Formules de probabilités et statistique

Probabilités et statistique pour la biologie (STAT1)

Jacques van Helden

2018-05-10

Table of Contents

# Combinatoire

|  |  |  |
| --- | --- | --- |
| Nom | Conditions | Formule |
| Permutations (factorielle) |  | $n!=n⋅(n−1)⋅…⋅2⋅1$ |
| Arrangements | Sans remise, ordonné | $A\_{n}^{x}=\frac{n!}{(n−x)!}=n⋅(n−1)⋅…⋅(n−x+1)$ |
| Combinaison (*choose*, *coefficient binomial*) | Sans remise, sans ordre | $\left(\genfrac{}{}{0pt}{}{n}{x}\right)=C\_{n}^{x}=\frac{n!}{x!(n−x)!}$ |
|  |  |  |

# Concepts de probabilité

|  |  |  |
| --- | --- | --- |
| Description | Conditions | Formule |
| Définition fréquentielle de la probabilité |  | $P(A)=lim\_{n\rightarrow \infty }\frac{n\_{A}}{n}$ |
|  |  |  |
| Probabilité de non-réalisation |  | $P(¬A)=1−P(A)$ |
|  |  |  |
| Probabilités conditionnelles |  | $P(A∣B)=\frac{P(A∧B)}{P(B)}$ |
|  |  | $P(B∣A)=\frac{P(A∧B)}{P(A)}$ |
|  |  |  |
| Probabilité de $A$ ou $B$ | En général | $P(A∨B)=P(A)+P(B)−P(A∧B)$ |
|  |  |  |
| Probabilité de $A$ ou $B$ | Evénements mutuellement exclusifs | $P(A∨B)=P(A)+P(B)$ |
|  |  |  |
| Probabilité de $A$ et $B$ | En général | $P(A∧B)=P(A)⋅P(B∣A)$ |
|  |  |  |
| Probabilité de $A$ et $B$ | Evénements indépendants | $P(A∧B)=P(A)⋅P(B)$ |
|  |  |  |
| Règle de Bayes |  | $P(A∧B)=P(A)⋅P(B∣A)=P(B)⋅P(A∣B)$ |
|  |  | $⟹P(A∣B)=\frac{P(A)⋅P(B∣A)}{P(B)}$ |
|  |  | $⟹P(B∣A)=\frac{P(B)⋅P(A∣B)}{P(A)}$ |
|  |  |  |

# Distributions de probabilité discrètes

## Géométrique

* Conditions : nombre d’échecs avant le premier succès dans un schéma de Bernoulli
* Densité :

$$P(X=x)=(1−p)^{x}p$$

* Répartition :

$$P(X\leq x)=\sum\_{i=0}^{x}(1−p)^{i}p$$

* Moyenne : $μ\_{G}=(1−p)/p$
* Variance : $σ\_{G}^{2}=\frac{(1−p)}{p^{2}}$

## Binomiale

* Conditions : Nombre de succès au cours d’une série d’essais indépendants avec probabilité constante de succès (Schéma de Bernoulli)
* Densité :

$$P(X=x)=C\_{n}^{x}p^{x}(1−p)^{n−x}$$

* Répartition :

$$P(X\leq x=\sum\_{i=0}^{x}C\_{n}^{i}p^{i}(1−p)^{n−i}$$

* Moyenne : $μ\_{B}=np$
* Variance : $σ\_{B}^{2}=np(1−p)$
* Rapport moyenne/variance: $σ\_{B}^{2}<μ\_{B}$

## Poisson

* Conditions : nombre de succès observés au cours d’un intervalle de temps, en fonction du nombre attendu ($λ$)
* Application : approximation de la binomiale quand $n\rightarrow \infty ,p\rightarrow 0$ et $μ=np$ faible ($μ\_{B}\rightarrow λ$)
* Densité :

$$P(X=x)=\frac{e^{−λ}λ^{x}}{x!}$$

* Répartition :

$$P(X\leq x)=\sum\_{i=0}^{x}\frac{e^{−λ}λ^{i}}{i!}$$

* Moyenne : $μ\_{P}=λ$
* Variance : $σ\_{P}^{2}=λ$
* Rapport moyenne/variance: $σ\_{P}^{2}=μ\_{P}$

## Hypergéométrique

* Conditions : Tirage non ordonné, sans remise dans un ensemble fini avec deux catégories.
* Exemple-type: urne avec boules de deux couleurs
* Densité :

$$P(X=x)=\frac{C\_{m}^{x}C\_{n}^{k−x}}{C\_{m+n}^{k}}$$

* Répartition :

$$P(X\leq x=\sum\_{i=x}^{min(k,m)}\frac{C\_{m}^{i}C\_{n}^{k−i}}{C\_{m+n}^{k}}$$

* Moyenne : $μ\_{H}=k⋅\frac{m}{m+n}$
* Variance : $σ\_{H}^{2}=\frac{k\frac{m}{N}(1−\frac{m}{N})(N−k)}{(N−1)};N=m+n$

# Echantillonnage et estimation

* Les symboles grecs ($μ$, $σ$) correspondent aux statistiques de population, les symboles romains ($\overline{x}$, $s$) aux statistiques d’échantillon.
* L’accent circonflexe ($\hat{}$) indique les estimateurs de paramètres de population calculés à partir de paramètres d’échantillons.

