Rappels de cours de Probabilités

Chapitre 1 : Élément du calcul des probabilités

Si A et B sont deux évènements (*ex :* avoir au moins un patient guéri dans l'échantillon, n'avoir aucune guérison...)

- On appelle **intersection** des évènements A et B l'ensemble des résultats de l'expérience qui réalise A **et** B. On le note $A \cap B$.
- L'**union** des évènements A et B est l'ensemble des résultats de l'expérience qui réalisent A, B ou les deux. On le note $A \cup B$.
- Si la réalisation de l'évènement A implique la non-réalisation de de l'évènement B, alors A et B sont **incompatibles** et $A \cap B = \varnothing$.
- L'évènement complémentaire de A, noté $\overline{A}$ est l'ensemble des résultats de l'expérience qui ne réalisent pas A.
- $P(A|B)$ est la probabilité de l'évènement A avec une information supplémentaire : l'évènement B s'est réalisé. On parle de « A sachant B ».

$$P(\overline{A}) = 1 - P(A) \qquad P(\varnothing) = 0$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad P(A|B) = \frac{P(A \cap B)}{P(B)} \overset{\textbf{Bayes}}{=} \frac{P(B|A)P(A)}{P(B)}$$

$$P(A \cup (B \cap C)) = P((A \cup B) \cap (A \cup C)) \qquad P(A \cap (B \cup C)) = P((A \cap B) \cup (A \cap C))$$

$$P(A \cap B) = P(B \cap A) = P(A|B)P(B) = P(B|A)P(A) \qquad P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}$$

$$P(A \cap B) = P(B)P(A) \Rightarrow A \text{ et } B \text{ sont indépendants} \qquad P(B|A) = P(B|\overline{A}) = P(B) \Rightarrow A \text{ et } B \text{ sont indépendants}$$

Probabilités totales : $P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$ avec (B_i) une partition de l'univers.

Chapitre 2 : Variables aléatoires

Associer à chaque évènement élémentaire d'une expérience aléatoire un nombre permet de définir une **variable aléatoire**.

- Variable aléatoire X discrète : les valeurs prises par X sont finies ou infinies et dénombrables. La loi de probabilité de X est donnée par : $\begin{cases} x_1, x_2, ...\text{(réalisations)} \\ p_1 = P(X = x_1), p_2, ...\text{(probabilités associées)} \end{cases}$
- Variable aléatoire X continue : X peut prendre théoriquement un nombre infini de valeurs réelles qui forment un ensemble continu (*ex* : la taille...). Elle admet une **densité de probabilité** f_X et une **fonction de répartition** F.

$$F(x) = P(X \leq x) \qquad \int_{-\infty}^{+\infty} f_X(x)dx = 1 \text{ et } \forall x, f_X(x) \geq 0$$

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx = F(b) - F(a)$$

- **Espérance** : $E(X) = \mu = \underbrace{\sum x_i \ P(X = x_i)}_{\text{cas discret}}$ ou $E(X) = \underbrace{\int_{-\infty}^{+\infty} x f_X(x)dx}_{\text{cas continu}}$
- **Variance** : $V(X) = E(X^2) - (E(X))^2 = \underbrace{\left(\sum x_i^2 \ P(X = x_i)\right) - \mu^2}_{\text{cas discret}}$ ou $V(X) = \underbrace{\int_{-\infty}^{+\infty} x^2 f_X(x)dx - \mu^2}_{\text{cas continu}}$

Calculs sur la variance et l'espérance : soient a et b deux réels, X et Y deux variables aléatoires.

$$E(aX + b) = aE(X) + b \qquad Var(aX + b) = a^2 Var(X)$$

$$E(X + Y) = E(X) + E(Y) \qquad Var(X + Y) = Var(X) + Var(Y) + 2cov(X, Y)$$

$$E(X - Y) = E(X) - E(Y) \qquad cov(X, Y) = E(XY) - E(X)E(Y)$$

$$cov(X, Y) = 0 \text{ si } X \text{ et } Y \text{ sont indépendantes}$$

$$cov(X, X) = Var(X)$$

Chapitres 3 et 4 : Loi binomiale et loi de Poisson

• Épreuve de **Bernoulli** : expérience aléatoire ayant deux résultats possibles : *succès* avec une probabilité associée p et *échec* avec une probabilité associée $1-p$.

Réalisations	Probabilités	Espérance	Variance
$X=0$ ou $X=1$	$P(X=1)=p$ et $P(X=0)=1-p$	$E(X)=p$	$Var(X)=p(1-p)$

• Distribution binomiale : somme de n variables de Bernoulli, on note $X \sim \mathcal{B}(n,p)$

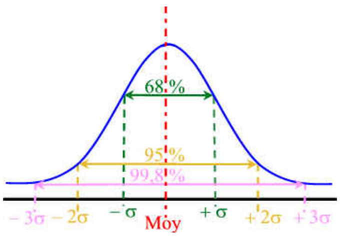
Probabilités	Espérance	Variance
$P(X=k)=\dfrac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$ $\rightarrow$ probabilité d'avoir k succès parmi n essais	$E(X)=np$	$Var(X)=np(1-p)$

• Distribution de Poisson : décrit le comportement du nombre d'évènements se produisant dans un intervalle de temps fixé, permet de caractériser les évènements rares. On note $X \sim \mathcal{P}(\lambda)$ où λ est le nombre moyen d'occurrences dans un intervalle fixé.

