Lecture 5

5. Continuous Random Variables and Probability Density Function

Recall the definition of a random variable:
Let $(S, \mathcal{K}, P)$ be a probability space. A *random variable* is a function $X : S \to \mathbb{R}$ satisfying the property that for every $x \in \mathbb{R}$, the event

$$(X \leq x) \quad := \quad \{e \in S \mid X(e) \leq x\} \in \mathcal{K}.$$

Then, for every random variable X (not necessarily discrete), we defined the *cumulative distribution function* of X: the function $F = F_X : \mathbb{R} \to \mathbb{R}$, defined by

$$F_X(x) \quad = \quad P(X \leq x).$$

Definition 5.1.

Let $(S, \mathcal{K}, P)$ be a probability space. A random variable $X : S \to \mathbb{R}$ is a **continuous random variable**, if the set of values $X(S)$ is any (finite or infinite) interval in $\mathbb{R}$.

Proposition 5.2.

Let X be a continuous random variable with cdf $F : \mathbb{R} \to \mathbb{R}$. Then F is *absolutely continuous*, i.e. there exists a real function $f : \mathbb{R} \to \mathbb{R}$, such that

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \tag{5.1}$$

for all $x \in \mathbb{R}$.

Definition 5.3.

Let X be a continuous random variable. Then the function f from Proposition 5.2 is called the **probability density function (pdf)** of X.

Remark 5.4.

So, we use "pdf" to describe any random variable, "probability *distribution* function" for discrete random variables and "probability *density* function" for the continuous case. The term "density" in the continuous case, extends in a natural way the notion of "distribution" from the discrete case, with summation being replaced by integration. Note that not all books or authors make that distinction (e.g. in Matlab, they are all called "densities").

Recall the properties of a cdf (Theorem 2.4., Chapter 3, Lecture 4).

Theorem (Properties of a cdf).

Let X be a random variable with cdf $F : \mathbb{R} \to \mathbb{R}$. Then F has the following properties:

a) If $a < b$ are real numbers, then $P(a < X \leq b) = F(b) - F(a)$.

b) F is monotonely increasing, i.e. if $a < b$, then $F(a) \leq F(b)$.

c) F is right continuous, i.e. $F(x+0) = F(x)$, for every $x \in \mathbb{R}$, where $F(x+0) = \lim\limits_{y \searrow x} F(y)$ is the limit from the right at x.

d) $\lim\limits_{x \to -\infty} F(x) = 0$ and $\lim\limits_{x \to \infty} F(x) = 1$.

e) For every $x \in \mathbb{R}$, $P(X < x) = F(x-0) = \lim\limits_{y \nearrow x} F(y)$ and $P(X = x) = F(x) - F(x-0)$.

From them, some common properties of a density function can be derived.

Theorem 5.5.

Let X be a continuous random variable with cdf F and density function f. Then the following properties hold:

a) $F'(x) = f(x)$, for all $x \in \mathbb{R}$.

b) $f(x) \geq 0$, for all $x \in \mathbb{R}$.

c) $\int\limits_{\mathbb{R}} f(t)dt = 1$.

d) For every $x \in \mathbb{R}$, $P(X = x) = 0$ and for every $a, b \in \mathbb{R}$ with $a < b$,

$$\begin{aligned} P(a < X \leq b) &= P(a < X \leq b) = P(a < X < b) = P(a \leq X \leq b) \\ &= \int\limits_a^b f(t)\, dt\,. \end{aligned} \tag{5.2}$$

Proof.

a) This property follows directly from the definition of a continuous random variable, by differentiating both sides of (5.1):

$$F(x) \;=\; \int_{-\infty}^{x} f(t)\,dt \;\Longrightarrow\; F'(x) \;=\; f(x).$$

b) Recall from Theorem 2.4. that F is monotonely increasing. Thus, its derivative is nonnegative, for every $x \in \mathbb{R}$.

c) Recall that $\lim_{x\to\infty} F(x) = 1$ (Theorem 2.4.d)). So, we have

$$\int_{\mathbb{R}} f(t)\,dt = \int_{-\infty}^{\infty} f(t)\,dt = \lim_{x\to\infty} \int_{-\infty}^{x} f(t)\,dt = \lim_{x\to\infty} F(x) \;=\; 1.$$

□

Proof.

d) To prove the first part, let $x \in \mathbb{R}$ be fixed and recall from Theorem 2.4.e) that

$$P(X = x) = F(x) - F(x - 0).$$

But for a continuous random variable, F is absolutely continuous, so continuous at every point, thus,

$$F(x) = F(x + 0) = F(x - 0).$$

Hence, $P(X = x) = 0$.

