Computational Physics

Topic 01 — Computational Problems Involving Probability

Lecture 20 — Continuous Random Variables

Dr Kieran Murphy

Department of Computing and Mathematics,
SETU (Waterford).
(kieran.murphy@setu.ie)

Autumn Semester, 2025/26

Outline

- Continuous random variables
- Uniform, Exponential, Normal

Continuous Random Variable

A discrete random variable can only take on a discrete set of values (e.g., 0,1,2, etc.). We now want to model experiments where the outcome can take any value in an interval. For example.

- The time that elapses between the installation of a new component and its failure.
- The percentage of impurity in a batch of chemicals.
- The value of a bitcoin token (1061.47€ as of 3/4/17)

While we could approximate the output using discrete random variables it is easier* and computationaly faster to use continuous random variables.

Definition 1

A random variable is a continuous random variable if it can take any value in an interval.

*There are a few technical issues to resolve first (next few slides)

Moving from Discrete to Continuous I

Consider the experiment of reading the time on a 24-hour digital clock. Assuming that all times are equally likely, then what is the probability that the clock reads, say 3am?

- This is apparently an easy question, and "most" people would say

$$\Pr(3\text{am}) = 1/\left(24 \times 60\right) = 0.0006944444$$

  since there are $24 * 60$ minutes in a 24-hour period
- But this assumes the clock display accuracy is 1 minute. What happens if it displays seconds? Then

$$\Pr(3\text{am}) = 1/\left(24 \times 60 \times 60\right) = 0.00001157407$$

- Or displays hundredths of a second?

$$\Pr(3\text{am}) = 1/\left(24 \times 60 \times 60 \times 100\right) = 0.0000001157407$$

- Or display mili-seconds, … or pico-seconds, or …

> The more accurate the measurement the smaller the probability of reading any particular value.

Moving from Discrete to Continuous II

In the limit with a clock with perfect display accuracy (where the time shown smoothy changes) we have

$$\Pr(3\text{am}) = 0$$

- To date we have taken that an event has probability zero if it cannot occur.
- But clearly at some point (exactly once a day) the clock will read "3am". So we seem to have a contradiction (when dealing with continuous random variables).

This contradiction can be resolved using Measure theory, but for our purposes we can:

- Interpret $Pr(X = x) = 0$, the probability of any particular value, as "arbitrary small" ($\neq$ "impossible").
- Always work with cumulative probabilities, i.e., $\Pr(X \leq x)$.

Probability Density Function (PDF)

For a discrete random variable we had $\Pr(X = x) = f(x)$ where $f(x)$ was the probability mass function (PMF). For a continuous random variable:

- $\Pr(X = x) = 0$
- The probability mass function (PMF) is replaced by its continuous analogue, the probability density function (PDF)

Definition 2 (Probability Density Function (PDF))

A probability density function (PDF) is any function, $f(x)$, with properties:

1. output is non-negative
$$0 \leq f(x) \tag{1}$$
2. Area under curve is one
$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \tag{2}$$

- Note we have lost the condition $f(x) < 1$ so here $f(x)$ is NOT a probability.

