02b - Random variables - 02

Master in Foundations of Data Science

Bayesian Statistics and Probabilistic Programming

Fall 2018-2019

Josep Fortiana

Monday, October 01, 2018

Universitat de Barcelona

02b - Random variables -02

Expectation, variance, higher moments

Law of Large Numbers (LLN) and Central Limit Theorem (CLT)

Bivariate discrete r.v. 's

Bayes formula for discrete r.v. 's

Bivariate continuous r.v. 's

Bayes' formula for continuous r.v. 's

02b - Random variables -02

Expectation of a discrete r.v.

For a discrete r.v. X, with values $\boldsymbol{x} = (x_1, \ldots, x_m)$ and probabilities $\boldsymbol{d} = (d_1, \ldots, d_m)$,

The *expectation* or *mean* of X is the average of the $\boldsymbol{x}$ values, weighted by their respective probabilities:

$$\mathrm{E}(X) = \sum_{j=1}^{m} x_j \, d_j.$$

Ordinary average of a list of numbers

For n numbers $\boldsymbol{x} = (x_1, \ldots, x_n)$, the arithmetic mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i \,\left(\frac{1}{n}\right),$$

is the expectation of the r.v. with values $\boldsymbol{x}$ and uniform probability (a probability mass $\dfrac{1}{n}$ on each of the n points x_i).

Weighted average of a list of numbers

For a list of numbers, grouped on a table with m different values, $\boldsymbol{x} = (x_1, \ldots, x_m)$, and vector of relative frequencies $\boldsymbol{f} = (f_1, \ldots, f_m)$, the weighted average:

$$\bar{x} = \sum_{j=1}^{m} x_j\, f_j,$$

is the expectation of the r.v. with values $\boldsymbol{x}$ and probabilities $\boldsymbol{d} = \boldsymbol{f}$.

Properties of the expectation

$\mathrm{E}(\,\cdot\,)$ is a linear operator:

$$\boxed{\begin{aligned} \mathrm{E}(X+Y) &= \mathrm{E}(X)+\mathrm{E}(Y), \\ \mathrm{E}(c\,X) &= c\,\mathrm{E}(X), \end{aligned}}$$

for $c \in \mathbb{R}$.

Expectation of a product of r.v.

If X and Y are independent r.v. , then:

$$\mathrm{E}(X\,Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y).$$

Non-independent r.v.'s do not satisfy this equality.

Example: expectation of a constant r.v.

Si c is a constant r.v. , with value $c \in \mathbb{R}$, then

$$\mathrm{E}(c) = c.$$

Expectation of an infinite discrete r.v.

When the set $\boldsymbol{x}$ is not finite, we substitute the sum of a series for the finite sum.

When the series $\sum x_n\, d_n$ is not summable the r.v. has no expectation.

Expectation of a function of a discrete r.v.

If X is a discrete r.v. , with values $\{x_j\}$ and probabilities $\{d_j\}$, and g is a function, *the expectation of* $g(X)$ is:

$$\mathrm{E}(g(X)) = \sum_j g(x_j)\, d_j,$$

(if this sum exists in the infinite case).

Variance of a discrete r.v.

If X is a discrete r.v. , with values $\{x_j\}$, probabilities $\{d_j\}$, and $m_x = \mathrm{E}(X)$, *the variance of* X is defined by:

$$\mathrm{var}(X) = \mathrm{E}((X - E(X))^2) = \sum_j (x_j - m_x)^2\, d_j,$$

(if this sum exists in the infinite case).

Relation with ${s_x}^2$ of a numbers list

For n numbers $\boldsymbol{x} = (x_1, \ldots, x_n)$,

$${s_x}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \left(\frac{1}{n}\right),$$

is the variance of the r.v. with values $\boldsymbol{x}$ and uniform probability.

Expectation of the square and squared expectation

$\mathrm{E}(X^2)$ and $(\mathrm{E}(X))^2$ do not coincide.

Always:

$$\mathrm{E}(X^2) \geq (\mathrm{E}(X))^2.$$

The equality holds if, and only if X is constant.

