SCI03 - - Analyse de données expérimentales

Rappels de probabilités

Thierry Denœux[1]

[1]Université de Technologie de Compiègne
tél : 44 96
tdenoeux@hds.utc.fr

Automne 2014

Expérience aléatoire

- On appelle expérience aléatoire une expérience qui, répétée plusieurs fois dans des conditions opératoires identiques, produit des résultats qui peuvent être différents.
- Les exemples classiques sont issus des jeux de hasard : lancer d'un ou plusieurs dés, tirage dans une urne, etc.

Modélisation

- Mathématiquement, la notion d'expérience aléatoire $\mathcal{E}$ se formalise en définissant :
  1. un ensemble fondamental Ω définissant l'ensemble des résultats possibles de $\mathcal{E}$, appelés événements élémentaires ;
  2. un ensemble $\mathcal{A}$ de parties de Ω (événements), clos par complémentation et union dénombrable (tribu ou σ-algèbre) ;
  3. une fonction $\mathbb{P} : \mathcal{A} \to [0, 1]$, appelée mesure ou distribution de probabilité, qui à tout événement A associe un nombre $\mathbb{P}(A)$ appelé probabilité de cet événement.

Probabilité

Définition

- $\mathbb{P}(\Omega) = 1$;
- Pour toute famille $(A_i)_{i \in I}$ finie ou dénombrable d'événements deux à deux disjoints ($A_i \cap A_j = \emptyset$, $\forall i, j \in I$),

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i)$$

  (additivité).
- La structure $(\Omega, \mathcal{A})$ est appelée espace probabilisable, et $(\Omega, \mathcal{A}, \mathbb{P})$ espace probabilisé.

Probabilité

Interprétation

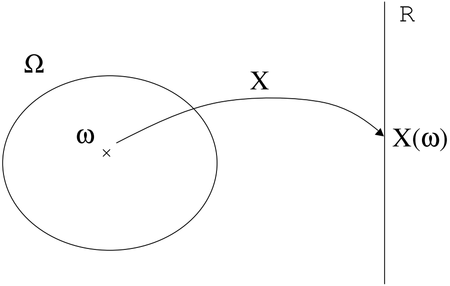
- $\mathbb{P}(A)$ s'interprète comme la limite de la fréquence de réalisation de l'événement A, lorsque le nombre n de répétitions de l'expérience aléatoire tend vers l'infini :

$$\mathbb{P}(A) = \lim_{n\to\infty} \frac{n_A}{n},$$

n_A étant le nombre de réalisations de l'événement A obtenu au cours de n répétitions de $\mathcal{E}$.

Définition

- On appelle variable aléatoire (réelle) une grandeur numérique dont la valeur est fonction du résultat d'une expérience aléatoire.

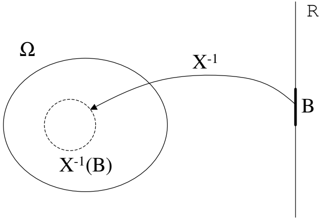


Loi de probabilité

- Soit B un intervalle de $\mathbb{R}$. On peut définir la probabilité que la v.a. X prenne sa valeur dans B comme

$$\mathbb{P}_X(B) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\}) = \mathbb{P}(X^{-1}(B)),$$

(On suppose que $X^{-1}(B) \in \mathcal{A}$ pour tout B).



- La donnée de $\mathbb{P}_X(B)$ pour tout intervalle B définit la loi (ou distribution) de probabilité de X.

Fonction de répartition

- Pour décrire complètement $\mathbb{P}_X$, il suffit de donner la fonction de répartition de X

$$F_X : \begin{array}{l} \mathbb{R} \to [0,1] \\ x \to \mathbb{P}_X(]-\infty, x]), \end{array}$$

ce que l'on note $F_X(x) = \mathbb{P}(X \leq x)$.

