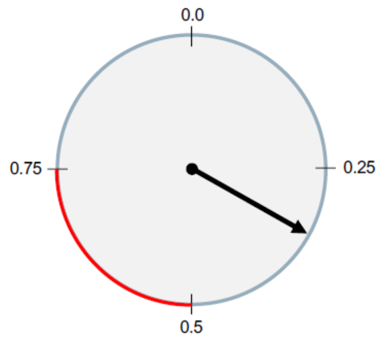
Lecture 6: Continuous Random Variables

Yunwen Lei

School of Computer Science, University of Birmingham

1 Continuous Random Variables

In the last class, we study discrete random variables. In this class, we will move on to continuous random variables. Informally, a random variable X with an uncountable range $X(\Omega)$ is a continuous random variable. An example is a random draw of a spinner, whose outcome can be any number in $[0, 1]$



What needs to change when working with continuous as opposed to discrete distributions

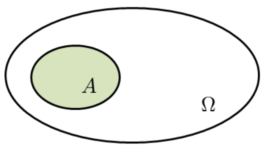
- the probability of a spinner outputting any particular number is 0 $\mathbb{P}(X = a) = 0$, since the number of values which may be assumed by the random variable is infinite. This extends to any countable collection of real numbers, i.e., for any sequence $\{a_j\}$ $\mathbb{P}(X = a_1 \text{ or } a_2 \text{ or } \cdots) = 0$.
- we can only think about intervals $\mathbb{P}[0.5 < X < 0.75] = 0.25$

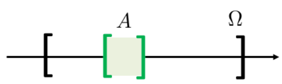
Formally, continuous random variables are defined in terms of CDFs.

Definition 1 (Continuous Random Variables)**.** A random variable having a continuous CDF is said to be a continuous random variable.

Since assigning a probability to any specific value is not meaningful, a natural question is how to define $\mathbb{P}(\{X \in A\})$?

A natural solution is to measure the size of a set: if the randomly variable has the same likelihood to take any value, then we should assign the probability according to its relative size





$$\mathbb{P}(\{x \in A\}) = \frac{\text{“size of } A\text{”}}{\text{“size of } \Omega\text{”}}$$

To motivate the definition of density function, we consider the following example.

Example 1. Suppose that the sample space is the interval $\Omega = [0, 5]$ and the event is $A = [2, 3]$. To measure the size of A, we can integrate A to determine the length

$$\mathbb{P}(\{X \in [2, 3]\}) = \frac{\text{“size of } A\text{”}}{\text{“size of } \Omega\text{”}} = \frac{\int_A dx}{\int_\Omega dx} = \frac{\int_2^3 dx}{\int_0^5 dx} = \frac{1}{5}.$$

More formally,

$$\mathbb{P}(\{X \in [2, 3]\}) = \frac{\int_A dx}{\int_\Omega dx} = \frac{\int_A dx}{|\Omega|} = \int_A \underbrace{\frac{1}{|\Omega|}}_{\text{equally important over } \Omega} dx. \tag{1.1}$$

In the above example, the probability for any interval is defined by an integral of the function $f(x) = 1/|\Omega|$ over the interval. What happens if we want to relax the "equiprobable" assumption? A possible solution is to replace the constant function $1/|\Omega|$ with $f_X(x)$. This will give us

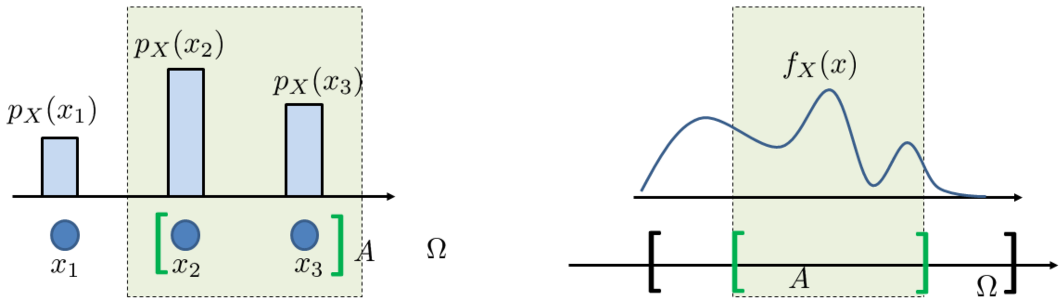
$$\mathbb{P}(\{X \in A\}) = \int_A \underbrace{f_X(x)}_{\text{replace } 1/|\Omega|} \, dx \tag{1.2}$$

If you compare it with PMF, we note that when X is discrete, then

$$\mathbb{P}(\{X \in A\}) = \sum_{x \in A} P_X(x) \tag{1.3}$$

This shows a clear difference between continuous and discrete random variables:

- the probability of $X \in A$ is computed via an integral over A for continuous random variables
- the probability of $X \in A$ is computed via a summation over A for discrete random variables



Left A probability mass function (PMF) tells us the relative frequency of a state when computing the probability. In this example, the "size" of A is $P_X(x_2) + P_X(x_3)$

Right A probability density function (PDF) is the infinitesimal version of the PMF. Thus, the "size" of A is the integration over the PDF.

