Compléments sur R
Notions de base en probabilités

Master M1 Informatique - Université d'Orléans

Christel Vrain
Christel.Vrain@univ-orleans.fr

LIFO (Laboratoire d'Informatique Fondamentale d'Orléans)
Département Informatique - Faculté Sciences
Université d'Orléans

Références

- An Introduction to R. W. N. Venables, D. M. Smith and the R Development Core Team
- Enseignements de Statistique en Biologie. A.B. Dufour D. Chessel J.R. Lobry Contributeurs S. Mousset S. Dray. Maintenance système S. Penel
  Site web : http://pbil.univ-lyon1.fr/R/enseignement.html
- Premiers pas en statistique. Y. Dodge. Springer 2001.

Probabilités :événements

- Expérience aléatoire : on connaît l'ensemble des résultats possibles, on ne peut prédire avec certitude le résultat
- Ensemble fondamental Ω : ensemble de tous les résultats possibles de l'expérience
- Evénement : résultat d'une expérience, sous-ensemble de l'ensemble fondamental
- 2 événements particuliers : événement certain ($= \Omega$), événement vide ($= \emptyset$)

Exemple : lancer successif de 2 dés
$\Omega = \{(1,1),(1,2),\dots,(3,3),\dots(6,6)\}$

- événement A : la somme des points vaut 6, $A = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$
- événement B : la somme des points est paire, $B = \{(1,1),(1,3),(1,5),(2,2),\dots,(6,6)\}$
- événement vide : la somme des points vaut 0
- événement certain : la somme des points est inférieure (au sens large) à 13

Probabilités : Opérations

- Opérations sur des événements
  - négation de A, $\overline{A}$: se réalise lorsque A ne se réalise pas
  - conjonction de 2 événements : $A \cap B$
  - disjonction de 2 événements : $A \cup B$
- 2 événements sont incompatibles si $A \cap B = \emptyset$

Lancer un dé : $\Omega = \{1,2,3,4,5,6\}$

- A: obtenir un nombre de points pair, $A = \{2,4,6\}$
- B: obtenir un chiffre inférieur (ou égal) à 5, $B = \{1,2,3,4,5\}$
- négation ou complémentaire de A : $\overline{A} = \{1,3,5\}$
- événement "A et B" : $A \cap B = \{2,4\}$
- événement "A ou B" : $A \cup B = \{1,2,3,4,5,6\}$

Probabilités : Axiomes

- Si Ω a n éléments $\{x_1, \ldots, x_n\}$, on associe à chaque x_i une probabilité $p(x_i)$
  - $p(x_i) \geq 0$, pour tout $i = 1, \ldots n$
  - $\Sigma_{i=1}^{n} p(x_i) = 1$
- $p(\Omega) = 1$
- $p(\emptyset) = 0$
- $p(\overline{A}) = 1 - p(A)$
- Si A et B sont incompatibles, $p(A \cup B) = p(A) + p(B)$
- sinon $p(A \cup B) = p(A) + p(B) - p(A \cap B)$
- Si $A_1, \ldots, A_m$ sont mutuellement exclusifs ($A_i \cap A_j = \emptyset$, pour tout i et j, $i \neq j$)
  $p(A_1 \cup A_2 \ldots \cup A_n) = p(A_1) + p(A_2) + \ldots + p(A_n)$

Un exemple

Les parents d'Albert (a), de Brigitte (b), de Charles (c) et de Danielle (d) ne peuvent utiliser leur billet d'abonnement au théâtre de la ville. Ils décident de donner leur billet à 2 enfants choisis aléatoirement. Quelle est la probabilité qu'un garçon et une fille soit choisis ?

Probabilités conditionnelles :

- Probabilité conditionnelle de B sachant A
  $$P(B|A) = \frac{P(A \cap B)}{P(A)}$$
- Théorème de Bayes : $P(B|A).P(A) = P(A|B).P(B)$

	rapide	*pas rapide*	*total*
confortable	40	10	50
inconfortable	20	30	50
total	60	40	100

Ω : ens. des 100 voitures $\quad p(v_i) = 1/100$ pour toute voiture i

- A : choisir une voiture rapide
- B : choisir une voiture confortable

$p(A) = \frac{60}{100}$ $\quad p(B) = \frac{50}{100}$ $\quad p(A \cap B) = \frac{40}{100}$ $\quad p(B|A) = \frac{0.4}{0.6}$