Now let $a, b \in \mathbb{R}$ with $a < b$. By Theorem 2.4.a), we have

$$P(a < X \leq b) = F(b) - F(a) = \int_{-\infty}^{b} f(t)\, dt - \int_{-\infty}^{a} f(t)\, dt = \int_{a}^{b} f(t)\, dt,$$

which, by the first part, is equal to all the other probabilities in (5.2).

□

Remark 5.6.

So, probabilities involving continuous random variables can be computed by integrating the density function over the given sets.

Furthermore, recall from Calculus that the integral $\int_a^b f(x)dx$ of a non-negative function f equals the area below the curve $y = f(x)$, above the x-axis, between the vertical lines $x = a$ and $x = b$.

Therefore, geometrically, probabilities are represented by areas (see Figure 1). This aspect will be important later on. Also, by Theorem 5.5d), the *total* area under the graph of a density function is equal to 1.

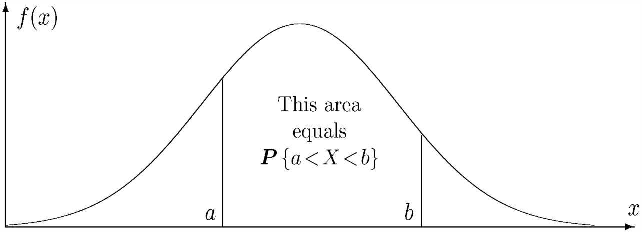


Figure 1: Probability as area, for continuous random variables

6. Common Continuous Distributions

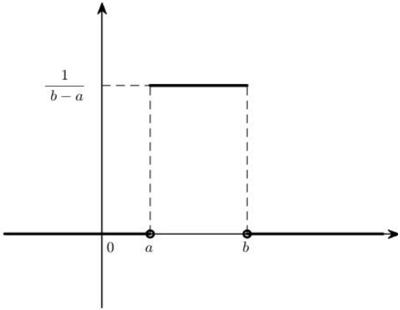
Uniform Distribution $U(a,b)$

A random variable X has a Uniform distribution ($\boxed{\text{unif}}$) with parameters $a, b \in \mathbb{R},\ a < b$, if its pdf is

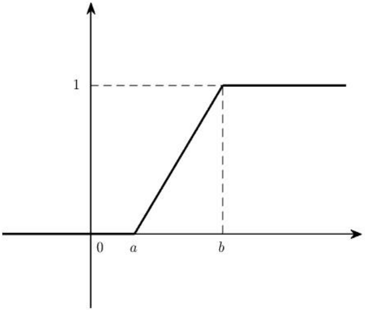
$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } x \in [a,b] \\ 0, & \text{if } x \notin [a,b]. \end{cases} \tag{6.1}$$

Then, by (5.1), its cdf is

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0, & \text{if } x \le a \\ \dfrac{x-a}{b-a}, & \text{if } a < x \le b \\ 1, & \text{if } x \ge b\,. \end{cases} \tag{6.2}$$



(a) Density Function (pdf)



(b) Cumulative Distribution Function (cdf)

Figure 2: Uniform Distribution

Remark 6.1.

1. The Uniform distribution is used when a variable can take *any* value in a given interval, equally probable. For example, locations of syntax errors in a program, birthdays throughout a year, etc.

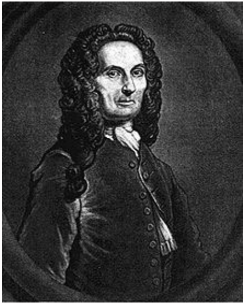
2. A special case is that of a **Standard Uniform Distribution**, where $a = 0$ and $b = 1$. The pdf and cdf are given by

$$f_U(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}, \quad F_U(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x \leq 1 \\ 1, & x \geq 1\,. \end{cases} \tag{6.3}$$

Standard Uniform variables play an important role in stochastic modeling; in fact, *any* random variable, with any thinkable distribution (discrete or continuous) can be generated from Standard Uniform variables.

Normal Distribution $N(\mu, \sigma)$

The Normal distribution is, by far, the most important distribution, underlying many of the modern statistical methods used in data analysis. It was first described in the 1700's by De Moivre, as a limiting case for the Binomial distribution (when n, the number of trials, becomes infinite), but did not get much attention.



