Probabilités et Statistiques

Introduction aux probabilités à l'analyse combinatoire

❒ **Univers de probabilités** – L'ensemble de toutes les issues possibles d'une expérience aléatoire est appelé l'univers de probabilités d'une expérience aléatoire et est noté S.

❒ **Évènement** – Toute partie E d'un univers est appelé un évènement. Ainsi, un évènement est un ensemble d'issues possibles d'une expérience aléatoire. Si l'issue de l'expérience aléatoire est contenue dans E, alors on dit que E s'est produit.

❒ **Axiomes de probabilités** – Pour chaque évènement E, on note $P(E)$ la probabilité que l'évènement E se produise.

$$(1)\quad \boxed{0\leqslant P(E)\leqslant 1}\qquad (2)\quad \boxed{P(S)=1}\qquad (3)\quad \boxed{P\left(\bigcup_{i=1}^{n}E_i\right)=\sum_{i=1}^{n}P(E_i)}$$

❒ **Permutation** – Une permutation est un arrangement de r objets parmi n objets, dans un ordre donné. Le nombre de tels arrangements est donné par $P(n,r)$, défini par :

$$\boxed{P(n,r)=\frac{n!}{(n-r)!}}$$

❒ **Combinaison** – Une combinaison est un arrangement de r objets parmi n objets, où l'ordre ne compte pas. Le nombre de tels arrangements est donné par $C(n,r)$, défini par :

$$\boxed{C(n,r)=\frac{P(n,r)}{r!}=\frac{n!}{r!(n-r)!}}$$

Remarque : on note que pour $0\leqslant r\leqslant n$, on a $P(n,r)\geqslant C(n,r)$.

Probabilité conditionnelle

❒ **Théorème de Bayes** – Pour des évènements A et B tels que $P(B)>0$, on a :

$$\boxed{P(A|B)=\frac{P(B|A)P(A)}{P(B)}}$$

Remarque : on a $P(A\cap B)=P(A)P(B|A)=P(A|B)P(B)$.

❒ **Partition** – Soit $\{A_i,i\in[\![1,n]\!]\}$ tel que pour tout i, $A_i\neq\varnothing$. On dit que $\{A_i\}$ est une partition si l'on a :

$$\boxed{\forall i\neq j, A_i\cap A_j=\emptyset\quad\text{et}\quad\bigcup_{i=1}^{n}A_i=S}$$

Remarque : pour tout évènement B dans l'univers de probabilités, on a $P(B)=\displaystyle\sum_{i=1}^{n}P(B|A_i)P(A_i)$.

❒ **Formule étendue du théorème de Bayes** – Soit $\{A_i,i\in[\![1,n]\!]\}$ une partition de l'univers de probabilités. On a :

$$\boxed{P(A_k|B)=\frac{P(B|A_k)P(A_k)}{\displaystyle\sum_{i=1}^{n}P(B|A_i)P(A_i)}}$$

❒ **Indépendance** – Deux évènements A et B sont dits indépendants si et seulement si on a :

$$\boxed{P(A\cap B)=P(A)P(B)}$$

Variable aléatoires

❒ **Variable aléatoire** – Une variable aléatoire, souvent notée X, est une fonction qui associe chaque élement de l'univers de probabilité à la droite des réels.

❒ **Fonction de répartition** – La fonction de répartition F (en anglais *CDF - Cumulative distribution function*), qui est croissante monotone et telle que

$$\lim_{x\to-\infty}F(x)=0\quad\text{et}\quad\lim_{x\to+\infty}F(x)=1$$

est définie de la manière suivante :

$$\boxed{F(x)=P(X\leqslant x)}$$

Remarque : on a $P(a<X\leqslant B)=F(b)-F(a)$.

❒ **Densité de probabilité** – La densité de probabilité f (en anglais *PDF - Probability density function*) est la probabilité que X prenne des valeurs entre deux réalisations adjacentes d'une variable aléatoire.

❒ **Relations vérifiées par les PDF et CDF** – Voici les propriétés importantes à savoir dans les cas discret (D) et continu (C).