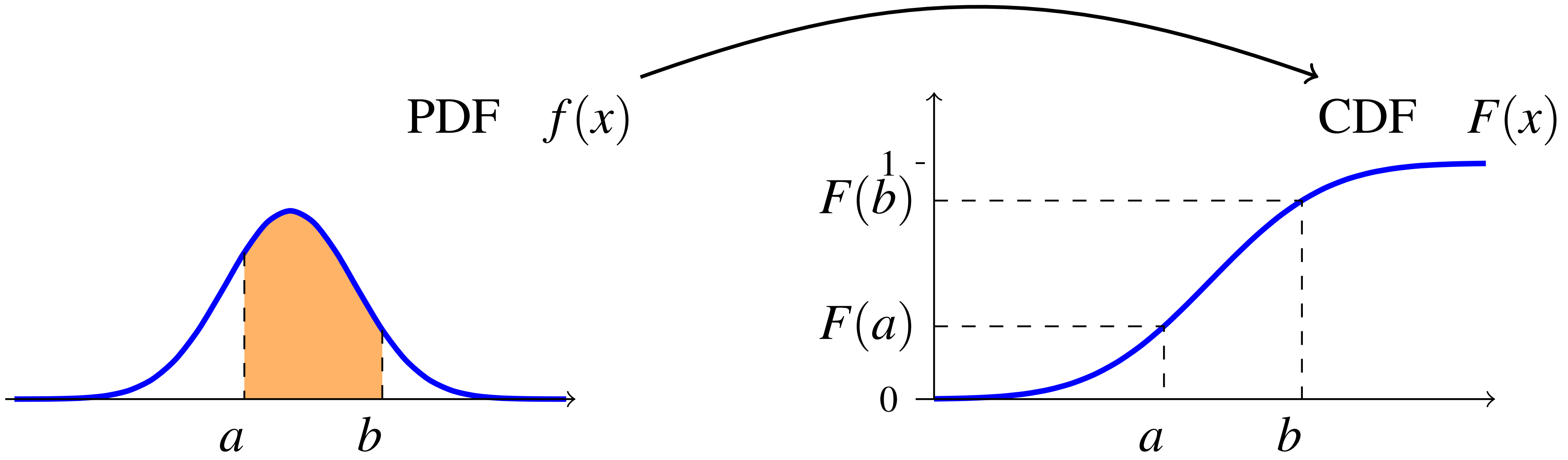
Cumulative Distribution Function (CDF) I

While the value of a probability density function, $f(x)$, does NOT represent probabilities, the area under $f(x)$ over an interval $a \leq x \leq b$ represents the probability

$$\Pr(a \leq X \leq b)$$

So we have

$$\Pr(X \leq x) = F(x) = \int_{-\infty}^{x} f(s)\, ds$$



$$\Pr(a \leq X \leq b) = \Pr(X \leq b) - \Pr(X \leq a) = F(b) - F(a)$$

Cumulative Distribution Function (CDF) II

Definition 3 (Cumulative Distribution Function (CDF))

Let X be a continuous random variable with a probability density function (PDF), $f(x)$ then the cumulative distribution function (CDF) is defined by

$$F(x) = \int_{-\infty}^{x} f(s)\, ds$$

and has properties:

- $\Pr(X \leq x) = F(x)$
- $\Pr(a \leq X \leq b) = F(b) - F(a)$ for any values a and b with $a < b$.
- $F(x) \to 0$ as $x \to -\infty$
- $F(x) \to 1$ as $x \to +\infty$

With continuous random variables we find probabilities for intervals of the random variable, not singular specific values.

Mean and Variance of Continuous Random Variables

The mean and variance of continuous random variables can be computed similar to those for discrete random variables, but for continuous random variables, we will be integrating over the domain of X rather than summing over the possible values of X.

Definition 4

Suppose X is a continuous random variable with probability density function $f(x)$. The expected value or mean of X, denoted is

$$\mathrm{E}[X] = \mu = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

and variance is

$$\mathrm{Var}[X] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\, dx$$

And we have $\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - \left(\mathrm{E}[X]\right)^2$

Example 5 I

Example 5

The lifetime, in years, of some electronic component is a continuous random variable with probability density function

$$f(x) = \begin{cases} \frac{k}{x^4} & \text{for } x \geq 1 \\ 0 & \text{for } x < 1 \end{cases}$$

Determine:

(a) the value of parameter, k.

(b) the cumulative distribution function,

(c) the probability for the lifetime to exceed 2 years.

(d) the expected value of X

(e) the variance of X

Example 5 I

(a), (b), (c)

- (d) The expected value is

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_{1}^{\infty} \frac{3}{x^3}\, dx = \left[-\frac{3}{2x^2}\right]_1^{\infty} = \frac{3}{2}$$

- (e) Using

$$\mathrm{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx = \int_{1}^{\infty} \frac{3}{x^2}\, dx = \left[-\frac{3}{x}\right]_1^{\infty} = 3$$

then the variance is

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - \left(\mathrm{E}[X]\right)^2 = 3 - (3/2)^2 = 3/4$$

The Uniform Distribution

Definition 6 (Uniform Distribution)

The uniform distribution, is a **two-parameter** continuous probability distribution defined by the probability density function (PDF)

