

III Probabilités et statistiques

Ω ensemble fondamental "tous résultats possibles d'une expérience aléatoire"
un peut pas faire de prévision

$E \subseteq \Omega$ Evénement : sous ensemble de Ω

Occurrence L'événement E s'est produit si : l'issue de l'épreuve $\in E$.

$A \cup B$ Evénement A "ou" B

$\emptyset \subseteq \Omega$ Evénement impossible

$A \subseteq B$ A implique l'événement B

Probabilité $P(A) \geq 0$ $P(\Omega) = 1$

3 axiomes $A \cup B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

► Probabilité conditionnelle

de E sachant F

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

► Indépendance des événements E et F

$E \perp F$

$$P(E \cap F) = P(E) \cdot P(F)$$

$$P(E|F) = P(E)$$

► Variables aléatoires

X variable aléatoire "avant tirage"

"Fonction du résultat d'un tirage $\omega, X(\omega)$ "

$X: \Omega \rightarrow \mathbb{R}$

$X \perp Y$ si: $P(X \leq a, Y \leq b) = P(X \leq a) \cdot P(Y \leq b)$

► Expected value

Expected value $\hat{=}$ moyenne pondérée

Variance $\hat{=}$ de combien l'issue s'écarte de $E[X]$

► Variance

$$E[X] = \sum_k x_k P(X=x_k) = \mu \quad V[X] = E[X^2] - \mu^2$$

$$\text{ou } \int_{-\infty}^{\infty} x f(x)$$

$$\text{ou } \int_{-\infty}^{\infty} (x-\mu)^2 g(x) dx$$

$$E[X+Y] = E[X] + E[Y]$$

$$V[X+Y] = V[X] + V[Y]$$

$$E[aX+b] = aE[X] + b$$

$$V[aX+b] = a^2 V[X]$$

$$E[g(x)] = \sum_i g(x_i) p(x_i)$$

$X \perp Y$

$$X \perp Y \Rightarrow E[XY] = E[X]E[Y]$$

$$E[X_j] = \mu \quad V[X_j] = \sigma^2 \quad X \text{ i.i.d.} \Rightarrow \perp, \text{ same dist.}$$

$$E[\bar{X}] = \mu$$

$$V[\bar{X}] = \frac{\sigma^2}{n}$$

$$\bar{X} := \frac{1}{n} \sum_{j=1}^n X_j \quad S^2 := \frac{1}{n} \sum_{j=1}^n (X_j - \mu)^2 \quad S^{*2} := \frac{1}{n-1} \sum_{j=1}^n (X_j - \mu)^2$$

► Distributions $P(X \in \mathbb{R}) = 1$ $0 \leq P(X \in B) \leq 1$

► Distribution discrète = histogramme normalisé à 1

X prend une valeur finie

$$p(x) = P(X=x)$$

$$\sum_{k=1}^{\infty} p(x_k) = 1$$

Fonction répartition:

$$F_X: \mathbb{R} \rightarrow [0, 1]$$

$$x \mapsto F_X(x) = P(\{X \leq x\})$$

Exemple: Bernoulli, Binomiale, Poisson
 $\lambda = n \cdot p, n$ grand, p petit

► Distribution continue X prend valeur $[a, b]$

Densité de probabilité $f(x)$

$$\int_{-\infty}^{\infty} f(x) dx = P(\{X \in \mathbb{R}\}) = 1$$

$$\frac{d}{dx} F_X(x) = f(x)$$

Fonction répartition:

$$P(\{X \in B\}) = \int_B f(x) dx$$

Exemple: Normale, uniforme, exponentielle

Distribution normale

v.a. normale de param μ, σ^2

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

v.a. centrée réduite

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

fct. répartition de Z

$$\Phi(x) = P(\{Z \leq x\})$$

Approx d'une dist binomiale avec une normale

$$n \rightarrow \infty \quad P(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b) \rightarrow \Phi(b) - \Phi(a)$$

Distribution χ^2 "Chi-carré" $X = X_1^2 + X_2^2 + \dots + X_n^2$

n degrés de liberté

$$X_i \sim \mathcal{N}(0, 1)$$

Distribution t (Student) $T = \frac{X}{\sqrt{Y/r}}$

r degrés liberté $Y \sim \chi_r^2, X \sim \mathcal{N}(0, 1)$

► Théorèmes limites

law large numbers:

$X_1, X_2, \dots, X_n$ i.i.d

$$E[X_j] = \mu \quad V_j$$

$$\bar{X} := \frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow \mu \quad \text{as } n \rightarrow \infty$$