Case	**CDF F**	**PDF f**	**Propriétés du PDF**
(D)	$F(x)=\displaystyle\sum_{x_i\leqslant x}P(X=x_i)$	$f(x_j)=P(X=x_j)$	$0\leqslant f(x_j)\leqslant 1$ and $\displaystyle\sum_j f(x_j)=1$
(C)	$F(x)=\displaystyle\int_{-\infty}^{x}f(y)dy$	$f(x)=\dfrac{dF}{dx}$	$f(x)\geqslant 0$ and $\displaystyle\int_{-\infty}^{+\infty}f(x)dx=1$

❒ **Variance** – La variance d'une variable aléatoire, souvent notée $\mathrm{Var}(X)$ ou σ^2, est une mesure de la dispersion de ses fonctions de distribution. Elle est déterminée de la manière suivante :

$$\boxed{\mathrm{Var}(X)=E[(X-E[X])^2]=E[X^2]-E[X]^2}$$

❒ **Écart-type** – L'écart-type d'une variable aléatoire, souvent notée σ, est une mesure de la dispersion de sa fonction de distribution, exprimée avec les même unités que la variable aléatoire. Il est déterminé de la manière suivante :

$$\boxed{\sigma=\sqrt{\textrm{Var}(X)}}$$

❒ **Espérance et moments de la distribution** – Voici les expressions de l'espérance $E[X]$, l'espérance généralisée $E[g(X)]$, $k^{i\grave{e}me}$ moment $E[X^k]$ et fonction caractéristique $\psi(\omega)$ dans les cas discret et continu.

Case	$E[X]$	$E[g(X)]$	$E[X^k]$	$\psi(\omega)$
(D)	$\displaystyle\sum_{i=1}^n x_if(x_i)$	$\displaystyle\sum_{i=1}^n g(x_i)f(x_i)$	$\displaystyle\sum_{i=1}^n x_i^kf(x_i)$	$\displaystyle\sum_{i=1}^n f(x_i)e^{i\omega x_i}$
(C)	$\displaystyle\int_{-\infty}^{+\infty}xf(x)dx$	$\displaystyle\int_{-\infty}^{+\infty}g(x)f(x)dx$	$\displaystyle\int_{-\infty}^{+\infty}x^kf(x)dx$	$\displaystyle\int_{-\infty}^{+\infty}f(x)e^{i\omega x}dx$

Remarque : on a $e^{i\omega x}=\cos(\omega x)+i\sin(\omega x)$.

❒ **Transformation de variables aléatoires** – Soit X,Y des variables liées par une certaine fonction. En notant f_X et f_Y les fonctions de distribution de X et Y respectivement, on a :

$$\boxed{f_Y(y)=f_X(x)\left|\frac{dx}{dy}\right|}$$

❒ **Loi d'intégration de Leibniz** – Soit g une fonction de x et potentiellement c, et a,b, les limites de l'intervalle qui peuvent dépendre de c. On a :

$$\boxed{\frac{\partial}{\partial c}\left(\int_a^bg(x)dx\right)=\frac{\partial b}{\partial c}\cdot g(b)-\frac{\partial a}{\partial c}\cdot g(a)+\int_a^b\frac{\partial g}{\partial c}(x)dx}$$

❒ **Inégalité de Tchebychev** – Soit X une variable aléatoire de moyenne μ. Pour $k,\sigma>0$, on a l'inégalité suivante :

$$\boxed{P(|X-\mu|\geqslant k\sigma)\leqslant\frac{1}{k^2}}$$

Variables aléatoires conjointement distribuées

❒ **Densité conditionnelle** – La densité conditionnelle de X par rapport à Y, souvent notée $f_{X|Y}$, est définie de la manière suivante :

$$\boxed{f_{X|Y}(x)=\frac{f_{XY}(x,y)}{f_Y(y)}}$$

❒ **Indépendance** – Deux variables aléatoires X et Y sont dits indépendantes si l'on a :

$$\boxed{f_{XY}(x,y)=f_X(x)f_Y(y)}$$

❒ **Densité marginale et fonction de répartition** – À partir de la densité de probabilité f_{XY}, on a :

Cas	**Densité marginale**	**Fonction de répartition**
(D)	$\displaystyle f_X(x_i)=\sum_j f_{XY}(x_i,y_j)$	$\displaystyle F_{XY}(x,y)=\sum_{x_i\leqslant x}\sum_{y_j\leqslant y}f_{XY}(x_i,y_j)$
(C)	$\displaystyle f_X(x)=\int_{-\infty}^{+\infty}f_{XY}(x,y)dy$	$\displaystyle F_{XY}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{XY}(x',y')dx'dy'$