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

where parameters

- a minimum value
- b maximum value

- The expected value and variance of a uniform random variable are

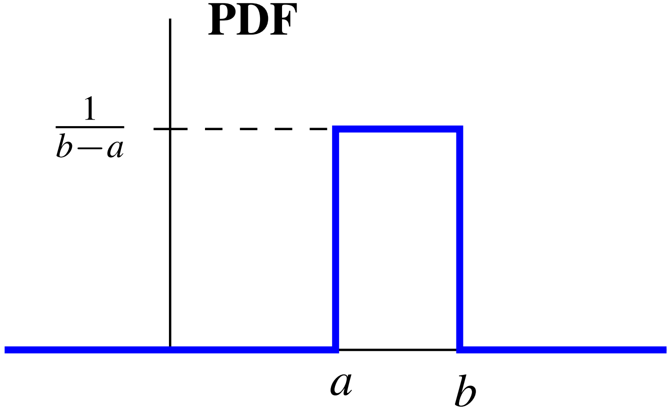
$$\mathrm{E}[X] = \mu = \frac{a+b}{2} \qquad \mathrm{Var}[X] = \sigma^2 = \frac{(b-a)^2}{12}$$

The Uniform Distribution — CDF

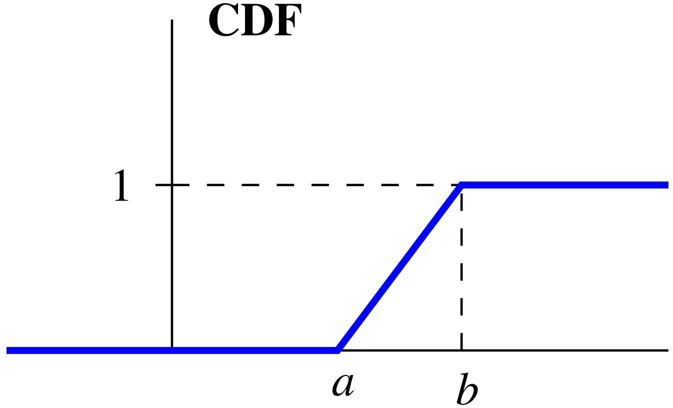
$$F(x) = \int_{-\infty}^{x} f(s)ds$$

The cumulative distribution function of the uniform distribution on (a, b) is

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \quad \Longrightarrow \quad F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & b < x \end{cases}$$



$\Longrightarrow$



$\boxed{f(x) \text{ is not a probability}}$

$$\Pr(X = x) = 0$$

$$\Pr(X = x) \neq f(x)$$

since X is a continuous random variable

$F(x)$ is a cumulative probability

$$\Pr(X \leq x) = F(x)$$

$$\Pr(a < X \leq b) = F(b) - F(a)$$

Example 7

Example 7

If X is a continuous random variable having the probability density function

$$f(x) = \begin{cases} 0.25 & 0 \leq x < 4 \\ 0 & \text{otherwise} \end{cases}$$

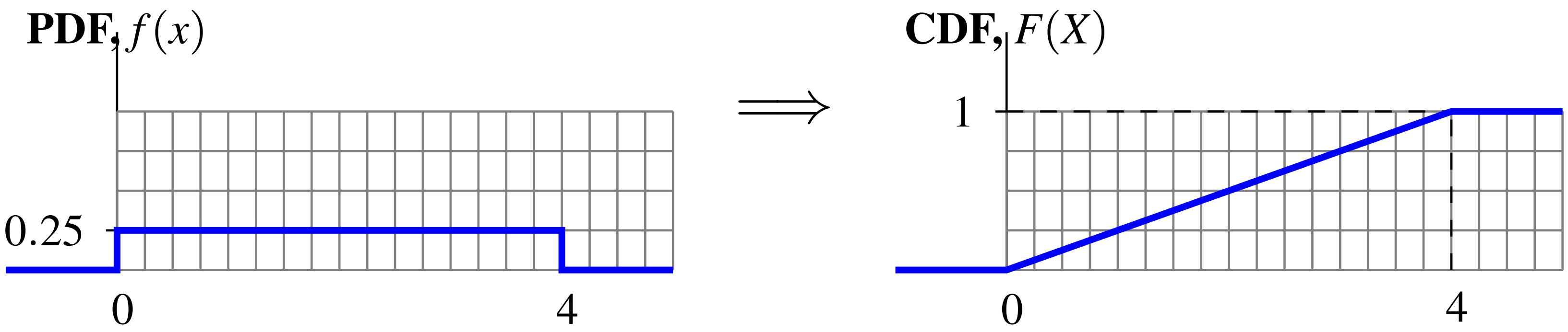
(a) Graph the probability density function,
(b) Determine and graph the cumulative distribution function.
(c) Calculate $\Pr(X < 2)$.
(d) Calculate $\Pr(1 < X < 3)$.
(e) Calculate the mean and variance of X.

