Université Paris I, Panthéon - Sorbonne

MAGISTÈRE FINANCE UFR06, 2010-2011

Cours d'introduction à la Statistique S1

JEAN-MARC BARDET (UNIVERSITÉ PARIS 1, SAMM)

U.F.R. 27, Equipe SAMM (Statistique, Analyse et Modélisation Multidisciplinaire)
Université Panthéon-Sorbonne, 90 rue de Tolbiac, 75013 Paris.

Plan du cours

1. Tribu et mesure de probabilité
2. Variables aléatoires
3. Vecteurs aléatoires
4. Théorèmes limites
5. Estimateurs et intervalles de confiance

Bibliographie

- Baillargeon, B. *Probabilités, statistiques et techniques de régression.* SMG.
- Barbe et Ledoux, *Probabilités*, Belin.
- Bercu, B., Pamphile, P. et Azoulay, E. *Probabilités et Applications - Cours Exercices.* Edisciences.
- Dacunha-Castelle et Duflo, *Probabilités et Statistiques (I)*, Masson
- Dress, F. *Probabilités et Statistique.* Dunod.
- Jacod, *Cours d'intégration*, http://www.proba.jussieu.fr/pageperso/jacod.html.
- Jacod, *Cours de Probabilités*, http://www.proba.jussieu.fr/pageperso/jacod.html.
- Lecoutre, J.-P. *Statistiques et Probabilités.* Dunod.

Introduction

La statistique et les probabilités: pour quoi faire?

1 Tribus et mesure de probabilités

1.1 Tribus d'événements

Définition.
- *Expérience aléatoire.*
- *Évènement élémentaire, ensemble fondamendal (univers).*
- *Événement et tribu.*

Définition.
- *Intersection, union.*
- *Évènements incompatibles.*
- *Évènement contraire.*

1.2 Mesure de probabilité d'un événement

Définition. *Une probabilité sur $(\Omega, \mathcal{A})$ est une application $P\colon \mathcal{A} \to [0,1]$, qui à un événement $E \in \mathcal{A}$ associe le réel $P(E) \in [0,1]$ et telle que:*

- $P(\Omega) = 1$.
- *Pour $(E_i)_{i\in J}$ événements de $\mathcal{A}$, incompatibles,* $P(\bigcup_{i\in J} E_i) = \sum_{i\in J} P(E_i)$.

Propriété.
- $P(\overline{E}) = 1 - P(E)$ *pour* $E \in \mathcal{A}$.
- $P(\emptyset) = 0$.
- *si* $A \subset B$, $0 \leq P(A) \leq P(B) \leq 1$.
- $P(E \cup F) + P(E \cap F) = P(E) + P(F)$, *pour* $(E,F) \in \mathcal{A}^2$.

Définition.
- *Cardinal d'un ensemble fini.*
- *Equiprobabilité (ou probabilité uniforme) sur un ensemble fini.*

Propriété. *Si Ω est fini et si P est une probabilité uniforme sur $(\Omega, \mathcal{A})$, alors pour tout $E \in \mathcal{A}$,* $P(E) = \dfrac{card(E)}{card(\omega)}$.

Remarque.

Pour calculer le cardinal d'un ensemble, on peut utiliser les résultats combinatoires suivants: on considère un ensemble de n éléments et on tire k éléments:

- S'il y a remise, et que l'ordre compte, un tirage est un k-uplet, et le nombre total de k-uplets est: n^k.
- S'il n'y a pas remise, et que l'ordre compte, un tirage est un arrangement, et le nombre total d'arrangements est: $A_n^k = n(n-1)..(n-k+1) = n!/k!$.
- S'il n'y a pas remise, et que l'ordre ne compte pas, un tirage est une combinaison, et le nombre total de combinaisons est: $C_n^k = A_n^k/k! = n!/(k!(n-k)!)$.

1.3 Probabilité conditionnelle et indépendance

Définition. *Soit P est une probabilité sur $(\Omega, \mathcal{A})$ et soit $(E, F) \in \mathcal{A}^2$. On appelle probabilité conditionnelle de B sachant A, la probabilité que B soit réalisé sachant que A est réalisé et* $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$ *si* $P(A) \neq 0$.

