Rappels de probabilités

Florence Perronnin

Univ. Grenoble Alpes, LIG, Inria

September 18, 2018

Outline

Understanding probability

Probability terminology may often be mistaken with common language: "fairly sure", "likely", "almost certain", "low chances"...
Everyone has a personal view of what is a probability. Somehow it involves assigning a number to characterize the outcome of some uncertain experiment (e.g., games of chance).
Most common perceptions:

- intuitive view (above)
- classical [Laplace 1814]
- frequentist [Bernoulli 1713]
- subjective (information)
- logical [Keynes 1921 and Carnap 1950]
- axiomatic [Kolmogorov 1933]

Understanding probability

Classical view

1654, Pascal, Fermat and the Chevalier de Méré.

How to divide the stakes if a game of chance is interrupted before the end? Compute favorable cases to possible outcomes and divide accordingly.

1663, Cardano:

compute

$$\frac{\#\text{favorable cases}}{\#\text{possible cases}}$$

to make a fair bet

The above formula is used by De Moivre and Laplace as a definition of probability.

However this definition relies implicitly on the *equiprobability assumption*.

Understanding probability

Classical view

De Méré's problem:

It seems equally likely to :

- event A: get a 6 after throwing 1 die 4 times
- event B: get a double-6 after throwing two dice 24 times

Argument #1 (Antoine Gombaud): $\mathbb{P}[A] = \frac{\text{4 chances to get a 6}}{\text{6 possibilities}} = \frac{2}{3}$

$\mathbb{P}[B] = \frac{\text{24 chances to get a double 6}}{\text{36 possibilities}}$

Here, probability was confused with expectation.

Argument #2 (Fermat) $\mathbb{P}[A] = 0.518$ and $\mathbb{P}[B] = 0.491$ (correct answer).

Understanding probability

Frequentist view

Your company has 30 license tokens for a popular mathematical software. 10 license tokens are in use today, 12 yesterday and 8 the day before. What is the probability that a given token is used?

Bernoulli, 1713 (Law of large numbers)

Relative frequency of a given even in a large number of trials should be close to the theoretical probability of that event .

Probability can then be redefined as the limit of the relative frequency if the experiment were repeated infinitely many times.

Limitations

- infinitely many is a lot :-)
- estimation of probability may vary, depending on samples
- sometimes, repetition is not possible

Understanding probability

Information view (from the introduction)

Information

What is the probability that the network went down yesterday?
It is a past event: there's no randomness after it happened. Still, we can model our uncertainty by using probabilities.

Examples:

- probability of finding survivors after earthquake
- probability that a message was transmitted with no errors
- probability that a patient has the flu

Probability depends on the available information on the event.

Understanding probability

Subjective view

So, probability is not that objective after all.

Bayes theorem, 1763

Suppose $A_1, \dots A_n$ are *incompatible events* such that the collection represents all possibilities : $\bigcup_{j=1}^{n} A_i = \Omega$. Then

$$\mathbb{P}[A_i|B] = \frac{\mathbb{P}[B|A_i]\,\mathbb{P}[A_i]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A_i]\,\mathbb{P}[A_i]}{\sum_{j=1}^{n}\mathbb{P}[B|A_j]\,\mathbb{P}[A_j]}$$

If we "know" the prior (first estimation) $\mathbb{P}[A_j]$ and the likelihood $\mathbb{P}[B|A_j]$ then the observed data B can change the perceived probabilities of the events A_j : $\mathbb{P}[B|A_i]$ (posterior probabilities)

- Repetition of the same is not necessary
- Arbitrary prior distribution

Understanding probability

Axiomatic view

XXth century : Need for a formal theory of probability.

- using Borel's work on measure theory
- and Lebesgue integral,
- Kolmogorov (1933) defined the axiomatic theory.

Events are viewed as subsets of possible outcomes.

