Distributions discrètes

Probabilités et statistique pour la biologie (STAT1)

Jacques van Helden

2019-10-03

Distribution de probabilité discrète

On parle de ***distribution discrète*** pour désigner la distribution de probabilité de variables ne pouvant prendre que des valeurs discrètes (par opposition aux distributions continues).

Notes:

- En probabilités la variable observée (x) représente généralement le nombre de succès d'une série d'observations. Elle prend donc généralement des valeurs naturelles ($x \in \mathbb{N}$).
- La probabilité $\mathrm{P}(x)$ prend des valeurs réelles entre 0 et 1, mais sa distribution est discrète puisqu'elle n'est définie que pour des valeurs discrètes de x. On la représente généralement par une fonction en escalier.

Distribution géométrique

Application: temps d'attente jusqu'à la première réalisation d'un évènement au cours d'un schéma de Bernoulli.

Exemples:

- Comptage du nombre de jets d'un dé (x) qui précèdent la première occurrence d'un 6 (l'occurrence n'est pas incluse dans le compte).
- Longueur d'une séquence d'ADN avant la première occurrence d'une cytosine

Fonction de masse de probabilité géométrique

La ***fonction de masse de probabilité*** (***Probability Mass Function***, ***PMF***) indique la probabilité d'observer un résultat élémentaire particulier.

Pour la distribution géométrique, elle indique la probabilité d'observer exactement x échecs avant le premier succès, au cours d'une série d'essais indépendants à probabilité de succès p.

$$\mathrm{P}(X = x) = (1 - p)^x \cdot p$$

Justification:

- Probabilité d'échec pour un essai $= q = 1 - p$ (évènements complémentaires)
- Schéma de Bernoulli $\rightarrow$ les essais sont indépendants $\rightarrow$ probabilité de la série est le produit des probabilités des résultats successifs.

Fonction de masse de probabilité géométrique

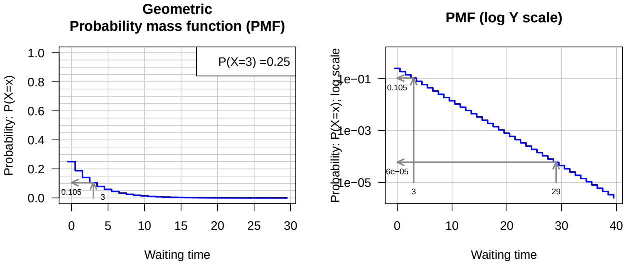


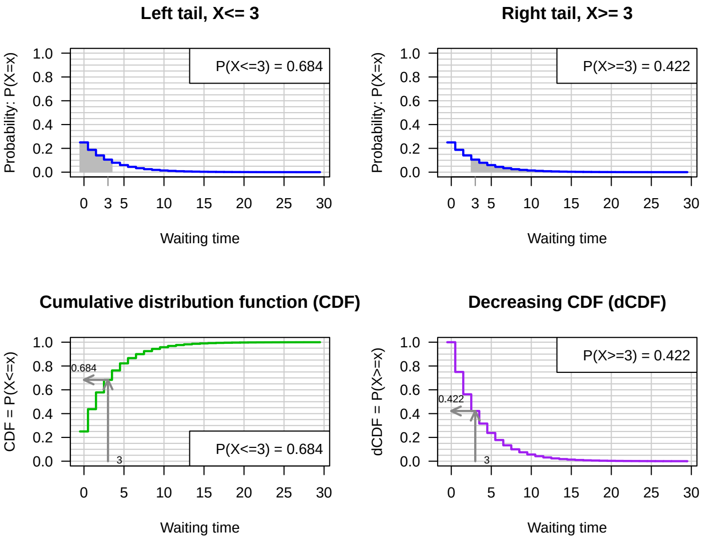
Figure 1: **Fonction de masse de la loi géométrique**. Gauche: ordonnée en échelle logarithmique.

Queues de distribution et fonction de répartition

Les queues de la distribution sont les aires comprises sous la courbe de densité jusqu'à une certaine valeur (***queue gauche***) ou à partir d'une certaine valeur (***queue droite***).

- La ***queue droite*** indique la probabilité d'obtenir un résultat (X) **inférieur ou égal** à une certaine valeur (x): $\mathrm{P}(X \leq x)$.
  - **Définition:** la ***fonction de répartition*** (***Cumulative Density Function***, ***CDF***) $\mathrm{P}(X \leq x)$ indique la probabilité qu'une variable aléatoire X prenne une valeur inférieure ou égale à une valeur donnée (x). Elle correspond à la queue gauche (en incluant la valeur x considérée).
- La **queue gauche** d'une distribution indique la probabilité d'observer un résultat **supérieur ou égal** à une certaine valeur: $\mathrm{P}(X \geq x)$.
  - Note: nous verrons ultérieurement l'utilisation de la queue droite de différentes distributions en tant que probabilité critique

Queues de distribution et fonction de répartition



Distribution binomiale

La **distribution binomiale** indique la probabilité d'observer un certain nombre (x) de succès au cours d'une série de n essais indépendants avec une probabilité de succès p constante (schéma de Bernoulli).

Fonction de masse de probabilité binomiale

$$\mathrm{P}(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} = C_n^x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n}$$

Fonction de répartition binomiale

$$\mathrm{P}(X \geq x) = \sum_{i=x}^{n} P(X = i) = \sum_{i=x}^{n} C_n^i p^i (1-p)^{n-i}$$

Propriétés

Distribution binomiale en *i*

La distribution binomiale peut prendre différentes formes selon les valeurs des paramètres (probabilité de succès p, et nombre d'essais n).