Remarque.

Calculer une probabilité sachant A revient à travailler avec une nouvelle probabilité sur l'ensemble fondamental A et la tribu qui lui est associée.

Définition. *A et B, événements de $(\Omega, \mathcal{A})$ sont indépendants pour P si $P(B \mid A) = P(B)$.*

Conséquence.

A et B, événements de $(\Omega, \mathcal{A})$ sont indépendants pour P si et seulement si $P(A \cap B) = P(A)P(B)$.

Remarque.

Ne pas confondre indépendance et incompatibilité. Plus généralement,

Définition.

- *Soit $(A_i)_{i \in I}$ une famille d'événements de $(\Omega, \mathcal{A})$. On dit que la famille $(A_i)_{i \in I}$ est une famille d'événements indépendants si et seulement si pour tout $k \in \mathbb{N}$, pour tout $j_1, \cdots, j_k \in I^k$ distincts, $P\Big(A_{j_1} \cap \cdots \cap A_{j_k}\Big) = P(A_{j_1}) \times \cdots \times P(A_{j_k})$.*
- *Si $(\Omega, \mathcal{A}_i)_{i \in I}$ est une famille de tribus sur Ω, on dit que ces tribus sont indépendantes si pour tout $A_i \in \mathcal{A}_i$ la famille $(A_i)_{i \in I}$ est indépendante.*

Définition. *Soit Ω un ensemble. On dit que $(E_i)_{i \in J}$ famille de sous-ensembles de Ω forme une partition de Ω si:*

- *Les E_i sont incompatibles deux à deux soit $E_i \cap E_j = \emptyset$ pour $i \neq j$.*
- *L'ensemble des E_i couvre Ω soit $\bigcup_{i \in J} E_i = \Omega$.*

Proposition. Formule des probabilités totales *Soit P est une probabilité sur $(\Omega, \mathcal{A})$ et soit $(E_i)_{i \in J}$ une partition de Ω alors:*

$$\text{pour } A \in \mathcal{A}, \quad P(A) = \sum_{i \in J} P(A \cap E_i).$$

Proposition. Formule de Bayes *Soit P est une probabilité sur $(\Omega, \mathcal{A})$ et soit $(E_i)_{i \in J}$ une partition de Ω. Soit A un événement de $\mathcal{A}$). On suppose que l'on connaît $P(E_i)$ et $P(A \mid E_i)$ pour tout $i \in J$. Alors, pour $k \in J$:*

$$P(E_k \mid A) = \frac{P(A \mid E_k)P(E_k)}{\sum_{i \in J} P(A \mid E_i)P(E_i)}.$$

2 Variables aléatoires

2.1 Définitions et propriétés générales

Définition. *Soit P une probabilité sur $(\Omega, \mathcal{A})$. On appelle variable aléatoire X dans $I \subset \mathbb{R}$, une application de $\Omega \to I$ telle que pour tout $x \in I$, l'événement $\{X \leq x\} = \{\omega \in \Omega \text{ tel que } X(\omega) \leq x\}$ soit un événement de $\mathcal{A}$. Deux cas particuliers importants sont à distinguer:*

- *Si* $I = \{x_i\}_{i \in J}$ *avec* $J \subset \mathbb{N}$ *(par exemple* $I = \{P, F\}$*,* $I = \mathbb{Z}$*,...),* X *est appelée variable aléatoire discrète.*
- *Si* I *est un intervalle de* $\mathbb{R}$ *(par exemple* $I = [0, 1]$*,* $I = \mathbb{R}^+$*,...),* X *est appelée variable aléatoire réelle.*

Définition. *Soit* P *une probabilité sur* $(\Omega, \mathcal{A})$ *et* X *une variable aléatoire de* $(\Omega, \mathcal{A}, P)$ *à valeurs dans* I*. On appelle fonction de répartition de* X *la fonction* $F_X : \mathbb{R} \to [0, 1]$ *telle que* $F_X(x) = P(X \leq x)$*, pour* $x \in \mathbb{R}$*.*

Propriété.

