CORRIGE DES EXERCICES du mercredi 9 avril 2003 ***Série 19***

Ex. 1 **a)** La fonction de fréquence d'une loi géométrique de paramètre p est donnée par

$$f(k) = (1-p)^{k-1}p = \exp\{\log p + (k-1)\log(1-p)\}\,,$$

pour $k \geq 1$. En posant $A(p) = \log(1-p)$, $B(p) = \log p$ et $T(k) = k-1$, on obtient bien une densité de la forme d'une famille exponentielle.
La densité conjointe s'écrit

$$f(y_1, \dots, y_n) = p^n \prod_{i=1}^{n} (1-p)^{y_i - 1} = p^n (1-p)^{\sum y_i - n}\,.$$

Ainsi, $S = \sum_{i=1}^n Y_i$ est une statistique exhaustive.

b) La somme $S = \sum_{i=1}^n Y_i$ suit une loi binomiale négative de paramètres n et p (*cf.* exercice 3, série 10). C'est-à-dire que

$$P(S = k) = \binom{k-1}{n-1}(1-p)^{k-n}p^k\,, \quad k \geq n\,.$$

Puisque

$$\mathrm{E}(g(S)) = \sum_{k=n}^{\infty} g(k)P(S=k) = 0\,, \quad \forall\, p \quad \Longleftrightarrow \quad g(k) \equiv 0\,,$$

S est complète.

Ex. 2 **a)** Puisque $Y_1 \sim \mathcal{N}(\mu, 1) \Rightarrow Y_1 - \mu \sim \mathcal{N}(0,1)$, on a que

$$\mathrm{E}(T(Y_1, \dots, Y_n)) = \mathrm{E}(I_{Y_1 \leq 0}) = P(Y_1 \leq 0) = P(Y_1 - \mu \leq -\mu) = \Phi(-\mu)\,.$$

b) La densité du vecteur $(Y_1, \dots, Y_n)$ s'écrit

$$f_{Y_1,\dots,Y_n}(y_1, \dots, y_n) = C \cdot h(y_1, \dots, y_n) \cdot e^{\mu \sum y_i - \frac{n\mu^2}{2}}\,.$$

Donc, $S = \sum Y_i$ est une statistique exhaustive pour μ. Comme $\theta = \Phi(-\mu)$, S est également une statistique exhaustive pour θ.
Pour démontrer que S est complète, il suffit de montrer que S est minimale exhaustive (puisque l'on est dans le cas d'une famille exponentielle). Soient $(Y_1, \dots, Y_n)$ et $(Z_1, \dots, Z_n)$ deux jeux de données. Le rapport $f_{(Y_1,\dots,Y_n)}(y_1, \dots, y_n)/f_{(Z_1,\dots,Z_n)}(z_1, \dots, z_n)$ est constant par rapport à μ (et donc par rapport à θ) si et seulement si $\sum Y_i = \sum Z_i$. Il s'ensuit que $S = \sum Y_i$ est minimale exhaustive.

c) Notez que $Y_1 - \overline{Y} \sim \mathcal{N}(0, \frac{n-1}{n})$.
D'après le théorème 2.5 du polycopié, $\mathrm{E}(T|S)$ est l'estimateur sans biais de variance minimale de θ. On calcule

$$\begin{aligned}
\mathrm{E}(I_{Y_1 \leq 0}|S = s) &= P(Y_1 \leq 0 | S = s) \\
&= P(Y_1 - \overline{Y} \leq -s/n | \overline{Y} = s/n) \\
&= P(Y_1 - \overline{Y} \leq -s/n) \\
&= P(\frac{\sqrt{n}}{\sqrt{n-1}}(Y_1 - \overline{Y}) \leq -\frac{s\sqrt{n}}{n\sqrt{n-1}}) \\
&= \Phi(-\frac{s\sqrt{n}}{n\sqrt{n-1}})\,,
\end{aligned}$$

et donc $\Phi(-\frac{\sqrt{n}}{\sqrt{n-1}}\overline{Y})$ est l'estimateur cherché.

Ex. 3 **a)** La fonction de sensibilité cherchée est une parabole :

$$\begin{aligned} SF_{\hat\sigma^2}(y) &= (n+1)\left[\frac{1}{n+1}(y^2+\sum_{i=1}^n y_i^2)-\frac{1}{n}\sum_{i=1}^n y_i^2\right] \\ &= y^2-\frac{1}{n}\sum_{i=1}^n y_i^2\,. \end{aligned}$$

b) Notons tout d'abord que $\mathrm{Var}(\hat\sigma^2)=\mathrm{cste}\cdot\frac{1}{n}$. Il faut donc multiplier $\hat\sigma^2$ par $\sqrt{n}$ pour avoir une variance asymptotique qui ne dépende pas de n :