Note: that this is just the uniform probability with parameters $a = 0$ and $b = 4$.

Example 7

(a,b)

$$f(x) = \begin{cases} 0.25 & 0 \le x < 4 \\ 0 & \text{otherwise} \end{cases} \implies F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{4} & 0 \le x < 4 \\ 1 & 4 < x \end{cases}$$



- (c) $\Pr(X < 2) = F(2) = 0.5$
- (d) $\Pr(1 < X < 30) = F(30) - F(1) = 1 - \frac{1}{4} = 0.75$
- (e) $\mathrm{E}[X] = \frac{0+4}{2} = 2, \qquad \mathrm{Var}[X] = \frac{(4-0)^2}{12} = 0.75$

The Exponential Distribution

The Poisson distribution was concerned with the number of (rare) events per unit time interval. The exponential distribution, is concerned with the length of time between events.

Definition 8 (Exponential Distribution)

The exponential distribution, is a **one-parameter** continuous probability distribution defined by the probability density function (PDF)

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where parameter

- λ is the mean number of independent events per unit time. $\lambda > 0$

- The expected value and variance of a exponential random variable are

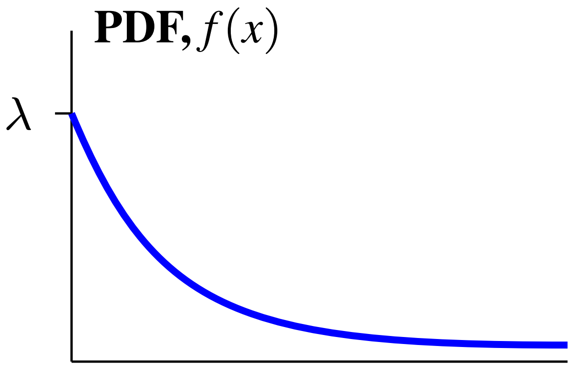
$$\mathrm{E}[X] = \mu = \frac{1}{\lambda} \qquad \mathrm{Var}[X] = \sigma^2 = \frac{1}{\lambda^2}$$

Exponential Distribution — CDF

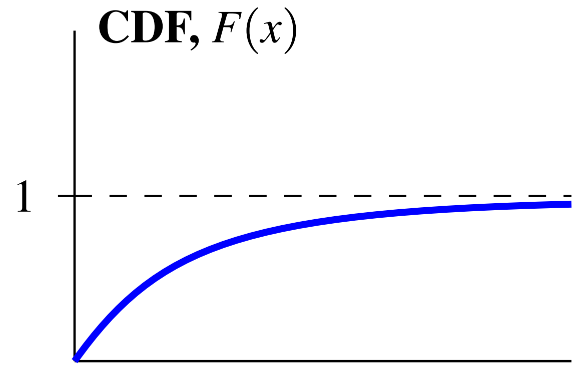
$$F(x) = \int_{-\infty}^{x} f(s)ds$$

The cumulative distribution function of the exponential distribution is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & 0 \leq x \\ 0 & \text{otherwise} \end{cases} \implies F(x) = \begin{cases} 1 - e^{-\lambda x} & 0 \leq x \\ 0 & \text{otherwise} \end{cases}$$



$\implies$



$\boxed{f(x) \text{ is not a probability}}$

$$\Pr(X = x) = 0$$

$$\Pr(X = x) \neq f(x)$$

since X is a continuous random variable

$F(x)$ is a cumulative probability

$$\Pr(X \leq x) = F(x)$$

$$\Pr(a < X \leq b) = F(b) - F(a)$$

Example 9

Example 9

Service times for customers at a library information desk can be modelled by an exponential distribution with a mean service time of 5 minutes. What is the probability that a customer service time will

(a) take longer than 10 minutes?

(b) take between 5 and 15 minutes?