- F_x *est une fonction croissante sur* $\mathbb{R}$*.*
- $\lim\limits_{x \to -\infty} F_X(x) = 0$ *et* $\lim\limits_{x \to +\infty} F_X(x) = 1$*.*
- $P(a < X \leq b) = F_X(b) - F_X(a)$*, pour tout* $-\infty \leq a < b \leq +\infty$*.*

Définition. *Soit* P *une probabilité sur* $(\Omega, \mathcal{A})$ *et* X *une variable aléatoire de* $(\Omega, \mathcal{A}, P)$ *à valeurs dans* I*.*

- *Si* X *est une variable aléatoire discrète* $(I = \{x_i\}_{i \in J})$*, on appelle loi de probabilité de* X *l'application* $P_X : \{x_i\}_{i \in J} \to [0, 1]$ *telle que* $P(X = x_i) = P_X(x_i)$*.*
- *Si* X *est une variable aléatoire réelle (*I *intervalle de* $\mathbb{R}$*), on appelle densité de probabilité de* X **si elle existe** *l'application* $f_X : I \to [0, +\infty[$ *telle que* $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$*.* X *est alors appelée variable aléatoire absolument continue.*

Propriété. *Si* X *est une variable aléatoire discrète à valeurs dans* $I = \{x_i\}_{i \in J}$ *alors* $\sum\limits_{i \in J} P(X = x_i) = 1$*.*

Propriété. *Si* X *est une variable aléatoire absolument continue de densité de probabilité* f_X *alors:*

- f_X *vérifie* $\int_{-\infty}^{+\infty} f_X(t)dt = 1$*.*
- $F'_X(x) = f_X(x)$ *pour presque tout* $x \in \mathbb{R}$*.*
- $P(a < X \leq b) = \int_a^b f_X(t)dt$*, pour tout* $-\infty \leq a < b \leq +\infty$*.*
- $P(X = x) = 0$ *pour tout* $x \in \mathbb{R}$*.*

2.2 Moments d'une variable aléatoire

Définition. *Soit* P *une probabilité sur* $(\Omega, \mathcal{A})$*,* X *une variable aléatoire de* $(\Omega, \mathcal{A}, P)$ *à valeurs dans* I *et* h *une fonction de* I *dans* $\mathbb{R}$*.*

- *Si* X *est une variable aléatoire discrète* $(I = \{x_i\}_{i \in J})$*, l'espérance de* X *est*

$$BBeX = \sum_{i \in J} x_i P(X = x_i) = \sum_{i \in J} x_i P_X(x_i) \quad \textbf{si cette somme existe}.$$

Plus généralement, $\mathbb{E}h(X) = \sum\limits_{i \in J} h(x_i) P(X = x_i)$ **si cette somme existe**.

- *Si X est une variable aléatoire aboslument continue à valeurs dans I et de densité f_X, l'espérance de X est*
$$\mathbb{E}X = \int_I x f_X(x)dx = \int_{-\infty}^{+\infty} x f_X(x)dx \quad \textbf{si cette intégrale existe}.$$
Plus généralement, $\mathbb{E}h(X) = \int_I h(x)f_X(x)dx$ **si cette intégrale existe**.

Définition. *Soit X une variable aléatoire de $(\Omega, \mathcal{A}, P)$.*

- *La variance de X,* **si elle existe**, *est* $var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}((X - \mathbb{E}X)^2) \geq 0$.
- *L'écart-type de X,* **si la variance existe**, *est* $\sigma(X) = \sqrt{var(X)}$.

Propriété. *Soit X une variable aléatoire (dont l'espérance et la variance existent), et a, b deux réels.*

- $\mathbb{E}b = b$ *et* $\mathbb{E}(aX + b) = a\mathbb{E}X + b$.
- $var(b) = 0$ *et* $var(aX + b) = a^2 var(X)$.

