

Maîtrises Math-Mass 2000-2001
Corrigé du partiel du 7 décembre 2000

1. (a) La densité de la loi de X se met sous la forme

$$f_{\theta}(x) = \exp(Q(\theta)T(x) - \phi(\theta))h(x)$$

avec $Q(\theta) = \ln(1 - \theta)$, $T(x) = x$, $\phi(\theta) = -\ln \theta$ et $h(x) = 1$. Il s'agit donc d'un modèle exponentiel. L'espace canonique est l'ensemble des paramètres η tels que

$$\sum_{x=0}^{+\infty} \exp(\eta T(x))h(x) < +\infty.$$

Or il s'agit d'une série géométrique dont la somme est bien connue :

$$\begin{aligned} \sum_{x=0}^{+\infty} \exp(\eta T(x))h(x) &= \sum_{x=0}^{+\infty} (\exp \eta)^x \\ &= \frac{1}{1 - \exp \eta} \text{ si } \exp \eta \in]0, 1[\\ &= +\infty \text{ si } \exp \eta \geq 1 \end{aligned}$$

L'espace canonique des paramètres est donc $\{\eta \in \mathbb{R} / \exp \eta \in]0, 1[\} = \mathbb{R}^*$. C'est un pavé ouvert de \mathbb{R} . D'autre part, il n'y a pas de relation affine entre les différents paramètres canoniques, puisqu'il n'y en a qu'un. Même chose pour $T(x)$. Le modèle est donc de rang plein.

(b)

$$\begin{aligned} \mathbb{E}_{\theta}[z^X] &= \sum_{x=0}^{+\infty} z^x \theta (1 - \theta)^x \\ &= \theta \sum_{x=0}^{+\infty} (z(1 - \theta))^x \\ &= \frac{\theta}{1 - z(1 - \theta)} \end{aligned}$$

car $|z(1 - \theta)| < 1$ sous les hypothèses de l'énoncé. Il s'agit d'une série entière de rayon de convergence $(1 - \theta)^{-1}$, qui est C^{∞} en 1. D'où

$$\begin{aligned} \mathbb{E}_{\theta}[X] &= \frac{d}{dz} \mathbb{E}_{\theta}[z^X] \Big|_{z=1} \\ &= \frac{\theta(1 - \theta)}{(1 - z(1 - \theta))^2} \Big|_{z=1} \\ &= \frac{1 - \theta}{\theta} \\ \mathbb{E}_{\theta}[X(X - 1)] &= \frac{d^2}{dz^2} \mathbb{E}_{\theta}[z^X] \Big|_{z=1} \\ &= \frac{2\theta(1 - \theta)^2}{(1 - z(1 - \theta))^3} \Big|_{z=1} \\ &= 2 \left(\frac{1 - \theta}{\theta} \right)^2 \\ \text{var}_{\theta}(X) &= \mathbb{E}_{\theta}[X(X - 1)] + \mathbb{E}_{\theta}[X] - \mathbb{E}_{\theta}[X]^2 \\ &= \frac{1 - \theta}{\theta^2} \end{aligned}$$

- (c) D'après la première question, le modèle est de rang plein, et $Q(\theta) = \ln(1 - \theta)$ est une fonction continuellement dérivable sur $]0, 1[$, dont la dérivée ne s'annule pas. Le modèle est donc régulier, ce qui signifie en particulier que l'information de Fisher est bien définie.

$$\begin{aligned} I_1(\theta) &= \text{var}_\theta \left(\frac{\partial}{\partial \theta} \ln f_\theta(X) \right) \\ &= \text{var}_\theta \left(\frac{1}{\theta} + \frac{X-1}{1-\theta} \right) \\ &= \frac{1}{(1-\theta)^2} \text{var}_\theta(X) \\ &= \frac{1}{\theta^2(1-\theta)} \end{aligned}$$

2. (a) Notons encore f_θ la densité de la loi de (X_1, \dots, X_n) par rapport à la mesure de comptage sur \mathbb{N}^n .

