Continuous random variables: part B

EE-209 - Eléments de Statistiques pour les Data Sciences

Outline

Joint cdfs and densities

We can define the joint cdf for a pair of r.v. (X, Y) by

$$F(x,y) := \mathbb{P}(X \leq x, Y \leq y) := \mathbb{P}\big((X \leq x)\&(Y \leq y)\big).$$

Any function $(x,y) \mapsto f(x,y)$ such that

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\xi, u)\, d\xi\, du$$

is a joint probability density for the pair (X, Y).

If F is piecewise $\mathcal{C}_2$, a joint probability density can be defined as

$$p_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} F(x,y).$$

Conditional density

If $p_{X,Y}(x,y)$ is a joint probability density for the pair of r.v. $(X,Y) \in \mathbb{R}^2$

- We can recover the marginal densities

$$p_X(x) = \int_{\mathbb{R}} p_{X,Y}(x,y)\,dy \quad \text{and} \quad p_Y(y) = \int_{\mathbb{R}} p_{X,Y}(x,y)\,dx.$$

- We can define the conditional density of Y given $X = x$, as follows:

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)} \quad \text{and} \quad p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

- As a consequence, we have Bayes' rule:

$$p_{Y|X}(y|x) = \frac{p_{X|Y}(x|y)\,p_Y(y)}{p_X(x)}.$$

Conditional expectation

If $p_{X|Y}(x|y)$ is the conditional probability density of X given Y, then

$$\mathbb{E}[X|Y=y] = \int x \, p_{X|Y}(x|y) \, dx$$

$$\mathbb{E}[X|Y] = \int x \, p_{X|Y}(x|Y) \, dx$$

Law of Total Expectation

$$\mathbb{E}\big[\,\mathbb{E}[X|Y]\,\big] = \mathbb{E}[X]$$

Pair of independent continuous random variables

Two random variables X and Y with a joint pdf are independent if one of the three equivalent properties hold

1. $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}$,
$$p_{X,Y}(x, y) = p_X(x)\, p_Y(y).$$
2. $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}$ such that $p_Y(y) > 0$, we have
$$p_{X|Y}(x|y) = p_X(x).$$
3. For any functions f and g,
$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\, \mathbb{E}[g(Y)].$$

The proofs are essentially the same as for discrete variables but replacing sums by integrals.

Independent continuous random variables

A collection of random variables $X_1, \ldots, X_n$ are independent if one of the three equivalent properties hold

1. $\forall (x_1, \ldots, x_n) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$,

$$p_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1) \ \ldots \ p_{X_n}(x_n).$$

2. $\forall (x_1, \ldots, x_n) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ such that $p_{X_{-i}}(x_{-i}) > 0$, we have for all i,

$$p_{X_i|X_{-i}}(x_i|x_{-i}) = p_{X_i}(x_i) \qquad \text{where} \quad X_{-i} := (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n).$$

3. For any functions $f_1, \ldots, f_n$,

$$\mathbb{E}[f_1(X_1) \ldots f_n(X_n)] = \prod_{i=1}^{n} \mathbb{E}[f_i(X_i)].$$

Variance, covariance and correlation

For real valued r.v. X and Y,

$$\mathrm{Var}[X] = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big],$$

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

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \qquad \text{with} \qquad \sigma_X^2 = \mathrm{Var}[X],\ \sigma_Y^2 = \mathrm{Var}[Y].$$

Now assuming that X is taking values in $\mathbb{R}^d$,

$$\mathrm{Cov}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top\big].$$

Properties of the Variance

The following properties can be verified immediately:

- $\text{Var}(aX + b) = a^2 \,\text{Var}(X)$.
- $\text{cov}(aX + b, cY + d) = ac \,\text{cov}(X, Y)$.
- $\text{corr}(aX + b, cY + d) = \text{corr}(X, Y)$.
- If X and Y are independent, $\text{cov}(X, Y) = 0$.
- In general $\text{Var}(X + Y) = \text{Var}(X) + 2\,\text{cov}(X, Y) + \text{Var}(Y)$
- If X and Y are independent, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Finally, we also have $|\text{corr}(X, Y)| \leq 1$. The proof is beyond scope.