Application : classifieur bayésien

Evénements indépendants :

- Deux événements sont indépendants si et seulement si $P(A \cap B) = P(A).P(B)$
- Si A et B sont indépendants, $P(B|A) = P(B)$ et $P(A|B) = P(A)$

Lancer d'un dé

- A : obtenir un nombre inférieur ou égal à 4
- B : obtenir un nombre pair

Les événements sont-ils indépendants ?
$p(A) = 4/6 \quad p(B) = 3/6 \quad p(A \cap B) = 2/6$

Classifieur bayésien

- un ensemble fini de classes possibles $C_1, \dots C_k$
- un langage de représentation des observations : des attributs (variables, caractéristiques) $A_1, \dots, A_n$
- une observation o décrite par ses valeurs sur $A_1, \dots, A_n$
  $A_1 = v_1, \dots A_n = v_n$

o : $M = faible \; \& \; A = moyen \; \& \; R = village \; \& \; E = oui$

classer o

trouver la classe C_i qui maximise $p(C_j | A_1 = v_1, \dots, A_n = v_n)$, $j = 1, \dots, k$

$$\begin{aligned} i &= argmax_{j=1,\dots,k} p(C_j | A_1 = v_1, \dots, A_n = v_n) \\ &= argmax_{j=1,\dots,k} \frac{p(A_1 = v_1, \dots, A_n = v_n | C_j) \times p(C_j)}{p(A_1 = v_1, \dots A_n = v_n)} \\ &= argmax_{j=1,\dots,k} p(A_1 = v_1, \dots, A_n = v_n | C_j) \times p(C_j) \end{aligned}$$

Classifieur bayésien naïf

- Trop de probabilités à calculer
  $Pr(A_1 = v_1, \dots, A_n = v_n | C_j) \times P(C_j)$
  - $p(C_j) \to k$ probabilités
  - $p(A_1 = v_1, \dots, A_n = v_n | C_j) \to D_{A_1} \times \dots \times D_{A_n} \times k$ probabilités, où D_{A_u} représente le nombre de valeurs de l'attribut A_u
- Hypothèse d'indépendance des attributs par rapport à la classe. A ne pas confondre avec l'indépendance des attributs.

$$p(A_1 = v_1, \dots, A_n = v_n | C_j) = p(A_1 = v_1 | C_j) \times \dots \times p(A_n = v_n | C_j)$$

classer o avec le classifieur bayésien naïf

trouver la classe C_i qui maximise
$p(A_1 = v_1 | C_j) \times \dots \times p(A_n = v_n | C_j) \times p(C_j)$, $j = 1, \dots, k$

m-estimation pour éviter d'avoir des probabilités nulles

Exemple

Client	M	A	R	E	I
1	moyen	moyen	village	oui	oui
2	élevé	moyen	bourg	non	non
3	faible	âgé	bourg	non	non
4	faible	moyen	bourg	oui	oui
5	moyen	jeune	ville	oui	oui
6	élevé	âgé	ville	oui	non
7	moyen	âgé	ville	oui	non
8	faible	moyen	village	non	non

Classer o : $M = faible \; \& \; A = moyen \; \& \; R = village \; \& \; E = oui$
On estime

- $p(M = faible | I = oui) \times p(A = moyen | I = oui) \times p(R = village | I = oui) \times p(E = oui | I = oui) \times p(I = oui) \simeq \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{3}{3} \times \frac{3}{8} = \frac{1}{36}$
- $p(M = faible | I = non) \times p(A = jeune | I = non) \times p(R = village | I = non) \times p(E = oui | I = non) \times p(I = non) \simeq \frac{2}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{5}{8} = \frac{1}{125}$

Variable aléatoire

- Variable aléatoire X : fonction à valeurs réelles définies sur Ω, la valeur de cette variable dépend du résultat de l'expérience.
- Loi de probabilité p : fonction qui associe à chaque valeur de x de X sa probabilité $P(X = x)$
- Fonction discrète lorsque l'ensemble $\mathcal{X}$ des valeurs prises par X est dénombrable ($\mathcal{X} = \{x_1, \dots, x_n, \dots\}$)
  $\Sigma_{x \in \mathcal{X}} p(x) = 1$