$$\begin{aligned} f_\theta(x_1, \dots, x_n) &= f_\theta(x_1) \dots f_\theta(x_n) \text{ car } X_1, \dots, X_n \text{ sont indépendants} \\ &= \exp(Q(\theta)T(x_1, \dots, x_n) + n \ln \theta) \end{aligned}$$

avec $T(x_1, \dots, x_n) = x_1 + \dots + x_n$. Comme à la première question, il s'agit d'un modèle exponentiel de rang plein, l'espace canonique des paramètres étant inchangé. On en déduit que $T_n = T(X_1, \dots, X_n)$ est une statistique exhaustive, complète et minimale. Les définitions générales de ces termes ont été vues en cours. Dans le cas présent, cela signifie respectivement (attention aux quantificateurs !)

- i. quel que soit $A \in \mathcal{A}^{\otimes n}$, il existe une version de $\mathbb{P}_\theta^{\otimes n}(A|T_n)$ indépendante de θ ;
- ii. quelle que soit f application de \mathbb{N} dans \mathbb{R} , on a l'implication $(\forall \theta \in]0, 1[, \mathbb{E}_\theta[f(T_n)] = 0) \implies (f = 0)$;
- iii. quelle que soit S exhaustive, il existe f borélienne telle que $T_n = f(S)$.

Comme $Q(\theta)$ est continuellement dérivable sur son ensemble de définition, de dérivée ne s'annulant pas, on en déduit aussi que T_n est une statistique régulière, et que c'est un estimateur UVMB et efficace de son espérance. Or $\mathbb{E}_\theta[T_n] = \mathbb{E}_\theta[X_1] + \dots + \mathbb{E}_\theta[X_n] = n\mathbb{E}_\theta[X] = n(\frac{1}{\theta} - 1)$.

- (b) L'information de Fisher $I(\theta)$ associée à un n -échantillon de variables aléatoires (indépendantes) est égale à n fois l'information de Fisher $I_1(\theta)$ associée à l'une d'entre elles. On a donc

$$I(\theta) = nI_1(\theta) = \frac{n}{\theta^2(1-\theta)}$$

- (c) Quel que soit $z \in]-\frac{1}{1-\theta}, \frac{1}{1-\theta}[$, on a

$$\begin{aligned} \mathbb{E}_\theta[z^X]^n &= \mathbb{E}_\theta[z^{X_1}] \dots \mathbb{E}_\theta[z^{X_n}] \\ &= \mathbb{E}_\theta[z^{X_1} \dots z^{X_n}] \text{ car les } X_i \text{ sont indépendants} \\ &= \mathbb{E}_\theta[z^{X_1 + \dots + X_n}] = \mathbb{E}_\theta[z^{T_n}] \\ &= \sum_{k=0}^{+\infty} z^k \mathbb{P}_\theta^{\otimes n}(T_n = k) \quad (*) \end{aligned}$$

Or, pour tout $u \in]-1, 1[$,

$$\begin{aligned} \frac{1}{(1-u)^n} &= \frac{1}{(n-1)!} \frac{d}{du^{n-1}} \left(\frac{1}{1-u} \right) \\ &= \frac{1}{(n-1)!} \frac{d}{du^{n-1}} \left(\sum_{k=0}^{+\infty} u^k \right) \\ &= \frac{1}{(n-1)!} \sum_{k=0}^{+\infty} (n+k-1) \dots (k+1) u^k = \sum_{k=0}^{+\infty} \frac{(k+n-1)!}{k!(n-1)!} u^k. \end{aligned}$$

D'où le développement de $\mathbb{E}_\theta [z^X]^n$:

$$\begin{aligned}\mathbb{E}_\theta [z^X]^n &= \frac{\theta^n}{(1 - z(1 - \theta))^n} \\ &= \sum_{k=0}^{+\infty} \frac{(k + n - 1)!}{k!(n - 1)!} z^k \theta^n (1 - \theta)^k \quad (**)\end{aligned}$$