Joint distribution between a discrete and a continuous variable

It is obvious possible to define the joint distribution between a discrete r.v. C and a continuous r.v. X.

$$p_{X|C}(x|c)\,P_C(c) \qquad \text{or} \qquad P_{C|X}(c|x)\,p_X(x)$$

The joint distribution is neither a pdf nor a pmf as defined in this course but is usually denoted like a pdf $p_{X,C}(x,c)$. (In a more abstract theory of probability the pmf is actually viewed as a “discrete” pdf.)

The **marginal distributions** can be computed using the **law of total probability** takes the forms:

$$p_X(x) = \sum_{c \in \mathcal{C}} p_{X|C}(x|c)\,P_C(c) \qquad \text{and} \qquad P_C(c) = \int P_{C|X}(c|x)\,p_X(x)\,dx.$$

The finite mixture of Gaussians

Finite mixtures of distributions are obtained as the marginal distribution $p_X(x)$ associated with a joint distribution between a discrete variable C and a continuous variables X.

The mixture of Gaussians is the most classical example:

- Let C take values in $\{1, \ldots, 3\}$ with pmf $P_C(k) = \pi_k$
- We can define the conditional density of X given C to be

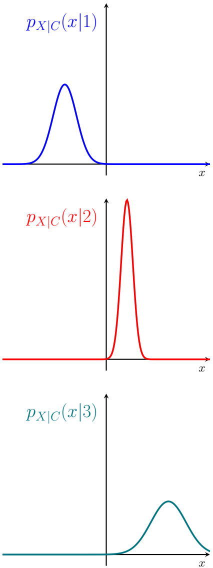
$$p_{X|C}(x|k) = \mathcal{N}(x; \mu_k, \sigma_k^2) := \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x-\mu_k)^2}{2\,\sigma_k^2}}.$$

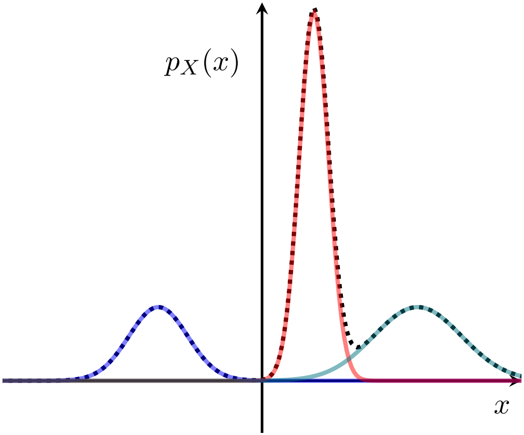
- The marginal is then

$$p_X(x) = \mathcal{N}(x; \mu_1, \sigma_1^2)\,\pi_1 + \mathcal{N}(x; \mu_2, \sigma_2^2)\,\pi_2 + \mathcal{N}(x; \mu_3, \sigma_3^2)\,\pi_3$$

- And by Bayes' rule, we have

$$P_C(k|x) = \frac{p_{X|C}(x|k)P_C(k)}{p_X(x)} = \frac{\mathcal{N}(x; \mu_k, \sigma_k^2)\,\pi_k}{\sum_{j=1}^{3} \mathcal{N}(x; \mu_j, \sigma_j^2)\,\pi_j}.$$





$$p_X(x) = p_{X|C}(x|1)\, P_C(1) + p_{X|C}(x|2)\, P_C(2) + p_{X|C}(x|3)\, P_C(3).$$

$$p_X(x) = \mathcal{N}(x; \mu_1, \sigma_1^2)\, \pi_1 + \mathcal{N}(x; \mu_2, \sigma_2^2)\, \pi_2 + \mathcal{N}(x; \mu_3, \sigma_3^2)\, \pi_3$$

Mixture model and weighted means of densities

We consider the heights distributions of two populations: men and women. Each can be modelled as Gaussian with the following parameters:

	μ	σ
women	161cm	7.1cm
men	175cm	8.5cm

- What is the distribution of height of a "mixed" population which has

$$\pi_1 = 40\% \text{ women and } \pi_0 = 60\% \text{ men?}$$

Let S be the sex variable, with $S = 1$ coding for female and $S = 0$ coding for male. We have

$$\begin{aligned} p(x) &= p(x|S=1)\,p(S=1) + p(x|S=0)\,p(S=0) \\ &= \pi_1\,\mathcal{N}(x;\mu_1,\sigma_1^2) + \pi_0\,\mathcal{N}(x;\mu_0,\sigma_0^2) \\ &= 0.4\,\mathcal{N}(x;161,7.1^2) + 0.6\,\mathcal{N}(x;175,8.5^2) \end{aligned}$$

Sampling from a pair of random variables

Let (X, Y) be a pair of r.v. with joint density $p_{(X,Y)}(x, y)$.

It is usually difficult to sample directly the pair.

However it is possible to

1. Sample $X \sim p_X$ to obtain x
2. Sample $Y \sim p_{(Y|X)(\cdot|x)}$ to obtain y

Note that each step is sampling a scalar random variable.

Pmfs vs pdfs

A number of formulas and results take the same form for pmfs and pdfs by simply replacing sums by integrals.
However it is important to always keep in mind that

- the pmf $P_X(x)$ is the probability of the set $\{x\}$, i.e. $\boxed{P_X(x) = \mathbb{P}(X = x)}$
- while for a pdf $p_X(x)$, we have $\boxed{p_X(x) \neq \mathbb{P}(X = x)}$
- instead $$\boxed{p_X(x) = F'_X(x) = \lim_{h \downarrow 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \downarrow 0} \frac{\mathbb{P}(x \leq X \leq x+h)}{h}}$$

Summary for Joint distribution over r.v.s

- For a pair of r.v.s, $F(x,y) := \mathbb{P}(X \leq x, y \leq y)$ is the joint pdf.
- The joint pdf is $p_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F(x,y)$.
- The marginal density is $p_X(x) = \displaystyle\int_{-\infty}^{\infty} p_{X,Y}(x,y)\,dy$.
- The conditional density is $p_{Y|X}(y|x) = \dfrac{p_{X,Y}(x,y)}{p_X(x)}$.
- Bayes's rule relates both conditionals and marginals: $p_{Y|X}(y|x) = \dfrac{p_{X|Y}(x|y)p_Y(y)}{p_X(x)}$.
- We saw 3 equivalent properties of independence, which generalize to n independent variables.
- We saw several properties of the variance and covariance

⚠ For pdfs, $p_X(x) \neq \mathbb{P}(X = x)$.

Outline

Pdf of the sum of two independent continuous random variables

The pdf of the sum of two independent continuous random variable is the convolution of the pdfs

- Let X and Y two independent r.v.s with pdfs p_X and p_Y
- Let $Z = X + Y$

Then Z has a probability density p_Z given by:

$$p_Z(z) = (p_X * p_Y)(z) := \int_{-\infty}^{+\infty} p_X(z-y)\, p_Y(y)\, dy = \int_{-\infty}^{+\infty} p_X(x)\, p_Y(z-x)\, dx.$$

We say that

- $p_X * p_Y$ is the convolution of p_X and p_Y
- p_X is convolved with p_Y.

proof

Application: pdf of the sum of two independent $\mathcal{U}[0,1]$

If U and V are independent uniform distributions on $[0,1]$, what is the distribution of $Y = U + V$?

We have $p_U(u) = p_V(u) = 1_{\{0 \le t \le 1\}}$, and by the previous theorem

$$\begin{aligned} f(y) &= \int_{-\infty}^{\infty} p_V(y-u)\, p_U(u)\, du \\ &= \int_{-\infty}^{\infty} 1_{\{0 \le y-u \le 1\}}\, 1_{\{0 \le u \le 1\}}\, du \\ &= 1_{\{0 \le y \le 1\}} \int_0^y du + 1_{\{1 \le y \le 2\}} \int_{y-1}^1 du \\ &= y\, 1_{\{0 \le y \le 1\}} + (2-y)\, 1_{\{1 \le y \le 2\}}. \end{aligned}$$

Pdf of a scaled version of a random variable $Y = aX$

If X is a continuous r.v. with pdf p_X what is the pdf of $Y = aX$ for $a > 0$?

