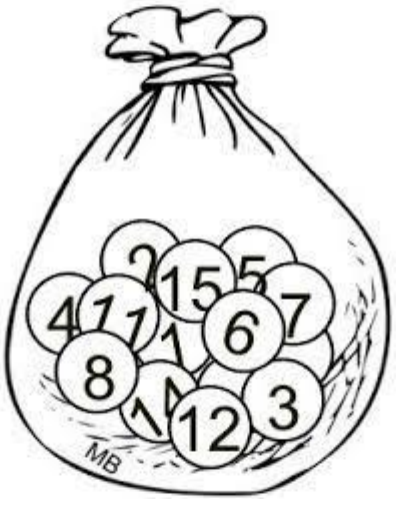
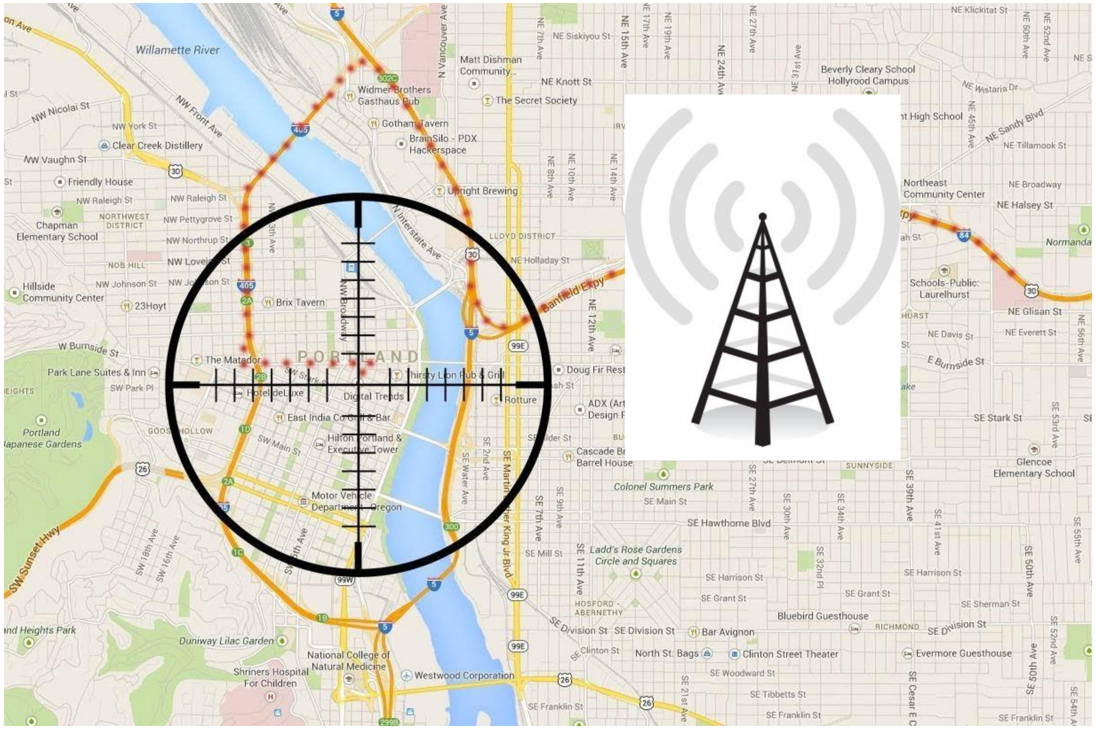
Lecture 05

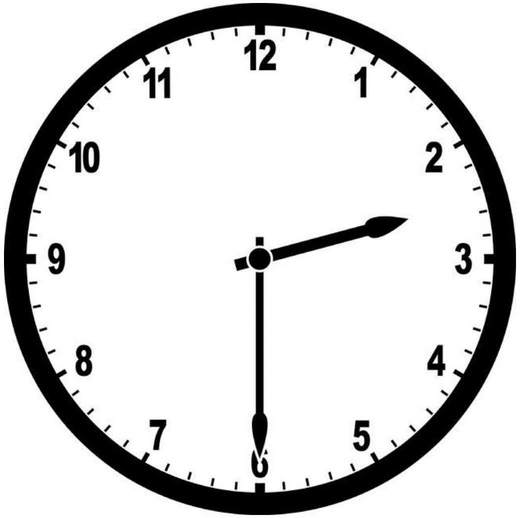
- Continuous Random Variables
- Probability Density Function (PDF)
- Cumulative Distribution Functions (CDF)
- Some Common PDFs
  (Uniform, Exponential, Normal)



Not all values are discrete



Distance



Time

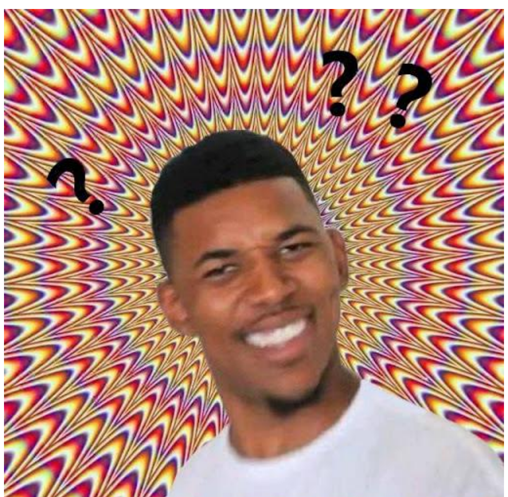
Definition of Continuous R.V.