En comparant (*) et (**), et du fait de l'unicité d'un développement en série entière, on en déduit le résultat demandé, i.e.

$$\mathbb{P}_\theta^{\otimes n}(T_n = k) = \theta^n (1 - \theta)^k \frac{(k + n - 1)!}{k!(n - 1)!}$$

3. (a) La vraisemblance du modèle au point (x_1, \dots, x_n) est $L(\theta, x_1, \dots, x_n) = \theta^n (1 - \theta)^{x_1 + \dots + x_n}$. Comme elle ne s'annule pas, on peut considérer plutôt la log-vraisemblance $\ln L(\theta, x_1, \dots, x_n) = n \ln \theta + (x_1 + \dots + x_n) \ln(1 - \theta)$. Il s'agit d'en trouver le maximum, en ayant θ variant dans $]0, 1[$. Si $x_1 + \dots + x_n = 0$, alors le maximum serait atteint pour $\theta = 1$, qui n'est pas une valeur de l'ensemble des paramètres. On peut alors dire qu'il n'y a pas d'estimateur du maximum de vraisemblance, ou plutôt rajouter à l'ensemble des lois possibles la masse de Dirac en 0 en posant $\mathbb{P}_1 = \delta_0$; notons qu'on n'a plus alors un modèle exponentiel. Si $x_1 + \dots + x_n > 0$, alors la log-vraisemblance tend vers $-\infty$ quand θ tend vers 0 ou vers 1. Le maximum est donc à rechercher dans un intervalle compact du type $[\varepsilon, 1 - \varepsilon]$ où la log-vraisemblance est continue. Elle y atteint donc son maximum. Comme elle est y aussi dérivable, ce maximum correspond à un point critique. L'estimateur du maximum de vraisemblance $\hat{\theta}_m$ doit donc vérifier l'équation de vraisemblance

$$\frac{\partial \ln L}{\partial \theta}(\hat{\theta}_m, x_1, \dots, x_n) = 0$$

Or

$$\begin{aligned}\frac{\partial}{\partial \theta} \ln L(\theta, x_1, \dots, x_n) = 0 &\iff \frac{n}{\theta} - \frac{x_1 + \dots + x_n}{1 - \theta} = 0 \\ &\iff \theta = \frac{n}{x_1 + \dots + x_n + n}\end{aligned}$$

Ainsi n'y a-t-il qu'une seule valeur de θ qui vérifie l'équation de vraisemblance. C'est donc le maximum de vraisemblance :

$$\hat{\theta}_m(x_1, \dots, x_n) = \frac{n}{x_1 + \dots + x_n + n}$$

Notons que cette formule reste valable pour le cas $x_1 + \dots + x_n = 0$ sous l'hypothèse qu'on a rajouté 1 à l'ensemble des paramètres.

Cet estimateur est fortement consistant si $\hat{\theta}_m(X_1, \dots, X_n)$ converge presque sûrement vers θ quand la taille de l'échantillon tend vers l'infini. Or les variables aléatoires X_i sont indépendantes, de même loi et possèdent un premier moment (attention aux hypothèses !). D'après la loi forte des grands nombres, on en déduit que $\frac{X_1 + \dots + X_n}{n}$ converge presque sûrement vers $\mathbb{E}_\theta[X_i] = \frac{1}{\theta} - 1$. Cela implique la convergence presque sûre de $\hat{\theta}_m(X_1, \dots, X_n) = \frac{1}{\frac{X_1 + \dots + X_n}{n} + 1}$ vers $\frac{1}{\frac{1}{\theta} - 1 + 1} = \theta$. Autrement dit, on vient de montrer que $\hat{\theta}_m(X_1, \dots, X_n)$ est fortement consistant.