We can again use the cdf

$$F_Y(y) := \mathbb{P}(Y \leq y) = \mathbb{P}(aX \leq y) = \mathbb{P}\left(X \leq \frac{y}{a}\right) = F_X\left(\frac{y}{a}\right)$$

$$p_Y(y) = F'_Y(y) = \frac{\partial F_X\left(\frac{y}{a}\right)}{\partial y} = \frac{\partial\left(\frac{y}{a}\right)}{\partial y} F'_X\left(\frac{y}{a}\right) = \frac{1}{a} p_X\left(\frac{y}{a}\right)$$

Pdf of $Y = aX$ when $a > 0$

$$p_Y(y) = \tfrac{1}{a} p_X\left(\tfrac{y}{a}\right)$$

Application:

Let X be an exponential r.v. with pdf $f_X(x) = e^{-x}$, what is the density of $Y = \frac{1}{\lambda}X$?

By the previous result, $f_Y(y) = \lambda e^{-\lambda y}$. This is the generic form of an exponential r.v.. We see that the different exponential r.v. are simply scaled versions of one another.

The square of a $\mathcal{N}(0,1)$ is a $\chi^2(1)$

Proposition

If $X \sim \mathcal{N}(0,1)$ and $Y = X^2$ then $Y \sim \chi^2(1) \equiv \Gamma(\frac{1}{2}, \frac{1}{2})$, and in particular

$$p_Y(y) = \frac{1}{\sqrt{2\pi y}}\, e^{-\frac{1}{2}y}$$

Proof that $X \sim \mathcal{N}(0,1) \;\Rightarrow\; X^2 \sim \chi^2(1)$

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(X^2 \le y) = \mathbb{P}(-\sqrt{y} \le X \le \sqrt{y}) = 2\,\mathbb{P}(0 \le X \le \sqrt{y}) \\ &= 2\big[F_X(\sqrt{y}) - F_X(0)\big] \end{aligned}$$

The function F_Y is differentiable for any $y > 0$ and so for any $y > 0$

$$\begin{aligned} p_Y(y) = F_Y'(y) &= 2\frac{d}{dy}\left[F_X(\sqrt{y}) - F_X(0)\right] = 2\frac{1}{2\sqrt{y}}F_X'(\sqrt{y}) \\ &= \frac{1}{\sqrt{y}}\,p_X(\sqrt{y}) = \frac{1}{\sqrt{2\pi y}}\,e^{-\frac{1}{2}y}. \end{aligned}$$

Summary

In this lecture we have seen

- $f_{X+Y} = f_X * f_Y$
- $f_{aX} = \frac{1}{a} f_X(\frac{\cdot}{a})$
- if $Z \sim \mathcal{N}(0,1)$ then $Z \sim \chi_1^2$.

Outline

I.i.d. sample

At the heart of statistic there is the idea of having a **sample** $\{X_1, \ldots, X_n\}$ of **observations** from a **population**. A "nice" sample is a sample which is i.i.d.

Identically distributed

All observations come from the same population, and so have same distribution/pmf/pdf:

$$\mathbb{P}(X_1 \leq t) = \mathbb{P}(X_i \leq t) \qquad \text{or equivalently} \qquad p_{X_1} = p_{X_i} \quad \text{with} \quad X_i \sim p_{X_i}.$$

Independent

Each observation is drawn purely at random without any dependence to the others, so

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1) \ldots p_{X_n}(x_n)$$

I.i.d. = Independent + Identically distributed

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = p(x_1) \ldots p(x_n), \qquad \text{with} \quad p = p_{X_1}.$$

Sample and samples

Statistics and CS/signal processing have different naming conventions.

	Statistics	CS
x_i	observation	sample
$\{x_1, \ldots, x_n\}$	sample	dataset

In this course, I will use the statistics terminology.

