Correction du TD 4 : Estimation Ponctuelle

Exercice 1 : Loi usuelles

Dans chacun des modèles suivants, dire quel est le paramètre, l'ensemble de ses valeurs et écrire sa vraisemblance (attention, "le" paramètre peut être de dimension supérieure à 1).

1. **N-échantillon d'une loi de Bernoulli :** X_1, X_2, ..., $X_N \approx$ *i.i.d.* $B(1,p)$.
   $E(X) = p \qquad V(X) = p(1-p)$
   $l(p; x_1, x_2, ..., x_N) = \prod_{i=1}^{N} p^{x_i}(1-p)^{1-x_i} = p^{\Sigma x_i}(1-p)^{N-\Sigma x_i}$
   $lnl(p; x_1, x_2, ..., x_N) = L(p; x_1, x_2, ..., x_N) = \sum_i x_i \ln(p) + \left(N - \sum_i x_i\right)\ln(1-p)$
2. **N-échantillon d'une loi de Poisson :** X_1, X_2, ..., $X_N \approx$ *i.i.d.* $P(\lambda)$.
   $E(X) = \lambda \qquad V(X) = \lambda$
   $l(\lambda; x_1, x_2, ..., x_N) = \prod_{i=1}^{N} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = \frac{1}{\prod x_i!}\lambda^{\Sigma x_i}e^{-n\lambda}$
   $lnl(\lambda; x_1, x_2, ..., x_N) = L(\lambda; x_1, x_2, ..., x_N) = \sum_i x_i \ln(\lambda) - N\lambda - \sum_i \ln(x_i!)$
3. **N-échantillon d'une loi de Exponentielle :** X_1, X_2, ..., $X_N \approx$ *i.i.d.* $\Gamma(1,\theta)$.
   $E(X) = \theta \qquad V(X) = \theta^2$
   $l(\theta; x_1, x_2, ..., x_N) = \prod_{i=1}^{N} \frac{1}{\theta}e^{-x_i/\theta} = \frac{1}{\theta^N}\exp\left(-\frac{1}{\theta}\sum_i x_i\right)$
   $lnl(\theta; x_1, x_2, ..., x_N) = L(\theta; x_1, x_2, ..., x_N) = -N\ln(\theta) - \frac{1}{\theta}\sum_i x_i$
4. **N-échantillon d'une loi de Normale :** X_1, X_2, ..., $X_N \approx$ *i.i.d.* $N(m,\sigma^2)$.
   $E(X) = m \qquad V(X) = \sigma^2$
   $l(m,\sigma^2; x_1, x_2, ..., x_N) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(x_i - m)^2\right] = (2\pi\sigma^2)^{-\frac{N}{2}}\exp\left[-\frac{1}{2\sigma^2}\sum(x_i - m)^2\right]$
   $lnl(m,\sigma^2; x_1, x_2, ..., x_N) = L(m,\sigma^2; x_1, x_2, ..., x_N) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_i (x_i - m)^2$
5. **N-échantillon d'une loi de Uniforme :** X_1, X_2, ..., $X_N \approx$ *i.i.d.* $U[0,\theta]$.
   $E(X) = \frac{\theta}{2} \qquad V(X) = \frac{\theta^2}{12}$
   $l(\theta; x_1, x_2, ..., x_N) = \prod_{i=1}^{N} \frac{1}{\theta}1_{0 \le x_i \le \theta} = \frac{1}{\theta^N}$ si pour tout i $0 \le x_i \le \theta$, et 0 sinon
   $lnl(x_1, x_2, ..., x_N; \theta) = L(x_1, x_2, ..., x_N; \theta) = -N\ln(\theta)$ si pour tout i $0 \le x_i \le \theta$, et $-\infty$ sinon.