Definition 2 (Probability density function)**.** Let X be a continuous random variable. The probability density function of X is a function $f_X : \mathbb{R} \mapsto \mathbb{R}_+$, when integrated over an interval $[a, b]$, yields the probability of obtaining $a \le X \le b$:

$$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x) dx.$$

The above definition shows a core requirement of a function to be a density function: its integral yields the probability. From the above definition, we know that a probability density function should be nonnegative and should integral to 1 over the range, i.e., $f_X(x) \ge 0$ and $\int_{\Omega(X)} f_X(x) dx = 1$.

Example 2. Let

$$f_X(x) = \begin{cases} 3x^2, & \text{if } x \in [0, 1] \\ 0, & \text{otherwise.} \end{cases}$$

Let $A = [0, 0.5]$. Then the probability $\mathbb{P}(\{X \in A\})$ is

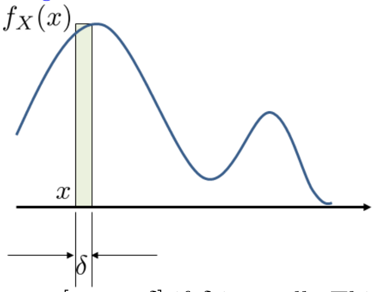
$$\mathbb{P}(0 \le X \le 0.5) = \int_0^{0.5} 3x^2 dx = \int_0^{0.5} dx^3 = 1/8.$$

Example 3. Let $f_X(x) = 1/|\Omega|$ with $\Omega = [0, 5]$. Let $A = [3, 5]$. Then the probability $\mathbb{P}(\{X \in A\})$ is

$$\mathbb{P}(3 \le X \le 5) = \int_3^5 \frac{1}{|\Omega|} dx = \int_3^5 \frac{1}{5} dx = \frac{2}{5}.$$

We know that the probability of an event should always be less than or equal to 1. Does this extends to PDF, i.e., can $f_X(x) \geq 1$? The answer is yes. Note $f_X(x)$ is not the probability of having $X = x$. Indeed, $f_X(x)$ can be interpreted as the probability per unit length:

$$\mathbb{P}(X \in [x, x+\delta]) = \int_x^{x+\delta} f_X(t)dt \approx f_X(x)\delta$$



In the approximate equation, we use the fact that $f_X(t) \approx f_X(x)$ for $t \in [x, x+\delta]$ if δ is small. This shows that the probability of $X \in [x, x+\delta]$ can be approximated by the PDF times the length of the interval. If the length is small, we can allow large PDF and still get a probability less than 1. Below is an example of random variable with infinitely large PDF.

Example 4. Consider a random variable X with PDF $f_X(x) = \frac{1}{2\sqrt{x}}$ for $x \in (0, 1]$ and 0 otherwise.

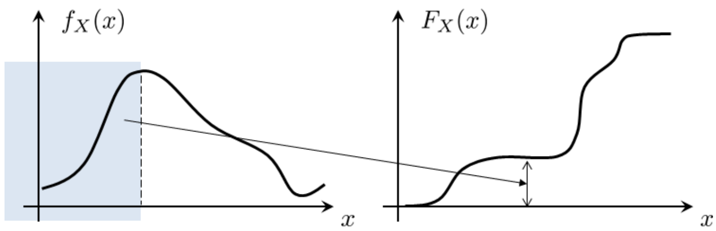
- It is clear that $f_X(x) \to \infty$ as $x \to 0$
- $f_X(x)$ remains a valid PDF because

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 \frac{1}{2\sqrt{x}}dx = \sqrt{x}|_0^1 = 1.$$

Analogous to discrete random variables, we can also define cumulative distribution function (CDF) for continuous random variables. Note the definition is the same as that for discrete random variables. This shows the motivation of introducing CDF: it provides a unifying concept applicable to both type of random variables.

Definition 3 (Cumulative Distribution Function, CDF)**.** Let X be a continuous random variable. The Cumulative Distribution Function of X is

$$F_X(x) = \mathbb{P}(X \leq x).$$



It is interesting to show the connection between PDF and CDF. One can get the other quantity if given an quantity: one direction is by integration and the other is by differentiation

Remark 1.

