambda) = [\dot{h}(\mu)]^2[I_n(\mu)]^{-1} = \lambda^2/n.$$

   où $\dot{h}(\mu)$ *est la dérivée de* h *par rapport à* μ.

   (b) Montrer que l'estimateur $\widehat{\lambda} = (n-1)/\sum X_i$ est non biaisé pour l'estimation de λ. Est-il efficace? Asymptotiquement efficace?
   On a $\lambda \sum_i X_i \sim \mathcal{G}a(n,1)$*. Et, en utilisant le rappel (en fin d'énoncé de TD) de la définition de la densité* f_Γ *de la loi Gamma,*

   $$\begin{aligned}\mathbb{E}\left(\frac{1}{\lambda\sum_i X_i}\right) &= \int \frac{1}{s} f_\Gamma(s) ds = \int \frac{1}{\Gamma(n)} s^{n-2} \exp(-s)\, ds \\ &= \frac{1}{n-1}\int \frac{1}{\Gamma(n-1)} s^{n-2}\exp(-s)\, ds = \frac{1}{n-1}\end{aligned}$$

   d'où $\mathbb{E}((n-1)/\sum_i X_i = \lambda$ *et* $\widehat{\lambda}$ *est non biaisé. Sa variance vaut* $\lambda^2/(n-2) > \lambda^2/n = BCR(\lambda)$*, il n'est donc pas efficace, mais asymptotiquement efficace. On peut retrouver le résultat asymptotique en utilisant la delta-méthode*

   $$\sqrt{n}(h(\widehat{\mu}) - h(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, [\dot{h}(\mu)]^2/I_1(\mu)\right)$$

(c) Existe-t-il un estimateur efficace de λ?
Il n'existe pas d'estimateur efficace à distance finie. La loi exponentielle fait partie de la famille exponentielle de loi: sa densité se met sous la forme $\exp(a(x)\alpha(\mu) + \beta(\mu) + c(x))$ *avec* $a(x) = x$, $\alpha(\mu) = 1/\mu$, $\beta(\mu) = -\log(\mu)$, $c(x) = 0$, *et le seul paramètre (à une fonction affine près) qu'on sait estimer efficacement est* $\mathbb{E}(a(X)) = \mu$.

3 Loi de Bernoulli

Soit un n-échantillon de loi de Bernoulli $\mathcal{B}(p)$. Déterminer la borne de Cramer-Rao pour un estimateur sans biais de la cote $p/(1-p)$. Peut-on estimer efficacement la cote à distance finie? asymptotiquement?

La log-vraisemblance s'écrit $\ell_\theta(X) = \sum_i X_i \log(p) + (n - \sum_i X_i)\log(1-p)$, *de dérivée par rapport à p:* $\dot{\ell}_\theta(X) = \sum_i X_i/p - (n - \sum_i X_i)\log(1-p)$, *d'où* $\widehat{p} = \bar{X}$. *On vérifie que la dérivée seconde est négative*

$$\ddot{\ell}_\theta(X) = -\frac{\sum_i X_i}{p^2} - \frac{n - \sum_i X_i}{(1-p)^2}$$

L'information de Fisher vaut $I_n(p) = -\mathbb{E}(\ddot{\ell}_\theta(X)) = n/(p-p^2)$ *et la borne de Cramér-Rao pour un estimateur sans biais de* $h(p) = p/(1-p)$ *vaut*

$$BCR(h(p)) = [\dot{h}(p)]^2[I_n(p)]^{-1} = \frac{p}{n(1-p)^3}$$

où $\dot{h}(p) = 1/(1-p)^2$ *est la dérivée de h par rapport à p.*
La loi de Bernoulli est aussi une loi de la famille exponentielle, avec $a(x) = x$, $\alpha(p) = \log(p/(1-p))$, $\beta(p) = \log(1-p)$ *et* $c(x) = 0$. *La seule fonction de p (à une transformation affine près) qu'on sait estimer efficacement est* $\mathbb{E}(a(X)) = p$. *On ne peut estimer efficacement la cote à distance finie, mais* $\bar{X}(1-\bar{X})$ *où* $\bar{X}$ *est l'EMV de p est asymptotiquement efficace.*

4 Loi de Weibull

La durée de vie d'un matériel peut être modélisée par une variable aléatoire X qui suit une loi de Weibull $\mathcal{W}(\lambda, 1/2)$ de densité

$$f(x,\lambda) = \frac{1}{2\lambda\sqrt{x}} \exp\left(-\frac{\sqrt{x}}{\lambda}\right) \mathbb{I}_{x>0}$$

Son espérance vaut $\mathbb{E}(X) = \lambda\Gamma(3) = 2\lambda$ et sa variance $\mathrm{Var}(X) = \lambda^2\Gamma(5) - \mathbb{E}(X)^2$.

