INTRODUCTION TO CALCULUS

MATH 1A

Unit 23: PDF and CDF

Lecture

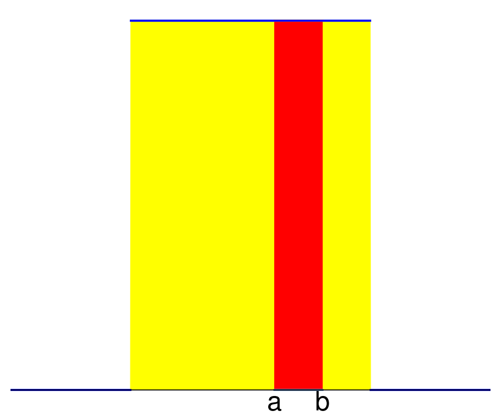
23.1. In **probability theory** one considers functions too:

Definition: A non-negative piece-wise continuous function $f(x)$ which has the property that $\int_{-\infty}^{\infty} f(x)\, dx = 1$ is called a **probability density function**. [1]. For every interval $A = [a, b]$, the number

$$\mathrm{P}[A] = \int_a^b f(x)\, dx$$

is the **probability** of the event that the data are in A.

23.2. An important case is the function $f(x)$ which is 1 on the interval $[0, 1]$ and 0 else. It is the **uniform distribution** on $[0, 1]$. **Random number generators** in computers first of all generate random numbers with that distribution. In Mathematica, you get such numbers by evaluating **Random[]**. In Python you get it with **import random; random.uniform(0,1)**. The probability $\int_{0.3}^{0.7} f(x)\, dx$ for example is 0.4. Here is the **function** $f(x)$:

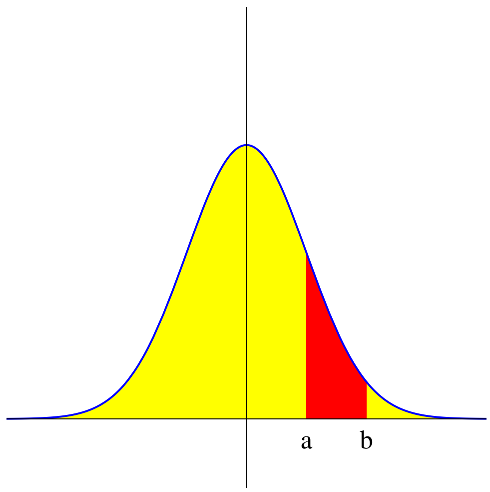


23.3. An other important probability density is the **standard normal distribution**, also called **Gaussian distribution**. You can get such random numbers in python by **import random; random.gauss(0,1)**.

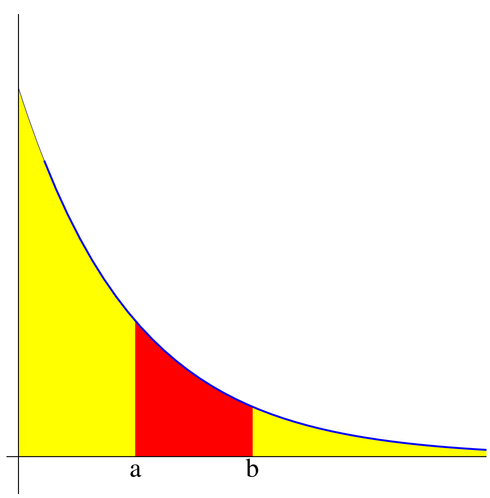
Definition: The **normal distribution** has the density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} .$$

23.4. It is the distribution which appears most often if data can take both positive and negative values. One reason why it appears so often is that if one observes different unrelated quantities then their sum, suitably normalized is close to the normal distribution. **Errors** for example often have normal distribution. Astronomers like Galileo noticed this already in 1630ies. Laplace in 1774 first defined probability distributions and Gauss in 1801 first looked at the normal distribution, also in the context of analyzing astronomical data when searching for the **dwarf planet** Ceres.



Example: The probability density function of the **exponential distribution** is defined as $f(x) = e^{-x}$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$. It is used to used measure lengths of arrival times like the time until you get the next email. The density is zero for negative x because there is no way we can travel back in time.
What is the probability that you get an email between times $x = 1$ and times $x = 2$? Answer: it is $\int_1^2 f(x)\, dx = e^{-1} - e^{-2} = 1/e - 1/e^2$.



