Master Program in *Data Science and Business Informatics*

Statistics for Data Science

Lesson 08 - Continuous random variables

Salvatore Ruggieri

Department of Computer Science
University of Pisa, Italy
salvatore.ruggieri@unipi.it

Discrete random variables

DEFINITION. Let Ω be a sample space. A *discrete random variable* is a function $X : \Omega \to \mathbb{R}$ that takes on a finite number of values $a_1, a_2, \ldots, a_n$ or an infinite number of values $a_1, a_2, \ldots$.

DEFINITION. The *probability mass function* p of a discrete random variable X is the function $p : \mathbb{R} \to [0, 1]$, defined by

$$p(a) = \mathrm{P}(X = a) \qquad \text{for } -\infty < a < \infty.$$

- Support finite or countable $\{a_1, \ldots, a_n, \ldots\}$
  - $p(a_i) > 0$ for $i = 1, 2, \ldots$
  - $p(a) = 0$ if $a \notin \{a_1, a_2, \ldots\}$
  - $\sum_i p(a_i) = 1$
- What happens when the support is uncountable? E.g., $[0, 1]$ or $\mathbb{R}^+$ or $\mathbb{R}$
  - Many observations belong to the continuum (time, height, weight, blood pressure, temperature, distance, speed, etc.)

Discrete random variables

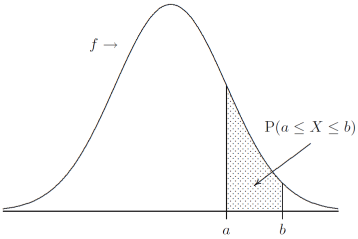
- Le X with support $\{0, 1\}$
  - $p(a) = P(X = a) = 1/2$ for a in the support
- Assume to expand the support to $\{0, 1/n, 2/n, \ldots, n/n\}$
  - $p(a) = P(X = a) = 1/(n+1)$ for a in the support
- Ok for $n \in \mathbb{N}$, but for $n \to \infty$, we have:

$$p(a) = P(X = a) = 0 \quad \text{for all } a$$

  which breaks the requirements of distribution functions! *[Trascurable but possible events]*
- Since $|\mathbb{R}| = 2^{\aleph_0} > \aleph_0 = |\mathbb{N}|$, $n = \infty$ is reached when considering the continuum!
- **Conclusion:** the idea of probability mass function does not extend to the continuum!

Continuous random variables

- We cannot assign a positive "mass" to a real number, but we can assign it to an interval!



DEFINITION. A random variable X is *continuous* if for some function $f : \mathbb{R} \rightarrow \mathbb{R}$ and for any numbers a and b with $a \leq b$,

$$\mathrm{P}(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x.$$

The function f has to satisfy $f(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$. We call f the *probability density function* (or *probability density*) of X.

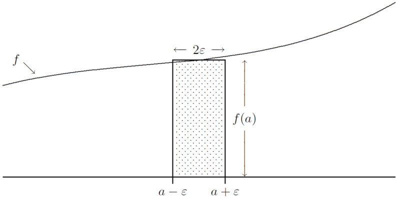
- Support of X is $\{x \in \mathbb{R} \mid f(x) > 0\}$
- $F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$ *[Cumulative Distribution Function]*
- $P(X \in A) = \int_{x \in A} f(x)dx$ for $A \subseteq \mathbb{R}$ measurable
  - There exist non-measurable subsets of $\mathbb{R}$, i.e., for which **we cannot assign a mass**
  - **Borel sets** are measurable: intervals over $\mathbb{R}$ closed under countable union and complement

Density function

$$P(X = a) \leq P(a - \epsilon \leq X \leq a + \epsilon) = \int_{a-\epsilon}^{a+\epsilon} f(x)dx = F(a+\epsilon) - F(a-\epsilon)$$

- for $\epsilon \to 0$, $P(a - \epsilon \leq X \leq a + \epsilon) \to 0$, hence $P(X = a) = 0$

- What is the meaning of the density function $f(x)$ then?
  - $f(a)$ is a (relative to other points) measure of how likely is X will be near a
  - "probability mass per unit length" around a: $f(a) \cdot 2\epsilon$



- Discrete vs Continuous Random Variables *[F(x) is a continuous function for continuous r.v.]*

$$F(a) = \sum_{a_i \leq a} p(a_i) \quad p(a_i) = F(a_i) - F(a_{i-1}) \qquad F(x) = \int_{-\infty}^{x} f(y)dy \quad f(x) = \frac{d}{dx}F(x)$$

$X \sim U(\alpha, \beta)$

DEFINITION. A continuous random variable has a *uniform distribution* on the interval $[\alpha, \beta]$ if its probability density function f is given by $f(x) = 0$ if x is not in $[\alpha, \beta]$ and

$$f(x) = \frac{1}{\beta - \alpha} \quad \text{for } \alpha \leq x \leq \beta.$$

We denote this distribution by $U(\alpha, \beta)$.