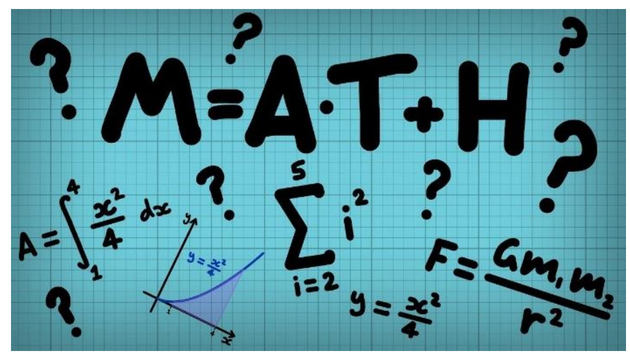
The **old-school** definition:

A random variable X is continuous if there is a probability density function (PDF) $f(x) \geq 0$ such that for $-\infty < x < \infty$:

$$P(a \leq X \leq b) = \int_a^b f(x)dx\,.$$

Not easy to understand!





Let's have a nice introduction.

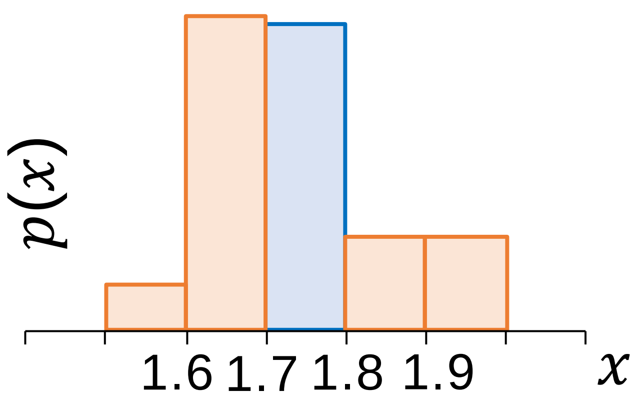
Heights of Students

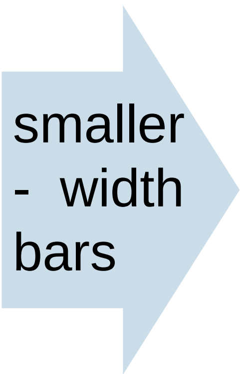
You are volunteering at SZU Opening day.

- To choose a t-shirt for your friends and yourself, you need to know their heights.

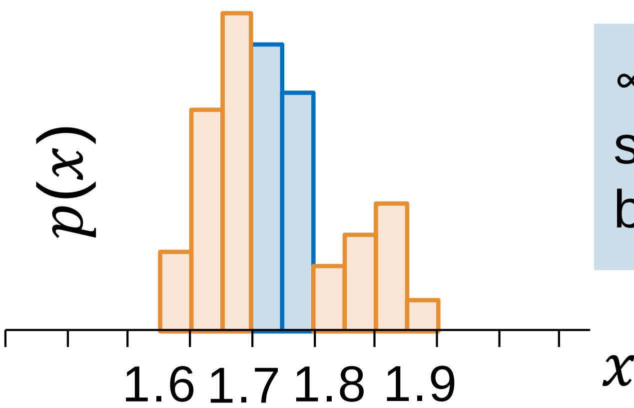
1) What is the probability that Tom is 1.75 m height? (not meaningful)
2) What is the probability that Tom is 1.70 to 1.80 m height?

$P(1.70 < X \leq 1.80)$





$P(1.70 < X \leq 1.75)+$
$P(1.75 < X \leq 1.80)$



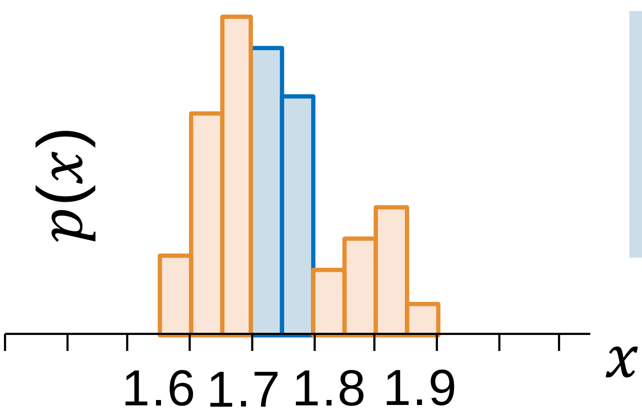
Heights of Students

You are volunteering at SZU Opening day.