|  |  |
| --- | --- |
| Symbole | Description |
| $N$ | Taille (nombre d’individus) de la population. |
|  |  |
| $μ=\frac{1}{N}\sum\_{i=1}^{N}x\_{i}$ | Moyenne de la population. |
|  |  |
| $σ^{2}=\frac{1}{N}\sum\_{i=1}^{N}(x\_{i}−μ)^{2}=\frac{1}{N}\sum\_{i=1}^{N}x\_{i}^{2}−μ^{2}$ | Variance de la population |
|  |  |
| $σ=\sqrt{σ^{2}}$ | Écart-type de la population |
|  |  |
| $n$ | Effectif (nombre d’individus) de l’échantillon. |
|  |  |
| $\overline{x}=\frac{1}{n}\sum\_{i=1}^{n}x\_{i}$ | Moyenne d’échantillon. |
|  |  |
| $s^{2}=\frac{1}{n}\sum\_{i=1}^{n}(x\_{i}−\overline{x})^{2}=\frac{1}{n}\sum\_{i=1}^{n}x\_{i}^{2}−\overline{x}^{2}$ | Variance de l’échantillon |
|  |  |
| $s=\sqrt{s^{2}}$ | Écart-type de l’échantillon |
|  |  |
| $\hat{σ^{2}}=\frac{n}{n−1}s^{2}=\frac{1}{n−1}\sum\_{i=1}^{n}(x\_{i}−\overline{x})^{2}$ | Estimateur non-biaisé de la variance de la population. |
|  |  |
| $\hat{σ}=\sqrt{\hat{σ}^{2}}$ | Estimateur non-biaisé de l’écart-type de la population. |
|  |  |
| $<σ\_{\overline{X}}>=\frac{\hat{σ}}{\sqrt{n}}$ | Erreur standard: écart-type attendu sur la moyenne d’échantillon. |
|  |  |
| $\overline{x}\pm \frac{\hat{σ}}{\sqrt{(}n)}⋅t\_{1−α/2}^{n−1}$ | Intervalle de confiance autour de la moyenne. |
|  |  |

# Test de comparaison de moyennes

|  |  |
| --- | --- |
| Symbole | Description |
| $μ\_{1},μ\_{2}$ | Moyennes respectives des populations 1 et 2. |
|  |  |
| $σ\_{1},σ\_{2}$ | Écarts-types respectifs des populations 1 et 2. |
|  |  |
| $n\_{1}$, $n\_{2}$ | Effectifs (nombre d’individus) des échantillons prélevés sur les populations 1 et 2. |
|  |  |
| $\overline{x}=\frac{1}{n}\sum\_{i=1}^{n}x\_{i}$ | Formule générale de la moyenne d’échantillon |
|  |  |
| $\overline{x}\_{1},\overline{x}\_{2}$ | Moyennes d’échantillons. |
|  |  |
| $δ=μ\_{2}−μ1$ | Différence entre les moyennes des populations. |
|  |  |
| $d=\hat{δ}=\hat{μ}\_{2}−\hat{μ}\_{1}=\overline{x}\_{2}−\overline{x}\_{1}$ | $d$ = **Taille d’effet**: dans un test de comparaison de moyennes, il s’agit de la différence entre les moyennes d’échantillons, utilisée comme estimateur de $δ$. |
|  |  |
| $s^{2}=\frac{1}{n}\sum\_{i=1}^{n}(x\_{i}−\overline{x})^{2}=\frac{1}{n}\sum\_{i=1}^{n}x\_{i}^{2}−\overline{x}^{2}$ | Formule générale de la variance d’échantillon |
|  |  |
| $s\_{1}^{2},s\_{2}^{2}$ | Variances mesurées sur les échantillons. |
|  |  |
| $\hat{σ}\_{p}=\sqrt{\frac{n\_{1}s\_{1}^{2}+n\_{2}s\_{2}^{2}}{n\_{1}+n\_{2}−2}}$ | Écart-type groupé (*pooled standard deviation*), utilisé comme estimateur de l’écart-type commun des deux populations, en supposant leurs variances égales (hypothèse de travail d’homoscédasticité). |
|  |  |
| $\hat{σ}\_{δ}=\hat{σ}\_{p}\sqrt{\left(\frac{1}{n\_{1}}+\frac{1}{n\_{2}}\right)}$ | Erreur standard sur la différence entre moyennes, en supposant que les populations ont la même variance (test de Student). |
|  |  |
| $t\_{S}=\frac{\hat{δ}}{\hat{σ}\_{δ}}=\frac{\overline{x}\_{2}−\overline{x}\_{1}}{\sqrt{\frac{n\_{1}s\_{1}^{2}+n\_{2}s\_{2}^{2}}{n\_{1}+n\_{2}−2}\left(\frac{1}{n\_{1}}+\frac{1}{n\_{2}}\right)}}$ | Statistique $t$ de Student, $ν=n\_{1}+n\_{2}−2$ d.l. |
|  |  |
| $t\_{W}=\frac{\overline{x}\_{1}−\overline{x}\_{2}}{\sqrt{\frac{\hat{σ}\_{1}^{2}}{n\_{1}}+\frac{\hat{σ}\_{2}^{2}}{n\_{2}}}}$ | Statistique $t$ de Welch, $ν=\frac{\left(\frac{s\_{1}^{2}}{n\_{1}}+\frac{s\_{2}^{2}}{n\_{2}}\right)^{2}}{\frac{s\_{1}^{4}}{n\_{1}^{2}(n\_{1}−1)}+\frac{s\_{2}^{4}}{n\_{2}^{2}(n\_{2}−1)}}$ d.l. |
|  |  |