1. La loi de Weibull $\mathcal{W}(\lambda, 1/2)$ appartient-elle à la famille exponentielle? Proposer une statistique exhaustive.

$$\log(f(x,\lambda)) = -\frac{\sqrt{x}}{\lambda} - \log(2\lambda) - \frac{1}{2}\log(x) = \alpha(\lambda)a(x) + \beta(\lambda) + c(x)$$

On reconnait la forme d'une loi de la famille exponentielle avec $\alpha(\lambda) = -\frac{1}{\lambda}$, $a(x) = \sqrt{x}$, $\beta(\lambda) = -\log(2\lambda)$, $c(x) = \frac{1}{2}\log(x)$.
Par le principe de factorisation, $\sum_i a(X_i) = \sum_i \sqrt{X_i}$ *est une statistique exhaustive.*

2. Déterminer l'estimateur du maximum de vraisemblance $\widehat{\lambda}$ de λ à partir d'un n-échantillon i.i.d. $(X_1, \dots, X_n)$ de loi $\mathcal{W}(\lambda, 1/2)$.

   On dérive la logvraisemblance

$$\frac{\partial}{\partial\lambda}\log f(x,\lambda) = \sum_i \frac{\sqrt{X_i}}{\lambda^2} - \frac{n}{\lambda} = 0$$

   d'où l'EMV $\widehat{\lambda} = \sum_i \sqrt{X_i}/n$. *Le hessien vaut*

$$H_n = \frac{\partial^2}{\partial\lambda^2}\log f(x,\lambda) = -2\frac{\sum_i \sqrt{X_i}}{\lambda^3} + \frac{n}{\lambda^2}$$

   négatif autour de l'EMV, c'est un maximum. L'étude de la log-vraisemblance indique d'ailleurs que c'est l'unique maximum.

3. Déterminer la loi de $\widehat{\lambda}$. Aide: montrer que $\sqrt{X}$ suit une loi exponentielle d'espérance à déterminer, et en déduire que si $(X_1, \dots, X_n)$ est un n-échantillon i.i.d. de loi $\mathcal{W}(\lambda, 1/2)$, alors $2\sum_i \sqrt{X_i}/\lambda \sim \chi^2(2n)$.

   Pour toute fonction φ *continue et bornée, on a*

$$\mathbb{E}(\varphi(\sqrt{X})) = \int_0^\infty \varphi(\sqrt{x})\frac{1}{2\lambda\sqrt{x}}\exp(-\frac{\sqrt{x}}{\lambda})\,dx = \int_0^\infty \varphi(y)\frac{1}{\lambda}\exp(-\frac{y}{\lambda})\,dy$$

   où on reconnaît la densité d'une loi exponentielle d'espérance λ. *On en déduit* $\sqrt{X}/\lambda \sim \mathcal{E}(1)$ *(même principe). Soit* $S_n = \sum_i Z_i$ *où* $Z_i \sim \mathcal{E}(1)$.

$$\mathbb{E}(\varphi(2S_n)) = \int_0^\infty \varphi(2s)\frac{s^{n-1}}{\Gamma(n)}\exp(-\frac{s}{\lambda})\,ds = \int_0^\infty 2^-n\frac{y^{n-1}}{\Gamma(n)}\exp(-\frac{y}{2})\,dy$$

   où on reconnaît la densité d'une loi du $\chi^2(2n)$. *Donc* $\frac{2n}{\lambda}\widehat{\lambda} \sim \chi^2(2n)$

4. Calculer l'information de Fisher. $\widehat{\lambda}$ est-il efficace?

   $\widehat{\lambda}$ *est non biaisé, de variance* $Var(\widehat{\lambda}) = Var(\sqrt{X_1})/n = \lambda^2/n$. *l'information de Fisher* $I_n = \mathbb{E}(-H_n) = n/\lambda^2$. *La variance atteint bien l'inverse de l'information de Fisher. Ce qui est cohérent avec le fait que la seule fonction de* λ *(à une transformation affine près) qu'on sait estimer efficacement dans le modèle exponentiel est une fonction affine de* $\mathbb{E}(\sum_i \sqrt{X_i})/n = \lambda$.

5 Loi uniforme

Soit $X_1, \dots, X_n$ un n-échantillon de loi uniforme $\mathcal{U}_{[0;\theta]}$, où $\theta > 0$ est le paramètre inconnu.

