Généralités

X est une variable aléatoire qui suit une loi binomiale B(n ; f). n est fixé et f inconnue.
E désigne l'ensemble { 0 ; 1 ; … ; n }. t est un réel de] 0 ; 1 [.
Pour tout événement A on note $P_r(A)$ la probabilité de A lorsque f vaut r.

Définition

Soient A et B deux applications de E dans [0 ; 1] . Soit I(X) l'intervalle [A(X) ; B(X)].

On dira que I(X) est un intervalle de confiance de f au risque t lorsque $P_r(r \in I(X)) = 1 - t$.

Remarques :

Soit Ξ l'ensemble des intervalles inclus dans [0 ; 1]
A(X) et B(X) désignent deux variables aléatoires à valeurs dans [0 ; 1].
Ainsi I(X) est une variable aléatoire à valeurs dans Ξ.
L'événement « $r \in I(X)$ » devrait donc plutôt de noter « $I(X) \ni r$ » .

Utilisation d'une loi normale pour la recherche d'un intervalle de confiance.

On suppose donc que les conditions justifient l'utilisation d'une loi normale…

1. Méthode de l'ellipse

Si l'on note Y la variable aléatoire égale à $\frac{X}{n}$, on peut alors considérer que, lorsque f = r, Y suit « à peu près » la loi normale de moyenne r et de variance $\sigma_r^2 = \frac{r\cdot(1-r)}{n}$.

α est le seuil de risque et u le réel tel que $\Pi(u) = 1 - \frac{\alpha}{2}$ où Π est la fonction de répartition de la loi normale centrée réduite. Soit I(X) l'ensemble des r tels que $r - \sigma_r\cdot u \le Y \le r + \sigma_r\cdot u$

I(X) est un intervalle de confiance de r au risque α puisque $P_r\left(-u \le \frac{Y-r}{\sigma_r} \le u\right) = 1 - \alpha$.

Ainsi, $r \in I(X)$ signifie $\quad \mathbf{(r - Y)^2 \le u^2 \cdot \frac{r\cdot(1-r)}{n}} \quad$ **(I)**

En notant $\quad a = 1 + \frac{u^2}{n}$; $b(X) = 2Y + \frac{u^2}{n}$ et $c(X) = Y^2$ cette contrainte devient :

$$\mathbf{a\cdot r^2 - b(X)\cdot r + c(X) \le 0}. \qquad \textbf{(I')}$$

Le discriminant de ce polynôme est $\Delta(X) = \ldots = 4Y\,\frac{u^2}{n}(1-Y) + \frac{u^4}{n^2} = \delta(X)^2$.

Les résultats classiques concernant le signe du trinôme du second degré donnent :

$$\mathbf{I(X) = \left[\frac{b(X) - \delta(X)}{2a}\ ,\ \frac{b(X) + \delta(X)}{2a}\right]}$$

En Remplaçant Y par $\frac{X}{n}$ on obtient :

$$\mathbf{a = 1 + \frac{u^2}{n}} \quad ; \quad \mathbf{b(X) = \frac{2X + u^2}{n}} \quad ; \quad c(X) = \frac{X^2}{n^2} \quad ; \quad \mathbf{\delta(X) = \frac{u}{n}\cdot\sqrt{u^2 + 4X - \frac{4X^2}{n}}}$$

Remarque :

Notons que b(0) = δ(0) et que b(n) + δ(n) = 2a . Ainsi, ces formules fournissent des bornes raisonnables.

C'est la forme de l'inégalité (I) qui est àl'origine du nom de cette méthode …

Correction de continuité : On remplace X par l'intervalle $[X -0,5 \; ; X + 0,5]$

L'encadrement désiré $n\cdot r - u \cdot n\cdot \sigma_r \leq X \leq n\cdot r + u\cdot n\cdot \sigma_r$ devient :

Si	$X \neq 0$ et $X \neq n$	$[X -0,5 \; ; X + 0,5]$	$\subset$	$[n\cdot r - u \cdot n\cdot\sigma_r \; ; n\cdot r + u\cdot n\cdot\sigma_r]$
Si	$X = 0$	$[0 \; ; 0,5]$	$\subset$	$[n\cdot r - u \cdot n\cdot\sigma_r \; ; n\cdot r + u\cdot n\cdot\sigma_r]$
Si	$X = n$	$[n -0,5 \; ; n]$	$\subset$	$[n\cdot r - u \cdot n\cdot\sigma_r \; ; n\cdot r + u\cdot n\cdot\sigma_r]$

Posons : $\mathbf{b_1(z)} = \dfrac{2z + u^2}{n} - \dfrac{u}{n}\cdot\sqrt{u^2 + 4z - \dfrac{4z^2}{n}}$ et $\mathbf{b_2(z)} = \dfrac{2z + u^2}{n} + \dfrac{u}{n}\cdot\sqrt{u^2 + 4z - \dfrac{4z^2}{n}}$

En dérivant, on trouve que b_1' et b_2' sont positives sur $[0 ; n]$ .Ainsi, la réunion des I(z) où z parcours une partie $[z_1 \; ; z_2]$ de $[0 \; ; n]$ est égale à $\left[\dfrac{b_1(z_1)}{2a} , \dfrac{b_2(z_2)}{2a} \right]$

En résumé :

Si	**X = 0**	**on prend**	**I(0) =**	$\left[0 \; , \; \dfrac{b_2(0,5)}{2a} \right]$
Si	**0 < X < n**		**I(X) =**	$\left[\dfrac{b_1(X - 0,5)}{2a} \; , \; \dfrac{b_2(X + 0,5)}{2a} \right]$
S	**X = n**		**I(n) =**	$\left[\dfrac{b_1(n - 0,5)}{2a} \; , \; 1 \right]$

2. Méthodes simplifiées

Dans ces deux cas, il faudra éventuellement tronquer l'intervalle obtenu par calcul pour ne garder que les valeurs effectivement contenues dans $[0 \; ; 1]$ …

a. La plus célèbre

La simplification consiste àremplacer $\sigma_r^2 = \dfrac{r\cdot(1 - r)}{n}$ par son <u>estimateur</u> $s(X)^2 = \dfrac{Y\cdot(1 - Y)}{n - 1}$.

<u>Cette méthode n'est donc pas utilisable lorsque X = 0 ou X = n</u>

L'intervalle de confiance obtenu est $\mathbf{J(X)} = \left[\dfrac{X}{n} - u\cdot s(X) \, , \dfrac{X}{n} + u\cdot s(X) \right]$.

La correction de continuité donne $\mathbf{J(X)} = \left[\dfrac{2X-1}{2n} - u\cdot s(X) \, , \dfrac{2X+1}{2n} + u\cdot s(X) \right]$

b. La plus simple

On peut aussi remplacer σ_r^2 par son majorant $\dfrac{1}{4n}$ …

On obtient alors l'intervalle $\mathbf{J^*(X)} = \left[\dfrac{X}{n} - \dfrac{u}{2\sqrt{n}} , \dfrac{X}{n} + \dfrac{u}{2\sqrt{n}} \right]$

ou, avec correction de continuité : $\mathbf{J^*(X)} = \left[\dfrac{2X-1}{2n} - \dfrac{u}{2\sqrt{n}} , \dfrac{2X+1}{2n} + \dfrac{u}{2\sqrt{n}} \right]$

Cette méthode, bien qu'assez grossière, possède l'avantage de son extrême simplicité…