Estimation, vraisemblance, confiance

Commençons par un exemple simple : on effectue n tirages successifs d'une pièce où la probabilité d'obtenir face est p (et celle d'avoir pile est $1 - p$). Signalons que ce problème est identique à celui-ci : dans une population où une proportion p des individus possède une certaine caractéristique ($1 - p$ ne l'ayant pas), on prélève n individus successivement avec remise. Dans ce contexte, on appelle X_i la variable aléatoire au $i^{è}$ tirage, telle que $X_i = 1$ si face sort, et $X_i = 0$ sinon. Puis on prend la variable $T_n = \frac{X_1 + X_2 + ... + X_n}{n}$, avec $n \geq 1$. Celle-ci correspond à la fréquence de sortie des face au cours des n tirages.

La notion d'estimateur, et d'estimateur convergent

Notre problème est alors le suivant : on ne connaît pas la valeur du paramètre p, mais en réalisant des expériences de n tirages successifs, on va pouvoir estimer quelle est la valeur de p. Que va-t-il se passer ? Si l'on réalise une seule expérience avec n petit, le résultat obtenu avec X_n sera plus ou moins proche de p, et en reproduisant cette expérience, on ne trouvera sans doute pas la même valeur pour p. Toujours avec n petit, si l'on fait un certain nombre d'expériences en calculant à la fin la moyenne des T_n obtenus, on aura une valeur proche de p. Enfin, si l'on fait une seule expérience, mais avec n grand, on obtient une valeur proche de p, en vertu de la loi des grands nombres.

Résumons la situation : en prenant un échantillon de n tirages, la valeur de T_n obtenue expérimentalement constitue une estimation du paramètre p, plus ou moins valable certes. Comme il existe une variable aléatoire T_n pour chaque valeur de n, on dit plus précisément que la suite de variables aléatoires (T_n) est un estimateur de ce paramètre. Cet estimateur est de bonne qualité dans la mesure où pour n suffisamment grand, il est très proche de p. On dit alors que l'estimateur est convergent, lorsque la réalisation d'une expérience sur un échantillon de n tirages a une probabilité très faible de s'écarter de la valeur du paramètre pour peu que n soit suffisamment grand.

Si la fréquence expérimentale de sortie des faces sur un échantillon de n tirages constitue un estimateur convergent du paramètre p, il existe d'autres types d'expériences où l'estimateur peut être choisi autrement. Prenons maintenant l'exemple d'une loi uniforme sur un intervalle [0 a], le paramètre inconnu étant maintenant a (a réel >0). Cela signifie que si l'on découpe cet intervalle en petits intervalles tous de même longueur, on a autant de chances de tomber dans l'un que dans un autre, et que la probabilité de tomber ailleurs est nulle. Dans ces conditions, si l'on réalise un échantillon de n tirages X_1, X_2, …, X_n, on peut prendre comme estimateur le maximum obtenu. Ce maximum est proche de a, et si n devient très grand, il s'approche d'autant plus de a. On a encore là un estimateur convergent.

Estimateur sans biais

Cela nous amène à préciser la qualité de l'estimateur. On dit que l'estimateur (T_n) est sans biais lorsque sa valeur moyenne (son espérance $E(T_n)$) est égale au paramètre, quel que soit le nombre n de tirages. c'est-à-dire quelle que soit la taille de l'échantillon. Dans l'exemple des n tirages à pile ou face, on a justement $E(T_n) = p$, et l'estimateur (T_n) est sans biais. Par contre, dans l'exemple de la loi uniforme sur [0 a], la valeur maximale obtenue en n tirages est toujours strictement inférieure à a, et si cet estimateur est convergent, il n'est pas sans biais.

On dispose aussi d'une propriété que nous admettrons ici : lorsque l'estimateur est sans biais et que sa variance tend vers 0 lorsque n tend vers l'infini, alors il est convergent. Il suffit donc de calculer l'espérance et la variance pour savoir si l'estimateur est sans biais, et ensuite convergent.

Tout ce qui précède peut être précisé par le calcul :

- Cas des n tirages à pile ou face, avec comme variable aléatoire la fréquence de sortie des face, $T_n = \dfrac{X_1 + X_2 + ... + X_n}{n}$. Pour chacun des n tirages, la variable X_i suit une loi de Bernoulli, et l'on a $E(X_i) = p$ et $V(X_i) = p(1 - p)$. On en déduit que :

$E(T_n) = \dfrac{1}{n}(E(X_1) + E(X_2) + ... + E(X_n)) = \dfrac{1}{n} np = p$. T_n est un estimateur sans biais.

En utilisant le fait que les variables X_i sont indépendantes, on a aussi :

$V(T_n) = \dfrac{1}{n^2}(V(X_1) + V(X_2) + ... + V(X_n)) = \dfrac{1}{n^2} np(1 - p) = \dfrac{p(1 - p)}{n}$, qui tend vers 0 pour n infini. L'estimateur T_n est convergent.

- Cas de la loi uniforme sur [0 a], avec comme densité de probabilité $d(x) = 1/a$ pour x dans [0 a], et $d(x) = 0$ ailleurs. La fonction de répartition est alors :

$F(x) = x / a$ sur l'intervalle [0 a], $F(x) = 0$ pour x négatif, et $F(x) = 1$ pour x supérieur à 1.

Pour la variable aléatoire X obéissant à cette loi uniforme, on en déduit son espérance et sa variance :

$$E(X) = \int_{-\infty}^{+\infty} t d(t) dt = \int_0^a \frac{t}{a} dt = \frac{a}{2}$$

$$V(X) = E(X^2) - (E(X))^2 = \int_0^a \frac{t^2}{a} dt - \frac{a^2}{4} = \frac{a^2}{3} - \frac{a^2}{4} = \frac{a^2}{12}$$

Prenons maintenant un échantillon de n tirages ayant tous cette loi uniforme de paramètre a, et construisons à partir de là une variable aléatoire T_n dont on va tester la qualité d'estimateur.

1) $T_n = \dfrac{2}{n}(X_1 + X_2 + ... + X_n)$.