MODEL Exponential, $\lambda = 1/5 = 0.2$ since $\mu = 1/\lambda$

$\implies$ CDF $F(x) = 1 - \mathrm{e}^{-0.2x}$

(a) *... take longer than 10 minutes*[†]

$$\Pr(X \geq 10) = 1 - \Pr(X \leq 10) = 1 - F(10) = 13.533\%$$

(b) *... take between 5 and 15 minutes*

$$\Pr(5 \leq X \leq 15) = F(15) - F(5) = 0.950 - 0.632 = 31.8\%$$

[†]Note: Since X is continuous $\Pr(X \geq 10) = \Pr(X > 10) + \cancel{\Pr(X = 10)} = \Pr(X > 10)$

Example 10

Example 10

An industrial plant with 2000 employees has a mean number of accidents per week equal to $\lambda = 0.4$ and the number of accidents follows a Poisson distribution. What is the probability that

(a) there are between 2 and 4 (inclusive) accidents in a week?

(b) the time between accidents is more than 2 and less than 4 weeks?

MODEL: Poisson (number of accidents) / Exponential (time between accidents),

- $\lambda = 4$ accidents per week

(a) *between 2 and 4 accidents...?* *number of accidents* $\Longrightarrow$ *Poisson* $\boxed{f(x) = \frac{\lambda^x e^{-\lambda}}{x!}}$

$$\begin{aligned}\Pr(2 \leq X \leq 4) &= \Pr(X = 2 \text{ OR } X = 3 \text{ OR } X = 4) \\ &= f(2) + f(3) + f(4) = 0.5372587\end{aligned}$$

(b) *the time between accidents ...?* *time between* $\Longrightarrow$ *Exponential* $\boxed{F(x) = 1 - e^{-\lambda x}}$

$$\Pr(2 \leq X \leq 4) = F(4) - F(2) = 0.2386513$$

The Normal Distribution

Definition 11 (Normal Distribution)

The normal distribution, is a **two-parameter** continuous probability distribution defined by the probability density function (PDF)

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

where parameters

- μ is the mean $\quad -\infty < \mu < \infty$
- σ is the standard deviation $\quad 0 < \sigma < \infty$

- The expected value and variance of a exponential random variable are

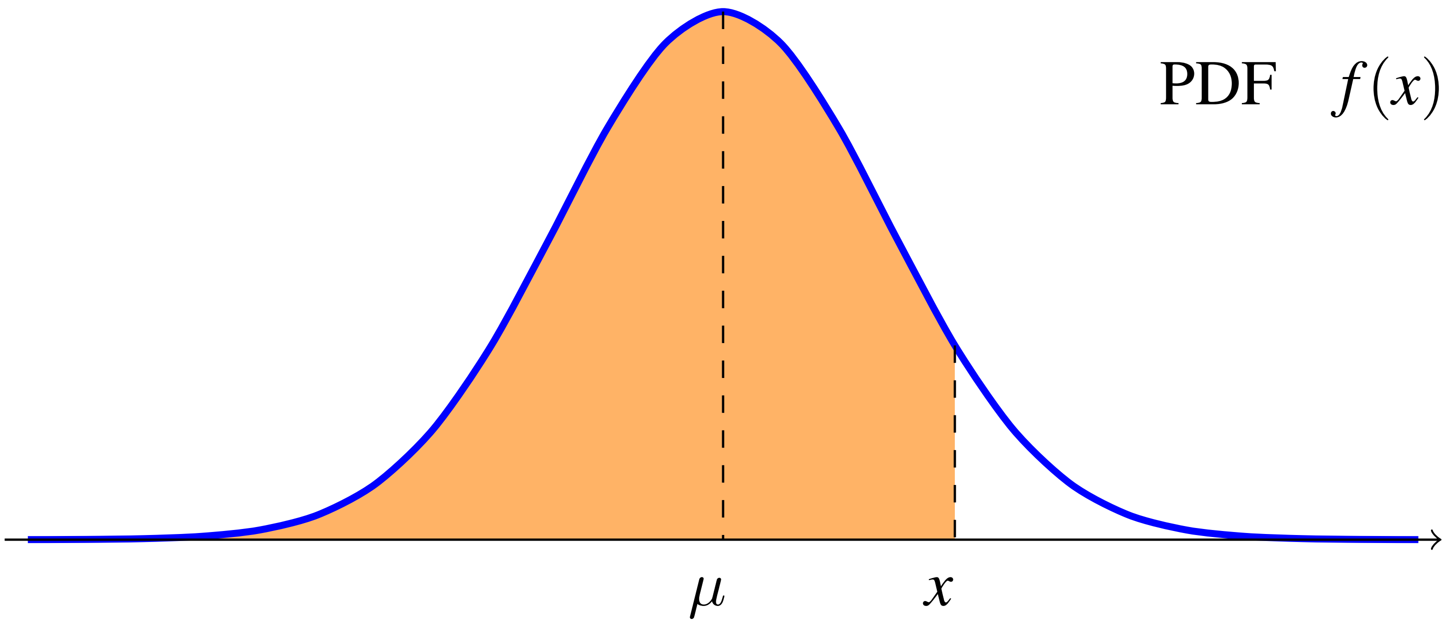
$$\mathrm{E}[X] = \mu \qquad\qquad \mathrm{Var}[X] = \sigma^2$$

