++ de Probabilité

Ou comment survivre dans la jungle des probas

Colin 'Alaska' Coërchon

IIE++

8 janvier 2024

Sommaire

1. Résumé des éléments importants
   - Les bases indispensables
   - Détermination d'une loi
   - Autres trucs importants
2. Correction du partiel de l'an dernier
3. Place aux questions !

1 Résumé des éléments importants
- Les bases indispensables
- Détermination d'une loi
- Autres trucs importants

2 Correction du partiel de l'an dernier

3 Place aux questions !

Les lois discrètes à connaître

- Loi de Bernoulli $\mathcal{B}(p)$:

$$\mathbb{P}(X = 1) = p \quad \text{et} \quad \mathbb{P}(X = 0) = 1 - p$$

- Loi Binomiale $\mathcal{B}(n, p)$:

$$\forall k \in [\![1, n]\!], \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- Loi Géométrique $\mathcal{G}(p)$:

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X = k) = p\,(1 - p)^{k-1}$$

- Loi de poisson $\mathcal{P}(\lambda)$:

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Les lois continues à connaître

- Loi normale standard $\mathcal{N}(0,1)$:

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- Loi normale $\mathcal{N}(\mu, \sigma^2)$:

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Loi exponentielle $\mathcal{E}(\lambda)$:

$$\forall x \in \mathbb{R}, \quad f(x) = \lambda e^{-\lambda x} \, \mathbb{1}_{\mathbb{R}_+}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Ne pas oublier les deux lois uniformes...

- Loi uniforme discrète sur $[\![1, n]\!]$ $(\mathbb{U}_{[\![1,n]\!]})$:

$$\forall k \in [\![1, n]\!], \quad \mathbb{P}(X = k) = \frac{1}{n}$$

- Loi uniforme continue sur $[a, b]$ $(\mathcal{U}(a, b))$:

$$\forall x \in \mathbb{R}, \quad f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

Pour chacune de ces lois, il est crucial d'apprendre au moins :

- Leur ensemble de définition ($X(\Omega)$).
- Leur **espérance** ($\mathbb{E}[X]$).
- Leur **variance** ($\mathrm{Var}(X)$).

Pour chacune de ces lois, il est crucial d'apprendre au moins :

- Leur ensemble de définition ($X(\Omega)$).
- Leur **espérance** ($\mathbb{E}[X]$).
- Leur **variance** ($\mathrm{Var}(X)$).

Et pour les tryhardeurs :

- Leur fonction caractéristique (au moins pour la loi normale $\mathcal{N}(\mu, \sigma^2)$) : $\psi_X(t) = \mathbb{E}\left[e^{\mathrm{i}tX}\right]$
- Leur fonction génératrice (*pour les lois discrètes*) : $G_X(t) = \mathbb{E}\left[t^X\right]$

Pour chacune de ces lois, il est crucial d'apprendre au moins :

- Leur ensemble de définition ($X(\Omega)$).
- Leur **espérance** ($\mathbb{E}[X]$).
- Leur **variance** ($\mathrm{Var}(X)$).

Et pour les tryhardeurs :

- Leur fonction caractéristique (au moins pour la loi normale $\mathcal{N}(\mu, \sigma^2)$) : $\psi_X(t) = \mathbb{E}\left[e^{\mathrm{i}tX}\right]$
- Leur fonction génératrice (*pour les lois discrètes*) : $G_X(t) = \mathbb{E}\left[t^X\right]$

Voici une super formulaire qui résume tout cela : lien

Quelques définitions cruciales

Pour les lois discrètes

On considère X v.a. discrète sur $\mathbb{N}$.

$$F_X(N) \triangleq \mathbb{P}(X \leq N) = \sum_{k=0}^{N} \mathbb{P}(X = k) \quad \text{et} \quad \mathbb{E}[X] = \sum_{k=0}^{+\infty} k\,\mathbb{P}(X = k)$$

Quelques définitions cruciales

Pour les lois discrètes

On considère X v.a. discrète sur $\mathbb{N}$.

$$F_X(N) \triangleq \mathbb{P}(X \leq N) = \sum_{k=0}^{N} \mathbb{P}(X = k) \quad \text{et} \quad \mathbb{E}[X] = \sum_{k=0}^{+\infty} k\,\mathbb{P}(X = k)$$