Exercice 2 : Estimateur du Maximum de vraisemblance

On dispose d'un échantillon (X_1, X_2, ..., X_N) de taille N d'une variable suivant une loi exponentielle de paramètre a (réel strictement positif), de densité :

$$f(x,a) = \begin{cases} \frac{1}{a}\, e^{-x/a} & \textbf{si } x > 0 \\ 0 & \textbf{sinon} \end{cases}$$

On rappelle que $E(X) = a$ et $V(X) = a^2$. (voir notes de cours sur les lois usuelles)

1. **Vérifier par le calcul que l'estimateur du maximum de vraisemblance de a est $\overline{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ (moyenne empirique des N valeurs prises par X).**

   $X_1, X_2, ..., X_N \approx i.i.d.\gamma(1; a)$, donc la vraisemblance de a est :
   $l(a; x_1, x_2, ..., x_N) = \frac{1}{a}e^{-\frac{x_i}{a}} = \frac{1}{a^N}\exp\left(-\frac{\sum_i x_i}{a}\right)$
   $L(a; x_1, x_2, ..., x_N;) = lnl(a, x_1, x_2, ..., x_N) = -N\ln(a) - \frac{\sum_i x_i}{a}$
   L'estimateur du maximum de vraisemblance $\hat{a}$ réalise le maximum en a de $L(a; x_1, x_2, ..., x_N)$. La fonction objectif est deux fois continument dérivable pour $a > 0$: nous pouvons appliquer les résultats classiques d'optimisation.

   $$\begin{array}{rcl} \text{La condition nécessaire d'ordre1 } (CN1) \text{ est } \dfrac{\partial}{\partial a}L(a; x_1, x_2, ..., x_N) & = & 0 \text{ en } a = \widehat{a} \\ \text{Une condition suffisante d'ordre2 } (CS2) \text{ est } \dfrac{\partial^2}{\partial a^2}L(a; x_1, x_2, ..., x_N) & < & 0 \text{ en } a = \widehat{a} \end{array}$$

   CN1 :

   $$\begin{array}{rcl} \dfrac{dL}{da} & = & -\dfrac{N}{a} + \dfrac{\sum_i x_i}{a^2} = 0 \\ \text{solution} & : & \hat{a} = \dfrac{\sum_i x_i}{N} \end{array}$$

   et CS2 :

   $$\frac{d^2L}{da^2} = \frac{N}{a^2} - 2\frac{\sum_i x_i}{a^3}$$

   Pour calculer la valeur prise par la dérivée seconde en $a = \widehat{a}$, nous remplaçons a par $\widehat{a}$ et $\sum_i x_i$ par $N\widehat{a}$. Nous obtenons

   $$\frac{d^2L}{da^2}\left(\hat{a}; x_1, x_2, ..., x_N\right) = \frac{N}{\hat{a}^2} - 2\frac{N\hat{a}}{\hat{a}^3} = -\frac{N}{\hat{a}^2} < 0$$

   La solution trouvée correspond bien à un maximum.
   L'estimateur du maximum de vraisemblance est donc $\hat{a}$, moyenne arithmétique des x_i.

2. **Montrer que $\overline{X}_N$ est un estimateur sans biais de a.**

   Dans un échantillon, les N variables sont identiquement distribuées : la moyenne empirique a pour espérance :
   $E\left[\frac{1}{N}\sum_i X_i\right] = \frac{1}{N}\sum_i E[X_i] = \frac{1}{N}NE[X] = E[X]$
   Ici, $E(X) = a$, on obtient bien que $E\left[\hat{a}\right] = a$: l'estimateur est sans biais.