- To choose a t-shirt for your friends and yourself, you need to know their heights.

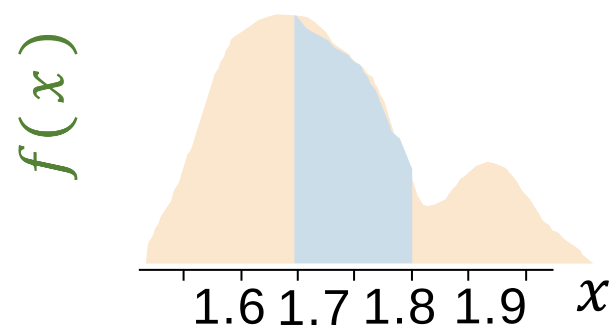
1) What is the probability that Tom is 1.75 m height? (not meaningful)
2) What is the probability that Tom is 1.70 to 1.80 m height?

$P(1.70 < X \le 1.75)+$
$P(1.75 < X \le 1.80)$



∞-ly small bars

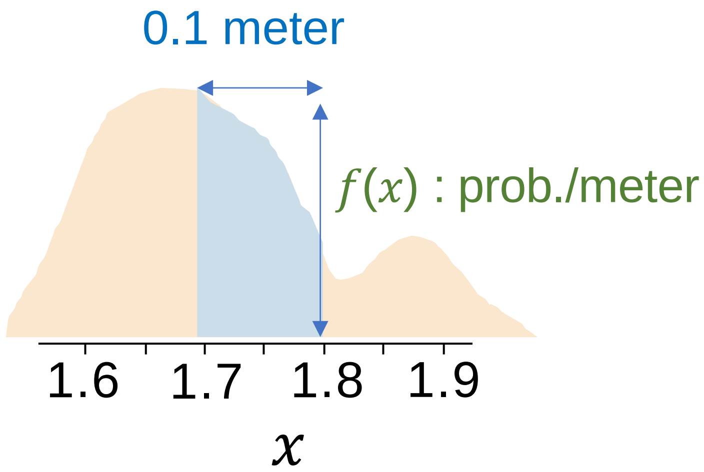
$P(1.70 < X \le 1.80)$



probability density function, PDF (概率密度函数)

From PDF to Probability

Integrate $f(x)$ to get probabilities.



PDF Units: probability per units of X

$$P(1.7 \leq X \leq 1.8) = \int_{1.7}^{1.8} f(x)dx$$

A random variable X is continuous if there is a probability density function (PDF) $f(x) \geq 0$ such that for $-\infty < x < \infty$:

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx\,.$$

Understanding PDF

- $m = \int_v \rho \, dv$

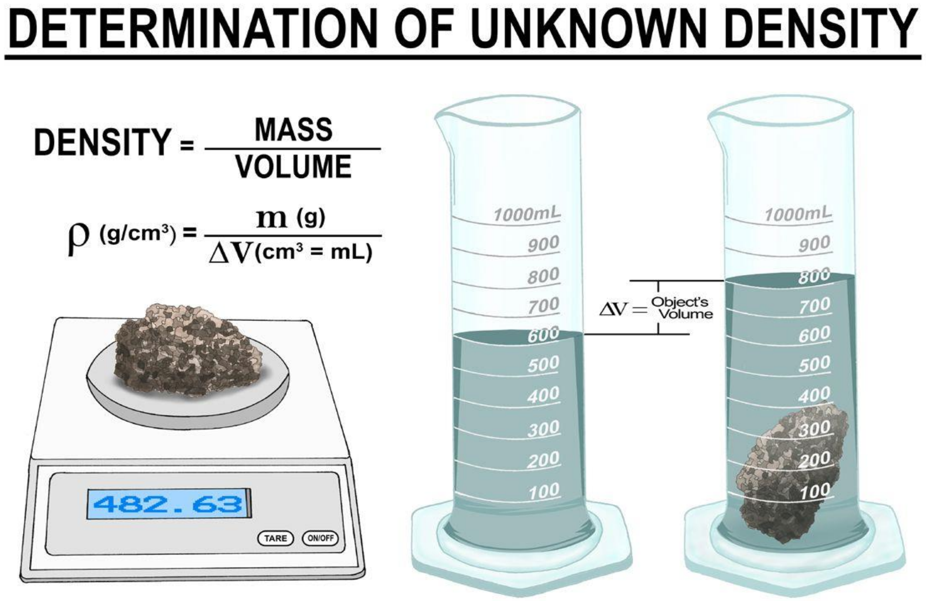
Mass = integral of density

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

Probability = integral of probability density.

- Density vs. Probability density

- Mass vs. Probability



How to calculate mass and density in a physics experiment.