Si l'on prend le double de la moyenne sur n expériences, c'est parce que cette moyenne est censée approcher l'espérance a / 2. On en déduit :

$E(T_n) = \frac{2}{n} n E(X) = \frac{2}{n} n \frac{a}{2} = a$. T_n est un estimateur sans biais.

$V(T_n) = \frac{4}{n^2} n V(X) = \frac{4}{n^2} n \frac{a^2}{12} = \frac{a^2}{3n}$ qui tend vers 0 lorsque n tend vers l'infini. L'estimateur T_n est convergent.

2) $T_n = \max(X_1, X_2, ..., X_n)$

La fonction de répartition correspondante est :

$$F_{T_n}(x) = p(X_1 < x \; et \; X_2 < x \; et ... \; et \; X_n < x)$$
$$= p(X_1 < x)\, p(X_2 < x) ... \, p(X_n < x)$$
$$= F(x)^n$$

où l'on retrouve la fonction de répartition $F(x)$ de la loi uniforme,

$= \left(\frac{x}{a}\right)^n$ sur [0 a], et sinon 0 pour $x < 0$ et 1 pour $x > a$.

On en déduit la densité $d(x) = F_{\mathrm{Tn}}(x)$ ' = $\frac{1}{a^n} n\, x^{n-1}$ sur [0 a] et 0 ailleurs. Alors :

$E(T_n) = \int_0^a \frac{1}{a^n} n\, t^n dt = \frac{n}{n+1} a$. L'estimateur T_n n'est pas sans biais.

$$V(T_n) = E(T_n{}^2) - (E(T_n))^2 = \int_0^a \frac{1}{a^n} n\, t^{n+1} dt - \frac{n^2}{(n+1)^2} a^2$$
$$= \frac{n}{n+2} a^2 - \frac{n^2}{(n+1)^2} a^2 = n\, a^2 \left(\frac{1}{n+2} - \frac{n}{(n+1)^2}\right)$$
$$= \frac{n\, a^2}{(n+1)^2 (n+2)}$$

Lorsque n tend vers l'infini, la variance tend vers 0.

3) $T_n = \frac{n+1}{n} \max(X_2, X_2, ..., X_n)$

En corrigeant ainsi l'estimateur trouvé au 2°, on trouve que $E(T_n) = \frac{n+1}{n} \frac{n}{n+1} a = a$, cet estimateur est sans biais, et

$V(T_n) = \frac{(n+1)^2}{n^2} \frac{n\, a^2}{(n+1)^2 (n+2)} = \frac{a^2}{n(n+2)}$ qui tend vers 0 pour n infini. L'estimateur sans biais est aussi convergent.

Enfin, dès que n dépasse 1, la variance obtenue ici est inférieure à celle obtenue avec l'estimateur sans biais convergent du 1° (on avait une variance égale à a^2 / 3n), on en conclut que l'estimateur du 3° est meilleur que celui du 1°. On dit qu'il est plus efficace.

4) $T_n = \min(X_1, X_2, ..., X_n)$

On peut prévoir que la valeur moyenne de T_n est proche de 0, et que l'on n'a vraiment pas un estimateur de a. Vérifions-le par le calcul.

La fonction de répartition de T_n est, compte tenu des résultats obtenus sur la variable X de la loi uniforme, et de sa fonction de répartition $F(x)$:

$$F_{T_n}(x) = p(T_n < x) = 1 - p(T_n > x) = 1 - p(X_1 > x \ et \ X_2 > x \ ... \ et \ X_n > x)$$

$$= 1 - p(X_1 > x)p(X_2 > x)...p(X_n > x) = 1 - (p(X > x))^n$$

$$= 1 - (1 - p(X < x))^n = 1 - (1 - F(x))^n = 1 - (1 - \frac{x}{a})^n \quad \text{sur } [0 \ a]$$

et $F_{Tn} = 0$ pour $x < 0$, $F_{Tn} = 1$ pour $x > 1$.

On en déduit par dérivation la densité $d(x) = \frac{n}{a}(1 - \frac{x}{a})^{n-1}$ sur $[0 \ a]$ et 0 ailleurs, puis :

$E(T_n) = \frac{n}{a}\int_0^a t(1 - \frac{t}{a})^{n-1} dt = \frac{a}{n+1}$ après une intégration par parties, et

comme la dispersion des résultats est la même avec le minimum des X_i qu' avec le maximum des X_i, pour des raisons de symétrie, on a comme au 2° $V(T_n) = \frac{n a^2}{(n+1)^2(n+2)}$.

5) $T_n = \max(X_1, X_2, ..., X_n) + \min(X_1, X_2, ..., X_n)$

Grâce aux calculs précédents,

$$E(T_n) = E(\max(X_1, X_2, ..., X_n)) + E(\min(X_1, X_2, ..., X_n))$$

$$= \frac{n}{n+1}a + \frac{1}{n+1}a = a$$

On obtient bien un estimateur sans biais. D'autre part, en notant $Y = \max(X_1, X_2, ..., X_n)$ et $Z = \min(X_1, X_2, ..., X_n)$, on a :

$V(T_n) = V(Y) + V(Z) + 2\operatorname{cov}(Y, Z) = 2V(Y) + 2\operatorname{cov}(Y, Z)$ car les variables Y et Z ne sont pas indépendantes et que $V(Y) = V(Z)$. Sachant que

$\operatorname{cov}(Y, Z)^2 \leq V(Y)\ V(Z)$ soit ici $|\operatorname{cov}(Y, Z)| \leq V(Y)$, on en déduit

$$|V(T_n)| \leq 2|V(Y)| + 2|\operatorname{cov}(Y, Z)|$$

$$V(T_n) \leq 2V(Y) + 2V(Y)$$

$$V(T_n) \leq 4V(Y)$$

$$V(T_n) \leq \frac{4n a^2}{(n+1)^2(n+2)}$$

L'estimateur T_n est lui aussi convergent. Comme sa variance est inférieure à celle obtenue avec l'estimateur du 1° (soit $a^2 / 3n$), et même infiniment plus petite pour n infini, on en déduit que l'estimateur du 5° est plus efficace que celui du 1°.