Pour les lois continues

On considère X v.a. continue sur $\mathbb{R}$.

$$F_X(x) \triangleq \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t \quad \text{et} \quad \mathbb{E}[X] = \int_{-\infty}^{+\infty} t\,f_X(t)\,\mathrm{d}t$$

Quelques séries à **impérativement** connaître

La série géométrique

Pour $|r| < 1$, la série géométrique est donnée par :

$$\sum_{k=0}^{+\infty} r^k = \frac{1}{1-r} \qquad \text{et} \qquad \sum_{k=q}^{+\infty} r^k = \frac{r^q}{1-r}$$

La série exponentielle

Pour tout $x \in \mathbb{R}$, la série exponentielle est donnée par :

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \qquad \text{et} \qquad e^1 = \sum_{k=0}^{\infty} \frac{1}{k!}$$

Quelques séries à **impérativement** connaître

Et une dernière pour la route

Pour $|x| < 1$, on a :

$$\sum_{k=1}^{+\infty} k\,x^{k-1} = \frac{1}{(1-x)^2} \qquad \text{et} \qquad \sum_{k=2}^{+\infty} k(k-1)\,x^{k-2} = \frac{2}{(1-x)^3}$$

Quelques séries à **impérativement** connaître

Et une dernière pour la route

Pour $|x| < 1$, on a :

$$\sum_{k=1}^{+\infty} k\,x^{k-1} = \frac{1}{(1-x)^2} \qquad \text{et} \qquad \sum_{k=2}^{+\infty} k(k-1)\,x^{k-2} = \frac{2}{(1-x)^3}$$

Exemple d'application

Calculer $\displaystyle\sum_{k=0}^{+\infty} k^2\,x^k$.

$$\left(= \frac{x(x-1)}{(1-x)^3}\right).$$

Astuce pour la variance

Définition

$$\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

Astuce pour la variance

Définition

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

L'astuce de calcul

Par le théorème de transfert, on a :

$$\mathbb{E}\left[X^2\right] = \sum_{k=0}^{+\infty} k^2\, \mathbb{P}(X = k) \quad \text{et} \quad \mathbb{E}\left[X^2\right] = \int_{-\infty}^{+\infty} t^2\, f_X(t)\, \mathrm{d}t$$

Vérification des fonctions densités

Cas des lois discrètes

Pour vérifier que X suit bien une loi de probabilité, on doit vérifier **deux propriétés** :

- Positivité :

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) \geq 0$$

- Normalisation :

$$\sum_{k=0}^{+\infty} \mathbb{P}(X = k) = 1$$

Vérification des fonctions densités

Cas des lois continues

Pour vérifier qu'une fonction f est une densité de probabilité, on doit vérifier **deux propriétés** :

- Positivité :

$$\forall x \in \mathbb{R}, \quad f(x) \geq 0$$

- Normalisation :

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$$

1 Résumé des éléments importants
- Les bases indispensables
- Détermination d'une loi
- Autres trucs importants

2 Correction du partiel de l'an dernier

3 Place aux questions !

Comment répondre à la question :

Soient X une variable aléatoire connue et $Y = f(X)$, quelle est la loi de Y ?

On connaît 3 méthodes :

- **Méthode 1** : Par les fonctions de répartition

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(f(X) \leq x) \quad \text{puis : } f_Y(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_Y(x)$$

- **Méthode 2** : Méthode de la **fonction muette**

  On calcule pour toute fonction continue bornée g : $\mathbb{E}[g(Y)]$. Et on essaye de se ramener à :

$$\mathbb{E}[g(Y)] = \int_{\mathbb{R}} g(x)h(x)\,\mathrm{d}x \qquad \text{(avec le chgt. de var. } y = f(x))$$

  Alors dans ce cas, la fonction h est une densité de Y.