1. Déterminer une statistique exhaustive pour estimer θ. Le modèle est-il régulier?

   Le support de la loi dépend du paramètre inconnu, le modèle n'est pas régulier. La densité d'un n-échantillon peut s'écrire $f(x_1, \dots x_n; \theta) = \theta^{-n} \mathbb{1}_{0 \le \min_i x_i \le \max_i x_i \le \theta} = \theta^{-n} \mathbb{1}_{0 \le \max_i x_i \le \theta}$. *Donc* $\max_i X_i$ *est une statistique exhaustive d'après le théorème de factorisation.*

2. Que penser des performances de l'estimateur des moments?

   L'estimateur des moments est $2\bar{X}$ *qui ne dépend pas d'une statistique exhaustive, il sera dominé.*

3. Montrer que l'estimateur du maximum de vraisemblance est $T_n = \max_i X_i$, de densité

$$h_\theta(t) = \frac{n}{\theta^n} t^{n-1} \mathbb{1}_{[0;\theta]}(t).$$

   $\theta \mapsto L(\theta, x)$ *est une fonction nulle pour* $\theta < \max_i x_i$. *Elle vaut* $1/\theta^n$ *(et donc décroissante) pour* $\theta > max_i x_i$. *Ainsi, son maximum est en* $\max_i x_i$, *doù l'EMV est* $T_n = \max_i X_i$. *En utilisant l'indépendance des observations* X_i *de loi* F*:*

$$\mathbb{P}(T_n \le x) = \mathbb{P}(\cap_{i=1}^n (X_i \le x) = \prod_{i=1}^n \mathbb{P}(X_i \le x) = F^n(x).$$

   La densité de T_n *est donc, par dérivation en* x

$$h_\theta(x) = nF^{n-1}(x) f(x) = \frac{n}{\theta^n} x^{n-1} \mathbb{1}_{[0;\theta]}(t)$$

4. Montrer que T_n est biaisé. En déduire un estimateur non biaisé $\tilde{T}_n$. $\tilde{T}_n$ est-il efficace?

   D'où

$$\mathbb{E}(T_n) = n\theta^{-n} \int_0^\theta x^n dx = \frac{n}{n+1}\theta; \quad Var(T_n) = \frac{n\theta^2}{(n+1)^2(n+2)}$$

   T_n *est donc biaisé, de biais* $-\theta/n$. *L'estimateur* $T_n^* = (n+1)T_n/n$ *est sans biais. Mais on ne peut calculer de borne de Cramer-Rao (ni parler d'efficacité) puisque le modèle n'est pas régulier. Le calcul montre que son risque est inférieur à celui de* T_n *(il est même UVMB). Mais il n'est pas admissible parce que* $\tilde{T}_n = (n+2)T_n/(n+1)$, *bien que légèrement biaisé, est de risque inférieur.*

5. Montrer que T_n est consistant, puis la convergence suivante: $n(1 - T_n/\theta) \xrightarrow{\mathcal{L}} \mathcal{E}(1)$.

   Consistance: soit F_T *la fonction de répartition de* T_n. *pour tout* $\varepsilon > 0$, *on a*

$$\mathbb{P}\Big(|T_n - \theta| > 0\Big) = \mathbb{P}\Big(\max_i X_i < \theta - \varepsilon\Big) = F_T(\theta - \varepsilon) = \left(\frac{\theta - \varepsilon}{\theta}\right)^n$$

christine.keribin@math.u-psud.fr
Chargés de TD: Baptiste Broto, Timothée Mathieu, Solène Thépaud

expression qui tend vers 0 *quand* $n \to \infty$
La convergence en loi se démontre par un théorème de convergence dominée: pour toute fonction g *continue et bornée, on a*

$$\begin{aligned} \mathbb{E}\Big(g(n(1 - T_n/\theta))\Big) &= \int_0^\theta g(n(\theta - t))\frac{n}{\theta^n}t^{n-1}dt \\ &= \int_{\mathbb{R}^+} g(u)f_n(u)du \end{aligned}$$

avec $f_n(u) = \left(1 - \frac{u}{n}\right)^{n-1} \mathbb{1}_{[0;n]}(u)$. *Or,* $f_n(u)$ *tend ponctuellement vers* $\exp(-u)$ *pour tout* u *et comme* $\log f_n(u) \leq -u/2$ *dès que* $n \geq 2$, *on a* $|g(u)f_n(u)| \leq |g|_\infty \exp(-u/2)$ *d'intégrale finie. On en déduit que*

$$\mathbb{E}\Big(g(n(1 - T_n/\theta))\Big) \to \int g(u)e^{-u}\mathbb{1}_{\mathbb{R}^+}(u)du$$

La loi limite de $n(1 - T_n/\theta)$ *est donc une loi exponentielle d'espérance* 1.

Note: La densité d'une loi du Khi-deux $\chi^2(k)$ à k degrés de liberté est définie pour $x \geq 0$ par $f_\chi(x) = \frac{2^{-k/2}}{\Gamma(k/2)}x^{k/2-1}\exp(-x/2)$ La densité d'une loi Gamma $\mathcal{G}a(n, \theta)$ est définie pour $x \geq 0$ par $f_\Gamma(x) = \frac{\theta^{-n}}{\Gamma(n)}x^{n-1}\exp(-x/\theta)$