Properties of sums and means of i.i.d. r.v.s

- Let $X_1, \ldots, X_n$ be independent, then

$$\mathsf{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \mathsf{Var}(X_i).$$

- If the r.v. $X_1, \ldots, X_n$ are i.i.d., then

$$\mathsf{Var}(\bar{X}) = \tfrac{1}{n}\mathsf{Var}(X_1).$$

The family of Gaussian distributions is stable

A family $\mathcal{F}$ of random variables/distributions such that
“if X and Y belong to $\mathcal{F}$, then $\alpha X + \beta Y$ also belongs to $\mathcal{F}$” is called a "stable" family of distributions.

Distribution of a linear combination of Gaussians

Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ two *independent* r.v.s, then

$$a_1 X_1 + a_2 X_2 \sim \mathcal{N}(a_1\mu_1 + a_2\mu_2,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2).$$

This shows that the family of Gaussian distributions is stable.

There are other stable families but most of them are complicated. The other one that we have encountered is the family of Cauchy distributions (not obvious).

Distributions of some sums of continuous r.v.s

Sum of $\Gamma(r_i, \lambda)$ random variables

- If $X_1, \dots, X_n$ are independent with $X_i \sim \Gamma(r_i, \lambda)$ and $Z = X_1 + \dots + X_n$ then

$$Z \sim \Gamma(r_1 + \dots + r_n, \lambda).$$

In particular, if we set $r_i = 1$ in the previous result, we get

Sum of i.i.d. Exponential random variables

- If $X_1, \dots, X_n$ be i.i.d. with $X_i \sim \mathcal{E}(\lambda)$ and $Z = X_1 + \dots + X_n$ then $Z \sim \Gamma(n, \lambda)$.

And if we consider the case $r_i = \lambda = \frac{1}{2}$, we get

Sum of squares of i.i.d. standard normal random variables

- If $X_1, \dots, X_n$ be i.i.d. with $X_i \sim \mathcal{N}(0, 1)$ and $Z = X_1^2 + \dots + X_n^2$ then $X_i^2 \sim \chi^2(1)$ and

$$\forall i, \quad X_i^2 \sim \chi^2(1) \equiv \Gamma(\tfrac{1}{2}, \tfrac{1}{2}), \quad \text{and} \quad Z \sim \chi^2(n) \equiv \Gamma(\tfrac{n}{2}, \tfrac{1}{2}).$$

Summary

In this lecture we have seen

- The concepts of sample and observation from a population
- The concepts of independent observations and of identically distributed observations
- The definition of **i.i.d.** random variables

For some random variables the distribution of the sum of a sample has a known form:

- If $X_i \sim \Gamma(r_i, \lambda)$ independent and $Y = X_1 + \ldots + X_n$, then $Y \sim \Gamma(r_1 + \ldots + r_n, \lambda)$.
- If $E_i \overset{\text{i.i.d.}}{\sim} \mathcal{E}(\lambda)$, then $E_1 + \ldots + E_n \sim \Gamma(n, \lambda)$.
- If $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, then $Z_1 + \ldots + Z_n \sim \chi^2_n$.

Finally, any linear combination of independent Gaussians is Gaussian.

Proofs and extra material

(beyond the scope of the course)

Sum of two indep. r.v.s: Proof

$$
\begin{aligned}
F_Z(t) = \mathbb{P}(Z \le t) = \mathbb{P}(X + Y \le t) &= \mathbb{P}(X \le t - Y) \\
&= \int_{-\infty}^{\infty} \mathbb{P}(X \le t - Y \mid Y = y)\, p_Y(y)\, dy && \text{law of total probability} \\
&= \int_{-\infty}^{\infty} \mathbb{P}(X \le t - y)\, p_Y(y)\, dy && \text{by independence of } X \text{ and } Y \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{t-y} p_X(x)\, dx\, p_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{t} p_X(z-y)\, dz\, p_Y(y)\, dy && \text{change of var.:} \quad x = z - y \\
&= \int_{-\infty}^{t} \int_{-\infty}^{\infty} p_X(z-y)\, p_Y(y)\, dy\, dz && \text{exchanging integration order} \\
&= \int_{-\infty}^{t} f(z)\, dz && \text{with} \quad f(z) := \int_{-\infty}^{\infty} p_X(z-y)\, p_Y(y)\, dy.
\end{aligned}
$$

This shows that f is actually a valid pdf for Z.