- **Méthode 3** : Par les fonctions caractéristiques [DIFFICILE]

  Rappel : $\forall t \in \mathbb{R}$, $\psi_X(t) = \mathbb{E}\left[e^{\mathrm{i}tX}\right]$. On utilise la propriété :

$$\psi_X = \psi_Y \Longleftrightarrow f_X = f_Y$$

Exemple : Méthode 1

On s'intéresse à :

$$X \sim \mathcal{N}(0,1) \quad \text{et} \quad Y = X^3$$

On a donc :

$$\forall x \in \mathbb{R}, \quad F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(X^3 \leq x) = \mathbb{P}(X \leq \sqrt[3]{x}) = F_X(\sqrt[3]{x})$$

D'où :

$$f_Y(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(F_X(\sqrt[3]{x})\right) = \frac{1}{3\sqrt[3]{x^2}} F_X'(\sqrt[3]{x}) = \frac{1}{3\sqrt{2\pi}\sqrt[3]{x^2}} e^{-\frac{\sqrt[3]{x^2}}{2}}$$

Autres exemples pour cette méthode

$$Y = \ln(X), \ Y = \exp(X), \ Y = \sqrt{X}, \ Y = 3X + 2, \ \text{etc...}$$

Exemple : Méthode 2

On reprend le même exemple.

Soit g une fonction bornée et mesurable. Alors

$$\mathbb{E}[g(Y)] = \mathbb{E}[g(X^3)] = \int_{\mathbb{R}} g(x^3) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x$$

En posant $y = x^3$ (ou plutôt $x = y^{\frac{1}{3}}$), il vient que :

$$\mathbb{E}[g(Y)] = \int_{\mathbb{R}} g(y) \frac{1}{3\sqrt{2\pi}\sqrt[3]{y^2}} e^{-\frac{\sqrt[3]{y^2}}{2}} \, \mathrm{d}y$$

D'où :

$$f_Y(y) = \frac{1}{3\sqrt{2\pi}\sqrt[3]{y^2}} e^{-\frac{\sqrt[3]{y^2}}{2}}$$

Exemple : Méthode 3

Exemple typique

Soient $(X_i)_{1\leq i\leq n}$ des v.a i.i.d chacune suivant une loi normale centrée réduite ($\mathcal{N}(0,1)$). Déterminer la loi de :

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

Solution

$$\psi_{\overline{X}_n}(t) = \mathbb{E}\left[e^{i\overline{X}_n t}\right] = \mathbb{E}\left[e^{i\frac{t}{n}\sum_k X_k}\right] = \mathbb{E}\left[\prod_{k=1}^{n} e^{i\frac{t}{n}X_k}\right]$$

Et donc, par indépendance des (X_k), on a :

$$\psi_{\overline{X}_n}(t) = \prod_{k=1}^{n} \mathbb{E}\left[e^{i\frac{t}{n}X_k}\right] = \prod_{k=1}^{n} \psi_{\mathcal{N}(0,1)}\left(\frac{t}{n}\right)$$

Comme on sait que :

$$\psi_{\mathcal{N}(\mu,\sigma^2)}(t) = e^{i\mu t - \frac{t^2\sigma^2}{2}} \quad \text{on en déduit que :} \quad \psi_{\overline{X}_n}(t) = e^{-\frac{t^2}{2n}}$$

Et finalement, par la propriété énoncé à la slide 15, on conclut que :

$$\boxed{\overline{X}_n \sim \mathcal{N}\left(0, \frac{1}{n}\right)}$$

Somme de deux variables aléatoires

Théorème

Soit X et Y deux variables aléatoires. On pose alors :

$$Z = X + Y$$

Sous réserve de convergence, la variable aléatoire Z admet pour densité la fonction $h = f_X * f_Y$, produit de convolution des densités de X et Y.
On a donc, $\forall x \in \mathbb{R}$:

$$h(x) = (f_X * f_y)(x) = \int_{\mathbb{R}} f_X(t)\, f_y(x-t)\, \mathrm{d}t$$

Somme de deux variables aléatoires

Théorème

Soit X et Y deux variables aléatoires. On pose alors :

$$Z = X + Y$$

Sous réserve de convergence, la variable aléatoire Z admet pour densité la fonction $h = f_X * f_Y$, produit de convolution des densités de X et Y.
On a donc, $\forall x \in \mathbb{R}$:

$$h(x) = (f_X * f_y)(x) = \int_{\mathbb{R}} f_X(t)\, f_y(x-t)\, \mathrm{d}t$$

Un exemple ? $\longrightarrow$ correction du partiel 2023 !