- $F(x) = \int_{-\infty}^{x} f(x)dx = \frac{1}{\beta-\alpha}\int_{\alpha}^{x} 1dx = \frac{x-\alpha}{\beta-\alpha}$ for $\alpha \leq x \leq \beta$
- Differently from p.m.f.'s, densities can be larger than 1 (and arbitrarily large)
  - E.g., for $U(0, 0.5)$ we have $f(x) = 2$

See R script

$X \sim Exp(\lambda)$

- For $X \sim Geo(p)$, we have: $F(x) = P(X \leq x) = 1 - (1-p)^{\lfloor x \rfloor}$ for $x \geq 0$
- extend to reals: $F(x) = P(X \leq x) = 1 - (1-p)^x = 1 - e^{x \cdot log(1-p)} = 1 - e^{-\lambda x}$ for $\lambda = -log(1-p)$
- $f(x) = \frac{dF}{dx}(x) = \lambda e^{-\lambda x}$

DEFINITION. A continuous random variable has an *exponential distribution* with parameter λ if its probability density function f is given by $f(x) = 0$ if $x < 0$ and

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0.$$

We denote this distribution by $Exp(\lambda)$.

- λ is the rate of events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate, e.g.,
  - $\lambda = 1/10$ number of bus arrivals per minute, or $1/\lambda = 10$ minutes to wait for bus arrival
  - $P(X > 1) = 1 - P(X \leq 1) = e^{-\lambda} = 0.9048$ probability of waiting more than 1 minute.

$X \sim Exp(\lambda)$

DEFINITION. A continuous random variable has an *exponential distribution* with parameter λ if its probability density function f is given by $f(x) = 0$ if $x < 0$ and

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0.$$

We denote this distribution by $Exp(\lambda)$.

- Plausible and empirically adequate model for:
  - time until a radioactive particle decays, time it takes before your next telephone call, ...
  - time until default (on payment to company debt holders) in reduced-form credit risk modeling, ...
  - time between animal roadkills, time between bank teller serves customers, ...
  - monthly and annual maximum values of daily rainfall, (some types of) surgery duration, ...
- Exponential is *memoryless*: $P(X > s + t | X > s) = e^{-\lambda \cdot (s+t)} / e^{-\lambda \cdot s} = e^{-\lambda \cdot t} = P(X > t)$

See R script and seeing-theory.brown.edu

$X \sim N(\mu, \sigma^2)$

Definition. A continuous random variable has a *normal distribution* with parameters μ and $\sigma^2 > 0$ if its probability density function f is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty < x < \infty.$$

We denote this distribution by $N(\mu, \sigma^2)$.

- "Normal" means "typical" or "common"
- Also called Gaussian distribution, after **Carl Friedrich Gauss**, but introduced by **De Moivre**
- Standard Normal/Gaussian is $N(0, 1)$
  - $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ sometimes written as $\phi(x)$
  - No closed form for $F(a) = \Phi(a) = \int_{-\infty}^{a} \phi(x)dx$
- Binomial approximation by a Normal distribution
  - $Bin(n, p) \approx N(np, np(1-p))$ for n large and $0 \ll p \ll 1$ [**De Moivre–Laplace theorem**]

CCDF of $Z \sim N(0,1)$

Table B.1. Right tail probabilities $1 - \Phi(a) = P(Z \geq a)$ for an $N(0,1)$ distributed random variable Z.

a	0	1	2	3	4	5	6	7	8	9
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379

- E.g., $P(Z \geq 1.04) = 0.1492$
- And in general for $X \sim N(\mu, \sigma^2)$?
  - Use identity $P(X \geq a) = P(Z \geq \frac{a-\mu}{\sigma})$ *[Proof in future lessons]*

See R script

Quantiles

DEFINITION. Let X be a continuous random variable and let p be a number between 0 and 1. The pth *quantile* or 100pth *percentile* of the distribution of X is the smallest number q_p such that

$$F(q_p) = \mathrm{P}(X \le q_p) = p.$$

The *median* of a distribution is its 50th percentile.

