Probabilité – Statistique ◆ Formulaire

http://www.chez.com/jeyland/iup

Probabilité

Axiomes de Kolmogorov : Une probabilité *P* est une application de $\mathcal{A}$ vers [0,1] vérifiant les 3 axiomes suivants :

1. La probabilité de l'événement certain est 1 :
$$P(\Omega)=1$$
2. Les probabilités de deux évènements disjoints s'ajoutent :
Si $A\cap B=0$,alors $P(A\cup B)=P(A)+P(B)$
3. Si Ω est infini, on complète l'axiome 2 par l'axiome 3 : si $(A_n)_{n\geq 0}$ est une famille d'évènements deux à deux incompatibles (c'est-à-dire, $\forall n,m,\ A_n\cap A_m=0$) alors :
$$P\left(\bigcup_{n=0}^{+\infty}A_n\right)=\sum_{n=0}^{+\infty}P(A_n)$$

Formule de Poincaré : Soient *A* et *B* deux évènements quelconques. On a :
$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

Dénombrement

Règle de multiplication : Lorsqu'une expérience est composée de *m* sous-expériences, et si chacune possède n_k résultats possibles, quels que soient les résultats des autres sous-expériences, alors le nombre total de résultats possibles est :
$$n=n_1n_2...n_m$$

Permutations : Le nombre de permutations d'un ensemble à *n* éléments est :
$$n!$$

Arrangements : Lorsqu'on extrait *k* éléments d'un ensemble de *n* éléments, et que l'ordre dans lequel sont extraits ces éléments *importe*, on parle d'arrangements de *k* éléments parmi *n*.
$$A_n^k=\frac{n!}{(n-k)!}\ \text{avec}\ (k\leq n)$$

Combinaisons : Lorsqu'on extrait *k* éléments d'un ensemble de *n* éléments, et que l'ordre dans lequel sont extraits ces éléments *n'importe pas*, on parle de combinaisons de *k* éléments parmi *n*.
$$C_n^k=\frac{A_n^k}{k!}=\frac{n!}{k!(n-k)!}\ \text{avec}\ (k\leq n)$$

Formule du binôme de Newton : Pour tout réel *x* et *y* et pour tout entier naturel *n*, on a :
$$(x+y)^n=\sum_{k=0}^{n}C_n^k x^x y^{n-k}$$

Probabilités conditionnelles – Indépendance

Probabilité conditionnelle : Soit (Ω,*A*,*P*) un espace probabilisé. Soient *A* et *B* deux évènements aléatoires avec $P(B)\neq 0$. On appelle probabilité conditionnelle de *A* sachant *B*, le rapport :
$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$

Indépendance : Soit (Ω,*A*,*P*) un espace probabilisé. Deux évènements *A* et *B* sont indépendants si et seulement si :
$$P(A\cap B)=P(A)P(B)$$

Incompatibilité : Soit (Ω,*A*,*P*) un espace probabilisé. Deux évènements *A* et *B* sont incompatibles si et seulement si :
$$A\cap B=0$$
Remarque : Deux évènements de probabilité non nulle ne peuvent être indépendants et incompatibles en même temps.

Formule des probabilités totales et composées :
$$P(B)=\sum_i P(B|A_i)P(A_i)$$

Formule de Bayes : Soit un partition $B_1,\dots,B_n$:
$$P(B_i|A)=\frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n}P(A|B_j)P(B_j)}$$

Variables aléatoires

Fonction de répartition : La fonction de répartition *F* d'une variable *X* (discrète ou continue) la caractérise entièrement. Elle possède les propriétés suivantes :

i) *F* est non décroissante.
ii) $\lim_{x\to-\infty}F(x)=0$
iii) $\lim_{x\to+\infty}F(x)=1$
iv) F est continue à droite : $\lim_{\substack{y\to x\\ y>x}}F(y)=F(x)$
v) Le saut de discontinuité de F au point x est égal à $P(X=x)$: $\lim_{\substack{y\to x\\ y>x}}F(y)-\lim_{\substack{y\to x\\ y<x}}F(x)=P(X=x)$

Théorème de caractérisation de la densité : Une fonction *f*(*x*) est la densité d'une certaine variable aléatoire continue *X* ssi :
$$\forall x,\ f(x)\geq 0\ \text{et}\ \int_{-\infty}^{+\infty}f(x)dx=1$$

Modélisation probabiliste discrète :

- Variable aléatoire uniforme discrète

$$P(X=x)=\frac{1}{n+1},k\leq x\leq k+n$$

- Variable aléatoire de Bernoulli

$$P(X=1)=P(''\,succès\,'')$$
$$E[X]=p\,,\ Var(X)=p(1-p)\,.$$

- Variable aléatoire géométrique

$$P(X=k)=(1-p)^{k-1}p,k=1,2,...$$
$$E[X]=\frac{1}{p}\,,\ Var(X)=\frac{1-p}{p^2}\,.$$

- Variable aléatoire binomiale

$$P(X=k)=p^k(1-p)^{n-k}C_n^k,k=0,...,n$$
$$E[X]=np\,,\ Var(X)=np(1-p)\,.$$

- Variable aléatoire de Poisson

$$\lambda>0,P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!},k=0,1,...$$
$$E[X]=\lambda\,,\ Var(X)=\lambda\,.$$