- Propriétés suivantes :
  1. croissance ;
  2. continuité à droite : $\lim_{\epsilon \to 0, \epsilon > 0} F_X(x+\epsilon) = F_X(x)$;
  3. conditions aux limites :
     $\lim_{x \to -\infty} F_X(x) = 0, \qquad \lim_{x \to \infty} F_X(x) = 1.$

Fonction de probabilité (v.a. discrète)

- Une loi de probabilité peut également être définie par la fonction qui à chaque élément de V_X associe sa probabilité (loi discrète), ou par un équivalent dans le cas continu.
- Soit X une v.a. discrète. On appelle fonction de probabilité de X la fonction :

$$p_X : \begin{array}{l} \mathbb{R} \to [0,1] \\ x \to \mathbb{P}_X(\{x\}). \end{array}$$

- Propriétés :

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \mathbb{P}_X(B) = \sum_{x \in B} p_X(x)$$

$$\sum_{x \in V_X} p_X(x) = \mathbb{P}(V_X) = 1.$$

Fonction de densité de probabilité (v.a. continue)

- Dans le cas continu, la notion de fonction de densité de probabilité remplace celle de fonction de probabilité.
- Une v.a. continue X est dite absolument continue s'il existe une fonction $f_X : \mathbb{R} \to \mathbb{R}_+$, (densité) t.q.

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \mathbb{P}_X(B) = \int_B f_X(t)dt.$$

- Propriétés :

$$\int_{-\infty}^{+\infty} f_X(t)dt = \mathbb{P}_X(\mathbb{R}) = 1 \quad \text{et} \quad F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

- La quantité $f_X(x)$ n'est pas une probabilité. C'est l'intégrale de f_X sur une intervalle qui est une probabilité.

Loi de Bernoulli

- Soit une expérience aléatoire $\mathcal{E}$, et A un événement associé à $\mathcal{E}$, de probabilité p.
- Soit X la fonction indicatrice de A, définie pour tout $\omega \in \Omega$ par :

$$X(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{sinon.} \end{cases}$$

- Fonction de probabilité :

$$p_X(1) = \mathbb{P}(A) = p, \qquad p_X(0) = \mathbb{P}(\overline{A}) = 1 - p$$

- Par définition, on dit que X suit une loi de Bernoulli de paramètre p, ce que l'on note $X \sim \mathcal{B}(p)$.

Loi binomiale

- Supposons que l'on répète n fois l'expérience précédente, de manière indépendante. On définit n v.a. de Bernoulli $X_i \sim \mathcal{B}(p)$, $i = 1, \ldots, n$.
- Le nombre de réalisations de l'événement A est

$$Y = \sum_{i=1}^{n} X_i,$$

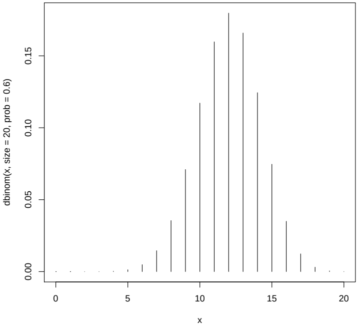
- Par définition, $Y \sim \mathcal{B}(n, p)$. Le domaine de définition est $V_X = \{0, \ldots, n\}$.
- Fonction de probabilité de Y :

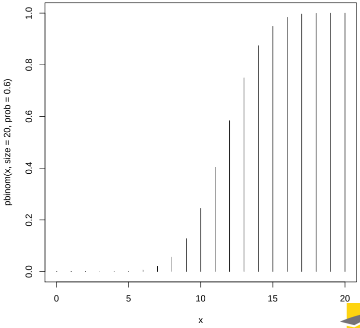
$$p_Y(y) = C_n^y p^y (1 - p)^{n-y}, \quad \forall y \in \{0, \ldots, n\}.$$

Loi binomiale en R

- Fonction de probabilité : `dbinom(x, size, prob)`.
- Fonction de répartition : `pbinom(q, size, prob)`.

```
x<-0:20
plot(x,dbinom(x,size=20,prob=0.6),type='h')
```

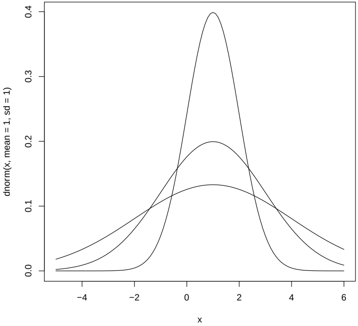




Loi normale

- La loi normale $\mathcal{N}(\mu, \sigma^2)$ est définie par la fonction de densité de probabilité :