Outline

Espace de probabilité

- Ω est l'ensemble des réalisations possibles $\{\omega_i\}$ (résultats d'une expérience aléatoire)
- Ω peut être fini ou non, dénombrable ou non.
- un événement est une partie A de Ω
- $\mathcal{F}$ est l'ensemble des événements:
  - $\emptyset \in \mathcal{F}$ et $\Omega \in \mathcal{F}$ *("aucun événement" et "tous les événements" sont dans $\mathcal{F}$)*
  - si $A \in \mathcal{F}$ alors $\bar{A} \in \mathcal{F}$ *(si un événement est dans $\mathcal{F}$, son contraire aussi)*
  - l'union d'événements est un événement:

$$\forall A_1, A_2, \ldots \in \mathcal{F}, \cup_{n=1}^{\infty} A_n \in \mathcal{F}$$

- $\mathbb{P}[.]$ est une mesure de probabilité sur $(\Omega, \mathcal{F})$:
  - $\mathbb{P}[\Omega] = 1$
  - Pour tout ensemble dénombrable d'événements incompatibles $A_1, A_2, \ldots$ on a:

$$\mathbb{P}[\cup_i A_i] = \sum_i \mathbb{P}[A_i]$$

Example of a probability space

Random experiment : throwing a die.

Give the corresponding probability space.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- $\mathcal{F}$ is the power set of Ω (set of all subsets)

For instance, event $E =$"the result is odd" is in $\mathcal{F}$:$\{1, 3, 5\}$.
Another event $E' =$"the result is smaller or equal to 3" is also in $\mathcal{F}$. As is $E \cup E'$, or $E \cap E'$.

- $\forall k \in \Omega, \mathbb{P}[k] = \frac{1}{6}$.

Therefore $\mathbb{P}[E \cup E'] =?$

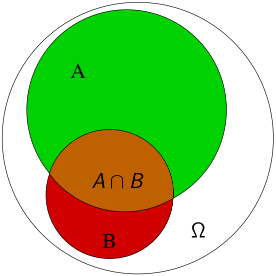
Probability space for humans (not mathematicians)

- Ω is the universe of possible outcomes (if we really did the experiment)
- $\mathcal{F}$ is the set of events, that is, a group (set) of possible outcomes.
- The probability measure $\mathbb{P}$ associates a positive number $\mathbb{P}[A]$ to any possible *event* $A \in \mathcal{F}$, such that:
  - $\mathbb{P}[\Omega] = 1$ (probability of the universe)
  - if two events A and E are incompatible then $\mathbb{P}[A \cup E] = \mathbb{P}[A] + \mathbb{P}[E]$ (union= OR)

In general $\mathbb{P}[A \cup E] \neq \mathbb{P}[A] + \mathbb{P}[E]$ since it would count twice the case $\{A \text{ and } E\}$ (noted $A \cap E$.)

Propriétés

- $\mathbb{P}[\emptyset] = 0$
- $\mathbb{P}\left[\overline{A}\right] = 1 - \mathbb{P}[A]$
- $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$
- $P(A) \leq P(B)$ si $A \subset B$



Incompatibilité

Deux événements A et B sont incompatibles ssi $\mathbb{P}[A \cap B] = \mathbb{P}[\emptyset] = 0$

Remarque

$P(A) = 0 \not\Rightarrow A = \emptyset$.
De même, $P(A) = 1 \not\Rightarrow A = \Omega$.

Application to software testing [Mitzenmacher and Upfal]

Polynomial computation testing

Given two polynoms $F(x) = \sum_{i=1}^{n} a_i x^i$ and $G(x) = \prod_{j=1}^{n} (b_j x + c_j)$. Can you check that F=G?

- compute G and F in canonical form and compare the coefficients
- Randomized algorithm : pick a number r uniformly in $\{0, 1, \ldots, 100n\}$ [1] and compare $F(r)$ and $G(r)$.

The algorithm can only give a wrong answer if $F \neq G$ and if $F(r) = G(r) \Leftrightarrow r$ is a root of $F - G$. Let's call E this event.
There are at most n such roots (n is the max possible degree of $F - G$)

$$\mathbb{P}[E] \leq \frac{n}{100 \times n} = 10^{-2}$$

[1] where n is the degree of the polynoms

Outline

Independance

Let us improve the above method to get better confidence in the algorithm output.