2.3 Lois à connaître

Définition. *Les lois à connaître sont:*

- *Lois discrètes: loi uniforme, loi de Bernoulli, loi binomiale, loi de Poisson.*
- *Lois continues: loi uniforme, loi exponentielle, loi gamma, loi normale, loi du χ^2.*

3 Vecteurs aléatoires

3.1 Définition et propriétés

Définition. *Un vecteur aléatoire X sur $(\Omega, \mathcal{A}, P)$ est une fonction de Ω à valeurs dans $\mathbb{R}^d$ telle que pour tout $B \in \mathcal{B}(\mathbb{R}^d)$, l'ensemble $X^{-1}(B) \in \mathcal{A}$.*

Définition. *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$. On peut définir la fonction de répartition de X par $F_X(x_1, \ldots, x_d) = P(X \in]-\infty, x_1] \times \cdots \times]-\infty, x_d])$ pour tout $(x_1, \ldots, x_d) \in \mathbb{R}^d$. Si $X = (X_1, \cdots, X_d)$ dans un repère de $\mathbb{R}^d$, on appelle loi marginale de la variable aléatoire X_i, la loi de probabilité donnée par la fonction de répartition $F_{X_i}(x_i) = \lim_{x_j \to \infty, j \neq i} F_X(x_1, \ldots, x_d)$.*

Définition. *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\{y^{(i)}\}_{i \in I} \in \mathbb{R}^d$, où I est dénombrable. On dit alors que X est un vecteur aléatoire discret et $P(X = y^{(j)}) \neq 0$.*

Définition. *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$ tel qu'il existe une fonction $f_X(x_1, \ldots, x_d)$ positive vérifiant $F_X(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_X(t_1, \ldots, t_d)dt_1 \ldots dt_d$ pour tout $(x_1, \ldots, x_d) \in \mathbb{R}^d$. On dit alors que X est un vecteur aléatoire absolument continu (par rapport à la mesure de Lebesgue) et f_X s'appelle sa densité de probabilité.*

Propriété. *Si $\frac{\partial^d F_X}{\partial x_1 \cdots \partial x_d}$ existe sur $\mathbb{R}^d$ sauf en un nombre fini de points, alors X est un vecteur aléatoire absolument continu et $f_X = \frac{\partial^d F_X}{\partial x_1 \cdots \partial x_d}$.*

Propriété. *Si $X = (X_1, \cdots, X_d)$ est un vecteur aléatoire absolument continu de densité de probabilité f_X alors $f_{X_i}(x_i) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{i-1}} \int_{-\infty}^{x_{i+1}} \int_{-\infty}^{x_d} f_X(t_1, \ldots, t_{i-1}, x_i, t_{i+1}, t_d)dt_1 \ldots dt_{i-1}dt_{i+1} \ldots dt_d$.*

Définition. *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$. Si cela existe, on définit l'espérance de X, $\mathbb{E}X = {}^t(\mathbb{E}X_1, \ldots, \mathbb{E}X_d)$ et sa matrice de variance-covariance $cov(X) = \mathbb{E}[(X - \mathbb{E}X)\,{}^t(X - \mathbb{E}X)] = (cov(X_i, X_j))_{1 \leq i,j \leq d}$.*

3.2 Variables aléatoires indépendantes

Définition. *Soit $(X_i)_{i\in I}$ une famille de variables aléatoires. On dit que ces variables sont indépendantes si la famille de tribus engendrées par les $X_i^{-1}(B)$ où $B \in \mathcal{B}(\mathbb{R})$ sont indépendantes.*

Propriété. *Soit $(X_i)_{i\in I}$ une famille de variables aléatoires. Ces variables sont indépendantes si et seulement si pour tout $k \in \mathbb{N}^*$, pour tout $i_1 < \ldots < i_k \in I^k$ et tout $(B_1, \ldots, B_k) \in \mathcal{B}(\mathbb{R})^k$,*

$$P(X_{i_1} \in B_1 \cap \cdots \cap X_{i_k} \in B_k) = P(X_{i_1} \in B_1) \times \cdots \times P(X_{i_k} \in B_k).$$

Propriété. *Si $X = (X_1, \cdots, X_d)$ est un vecteur aléatoire absolument continu de densité de probabilité f_X. Alors la famille $(X_1, \cdots, X_d)$ est une famille de variables aléatoires indépendantes si et seulement s'il existe d fonctions positives f_{X_i} telles que $f_X = \prod_{i=1}^d f_{X_i}$.*

Propriété. *Si $(X_1, \cdots, X_d)$ est une famille de variables aléatoires indépendantes alors $cov(X_i, X_j) = 0$ si $i \neq j$. La réciproque est fausse sauf dans le cas où $(X_1, \cdots, X_d)$ est un vecteur gaussien.*

3.3 Vecteurs aléatoires gaussiens

Le cas gaussien est un des seuls cas où il est facile d'exprimer la loi d'un vecteur aléatoire.