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$



Loi normale
Fonction de répartition

- La fonction de répartition correspondante n'a pas d'expression analytique.
- On l'exprime communément à l'aide de la fonction ϕ (fonction de répartition de la loi $\mathcal{N}(0,1)$, appelée loi normale centrée-réduite) :

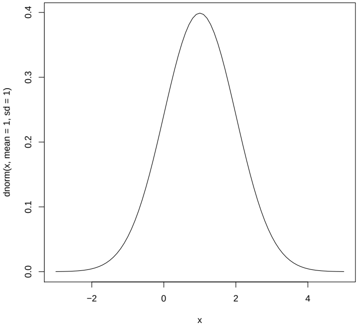
$$X \sim \mathcal{N}(\mu, \sigma^2) \Leftrightarrow F_X(x) = \phi\left(\frac{x-\mu}{\sigma}\right), \forall x \in \mathbb{R}.$$

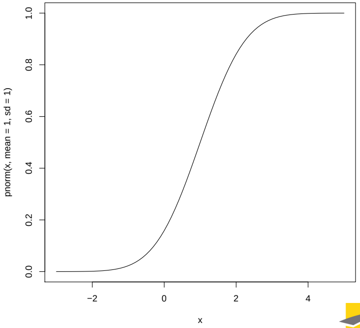
- Propriété : $\phi(-x) = 1 - \phi(x)$ pour tout $x \in \mathbb{R}$.

Loi normale en R

- Fonction de probabilité : `dnorm(x, mean, sd)`.
- Fonction de répartition : `pnorm(q, mean, sd)`.

```
x<-seq(-3,5,0.1)
plot(x,dbinom(x,mean=1,sd=1),type='l')
```





Espérance mathématique
Définition

- L'espérance (mathématique) d'une variable aléatoire réelle représente la « valeur moyenne »prise par cette variable aléatoire.
- Définition :

$$\mathbb{E}(X) = \begin{cases} \displaystyle\sum_{x \in V_X} x p_X(x) & \text{si } X \text{ est une v.a. discrète,} \\ \displaystyle\int_{\mathbb{R}} x f_X(x) dx & \text{si } X \text{ est une v.a. continue} \end{cases}$$

si ces quantités existent.

Espérance mathématique

Propriétés

- Si le graphe de la fonction de densité est symétrique par rapport à une valeur a (c'est-à-dire si $f_X(a-x) = f_X(a+x)$ pour tout x), alors $\mathbb{E}(X) = a$.
- Propriétés :
  - $\forall \alpha, \beta \in \mathbb{R},\ \mathbb{E}(\alpha X + \beta) = \alpha \mathbb{E}(X) + \beta$.
  - Soit X une v.a. et φ une fonction $\mathbb{R} \to \mathbb{R}$. On a :

$$\mathbb{E}(\varphi(X)) = \begin{cases} \displaystyle\int_{\mathbb{R}} \varphi(x) f_X(x) dx & \text{si } X \text{ est une v.a. continue,} \\ \displaystyle\sum_{x \in V_X} \varphi(x) p_X(x) & \text{si } X \text{ est une v.a. discrète.} \end{cases}$$

Variance
Définition

- C'est une mesure de dispersion de la v.a. autour de son espérance :
$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right]$$
- La racine carrée de la variance est appelée écart-type de la v.a. X et notée σ.

Variance
Propriétés

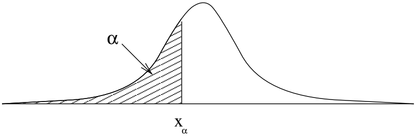
- $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.
- $\forall \alpha, \beta \in \mathbb{R}$, $\mathrm{Var}(\alpha X + \beta) = \alpha^2 \mathrm{Var}(X)$.
- Inégalité de Bienaymé-Tchebycheff :

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathrm{Var}(X).$$

(Plus la variance est petite, plus faible est la probabilité que X s'écarte de son espérance d'une valeur donnée.)