Definition: Assume f is a probability density function (PDF). The anti-derivative $F(x) = \int_{-\infty}^{x} f(t)\, dt$ is called the **cumulative distribution function** (CDF).

Example: For the exponential function the cumulative distribution function is

$$\int_{-\infty}^{x} f(x)\, dx = \int_{0}^{x} f(x)\, dx = -e^{-x}|_0^x = 1 - e^{-x} .$$

Definition: The probability density function $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$ is called the **Cauchy distribution**.

Example: Find the cumulative distribution function of the Cauchy distribution.
Solution:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \frac{1}{\pi}\arctan(x)|_{-\infty}^{x} = (\frac{1}{\pi}\arctan(x) + \frac{1}{2}) .$$

Definition: The **mean** of a distribution is the number

$$m = \int_{-\infty}^{\infty} x f(x)\, dx .$$

Example: The mean of the distribution $f(x_) = e^{-x}$ on $[0, \infty)$ is

$$\int_{0}^{\infty} x e^{-x}\, dx .$$

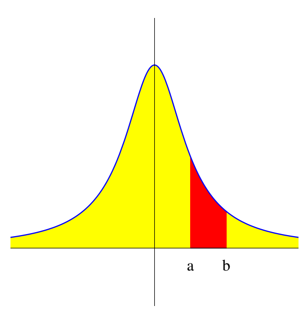
We do not know yet how to compute this but learn a technique later. For now, we have to guess the anti derivative or being told that it is $(-1-x)e^{-x}$. We can check that the derivative of this function is indeed e^{-x}. So,

$$\int_{0}^{\infty} x e^{-x}\, dx = \lim_{t\to\infty}(-1-x)e^{-x}|_0^t = \lim_{t\to\infty}(-1-t)e^{-t} + 1 = 1 .$$

23.5. The distribution looks similar to the Gaussian distribution, but it has more risk. The **variance** of this distribution

$$\int_{\infty}^{\infty} x^2 f(x)\, dx = (1/\pi)\int_{-\infty}^{\infty} \frac{x^2}{1+x^2}\, dx$$

is infinite. The function $\frac{x^2}{1+x^2}$ is asymptotically 1 and has a divergent integral from $-\infty$ to ∞.



Homework

Problem 23.1: Assume the probability density for the time you have to wait for your next shopping delivery you get is $f(x) = 7e^{-7x}$ where x is time in hours.
What is the probability that you get your next text message in the next 5 hours but not before 1 hour?

Problem 23.2: Assume the probability distribution for the waiting time to the next rain is $f(x) = (1/5)e^{-x/5}$, where x has days as unit.
What is the probability to get a warm day between tomorrow and after tomorrow that is between $x = 1$ and $x = 2$?

Problem 23.3: a) Verify that the function $f(x)$ which is defined to be zero outside the interval $[-1, 1]$ and given as $\frac{1}{\pi\sqrt{1-x^2}}$ inside the interval $[-1, 1]$ is a probability distribution.
b) What is the cumulative distribution function $F(x)$?
c) What is the expectation $\int_{-1}^{1} xf(x)\, dx$?

Problem 23.4: Assume some risky experiment leads to discrepancies (errors) which are distributed according to the Cauchy distribution $f(x) = 1/(\pi(1 + x^2))$.
a) Find the probability that the error is in absolute value larger than 1.
b) Find the probability that the error is smaller than $-\sqrt{3}/2$. In other words, what is $\int_{-\infty}^{-\sqrt{3}/2} f(x)\, dx$?

Problem 23.5: If $f(x)$ is a probability distribution, then $m = \int_{-\infty}^{\infty} xf(x)\, dx$ is called the **mean** of the distribution. a) Compute the mean for the standard normal distribution.
b) Compute the mean for the Cauchy distribution $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.
c) Compute the mean for the Exponential distribution $f(x) = e^{-x}$ on $[0, \infty)$. You might want to guess an anti-derivative of $xf(x)$. .

Oliver Knill, knill@math.harvard.edu, Math 1a, Harvard College, Spring 2021