Exercice d'application : la loi de Pareto

On dit qu'une variable aléatoire X obéit à la loi de Pareto de paramètres a et x_0 (tous deux > 0) lorsqu'elle admet pour densité

$d(t) = a\frac{x_0^a}{t^{a+1}}$ pour $t \geq x_0$ et 0 sinon ($t \in \mathbf{R}$)

A- Quelques propriétés de la loi de Pareto

1) Vérifier que d est bien une densité, et déterminer l'espérance et la variance de X, sous certaines conditions d'existence que l'on précisera.

- La fonction d est continue sur R sauf au point de discontinuité x_0 où elle admet une limite à gauche et une limite à droite, elle est partout ≥ 0, et l'intégrale $\int_{-\infty}^{+\infty} d(t)dt$ existe et vaut 1, en effet :

$\int_{-\infty}^{+\infty} d(t)dt = a x_0^a \int_{x_0}^{+\infty} \frac{1}{t^{a+1}} dt$, cette intégrale convergeant en $+\infty$ puisque l'exposant de t, $a+1 > 0$

$$= a x_0^a \left[\frac{t^{-a}}{-a} \right]_{x_0}^{+\infty} = a x_0^a \frac{x_0^{-a}}{a} = 1$$

- $E(X) = \int_{x_0}^{+\infty} t d(t)dt = a x_0^a \int_{x0}^{+\infty} \frac{1}{t^a} dt$. Cette intégrale n'existe que pour $a > 1$.

$$= a x_0^a \left[\frac{t^{-a+1}}{-a+1} \right]_{x_0}^{+\infty} = \frac{a x_0}{a-1}$$ pour $a > 1$, sinon X n'a pas d'espérance.

- Pour la variance, on commence par chercher $E(X^2)$:

$E(X^2) = \int_{x_0}^{+\infty} t^2 d(t)dt = a x_0^a \int_{x_0}^{+\infty} \frac{1}{t^{a-1}} dt$, qui n'existe que pour $a > 2$

$$= a x_0^a \left[\frac{t^{-a+2}}{-a+2} \right]_{x_0}^{+\infty} = \frac{a x_0^2}{a-2}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{a x_0^2}{a-2} - \frac{a^2 x_0^2}{(a-1)^2}$$

$$= \frac{a x_0^2}{(a-1)^2 (a-2)}$$ pour $a > 2$, sinon X n'admet pas de variance

2) Déterminer la fonction de répartition $F(x)$ de X.

Par définition $F(x) = p(X < x) = \int_{-\infty}^{x} d(t)dt$.

Pour $x < x_0$, avec une densité nulle, $F(x) = 0$.

Pour $x \geq x_0$,

$$F(x) = a x_0^a \int_{x_0}^{x} \frac{1}{t^{a+1}} dt = a x_0^a \left[\frac{t^{-a}}{-a} \right]_{x_0}^{x} = a x_0^a \left(\frac{x^{-a}}{-a} - \frac{x_0^{-a}}{-a} \right)$$

$$= 1 - \left(\frac{x_0}{x} \right)^a$$

3) Montrer que la probabilité conditionnelle $p(X > x + y \mid X > x)$ tend vers 1 lorsque x tend vers $+\infty$, y étant un nombre réel positif.

Par définition de la probabilité conditionnelle :

$$p\left(X > x+y \mid X > x\right)=\frac{p\left(X > x+y \text{ et } X > x\right)}{p(X > x)}=\frac{p\left(X > x+y\right)}{p(X > x)}$$

En supposant que x est > x_0, ce qui sera le cas lorsque x va tendre vers l'infini, on sait que $p\left(X > x\right)=1-p\left(X < x\right)=1-F\left(x\right)=\left(\frac{x_0}{x}\right)^a$. Ainsi :

$p\left(X > x+y \mid X > x\right)=\left(\frac{x}{x+y}\right)^a$. Lorsque x tend vers +∞, cette quantité va tendre vers 1.

Qu'est-ce que cela signifie concrètement ? Si la loi de Pareto modélise la durée de vie d'un phénomène, il est naturel d'appeler $p(X > x)$ la fonction de survie. Que signifie alors $p\left(X > x+y \mid X > x\right)\to 1$ pour x infini ? Plus le phénomène a une durée de vie longue, plus il a de chances de prolonger cette durée de vie longtemps. Autrement dit, plus on vieillit, plus on a de chances de vivre encore longtemps. Le « phénomène » en question n'est pas l'être humain ! [1]

B- Détermination d'estimateurs de x_0

On se donne une suite de variables aléatoires (X_n), toutes indépendantes, et de même loi que X, chacune obéissant donc à la loi de Pareto de paramètres a et x_0, avec $a > 2$ (il existe une espérance et une variance). On suppose que a est connu, et l'on veut estimer la valeur de x_0.

1) Première méthode : on utilise la moyenne expérimentale : $Y_n=\frac{X_1+X_2+...+X_n}{n}$. Déterminer la constante K telle que la variable aléatoire $Y'_n = K\ Y_n$ soit un estimateur sans biais de x_0. Montrer que cet estimateur (Y'_n) est aussi convergent.

$E(Y_n)=\frac{1}{n}(X_1+X_2+...+X_n)=\frac{n}{n}\frac{ax_0}{a-1}=\frac{ax_0}{a-1}$. Il suffit de prendre $Y_n'=\frac{a-1}{a}Y_n$ pour avoir $E(Y_n') = x_0$. Cela signifie que Y_n' est un estimateur sans biais de x_0. D'autre part,

$$\begin{aligned}V(Y_n')&=V\left(\frac{a-1}{a}Y_n\right)=V\left(\frac{a-1}{na}(X_1+X_2+...+X_n)\right)\\&=\frac{(a-1)^2}{n^2a^2}nV(X)=\frac{(a-1)^2}{n\,a^2}\frac{a\,x_0^2}{(a-1)^2(a-2)}\\&=\frac{x_0^2}{n\,a(a-2)}\end{aligned}$$

---

[1] Signalons que pour la loi exponentielle de paramètre k on a : $p\left(X > x+y \mid X > x\right)=\frac{e^{-k(x+y)}}{e^{-kx}}=e^{-ky}=p(X > y)$. Cela signifie que quel que soit l'âge que l'on a, on a autant de chances de vivre encore un certain temps.