Fractile

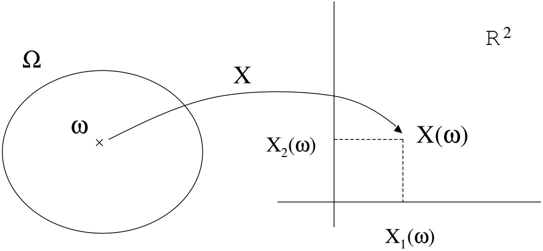
- Soit X une v.a. continue de fonction de répartition F_X.
- Pour tout $\alpha \in]0, 1[$, on appelle fractile (quantile) d'ordre α de X la quantité
$$x_\alpha = F_X^{-1}(\alpha).$$
- On a donc, par définition, $F_X(x_\alpha) = \alpha$, ou encore $\int_{-\infty}^{x_\alpha} f_X(x)dx = \alpha$.



- Fractiles de la loi normale en R :
`qnorm(p, mean = 0, sd = 1).`

Vecteur aléatoire
Définition

- On appelle vecteur aléatoire (réel) un vecteur de $\mathbb{R}^n$ dont les composants sont fonctions du résultat d'une expérience aléatoire $\mathcal{E}$.
- Mathématiquement, c'est une application $\mathbf{X}$ mesurable d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ dans $\mathbb{R}^n$.



Vecteur aléatoire

Loi de probabilité

- La loi de probabilité de **X** se définit comme une application

$$\mathbb{P}_{\mathbf{X}}(B) = \mathbb{P}(\{\omega \in \Omega \mid \mathbf{X}(\omega) \in B\}) = \mathbb{P}(\mathbf{X}^{-1}(B)),$$

pour tout sous-ensemble B de $\mathbb{R}^n$ pour lequel cette probabilité est définie.

- $\mathbb{P}_{\mathbf{X}}$ est appelée loi jointe du vecteur aléatoire **X**.
- La loi $\mathbb{P}_{X_i}$ de chaque composante X_i est appelée loi marginale de X_i.
- En général, la loi de **X** ne se déduit pas des lois marginales de ses composants.

Vecteur aléatoire

Fonction de répartition

Comme dans le cas monodimensionnel, on peut décrire $\mathbb{P}_{\mathbf{X}}$ de deux façons :

1. Par la fonction de répartition de **X**, $F_{\mathbf{X}} : \mathbb{R}^n \to [0, 1]$ définie par

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1; X_2 \leq x_2; \ldots; X_n \leq x_n).$$

2. Par la fonction de masse (cas discret) ou de densité de probabilité (cas continu) de **X**.

Vecteur aléatoire
Masse et densité de probabilité

- Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un vecteur aléatoire.
- Si $\mathbf{X}$ est discret (chaque composante est une v.a. discrète), on appelle fonction de masse de probabilité de $\mathbf{X}$ la fonction

$$p_{\mathbf{X}}(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1; X_2 = x_2; \ldots; X_n = x_n).$$

- Si $\mathbf{X}$ est continu (chaque composante est une v.a. continue), on appelle fonction de densité de probabilité de $\mathbf{X}$ la fonction $f_{\mathbf{X}} : \mathbb{R}^n \to \mathbb{R}_+$ telle que

$$\mathbb{P}_{\mathbf{X}}(B) = \int_B f_{\mathbf{X}}(x_1, \ldots, x_n) dx_1 \ldots dx_n,$$

pour tout sous-ensemble B de $\mathbb{R}^n$ pour lequel cette intégrale est définie.