Probabilités	Espérance	Variance
$P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$ $\rightarrow$ probabilité d'observer k évènements	$E(X)=\lambda$	$Var(X)=\lambda$

Chapitre 5 : Loi Normale, TCL et lois associées

• **Loi Normale** : $X \sim \mathcal{N}(\mu,\sigma^2)$, la variable X a une probabilité de 95% d'être comprise entre $\mu+2\sigma$ et $\mu-2\sigma$.



- Sa densité est symétrique par rapport à l'axe vertical $y=\mu$
- Par symétrie de sa densité, on déduit $F(-z)=1-F(z) \Leftrightarrow P(X\leq -z)=P(X\geq z)$
- Pour $X\sim\mathcal{N}(\mu,\sigma^2)$ on a $Z=\frac{X-\mu}{\sigma}\sim\mathcal{N}(0,1)$

• **Théorème Limite Centrale** : Soient $X_1, X_2, .., X_n$ des variables aléatoires de même loi, indépendantes, de moyenne μ et de variance σ^2. Pour n suffisamment grand, la moyenne de ces variables a une distribution normale, peu importe la loi suivie par ces variables.

$$\overline{X}=\frac{1}{n}\sum X_i \sim \mathcal{N}\left(\mu,\frac{\sigma^2}{n}\right) \Rightarrow \frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}\sim\mathcal{N}(0,1)$$

• **Loi du Khi-deux (χ^2)** : Soient $X_1, X_2, .., X_n$ n variables normales centrées réduites indépendantes, la somme de leur carré suit une loi du χ^2 à n degrés de liberté (ddl). On note $Z=X_1^2+X_2^2+...+X_n^2\sim\chi^2_n$. Plus généralement, si $X_1, X_2, .., X_n$ sont n variables normales $\mathcal{N}(\mu,\sigma^2)$ indépendantes alors on a :

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2_n$$

• **Loi de Student** : Soient $Z\sim\mathcal{N}(0,1)$ et $U\sim\chi^2_k$ deux variables aléatoires indépendantes.

$$T=\frac{Z}{\sqrt{\frac{U}{k}}}\sim \text{ loi de Student à } k \text{ degrés de liberté}$$

$\rightarrow$ voir le test de Student pour une application pratique.

Chapitre 6 : Estimation

Soient $X_1, X_2, .., X_n$ des variables aléatoires i.i.d. de moyenne μ et de variance σ^2.

- Estimation de μ

$$\widehat{\mu} = \overline{X} = \frac{\sum_{i=1}^n X_i}{n} \qquad \begin{array}{l} E(\overline{X}) = \mu \\ Var(\overline{X}) = \frac{\sigma^2}{n} \end{array} \qquad \overline{X} \text{ est un estimateur sans biais (et convergent) de } \mu$$

- Estimation de σ

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1} = \frac{1}{n-1}\left(\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right) \qquad \underbrace{\begin{array}{l} E(\widehat{\sigma}^2) = \sigma^2 \\ Var(\widehat{\sigma}^2) = \dfrac{2\sigma^4}{n-1} \end{array}}_{\text{pour} X \sim \mathcal{N}(\mu,\sigma^2)} \qquad \widehat{\sigma}^2 \text{ est un estimateur sans biais (et convergent) de } \sigma^2$$

Biais d'un estimateur : On définit le biais de l'estimateur $\widehat{\theta}$ par $B(\widehat{\theta}) = E(\widehat{\theta}) - \theta \Rightarrow \widehat{\theta}$ est sans biais si $E(\widehat{\theta}) = \theta$

Intervalles de confiance au niveau $1 - \alpha$

- d'une moyenne (variance connue) : $(X_n)_{n>1}$ n normales, $n > 30$ et $\varepsilon_{\alpha/2}$ quantile d'une loi normale

$$IC_{1-\alpha}(\mu) = \left[\overline{X} \pm \varepsilon_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- d'une moyenne (variance inconnue) : $(X_n)_{n>1}$ n normales, $n > 30$ et $t_{\alpha/2}^{n-1}$ quantile d'une loi de Student à $n-1$ ddl.

$$IC_{1-\alpha}(\mu) = \left[\overline{X} \pm t_{\alpha/2}^{n-1}\sqrt{\frac{\widehat{\sigma}^2}{n}}\right]$$

- d'un pourcentage : $(X_n)_{n>1}$ n Bernoulli de paramètre θ et $\varepsilon_{\alpha/2}$ quantile d'une loi normale

$$IC_{1-\alpha}(\theta) = \left[\overline{X} \pm \varepsilon_{\alpha/2}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right]$$

Chapitre 7 : Maximum de vraisemblance (Facultatif)

On cherche à écrire la vraisemblance des données observées en fonction du (ou des) paramètre(s) d'intérêt, puis à la maximiser afin d'estimer ce (ces) paramètre(s).