- We write $X \sim \mathcal{N}(\mu, \sigma^2)$ to indicate that the random variable, X, follows the normal distribution with mean, μ, and variance, σ^2.

Graph of the Probability Density Function (PDF)

- The shape of the normal probability density function is a symmetric bell-shaped curve centred on the mean μ.
- If $X \sim \mathcal{N}(\mu, \sigma^2)$ then the area under the normal probability density function to the left of x is the cumulative probability

$$\Pr(X \leq x)$$



Standard Normal Distribution

Definition 12 (Standard Normal Distribution)

Let Z be a normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$, that is, $Z \sim \mathcal{N}(0, 1)$. We say that Z follows the standard normal distribution.

We can obtain probabilities for any normally distributed random variable by

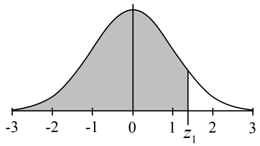
1. Convert the random variable to the standard normally distributed random variable Z using the transformation

$$Z = \frac{X - \mu}{\sigma}$$

2. Use the standard normal distribution tables (page 20 and 21).

Area under the standard normal curve

$$P(z \le z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{1}{2}z^2} dz$$



z_1	0·00	0·01	0·02	0·03	0·04	0·05	0·06	0·07	0·08	0·09
0·0	0·5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0·1	0·5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0·2	0·5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0·3	0·6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0·4	0·6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0·5	0·6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0·6	0·7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0·7	0·7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0·8	0·7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0·9	0·8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1·0	0·8413	8438	8461	8485	8508	8531	8554	8577	8599	8621

Page 20

Area under the standard normal curve (continued)

z_1	0·00	0·01	0·02	0·03	0·04	0·05	0·06	0·07	0·08	0·09
1·1	0·8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1·2	0·8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1·3	0·9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1·4	0·9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1·5	0·9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1·6	0·9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1·7	0·9554	9564	9573	9582	9591	9599	9608	9616	9625	9633

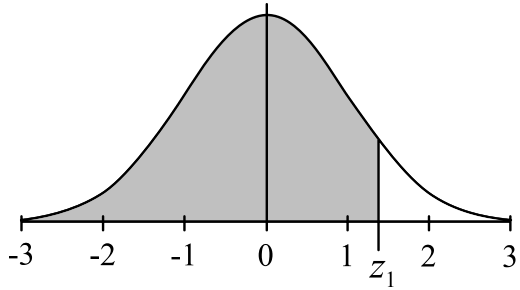
Using Standard Normal Distribution Tables I

The standard normal distribution table gives the values of $F(z) = \Pr(Z \leq z)$ for **non-negative** values of z.