Lancer d'une pièce 3 fois. Chaque face est équiprobable.
$\Omega = \{PPP, PPF, PFP, PFF, FPP, FPF, FFP, FFF\}$

- Nombre de piles : variable pouvant prendre 4 valeurs (0, 1, 2 ou 3)
  $p(0) = p(X = 0) = 1/8$ $p(1) = p(X = 1) = 3/8$ $p(2) = p(X = 2) = 3/8$ $p(3) = p(X = 3) = 1/8$
- Nombre de piles dans les 2 premiers essais : variable pouvant prendre 3 valeurs (0,1, 2)
  $p(0) = p(X = 0) = 2/8$ $p(1) = p(X = 1) = 4/8$ $p(2) = p(X = 2) = 2/8$

Variable aléatoire

Fonction de répartition d'une variable aléatoire X : la fonction F définie par :
$F(x) = P(X \leq x)$

- F est croissante
- pour tout x, $F(x) \in [0, 1]$
- $F(-\infty) = 0$
- $F(+\infty) = 1$

Lancer d'une pièce 3 fois. Chaque face est équiprobable.
$\Omega = \{PPP, PPF, PFP, PFF, FPP, FPF, FFP, FFF\}$

- Nombre de piles : variable pouvant prendre 4 valeurs (0, 1, 2 ou 3)
  $p(0) = 1/8 \quad p(1) = 3/8 \quad p(2) = 3/8 \quad p(3) = 1/8$
  $F(0) = p(X \leq 0) = 1/8 \quad F(1) = p(X \leq 1) = 4/8$
  $F(2) = p(X \leq 2) = 7/8 \quad F(3) = p(X \leq 3) = 8/8$

Espérance - Variance d'une variable discrète

Pour une variable X discrète :

- Espérance : $E(X) = \Sigma_{x_i \in \mathcal{X}} x_i.p(x_i)$
- Variance : $\sigma^2 = Var(X) = \Sigma_{x \in \mathcal{X}} (x_i - \mu)^2.p(x_i) = E(X - \mu)^2$

Lancer d'une pièce 3 fois. Chaque face est équiprobable.
$\Omega = \{PPP, PPF, PFP, PFF, FPP, FPF, FFP, FFF\}$

- Nombre de piles : variable X pouvant prendre 4 valeurs (0, 1, 2 ou 3)
  $p(0) = 1/8 \quad p(1) = 3/8 \quad p(2) = 3/8 \quad p(3) = 1/8$
- On lance 3 fois les dés. On ne gagne rien s'il n'y a pas de pile. On gagne 1 euro s'il y a un ou deux piles. On perd 2 euros s'il y a 3 piles $\rightarrow$ G variable aléatoire du gain. Espérance de G ?
  $E(G) = 0 \times p(G = 0) + 1 \times p(G = 1) - 2 \times p(G = -2)$
  $= 0 \times 1/8 + 1 \times 6/8 - 2 \times 1/8 = 1/2$
  Variance de G ?
  $Var(G) = (0-1/2)^2 \times p(G = 0) + (1-1/2)^2 \times p(G = 1) + (-2-1/2)^2 \times p(G = -2)$
  $= 1/4 \times 1/8 + 2/4 \times 3/8 + 25/4 \times 1/8$

Espérance - Variance d'une variable discrète

- $E(a.X + b) = a.E(X) + b$
- $E(X + Y) = E(X) + E(Y)$
- Si X et Y sont indépendantes, $E(X.Y) = E(X).E(Y)$
- $Var(aX + b) = a^2.Var(X)$
- Si X et Y sont indépendantes, $Var(X + Y) = Var(X) + Var(Y)$
- Se généralise pour n variables indépendantes

Loi conjointe - Loi marginale

Soient 2 variables aléatoires X à valeurs dans $\{x_1, \ldots, x_n\}$ et Y à valeurs dans $\{y_1, \ldots, y_p\}$.

