Espérance d'une loi binomiale

Soit X une variable aléatoire suivant une loi binomiale de paramètres n et p.
L'espérance de X est égale à :

$$E(X)=\sum_{k=0}^{n} kp(X=k)=\sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k}$$
$$=\sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k}$$

car le premier terme est clairement nul.
D'autre part, on a pour k supérieur ou égal à 1 :

$$k\binom{n}{k}=k\frac{n!}{k!(n-k)!}=\frac{n!}{(k-1)!(n-k)!}$$
$$=n\frac{(n-1)!}{(k-1)!(n-k)!}=n\binom{n-1}{k-1}$$

Par conséquent :

$$E(X)=\sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k}=\sum_{k=1}^{n} n\binom{n-1}{k-1}p^k(1-p)^{n-k}$$
$$=np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}$$
$$=np\sum_{K=0}^{n-1}\binom{n-1}{K}p^K(1-p)^{n-1-K}=np(p+1-p)^{n-1}=np$$

Variance d'une loi binomiale

On calcule d'abord $E(X(X-1))$ qui est égale par définition à :

$$E(X(X-1))=\sum_{k=0}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k}$$
$$=\sum_{k=2}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k}.$$

Comme précédemment, on remarque que, pour k supérieur ou égal à 2 :

$$k(k-1)\binom{n}{k}=k(k-1)\frac{n!}{k!(n-k)!}=n(n-1)\frac{(n-2)!}{(k-2)!(n-k)!}$$
$$=n(n-1)\binom{n-2}{k-2}$$

D'où l'on déduit :

$$E(X(X-1))=\sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2}p^k(1-p)^{n-k}$$
$$=n(n-1)p^2\sum_{k=2}^{n}\binom{n-2}{k-2}p^{k-2}(1-p)^{n-k}$$
$$=n(n-1)p^2\sum_{k=2}^{n}\binom{n-2}{k-2}p^{k-2}(1-p)^{n-2-(k-2)}$$
$$==n(n-1)p^2$$

On en déduit que :

$$E(X^2-X)=E(X^2)-E(X)=n(n-1)p^2$$

puis

$$E(X^2)=n(n-1)p^2+E(X)=n(n-1)p^2+np$$

et enfin :

$$V(X)=E(X^2)-(E(X))^2=n(n-1)p^2+np-n^2p^2$$
$$=np-np^2=np(1-p)$$

Théorème de Moivre-Laplace

Si X_n suit la loi binomiale de paramètres n et p, alors la loi centrée réduite associée $R_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$ est telle que :

$$\lim_{n\to+\infty} P(a \le R_n \le b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx$$

pour tout intervalle $[a\,;b]$ de $\mathbb{R}$.

Démonstration

Nous allons considérer le cas où $p = \dfrac{1}{2}$ pour simplifier les notations mais le cas général se démontre de la même façon.

On sait que dans ce cas l'espérance de X_n est égale à $\mu = np = \dfrac{n}{2}$ et que l'écart-type de X_n vaut $\sigma = \sqrt{np(1-p)} = \dfrac{\sqrt{n}}{2}$.

Il s'agit donc de calculer :

$$P(a \le R_n \le b) = P\left(a \le \frac{X_n - \frac{n}{2}}{\frac{\sqrt{n}}{2}} \le b\right) = P\left(\frac{n}{2} + \frac{\sqrt{n}}{2}a \le X_n \le \frac{n}{2} + \frac{\sqrt{n}}{2}b\right)$$

$$= \sum_{k\in A}\binom{n}{k}\frac{1}{2^k}\frac{1}{2^{n-k}} = \sum_{k\in A}\frac{n!}{k!(n-k)!}\frac{1}{2^n} = \sum_{k=k_n}^{l_n}\frac{n!}{k!(n-k)!}\frac{1}{2^n}.$$

où $A = \{0,1,2,...,n\} \cap \left[\dfrac{n}{2} + \dfrac{\sqrt{n}}{2}a\,;\dfrac{n}{2} + \dfrac{\sqrt{n}}{2}b\right] = [k_n\,;l_n]$ où k_n et l_n sont deux entiers naturels dépendant de l'entier n.

Tentons de simplifier le terme de cette somme, en utilisant la formule de Stirling On rappelle qu'elle affirme que :

$$N! \sim N^N e^{-N}\sqrt{2\pi N}$$

au voisinage de l'infini.

Appliquons-la trois fois dans le terme $\dfrac{n!}{k!(n-k)!}\dfrac{1}{2^n}$; remarquons que si n tend vers l'infini, il en est de même de $k \ge \dfrac{n}{2} + \dfrac{\sqrt{n}}{2}a$ et de $n-k \ge \dfrac{n}{2} - \dfrac{\sqrt{n}}{2}b$.