Comme l'estimateur est sans biais et que sa variance tend vers 0 pour n infini, il est aussi convergent.

2) Deuxième méthode : On prend comme variable aléatoire
$Z_n = \min(X_1, X_2, \ldots, X_n)$.

a) Déterminer sa fonction de répartition puis son espérance et sa variance.

Appelons G la fonction de répartition de Z_n. Par définition,
$G(x) = p(Z_n < x) = 1 - p(Z_n > x) = 1 - p(X_1 > x$ et $X_2 > x$ et … et $X_n > x)$
$= 1 - p(X_1 > x)\, p(X_2 > x) \ldots p(X_n > x) = 1 - p(X > x)^n$
$= 1 - (1 - p(X < x))^n$
où l'on a introduit la variable aléatoire X obéissant à la loi de Pareto de paramètres a et x_0, dont on connaît la fonction de répartition $F(x) = p(X < x) = 1 - (\frac{x_0}{x})^a$. D'où

$$G(x) = 1 - \left((\frac{x_0}{x})^a \right)^n = 1 - (\frac{x_0}{x})^{na}$$

On reconnaît une loi de Pareto de paramètres na et x_0. On en déduit :

$$E(Z_n) = \frac{n\,a\,x_0}{na - 1} \quad \text{et} \quad V(Z_n) = \frac{n\,a\,x_0^2}{(na-1)^2(na-2)}$$

b) Déterminer la variable aléatoire Z_n' de la forme $b_n\, Z_n$ telle que la suite (Z_n') donne un estimateur sans biais de x_0, avec b_n fonction de n. Montrer que cet estimateur est aussi convergent.

Au vu de la courbe de la densité de la loi de Pareto, c'est dans une petite zone au-delà de x_0 que l'on a de très fortes probabilités. Il est donc naturel de prendre comme estimateur de x_0 le minimum des variables X_i (i de 1 à n), à savoir Z_n. On remarque d'ailleurs que pour n grand, l'espérance de Z_n tend vers x_0, grâce à la formule précédemment trouvée. Mais cet estimateur est biaisé, dans la mesure où il donne une valeur supérieure à x_0. Pour avoir un estimateur non biaisé, il suffit de prendre :

$$Z_n' = \frac{na-1}{na} Z_n \quad \text{puisqu'alors} \quad E(Z_n') = \frac{na-1}{na} E(Z_n) = \frac{na-1}{na}\frac{n\,a\,x_0}{na-1} = x_0$$

D'autre part

$$V(Z_n') = \left(\frac{na-1}{na}\right)^2 V(Z_n) = \frac{(na-1)^2}{n^2 a^2}\frac{n\,a\,x_0^2}{(na-1)^2(na-2)}$$

$$= \frac{x_0^2}{n\,a(na-2)}$$

La variance tend vers 0 pour n infini. L'estimateur sans biais est aussi convergent.

c) Montrer que l'estimateur (Z_n') est plus efficace que (Y_n')

Il s'agit de comparer $V(Y_n') = \frac{x_0^2}{n\,a(a-2)}$ et $V(Z_n') = \frac{x_0^2}{n\,a(na-2)}$.

On voit aussitôt que $V(Z_n') \leq V(Y_n')$ dès que n est ≥ 1. L'estimateur Z_n' est plus efficace que Y_n' , et quand n augmente, il l'est de plus en plus , avec $V(Z_n') / V(Y_n')$ tendant vers 0 quand n tend vers l'infini.

Exercice : Loi de Poisson et estimateur d'une exponentielle

Considérons n variables aléatoires indépendantes $X_1, X_2, \ldots, X_n$ obéissant toutes à une loi de Poisson de paramètre λ (réel positif). L'objectif est de trouver une estimation de $e^{-\lambda}$.

Pour cela on introduit les variables $Y_1, Y_2, \ldots, Y_n$, telles que $Y_i = 1$ si $X_i = 0$, et $Y_i = 0$ sinon, ainsi que leur somme $Z = \sum_{i=1}^{n} Y_i$

1) Montrer que Z / n est un estimateur de $e^{-\lambda}$, sans biais et convergent.

Comme $p(X_i = 0) = e^{-\lambda}$ et que Z / n est la proportion des variables X_i nulles parmi toutes les X_i, il est logique que l'on ait là un estimateur de $e^{-\lambda}$. Vérifions-le. Les variables aléatoires Y_i suivent une loi de Bernoulli de paramètre $p(Y_i) = 1) = p(X_i = 0) = e^{-\lambda}$. D'où l'espérance $E(Z / n) = n\, e^{-\lambda} / n = e^{-\lambda}$.

D'autre part, $V(Y_i) = e^{-\lambda}(1 - e^{-\lambda})$. Comme les X_i, les variables Y_i sont indépendantes, et $V(Z / n) = (1 / n^2)\, n\, e^{-\lambda}(1 - e^{-\lambda}) = e^{-\lambda}(1 - e^{-\lambda}) / n$ qui tend vers 0 pour n infini. La variable aléatoire Z / n est bien un estimateur sans biais et convergent de $e^{-\lambda}$.

2) Montrer que l'on a $p(X_1 = 0 \mid X_1 + X_2 + \ldots + X_n = j) = \left(\frac{n-1}{n}\right)^j$ pour tout j entier naturel. On rappelle qu'une somme de n variables aléatoires vérifiant la même loi de Poisson de paramètre λ suit une loi de Poisson de paramètre $n\lambda$.

$$p(X_1 = 0 \mid X_1 + X_2 + \ldots + X_n = j) = \frac{p(X_1 = 0 \text{ et } (X_1 + X_2 + \ldots + X_n = j))}{p(X_1 + X_2 + \ldots + X_n = j)}$$

$$= \frac{p(X_1 = 0 \text{ et } (X_2 + \ldots + X_n = j))}{p(X_1 + X_2 + \ldots + X_n = j)} = \frac{p(X_1 = 0)\; p(X_2 + \ldots + X_n = j)}{p(X_1 + X_2 + \ldots + X_n = j)}$$

$$= \frac{e^{-\lambda}\; e^{-(n-1)\lambda} \dfrac{(n-1)^j \lambda^j}{j!}}{e^{-n\lambda} \dfrac{n^j \lambda^j}{j!}} = \left(\frac{n-1}{n}\right)^j$$