Abraham de Moivre (1667 - 1754)

Half a century later, both Laplace and Gauss (independently of each other) rediscovered it in conjunction with the behavior of errors in astronomical measurements. It is also referred to as the “Gaussian” distribution.

Pierre-Simon de Laplace
(1749 - 1827)

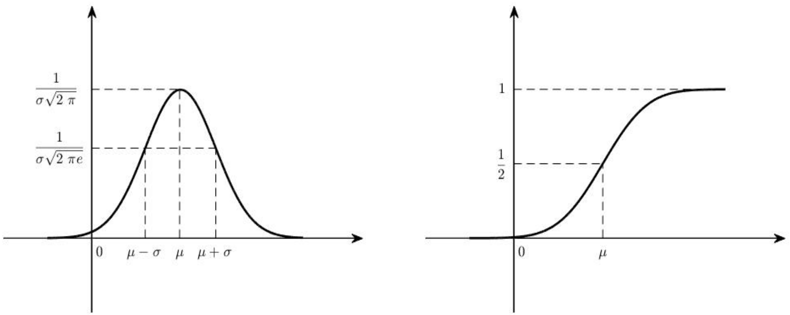
Carl Friedrich Gauss
(1777 - 1855)

A random variable X has a Normal distribution (norm) with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, if its pdf is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \tag{6.4}$$

The cdf of a Normal variable is then given by

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt. \tag{6.5}$$



(a) Density Function (pdf) (b) Cumulative Distribution Function (cdf)

Figure 3: Normal Distribution

The graph of the Normal density is a symmetric, bell-shaped curve (known as "Gauss's bell" or "Gauss's bell curve") centered at the value of the first parameter μ, as can be seen in Figure 3(a). The graph of the cdf of a Normally distributed random variable is given in Figure 3(b) and this is approximately what the graph of the cdf of *any* continuous random variable looks like.

Remark 6.2.

1. There is an important particular case of a Normal distribution, namely $N(0, 1)$, called the **Standard (or Reduced) Normal Distribution**. A variable having a Standard Normal distribution is usually denoted by Z. The density and cdf of Z are given by

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \;\; x \in \mathbb{R} \;\; \text{and} \;\; F_Z(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt. \qquad (6.6)$$

The function F_Z given in (6.6) is known as *Laplace's function* (or *the error function*) and its values can be found in tables or can be computed by most mathematical software.

2. As noticed from (6.5) and (6.6), there is a relationship between the cdf of any Normal $N(\mu, \sigma)$ variable X and that of a Standard Normal variable Z, namely

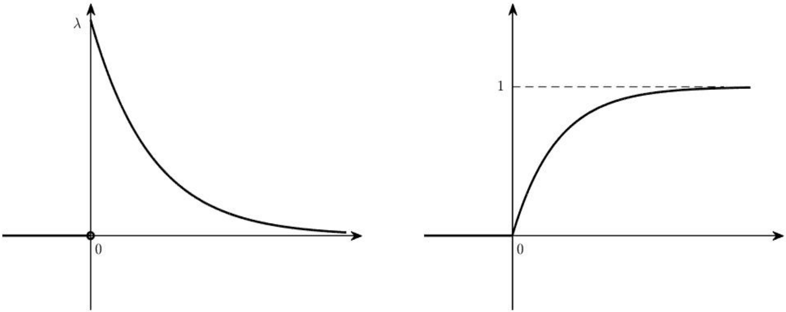
$$F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right).$$

Exponential Distribution $Exp(\lambda)$

A random variable X has an Exponential distribution ($\boxed{\text{exp}}$) with parameter $\lambda > 0$, if its density function and cdf are given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \text{ and } F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}, \quad (6.7)$$

respectively. Their graphs are given in Figure 4.



(a) Density Function (pdf) (b) Cumulative Distribution Function (cdf)

Figure 4: Exponential Distribution

Remark 6.3.

1. The Exponential distribution is often used to model time: lifetime, waiting time, halftime, interarrival time, failure time, time between rare events, etc.

In a sequence of rare events (where the number of rare events has a Poisson distribution), the time between two consecutive rare events (as well as the time of the occurrence of the first rare event) is Exponential.

The parameter λ represents the frequency of rare events, measured in time^{-1} units.

