

Cours IFT6266,

Petit rappel sur les probabilités

Soit $\mathcal{X}$ un ensemble (de valeurs possibles). Une variable aléatoire X (en général écrite en majuscule) associe aux événements aléatoires possibles une valeur dans $\mathcal{X}$, et à chaque événement aléatoire possible correspond une valeur x dans $\mathcal{X}$. Soit $\xi \subseteq \mathcal{X}$, alors $P(X \in \xi)$ dénote la probabilité qu'un des événements aléatoires possibles corresponde à une valeur qui soit dans ξ .

- **Axiomes:**

- $0 \leq P(X \in \xi) \leq 1$.
- $P(X \in \mathcal{X}) = 1$
- Si $\xi_1 \subset \mathcal{X}$ et $\xi_2 \subset \mathcal{X}$ et $\xi_1 \cap \xi_2 = \emptyset$ alors $P(X \in \xi_1 \text{ ou } X \in \xi_2) = P(X \in \xi_1) + P(X \in \xi_2)$.

- **Probabilité cumulative :**

Si $\mathcal{X}$ est un ensemble ordonné (pour nous les réels), alors on définit la probabilité cumulative comme $P(X \leq x)$. La fonction de probabilité cumulative $F_X(x) = P(X \leq x)$. Dans le cas multivarié $X = (X_1, \dots, X_n) \in \mathcal{R}^n$ on définit $F_X(x) = P(X \leq x) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$. La **distribution cumulative empirique** d'un échantillon $(x_1, \dots, x_n)$ est

$$\hat{F}(x) = \frac{\#x_i \leq x}{n}.$$

- **Continu vs discret** Si $F_X(x)$ est une fonction **continue** alors on dit que la v.a. X est **continue**. Si $F_X(x)$ est une somme de "pas" (i.e. est constante par parties) alors la v.a. X est dite **discrète**. Sinon on a une v.a. qui est une mixture de discret et de continu.

- **Densité:**

La fonction $f(x)$ est la fonction de densité de la variable aléatoire X ssi $\forall \xi \subseteq \mathcal{X}$,

$$P(X \in \xi) = \int_{\xi} f(x) dx.$$

Donc, pour $\mathcal{X} \subseteq \mathbb{R}$ la dérivée de la cumulative est la densité: $\frac{dF_X(x)}{dx} = f(x)$. Dans le cas plus général où $\mathcal{X} \subseteq \mathbb{R}^n$, la densité est obtenue en dérivant la cumulative par rapport à chaque composante de x . Par exemple pour $n = 2$, en notant $x = (x_1, x_2)$, $f(x) = \frac{\partial^2 F_X(x)}{\partial x_1 \partial x_2}$. La **densité empirique** est zéro partout sauf sur les échantillons, qui sont des points de masse $1/n$.

- **Abus de notation:** Par abus de notation je vais parfois utiliser la même notation $\mathcal{P}(X = x)$ pour faire référence à une fonction de probabilité (quand X est discrète) ou pour faire référence

à une densité (quand X est continue). Cependant il ne faut pas oublier que la densité n'est pas une probabilité: en gros c'est $f(x)dx$ qui est une probabilité. Notons aussi qu'on peut avoir une densité $\mathcal{P}(x) > 1$ mais pas une probabilité $\mathcal{P}(x) > 1$. Si X est une v.a. continue, alors la véritable probabilité $P(X = x)$ que $X = x$ est 0, et $f(x) < \infty$. Si X est une v.a. discrète, alors $f(x)$ est zéro presque partout, sauf pour les valeurs de x dans $\mathcal{X}$ (qui est alors un ensemble dénombrable): $f(x)$ a des "points de masse" pour les valeurs de $x \in \mathcal{X}$, i.e. pour un tel x l'intégrale de $f(x)$ dans un voisinage $V(x)$ autour de x donne $\int_{V(x)} f(x)dx = P(X = x)$.

• Union:

En général $\mathcal{P}(X \in \xi_1 \text{ ou } X \in \xi_2) = \mathcal{P}(X \in \xi_1) + \mathcal{P}(X \in \xi_2) - \mathcal{P}(X \in \xi_1 \text{ et } X \in \xi_2)$.

• Probabilité jointe

$\mathcal{P}(X = x, Y = y) = \mathcal{P}(X = x \text{ et } Y = y)$, avec $x \in \mathcal{X}, y \in \mathcal{Y}$ est définie comme précédemment, mais pour la variable aléatoire $Z = (X, Y)$, qui prend ses valeurs dans le produit cartésien de $\mathcal{X}$ et $\mathcal{Y}$, $\mathcal{X} \times \mathcal{Y}$.

• Probabilité conditionnelle

- $\mathcal{P}(X = x|Y = y) = \frac{\mathcal{P}(X=x, Y=y)}{\mathcal{P}(Y=y)}$ (définie seulement quand $\mathcal{P}(Y = y)$ est non-nul).