Recall that the algorithm is always right when it outputs that $F \neq G$. If the output is $F = G$, we could repeat the experiment to increase the certainty.

Let us pick a second number r_2 and check whether $F(r_2) = G(r_2)$. Let us call E_2 the event $F(r_2) = G(r_2)$ when $F \neq G$. Then as above, $\mathbb{P}[E_2] \leq 10^{-2}$.

If r_2 is chosen *independently*[2] of r, then

$$\mathbb{P}[E \cap E_2] \leq (10^{-2})^2 = 10^{-4}$$

[2] With replacement, so we need to treat properly the case $r = r_2$

Outline

Probabilités conditionnelles

La probabilité qu'un événement A se produise sachant qu'un événement B s'est produit est notée $\mathbb{P}[A|B]$.

Formule de Bayes

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

En généralisant (par récurrence) on a:

$$P\left(\bigcap_{k=0}^{n} A_k\right) = P(A_1|A_0)P(A_2|......A_1)\ldots P(A_n|A_0 \cap A_1 \ldots \cap A_{n-1})P(A_0)$$

Causality and conditioning

$\mathbb{P}[A|B] \neq \mathbb{P}[A]$ does not mean that B has a causal effect on A.

Example

- $A = \{$patient is infected by a virus$\}$
- $B = \{$disease test result is positive$\}$

Then $\mathbb{P}[B|A]$ indicates a causal explanation (test is positive partly due to the fact that the patient is infected.
But $\mathbb{P}[A|B]$ does not (the test has no effect on the presence of the virus in the patient).

Loi des probabilités totales

Soit $B_1, \ldots, B_n$ une partition de Ω ($B_i \cap B_j = \emptyset$et $\cup_{1 \leq i \leq n} B_i = \Omega$) telle que $\forall 1 \leq i \leq n, \mathbb{P}[B_i] > 0$. Alors pour tout événement A:

$$\mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A|B_i]\, \mathbb{P}[B_i]$$

Indépendance

Deux événements A et B sont indépendants ssi $P(A \cap B) = P(A) \times P(B)$.

Remarque : si A et B sont indépendants alors $\mathbb{P}[A|B] = \mathbb{P}[A]$.

Application of Bayes formula

Rare illness paradox

https://sciencetonnante.wordpress.com/2012/10/08/les-probabilites-conditionnelles-bayes-level-1/

Yule-Simpson paradox

https://sciencetonnante.wordpress.com/2013/04/29/le-paradoxe-de-simpson/

Exercice

Outline

Variables aléatoires

Une variable aléatoire X est une application $X : \Omega \to F$ ou F est un ensemble ordonné, t.q.

$$\forall x \in F, \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$$

Human langage translation

A random variable is a function $X : \Omega \to \mathbb{R}$ (or $\mathbb{N}$) characterizing the possible outcomes in Ω such that intervals in $\mathbb{R}$ have a probability.

- X est une v.a. discrète si F est dénombrable (typ. $\mathbb{N}$)
- X est une v.a. continue si F est indénombrable (typ. $\mathbb{R}$)

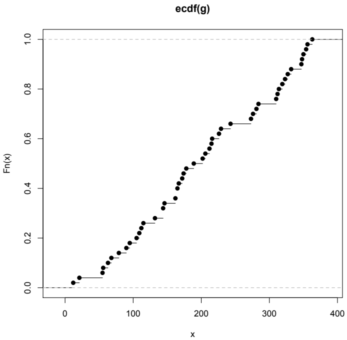
Cumulative Distribution function (C.D.F)

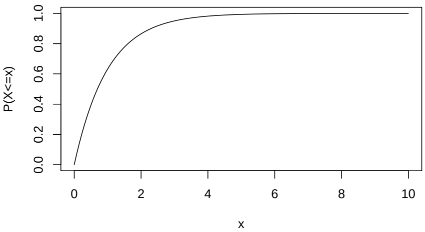
Pour toute v.a. X on définit sa fonction de répartition (CDF en anglais):

$$\forall x, \quad F(x) = \mathbb{P}[X \leq x]$$

F donne les probabilités *cumulées*.