PMF vs. PDF

Discrete random variable X

离散型随机变量概率分布律
Probability Mass Function (PMF):

$$p(x)$$

To get probability:

$$P(X = x) = P(x)$$

Continuous random variable X

连续型随机变量概率密度函数
Probability Density Function (PDF):

$$f(x)$$

To get probability:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Both are measures of how likely X is to take on a **value**.

CDF for continuous RVs

For a <u>continuous</u> random variable X with PDF $f(x)$, the CDF of X is

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

$$\text{PDF} \xRightarrow{\text{Integral}} \text{CDF, CDF} \xRightarrow{\text{Derivative}} \text{PDF,}$$

Matching:

1. $P(X < a)$	A. $F(a)$
2. $P(X > a)$	B. $1 - F(a)$
3. $P(X \geq a)$	C. $F(a) - F(b)$
4. $P(a \leq X \leq b)$	D. $F(b) - F(a)$

Notes on PDF and CDF

(1) The CDF $F(x)$ of a continuous R.V. is also a continuous function.

(2) The probability of a continuous R.V. taking a specific value is 0, $P\{X = a\} = 0$. This is due to

$$
\begin{aligned}
P\{X = a\} &= \lim_{\sigma \to 0+} P\{a - \sigma < X \leq a\} \\
&= \lim_{\sigma \to 0+} [F(a) - F(a - \sigma)] = 0
\end{aligned}
$$

Therefore, for a continuous R.V. X

$$P\{a < x < b\} = P\{a < x \leq b\} = P\{a \leq x < b\} = P\{a \leq x \leq b\}$$

Notes on PDF and CDF

(3) Changing the values of $f(x)$ at some finite points does not affect the value of $F(x)$, so $f(x)$ is not unique.

(4) A is an impossible event $\Rightarrow P(A) = 0$;

$P(A) = 0 \nRightarrow A$ is an impossible event.

(5) The PDF must satisfy

$$P(a < x < b) \geq 0$$

$$\int_{-\infty}^{\infty} f(x)dx = P(-\infty < X < \infty) = 1$$

Ex. Given a CDF of the continuous R.V. X as

$$F(x) = \begin{cases} 0, & x < 0 \\ \sin x, & 0 \le x < \pi/2 \\ 1, & \pi/2 \le x \end{cases}$$

Find the PDF $f(x)$ and $P\{\pi/4 \le X \le 2\}$.

Solve: $f(x) = F'(x) = \begin{cases} \cos x\,, & 0 \le x < \pi/2 \\ 0, & \text{otherwise} \end{cases}$

$$P\left\{\frac{\pi}{4} \le X < \frac{\pi}{2}\right\} = F\left(\frac{\pi}{2}\right) - F\left(\frac{\pi}{4}\right) = 1 - \sin \pi/4 = 1 - \sqrt{2}/2$$

or $$= \int_{\pi/4}^{2} f(x)dx = \int_{\pi/4}^{\pi/2} \cos x\, dx = 1 - \sqrt{2}/2$$

Ex. Let the PDF of a continuous R.V. x be

$$f(x) = \begin{cases} kx^2, & 0 \le x < 2 \\ kx, & 2 \le x \le 3, \\ 0, & \text{otherwise} \end{cases}$$

Find (1) the value of k; (2) $F(x)$; (3) $P\{1 < X < 5/2\}$.

(1) $$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_0^2 kx^2 dx + \int_2^3 kxdx = \frac{31}{6}k \quad \Rightarrow \quad k = \frac{6}{31}$$

(2) $$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} \int_{-\infty}^{x} 0dt = 0 & x < 0 \\ \int_0^x kt^2 dt = \frac{x^3}{3}k & 0 \le x < 2 \\ \int_0^2 kt^2 dt + \int_2^x ktdt = \frac{4 + 3x^2}{6}k & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

Ex. Let the PDF of a continuous R.V. x be

$$f(x) = \begin{cases} kx^2, & 0 \le x < 2 \\ kx, & 2 \le x \le 3, \\ 0, & \text{otherwise} \end{cases}$$

Find (1) the value of k; (2) $F(x)$; (3) $P\{1 < X < 5/2\}$.

(3)

$$P\left\{1 < X < \frac{5}{2}\right\} = P\left\{1 < X \le \frac{5}{2}\right\} = P\left\{x \le \frac{5}{2}\right\} - P\{x \le 1\}$$
$$= \int_0^2 kt^2 dt + \int_2^{5/2} ktdt - \int_0^1 kt^2 dt$$
$$= k\left[\int_1^2 t^2 dt + \int_2^{5/2} tdt\right] = \frac{83}{124}$$

Ex. Given the PDF of R.V. x as

$$f(x) = \begin{cases} \dfrac{A}{\sqrt{1-x^2}} \quad, & |x| < 1 \\ 0 \quad, & |x| \geq 1 \end{cases}$$

find (1) the value of A, (2) $P\{-\frac{1}{2} < X < \frac{1}{2}\}$, (3) $F(x)$.