Central limit thm:

$$Z = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma \cdot \sqrt{n}}$$

$$P(Z \leq a) \rightarrow \Phi(a) \quad \text{as } n \rightarrow \infty$$

► Estimation

X v.a dont on ne connaît pas la distribution

On suppose $E[X] = \lambda$ et $Var[X] = \gamma$

On veut estimer λ ou γ

Choix d'un estimateur: $\hat{\lambda} = f(X_1, X_2, \dots, X_n)$

vraie valeur avec échantillons $\hat{\lambda} = f(x_1, x_2, \dots, x_n)$

estimateur sans biais si: par ex $E[\hat{\lambda}] = \lambda$

► Intervalles de confiance

$\hat{\lambda}$ change à chaque échantillonnage, dans quelle probabilité $\hat{\lambda}$ est comprise entre $\lambda \in [\hat{\lambda}_1, \hat{\lambda}_2]$

Pour l'espérance la v. théorique est avec $(1-\alpha) \cdot 100\%$ dans l'intervalle de confiance

σ connu $U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$\bar{X} - u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

σ inconnu $T = \frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim \text{Student } v = n-1$

$$\bar{X} - t_{1-\alpha/2} \cdot \frac{S'}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{1-\alpha/2} \cdot \frac{S'}{\sqrt{n}}$$

La variance $\gamma = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2$ $n-1$ degrés lib.

$$\frac{(n-1)S'^2}{\chi^2_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S'^2}{\chi^2_{\alpha/2}}$$

Une proportion proportion p des éléments qui possèdent le caractère

$X \sim \text{Bernoulli}$: $n\bar{X} = S_n$ estimateur de p

$$P(n\bar{X} \leq \hat{k}) = \sum_{j=0}^{\hat{k}} \binom{n}{j} p_u^j (1-p_u)^{n-j} = \frac{\alpha}{2}$$

$$P(n\bar{X} \geq \hat{k}) = \sum_{j=\hat{k}}^n \binom{n}{j} p_l^j (1-p_l)^{n-j} = \frac{\alpha}{2}$$

► Tests statistiques

But: tester si la réalisation se trouve dans l'intervalle de confiance et valider l'hypothèse

Tests paramétriques: distribution connue, μ ou σ^2

Risque d'erreur: i) 1^{ère} espèce $\rightarrow H_0$ vraie mais rejetée

ii) 2^{ème} espèce $\rightarrow H_1$ vraie mais H_0 acceptée

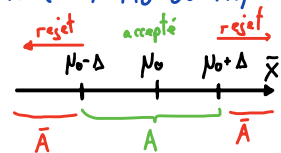
Proba commettre erreur

Risque 1^{ère} espèce
Risque 2^{ème} espèce

Décision versus Situation	H_0 vraie	H_1 vraie
accepter H_0	$1 - \alpha$	β
rejeter H_0	α	$1 - \beta$

$T(X_1, \dots, X_n) \rightarrow$ fonction test, trancher H_0 ou H_1

Région critique: $A \cup \bar{A} = \mathbb{R}$



► Test de la moyenne

Tester $\rightarrow H_0: \mu = \mu_0$ contre $H_1: \mu \neq \mu_0$

σ connu:

rejetée H_0 si:

$$\bar{x} < \mu_0 - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{ou} \quad \bar{x} > \mu_0 + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\beta_\mu = P\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - u_{1-\alpha/2} \leq T \leq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + u_{1-\alpha/2}\right)$$

σ inconnu:

rejetée H_0 si:

$$\bar{x} < \mu_0 - t_{1-\alpha/2} \frac{S'}{\sqrt{n}} \quad \text{ou} \quad \bar{x} > \mu_0 + t_{1-\alpha/2} \frac{S'}{\sqrt{n}}$$

$$\beta_\mu = P\left(\frac{\mu_0 - \mu}{S'/\sqrt{n}} - t_{1-\alpha/2} \leq T \leq \frac{\mu_0 - \mu}{S'/\sqrt{n}} + t_{1-\alpha/2}\right)$$

$\hookrightarrow$ risque seconde espèce

P-value: $p := P(T < -|t| | H_0) + P(T \geq |t| | H_0)$

$t = \frac{\bar{x} - \mu_0}{S'/\sqrt{n}}$
 $p < \alpha \rightarrow H_0$ à rejeter
 $p \geq \alpha \rightarrow H_0$ pas à rejeter