Comme les variables aléatoires X_i possèdent aussi un deuxième moment (attention aux hypothèses (bis) !), on sait décrire les fluctuations de leur moyenne empirique autour de leur moyenne grâce au théorème central limite : la variable aléatoire

$$\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - \left(\frac{1}{\theta} - 1 \right) \right)$$

converge en loi vers une gaussienne centrée de variance $\text{var}_\theta(X_i) = \frac{1 - \theta}{\theta^2}$. Or

$$\hat{\theta}_m(X_1, \dots, X_n) = \psi \left(\frac{X_1 + \dots + X_n}{n} \right)$$

avec $\psi(u) = \frac{1}{1+u}$. On en déduit que

$$\sqrt{n}(\hat{\theta}_m(X_1, \dots, X_n) - \theta) = \sqrt{n}\left(\psi\left(\frac{X_1 + \dots + X_n}{n}\right) - \psi\left(\frac{1}{\theta} - 1\right)\right)$$

converge en loi vers une gaussienne centrée de variance

$$\psi'^2\left(\frac{1}{\theta} - 1\right) \frac{1 - \theta}{\theta^2} = \left(-\frac{1}{\left(\frac{1}{\theta} - 1\right)^2}\right)^2 \frac{1 - \theta}{\theta^2} = \theta^2(1 - \theta) = \frac{1}{I_1(\theta)}.$$

Or S_n est un estimateur asymptotiquement efficace de $g(\theta)$ si $\sqrt{n}(S_n - g(\theta))$ converge en loi vers une gaussienne centrée de variance $\frac{g'^2(\theta)}{I_1(\theta)}$. Ceci montre que $\hat{\theta}_m(X_1, \dots, X_n)$ est un estimateur asymptotiquement efficace de θ .

(b) D'après la première question, on sait que

$$\theta = \frac{\mathbb{E}_\theta[X]}{\mathbb{E}_\theta[X^2] - \mathbb{E}_\theta[X]^2}.$$

On en déduit, en remplaçant les moments de X par les moments empiriques, un estimateur de θ (dont on peut vérifier qu'il est fortement consistant d'après la loi des grands nombres) :

$$\begin{aligned} \theta^*(X_1, \dots, X_n) &= \frac{\frac{X_1 + \dots + X_n}{n}}{\frac{X_1^2 + \dots + X_n^2}{n} - \left(\frac{X_1 + \dots + X_n}{n}\right)^2} \\ &= \frac{T_n}{X_1^2 + \dots + X_n^2 - \frac{T_n^2}{n}} \end{aligned}$$

On peut construire par cette méthode d'autres estimateurs de θ .

(c) D'après la deuxième question, T_n est une statistique exhaustive et complète. D'après le théorème de Lehmann-Scheffé, la variable aléatoire $\mathbb{E}_\theta[1_{X_1=0}|T_n]$ est un estimateur UVMB de $\mathbb{E}_\theta[1_{X_1=0}] = \mathbb{P}_\theta^{\otimes n}(X_1 = 0) = \theta$. Or pour tout $k \in \mathbb{N}$

$$\begin{aligned} \mathbb{E}_\theta[1_{X_1=0}|T_n = k] &= \frac{\mathbb{P}_\theta^{\otimes n}(X_1 = 0|T_n = k)}{\mathbb{P}_\theta^{\otimes n}(T_n = k)} \\ &= \frac{\mathbb{P}_\theta^{\otimes n}(X_1 = 0, X_1 + \dots + X_n = k)}{\mathbb{P}_\theta^{\otimes n}(X_1 + \dots + X_n = k)} \\ &= \frac{\mathbb{P}_\theta^{\otimes n}(X_1 = 0, X_2 + \dots + X_n = k)}{\mathbb{P}_\theta^{\otimes n}(X_1 + \dots + X_n = k)} \\ &= \frac{\mathbb{P}_\theta^{\otimes n}(X_1 = 0) \times \mathbb{P}_\theta^{\otimes n}(X_2 + \dots + X_n = k)}{\mathbb{P}_\theta^{\otimes n}(X_1 + \dots + X_n = k)} \\ &\quad \text{car } X_1 \text{ et } X_2 + \dots + X_n \text{ sont indépendants} \\ &= \frac{\mathbb{P}_\theta^{\otimes n}(X_1 = 0) \times \mathbb{P}_\theta^{\otimes n}(T_{n-1} = k)}{\mathbb{P}_\theta^{\otimes n}(T_n = k)} \\ &\quad \text{car } T_{n-1} \text{ et } X_2 + \dots + X_n \text{ ont même loi} \\ &= \frac{\theta \times \theta^{n-1}(1 - \theta)^k \frac{(k+n-2)!}{k!(n-2)!}}{\theta^n(1 - \theta)^k \frac{(k+n-1)!}{k!(n-1)!}} \text{ d'après la question 2.(c)} \\ &= \frac{n-1}{k+n-1} \end{aligned}$$

