Special Continuous Random Variables

CSGE602013 –STATISTICS AND PROBABILITY

FACULTY OF COMPUTER SCIENCE UNIVERSITAS INDONESIA

References

- Introduction to Probability and Statistics for Engineers & Scientists, 4th ed., Sheldon M. Ross, Elsevier, 2009.
- A First Course in Probability, 8th Edition. Sheldon M. Ross
- Applied Statistics for the Behavioral Sciences, 5th Edition, Hinkle., Wiersma., Jurs., Houghton Mifflin Company, New York, 2003.
- Probability and Statistics for Engineers & Scientists, 4th Edition. Anthony J. Hayter, Thomson Higher Education

Menu

01

Uniform Random Variables

02

Exponential Random Variables

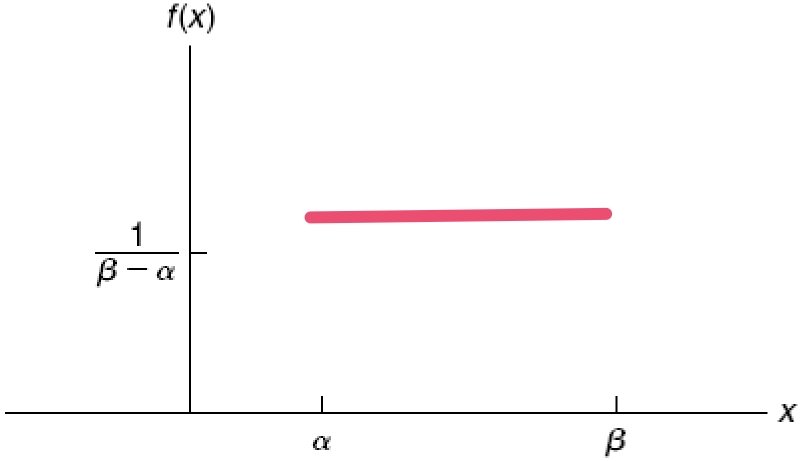
03

Normal (Gaussian) Random Variables

Uniform Random Variables

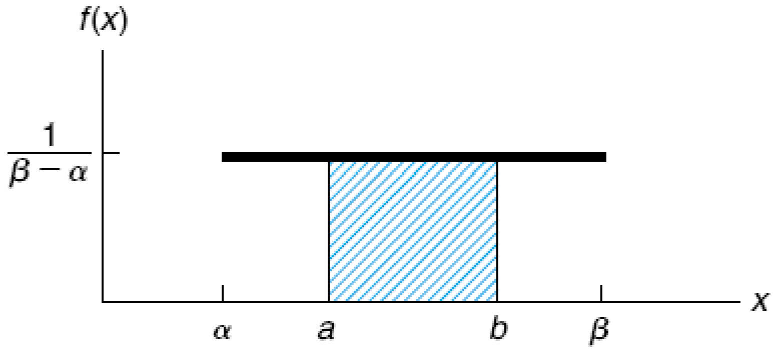
- A random variable **X** is said to be **uniformly** distributed over the interval **[α, β]** if its probability density function is given by:

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \alpha \le x \le \beta \\ 0 & otherwise \end{cases}$$



Graph of $f(x)$ is uniform [α, β].

Uniform Random Variables (2)



$P\{a < X < b\}$,
(a, b) is a subinterval of $[\alpha, \beta]$

$$P(a < x < b) = \frac{1}{\beta-\alpha}\int_a^b dx = \frac{b-a}{\beta-\alpha}$$

$$F(x) = P(X \le x) = \begin{cases} 0 & x < \alpha \\ \dfrac{1}{\beta-\alpha}\displaystyle\int_\alpha^x dx = \dfrac{x-\alpha}{\beta-\alpha} & \alpha \le x \le \beta \\ 1 & x > \beta \end{cases}$$

E[X] of a Uniform RV

The mean of a uniform $[\alpha, \beta]$ random variable is

$$E[X] = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha}\, dx$$
$$= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)}$$
$$= \frac{(\beta - \alpha)(\beta + \alpha)}{2(\beta - \alpha)}$$

or

$$E[X] = \frac{\alpha + \beta}{2}$$

Var(X) of a Uniform RV

The variance is computed as follows.

$$E[X^2] = \frac{1}{\beta - \alpha}\int_\alpha^\beta x^2 dx = \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} = \frac{\beta^2 + \alpha\beta + \alpha^2}{3}$$

and so

$$\mathrm{Var}(X) = \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \left(\frac{\alpha + \beta}{2}\right)^2 = \frac{\alpha^2 + \beta^2 - 2\alpha\beta}{12} = \frac{(\beta - \alpha)^2}{12}$$

Example

If X is **uniformly** distributed over the interval [0, 10], compute:

1. $F(x)$
2. $P(1 < X < 4)$
3. $P(X < 5)$
4. $P(X > 6)$
5. $P(X > 6 \mid X > 5)$

Example

If X is **uniformly** distributed over the interval **[0, 10]**, compute:

1) $F(x)$

$$F(x) = P(X \le x) = \begin{cases} 0 & x < 0 \\ \dfrac{1}{10-0}\displaystyle\int_0^x dx = \dfrac{x}{10} & 0 \le x \le 10 \\ 1 & x > 10 \end{cases}$$

2) $P(1 < X < 4)$

$$P(1 < X < 4) = F(4) - F(1) = \frac{4-1}{10} = 0.3$$

Example

If X is **uniformly** distributed over the interval **[0, 10]**, compute:

3) $P(X < 5)$

$$P(X < 5) = F(5) = \frac{5}{10} = 0.5$$

4) $P(X > 6)$

$$P(X > 6) = 1 - P(X \leq 6) = 1 - F(6) = 1 - \frac{6}{10} = 0.4$$

5) $P(X > 6 \mid X > 5)$

$$P(X > 6 \mid X > 5) = \frac{P(X > 6, X > 5)}{P(X > 5)}$$

$$= \frac{P(X > 6)}{P(X > 5)} = \frac{1 - F(6)}{1 - F(5)} = \frac{4}{5}$$

Exercise

- 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) at least **12 minutes** for a bus

Exercise

- 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.
- In this example $\alpha = 0, \beta = 30$;

$$f(x) = \begin{cases} \frac{1}{30}, & if\ 0 \le x \le 30 \\ 0, & otherwise \end{cases}$$

- Note that between 7 and 7:30 there are two schedules for the bus, i.e. 7.15 and 7:30
- 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) at least **12 minutes** for a bus

Exercise

- 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.
- In this example $\alpha = 0, \beta = 30;$

$$f(x) = \begin{cases} \frac{1}{30}, if\ 0 \leq x \leq 30 \\ 0, \qquad otherwise \end{cases}$$

- Note that between 7 and 7:30 there are two schedules for the bus, i.e. 7.15 and 7:30
- 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

  - for this passenger to wait less than 5 minutes he must arrive between 7:10 and 7:15 or 7:25 and 7:30.
  - Thus the probability is

$$P(10 < X \ \leq 15) + \ P(25 < X \leq 30) = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}.$$

Exercise

- 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.
- In this example $\alpha = 0, \beta = 30$;

$$f(x) = \begin{cases} \frac{1}{30}, & if\ 0 \le x \le 30 \\ 0, & otherwise \end{cases}$$

- Note that between 7 and 7:30 there are two schedules for the bus, i.e. 7.15 and 7:30
- 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:

  b) at least **12 minutes** for a bus:

  - he must arrive between 7:0 and 7:03 or 7:15 and 7:18;
  - Thus the probability is

$$P(0 < X \le 3) + P(15 < X \le 18) = \frac{3}{30} + \frac{3}{30} = \frac{1}{5}.$$

Recall: What Was the Poisson distribution?

The number of arrivals within an interval follows the **Poisson** distribution.

Recall: What Was the Poisson distribution?

The number of arrivals within an interval follows the **Poisson** distribution.

The amount of time between two successive arrivals follows the **exponential** distribution.

Exponential Random Variables

- The exponential distribution is often used to describe the amount of time until some specific event occurs.
- For example:
  - The amount of time until an earthquake occurs
  - The amount of time until a new war breaks out
  - The amount of time until a telephone call you receive turns out to be a wrong number
  - Time between two successive job arrivals to a file server (interarrival time)
  - Time to failure of a component (lifetime of a component)

Exponential Random Variables (2)

- The random variable X is said to be an exponential random variable (exponentially distributed) with parameter λ if it's PDF is given by:

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

$$X \sim Exp(\lambda)$$

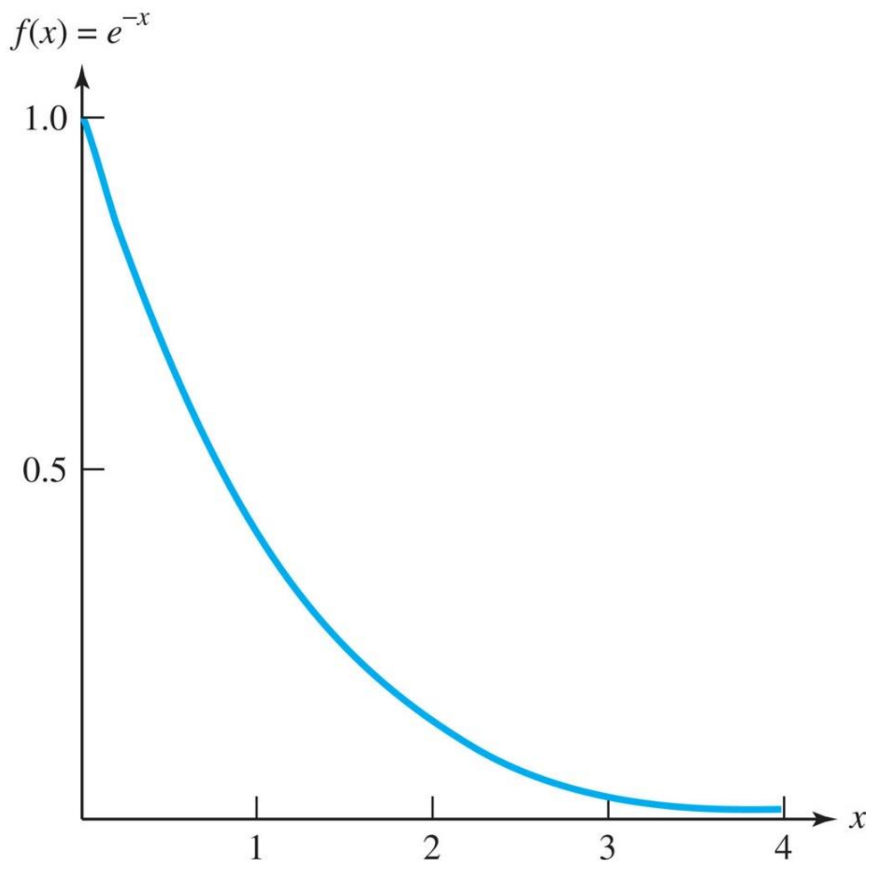
- Where, λ is
  - Arrival rate per unit interval
  - Average number of events occurring per unit of interval

Exponential Random Variables (3)

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

$$X \sim Exp(\lambda)$$

- Exponential distribution with **λ** = 1



Exponential Random Variables (4)

Expectation / Mean	$E[X] = \frac{1}{\lambda}$
Variance	$Var(X) = \frac{1}{\lambda^2}$
PDF	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$
CDF	$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$

Example

- The arrival numbers of the customer in a store is a Poisson RV with rate $\lambda = 5$ visitors per hour.

1. What is the probability that the next visitor coming in less than 10 minutes ?

2. What is the probability that the next visitor coming between minute 10 to 12 from now ?

Example

- The arrival numbers of the customer in a store is a Poisson RV with rate λ = 5 visitors per hour.
- Let X be a R.V. that denotes the amount of time until a customer arrive (or inter-arrival time)

1. What is the probability that the next visitor coming in less than 10 minutes ?

$$\lambda = 5 \, {visitors}/{hour} \qquad X \sim Exp(5) \qquad 10 \min = 1/6 \; hour$$

$$P(X < 1/6) = F(1/6) = 1 - e^{-5/6} = 0.5654$$

2. What is the probability that the next visitor coming between minute 10 to 12 from now ?

$$10 \min = 1/6 \; hour \qquad 12 \min = 1/5 \; hour$$

$$\begin{aligned} P(1/6 < X < 1/5) &= F(1/5) - F(1/6) \\ &= (1 - e^{-5/5}) - (1 - e^{-5/6}) \\ &= 0.0667 \end{aligned}$$

Markov (Memoryless) Property

That is,

$$P(X > s+t \mid X > t) = P(X > s) \qquad s,t \geq 0$$

The **exponential distribution** is the only continuous distribution that has the memoryless property.

The distribution of additional functional life of an item of age t is the same as that of a new item.

There is **no need to remember** the age of a functional item since as long as it is still functional it is "**as good as new.**"

Markov (Memoryless) Property

Proof:

$$
\begin{aligned}
& P(X > s + t \mid X > t) \\
& = \frac{P(X > s + t, X > t)}{P(X > t)} \\
& = \frac{P(X > s + t)}{P(X > t)} \\
& = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} \qquad P(X > s + t, X > t) = P(X > s)P(X > t) \\
& = e^{-\lambda s} = P(X > s)
\end{aligned}
$$

Example

- Suppose that a number of miles that a car run before its battery wears out is exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5,000 mile trip, what is the probability that he will be able to complete his trip without to replace the battery?