Sol.

(1) $A = 1/\pi$, hint： $\int \frac{1}{\sqrt{1-x^2}} dx = \sin^{-1} x$

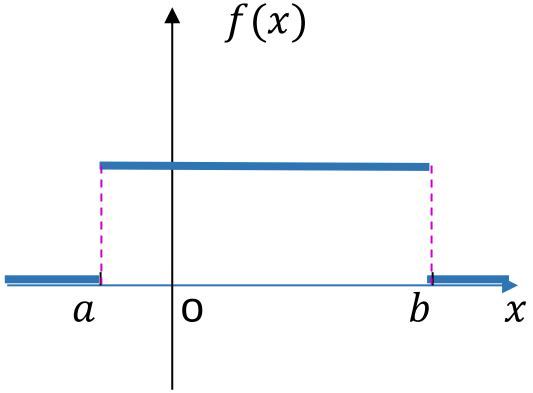
(2) $P\{-\frac{1}{2} < X < \frac{1}{2}\} = \frac{1}{3}$

(3) $F(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2} + \frac{1}{\pi} \arcsin x\,, & -1 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$

Some common PDFs

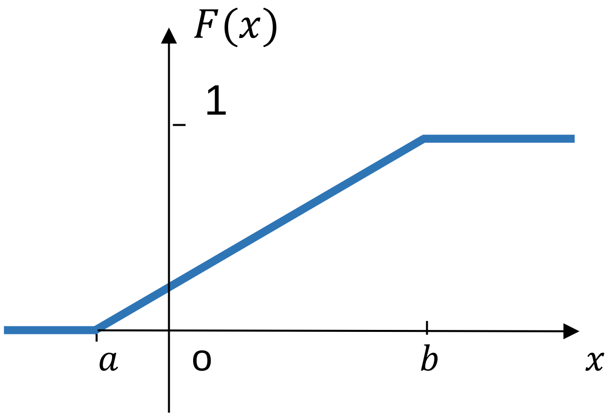
Uniform distribution (均匀分布), $X \sim U(a, b)$

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



The CDF of uniform distribution is

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b \end{cases}$$



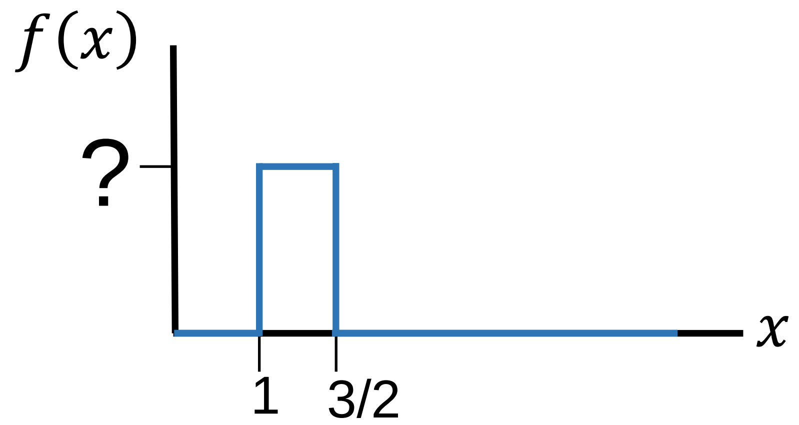
Quick check

$$X \sim U(a,b),\ f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

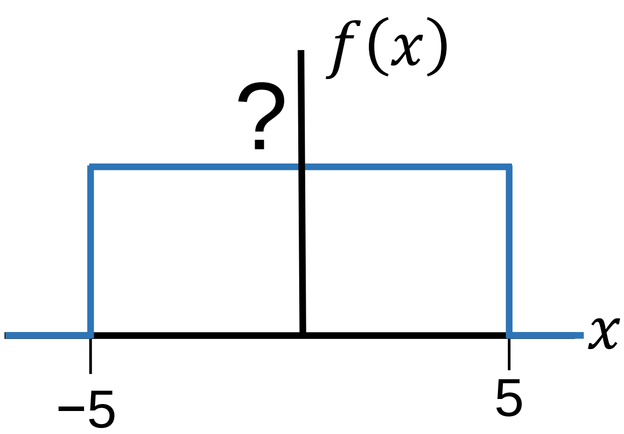
Complete the following graphs

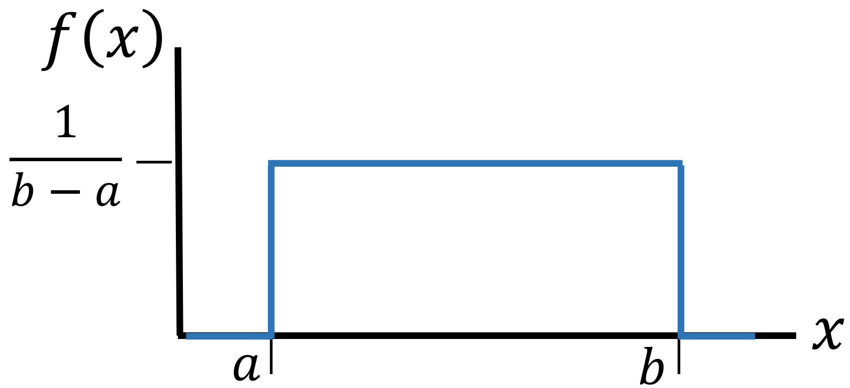
1.