3) On pose $S_n = X_1 + X_2 + \ldots + X_n$ qui prend ses valeurs dans **N**, et l'on considère la nouvelle variable aléatoire $\left(\frac{n-1}{n}\right)^{S_n}$. Montrer que celle-ci est un estimateur sans biais et convergent de $e^{-\lambda}$.

$$E\left(\left(\frac{n-1}{n}\right)^{S_n}\right) = \sum_{j=0}^{+\infty}\left(\frac{n-1}{n}\right)^j p(S_n = j) = \sum_{j=0}^{+\infty}\left(\frac{n-1}{n}\right)^j e^{-n\lambda}\,\frac{n^j\lambda^j}{j!}$$

$$= e^{-n\lambda}\sum_{j=0}^{+\infty}\frac{(n-1)^j\lambda^j}{j!} = e^{-n\lambda}e^{(n-1)\lambda} = e^{-\lambda}$$

D’autre part :

$$E\left(\left(\left(\frac{n-1}{n}\right)^{S_n}\right)^2\right) = \sum_{j=0}^{+\infty}\left(\frac{n-1}{n}\right)^{2j} p(S_n = j) = \sum_{j=0}^{+\infty}\left(\frac{n-1}{n}\right)^{2j} e^{-n\lambda}\,\frac{n^j\lambda^j}{j!}$$

$$= e^{-n\lambda}\sum_{j=0}^{+\infty}\frac{1}{j!}\left(\frac{(n-1)^2\lambda}{n}\right)^j = e^{-n\lambda}e^{\frac{(n-1)^2\lambda}{n}} = e^{-2\lambda+\frac{\lambda}{n}}$$

$$V\left(\left(\frac{n-1}{n}\right)^{S_n}\right) = e^{-2\lambda+\frac{\lambda}{n}} - e^{-2\lambda} = e^{-2\lambda}\left(e^{\frac{\lambda}{n}}-1\right)$$ qui tend bien vers 0 pour *n* infini.

4) Montrer que quelle que soit la valeur du paramètre λ, $\left(\frac{n-1}{n}\right)^{S_n}$ est un meilleur estimateur de $e^{-\lambda}$ que Z/n.

Formons la différence de leurs variances :

$$V\left(\left(\frac{n-1}{n}\right)^{S_n}\right) - V(Z/n) = \frac{e^{-2\lambda}}{n}\left(ne^{\frac{\lambda}{n}} - e^{\lambda} - n + 1\right) = \frac{e^{-2\lambda}}{n}\,g(\lambda)$$

On constate que la dérivée $g(\lambda)' = e^{\lambda/n} - e^{\lambda}$ est négative, la fonction $g(\lambda)$ décroît à partir de $g(0) = 0$, elle est aussi négative. D’où $V\left(\left(\frac{n-1}{n}\right)^{S_n}\right) < V(Z/n)$.

La méthode du maximum de vraisemblance, pour déterminer un estimateur

Reprenons notre jeu de *n* tirages successifs à pile ou face, avec ses *n* variables de Bernoulli prenant la valeur 1 avec la probabilité *p* quand face (*F*) sort et la valeur 0 avec la probabilité 1 – *p* quand c’est pile (*P*). La valeur du paramètre *p* nous est inconnue.

On va alors faire une seule expérience. Par exemple pour $n = 10$, on obtient comme résultat *FPPFFPPPFP* ou en binaire 1001100010. La probabilité de cet évènement est $p^4(1-p)^6$. En essayant les valeurs de *p* de 0 à 1, on s’aperçoit que la quantité $p^4(1-p)^6$ admet un maximum au voisinage de $p = 0{,}4$. Que conclure ? Qu’il est plus vraisemblable d’avoir une valeur de *p* égale à 0,4 qu’une valeur égale à 0,7 par exemple. Ainsi, à l’issue d’une expérience avec ses *n* tirages obtenus, il est normal de prendre comme estimation de *p* celle qui correspond à la probabilité maximale pour l’expérience réalisée, même si l’on sait qu’une deuxième expérience pourrait donner une estimation de *p* plus ou moins différente. Cela nous amène à définir une fonction vraisemblance $L(x_1, x_2, \ldots, x_n, p)$:

A partir d'un échantillon de n tirages dont les variables aléatoires sont X_1, X_2, …, X_n, toutes de même loi (de paramètre p) et indépendantes, la réalisation d'une expérience conduit à un résultat x_1, x_2, …, x_n dont la probabilité est :

$p(X_1 = x_1$ et $X_2 = x_2$ et … et $X_n = x_n) = p(X_1 = x_1)\, p(X_2 = x_2) \ldots p(X_n = x_n)$, et l'on prend comme fonction de vraisemblance ce produit de probabilités, soit :

$$L(x_1, x_2, \ldots, x_n, p) = p(X_1 = x_1)\, p(X_2 = x_2) \ldots p(X_n = x_n).$$

Cette définition, valable pour des probabilités discrètes, se généralise aux probabilités continues, en faisant intervenir la densité de probabilité $d(x)$ de la loi de probabilité associée aux n variables aléatoires de l'échantillon, soit

$$L(x_1, x_2, \ldots, x_n, p) = d(x_1)\, d(x_2) \ldots d(x_n)$$

On obtient alors une estimation du paramètre p en prenant comme valeur de p celle qui rend maximale la vraisemblance. Il reste à chercher ce maximum en prenant la dérivée par rapport à p de la fonction vraisemblance, de façon qu'elle s'annule en passant du positif au négatif. C'est cela la méthode du maximum de vraisemblance. Plus précisément, comme la fonction L fait intervenir un produit, on préfère prendre son logarithme, $\ln L$, de façon à obtenir une somme, et de chercher le maximum de cette somme, ce qui correspond exactement au maximum du produit dans L, puisque la fonction logarithme est croissante.