- Median m_X is $q_{0.5}$
- If $F()$ is *strictly* increasing, $q_p = F^{-1}(p)$
- E.g., for $Exp(\lambda)$, $F(a) = 1 - e^{-\lambda x}$, hence $F^{-1}(p) = \frac{1}{\lambda} \log \frac{1}{(1-p)}$

See R script

- General definition (also for discrete r.v.):

$$q_p = \inf_x \{P(X \le x) \ge p\}$$

Joint distributions: continuous random variables

DEFINITION. Random variables X and Y have a *joint continuous distribution* if for some function $f : \mathbb{R}^2 \to \mathbb{R}$ and for all numbers a_1, a_2 and b_1, b_2 with $a_1 \leq b_1$ and $a_2 \leq b_2$,

$$\mathrm{P}(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

The function f has to satisfy $f(x,y) \geq 0$ for all x and y, and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$. We call f the *joint probability density function* of X and Y.

- The marginal density functions of X and Y are:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

- Moreover, as in the univariate case:

$$F(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y)dxdy \qquad f(x,y) = \frac{d}{dx}\frac{d}{dy}F(x,y) = \frac{d^2}{dxdy}F(x,y)$$

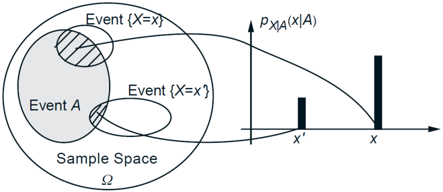
See R script

Recalling conditional distribution

Conditional distribution

Consider the joint distribution P_{XY} of X and Y. The conditional distribution of X given $Y \in B$ with $P_Y(Y \in B) > 0$, is the function $F_{X|Y \in B} : \mathbb{R} \to [0, 1]$:

$$F_{X|Y \in B}(a) = P_{X|Y}(X \leq a | Y \in B) = \frac{P_{XY}(X \leq a, Y \in B)}{P_Y(Y \in B)} \quad \text{for} \quad -\infty < a < \infty$$



- Distribution of X after knowing $Y \in B$.
- Chain rule: $P_{XY}(X \leq a, Y \in B) = P_{X|Y}(X \leq a | Y \in B) P_Y(Y \in B)$
- What if the distribution does not change w.r.t. the prior P_X?

Independence of two random variables

Independence $X \perp\!\!\!\perp Y$

A random variable X is independent from a random variable Y, if for all $P(Y \leq b) > 0$:

$$P_{X|Y}(X \leq a | Y \leq b) = P_X(X \leq a) \quad \text{for } -\infty < a < \infty$$

- Properties
  - $X \perp\!\!\!\perp Y$ iff $P_{XY}(X \leq a, Y \leq b) = P_X(X \leq a) \cdot P_Y(Y \leq b) \quad \text{for } -\infty < a, b < \infty$
  - $X \perp\!\!\!\perp Y$ iff $Y \perp\!\!\!\perp X$ *[Symmetry]*
- For X, Y **continuous** random variables:
  - $X \perp\!\!\!\perp Y$ iff $f_{XY}(x, y) = f_X(x) \cdot f_Y(y) \quad \text{for } -\infty < x, y < \infty$
  - **Exercise at home.** Prove it!
  - $X \perp\!\!\!\perp Y$ iff $P_{XY}(X \in \mathcal{A}, Y \in \mathcal{B}) = P_X(X \in \mathcal{A}) \cdot P_Y(Y \in \mathcal{B}) \quad \text{for } \mathcal{A}, \mathcal{B} \subseteq \mathbb{R}$ measurable

See R script

Independence of multiple random variables

Independence (factorization formula)

Random variables $X_1, \ldots, X_n$ are independent, if:

$$P_{X_1,\ldots,X_n}(X_1 \leq a_1, \ldots, X_n \leq a_n) = \prod_{i=1}^{n} P_{X_i}(X_i \leq a_i) \quad \text{for } -\infty < a_1, \ldots, a_n < \infty$$

- $X_1, \ldots, X_n$ **continuous** random variables are independent iff:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \quad \text{for } -\infty < x_1, \ldots, x_n < \infty$$

- **Definition:** $X_1, \ldots, X_n$ are **i.i.d.** (independent and identically distributed) if $X_1, \ldots, X_n$ are independent and $X_i \sim F$ for $i = 1, \ldots, n$ for some distribution F

Sum of independent continuous random variables

ADDING TWO INDEPENDENT CONTINUOUS RANDOM VARIABLES. Let X and Y be two independent continuous random variables, with probability density functions f_X and f_Y. Then the probability density function f_Z of $Z = X + Y$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, \mathrm{d}y$$

for $-\infty < z < \infty$.