3. *****(facultatif) Montrer que $\overline{X}_N$ est un estimateur efficace de a.**

   Dans un échantillon, les N variables sont indépendantes : la moyenne empirique a pour variance,
   $V\left[\frac{1}{N}\sum_i X_i\right] = \frac{1}{N^2}\sum_i V\left[X_i\right]$
   et comme les variances sont toutes égales :
   $V\left[\frac{1}{N}\sum_i X_i\right] = \frac{1}{N^2}NV\left[X\right] = \frac{V[X]}{N}$
   Ici, $V\left(X\right) = a^2$, on obtient $V\left[\hat{a}\right] = \frac{a^2}{N}$.
   Les hypothèses de l'inégalité de Cramer-Rao sont satisfaites (en particulier, le support de la densité ne dépend pas du paramètre). On sait donc, que pour tout estimateur $\widetilde{a}$ sans biais : $V\left[\widetilde{a}\right] \geq \frac{1}{I(a)}$
   où $I\left(a\right)$ est la quantité d'information définie par :

$$I\left(a\right) = E\left\{\left[\frac{\partial}{\partial a}L\left(a; X_1, X_2, ..., X_N\right)\right]^2\right\} = E\left[-\frac{\partial^2}{\partial a^2}L\left(a; X_1, X_2, ..., X_N\right)\right]$$

   Ici, $\frac{\partial^2}{\partial a^2}L\left(a; X_1, X_2, ..., X_N\right) = \frac{N}{a^2} - 2\frac{\sum_i X_i}{a^3}$ et $E\left[\sum_i X_i\right] = Na$. On obtient facilement $I\left(a\right) = -\frac{N}{a^2} + 2\frac{Na_i}{a^3} = \frac{N}{a^2}$
   On constate que $V\left[\hat{a}\right] = \frac{1}{I(a)}$: le minimum de Cramer Rao est atteint pour cet estimateur, qui est donc efficace.

Exercice 3 : Estimateur du Maximum de vraisemblance

Soit $(Y_1, ...Y_n)$ un échantillon de taille n provenant d'une variable aléatoire de densité $f(y;\theta)$. Donnez l'estimateur du maximum de vraisemblance lorsque

$$f(y;\theta) = \theta y^{\theta-1} \qquad \text{avec} \quad 0 < y < 1$$

Le paramètre du modèle est θ, car lorsque θ est connu, le modèle est entièrement connu.
La vraisemblance du paramètre s'écrit $l(\theta)$, et sa log-vraisemblance $L(\theta) = ln\ l(\theta)$:
$l(\theta) = \prod_{i=1}^{n} f(y_i;\theta) = \prod_{i=1}^{n} \theta y_i^{\theta-1} = \theta^n \prod_{i=1}^{n} y_i^{\theta-1}$
$L(\theta) = ln(l(\theta)) = n\ ln(\theta) + (\theta - 1)\sum_{i=1}^{n} ln(y_i)$
L'estimateur du maximum de vraisemblance réalise le maximum en θ de $L(\theta)$:
La condition nécessaire d'ordre 1 (CN1) est $\frac{\partial L(\theta)}{\partial\theta} = 0$ en $\theta = \widehat{\theta}$
Une condition suffisante d'ordre 2 (CS2) est $\frac{\partial^2 L(\theta)}{\partial\theta^2} < 0$ en $\theta = \widehat{\theta}$
(CN1) : $\frac{\partial L(\theta)}{\partial\theta} = 0 \quad \Leftrightarrow \quad \frac{n}{\theta} + \sum_{i=1}^{n} ln(y_i) = 0$
$\widehat{\theta} = -\frac{n}{\sum_{i=1}^{n} ln(y_i)}$

Exercice 4 : Méthode des moments

Le gérant d'un supermarché désire modéliser le comportement des acheteurs de couches O'Sek, une société irlandaise qui exporte dans toute l'Europe. Il pense que la fréquence d'achat depend (entre autres) du nombre de paquets achetés à chaque passage. Plus précisément, il suppose que le nombre journalier d'acheteurs d'un seul paquet est une variable aléatoire X suivant une loi de Poisson P (λ_1), alors que le nombre journalier d'acheteurs de deux paquets est une variable aléatoire Z suivant une loi de Poisson P (λ_2). Nous supposons (pour simplifier) qu'il n'y a pas d'acheteur de plus de deux paquets à la fois, et que X et Z sont indépendantes en probabilité. La seule observation disponible chaque jour est le nombre total de paquets vendus : $Y = X + 2Z$. Les observations (i.i.d.) de N jours conduisent à l'échantillon $(Y_1, Y_2, ..., Y_N)$ de la loi de Y.