F décrit la loi ou distribution de probabilité de X.

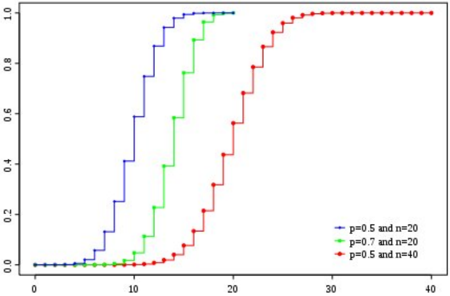




- $0 \leq F(X) \leq 1$
- F is nondecreasing

Cas particulier d'une variable aléatoire discrète

La fonction de répartition (C.D.F) d'une variable aléatoire discrète est une fonction en escalier (sommes cumulées).



Probability density function (P.D.F)

La distribution d'une variable aléatoire peut également être définie de façon plus "élémentaire" par :

Cas d'une variable discrète

Fonction de distribution :

$$f(k) = \mathbb{P}[X = k]$$

pour toute valeur k possible.

Cas d'une variable continue (réelle)

Densité de probabilité (PDF en anglais) $f(x)$

$$\mathbb{P}[x \leq X \leq y] = \int_x^y f(u)du$$

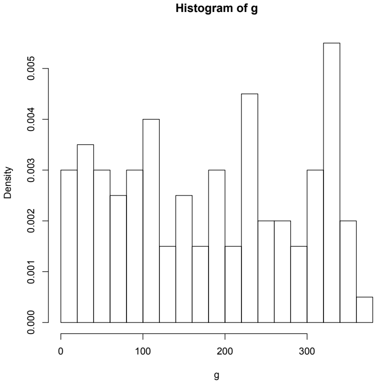
attention aux bornes d'intégration...

Variables aléatoires discrètes

Pour une v.a. discrète (à valeurs entières par exemple) on définit la fonction de distribution (ou loi) par

$$f(k) = \mathbb{P}[X = k]$$

. C'est la fonction qui donne la probabilité de chaque valeur.



- $0 \leq f(k) \leq 1$
- $F(k) = \sum_{m=-\infty}^{k} f(m)$
- f peut être approchée par l'histogramme d'un échantillon

Outline

Bernoulli trials

Bernoulli

A binary random variable $X \in \{0, 1\}$ such that $\mathbb{P}[X = 1] = p$ is a Bernoulli random variable $\mathcal{B}(p)$.

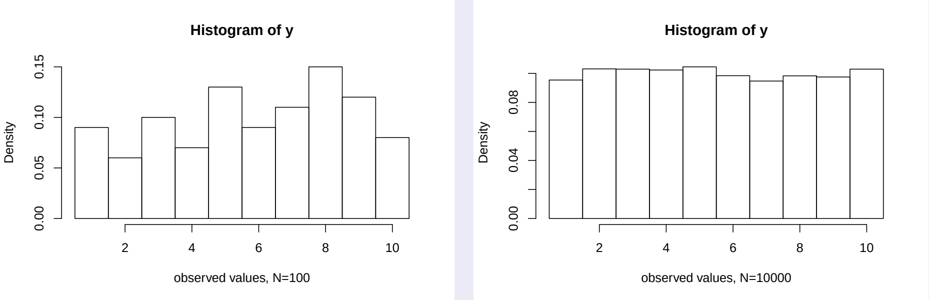
Typical example : coin tossing, test result (pass/fail)...

Uniform distribution

Une variable $X \in \{n_1, n_2, n_3, \ldots, n_k\}$ pouvant prendre k valeurs est uniforme si chacune de ces valeurs a la même probabilité de se réaliser :

$$\mathbb{P}[X = n_i] = \frac{1}{k}$$

```
 y=sample.int(10,size=N,replace=TRUE)
hist(y,breaks=br,freq=FALSE)
```



Binomial distribution

Repeating bernoulli trials...