Reprenons l'exemple des n tirages répétés à pile ou face, chacun obéissant à la loi de Bernoulli de paramètre p. Or la variable aléatoire X de la loi de Bernoulli est telle que $p(X = x) = p^x (1-p)^{1-x}$, puisque cela donne $p(X = 1) = p$ et $p(X = 0) = 1 - p$. La fonction de vraisemblance s'écrit :

$$\begin{aligned} L(x_1, x_2, \ldots, x_n, p) &= p(X_1 = x_1)\, p(X_2 = x_2) \ldots p(X_n = x_n) \\ &= p^{\,x1+x2+\ldots+xn}\,(1-p)^{(1-x1)+(1-x2)+\ldots+(1-xn)} \end{aligned}$$

$\ln L = (x_1 + x_2 + \ldots + x_n) \ln p + (n - (x_1 + x_2 + \ldots + x_n)) \ln(1-p)$

En dérivant par rapport à p :

$$\begin{aligned} (\ln L)' &= (x_1 + x_2 + \ldots + x_n)\frac{1}{p} - (n - (x_1 + x_2 + \ldots + x_n))\frac{1}{1-p} \\ &= \frac{(x_1 + x_2 + \ldots + x_n) - np}{p(1-p)} \end{aligned}$$

Cette dérivée s'annule pour $p = \dfrac{(x_1 + x_2 + \ldots + x_n)}{n}$ et elle passe du signe + au signe – (le numérateur de $(\ln L)'$ étant une fonction affine, et le dénominateur restant positif). Au terme de ce calcul, on retrouve le fait que l'estimateur pour p, issu de la méthode du maximum de vraisemblance, est la fréquence de sortie des faces, soit $\dfrac{X_1 + X_2 + \ldots + X_n}{n}$

Quelques exemples

1) Loi géométrique

Rappelons qu'une variable X suit une loi géométrique de paramètre p lorsque $p(X = k) = (1 - p)^{k-1} p$ avec k entier ≥ 1.

Prenons maintenant une suite de variables aléatoires (X_n) indépendantes et obéissant toutes à la loi géométrique de paramètre p. La réalisation d'un évènement sur n tirages conduit au résultat $x_1, x_2, \ldots, x_n$. La fonction vraisemblance est :

$$L(x_1, x_2, \ldots, x_n. p) = (1 - p)^{x1 - 1}\, p\, (1 - p)^{x2 - 1}\, p \ldots (1 - p)^{xn - 1}\, p$$
$$= p^n (1-p)^{\sum x_i - n}$$
$$\ln L = n \ln p + (\Sigma x_i - n) \ln (1 - p)$$

Dérivons par rapport à p :
$$(\ln L)' = \frac{n}{p} - \frac{\sum x_i - n}{1-p} = \frac{n - (\sum x_i) p}{p(1-p)} .$$

Elle s'annule pour $p = \dfrac{n}{\sum x_i}$ tout en passant du positif au négatif. L'estimateur du maximum de vraisemblance de p est :

$\dfrac{n}{\sum X_i}$, soit l'inverse de la moyenne expérimentale. Cela était prévisible puisque l'espérance de la loi géométrique est $1 / p$.

2) Loi exponentielle

Par définition, une variable aléatoire qui suit la loi exponentielle de paramètre k a pour densité $d(x) = ke^{-kx}$ pour $x \geq 0$, et 0 sinon. On en déduit notamment son espérance $E(X) = 1 / k$. Prenons maintenant une suite de variables aléatoires (X_n) (avec $n \geq 1$) obéissant à cette même loi, et indépendantes.
La fonction de vraisemblance est :

$$L(x_1, x_2, ..., x_n, k) = ke^{-kx_1} ke^{-kx_2} ... ke^{-kx_n} = k^n e^{-k \sum_{i=1}^{n} x_i}$$
$$\ln L = n \ln k - k \sum_{i=1}^{n} x_i$$

La dérivée par rapport à k est $\dfrac{n}{k} - \sum_{i=1}^{n} x_i$. Elle s'annule pour

$k = \dfrac{n}{\sum_{i=1}^{n} x_i}$, en passant du signe plus ou signe moins. La suite de variables

aléatoires ($\dfrac{n}{\sum_{i=1}^{n} X_i}$), inverse de la moyenne expérimentale, est un estimateur du paramètre a, comme on pouvait s'y attendre, puisque l'espérance de X est justement $1 / k$.[2]

3) Loi de Pareto (suite)

a) Déterminer un estimateur du maximum de vraisemblance.

La loi de Pareto a deux paramètres a et x_0 tous deux >0. On veut maintenant déterminer un estimateur du paramètre a supposé inconnu. A partir de la densité $d(t) = a\dfrac{x_0^a}{t^{a+1}}$ pour $t \geq x_0$ et 0 sinon, on prend la fonction de vraisemblance :

$L(x_1, x_2, \ldots, x_n, a) = d(x_1)\, d(x_2) \ldots d(x_n)$ en supposant que tous les x_i sont $> x_0$.

$$= a\frac{x_0^a}{x_1^{a+1}}\; a\frac{x_0^a}{x_2^{a+1}} \ldots a\frac{x_0^a}{x_n^{a+1}} = a^n \frac{x_0^{na}}{x_1^{a+1}x_2^{a+1}\ldots x_n^{a+1}}$$

$\ln L = n \ln a + na \ln x_0 - (a + 1)(\ln x_1 + \ln x_2 + \ldots + \ln x_n)$

Dérivons par rapport à a : $(\ln L)' = \dfrac{n}{a} + n\ln x_0 - (\ln x_1 + \ln x_2 + \ldots + \ln x_n)$

La dérivée s'annule pour

$\dfrac{n}{a} = (\ln x_1 + \ln x_2 + \ldots + \ln x_n) - n\ln x_0 = \ln\dfrac{x_1}{x_0} + \ln\dfrac{x_2}{x_0} + \ldots + \ln\dfrac{x_n}{x_0}$, qui est un nombre positif puisque tous les x_i sont $> x_0$ (>0), d'où la valeur de a correspondante :

$$a = \frac{n}{\sum_{i=1}^{n} \ln\frac{x_i}{x_0}}$$

Cette valeur est bien un maximum pour $\ln L$, et par suite pour L, puisque la dérivée de $\ln L$ par rapport à a, de la forme $n / a - B$, est décroissante, passant donc du positif au négatif. Ainsi la suite (W_n), avec

$W_n = \dfrac{n}{\sum_{i=1}^{n} \ln\frac{X_i}{x_0}}$, est un estimateur du maximum de vraisemblance.