Expectation of the square and squared expectation

When both quantities exist:

$$\boxed{\operatorname{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2.}$$

Proof

$$\begin{aligned}
\mathrm{var}(X) = \mathrm{E}((X - m_x)^2) &= \mathrm{E}(X^2 - 2\, m_x\, X + m_x^2) \\
&= \mathrm{E}(X^2) - 2\, m_x\ \mathrm{E}(X) + m_x^2 \\
&= \mathrm{E}(X^2) - 2\, m_x^2 + m_x^2 \\
&= \mathrm{E}(X^2) - m_x^2.
\end{aligned}$$

Properties of the variance

$\mathrm{var}(\,\cdot\,)$ is a quadratic operator:

$$\mathrm{var}(c\,X) = c^2\,\mathrm{var}(X), \qquad \text{for } c \in \mathbb{R}.$$

When X and Y are independent r.v. , then

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y).$$

Expectation and variance of a Bernoulli distribution

If $X \sim \mathrm{Ber}(p)$,

$$\mathrm{E}(X) = p,$$

$$\mathrm{var}(X) = p\,(1 - p).$$

Computing the expectation of a Bernoulli distribution

If $X \sim \mathrm{Ber}(p)$, the values vector is $\boldsymbol{x} = (0, 1)$ and the probabilities vector is $\boldsymbol{d} = (1 - p, p)$

$$\mathrm{E}(X) = 0 \times (1 - p) + 1 \times p = p.$$

Computing the variance of a Bernoulli distribution

$$\mathrm{E}(X^2) = 0 \times (1-p) + 1 \times p = p,$$

$$\mathrm{var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = p - p^2 = p\,(1-p).$$

Expectation and variance of a binomial distribution

If $X \sim \mathrm{B}(n, p)$,

$$\mathrm{E}(X) = n\,p,$$

$$\mathrm{var}(X) = n\,p\,(1 - p).$$

Expectation and variance of a binomial distribution

$X \sim \mathrm{B}(n, p)$, is the number of occurences of an event A in n independent repetitions of a binary experiment.

X is the sum of the n independent indicators of A occurrence.

Hence, $E(X)$ and $\mathsf{var}(X)$ are those of a Bernoulli distribution times n.

Expectation of a continuous r.v.

If X is a continuous r.v. with pdf f, we define:

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx,$$

if this integral exists.

As in the case of infinite discrete r.v.'s, there are continuous r.v. with no expectation.

Expectation of a function of a continuous r.v.

If X is a continuous r.v. with pdf f, and g is a function,

$$\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx,$$

if this integral exists.

Variance of a continuous r.v.

If X is a continuous r.v. with pdf f,

$$\operatorname{var}(X) = \int_{-\infty}^{\infty} (x-m)^2\, f(x)\, dx,$$

where $m = \operatorname{E}(X)$, if this integral exists.

Expectation and variance of a $\mathsf{Unif}(0,1)$ distribution

The pdf of $X \sim \mathsf{Unif}(0,1)$:

$$f(x) = \begin{cases} 1, & \text{if } x \in (0,1), \\ 0, & \text{if } x \notin (0,1). \end{cases}$$

Expectation and variance of a $\mathsf{Unif}(0,1)$ distribution

For $k \neq -1$,

$$\mathsf{E}(X^k) = \int_{-\infty}^{\infty} x^k\, f(x)\, dx = \int_0^1 x^k\, dx$$

$$= \frac{1}{k+1}.$$

In particular, $\mathsf{E}(X) = \dfrac{1}{2}$, $\mathsf{var}(X) = \dfrac{1}{12}$.

Expectation and variance of a normal distribution

If $X \sim \mathrm{N}(\mu, \sigma^2)$,

- μ is the *expectation* of X, $\mathrm{E}(X) = \mu$.
- σ^2 is the *variance* of X, $\mathrm{var}(X) = \sigma^2$.

02b - Random variables -02

Law of Large Numbers (LLN)

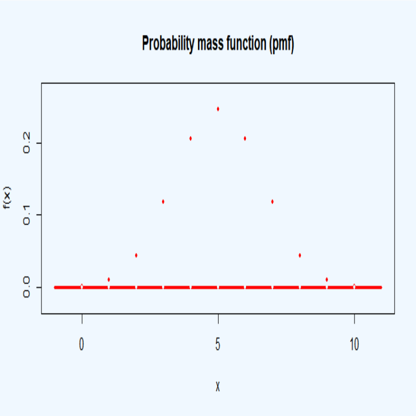
Dividing $X_n \sim \mathrm{B}(n, p)$ by n, we obtain the *relative frequency* of an event in n independent repetitions:

$$f_n = \frac{X_n}{n}.$$

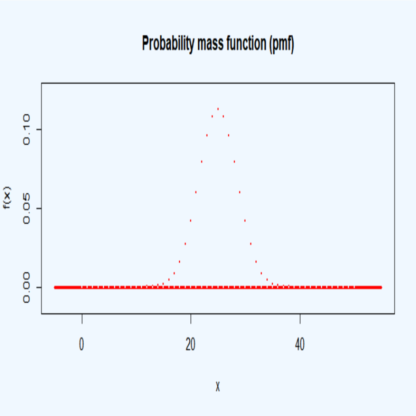
Its values are: $\{k/n : 0 \leq k \leq n\}$.