2.



3.





Ex. 3c (P199) Buses arrive at a specified stop at 15-minute intervals starting at 7 a.m. That is, they arrive at 7, 7:15, 7:30, 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability that he waits

(a) less than 5 minutes for a bus;

(b) more than 10 minutes for a bus.

Sol.

(a) $P\{10 < X < 15\} + P\{25 < X < 30\}$

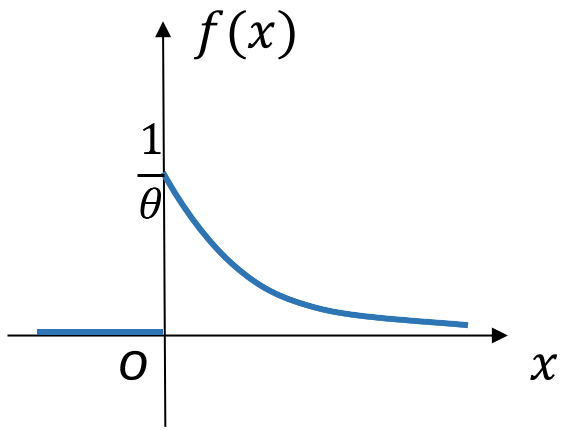
$$= \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = 1/3$$

(b) $P\{0 < X < 5\} + P\{15 < X < 20\}$

$$= \int_{0}^{5} \frac{1}{30} dx + \int_{15}^{20} \frac{1}{30} dx = 1/3$$

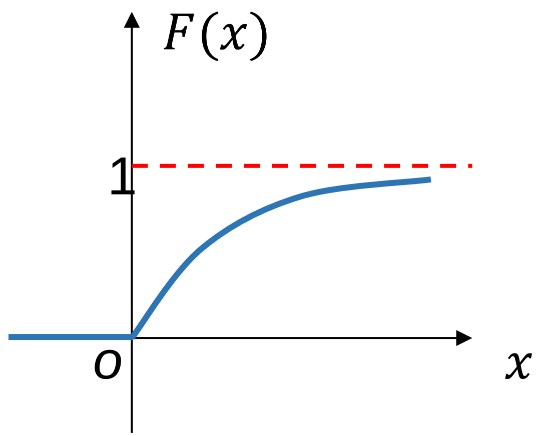
Some common PDFs

Exponential distribution (指数分布), $X \sim \exp(\theta)$



$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$, with $\theta > 0$

- Model the time interval between (Poisson) events.
- The meaning of parameter θ?
- The CDF of uniform distribution is



$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

- Memoryless property:

$\forall s, t > 0, P\{X > s + t | X > s\} = P\{X > t\}$

Example of Exp. distribution

Examples:

- Time until next earthquake
- Time for request to reach web server
- Time until end of phone call

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$$
$$F(X) = 1 - e^{-\frac{x}{\theta}}$$
$$x > 0$$

Ex. 5b (P212) Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\theta = 10$. If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait

(a) More than 10 minutes; $P\{X > 10\} = 1 - F(10)$

(b) Between 10 and 20 minutes. $P\{10 < X < 20\} = F(20) - F(10)$

Modelling earthquakes

$$F(X) = 1 - e^{-\frac{x}{\theta}}$$
$$x > 0$$

Suppose major earthquakes (8.0+) occur on average once every 500 years, what is the probability of a major earthquake in next 30 years?

Define events & state goal

X: when next earthquake happens

$X \sim \exp(\theta)$, $\theta = 500$

Want: $P(X < 30)$

Solve:

$$P(X < 30) = 1 - e^{-\frac{30}{500}}$$

2011 Tōhoku earthquake and tsunami

March 11, 2011

The 2011 Tōhoku earthquake and tsunami occurred at 14:46 JST on 11 March. The magnitude 9.0–9.1 undersea megathrust earthquake had an epicenter in the Pacific Ocean, 72 km east of the Oshika Peninsula of the Tōhoku region, and lasted approximately six minutes, causing a tsunami. Wikipedia

Exp. Distribution vs. Poisson Distribution

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0,1,2, \dots$$

Describe the time interval between two occurrences.

$\theta = 1/\lambda$

Describe # occurrences in a unit time interval.

