Eric Depiereux, Benoît DeHertogh, Grégoire Vincke

www.fundp.ac.be/biostats ***Module 50***

50. La loi binomiale

50.1 Principe

Dans l'exemple du nombre de garçons par famille de trois enfants nous venons de voir que la probabilité que (X = x) est égale à « un certain nombre de fois » une probabilité élémentaire (1/8), ce nombre de fois dépendant du nombre de combinaisons qui engendrent un résultat particulier (trois combinaisons donnent un garçon sur 3 enfants).

Notez bien que cette prédiction a priori de la probabilité, non seulement n'est valable que sous certaines hypothèses, mais ne s'applique pas à toutes les variables. La variable X « le nombre d'enfants par famille » ne se prête pas à une modélisation a priori. La seule façon d'en établir la fonction de probabilité est d'estimer celle -ci par l'observation.

On montre en analyse combinatoire que ce nombre de combinaisons se calcule de façon générale par la formule suivante :

$$C_x^n = \frac{n!}{x!(n-x)!}$$ **Équation 50-1**

expression dans laquelle n représente le nombre de réalisations de l'expérience (ici, n = 3 enfants), et x (0 ≤ x ≤ n) le nombre de succès de l'expérience (par exemple x = 1 garçon), ce qui donne

$$C_1^3 = \frac{1.2.3}{1.(1.2)} = 3$$

Cette formule permet de traiter facilement des situations plus complexes : le nombre de façons d'avoir 4 filles dans une famille de 12 enfants est $C_4^{12} = \frac{12!}{4!8!} = 495$

A chacune de ces combinaisons correspond une probabilité élémentaire qui peut s'écrire de façon générale :

P(A1 ∩ A2 ∩ A3…∩ Ai…∩ An)

n représentant le nombre de fois que l'on répète l'expérience.

Si les événements sont indépendants, suivant la loi des probabilités composées, cette probabilité se calcule de la façon suivante :

P(A1)P(A2)P(A3)…P(Ai)…P(An).

Si les événements peuvent se classer en deux catégories exclusives A et A * que nous appellerons succès A et échec A * et que la probabilité de succès π est constante durant l'expérience, la probabilité élémentaire est

$\pi^x(1 - \pi)^{n - x}$

Soit, pour 4 filles parmi 12 enfants, si la probabilité constante d'avoir une fille est 0,5, si les naissances sont indépendantes et si tout individu est classé sans ambiguïté dans l'un des deux sexes : $0{,}5^4(1 - 0{,}5)^8 = 0{,}000244$.

La probabilité de l'ensemble de ces événements élémentaires s'obtient en multipliant la probabilité élémentaire par le nombre de combinaisons qui donnent 4 succès :

$P(X = 4) = C_4^{12}\pi^4(1-\pi)^8 = 0.12$

Cette loi s'appelle la loi binomiale, elle s'applique sous les conditions énoncées :

résultats incompatibles A et A *
n répétitions indépendantes
probabilité π constante

Donc, pour

$$X = Bi(n;\pi)$$

$$C_x^n = \frac{n!}{x!(n-x)!}$$

P(X = x) s'écrit de façon générale

$$P(X = x) = C_x^n \pi^x (1-\pi)^{(n-x)}$$ **Équation 50-2**

Il est peut -être utile de rappeler que x^0 pour x > 0 et 0 ! valent 1.

50.2 Exemple

Supposons la séquence d'un génome[1] dans laquelle les nucléotides[2] A,C,G,T ont une égale probabilité de se rencontrer. Le fait de trouver AAAAA est -il exceptionnel ?

Si les nucléotides se suivent au hasard indépendamment l'un de l'autre, la probabilité recherchée correspond a P(X = 5) pour X = Bi(5 ;0,25) soit

$$P(X=5)=C_5^5 0.25^5(0.75)^0 = 0.25^5 = 0.00097$$

Est -il exceptionnel de trouver plus de 10 occurrences de A dans une séquence de 20 nucléotides ?

La probabilité recherchée correspond à P(X > 10) pour X = Bi(20 ;0,25)soit

$$\begin{aligned}
&(P(X=11)=C_{11}^{20}0.25^{11}(0.75)^9 = 0.003006)\\
&+(P(X=12)=C_{12}^{20}0.25^{12}(0.75)^8 = 0.000751)\\
&+(P(X=13)=0.000154)\\
&...\\
&+(P(X=19)=5.46\times 10^{-11})\\
&+(P(X=20)=0.25^{20} = 9.09\times 10^{-13})\\
&=0.00394
\end{aligned}$$

---

[1] *Génome : ensemble du patrimoine héréditaire porté par les chromosomes.*

[2] *Nucléotide : base azotée. Produit de l'union d'un nucléoside avec l'acide phosphorique, entrant dans la composition des acides nucléiques. Les quatre bases azotées de l'A. D. N. sont :*
G : Guanine (de guano, matière de laquelle fut extraite cette substance).
C : Cytosine.
A : Adénine dérivée de la purine.
T : Thymine.