Values closer to p have higher probability.

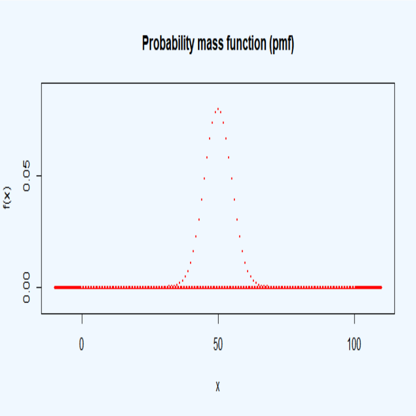
Pmf of f_n for $p = 0.5$ and $n = 10$



Pmf of f_n for $p = 0.5$ and $n = 50$



Pmf of f_n for $p = 0.5$ and $n = 100$



Law of Large Numbers

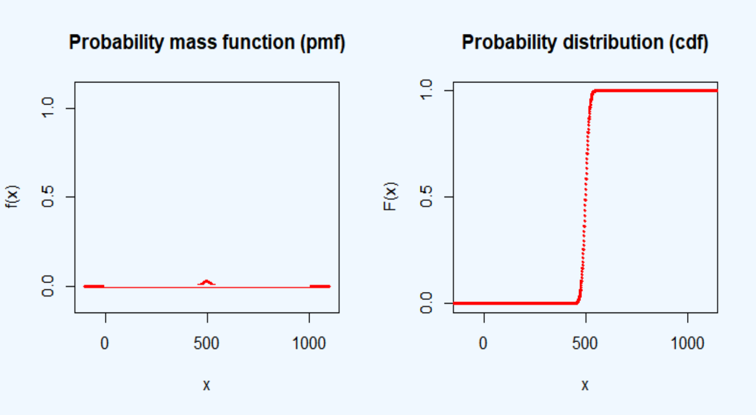
When $n \to \infty$ values with a significant probability are few and close to p.

E.g., when $p = 0.5$ and $n = 1000$ only 81 values of f_n have probability > 0.001.

The distribution tends to a constant r.v. with value p.

$$\{f_n\} \xrightarrow{n \to \infty} p.$$

pmf and cdf of f_n for $p = 0.5$ and $n = 1000$



Sum of independent normal r.v.

Normal r.v. are closed under the operation "Sum of independent r.v.":

If $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ and they are independent:

$$X_1 + X_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Hence the sum of n r.v. , i.i.d.$\sim \mathrm{N}(\mu, \sigma^2)$ is a normal r.v. , with mean $n\,\mu$ and variance $n\,\sigma^2$.

Central Limit Theorem (CLT)

What happens when non-normal r.v. are added?

The sum of n independent r.v. is approximately normal, assuming some regularity conditions.

The approximation improves as $n \to \infty$.

In particular, if $X_n \sim \mathrm{B}(n, p)$,

$$X_n \approx \mathrm{N}(np, np(1-p)),$$

the larger is n the better is the approximation.

02b - Random variables -02

Joint pmf for two discrete variables

The joint pmf of a pair of discrete r.v. (X, Y) is:

$$h(x, y) \equiv \mathrm{P}(\{X = x, Y = y\}), \quad (x, y) \in \mathbb{R}^2.$$

	$X = 1$	$X = 2$	$X = 3$	
$Y = 0$	0.90	0.03	0.02	0.95
$Y = 1$	0.01	0.02	0.02	0.05
	0.91	0.05	0.04	1

As a matrix, each (i, j)-th entry is $\mathrm{P}(X = j, Y = i)$.

Joint cdf for two r.v.

The *Joint cumulative probability distribution function (cdf)* of a pair of r.v. 's (X, Y) is:

$$\begin{aligned} H: \ \mathbb{R}^2 &\to [0,1] \\ (x,y) &\mapsto H(x,y) \equiv \mathrm{P}\left(\{X \le x, Y \le y\}\right). \end{aligned}$$

This definition is general, for any type of r.v., either discrete, or continuous, or anything else.

Probabilities from the joint cdf for two r.v.