2. A word of **caution** here: The parameter μ in Matlab (where the Exponential pdf is defined as $\frac{1}{\mu}e^{-\frac{1}{\mu}x}, x \geq 0$) is actually $\mu = 1/\lambda$. It all comes from the different interpretation of the "frequency". For instance, if the frequency is "2 per hour", then $\lambda = 2/\text{hr}$, but this is equivalent to " one every half an hour", so $\mu = 1/2$ hours.

The parameter μ is measured in time units.

Remark 6.4.

1. The Exponential distribution is a special case of a more general distribution, namely the *Gamma*(a, b), $a, b > 0$, distribution ($\boxed{\text{gam}}$). The Gamma distribution models the *total* time of a multistage scheme.

2. If $\alpha \in \mathbb{N}$, then the sum of α independent *Exp*(λ) variables has a *Gamma*$(\alpha, 1/\lambda)$ distribution.

Remark 6.5.

In Statistics, the most widely used distributions are the following:
– the Normal distribution, $N(\mu, \sigma)$, especially $N(0, 1)$,
– the Student (T) distribution, $T(n)$,
– the χ^2 distribution, $\chi^2(n)$,
– the Fisher (F) distribution, $F(m, n)$.

7. Continuous Random Vectors, Joint Density Function and Marginal Densities

Again, we will restrict our study to the two-dimensional case.

Definition 7.1.

Let $(S, \mathcal{K}, P)$ be a probability space.

- A two-dimensional **random vector** is a function $(X, Y) : S \to \mathbb{R}^2$ satisfying the condition

$$(X \leq x, Y \leq y) = \{e \in S \mid X(e) \leq x, Y(e) \leq y\} \in \mathcal{K}, \qquad (7.1)$$

  for all $(x, y) \in \mathbb{R}^2$.
- The function $F : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$F(x, y) = P(X \leq x, Y \leq y) \qquad (7.2)$$

  is called the **joint cumulative distribution function (joint cdf)** of the vector (X, Y).

The properties of the cdf of a random variable translate very naturally to a random vector, as well.

Theorem 7.2.

Let (X, Y) be a random vector with joint cdf $F : \mathbb{R}^2 \to \mathbb{R}$ and let $F_X, F_Y : \mathbb{R} \to \mathbb{R}$ be the cdf's of X and Y, respectively. Then following properties hold:

a) If $a_k < b_k,\ k = \overline{1, 2}$, then

$$P(a_1 < X \le b_1, a_2 < Y \le b_2) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2). \tag{7.3}$$

b) F is monotonically increasing in each variable.

c) F is right continuous in each variable.

d) $\lim\limits_{x,y\to\infty} F(x, y) = 1,$

$\lim\limits_{y\to-\infty} F(x, y) = \lim\limits_{x\to-\infty} F(x, y) = 0,\ \forall x, y \in \mathbb{R},$

$\lim\limits_{y\to\infty} F(x, y) = F_X(x),\ \forall x \in \mathbb{R},\ \lim\limits_{x\to\infty} F(x, y) = F_Y(y),\ \forall y \in \mathbb{R}.$

Proof.

We give a selected proof.

a) This proof is similar to the proof for random variables.

d) Let $x \in \mathbb{R}$. We have

$$\lim_{y \to \infty} F(x, y) = P(X \leq x, Y \leq \infty) = P(X \leq x) = F_X(x)$$

and by symmetry, $\lim_{x \to \infty} F(x, y) = F_Y(y), \ \forall y \in \mathbb{R}$. Then, it follows that

$$\lim_{x, y \to \infty} F(x, y) = \lim_{y \to \infty} F_Y(y) = \lim_{x \to \infty} F_X(x) = 1.$$

For any $x \in \mathbb{R}$,

$$\lim_{y \to -\infty} F(x, y) = P(X \leq x, Y \leq -\infty) = P(\emptyset) \ = \ 0$$

and by symmetry, $\lim_{x \to -\infty} F(x, y) = 0, \ \forall y \in \mathbb{R}$, also. □

Definition 7.3.

Let $(S, \mathcal{K}, P)$ be a probability space. A random vector $(X, Y) : S \to \mathbb{R}^2$ is a **continuous random vector**, if the set of values $(X, Y)(S)$ is a (finite or infinite) continuous subset of $\mathbb{R}^2$.

Proposition 7.4.