- The integral is called the **convolution** of $f_X()$ and $f_Y()$
- $X, Y \sim Exp(\lambda)$, $Z = X + Y$, $\quad X, Y, Z \geq 0$ implies $0 \leq Y \leq Z$

$$f_Z(z) = \int_{-\infty}^{\infty} \lambda e^{-\lambda(z-y)} \lambda e^{-\lambda y} dy = \lambda^2 e^{-\lambda z} \int_0^z 1 dy = \lambda(\lambda z) e^{-\lambda z}$$

- $Z = X_1 + \ldots + X_n$ for $X_i \sim Exp(\lambda)$ independent: *[Earlang $Erl(n, \lambda)$ distribution]*

$$f_Z(z) = \frac{\lambda(\lambda z)^{n-1} e^{-\lambda z}}{(n-1)!}$$

$Gam(\alpha, \lambda)$

- Let λ be some average rate of an event, e.g., $\lambda = 1/10$ number of buses in a minute
  - The waiting time to see **one** event is exponentially distributed. E.g., probability of waiting x minutes to see one bus.
  - The waiting time to see n **events** is Erlang distributed. E.g., probability of waiting x minutes to see n buses.

DEFINITION. A continuous random variable X has a *gamma distribution* with parameters $\alpha > 0$ and $\lambda > 0$ if its probability density function f is given by $f(x) = 0$ for $x < 0$ and

$$f(x) = \frac{\lambda (\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \qquad \text{for } x \geq 0,$$

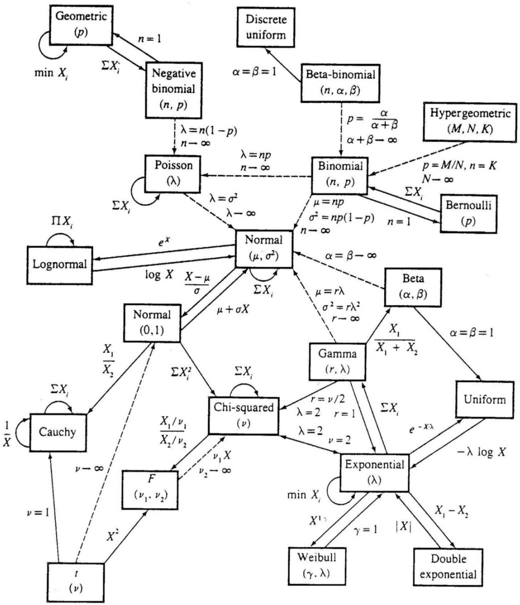
where the quantity $\Gamma(\alpha)$ is a normalizing constant such that f integrates to 1. We denote this distribution by $Gam(\alpha, \lambda)$.

- Extends $Erl(n, \lambda)$ from $n \in \mathbb{N}^+$ to $\alpha \in \mathbb{R}^+$ by **Euler's $\Gamma(\alpha)$** [$\Gamma(n) = (n-1)!$, see Lesson 06]
  - The waiting time to see α **quantities** is Gamma distributed. E.g., probability of waiting x minutes to see α volume of rain.

See R script

Common distributions

- **Probability distributions at Wikipedia**
- **Probability distributions in R**
- C. Forbes, M. Evans,
  N. Hastings, B. Peacock (2010)
  Statistical Distributions, 4th Edition
  Wiley



Relationships among common distributions. Solid lines represent transformations and special cases, dashed lines represent limits. Adapted from Leemis (1986).

The continuous Bayes' rule

BAYES' RULE. Suppose the events $C_1, C_2, \ldots, C_m$ are disjoint and $C_1 \cup C_2 \cup \cdots \cup C_m = \Omega$. The conditional probability of C_i, given an arbitrary event A, can be expressed as:

$$\mathrm{P}(C_i \mid A) = \frac{\mathrm{P}(A \mid C_i) \cdot \mathrm{P}(C_i)}{\mathrm{P}(A \mid C_1)\mathrm{P}(C_1) + \mathrm{P}(A \mid C_2)\mathrm{P}(C_2) + \cdots + \mathrm{P}(A \mid C_m)\mathrm{P}(C_m)}.$$

- **Definition.** Conditional density of X given $Y = y$ with $f_Y(y) > 0$:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

- Continuous Bayes' rule:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|t) f_X(t) dt}$$

- **Exercise at home.** A light bulb has a life-time $X \sim Exp(\lambda)$. λ is known to be $\sim U(1, 1.5)$. What can we say about the distribution of λ give observed life-time x? Code your solution also in R.