From H, the probability of a rectangle $R \subset \mathbb{R}^2$, determined by the lower left vertex (x_1, y_1) and the upper right vertex (x_2, y_2) is:

$$\begin{aligned} \mathsf{P}(R) = H((x_2, y_2)) - H((x_1, y_2)) \\ - H((x_2, y_1)) + H((x_1, y_1)). \end{aligned}$$

Joint cdf for the example pmf

Given $(x, y) \in \mathbb{R}^2$,

$$H(x,y) = \begin{cases} 0, & \text{if} \quad x < 1 \quad \text{or} \quad y < 0, \\ 0.9, & \text{if} \quad 1 \le x < 2, \ 0 \le y < 1, \\ 0.93, & \text{if} \quad 2 \le x < 3, \ 0 \le y < 1, \\ 0.95, & \text{if} \quad 3 \le x, \qquad 0 \le y < 1, \\ 0.91, & \text{if} \quad 1 \le x < 2, \ 1 \le y, \\ 0.96, & \text{if} \quad 2 \le x < 3, \ 1 \le y, \\ 1, & \text{if} \quad 3 \le x, \qquad 1 \le y. \end{cases}$$

Marginal univariate pmf's

f, the *marginal pmf* of X, is obtained by adding the columns, giving:

$\mathrm{P}(X = 1) = 0.91, \ \mathrm{P}(X = 2) = 0.05, \ \mathrm{P}(X = 3) = 0.04.$

g, the *marginal pmf* of Y, is obtained by adding the rows, giving:

$$\mathrm{P}(Y = 0) = 0.95, \quad \mathrm{P}(Y = 1) = 0.05.$$

Marginal univariate cdf's

For X and Y, respectively

$$
\begin{aligned}
F(x) &= \lim_{y \to \infty} H(x, y), \\
G(y) &= \lim_{x \to \infty} H(x, y).
\end{aligned}
$$

This limit is equivalent to "adding", for each value of one of the variables, the probabilities of all possible values of the other.

Independent random variables

X, Y are (stochastically) independent if, and only if,

$$H(x, y) = H_0(x, y) \equiv F(x) \cdot G(y), \quad (x, y) \in \mathbb{R}^2.$$

In terms of the pmf's:

$$h(x, y) = h_0(x, y) \equiv f(x) \cdot g(y), \quad (x, y) \in \mathbb{R}^2.$$

Joint pmf for two independent r.v.

h_0, the independence joint pmf with the same marginals as the table in the example:

	$X = 1$	$X = 2$	$X = 3$	
$Y = 0$	0.8645	0.0475	0.038	0.95
$Y = 1$	0.0455	0.0025	0.002	0.05
	0.91	0.05	0.04	1

Conditional table, given Y

The *table of probabilities conditional to* Y is obtained by dividing each row in the original table by its total (marginal probability of that value of Y).

	$X=1$	$X=2$	$X=3$	
$Y=0$	0.9474	0.0316	0.0211	1
$Y=1$	0.2000	0.4000	0.4000	1

Now the (i,j)-th entry is $\mathrm{P}(X=j|Y=i)$. This table is also called *the matrix of row profiles.*

Conditional table, given X

The *table of probabilities conditional to* X, or *matrix of column profiles*, is obtained by dividing each column by its total (marginal probability of that value of X).

	$X=1$	$X=2$	$X=3$
$Y=0$	0.9890	0.6000	0.5000
$Y=1$	0.0110	0.4000	0.5000
	1	1	1

Now the (i,j)-th entry is $\mathrm{P}(Y=i|X=j)$.

Covariance of two discrete r.v. (X, Y)

Given h, joint pmf, we get $\mathrm{E}(X)$, $\mathrm{E}(Y)$, $\mathrm{var}(X)$, $\mathrm{var}(Y)$, from the marginal pmf's f and g.

The *covariance* of (X, Y) is:

$$\mathrm{cov}(X, Y) \stackrel{\text{def}}{=\!=} \sum_x \sum_y h(x, y)\,(x - \mathrm{E}(X))\,(y - \mathrm{E}(Y)),$$

where x and y take all values of X and Y, respectively.

Covariance of two discrete r.v. (X, Y)

Alternatively, compute the expectation of the product:

$$\mathrm{E}(X \cdot Y) \stackrel{\text{def}}{=\!=} \sum_x \sum_y h(x, y)\, x\, y,$$

and, then, use the equality:

$$\operatorname{cov}(X, Y) = \mathrm{E}(X \cdot Y) - \mathrm{E}(X) \cdot \mathrm{E}(Y).$$

Covariance of two discrete r.v. (X, Y)

For the example:

$$\mathrm{E}(X) = 1.13, \qquad \mathrm{E}(Y) = 0.05,$$

$$\operatorname{var}(X) = 0.1931, \qquad \operatorname{var}(Y) = 0.0475,$$

and

$$\operatorname{cov}(X, Y) = 0.0535.$$

Correlation coefficient of two discrete r.v. (X, Y)

$$\operatorname{cor}(X, Y) \stackrel{\text{def}}{=\!=} \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var}(X) \cdot \operatorname{var}(Y)}}.$$

Normalization of $\operatorname{cov}(X, Y)$, $-1 \leq \operatorname{cor}(X, Y) \leq 1$.