On peut déjà prévoir que pour n suffisamment grand l'espérance de W_n va être proche de a. Cela demande à être précisé.

b) Déterminer $E(W_n)$ pour $n \geq 2$, puis trouver un estimateur sans biais de a. On rappelle un résultat vu auparavant : Lorsque X est une variable aléatoire suivant une loi de Pareto de paramètres a et x_0, la variable aléatoire $Y = \ln\dfrac{X}{x_0}$ suit une loi exponentielle de paramètre a.

---

[2] En procédant comme on le fera dans le paragraphe suivant avec la loi de Pareto(3°-b), on démontre que la suite de variables aléatoires $\dfrac{n-1}{\sum_{i=1}^{n} X_i}$ est un estimateur sans biais de k.

On rappelle aussi qu'une somme de n variables aléatoires obéissant toutes à la même loi exponentielle de paramètre k obéit elle-même à la loi gamma de paramètres n et k.

Posons $Y_i = \ln \frac{X_i}{x_0}$. Puisque les X_i suivent une loi de Pareto de paramètres a et x_0, Y_i suit une loi exponentielle de paramètre a. Dans ces conditions $\sum_{i=1}^{n} \ln \frac{X_i}{x_0} = \sum_{i=1}^{n} Y_i$, et l'on a une somme de variables indépendantes obéissant toutes à la loi exponentielle de paramètre a. On sait qu'alors la variable aléatoire $\sum_{i=1}^{n} Y_i$ a pour densité $d(t) = \frac{a^n}{(n-1)!} e^{-at} t^{n-1}$ pour $t \geq 0$ et 0 sinon. A son tour

$$W_n = \frac{n}{\sum_{i=1}^{n} \ln \frac{X_i}{x_0}} = \frac{n}{\sum_{i=1}^{n} Y_i}$$ a pour espérance :

$E(W_n) = \int_{-\infty}^{+\infty} \frac{n}{t} d(t)\, dt = \frac{n\, a^n}{(n-1)!} \int_0^{+\infty} e^{-at} t^{n-2} dt$, sous réserve que cette intégrale existe. Comme $n \geq 1$, l'intégrale existe en 0, et en $+\infty$, $e^{-at} t^{n-1}$ tend vers 0 et l'intégrale existe encore.

On constate aussi que $\frac{a^{n-1}}{(n-2)!} e^{-at} t^{n-2}$, pour $n \geq 2$, est la densité de $\sum_{i=1}^{n-1} Y_i$, d'où :

$$\int_0^{+\infty} \frac{a^{n-1}}{(n-2)!} e^{-at} t^{n-2} dt = 1.$$

$$E(W_n) = \frac{n\, a^n}{(n-1)!} \int_0^{+\infty} e^{-at} t^{n-2} dt = \frac{n\, a}{n-1} \int_0^{+\infty} \frac{a^{n-1}}{(n-2)!} e^{-at} t^{n-2} dt = \frac{n\, a}{n-1}$$

Pour avoir un estimateur sans biais de a, il suffit de prendre la suite (W_n') avec :

$$W_n' = \frac{n-1}{n} W_n$$

puisque $E(W_n') = \frac{n-1}{n} E(W_n) = \frac{n-1}{n} \frac{n\, a}{n-1} = a$.

4) Loi normale

Pour une variable aléatoire X suivant une loi normale de paramètres m et σ^2, la densité est :

$$d(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Pour le paramètre m, la fonction de vraisemblance s'écrit :

$$L(x_1, x_2, ..., x_n, m) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{(x_1-m)^2}{2\sigma^2} - \frac{(x_2-m)^2}{2\sigma^2} - ... - \frac{(x_n-m)^2}{2\sigma^2}}$$

$$\ln L = -n\ln(\sigma\sqrt{2\pi}\,) - \frac{1}{2\sigma^2}((x_1-m)^2+(x_2-m)^2+...+(x_n-m)^2)$$

Cette quantité, considérée comme une fonction de m, est maximale lorsque la somme $f(m) = (x_1-m)^2+(x_2-m)^2+...+(x_n-m)^2$ est minimale.

$f'(m) = 2(m - x_1 + m - x_2 + \dots + m - x_n) = 2(n\,m - \sum_{i=1}^{n} x_i)$. Cette dérivée s'annule pour :

$m = \frac{\sum_{i=1}^{n} x_i}{n}$ et son signe passe de – à +, ce qui correspond à un minimum pour f.

Prenons comme variable aléatoire la moyenne $Y_n = \frac{\sum_{i=1}^{n} X_i}{n}$.

On vérifie que $E(Y_n) = m$ et que $V(Y_n) = \sigma^2 / n$ qui tend vers 0 pour n infini. La suite de variables (Y_n) constitue un estimateur sans biais et convergent de m.

- Passons à l'estimation du paramètre $T = \sigma^2$. On a comme précédemment

  $$\ln L = -n\ln(\sqrt{2\pi T}\,) - \frac{1}{2T}((x_1-m)^2+(x_2-m)^2+...+(x_n-m)^2)$$

  $$= -\frac{n}{2}\ln(2\pi T) - \frac{1}{2T}((x_1-m)^2+(x_2-m)^2+...+(x_n-m)^2)$$

  En dérivant par rapport à T :

  $(\ln L)' = -\frac{1}{2}\left(\frac{n}{T} - \frac{\sum_{i=1}^{n}(m-x_i)^2}{T^2}\right) = -\frac{1}{2T^2}\left(nT - \sum_{i=1}^{n}(m-x_i)^2\right)$ qui s'annule pour $T = \frac{\sum_{i=1}^{n}(m-x_i)^2}{n}$ en passant du signe + au signe – . En posant $Z_n = \frac{\sum_{i=1}^{n}(m-X_i)^2}{n}$, on vérifie que $E(Z_n) = \sigma^2$ et $V(Z_n) = \frac{V((X-m)^2)}{n}$ qui tend vers 0 pour n infini. La suite (Z_n) constitue bien une suite d'estimateurs sans biais et convergents de σ^2.