Quand l'espérance $(p \cdot n)$ est inférieure à 1, la distribution est monotone décroissante, et on la qualifie de ***distribution en forme de** i*.

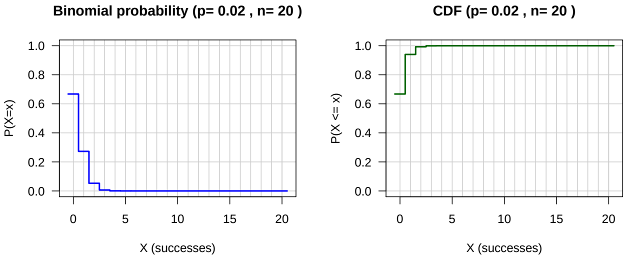


Figure 3: Distribution binomiale en forme de i

Distribution binomiale en cloche asymétrique

Quand la probabilité de succès relativement élevée mais inférieure à 0.5, la distribution prend une forme en cloche asymétrique.

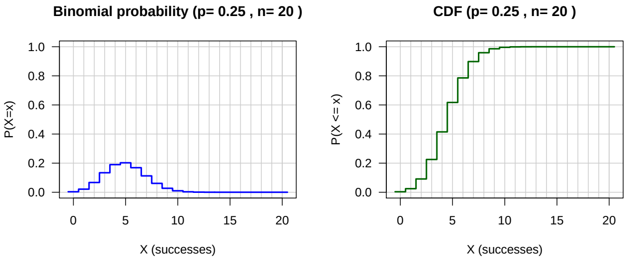


Figure 4: Distribution binomiale en forme de cloche asymétrique.

Distribution binomiale en cloche symétrique

Quand la probabilité de succès vaut 0.5, la distribution prend une forme en cloche symétrique.

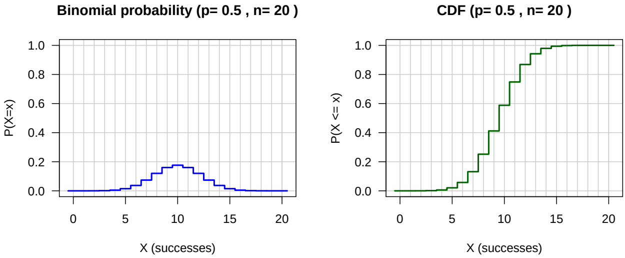


Figure 5: Distribution binomiale en forme de cloche symétrique (p=0.5).

Distribution binomiale en j

Quand la probabilité de succès est proche de 1, la distribution et monotone croissante et on la qualifie de ***distribution en forme de*** j.

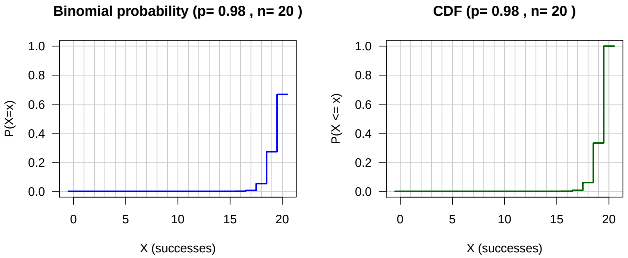


Figure 6: Distribution binomiale en forme de j.

Exemples d'applications de la binomiale

1. **Jeu de dés**: nombre de 6 observés lors d'une série de 10 tirages.
2. **Alignement de séquences**: nombre d'identités entre deux séquences alignées sans gap.
3. **Analyse de motifs**: nombre d'occurrences d'un motif dans un génome.

Note: le recours à la binomiale présuppose un schema de Bernoulli. Pour les exemples 2 et 3 ceci revient à considérer que les nucléotides se succèdent de façon indépendante, ce qui est assez peu réaliste.

Loi de Poisson

La loi de Poisson décrit la probabilité du nombre d'occurrences d'un événement pendant un intervalle de temps fixé, en supposant que le nombre moyen d'événements par unité de temps est constant, et que les événements sont indépendants (les réalisations précédentes n'affectement pas la probabilité des occurrences suivantes).

Fonction de masse

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

- x est le nombre d'événements observés;
- λ (lettre grècque "lambda") représente l'espérance, autrement dit la moyenne attendue pour le nombre d'événements;
- e est la base de l'exponentielle ($e = 2.718$).

Propriétés de la distribution de Poisson

- **Espérance** (nombre de succès attendus au hasard): $< X >= \lambda$ (par construction)
- **variance**: $\sigma^2 = lambda$ (***variance égale à la moyenne!***)
- **Ecart-type**: $\sigma = \sqrt{\lambda}$

Application : mutagenèse

- On soumet une population à un mutagène (agent chimique, irratiations). Chaque individu subit un certain nombre de mutations.
- En tenant compte de la dose de mutagène (temps d'exposition, intensité/concentration), on peut estimer empiriquement le nombre moyen de mutations par individu (*espérance*, λ).
- La loi de Poisson peut être utilisée pour décrire la probabilité d'observer un nombre donné de mutations ($x = 0, 1, 2, ...$).

Expérience historique de Luria-Delbruck (1943)

En 1943, Salvador Luria et Max Delbruck démontrent que les mutations ne sont pas induites par l'agent mutagène, (Luria & Delbruck, 1943, Genetics 28:491–511).

Convergence de la loi binomiale vers la Poisson

A FAIRE

Exercices

- html
- pdf
- Rmd