- If X is a continuous random variable and $a \leq b$, then (integration)

$$\int_a^b f_X(x)dx = \mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a) \tag{1.4}$$

- If F_X is differentiable at x, then (differentiation)

$$f_X(x) = \frac{dF_X(x)}{dx} = \frac{d}{dx}\int_{-\infty}^x f_X(y)dy. \tag{1.5}$$

2 Common Continuous Random Variables

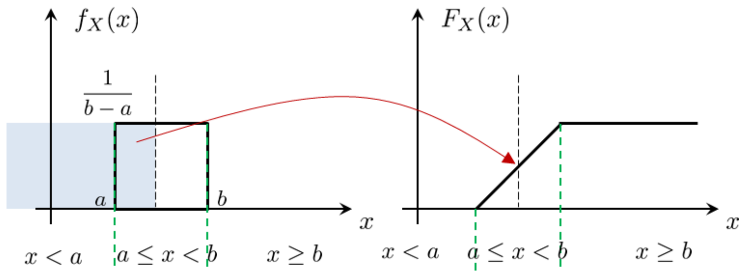
In this section, we will introduce several common continuous random variables.

2.1 Uniform Random Variable

Definition 4 (Uniform Random Variable)**.** We say X is a continuous uniform random variable on $[a, b]$ if the PDF is

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \le x \le b \\ 0, & \text{otherwise.} \end{cases}$$

We write $X \sim \text{Uniform}(a, b)$.



The CDF of a uniform random variable is

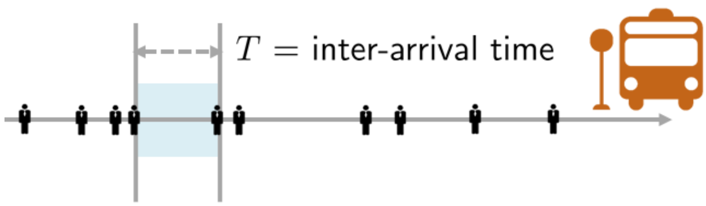
$$F_X(x) = \begin{cases} 0, & \text{if } x < a \\ \frac{x-a}{b-a}, & \text{if } a \le x \le b \\ 1, & \text{otherwise.} \end{cases}$$

2.2 Exponential Random Variable

Exponential random variable occurs naturally if a random variable has memoryless property

$$\mathbb{P}\big(X > x + a | X > a\big) = \mathbb{P}(X > x). \tag{2.1}$$

Note that $\mathbb{P}\big(X > x + a | X > a\big)$ is the probability of $X > x + a$ under the condition $X > a$. If we interpret X as the waiting time, it shows the probability of waiting for at least another x minutes under the condition that we have already waited for a minutes. $\mathbb{P}(X > x)$ is the probability of waiting for at least x minutes. The above identity then shows that the probability of waiting for at least another x minutes remains the same no matter how long we have already waited. This is called the *memoryless of exponential random variable.* Exponential random variable are often used to model the waiting time of some events: photon arrival time, passenger arrival time



Definition 5 (Exponential Random Variable)**.** We say X is an exponential random variable of parameter λ if the PDF is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \ge 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

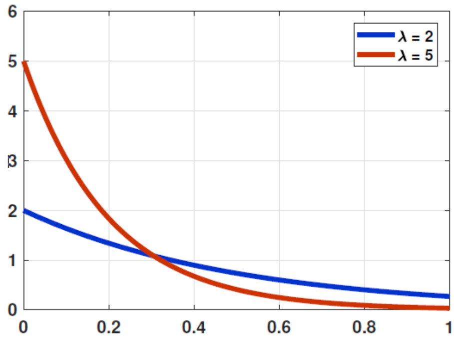
We write $X \sim \text{Exponential}(\lambda)$.

The CDF of an exponential random variable can be determined by

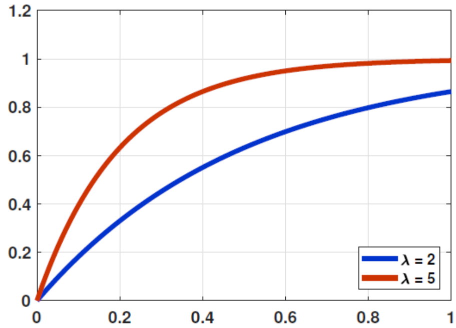
$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, \ x \ge 0 \tag{2.3}$$

Therefore, the CDF is

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{otherwise.} \end{cases}$$



(a) PDF



(b) CDF

Figure: The PDF and CDF of $X \sim \text{Exponential}(\lambda)$.