$\operatorname{cor}(X, Y)$ measures linear dependence between X, Y.

For the example, $\operatorname{cor}(X, Y) = 0.5586$.

02b - Random variables -02

Bayes formula for two discrete variables

Each entry in the conditional table, given X, is the result of applying Bayes' formula,

$$\mathrm{P}(Y=i|X=j)=\frac{\mathrm{P}(X=j|Y=i)\,\mathrm{P}(Y=i)}{\mathrm{P}(X=j)},$$

operating on entries in the Y marginal in the original table, with the conditional table given Y. Note the denominator:

$$\mathrm{P}(X=j)=\sum_i \mathrm{P}(X=j|Y=i)\,\mathrm{P}(Y=i)=\sum_i \mathrm{P}(X=j,Y=i).$$

Interpreting Bayes formula

Each j-th column in the the conditional table, given X, is the transform of the *a priori* Y probabilities column vector, that is, the marginal pmf of Y, resulting from entering the knowledge that $\{X = j\}$ has been observed.

02b - Random variables -02

Joint probability density function

Let (X, Y) be a pair of r.v. with an absolutely continuous joint probability distribution. It has a joint probability density (pdf):

$$h(x, y), \quad \text{for } (x, y) \in \mathbb{R}^2.$$

To compute the probability of some region $A \subset \mathbb{R}^2$,

$$\mathrm{P}(A) = \iint_A h(x, y)\, dx\, dy.$$

From the joint cdf to the joint pdf and back

Given the joint cdf, the joint pdf is given by:

$$h(x, y) = \frac{\partial^2 H(x, y)}{\partial x \, \partial y}.$$

Given the joint pdf, the joint cdf is given by:

$$H(x, y) = \int_{-\infty}^{x} ds \int_{-\infty}^{y} dt \, h(s, t).$$

Marginal pdf's

The marginal pdf of X is:

$$f(x) = \int_{\mathbb{R}} h(x,y)\, dy, \quad x \in \mathbb{R},$$

and the marginal pdf of Y is:

$$g(y) = \int_{\mathbb{R}} h(x,y)\, dx, \quad y \in \mathbb{R}.$$

Marginal cdf's

As in the discrete case

$$
\begin{aligned}
F(x) &= \lim_{y\to\infty} H(x,y), \\
G(y) &= \lim_{x\to\infty} H(x,y).
\end{aligned}
$$

Independent continuous r.v. 's

(X, Y) are *independent:*

$$\iff h(x, y) = f(x) \cdot g(y)$$

$$\iff H(x, y) = F(x) \cdot G(y)$$

Covariance of two continuous r.v. (X, Y)

$$\operatorname{cov}(X, Y) = \iint_{\mathbb{R}^2} (x - \operatorname{E}(X)) \cdot (y - \operatorname{E}(Y)) \cdot h(x, y)\, dx\, dy.$$

Alternatively, as in the discrete case, obtain:

$$\operatorname{E}(X \cdot Y) = \iint_{\mathbb{R}^2} x \cdot y \cdot h(x, y)\, dx\, dy,$$

and

$$\operatorname{cov}(X, Y) = \operatorname{E}(X \cdot Y) - \operatorname{E}(X) \cdot \operatorname{E}(Y).$$

Conditional probability density function

The pdf of X, conditional to $\{Y = y\}$ is:

$$f(x|y) = \frac{h(x, y)}{g(y)}, \quad x \in \mathbb{R}.$$

The pdf of Y, conditional to $\{X = x\}$ is:

$$g(y|x) = \frac{h(x, y)}{f(x)}, \quad y \in \mathbb{R}.$$

02b - Random variables -02

Bayes' formula for pdf's

Combining both expressions we obtain:

$$g(y|x) = \frac{h(x,y)}{f(x)} = \frac{f(x|y)\,g(y)}{f(x)} = \frac{f(x|y)\,g(y)}{\int_{\mathbb{R}} f(x|y)\,g(y)\,dy}$$

Interpretation of Bayes' formula

As in the discrete case, the interpretation of this formula is:

The *final*, or *a posteriori* pdf $g(y|x)$ is the transformation of the *initial*, or *a priori* pdf $g(y)$, as a result of blending in the evidence that the observed value of the r.v. X is x.