For continuous R.V.

Written as $X \sim \exp(\theta)$,

is a Probability Density Function (PDF).

For discrete R.V.

Written as $X \sim \pi(\lambda)$,

is a Probability Mass Function (PMF).

Exp. Distribution vs. Poisson Distribution

Ex. At a supermarket checkout counter, an average of 2 customers checkout every minute.

1) Calculate the probability of no customer checkout in 1 minute by Poisson distribution and exponential distribution, respectively.

Solve:

By Poisson distribution:

$$\lambda = 2 \;\; \Rightarrow \;\; P(k=0) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{1}{e^2}$$

By Exponential distribution:

$$\theta = \frac{1}{2} \;\; \Rightarrow \;\; F(x>1) = \int_1^{+\infty} \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = -\,e^{-2t}\Big|_1^{\infty} = \frac{1}{e^2}$$

Exp. and Poisson are equivalent in unit time probability calculation with $\lambda = 1/\theta$.

Exp. Distribution vs. Poisson Distribution

Ex. At a supermarket checkout counter, an average of 2 customers checkout every minute.

2) Calculate the probability of no customer checkout in 2 minutes by Poisson distribution and exponential distribution, respectively.

Solve:

By Poisson distribution:

$$\lambda = 4 \;\; \Rightarrow \;\; P(k=0) = \frac{\lambda^k e^{-\lambda}}{k!} = \frac{1}{e^4}$$

By Exponential distribution:

$$\theta = \frac{1}{2} \;\; \Rightarrow \;\; F(x>2) = \int_2^{+\infty} \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = -e^{-2t}\Big|_2^{\infty} = \frac{1}{e^4}$$

$1/\theta$ in Exp. and λ in Poisson are not equivalent anymore, $\lambda \neq 1/\theta$.

Some common PDFs

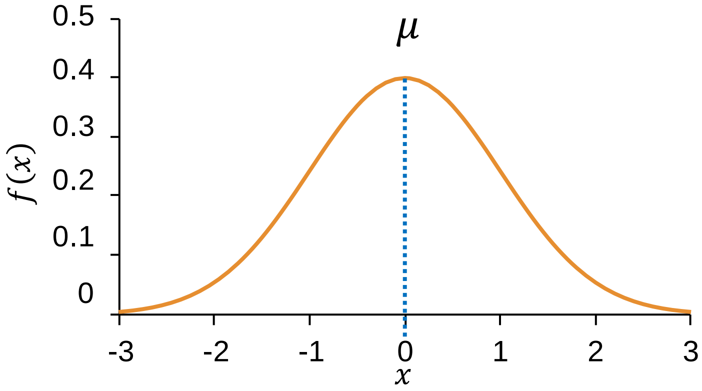
- Normal (Gaussian) distribution (正态分布), $X \sim \mathcal{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

where μ and σ^2 are mean and variance of x.

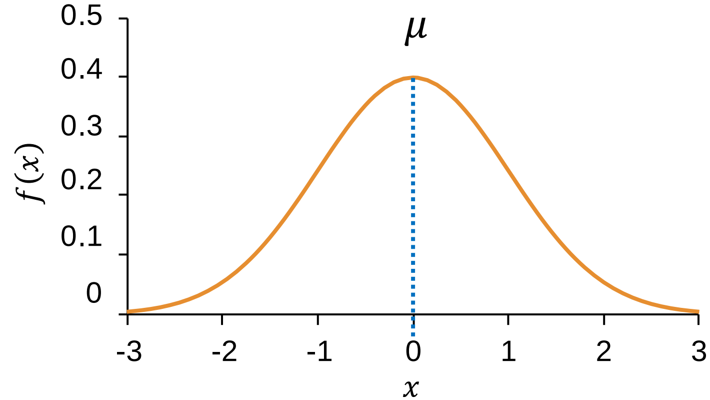
Why it is important?

- Common for natural phenomena (due to central limit theorem);



Anatomy of a beautiful equation

Let $X \sim \mathcal{N}(\mu, \sigma^2)$



The PDF of X is defined as:

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

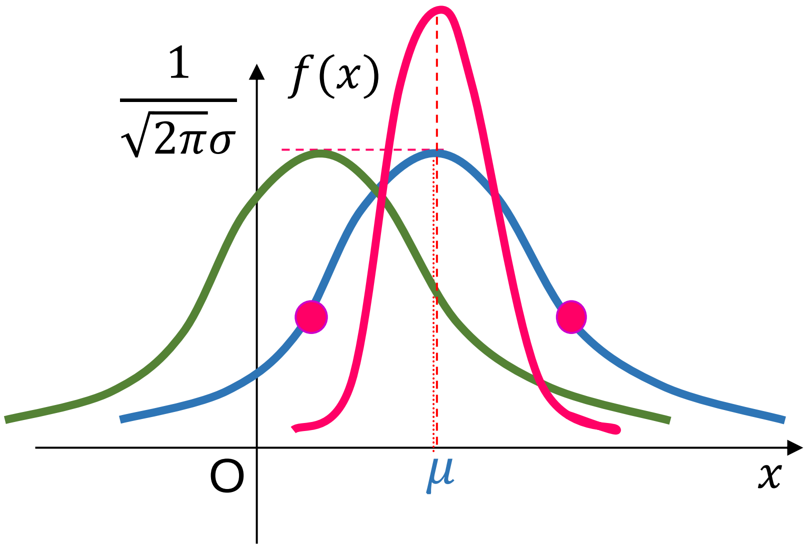
symmetric around μ

Normalization constant

exponential tail

variance σ^2 manages spread

Property of Normal PDF



- The curve is symmetric by $x = \mu$.

$$P\{\mu - a < X \leq \mu\} = P\{\mu < X < \mu + a\}$$

- $f(x)$ reaches maximum at $x = \mu$.

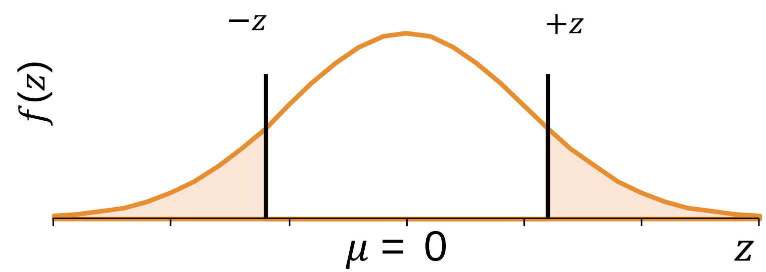
- Turning points at $x = \mu \pm \sigma$, approaches x axis asymptotically.

- Fix σ, varies μ: the curve moves along x axis.

- Fix μ, varies σ: a smaller $\sigma \Rightarrow$ a sharper bell-shape curve, a larger probability around μ.

Symmetry of the Normal R.V.

Let $Z \sim \mathcal{N}(0,1)$ with CDF $P(Z \leq z) = F(z)$.



Suppose we only knew numeric values for $F(z)$ and $F(y)$, for some z, $y \geq 0$.

How do we compute the following probabilities?

1. $P(Z \leq z)$
2. $P(Z < z)$
3. $P(Z \geq z)$
4. $P(Z \leq -z)$
5. $P(Z \geq -z)$
6. $P(y < Z < z)$

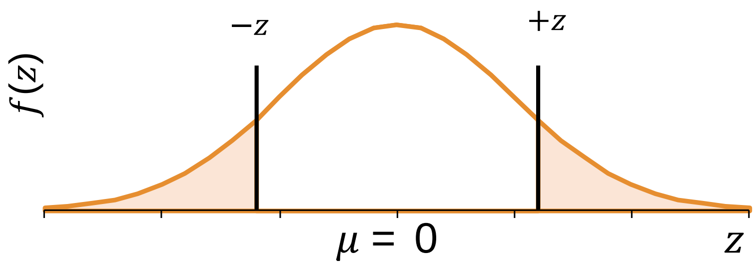
A. $F(z)$

B. $1 - F(z)$

C. $F(z) - F(y)$

Symmetry of the Normal R.V.

Let $Z \sim \mathcal{N}(0,1)$ with CDF $P(Z \leq z) = F(z)$.



Suppose we only knew numeric values for $F(z)$ and $F(y)$, for some z, $y \geq 0$.

How do we compute the following probabilities?

1. $P(Z \leq z) \quad = F(z)$
2. $P(Z < z) \quad = F(z)$
3. $P(Z \geq z) \quad = 1 - F(z)$
4. $P(Z \leq -z) \quad = 1 - F(z)$
5. $P(Z \geq -z) \quad = F(z)$
6. $P(y < Z < z) \quad = F(z) - F(y)$

A. $F(z)$

B. $1 - F(z)$

C. $F(z) - F(y)$

Symmetry is particularly useful when computing probabilities of zero-mean Normal RVs.

Linear transformation of Normal R.V.s

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with CDF $P(X \le x) = F(x)$. Linear transformations of X are also Normal.

If $Y = aX + b$, then $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

$$F_Y(y) = P\{Y \le y\}$$
$$= P\{aX + b \le y\}$$
$$= P\left\{X \le \frac{y-b}{a}\right\}$$
$$= F_X\left(\frac{y-b}{a}\right)$$

Differentiation

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$
$$= \frac{1}{\sqrt{2\pi}a\sigma} \exp\left\{-\left(\frac{y-b}{a} - \mu\right)^2 / 2\sigma^2\right\}$$
$$= \frac{1}{\sqrt{2\pi}a\sigma} \exp\left\{-\frac{(y-b-a\mu)^2}{2(a\sigma)^2}\right\}$$

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

Consider some specific values of a and b