► Convolution

$f(t), g(t)$ T périodique dans l'intervalle $[a, b]$

T périodique! $(f * g)(s) = \int_a^b f(t)g(s-t)dt$

Discret: $\vec{f} := (\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots)$ $\vec{g} :=$ same

$(\vec{f} * \vec{g})_n := \sum_{k=-\infty}^{+\infty} f_k g_{n-k}$ $(\vec{a} * \vec{b}) := \sum_{m+n-1} a_m b_n$

$$\begin{aligned} f * g &= g * f & f * (g * h) &= (f * g) * h \\ \alpha(f * g) &= (\alpha f) * g & f * (g + h) &= (f * g) + (f * h) \end{aligned}$$

► Somme de variables aléatoires indép.

$$Z = X + Y \quad f_Z(z) = (f_X * f_Y)(z)$$

Densité prob. $= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y) dy$

IV Combinatoire

► Permutations possibles

Combien permutations possibles d'un ensemble de n éléments? $n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$, $n > 10$

ABCDE, ABCED, ABECD... $n = 5$

► Arrangements sans répétition

$A, B, C, D, E \Rightarrow \overbrace{ABC, ACB, BAD}^k$... avec ordre
 $A_n^k = \frac{n!}{(n-k)!}$

► Combinaisons différentes

$A, B, C, D, E \Rightarrow \overbrace{ABC, ABD, AEC}^k$... sans ACB
 $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

► Régression linéaire

Chercher une relation affine $y = ax + b$
pour corréler deux variables X, Y

Méthode des moindres carrés

→ Trouver a, b t.q $S = \sum_{j=1}^n \epsilon_j^2$ plus petit

$$\epsilon_j = y_j - Y^{\text{"ax+b"}}$$

On trouve: $\hat{y} = ax + b$ "meilleur"

$$a = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\frac{1}{n-1} \sum (x_j - \bar{x})(y_j - \bar{y})}{\frac{1}{n-1} \sum (x_j - \bar{x})^2}$$

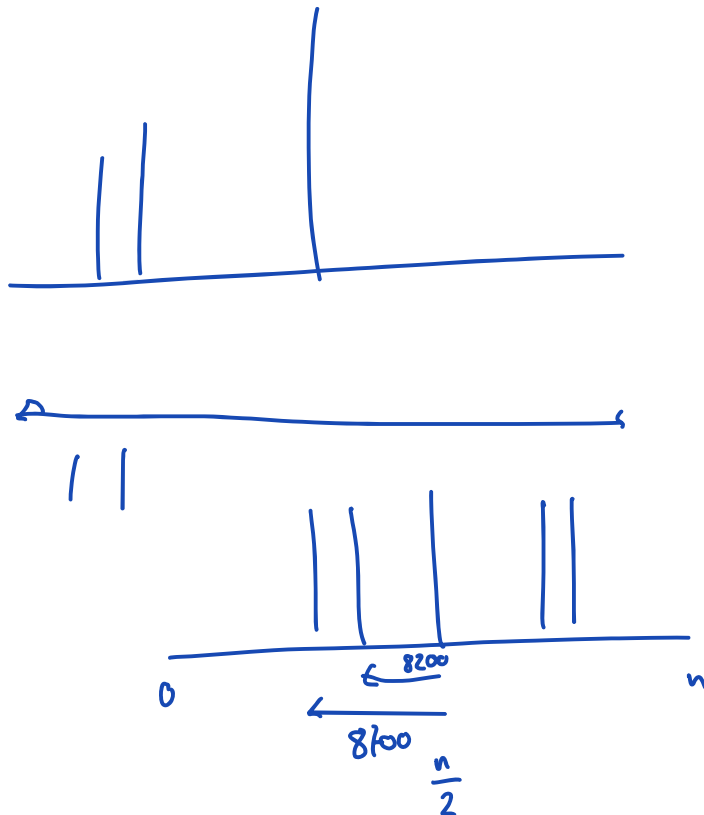
$$b = \bar{y} - a\bar{x} \quad (ax+b \text{ passe par le point } \bar{x}, \bar{y})$$

Coefficient corrélation linéaire entre X et Y

$$r_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}} \quad r \in [-1, 1]$$

Généralisation: $\begin{pmatrix} n & S_x \\ S_x & S_{xx} \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$

$\sum_{j=1}^n x_j^2 \leftarrow S_{xx}$



$$20 \quad 19 + 1$$

$$n = 2N + 1$$

$$n - 1 = 2N$$

$$\frac{n-1}{2} = N$$