Autrement dit,

$$\mathbb{E}_\theta[1_{X_1=0}|T_n] = \frac{n-1}{T_n + n - 1}$$

est un estimateur UVMB de θ .

4. Déterminons un intervalle de confiance pour $\frac{1}{\theta} - 1$ au seuil α . D'après la question 2.(a), l'espérance de $\frac{T_n}{n}$ vaut $\frac{1}{\theta} - 1$. On en déduit, en utilisant l'inégalité de Bienaymé-Tchebycheff, que pour tout $\beta > 0$ on a

$$\begin{aligned} \mathbb{P}_\theta^{\otimes n} \left(\left| \frac{T_n}{n} - \frac{1}{\theta} + 1 \right| \geq \beta \right) &\leq \frac{1}{\beta^2} \text{var}_\theta \left(\frac{T_n}{n} \right) = \frac{1}{\beta^2} \frac{1}{n^2} \text{var}_\theta(T_n) \\ &= \frac{1}{n^2 \beta^2} n \text{var}_\theta(X_1) \\ &= \frac{1 - \theta}{n \beta^2 \theta^2} \end{aligned}$$

D'où l'inégalité suivante :

$$\begin{aligned} \mathbb{P}_\theta^{\otimes n} \left(\frac{1}{\theta} - 1 \in \left[\frac{T_n}{n} - \beta, \frac{T_n}{n} + \beta \right] \right) &= \mathbb{P}_\theta^{\otimes n} \left(\left| \frac{T_n}{n} - \frac{1}{\theta} + 1 \right| \leq \beta \right) \\ &= 1 - \mathbb{P}_\theta^{\otimes n} \left(\left| \frac{T_n}{n} - \frac{1}{\theta} + 1 \right| \geq \beta \right) \\ &\geq 1 - \frac{1 - \theta}{n \beta^2 \theta^2} \end{aligned}$$

La fonction $\theta \mapsto 1 - \frac{1 - \theta}{n \beta^2 \theta^2}$ est une fonction croissante. Si $\theta \geq 0,5$, alors

$$1 - \frac{1 - \theta}{n \beta^2 \theta^2} \geq 1 - \frac{1 - 0,5}{n \beta^2 (0,5)^2} = 1 - \frac{2}{n \beta^2}$$

Autrement dit,

$$\forall \theta \geq 0,5 \quad \mathbb{P}_\theta^{\otimes n} \left(\frac{1}{\theta} - 1 \in \left[\frac{T_n}{n} - \beta, \frac{T_n}{n} + \beta \right] \right) \geq 1 - \frac{2}{n \beta^2}$$

L'intervalle aléatoire $\left[\frac{T_n}{n} - \beta, \frac{T_n}{n} + \beta \right]$ est donc un intervalle de confiance pour $\frac{1}{\theta} - 1$ au seuil $\frac{2}{n \beta^2}$.