50.3 Tables et graphique

Ce calcul récurrent est simplifié par l'usage de tables dans lesquelles la formule a été pré -calculée :

X = Bi(20;0.25)		
x	P(X = x)	P(X ≤ x)
0	0,003	0,003
1	0,021	0,024
2	0,067	0,091
3	0,134	0,225
4	0,190	0,415
5	0,202	0,617
6	0,169	0,786
7	0,112	0,898
8	0,061	0,959
9	0,027	0,986
10	0,010	0,996

Tableau 50 -50-1 Extrait de la table de la fonction de densité de probabilité (P(X = x)) et fonction de répartition (P(X ≤ x)) pour X v.a. Binomiale de paramètres n = 20 et π = 0.25. Les valeurs de X se trouvent dans la colonne de gauche.

P(X = x) est obtenue en appliquant la formule vue plus haut. P(X ≤ x) est obtenue en additionnant P(X = x) + P(X = x -1) +… + P(X = 0). Classiquement les tables ne publient que la fonction de répartition, la fonction de densité de probabilité pouvant en être déduite par une seule opération :
P(X = x) = P(X ≤ x) - P(X ≤ x -1).

Il faut noter que pour une variable discontinue :
P(X ≤ x) ≠ P(X < x) ;
P(X > x) = 1 - P(X ≤ x).

Nous trouvons donc la réponse à la question « Est -il exceptionnel de trouver plus de 10 occurrences de A dans une séquence de 20 nucléotides ? »

En une seule opération :
P(X > 10) = 1 -P(X ≤ 10)
= 1 -0,996 = 0,004

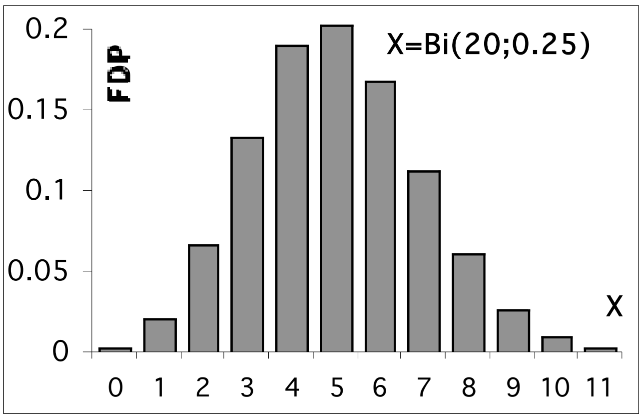

Figure 50 -1 Représentation graphique de la fonction de densité de probabilité pour X v.a.Bi(20 ;0,25).

Les distributions pour lesquelles n est grand et π pas trop éloigné de 0.5 sont des distributions symétriques et tendent vers des distributions normales lorsque n augmente (voir plus loin : variable aléatoire normale).

50.4 Valeurs caractéristiques

Pour X = Bi(n ; π) E(X) = n. π et VAR(X) = n. π.(1 -π)

Soit une v.a. Bi(105 ;0,6), entre quelles bornes s'attend -t -on à trouver 95% des valeurs ? Utilisons l'approximation $\mu \pm 2\sigma$ valable pour les distributions symétriques (n est grand et π pas trop éloigné de 0,5), avec $\sigma = \sqrt{n\pi(1-\pi)} \approx 5$.

La moyenne étant 105x0,6 = 63, 95% des valeurs de X sont donc comprises approximativement entre 53 et 73.

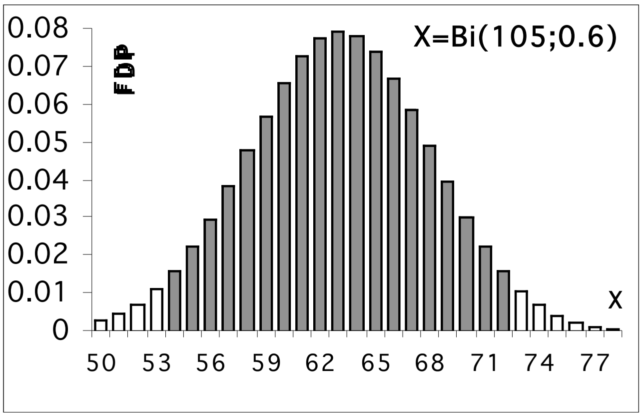

Figure 50 -2 Distribution Bi(105 ;0,6). La surface grisée correspond approximativement à 95% des valeurs de X.

50.5 Dans le tableur Excel

la fonction LOI.BINOMIALE(x; n ; π ;VRAI) pour X v.a. Bi(n ; π)

Dans notre exemple,
LOI.BINOMIALE(10;20;0,25;VRAI)
renvoie 0,996
Prenez garde, les conventions d'Excel varient d'une fonction à l'autre.