❒ **Covariance** – On définit la covariance de deux variables aléatoires X et Y, que l'on note σ_{XY}^2 ou plus souvent $\textrm{Cov}(X,Y)$, de la manière suivante :

$$\boxed{\textrm{Cov}(X,Y)\triangleq\sigma_{XY}^2=E[(X-\mu_X)(Y-\mu_Y)]=E[XY]-\mu_X\mu_Y}$$

❒ **Corrélation** – En notant σ_X,σ_Y les écart-types de X et Y, on définit la corrélation entre les variables aléatoires X et Y, que l'on note ρ_{XY}, de la manière suivante :

$$\boxed{\rho_{XY}=\frac{\sigma_{XY}^2}{\sigma_X\sigma_Y}}$$

Remarques : on note que pour toute variable aléatoire X, Y, on a $\rho_{XY}\in[-1,1]$. Si X et Y sont indépendants, alors $\rho_{XY}=0$.

❒ **Distributions importantes** – Voici les distributions importantes à savoir :

Type	**Distribution**	**PDF**	$\psi(\omega)$	$E[X]$	$\textrm{Var}(X)$
(D)	$X\sim\mathcal{B}(n,p)$ Binomial	$P(X=x)=\displaystyle\binom{n}{x}p^xq^{n-x}$ $x\in[\![0,n]\!]$	$(pe^{i\omega}+q)^n$	np	npq
	$X\sim\textrm{Po}(\mu)$ Poisson	$P(X=x)=\displaystyle\frac{\mu^x}{x!}e^{-\mu}$ $x\in\mathbb{N}$	$e^{\mu(e^{i\omega}-1)}$	μ	μ
(C)	$X\sim\mathcal{U}(a,b)$ Uniforme	$f(x)=\displaystyle\frac{1}{b-a}$ $x\in[a,b]$	$\displaystyle\frac{e^{i\omega b}-e^{i\omega a}}{(b-a)i\omega}$	$\displaystyle\frac{a+b}{2}$	$\displaystyle\frac{(b-a)^2}{12}$
	$X\sim\mathcal{N}(\mu,\sigma)$ Gaussien	$f(x)=\displaystyle\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $x\in\mathbb{R}$	$e^{i\omega\mu-\frac{1}{2}\omega^2\sigma^2}$	μ	σ^2
	$X\sim\textrm{Exp}(\lambda)$ Exponentiel	$f(x)=\lambda e^{-\lambda x}$ $x\in\mathbb{R}_+$	$\displaystyle\frac{1}{1-\frac{i\omega}{\lambda}}$	$\displaystyle\frac{1}{\lambda}$	$\displaystyle\frac{1}{\lambda^2}$

Estimation des paramètres

❒ **Échantillon aléatoire** – Un échantillon aléatoire est une collection de n variables aléatoires $X_1,...,X_n$ qui sont indépendantes et identiquement distribuées avec X.

❒ **Estimateur** – Un estimateur est une fonction des données qui est utilisée pour trouver la valeur d'un paramètre inconnu dans un modèle statistique.

❒ **Biais** – Le biais d'un estimateur $\hat{\theta}$ est défini comme étant la différence entre l'espérance de la distribution de $\hat{\theta}$ et de la valeur vraie, i.e. :

$$\boxed{\textrm{Bias}(\hat{\theta})=E[\hat{\theta}]-\theta}$$

Remarque : un estimateur est dit non biaisé lorsque l'on a $E[\hat{\theta}]=\theta$.

❒ **Moyenne empirique et variance empirique** – La moyenne empirique et la variance empirique d'un échantillon aléatoire sont utilisées pour estimer la valeur vraie μ et la variance vraie σ^2 d'une distribution, notés $\overline{X}$ et s^2 et sont définies de la manière suivante :

$$\boxed{\overline{X}=\frac{1}{n}\sum_{i=1}^nX_i}\quad\textrm{and}\quad\boxed{s^2=\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\overline{X})^2}$$

❒ **Théorème de la limite centrale** – Soit un échantillon aléatoire $X_1,...,X_n$ suivant une distribution donnée de moyenne μ et de variance σ^2, alors on a :

$$\boxed{\overline{X}\underset{n\rightarrow+\infty}{\sim}\mathcal{N}\left(\mu,\frac{\sigma}{\sqrt{n}}\right)}$$