1 Résumé des éléments importants
- Les bases indispensables
- Détermination d'une loi
- Autres trucs importants

2 Correction du partiel de l'an dernier

3 Place aux questions !

Vecteurs gaussiens

Définition

On appelle **vecteur gaussien** tout vecteur de variables aléatoires $(X_1, ..., X_n)$ tel que toute combinaison linéaire $Y = \sum_i a_i X_i$ suive une loi normale.

On appelle alors **matrice de covariance** la matrice M, associée au vecteur gaussien, tel que $\forall (i,j) \in [\![1, n]\!]^2,\ m_{ij} = \text{Cov}(X_i, X_j)$

Propriété 1

Soit $(X_i)_{1\leq i\leq n}$ des v.a. suivant des lois normales, on a :

$(X_i)_{1\leq i\leq n}$ indépendantes $\Longrightarrow (X_1, \ldots, X_n)$ est un vecteur gaussien

ATTENTION, on n'a pas $\Longleftarrow$! ! !

Propriété 1

Soit $(X_i)_{1\leq i\leq n}$ des v.a. suivant des lois normales, on a :

$(X_i)_{1\leq i\leq n}$ indépendantes $\Longrightarrow (X_1, \ldots, X_n)$ est un vecteur gaussien

ATTENTION, on n'a pas $\Longleftarrow$! ! !

Propriété 2

Soit $(X_1, \ldots, X_n)$ un vecteur gaussien.
Et soit $i, j \in [\![1, n]\!]$, $i \neq j$, on a :

$$X_i \text{ et } X_j \text{ sont indépendantes} \Longleftrightarrow \text{Cov}(X_i, X_j) = 0$$

Note : On a toujours $\Longrightarrow$, mais rarement $\Longleftarrow$, d'où la force de cette propriété !

Les différents types de convergence

- **Convergence presque sûre** : on dit que la suite de variables aléatoires X_n converge presque sûrement vers X ssi il existe un ensemble négligeable N tel que

$$\forall \omega \in \Omega \backslash N\,, \quad X_n(\omega) \xrightarrow[n\to+\infty]{} X(\omega)$$

- **Convergence en probabilité** :

$$X_n \xrightarrow{\mathbb{P}} X \quad \Longleftrightarrow \quad \forall \varepsilon > 0\,, \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

- **Convergence en loi** : pour toute fonction numérique f continue et bornée,

$$X_n \xrightarrow{\mathcal{L}} X \quad \Longleftrightarrow \quad \lim_{n\to+\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$$

Ordre d'importance :

Convergence p.s. $\Longrightarrow$ Convergence en proba $\Longrightarrow$ Convergence en Loi

Théorèmes de convergence

Loi forte des grands nombres

Soit (X_n) une suite de variables aléatoire i.i.d tel que $\mathbb{E}[|X_1|] < +\infty$. Alors

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{p.s} \mathbb{E}[X_1]$$

Théorème Central Limite (TCL)

Soit (X_n) une suite de variables aléatoire i.i.d tel que $\mathbb{E}[|X_1|] < +\infty$ et $\mathrm{Var}(|X_1|) < +\infty$. Alors :

$$\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

FIN DU ++, merci à vous d'être venu !

Si le ++ vous a plu, n'hésitez pas à donner un peu d'argent au compte du bar de IIE++, ça fait toujours plaisir !

FIN DU ++, merci à vous d'être venu !

Si le ++ vous a plu, n'hésitez pas à donner un peu d'argent au compte du bar de IIE++, ça fait toujours plaisir !

Un grand merci à Chanterelle de gérer toute l'orga, et d'aider en conséquence les 1A à valider !

FIN DU ++, merci à vous d'être venu !

Si le ++ vous a plu, n'hésitez pas à donner un peu d'argent au compte du bar de IIE++, ça fait toujours plaisir !

Un grand merci à Chanterelle de gérer toute l'orga, et d'aider en conséquence les 1A à valider !

Place à la correction du partiel maintenant ! !

1 Résumé des éléments importants

2 Correction du partiel de l'an dernier

3 Place aux questions !

1 Résumé des éléments importants

2 Correction du partiel de l'an dernier

3 Place aux questions !

Des questions ?