- Loi de probabilité conjointe :
  $p(x, y) = p((X = x) \cap (Y = y))$ pour $x = x_1, \ldots, x_n$ et $y = y_1, \ldots, y_p$
- Loi marginale de X , composante d' une loi conjointe (X, Y)
  $P_X(x_i) = P(X = x_i) = \Sigma_j P((X = x_i) \cap (Y = y_j))$
- Si X et Y sont indépendantes, $p(x, y) = p(x) \times p(y)$

Loi conjointe - Loi marginale. Exemple

Lancer d'une pièce 3 fois. Chaque face est équiprobable.
$\Omega = \{PPP, PPF, PFP, PFF, FPP, FPF, FFP, FFF\}$

- Nombre de piles :
  $p(0) = 1/8 \quad p(1) = 3/8 \quad p(2) = 3/8 \quad p(3) = 1/8$
- Nombre de piles dans les 2 premiers essais : $p(0) = 2/8 \quad p(1) = 4/8 \quad p(2) = 2/8$

Loi conjointe :

		X			
		0	1	2	3
	0	1/8	1/8	0	0
Y	1	0	2/8	2/8	0
	2	0	0	1/8	1/8

Covariance

Soients 2 variables aléatoires X à valeurs dans $\{x_1, \ldots, x_n\}$ et Y à valeurs dans $\{y_1, \ldots, y_p\}$.

- Covariance de X et Y
$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = \Sigma_{i,j}(x_i - \mu_X).(y_j - \mu_Y).p(x_i, y_j)$
- Définition équivalente
  $Cov(X, Y) = E(XY) - \mu_X.\mu_Y = \Sigma_{i,j}x_i.y_j.p(x_i, y_j) - \Sigma_i x_i.p(x_i)\Sigma_j y_j.p(y$
- Propriétés :
  - La covariance peut être positive ou négative.
  - $Cov(X, X) = Var(X, X)$
  - Si X et Y sont indépendantes : $Cov(X, Y) = 0$

Epreuves de Bernouilli

- Une suite d'épreuves est de Bernouilli si
  - A chaque épreuve on a le même ensemble fondamental $\Omega = \{échec, succès\}$.
  - La probabilité de chaque événement reste constante : $P(succès) = p, P(échec) = 1 - p$
  - Les épreuves sont mutuellement indépendantes.
- Une variable aléatoire X suit une loi de Bernouilli de paramètre p si $P(X = 1) = p$ et $P(X = 0) = q = 1 - p$
  - $\mu = E(X) = p = (1 \times P(X = 1) + 0 \times P(X = 0))$
  - $var(X) = p \times (1 - p)$
- Une suite d'épreuves de Bernouilli est représentée par les variables aléatoires X_i, $i = 1, 2, \ldots$, où chaque X_i suit la même loi de Bernouilli.

Loi binomiale

- Loi binômiale = somme d'une série d'épreuves de Bernouilli
- Probabilité d'obtenir x succès ($x = 0, 1, 2, \ldots, n$) parmi n tirages : $P(X = x) = C_n^x \times p^x \times q^{n-x}$ avec $C_n^x = \frac{n!}{x!(n-x)!}$
- $\mu = n \times p$ (somme de n variables de Benoulli)
- $\sigma^2 = n \times p \times q$ (somme de n variables de Benoulli indépendantes)

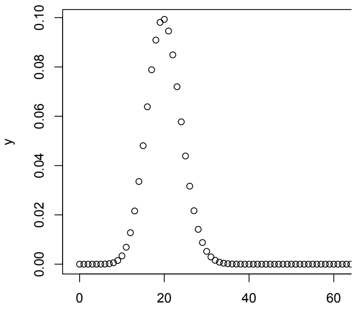
Exemples avec R

- *rxxxx* : génération de nombres aléatoires suivant la loi *xxxx*
- d*xxxx* : étant donné un ensemble de valeurs, retourne la hauteur de la probabilité de distribution en chaque point.

```
> rbinom(5,size=100,prob=0.2)
[1] 23 19 22 18 23
 # mu=100x0.2
 # var=100x0.2x0.8=16 sigma=4

> rbinom(5,size=100,prob=0.8)
[1] 79 80 80 81 87
> rbinom(5,size=100,prob=0.5)
[1] 40 58 52 41 53

> x<-seq(0,100,by=1)
> y<-dbinom(x,100,0.2)
> y
[1] 2.037036e-10 5.092590e-09 6.302080e-08 5.146699e-07 3.120186e-06
```



Exemples avec R

- p*xxxx* : probabilité $P(X \leq x)$ qu'un nombre tiré aléatoirement suivant cette loi soit inférieur à un nombre x donné
- q*xxxx* : nombre x tel que $P(X \leq x) = n$ (n nombre donné)

```
> pbinom(20, size=100,prob=0.2)
[1] 0.5594616
> pbinom(50,size=100,prob=0.2)
[1] 1
```

```
> qbinom(0.8,size=100,prob=0.2)
[1] 23
> pbinom(23,size=100,prob=0.2)
[1] 0.8109128
> pbinom(22,size=100,prob=0.2)
[1] 0.7389328
```


Variable aléatoire continue

Soit X une variable aléatoire continue à valeurs dans les réels.
On associe une probabilité à chaque intervalle de valeurs.