For example $\Pr(Z \leq 0.94) = 0.8264$

Area under the standard normal curve

$$P(z \leq z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{1}{2}z^2}\, dz$$

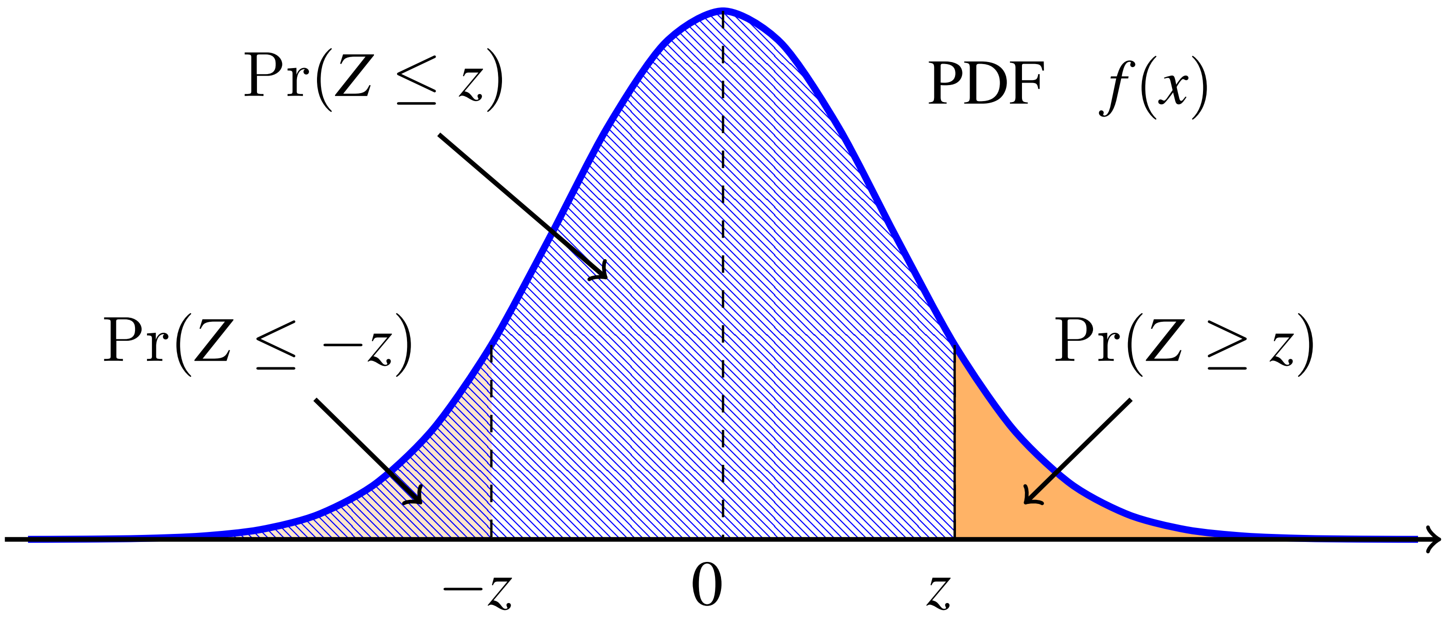


z_1	0·00	0·01	0·02	0·03	0·04	0·05	0·06	0·07	0·08	0·09
0·0	0·5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0·1	0·5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0·2	0·5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0·3	0·6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0·4	0·6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0·5	0·6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0·6	0·7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0·7	0·7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0·8	0·7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0·9	0·8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1·0	0·8413	8438	8461	8485	8508	8531	8554	8577	8599	8621

Using Standard Normal Distribution Tables II

To use the standard normal distribution table for negative z, use the symmetry property and the total law of probability

$$\Pr(Z \leq -z) = 1 - \Pr(Z \leq z)$$



For example

$$\Pr(Z \leq -0.94) = \Pr(Z \geq 0.94) = 1 - \Pr(Z \leq 0.94) = 0.1736$$

Example 13

Example 13 (Computing Standard Normal Probabilities)

Find the area under the standard normal curve that lies

(a) to the left of $z = 2.5$.
(b) to the left of $z = -1.55$.
(c) between $z = 0$ and $z = 0.78$.
(d) between $z = -0.33$ and $z = 0.66$.