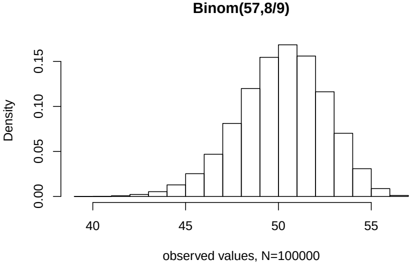
Binomial

La somme de n v.a. de Bernoulli indépendantes de même paramètre p est une v.a. binomiale $Binom(n, p)$. Sa distribution est:

$$\mathbb{P}[X_1 + ... + X_n = k] = C_n^k p^k (1-p)^{n-k}$$

Exercice :

1. Preuve de la distribution?
2. Montrer que $\sum_{i=0}^{n} \mathbb{P}[X = k] = 1$
3. Montrer que l'espérance est $\mathbb{E}[X] = np$



Geometric distribution (1)

Sequence of independent Bernoulli trials $\mathcal{B}(p)$ until first success:

$$\mathbb{P}[X = k] = p \times (1 - p)^{k-1}$$

Warning

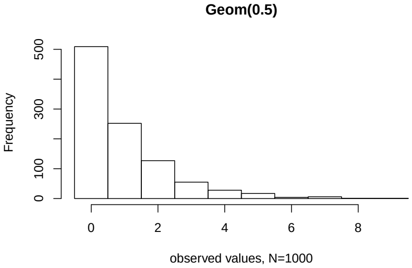
The geometric distribution in the R language (`rgeom`) is slightly different (number of failures before first success)

Exercice:

- Prove that $\sum_{i=0}^{\infty} \mathbb{P}[X = k] = 1$
- Prove that $\mathbb{E}[X] = \frac{1}{p}$

Geometric distribution (2)

Note that the geometric r.v. has infinite support, i.e. can take an infinite number of values.




```
y=rgeom(1000,prob=0.5)
br=seq(from=min(y)-0.5,to=max(y)+0.5,by=1)
hist(y,xlab="observed values, N=1000",main='Geom(0.5)',breaks=br)
```

Poisson distribution

Limite de la binomiale pour de grandes valeurs de n (et petites valeurs de p)

Paramètre: λ

$$\mathbb{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda x}$$

- Prove that $\mathbb{E}[X] = \lambda$
- Link with binomial ? (hint: use $\lambda = np$)

`rpois(n,lambda)`

Other classical distributions that you should know about

- Hypergeometric
- Negative binomial distribution
- Multinomial distribution

Outline

Expectation

Espérance

La moyenne (mean) ou espérance (expectation en anglais) d'une v.a. discrète X à valeurs dans I (dénombrable) est:

$$\mathbb{E}[X] = \sum_{x \in I} x \mathbb{P}[X = x]$$

Example

Soit $0 < p < 1$. Une variable aléatoire X à valeurs dans $\{0, 1\}$ suit une loi de Bernoulli $\mathcal{B}(p)$ ssi $\mathbb{P}[X = 1] = p$ et $\mathbb{P}[x = 0] = 1 - p$.

Son espérance est $\mathbb{E}[X] = 0 \times (1 - p) + 1 \times p = p$.

Exercise

What is the expected value of the sum of two fair dice?

Expectation

Linearity

Given two random variables X and Y we have:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

even if X and Y are dependent variables.

Also for any constant c we have:

$$\mathbb{E}[cX] = c\mathbb{E}[X]$$

Expectation

- Compute the expectation of a binomial random variable Z.
- Let $Y = X - \mathbb{E}[X]$. Compute $\mathbb{E}[Y]$ as a function of the moments of X.

Conditional expectation

$$\mathbb{E}\left[Y|Z=z\right]=\sum_{y} y\mathbb{P}[Y=y|Z=z]$$

Law of total probability:

$$\mathbb{E}\left[Y\right]=\mathbb{E}\left[\mathbb{E}\left[Y|Z\right]\right]$$

where $\mathbb{E}\left[Y|Z\right]$ is a random variable ($f(Z)$).