1. **Préciser le modèle statistique ainsi postulé.**
   Les observations sont $(Y_1, Y_2, ...Y_N) \approx i.i.d.Loi£$.
   Noter que cette loi n'est pas une loi de Poisson. La vraisemblance de (λ_1, λ_2) est calculable, mais pas vraiment très simple à manipuler (il n'est pas question de faire ce calcul en TD).

2. **Nous notons respectivement m et V l'espérance mathématique et la variance de Y. Calculer m et V en fonction de λ_1 et λ_2 et en déduire l'expression de λ_1 et λ_2 en fonction de m et V.**
   $E[Y] = E[X + 2Z] = E[X] + 2E[Z] = \lambda_1 + 2\lambda_2$
   $V[Y] = V[X + 2Z] = V[X] + 4V[Z] = \lambda_1 + 4\lambda_2$. (indépendance de X et Z)
   Nous avons donc les relations $\begin{cases} m = \lambda_1 + 2\lambda_2 \\ V = \lambda_1 + 4\lambda_2 \end{cases}$
   La résolution est immédiate : $\begin{cases} \lambda_1 = 2m - V \\ \lambda_2 = \frac{V-m}{2} \end{cases}$

3. **En déduire des estimateurs convergents de λ_1 et de λ_2. Quel défaut ces estimateurs peuvent-ils présenter pour des échantillons de petite taille ? Quel est leur avantage par rapport aux estimateurs du maximum de vraisemblance ?**

   On estime m et V par les moments empiriques : $\widehat{m} = \frac{1}{N}\sum_i Y_i \qquad \widehat{V} = \frac{1}{N}\sum_i (Y_i - \widehat{m})^2$

   On en déduit les estimateurs : $\begin{cases} \widehat{\lambda_1} = 2\widehat{m} - \widehat{V} \\ \widehat{\lambda_2} = \frac{\widehat{V}-\widehat{m}}{2} \end{cases}$
   La convergence de $\widehat{m}$ vers m et de $\widehat{V}$ vers V assure la convergence de $\widehat{\lambda_1}$ et $\widehat{\lambda_2}$.
   Le problème est que, pour N fini, rien n'assure que les valeurs trouvées soient positives : dans ce cas l'estimateur n'existe pas. On peut seulement dire que lorsque N grandit, la probabilité que l'estimateur existe tend vers 1.
   L'avantage des estimateurs ainsi définis est qu'ils sont faciles à calculer, alors que l'estimation par le maximum de vraisemblance est très difficile à mettre en oeuvre (surtout parceque cette fonction admet de nombreux maxima locaux). De toutes façons, ce sont des valeurs préliminaires à utiliser pour la résolution numérique du programme d'optimisation.

Exercice 5 : Méthode des moments

Donnez un estimateur de θ suivant la méthode des moments pour

$$f(y;\theta) = (\theta + 1)y^{-\theta-2} \qquad avec \;\; 0 < y < 1$$

$$E(\overline{Y}) = \int_0^1 f(y;\theta)\ y\ dy = \int_0^1 (\theta+1) y^{-\theta-2}\ y\ dy = \int_0^1 (\theta+1) y^{-\theta-1}\ dy = (\theta+1) \int_0^1 y^{-\theta-1}\ dy = (\theta+1)[\frac{x_i^{-\theta}}{-\theta}]_0^1$$

$E(\overline{Y}) = -\frac{\theta+1}{\theta}$. Remarque : $E(\overline{Y}) \geq 0$ ssi $\theta \leq -1$.

Conclusion : L'estimateur de θ selon la méthode des moments est : $\overset{\wedge}{\theta} = \frac{-1}{E(\overline{Y})+1}$