Modélisation probabiliste continue :

- Variable aléatoire uniforme

$$f(x)=c=\frac{1}{b-a}\text{, lorsque } a<x<b\text{ ; sinon } f(x)=0$$
$$F(x)=\frac{x-a}{b-a}$$
$$E[X]=\frac{a+b}{2}\,,\ Var(X)=\frac{(a-b)^2}{12}\,.$$

- Variable aléatoire exponentielle

$f(x) = \lambda e^{-\lambda x}$, lorsque $x > 0$; sinon $f(x) = 0$

$F(x) = 1 - e^{-\lambda x}$

$E[X] = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.

- Variable aléatoire gaussienne

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, pour tout *x*.

$E[X] = 0$, $Var(X) = 1$.

Espérance, Variance

Espérance mathématique :
Si *X* est une variable aléatoire discrète, alors :

$$E[X] = \sum_{x \in \chi} xP(X = x).$$

Si *X* est une variable aléatoire continue, alors :

$$E[X] = \int_R xf(x)dx$$

Variance : Pour une variable aléatoire *X* dont $E[X^2] < \infty$, la variance est :

$$Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

Covariance : La covariance entre *X* et *Y* est la grandeur :

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

Si X et Y sont indépendantes, alors $Cov(X,Y) = 0$.

Coefficient de corrélation linéaire :

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Relation de Cauchy-Schwartz :

$$E[UV]^2 \le E[U^2]E[V^2]$$

Moyenne arithmétique :

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Convergence

Inégalité de Markov : Soit *X* une variable aléatoire positive et d'espérance finie. Alors :

$$P(X \ge a) \le \frac{E[X]}{a}$$

Inégalité de Chebychev : Soit *X* une variable aléatoire d'espérance μ et de variance σ^2 finies. Alors, pour tout *t* positif :

$$P(|X - \mu| \ge t) \le \frac{\sigma^2}{t^2}$$

Loi des grands nombres : Soit une famille de variables aléatoires indépendantes et identiquement distribuées (c'est-à-dire, toutes de même loi de probabilité), d'espérance μ et de variance σ^2 finies. Alors, pour tout ε positif :

$$P(|\overline{X}_n - \mu| \ge \varepsilon) \to 0, \text{ lorsque } n \to +\infty$$

Théorème Central Limite : Soit une famille de variables aléatoires telle que précédemment, et *U* une variable aléatoire gaussienne de paramètres (0,1), indépendante. Alors, on a :

$$P\left(\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \le u\right) \to P(U \le u), \text{ lorsque } n \to +\infty$$

Autre écriture : $P\left(\frac{X - E[X]}{\sqrt{Var(X)}} \le u\right) \to P(U \le u)$

Statistiques descriptives

Caractéristiques à tendance centrale :
- *La moyenne* : C'est la somme de toutes les valeurs divisée par leur nombre.
- *La médiane* : C'est la valeur qui partage les données (ordonnées) en deux groupes d'effectifs égaux.
- *Le mode* : C'est la modalité la plus fréquente (d'un caractère qualitatif).

Caractéristiques de dispersion :
- *L'étendue* : C'est la différence entre la plus petite et la plus grande valeur.
- *L'écart-type* : C'est la racine carrée de la variance, qui est la moyenne des écarts quadratiques (au carré) à la moyenne.
- *L'intervalle inter-quartile* : C'est l'intervalle qui contient 50% des données en en laissant 25% à gauche et 25% à droite.
- *L'intervalle inter-décile* : C'est l'intervalle qui contient 80% des données en en laissant 10% à gauche et 10% à droite.

Estimation

Méthode des moments : On identifie les moments théoriques (espérance mathématique, variance) issus de la loi de probabilité à ceux de l'échantillon (moyenne, variance).

Méthode du maximum de vraisemblance : On cherche à maximiser la fonction suivante en annulant sa dérivée :

$$L(X_1,...X_n;\theta) = \prod_{i=1}^{n} f_\theta(X_i)$$

Intervalle de confiance d'une proportion :

$$\left[\hat{p} - u_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + u_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Intervalle de confiance de la moyenne :

Variance connue : $\left[\overline{X} - u_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

Variance inconnue : $\left[\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right]$

Tests d'hypothèses

Démarche :
1. Définir le cadre statistique (type d'échantillon, loi de probabilité, etc.).
2. Définir mathématiquement H_0 (par exemple, les espérances sont égales).
3. Définir mathématiquement H_1 (par exemple, les espérances sont différentes).
4. Fixer une valeur pour α.
5. En déduire le test et rejeter ou non H_0.

Rejet de H_0 : H_0 sera rejeté si :

$$F > p_0 + u_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}$$

Modélisation bivariable

Coefficient de corrélation linéaire :

$$r = \frac{1/n\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{s_x s_y} \quad \text{où } s_x = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2$$

Régression linéaire : Calcul des coefficients de *ax+b* :

$$b = r\frac{s_y}{s_x}, \; a = \overline{y} - b\overline{x}$$