• **Écriture de la vraisemblance**

Exemple : Pièce de monnaie avec pile et face. Je lance 10 fois ma pièce, je tombe deux fois sur pile. Je cherche le meilleur estimateur $\widehat{p}$ pour décrire ma probabilité de tomber sur pile. Si je choisis $\widehat{p} = 0,6$, je suis loin de la réalité et la probabilité d'obtenir des données identiques à mon expérience est faible. Un $\widehat{p} = 0,6$ est peu **vraisemblable**. Dans ce cas, $\widehat{p} = 0,2$ est plus vraisemblable. C'est l'**estimateur du maximum de vraisemblance**.

— Vraisemblance d'une observation discrète : $P_\theta(X = x) \longleftarrow$ probabilité où θ apparait
— Vraisemblance d'un observation continue : $f_\theta(x) \longleftarrow$ densité de X où θ apparait
— Vraisemblance d'un échantillon : $P_\theta(X = x_1) \times P_\theta(X = x_2) \times ... \times P_\theta(X = x_n)$ dans le cas discret et $f_\theta(x_1) \times f_\theta(x_2) \times .... \times f_\theta(x_n) = \prod_{i=1}^n f_\theta(x_i)$ dans le cas continu

• **Maximisation de la vraisemblance** : on cherche le $\widehat{\theta}$ tel que la dérivée de la vraisemblance s'annule (dérivée nulle en $\widehat{\theta} \Rightarrow$ changement de variation de la fonction $\Rightarrow$ présence d'un maximum (où d'un minimum) en $\widehat{\theta}$)
— passage à la log-vraisemblance (dérivation plus simple) $l = log(L)$
— estimation $\widehat{\theta}$ du maximum de vraisemblance : solution de l'équation $\frac{\partial l}{\partial \theta} = 0$

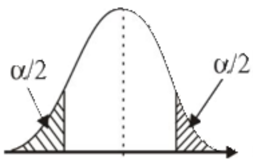
Exemple : Loi de Poisson $P(X = k) = \frac{\theta^x}{x!}e^{\theta}$. Pour n observations indépendantes, la vraisemblance de l'échantillon est $L(\theta) = \frac{\theta^{\Sigma x_i}}{\prod x_i!}e^{-n\theta}$, la log-vraisemblance est $l(\theta) = \sum x_i ln(\theta) - n\theta + cste$, sa dérivée est $\frac{\partial l}{\partial \theta} = \frac{\sum x_i}{\theta} - n$ et $\widehat{\theta}_{MV} = \frac{\sum x_i}{n}$

Chapitre 8 : Tests d'hypothèses

• Avant de faire un test statistique, il est nécessaire de :
— poser l'**hypothèse** nulle H_0 et l'hypothèse alternative H_1,
— décider de la valeur du **risque** de première espèce α.

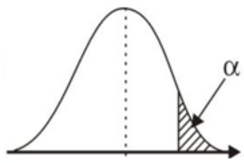
• **Risque** α : probabilité de rejeter H_0 si H_0 est vraie, $P_{H_0}(H_0 \in Rejet)$
Risque β : probabilité de ne pas rejeter H_0 si H_1 est vraie, $P_{H_1}(H_0 \notin Rejet)$
Puissance : probabilité de rejeter H_0 si H_1 est vraie $= 1-\beta$, $P_{H_1}(H_0 \in Rejet)$ $\Rightarrow$ on cherche à ce quelle soit la plus grande possible

Test bilatéral



ex : $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$
Ici on prend $\alpha/2$ de chaque côté
ce qui fait α en tout

Test unilatéral



ex : $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 > \mu_2$
Pour un risque unilatéral, je prends
α du côté qui m'intéresse

• Comparaison d'un **pourcentage observé** p_0 à un pourcentage théorique p, pour np et $n(1-p) \geq 5$ (pour que l'approximation par la loi normale soit possible).

$$\varepsilon = \frac{|p_0 - p|}{\sqrt{\frac{p(1-p)}{n}}}$$

— si $\varepsilon < 1,96$ $\rightarrow$ la différence n'est pas significative à 5%
— si $\varepsilon \geq 1,96$ $\rightarrow$ la différence est significative

• Comparaison d'une **moyenne observée** m à une **moyenne théorique** μ

$$t = \frac{|m - \mu|}{\sqrt{\frac{\widehat{\sigma}^2}{n}}}$$

avec $m = \frac{\sum x_i}{n}$.

1. Distribution normale ou grands échantillons de distribution quelconque avec variance **inconnue**
$\rightarrow$ il faut comparer t à la valeur lue dans la table de la loi de Student à $n-1$ ddl pour une valeur fixée de α
2. Distribution normale ou grands échantillons de distribution quelconque avec variance **connue**
$\rightarrow$ S^2 vaut donc σ^2 et il faut comparer t à la valeur lue dans la table de la loi normale $\mathcal{N}(0,1)$ pour une valeur fixée de α