La question de confiance

Jusqu'à présent nous avons donné une estimation d'un paramètre, c'est-à-dire une valeur fixe, sans savoir le degré de précision du résultat. Si l'on fait 1000 lancers répétés à pile ou face, cette seule expérience donnera une idée du paramètre p, correspondant à la fréquence de sortie des faces, mais si l'on fait une deuxième expérience de 1000 lancers, on aura un résultat légèrement différent. Aussi, à défaut de chercher une valeur ponctuelle du paramètre, on va se donner un intervalle, si petit soit-il, dans lequel va se trouver le paramètre avec un niveau de confiance donné, par exemple 95% de chances, dès que l'on pratique au moins N expériences élémentaires, cette valeur de N étant à calculer.

Reprenons l'exemple des tirages répétés à pile ou face avec le paramètre p correspondant à la probabilité d'avoir face, 1 – p étant celle d'avoir pile. On prend alors une suite de variables de Bernoulli (X_n) avec $n \geq 1$. On sait que, lorsque l'on fait n tirages, la moyenne expérimentale $S_n = \frac{X_1 + X_2 + ... + X_n}{n}$ constitue un estimateur de p. On sait aussi que plus n est grand, plus S_n va s'approcher de p.

Mais maintenant, donnons-nous une amplitude d'intervalle égale à 2 / 10 par exemple. Il s'agit de déterminer un entier N tel que pour tout $n \geq N$, l'intervalle $[S_n - 1/10,\ S_n + 1/10]$ soit un intervalle de confiance du paramètre p au niveau de 95%. Autrement dit, dès que l'on pratique au moins N lancers, soit n lancers avec $n \geq N$, et que l'on a une valeur expérimentale de S_n obtenue au bout de ces n lancers, on aura p compris entre $S_n - 1/10$ et $S_n + 1/10$ avec une probabilité d'au moins 95%. Il s'agit de trouver cette valeur de N.

On veut avoir $p(S_n - 1/10 \leq p \leq S_n + 1/10) \geq 0{,}95$. Commençons par appliquer l'inégalité de Bienaymé-Tchebichef : avec ε positif aussi petit qu'on veut,

$$p(|S_n - p| \geq \varepsilon) \leq \frac{V(S_n)}{\varepsilon^2}$$

$$\leq \frac{p(1-p)}{n\varepsilon^2} \quad \text{puisque } V(S_n) = \frac{p(1-p)}{n}$$

$$p(W_n' - \frac{1}{10} < p < W_n' + \frac{1}{10}) \geq 1 - \frac{V(W_n')}{n \times 0{,}01}$$

$$1 - p(|S_n - p| < \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$$

$$p(|S_n - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

$$p(-\varepsilon < S_n - p < \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

$$p(-S_n - \varepsilon < -p < -S_n + \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

$$p(S_n + \varepsilon > p > S_n - \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

Dans le cas présent :

$$p(S_n - \frac{1}{10} < p < S_n + \frac{1}{10}) \geq 1 - \frac{p(1-p)}{n \times 0{,}01}$$

Ainsi, dès que $1 - \frac{p(1-p)}{n \times 0{,}01} \geq 0{,}95$, on aura $p(S_n - 0{,}1 < p < S_n + 0{,}1) \geq 0{,}95$

La contrainte $1 - \frac{p(1-p)}{n \times 0{,}01} \geq 0{,}95$ s'écrit

$$\frac{p(1-p)}{0,01n} \leq 0,05 \text{ soit } n \geq \frac{p(1-p)}{0,05 \times 0,01}, \text{ ou } n \geq 2000\,p(1-p)$$

Lorsque p décrit]0 1[, $p(1-p)$ atteint son maximum pour $p = 0,5$, et ce maximum vaut 0,25 : $p(1-p) \leq 0,25$. Dès que l'on aura $n \geq 2000 \times 0,25$, on aura aussi $n \geq 2000\, p\,(1-p)$. Finalement il suffit de prendre $n \geq 500$. Autrement dit, si l'on fait $N = 500$ lancers, aboutissant à S_n, on aura la valeur de p à proximité de S_n, plus précisément avec $S_n - 0,1 < p < S_n + 0,1$, avec cela avec plus de 95 chances sur 100.

Intervalle de confiance pour la loi de Pareto

On veut connaître le paramètre a de la loi de Pareto avec une précision de ± 0,1, avec un niveau de confiance de 95%. On suppose aussi que a est compris entre 1 et 2. Pour cela on reprend la suite (W_n') qui est un estimateur sans biais de a (voir ci-dessus). On admettra que la variance de W_n' est : $V(W_n') = \dfrac{a^2}{n-2}$ lorsque n est > 2. Il s'agit de trouver un entier N tel que pour tout $n \geq N$, [W_n – 0,1 , W_n'+ 0,1] soit un intervalle de confiance de a au niveau de 95%.

En appliquant l'inégalité de Bienaymé-Tchebichef, et en faisant le même calcul que précédemment, on arrive à :

$$p(W_n' - \frac{1}{10} < p < W_n' + \frac{1}{10}) \geq 1 - \frac{V(W_n')}{0,01}$$

$$p(W_n' - \frac{1}{10} < p < W_n' + \frac{1}{10}) \geq 1 - \frac{a^2}{0,01(n-2)}$$

En prenant $1 - \dfrac{100a^2}{(n-2)} \geq 0,95$ on est assuré que $p(W_n' - \dfrac{1}{10} < p < W_n' + \dfrac{1}{10}) \geq 0,95$.

Or $1 - \dfrac{100a^2}{(n-2)} \geq 0,95$ s'écrit $n \geq 2000\, a^2 + 2$. Puisque l'on a supposé que a est compris entre 1 et 2, dès que $n \geq 2000 \times 4 + 2$, on aura $n \geq 2000\, a^2 + 2$. Finalement il convient de prendre $n \geq 8002$, et même sans véritable risque, au vu des majorations que nous avons faites dans le calcul : $n \geq 8000$.

Références :

- *Epreuve de mathématique au concours d'entrée de l'ESSEC*, 2007.
- *Wikipedia*, *Estimateur (statistique)*, avec la bibliographie et les sources jointes.