Pour que ce seuil soit égal à α , il suffit de prendre $\beta = \sqrt{\frac{2}{n \alpha}}$. Comme de plus

$$\frac{1}{\theta} - 1 \in \left[\frac{T_n}{n} - \beta, \frac{T_n}{n} + \beta \right] \iff \theta \in \left[\frac{1}{\frac{T_n}{n} + \beta + 1}, \frac{1}{\frac{T_n}{n} - \beta + 1} \right],$$

on en déduit finalement que

$$\forall \theta \geq 0,5 \quad \mathbb{P}_\theta^{\otimes n} \left(\theta \in \left[\frac{1}{\frac{T_n}{n} + \sqrt{\frac{2}{n \alpha}} + 1}, \frac{1}{\frac{T_n}{n} - \sqrt{\frac{2}{n \alpha}} + 1} \right] \right) \geq 1 - \alpha.$$

On a ainsi bien déterminé un intervalle de confiance pour θ au seuil α , sachant $\theta \geq 0,5$.

5. (a) Le modèle statistique est un modèle exponentiel à rapport de vraisemblance croissant en $-T_n = -(X_1 + \dots + X_n)$; en effet, d'après la question 2.(a), la densité de la loi $\mathbb{P}_\theta^{\otimes n}$ par rapport à la mesure de comptage sur \mathbb{N}^n s'écrit

$$f_\theta(x_1, \dots, x_n) = \exp(-\ln(1 - \theta)(-(x_1 + \dots + x_n)) + n \ln \theta),$$

et la fonction $\theta \mapsto -\ln(1 - \theta)$ est bien strictement croissante sur $]0, 1[$ (contrairement à l'avis général...). Or on sait que pour les modèles statistiques à rapport de vraisemblance croissant en une statistique donnée, il existe pour un problème de test d'hypothèses unilatères un test uniformément plus puissant parmi les tests de seuil α fixé. Dans le cas présent, sa fonction critique est de la forme :

$$\Phi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } -(x_1 + \dots + x_n) > \lambda \\ \gamma & \text{si } -(x_1 + \dots + x_n) = \lambda \\ 0 & \text{si } -(x_1 + \dots + x_n) < \lambda \end{cases}$$

avec γ, λ déterminés par $\mathbb{E}_{\theta_0}[\Phi(X_1, \dots, X_n)] = \alpha$.

- (b) On commence par simplifier la fonction critique en supposant $\gamma = 1$, ce qui revient à supposer que l'événement $-(X_1 + \dots + X_n) = \lambda$ est de probabilité suffisamment petite pour n'intervenir que de façon négligeable dans le calcul du niveau du test. Ce niveau est donc égal à

$$\mathbb{E}_{\theta_0} [\Phi(X_1, \dots, X_n)] = \mathbb{P}_{\theta_0}^{\otimes n}(-T_n \geq \lambda)$$

Pour évaluer cette probabilité, on utilise une approximation gaussienne obtenue grâce au théorème central limite. En effet, $\frac{T_n}{n}$ est la moyenne d'une somme de variables aléatoires indépendantes, de même loi et de carré intégrable. D'après le théorème central limite, la variable aléatoire

$$\left(\frac{T_n}{n} - \mathbb{E}_{\theta_0}[X_1]\right) \sqrt{n} \frac{1}{\sqrt{\text{var}_{\theta_0}(X_1)}} = \left(\frac{T_n}{n} - \frac{1}{\theta_0} + 1\right) \sqrt{n} \sqrt{\frac{\theta_0^2}{1 - \theta_0}}$$

converge en loi vers Z gaussienne centrée réduite, quand n tend vers l'infini. Ici, n étant égal à 100, on considère que l'on peut approcher la loi de l'un par la loi de l'autre. D'après la table des quantiles de la loi $\mathcal{N}(0, 1)$, on sait que $\mathbb{P}(Z \leq -1, 2816) \simeq 10\%$. On en déduit que

$$\begin{aligned} 10\% &\simeq \mathbb{P}(Z \leq -1, 2816) \\ &\simeq \mathbb{P}_{\theta_0}^{\otimes n} \left(\left(\frac{T_n}{n} - \frac{1}{\theta_0} + 1 \right) \sqrt{n} \sqrt{\frac{\theta_0^2}{1 - \theta_0}} \leq -1, 2816 \right) \\ &\simeq \mathbb{P}_{\theta_0}^{\otimes n} \left(-T_n \geq n \left(1, 2816 \sqrt{\frac{1 - \theta_0}{n\theta_0^2}} - \frac{1}{\theta_0} + 1 \right) \right) \\ &\simeq \mathbb{P}_{0,4}^{\otimes 100}(-T_n \geq -125, 18) \end{aligned}$$

La zone de rejet pour un test de niveau 10% est donc de la forme $T_n \leq 125$.