$$\begin{aligned} P(\sqrt{n}(\hat\sigma^2-\sigma^2)\le t) &= P(\sqrt{n}(\frac{1}{n}\sum\underbrace{(Y_i^2-\sigma^2)}_{Z_i})\le t)=P(\sqrt{n}\cdot\overline{Z}\le t)\,. \\ \mathrm{Var}(\overline{Z}) &= \frac{1}{n}\mathrm{Var}(Z_i)\,. \\ \mathrm{Var}(Z_i) &= \mathrm{Var}(Y_i^2-\sigma^2)=\mathrm{Var}(Y_i^2)=E(Y_i^4)-\underbrace{E^2(Y_i^2)}_{=\sigma^4}\,. \\ E(Y_i^4) &= \int y^4\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{y^2}{2\sigma^2}\}dy=\int\underbrace{y^3}_{=v}\underbrace{y\frac{1}{\sqrt{2\pi\sigma}}\exp\{-\frac{y^2}{2\sigma^2}\}\,dy}_{=u'} \\ &= \underbrace{-\sigma^2y^3\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{y^2}{2\sigma^2}\}\Big|_{-\infty}^{\infty}}_{=0}+3\sigma^2\underbrace{\int y^2\frac{1}{\sqrt{2\pi\sigma}}\exp\{-\frac{y^2}{2\sigma^2}\}dy}_{=\sigma^2}=3\sigma^4\,. \\ \Rightarrow\mathrm{Var}(Z_i) &= 2\sigma^4 \text{ et } \mathrm{Var}(\sqrt{n}\cdot\overline{Z})=2\sigma^4\,. \end{aligned}$$

$$\Rightarrow\sqrt{n}(\hat\sigma^2-\sigma^2)\sim\mathcal{N}(0,2\sigma^4)\,.$$

Ex. 4 La densité conjointe est telle que

$$f(y_1,\ldots,y_n\,;m,\pi)\propto\pi^{\sum y_i}(1-\pi)^{mn-\sum y_i}\,.$$

Ainsi, en posant $S=\sum_{i=1}^n Y_i$, on obtient une statistique exhaustive. Puisque S suit une loi binomiale de paramètres mn et π, c'est une famille exponentielle et puisque que S est minimale exhaustive, elle est donc forcément complète.
Le fait que $\mathrm{E}[S(S-1)\cdots(S-k+1)]=mn(mn-1)\cdots(mn-k+1)\pi^k$ s'obtient en considérant la fonction génératrice des probabilités associée à S

$$G_S(t)=\mathrm{E}(t^S)=(1-\pi+\pi t)^{mn}\,.$$

En utilisant la propriété que

$$G_s^{(k)}(t)\Big|_{t=1}=\mathrm{E}(S(S-1)\cdots(S-k+1))\,,$$

on obtient l'égalité désirée. En particulier, $\mathrm{E}[S(S-1)]=mn(mn-1)\pi^2$. Ainsi,

$$T=\frac{S}{mn}\Big(1-\frac{S-1}{mn-1}\Big)$$

est un estimateur non biaisé pour $\psi(\pi)=\pi(1-\pi)$. Puisque la statistique S est exhaustive et complète, $W=\mathrm{E}(T|S)$ est l'unique estimateur non biaisé de variance minimale. De plus, puisque $T=T(S)$ est non biaisé, on a directement que T est l'estimateur non biaisé de variance minimale.

On peut le vérifier par calcul.

$$\begin{aligned}
W &= \mathrm{E}(\alpha S(1-\beta(S-1))|S) \\
&= \alpha \mathrm{E}(S|S) - \alpha\beta \mathrm{E}(S(S-1)|S) \\
&= \alpha S - \alpha\beta S(S-1) \\
&= \alpha S(1-\beta(S-1)) \\
&= T
\end{aligned}$$

où $\alpha = (mn)^{-1}$ et $\beta = (mn-1)^{-1}$.

Ex. 5 La fonction de vraisemblance vaut

$$V(\theta\,; y_1, \ldots, y_n) = \theta^n \Big(\prod_{i=1}^{n} y_i\Big)^{\theta-1}.$$

L'estimateur du maximum de vraisemblance pour θ est donc égal à

$$\hat{\theta}_{\mathrm{MV}} = -\frac{n}{\sum \log Y_i}.$$

L'équation (2.29) du polycopié nous apprend que la loi limite de l'estimateur du maximum de vraisemblance est la loi normale

$$\mathcal{N}(\theta, \frac{1}{nI(\theta)}),$$

où $I(\theta)$ est l'information de Fisher associée à la densité $f_\theta(y)$. Une estimation de l'écart-type de l'estimateur du maximum de vraisemblance est donc

$$\hat{\sigma}_{\mathrm{MV}} = \Big(nI(\theta)\Big)^{-\frac{1}{2}} = \frac{\theta}{\sqrt{n}}.$$