(a) *to the left of* $z = 2.5$
$$\Pr(Z \le 2.5) = 0.9948$$
(b) *to the left of* $z = -1.55$
$$\Pr(Z \le -1.55) = 1 - \Pr(Z \le 1.55) = 0.0605$$
(c) *between* $z = 0$ *and* $z = 0.78$
$$\Pr(Z \le 0.78) - \Pr(Z \le 0) = 0.7823 - 0.5 = 0.2823$$
(d) *between* $z = -0.33$ *and* $z = 0.66$
$$\begin{aligned}\Pr(Z \le 0.66) - \Pr(Z \le -0.33) &= 0.7453 - (1 - \Pr(Z \le 0.33)) \\ &= 0.7453 - 0.3707 = 0.3746\end{aligned}$$

Example 14

Example 14 (Converting to Standard Normal)

A random variable has a normal distribution with $\mu = 69$ and $\sigma = 5.1$. What are the probabilities that the random variable will take a value

(a) less than 74.1

(b) greater than 63.1

(c) between 65 and 72.3

$$Z = \frac{X - \mu}{\sigma}$$

(a) *less than 74.1*

$$\Pr(X \leq 74.1) = \Pr\left(Z \leq \frac{74.1 - 69}{5.1}\right) = \Pr\left(Z \leq 1\right) = 0.8413$$

(b) *greater than 63.1*

$$\Pr(X \geq 63.1) = \Pr\left(Z \geq -1.16\right) = 0.8770$$

(c) *between 65 and 72.3*

$$\Pr(65 \leq X \leq 72.3) = \Pr(-0.78 \leq Z \leq 0.64) = 0.5212$$

Example 15

Example 15

A very large group of students obtains test scores that are normally distributed with mean 60 and standard deviation 15. Find the cutoff point for the top 10% of all students for the test scores.

Here we are using the tables in reverse ... we have a probability (0.10) and want to find the particular value for z on the standard normal curves, and then find the corresponding particular value for x on the normal curve (corresponding to $\mu = 60$ and $\sigma = 15$).

- standard normal probability $\rightarrow z$

$$\text{If } \Pr(Z \leq z) = 0.90 \text{ then } z = 1.28$$

  since $z = 1.28$ has probability value closest to 0.9.

- standard normal $\mathcal{N}(0, 1) \rightarrow$ normal $\mathcal{N}(60, 15^2)$

$$z = 1.28 \quad \rightarrow \quad x = (1.28)(15) + 50 = 79.2$$

$$Z = \frac{X - \mu}{\sigma}$$

Example 16

Example 16

A random variable has a normal distribution with variance 100. Find its mean if the probability that it will take on a value less than 77.5 is 0.8264.

- From the problem statement, we know

$$\Pr(X \leq 77.5) = 0.8264 = \Pr(Z \leq z)$$

- We can determine z using the standard normal tables (in reverse)

$$\Pr(Z \leq z) = 0.8264 \quad \Longrightarrow \quad z = 0.94$$

- The relationship between random variables $X \sim \mathcal{N}(\mu, 100)$ and Z means

$$Z = \frac{X - \mu}{\sigma} \quad \Longleftrightarrow \quad 0.94 = \frac{77.5 - \mu}{10} \quad \Longrightarrow \quad \mu = 68.1$$

Normal Distribution Approximation to Binomial

Normal distribution can be used to approximate the binomial distribution. This approximation allows to speed up the computation of probabilities when the number of trials is large:

- Recall, that the binomial has

$$\mathrm{E}[X] = \mu = np \qquad \text{and} \qquad \mathrm{Var}[X] = \sigma^2 = np(1-p)$$

- If the number of trials (n) is large and probability of success, p is close to 0.5, so that $np(1-p) > 5$, then the distribution of the random variable

$$Z = \frac{X-\mu}{\sigma} = \frac{X-np}{\sqrt{np(1-p)}}$$

is approximately a standard normal distribution.

Example 17

Example 17

An election forecaster has obtained a random sample of 900 voters in which 460 indicate that they will vote for Candidate A. Assuming that there are only two candidates, what is the probability of Candidate A winning the election?
What is the normal approximation to this probability?

MODEL Binomial, 'success' = 'vote for Candidate A'

- n = number of voters $n = 900$
- p = probability of 'success' $p = 460/900 = 0.511$

$$\Pr(\text{winning election}) = \Pr(X \geq 451) = 73.68\%$$

MODEL Normal

- Mean $\mu = np = 500$
- Variance $\sigma^2 = np(1-p) = 224.89$ $np(1-p) > 5$? ✔

$$\Pr(\text{winning election}) = \Pr(X \geq 451) = \Pr\left(Z \geq \frac{451 - 460}{\sqrt{224.89}}\right) = 72.59\%$$