6. Soit $S(X_1, \dots, X_n)$ un estimateur de θ . Le risque bayésien de S est donné par

$$R(\mu, S) = \int_0^1 \mathbb{E}_{\theta} [(S(X_1, \dots, X_n) - \theta)^2] \mu(d\theta).$$

Comme T_n est une statistique exhaustive, on déduit du théorème de Rao-Blackwell que $\mathbb{E}_{\theta}[S|T_n]$ est un estimateur de risque bayésien plus faible : en effet, la fonction de coût quadratique est bien convexe, donc le risque quadratique associé à $\mathbb{E}_{\theta}[S|T_n]$ est plus faible que celui associé à S pour tout θ ; d'où l'assertion. $\mathbb{E}_{\theta}[S|T_n]$ étant une fonction de T_n (plus exactement il en existe une version qui soit telle), on peut donc se contenter de rechercher les estimateurs qui minimisent le risque bayésien, autrement dit les estimateurs bayésiens, parmi les fonctions de T_n .

Soit donc g une fonction de \mathbb{N} dans \mathbb{R} . Calculons le risque bayésien associé à $g(T_n)$.

$$\begin{aligned} R(\mu, g(T_n)) &= \int_0^1 \mathbb{E}_{\theta} [(g(T_n) - \theta)^2] \mu(d\theta) \\ &= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \left(\sum_{k=0}^{+\infty} (g(k) - \theta)^2 \theta^n (1-\theta)^k \frac{(k+n-1)!}{k!(n-1)!} \right) d\theta \\ &= \sum_{k=0}^{+\infty} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{(k+n-1)!}{k!(n-1)!} \int_0^1 \theta^{n+a-1} (1-\theta)^{k+b-1} (g(k) - \theta)^2 d\theta \end{aligned}$$

où l'on a pu intervertir l'ordre de la sommation et de l'intégration dans la mesure où tous les termes étaient positifs. Minimiser le risque bayésien revient donc à minimiser pour chaque k l'intégrale

$$\begin{aligned} &\int_0^1 \theta^{n+a-1} (1-\theta)^{k+b-1} (g(k) - \theta)^2 d\theta \\ &= g(k)^2 \int_0^1 \theta^{n+a-1} (1-\theta)^{k+b-1} d\theta - 2g(k) \int_0^1 \theta^{n+a} (1-\theta)^{k+b-1} d\theta + \int_0^1 \theta^{n+a+1} (1-\theta)^{k+b-1} d\theta \end{aligned}$$

Ce polynôme du deuxième degré en $g(k)$ trouve est minimal pour

$$\begin{aligned}g(k) &= \frac{\int_0^1 \theta^{n+a}(1-\theta)^{k+b-1} d\theta}{\int_0^1 \theta^{n+a-1}(1-\theta)^{k+b-1} d\theta} \\&= \frac{\beta(n+a+1, k+b)}{\beta(n+a, k+b)} \\&= \frac{\Gamma(n+a+1)\Gamma(k+b)\Gamma(n+k+a+b)}{\Gamma(n+a)\Gamma(k+b)\Gamma(n+k+a+b+1)} \\&= \frac{n+a}{n+k+a+b}\end{aligned}$$

où l'on utilise le fait que $\Gamma(n+a+1) = (n+a)\Gamma(n+a)$ et $\Gamma(n+k+a+b+1) = (n+k+a+b)\Gamma(n+k+a+b)$.

On a ainsi montré que $\frac{n+a}{n+T_n+a+b}$ est un estimateur bayésien de θ .