Remark 2 (Memoryless of Exponential Random Variable)**.** Now we show how the exponential random variable satisfies the memoryless property. Since $F_X(x) = 1 - e^{-\lambda x}$ for $x > 0$ we know

$$\mathbb{P}(X > x) = 1 - F_X(x) = e^{-\lambda x}.$$

It then follows the definition of conditional probability that

$$\begin{aligned}\mathbb{P}(X > x + a | X > a) &= \frac{\mathbb{P}(X > x + a, X > a)}{\mathbb{P}(X > a)} \\ &= \frac{\mathbb{P}(X > x + a)}{\mathbb{P}(X > a)} = \frac{e^{-\lambda(x+a)}}{e^{-\lambda a}} = e^{-\lambda x} = \mathbb{P}(X > x),\end{aligned}$$

where the second identity holds since the event $\{X > x + a, X > a\}$ is the same as $\{X > x + a\}$. This shows that $\mathbb{P}(X > x + a | X > a)$ only depends on x and not on a!

2.3 Gaussian Random Variable

The final random variable we consider is the Gaussian random variable. The definition is given below.

Definition 6 (Gaussian Random Variable)**.** We say X is a Gaussian random variable if the PDF is

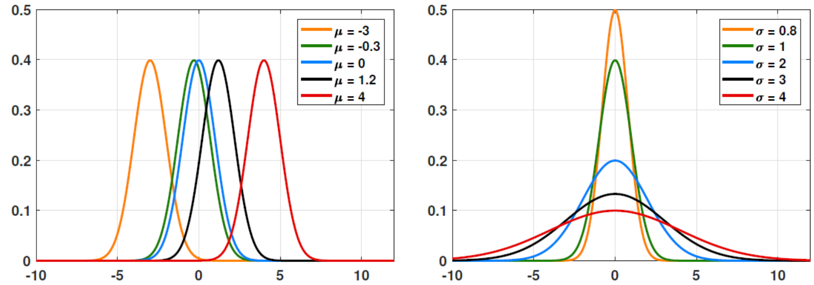
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{2.4}$$

where (μ, σ^2) are parameters of the distribution. We write

$$X \sim \text{Gaussian}(\mu, \sigma^2) \quad \text{or} \quad X \sim \mathcal{N}(\mu, \sigma^2).$$

Remark 3. The two parameters control the shape of Gaussian random variable. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

- it is symmetric around μ
- σ^2 determines how sharply the variable is around its center: if σ is small then the random variable is sharply concentrated around μ; if σ is large then the randomly variable becomes more flat.



μ changes, $\sigma = 1$ $\qquad$ $\mu = 0$, σ changes

Figure: A Gaussian random variable with different μ and σ

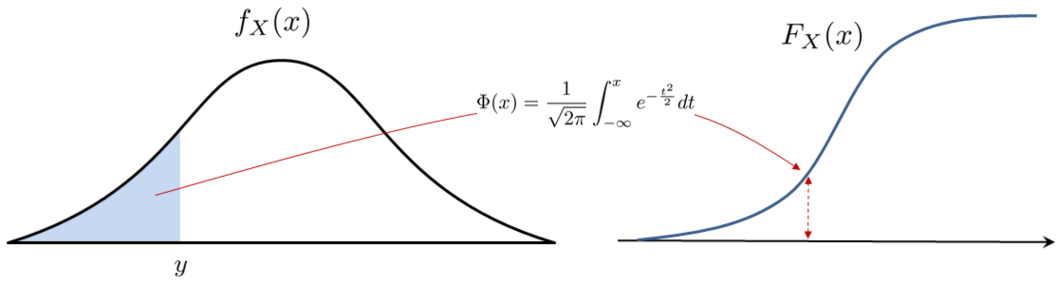
A Gaussian random variable of special interest is the standard Gaussian random variable.

Definition 7 (Standard Gaussian Random Variable)**.** We say X is a **standard** Gaussian random variable if the PDF is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}},$$

That is, $X \sim \mathcal{N}(0, 1)$ is Gaussian with $\mu = 0$ and $\sigma^2 = 1$. The CDF of the standard Gaussian is

$$\Phi(x) \stackrel{\text{def}}{=} F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$



Definition of the CDF of the standard Gaussian.

One can relate a Gaussian random variable to a standard Gaussian random variable after a linear transformation. That is, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then one can show that

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

A natural question is why we need to consider Gaussian random variable and how it comes. Naturally, Gaussian random variable appears if we consider a summation of independent random variables.