On a donc :

$$\frac{n!}{k!(n-k)!}\frac{1}{2^n} \sim \frac{n^n e^{-n}\sqrt{2\pi n}}{k^k e^{-k}\sqrt{2\pi k}\,(n-k)^{n-k}e^{-(n-k)}\sqrt{2\pi(n-k)}}\frac{1}{2^n}$$

$$= \frac{1}{2^n\sqrt{2\pi}}\sqrt{\frac{n}{k(n-k)}}\frac{n^n}{k^k(n-k)^{n-k}}$$

$$= \frac{1}{2^n\sqrt{2\pi n}}\frac{1}{\sqrt{\frac{k}{n}\left(1-\frac{k}{n}\right)}}\frac{1}{\left(\frac{k}{n}\right)^k\left(1-\frac{k}{n}\right)^{n-k}}$$

Écrivons maintenant k sous la forme $k = \dfrac{n}{2} + x\dfrac{\sqrt{n}}{2}$ avec $x \in [a\,;b]$. Remarquons que k varie entre n_1 et n_2, donc que x varie entre $\dfrac{n_1 - n/2}{\sqrt{n}/2}$ et $\dfrac{n_2 - n/2}{\sqrt{n}/2}$ par pas de $\dfrac{2}{\sqrt{n}}$.

On aura donc :

$$\frac{k}{n} = \frac{1}{2}\left(1 + \frac{x}{\sqrt{n}}\right)$$

$$\text{et } 1 - \frac{k}{n} = \frac{1}{2}\left(1 - \frac{x}{\sqrt{n}}\right).$$

Par conséquent :

$$\frac{n!}{k!(n-k)!}\frac{1}{2^n} \sim \frac{1}{\sqrt{2\pi n}}\frac{2}{\sqrt{1-\frac{x^2}{n}}}\frac{1}{\left(1+\frac{x}{\sqrt{n}}\right)^{\frac{n}{2}+x\frac{\sqrt{n}}{2}}\left(1-\frac{x}{\sqrt{n}}\right)^{\frac{n}{2}-x\frac{\sqrt{n}}{2}}}$$

$$= \frac{1}{\sqrt{2\pi n}}\frac{2}{\sqrt{1-\frac{x^2}{n}}}\frac{1}{\left(1+\frac{x}{\sqrt{n}}\right)^{\frac{n}{2}+x\frac{\sqrt{n}}{2}}\left(1-\frac{x}{\sqrt{n}}\right)^{\frac{n}{2}-x\frac{\sqrt{n}}{2}}}$$

$$= \frac{1}{\sqrt{2\pi n}}\frac{2}{\sqrt{1-\frac{x^2}{n}}}\Psi(x,n)$$

en posant $\psi(x,n) = \dfrac{1}{\left(1+\frac{x}{\sqrt{n}}\right)^{\frac{n}{2}\left(1+\frac{x}{\sqrt{n}}\right)}\left(1-\frac{x}{\sqrt{n}}\right)^{\frac{n}{2}\left(1-\frac{x}{\sqrt{n}}\right)}} = \left(1+\dfrac{x}{\sqrt{n}}\right)^{-\frac{n}{2}\left(1+\frac{x}{\sqrt{n}}\right)}\left(1-\dfrac{x}{\sqrt{n}}\right)^{-\frac{n}{2}\left(1-\frac{x}{\sqrt{n}}\right)}$

Considérons le logarithme népérien de $\Psi(x,n)$:

$$\ln\psi(x,n) = \ln\left(\left(1+\frac{x}{\sqrt{n}}\right)^{-\frac{n}{2}\left(1+\frac{x}{\sqrt{n}}\right)}\left(1-\frac{x}{\sqrt{n}}\right)^{-\frac{n}{2}\left(1-\frac{x}{\sqrt{n}}\right)}\right)$$

$$= -\frac{n}{2}\left(\left(1+\frac{x}{\sqrt{n}}\right)\ln\left(1+\frac{x}{\sqrt{n}}\right)+\left(1-\frac{x}{\sqrt{n}}\right)\ln\left(1-\frac{x}{\sqrt{n}}\right)\right)$$

Un développement limité en $\frac{1}{\sqrt{n}}$ peut être obtenu avec un logiciel de calcul formel (ci-dessous les premiers termes sont corrects) :

taylor(ln(1+x),x,3)	$x-\frac{x^2}{2}+\frac{x^3}{3}$
p(x):=taylor(ln(1+x),x,12)	*Terminé*
$\text{expand}\left(\frac{-n}{2}\cdot\left(\left(1+\frac{x}{\sqrt{n}}\right)\cdot p\left(\frac{x}{\sqrt{n}}\right)+\left(1-\frac{x}{\sqrt{n}}\right)\cdot p\left(\frac{-x}{\sqrt{n}}\right)\right),x\right)$	$\frac{-x^{12}}{132\cdot n^5}-\frac{x^{10}}{90\cdot n^4}-\frac{x^8}{56\cdot n^3}-\frac{x^6}{30\cdot n^2}-\frac{x^4}{12\cdot n}-\frac{x^2}{2}$

On peut donc être sûr que $\ln(\Psi(x,n)) = -\frac{x^2}{2}+O\left(\frac{x^4}{n}\right)$ donc que $\Psi(x,n)\sim e^{-\frac{x^2}{2}}$, un équivalent indépendant de n.