Let (X, Y) be a continuous random vector with joint cdf $F : \mathbb{R}^2 \to \mathbb{R}$. Then F is *absolutely continuous*, i.e. there exists a real function $f : \mathbb{R}^2 \to \mathbb{R}$, such that

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, du\, dv, \ \forall x, y \in \mathbb{R}. \qquad (7.4)$$

Definition 7.5.

Let (X, Y) be a continuous random vector. Then the function f from Proposition 7.4 is called the **joint probability density function (joint pdf)** of (X, Y).

Theorem 7.6.

Let (X, Y) be a continuous random vector with joint cdf F and joint density function f. Let $F_X, F_Y : \mathbb{R} \to \mathbb{R}$ be the cdf's of X and Y and $f_X, f_Y : \mathbb{R} \to \mathbb{R}$ be the pdf's of X and Y, respectively. Then the following properties hold:

a) $\dfrac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$, for all $(x,y) \in \mathbb{R}^2$.

b) $f(x,y) \geq 0$, for all $(x,y) \in \mathbb{R}^2$.

c) $\displaystyle\iint_{\mathbb{R}^2} f(x,y)\, dxdy = 1.$

d) For any domain $D \subseteq \mathbb{R}^2$, $P\big((X,Y) \in D\big) = \displaystyle\iint_D f(x,y)\, dxdy.$

e) $f_X(x) = \displaystyle\int_{\mathbb{R}} f(x,y)\, dy,\ \forall x \in \mathbb{R}$ and $f_Y(y) = \displaystyle\int_{\mathbb{R}} f(x,y)\, dx,\ \forall y \in \mathbb{R}.$

Proof.

These properties follow easily from Proposition 7.4 and the properties of the joint cdf stated in Theorem 7.2. □

Remark 7.7.

When obtained from the joint pdf of the vector (X, Y), the pdf's f_X and f_Y are called *marginal* densities.

Definition 7.8.

Two continuous random variables X and Y are **independent** if

$$f_{(X,Y)}(x,y) = f_X(x) f_Y(y), \tag{7.5}$$

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

8. Functions of Continuous Random Variables

Proposition 8.1.

Let $g : \mathbb{R} \to \mathbb{R}$ be a strictly monotone and differentiable function, with $g'(x) \neq 0, \forall x \in \mathbb{R}$. Let X be a continuous random variable with pdf f_X and let $Y = g(X)$. Then for $y \in \mathbb{R}$, the pdf of Y is given by

$$f_Y(y) = \begin{cases} \dfrac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}, & \text{if } y \in g(\mathbb{R}) \\ 0, & \text{if } y \notin g(\mathbb{R}) . \end{cases} \tag{8.1}$$

Proof.

Note that g being strictly monotone implies being injective. Then $g : \mathbb{R} \to g(\mathbb{R})$ is bijective and, hence, invertible. Thus, g^{-1} exists on $g(\mathbb{R})$.

Proof.

Case I. Assume g is strictly increasing. Then so is g^{-1}, $g' > 0$ and we have

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

$$= \begin{cases} P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)), & \text{if } y \in g(\mathbb{R}) \\ 0, & \text{if } y < \inf g(\mathbb{R}) \\ 1, & \text{if } y > \sup g(\mathbb{R}) \end{cases}.$$

Differentiate to obtain

$$f_Y(y) = \begin{cases} F_X'(g^{-1}(y)) \cdot (g^{-1}(y))', & \text{if } y \in g(\mathbb{R}) \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \dfrac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}, & \text{if } y \in g(\mathbb{R}) \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \dfrac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}, & y \in g(\mathbb{R}) \\ 0, & y \notin g(\mathbb{R}) \end{cases}.$$

Proof.

Case II. If g is strictly decreasing, then so is g^{-1}, $g' < 0$ and we find that $F_Y(y)$ is

$$P(g(X) \leq y) = \begin{cases} P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)), & \text{if } y \in g(\mathbb{R}) \\ 0, & \text{if } y < \inf g(\mathbb{R}) \\ 1, & \text{if } y > \sup g(\mathbb{R}) \end{cases}$$

and

$$f_Y(y) = \begin{cases} -\dfrac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}, & \text{if } y \in g(\mathbb{R}) \\ 0, & \text{else} \end{cases} = \begin{cases} \dfrac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}, & y \in g(\mathbb{R}) \\ 0, & y \notin g(\mathbb{R}) \end{cases}.$$

Thus, in both cases, we have (8.1). □