Loi multinomiale

- Soient $\mathcal{E}$ une expérience aléatoire modélisée par un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, et $A_1, \ldots, A_K$ un système complet d'événements, vérifiant $\bigcup_{k=1}^{K} A_k = \Omega$ et $A_i \cap A_j = \emptyset \; \forall i, j$. Soit $p_k = \mathbb{P}(A_k)$.
- On répète n fois l'expérience, et on note N_k le nombre de réalisation de l'événement A_k.
- Le vecteur aléatoire $N = (N_1, \ldots, N_K)$ suit une loi multinomiale de paramètre $(n, p_1, \ldots, p_K)$,

$$N \sim \mathcal{M}(n, p_1, \ldots, p_K).$$

- Pour tout $(n_1, \ldots, n_k) \in \{0, \ldots, n\}^K$:

$$p_N(n_1, \ldots, n_K) = \begin{cases} \dfrac{n!}{n_1! \ldots n_k!} p_1^{n_1} \ldots p_K^{n_K} & \text{si } \sum_{k=1}^{K} n_k = n, \\ 0 & \text{sinon.} \end{cases}$$

Loi multinomiale

Exemple en R

- Soit une urne contenant 20% de boules noires, 30% de boules blanches et 50% de boules rouges.
- Quelle est la probabilité d'obtenir 5 boules noires, 5 boules blanches et 10 boules rouges sur 20 tirages avec remise ?

```
> dmultinom(c(5,5,10),size=20,prob=c(0.2,0.3,0.5))

[1] 0.03535537
```

Loi normale bidimensionnelle

- Soit $\mathbf{X} = (X_1, X_2)$ le vecteur aléatoire bidimensionnel de fonction de densité

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right]$$

  avec $\sigma_1, \sigma_2 > 0$ et $\rho \in [-1, 1]$.
- Par définition, $\mathbf{X}$ suit une loi normale bidimensionnelle.
- On montre que $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$.

Covariance, corrélation

- Étant données deux variables aléatoires X et Y, on appelle covariance entre X et Y la quantité

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))].$$

- Propriétés :
  1. $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
  2. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$.
  3. $\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X, Y)$.
- On appelle matrice de variance du vecteur aléatoire $X = (X_1, \ldots, X_n)$ la matrice $\mathrm{Var}(X) = \Sigma$ de dimension (n, n) et de terme général $\mathrm{Cov}(X_i, X_j)$.
- On appelle coefficient de corrélation ρ entre X et Y la covariance $\mathrm{Cov}(X, Y)$ divisée par le produit des écarts-types de X et de Y. On a : $-1 \leq \rho \leq 1$.

utc Université de Technologie Compiègne

Indépendance de variables aléatoires
Exemple

- Soient X la taille et Y le poids d'un individu choisi au hasard dans une population $\mathcal{P}$.
- Supposons que 10% d'individus mesurent moins d'1,60 m et 10% d'individus pèsent moins de 60 kg.
- On a donc $\mathbb{P}(X \leq 1.6) = 0.1$ et $\mathbb{P}(Y \leq 60) = 0.1$.
- Cependant, on observera en général

$$\mathbb{P}(X \leq 1.6; Y \leq 60) > (0.1)^2 = 0.01.$$

- Les individus mesurant moins d'1.6 m et ceux pesant moins de 60 kg ont tendance à être les mêmes : il y a un lien entre ces deux caractères, les v.a. X et Y ne sont pas indépendantes.

Indépendance de variables aléatoires

Définition

- Les v.a. $X_1, \dots, X_n$ sont indépendantes si et seulement si

$$F_X(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n,$$

ou encore si et seulement si

$$p_X(x_1, \dots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i) \quad \text{[cas discret]}$$

$$ou \quad f_X(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \quad \text{[cas continu]}$$

- Intuitivement, la notion d'indépendance entre v.a. correspond à l'absence de relation entre ces variables.

Indépendance de variables aléatoires

Lien avec la covariance

- Si X et Y indépendantes, alors $\mathrm{Cov}(X, Y) = 0$;
- La réciproque est fausse ;
- Dans le cas où (X, Y) suit une loi normale bidimensionnelle, alors X et Y indépendantes si et seulement si $\mathrm{Cov}(X, Y) = 0$.
- On a alors $\rho = 0$ et $f_X(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$.

Somme de variables aléatoires

- Soit $(X_1, \ldots, X_n)$ un vecteur aléatoire.
- On a

$$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i).$$

- Si les v.a. $X_1, \ldots, X_n$ sont indépendantes, alors

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$