- Fonction de répartition :
  $$F(x) = P(X \leq x)$$
- Propriétés
  - $lim_{x \to +\infty} = 1$
  - $lim_{x \to -\infty} = 0$
  - F est continue dérivable
  - F est croissante
  - $P(a \leq X \leq b) = F(b) - F(a)$

Exemple : Position exacte de la grande aiguille lorsqu'une montre tombe en panne
$P(30 \leq X \leq 90) = \frac{90-30}{360}$
Quelle est la fonction de répartition F ?
Tracer cette fonction de répartition.

Fonction de densité

- densité f de la variable aléatoire X de fonction de répartition F : dérivée de la fonction F au point x
  $$f(x) = lim_{\Delta \to 0} \frac{F(x+\Delta) - F(x-\Delta)}{2\Delta}$$
- Propriété :
  $$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Exemple : Position exacte de la grande aiguille lorsqu'une montre tombe en panne
Quelle est la fonction de densité f ?

Espérance - Variance

- Espérance de X de densité f sur une intervalle I
$$E(X) = \mu = \int_I x.f(x)dx$$
- Variance de X de densité f sur une intervalle I
$$Var(X) = \int_I (x-\mu)^2.f(x)dx = \int_I x^2.f(x) - \mu^2$$

Exemple : Position exacte de la grande aiguille lorsqu'une montre tombe en panne
Quelle est son espérance mathématique ?
Quelle est sa variance ?

Loi uniforme

- Loi uniforme sur une intervalle $[a, b]$: loi d densité constante sur $[a, b]$
- Fonction de densité : $f(x) = \frac{1}{b-a}$
- $F(x) = \int_a^x \frac{dx}{b-a} = \frac{x-a}{b-a}$, pour $a \leq x \leq b$
- $\mu = \frac{a+b}{2}$
- $var(X) = \frac{(b-a)^2}{12}$

Loi normale

Soit X une variable aléatoire à valeurs réelles. La loi normale est définie par 2 paramètres, la moyenne μ et σ^2, notée $X \sim \mathcal{N}(\mu, \sigma^2)$

- Fonction de densité
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2), \; -\infty < x < +\infty$$
- Fonction de répartition
$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)dx, \; -\infty < x < +\infty$$
- Loi standard ou loi normale centrée réduite $\mathcal{N}(0, 1)$
$$f(z) = \frac{1}{\sqrt{2\pi}} exp(-\frac{z^2}{2}), \; -\infty < z < +\infty$$
  - $f(z) = f(-z)$
  - Si $\Phi(z) = \frac{1}{\sqrt{2\Pi}}.\int_{-\infty}^{z} exp(-\frac{z^2}{2})dz$, alors $\Phi(-z) = 1 - \Phi(z)$
- Propriété : Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$ et $P(a < Z < b) = \Phi(b) - \Phi(a)$

Soit $X \sim \mathcal{N}(23, 1.5^2)$. Calculez en fonction de Φ, $P(20 < X < 25)$.

Exemples avec R

- r*xxxx* : génération de nombres aléatoires suivant la loi *xxxx*
- d*xxxx* : étant donné un ensemble de valeurs, retourne la hauteur de la probabilité de distribution à caque point. (fonction densité)
- p*xxxx* : probabilité $P(X \leq x)$ qu'un nombre tiré aléatoirement suivant cette loi soit inférieur à un nombre x donné
- q*xxxx* : nombre x tel que $P(X \leq x) = n$ (n nombre donné)

```
> x<-rnorm(100)
> hist(x, proba=TRUE)
> pnorm(2)
[1] 0.9772499
> pnorm(-2)
[1] 0.02275013
> qnorm(0.97)
[1] 1.880794
```