Définition. *On dit que $X = (X_1, \ldots, X_d)$ est un vecteur gaussien à valeurs dans $\mathbb{R}^d$ si et seulement si pour tout $(u_1, \ldots, u_d) \in \mathbb{R}^d$, la variable aléatoire $\sum_{i=1}^d u_i X_i$ est une variable gaussienne (toute combinaison linéaire des coordonnées de X est une variable gaussienne).*

Propriété. *La loi d'un vecteur gaussien $X = (X_1, \ldots, X_d)$ est entièrement déterminée par son espérance $\mathbb{E}X = {}^t(\mathbb{E}X_1, \ldots, \mathbb{E}X_d)$ et sa matrice de variance-covariance $\Sigma = \mathbb{E}[(X - \mathbb{E}X)\,{}^t(X - \mathbb{E}X)]$ et si le rang de Σ est d alors X est un vecteur aléatoire absolument continu de densité de probabilité:*

$$f_X(x_1, \ldots, x_d) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \exp\Big(-\frac{1}{2}\,{}^t({}^t(x_1, \ldots, x_d) - \mathbb{E}X)\Sigma^{-1}({}^t(x_1, \ldots, x_d) - \mathbb{E}X)\Big).$$

Propriété. *Si $X = (X_1, \ldots, X_d)$ est une suite de v.a. gaussiennes indépendantes alors X est un vecteur gaussien. Si les (X_i) sont gaussiennes mais ne sont pas indépendantes, ceci n'est plus forcément vrai.*

4 Théorèmes limites

4.1 Convergence d'une suite variables aléatoires

Propriété (Inégalité de Markov)**.** *Soit X une variable aléatoire positive dont l'espérance existe. Alors pour tout $\varepsilon > 0$,*

$$P(X > \varepsilon) \leq \frac{\mathbb{E}X}{\varepsilon}.$$

Voici une conséquence directe de cette inégalité:

Propriété (Inégalité de Bienaymé-Tchebitchev)**.** *Soit X une variable aléatoire dont l'espérance m et la variance σ^2 existent. Alors pour tout $\varepsilon > 0$,*

$$P(|X - m| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Définition. • *Convergence en probabilité.*

- *Convergence en loi.*

4.2 Théorèmes limite

Théorème (Loi Faible des Grands Nombres). *Soit* $(X_i)_{i\in\mathbb{N}^*}$ *une suite de variables aléatoires indépendantes et identiquement distribuées par rapport à une loi dont l'espérance* m *existe. Alors :*

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} \underset{n\to+\infty}{\overset{\mathcal{P}}{\longrightarrow}} m.$$

Théorème (Théorème de la limite central). *Soit* $(X_i)_{i\in\mathbb{N}^*}$ *une suite de variables aléatoires indépendantes et identiquement distribuées par rapport à une loi dont l'espérance* m *et la variance* σ^2 *existent. Alors :*

$$\sqrt{n}\left(\frac{\overline{X}_n - m}{\sigma}\right) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0,1).$$

5 Estimation et intervalles de confiance

Définition.

- *Estimateur d'un paramètre.*
- *Biais d'un estimateur.*
- *Risque quadratique d'un estimateur.*

Exemple.

Trois exemples sont à connaître:

1. Estimateur de l'espérance par la moyenne empirique.
2. Estimateur de la variance par la variance empirique.
3. Estimateurs du paramètre θ d'un n-échantillon $(X_1, \cdots, X_n)$ de loi uniforme sur $[0, \theta]$.

Définition.

- *Intervalle de confiance théorique de niveau* α *d'un paramètre.*
- *Réalisation d'un intervalle de confiance.*

Exemple.

Intervalle de confiance dans le cadre du théorème de la limite central.