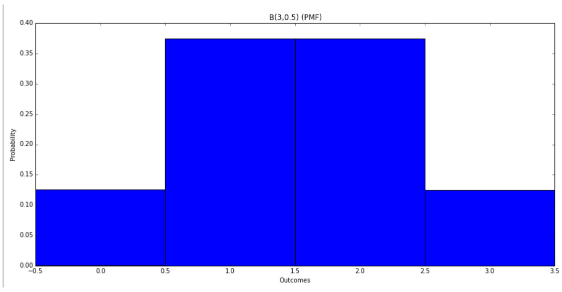
- When we sum many independent random variables, the resulting random variable is a Gaussian. That is, if $X_1, X_2, \ldots, X_n$ is a sequence of independent random variables, then

$$X_1 + X_2 + \cdots + X_n \to \mathcal{N}(\mu, \sigma^2) \quad \text{as } n \to \infty$$

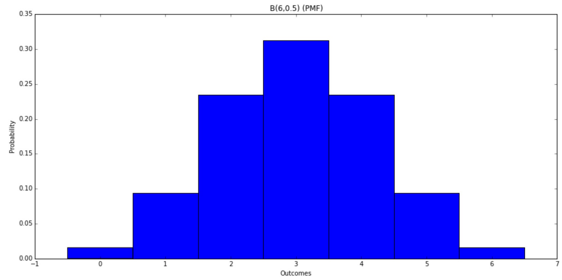
  for some (μ, σ^2).

- This is known as the Central Limit Theorem. The theorem applies to *any* random variable: if holds no matter X_i are discrete or continuous. The only thing it requires is that X_i's are independent. This shows the speciality of Gaussian random variables: we naturally get Gaussian random variables when we consider a summation of independent random variables.

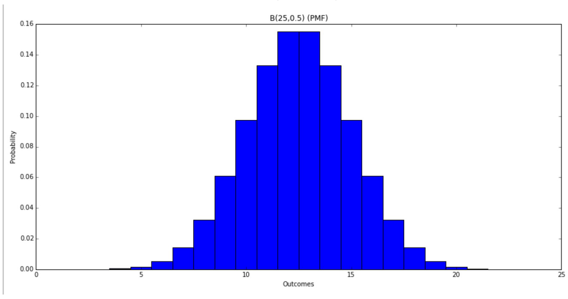
Example 5 (Gaussian Distribution as Limit of Binomial)**.** When we observe the characteristic shape of the Binomial Distribution $B(n, 0.5)$ as n approaches Infinity, we see something interesting



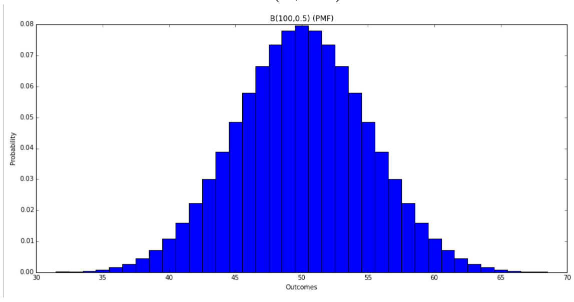
B(3,0.5)



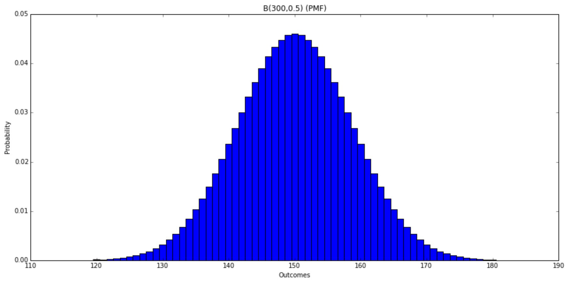
B(6,0.5)



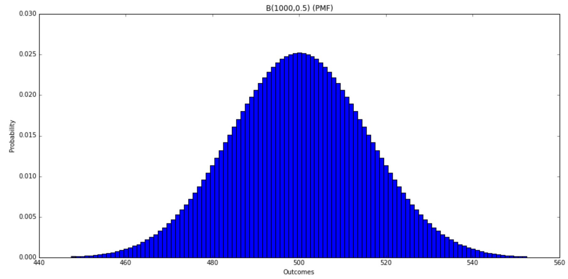
B(25,0.5)



B(100,0.5)



B(300,0.5)



B(1000,0.5)

Recall that Binomial Distribution $B(n, 0.5)$ is a summation of Bernoulli random variables Bernoulli(0.5). Therefore, the above figures show that the summation of n independent Bernoulli random variables converge to a Gaussian random variable as n goes to infinity. This verifies empirically the Central Limit Theorem.