On a une définition similaire pour $\mathcal{P}(X \in \xi|Y \in \zeta)$.

• Indépendance

X et Y sont indépendants si et seulement si $\mathcal{P}(X \in \xi, Y \in \zeta) = \mathcal{P}(X \in \xi)\mathcal{P}(Y \in \zeta)$ (pour tous les ξ et ζ possibles), ou, ce qui est équivalent, si et seulement si $\mathcal{P}(X = x, Y = y) = \mathcal{P}(X = x)\mathcal{P}(Y = y)$ pour tous x, y .

• Espérance

$E[h(X)] = \int h(x)\mathcal{P}(x)dx$
(ou dans le cas discret, $\sum_i h(x_i)\mathcal{P}(x_i)$).

• Variance et matrice de covariance, pour $X \in \mathcal{R}^n$:

$Var[X] = \int (x - E[X])'(x - E[X])\mathcal{P}(x)dx$
(ou dans le cas discret, $\sum_i (x_i - E[X])'(x_i - E[X])\mathcal{P}(x_i)$).

Notons que

$$Var[X] = E[XX'] - E[X]E[X'].$$

• Moments

Le k -ième moment de X est $E[X^k]$, et le k -ième moment centré est $E[(X - E[X])^k]$ (donc la variance est le deuxième moment centré). On peut caractériser une distribution avec tous ses moments, qu'on peut obtenir facilement d'une fonction, la fonction génératrice des moments: $M(t) = \int e^{tx}\mathcal{P}(x)dx$, avec $E[X^k] = M^{(k)}(0)$. Deux variables aléatoires qui ont la même fonction génératrice des moments ont la même distribution et vice-versa.

- **Loi Normale:** seuls ses deux premiers moments sont non-nuls.

$$\mathcal{P}(X = x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-.5(x-\mu)^2/\sigma^2}.$$

Pour $X \in \mathcal{R}^n$, on obtient la loi multivariée:

$$\mathcal{P}(X = x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-.5(x-\mu)'\Sigma^{-1}(x-\mu)}$$

où Σ est la matrice de covariance et $|\Sigma|$ son déterminant (Σ est positive définie, i.e., $\forall x, x'\Sigma x > 0$, i.e. ses valeurs propres sont positives).

- **Loi des grands nombres**

La moyenne $\bar{X}$ d'une séquence de n variables indépendentes ($X_1, \dots, X_n$) de même espérance μ et de même variance σ^2 converge vers cette espérance μ :

$$\lim_{n \rightarrow \infty} E[\bar{X}] = \mu.$$

La preuve est basé sur le simple fait que $Var[\bar{X}] = \frac{\sigma^2}{n}$.

- **Théorème limite centrale:**

La moyenne $\bar{X}$ d'une séquence de n variables *iid* ($X_1, \dots, X_n$) converge en probabilité (sa distribution converge) vers une loi Normale avec espérance $E[X]$ et variance $Var[X]/n$.

- **Théorème de Bayes**

$$\mathcal{P}(X = x|Y = y) = \frac{\mathcal{P}(Y=y|X=x)\mathcal{P}(X=x)}{\mathcal{P}(Y=y)}$$

dérive trivialement de la définition de probabilité conditionnelle. Une extension utile est

$$\mathcal{P}(X = x|Y = y) = \frac{\mathcal{P}(Y=y|X=x)\mathcal{P}(X=x)}{\sum_x \mathcal{P}(Y=y|X=x)\mathcal{P}(X=x)}.$$

On a donné un nom aux éléments de ce théorème:

$$\text{postérieur} = \frac{\text{vraisemblance} \times \text{priori}}{\text{evidence}}.$$

- **Fonction d'une variable aléatoire**

Soit X de distribution cumulative $F_X(x)$ et densité $\mathcal{P}(X = x)$ connues et soit h une fonction; on voudrait connaître la distribution de $Y = h(X)$. La méthode la plus générale est basée sur la cumulative: $\mathcal{P}(h(X) \leq t) = \int_A \mathcal{P}(x)dx$ pour $A = \{x : h(X) \leq t\}$. On obtient ensuite $\mathcal{P}(Y = y)$ par différentiation. Si $h : \mathcal{R}^n \rightarrow \mathcal{R}^n$ est inversible, avec $g = h^{-1}$, alors

$$\mathcal{P}(Y = y) = \mathcal{P}(X = g(y)) |det(g'(g(y)))|$$

ou $det()$ est le determinant et g' est la matrice des dérivées partielles des éléments de $g(x)$ par rapport aux éléments de x . Une méthode approximative permet d'estimer les deux premiers moments (en se basant sur l'expansion de Taylor de h): $E[Y] \approx h(E[X])$, et $Var[Y] \approx Var[X]h'(E[X])^2$. Une meilleure approximation est $E[Y] \approx h(E[X]) + 0.5Var[X]h''(E[X])$. Une méthode approximative parfois plus précise est l'intégration numérique (Monte-Carlo si la dimension de X est élevée, > 4 ou 5).