Comme par ailleurs $1-\frac{x^2}{n}\sim 1$ au voisinage de l'infini, on peut écrire que :

$$\frac{n!}{k!(n-k)!}\frac{1}{2^n}\sim\frac{2}{\sqrt{2\pi n}}e^{-\frac{x^2}{2}}$$

On peut ici sommer les équivalents car ils concernent des nombres réels positifs.
Revenons donc au calcul initial :

$$P(a\le R_n\le b)=\sum_{k=k_n}^{l_n}\frac{n!}{k!(n-k)!}\frac{1}{2^n}\sim\frac{1}{\sqrt{2\pi}}\sum_{x=\frac{k_n-n/2}{\sqrt{n}/2}}^{\frac{l_n-n/2}{\sqrt{n}/2}}\frac{2}{\sqrt{n}}e^{-\frac{x^2}{2}}\qquad\textbf{(E)}$$

On se propose de montrer que cette somme peut s'interpréter comme une somme de Riemann entre a et b.

Remarquons que les valeurs de x progressent par pas de $\frac{2}{\sqrt{n}}$.

Dans quel intervalle se trouve la première valeur ? Il faut revenir à la définition de k_n :c'est le plus petit des entiers tels que $k_n\ge\frac{n}{2}+\frac{\sqrt{n}}{2}a$. On a donc :

$$k_n-1<\frac{n}{2}+\frac{\sqrt{n}}{2}a\le k_n$$

ce que l'on peut aussi traduire par :

$$\frac{k_n-1-n/2}{\sqrt{n}/2}<a\le\frac{k_n-n/2}{\sqrt{n}/2}.$$

En d'autres termes : $0\le\frac{k_n-n/2}{\sqrt{n}/2}-a<\frac{2}{\sqrt{n}}$.

La première valeur est dans l'intervalle $\left[a\,;a+\frac{2}{\sqrt{n}}\right]$.

Par suite, la deuxième valeur se trouve dans $\left[a+\frac{2}{\sqrt{n}}\,;a+2\frac{2}{\sqrt{n}}\right]$ et ainsi de suite jusqu'à la dernière valeur.

La somme **(E)** est bien une somme de Riemann, dont on sait qu'elle converge vers $\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx$, ce qu'il fallait démontrer.

Programmation de la loi normale centrée réduite

(Extrait de L. Gacôgne et G. Frugier)

On peut programmer le calcul des valeurs de la fonction de répartition Π en utilisant l'approximation numérique suivante, pour $x \geq 0$.

$$f(x) = (2\,\pi)^{-1/2}\, e^{-x^2/2}$$
$$t = 1/(1 + px)$$
$$Q(x) = f(x)(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5)$$
$$\Pi(x) = 1 - Q(x)$$

$p = 0{,}2316419$
$b_1 = 0{,}319381530$
$b_2 = -\,0{,}356563782$
$b_3 = 1{,}781477937$
$b_4 = -\,1{,}821255978$
$b_5 = 1{,}330274429$

L'erreur commise est inférieure à 10^{-7}. Pour les valeurs négatives de la variable, on utilise la symétrie de la variable.

```
Function Cumnorm(x As Double) As Double
  XAbs = Abs(x)
  If XAbs > 37 Then
    Cumnorm = 0
  Else
    Exponential = Exp(-XAbs ^ 2 / 2)
    If XAbs < 7.07106781186547 Then
      build = 3.52624965998911E-02 * XAbs + 0.700383064443688
      build = build * XAbs + 6.37396220353165
      build = build * XAbs + 33.912866078383
      build = build * XAbs + 112.079291497871
      build = build * XAbs + 221.213596169931
      build = build * XAbs + 220.206867912376
      Cumnorm = Exponential * build
      build = 8.83883476483184E-02 * XAbs + 1.75566716318264
      build = build * XAbs + 16.064177579207
      build = build * XAbs + 86.7807322029461
      build = build * XAbs + 296.564248779674
      build = build * XAbs + 637.333633378831
      build = build * XAbs + 793.826512519948
      build = build * XAbs + 440.413735824752
      Cumnorm = Cumnorm / build
    Else
      build = XAbs + 0.65
      build = XAbs + 4 / build
      build = XAbs + 3 / build
      build = XAbs + 2 / build
      build = XAbs + 1 / build
      Cumnorm = Exponential / build / 2.506628274631
    End If
  End If
  If x > 0 Then Cumnorm = 1 - Cumnorm
End Function
```

FIGURE 2. A double precision univariate normal function. (All variables which are not declared here are set as private variables elsewhere.)

Par définition, $\operatorname{erf}\ \ x\ = \dfrac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$.

Caculons maintenant :

$$P\ \ -u \le X \le u\ \ = \frac{1}{\sqrt{2\pi}}\int_{-u}^{u} e^{-\frac{t^2}{2}}dt = \frac{2}{\sqrt{2\pi}}\int_0^u e^{-\frac{t^2}{2}}dt$$

$$= \frac{2}{\sqrt{2\pi}}\int_0^{u/\sqrt{2}} e^{-y^2}\sqrt{2}dy = \frac{2}{\sqrt{\pi}}\int_0^{u/\sqrt{2}} e^{-y^2}dy = \operatorname{erf}\left(\frac{u}{\sqrt{2}}\right)$$