Example

- Suppose that a number of miles that a car run before its battery wears out is exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5,000 mile trip, what is the probability that he will be able to complete his trip without to replace the battery?
- Let X be a random variable including the remaining lifetime (in thousand miles) of the battery. Then,

$$E[X] = 1/\lambda = 10 \qquad \Rightarrow \lambda = 1/10$$

$$P(X > 5) = 1 - F(5) = e^{-5\lambda} = e^{-1/2} = 0.604$$

Example

- Suppose that a number of miles that a car run before its battery wears out is exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5,000 mile trip, what is the probability that he will be able to complete his trip without to replace the battery?
- Let X be a random variable including the remaining lifetime (in thousand miles) of the battery. Then,

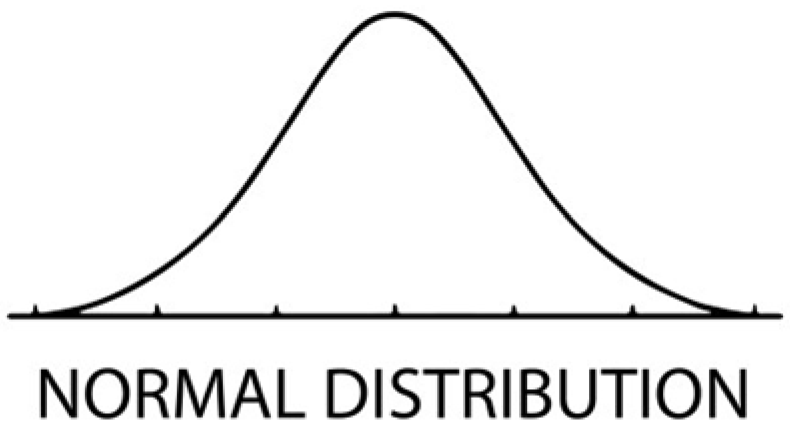
$$E[X] = 1/\lambda = 10 \qquad \Rightarrow \lambda = 1/10$$
$$P(X > 5) = 1 - F(5) = e^{-5\lambda} = e^{-1/2} = 0.604$$

- What if X is NOT exponential R. V. ?

$$P(X > t+5 \mid X > t) = \frac{1 - F(t+5)}{1 - F(t)}$$ Additional information **t** is needed

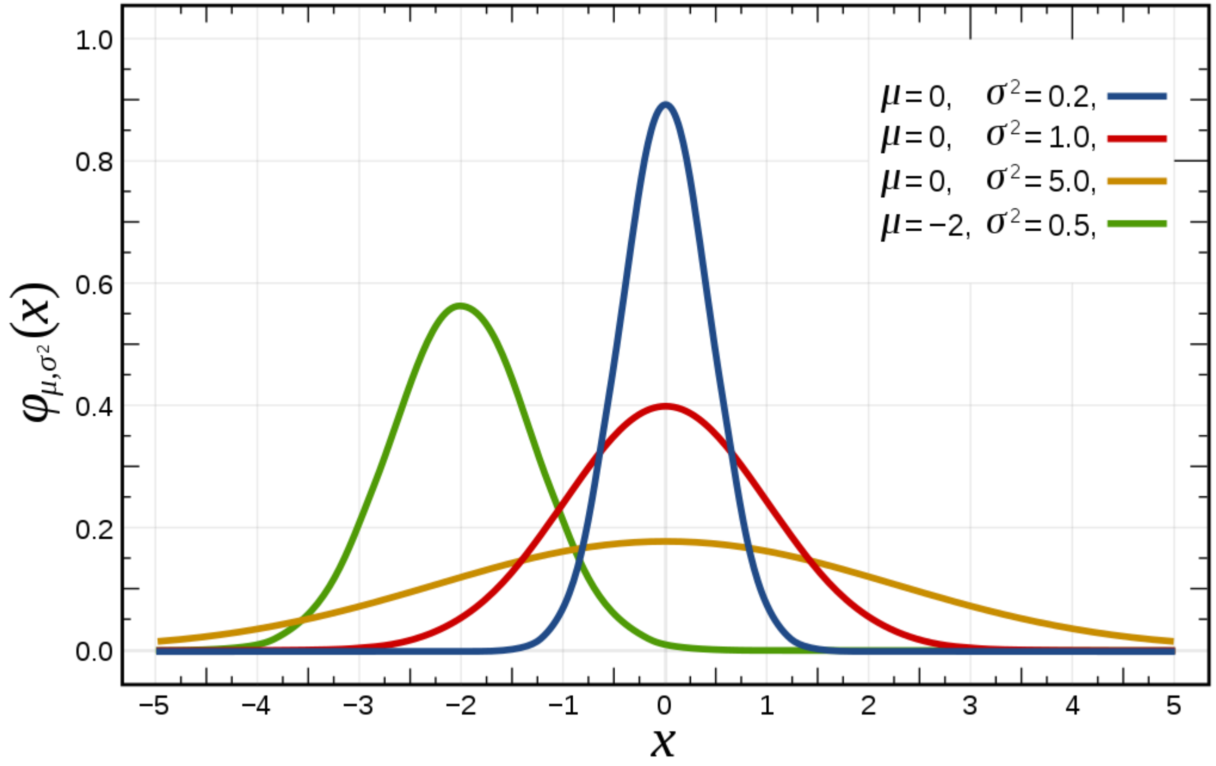
Normal Random Variables

- What does a normal distribution look like?



Normal Random Variables (2)

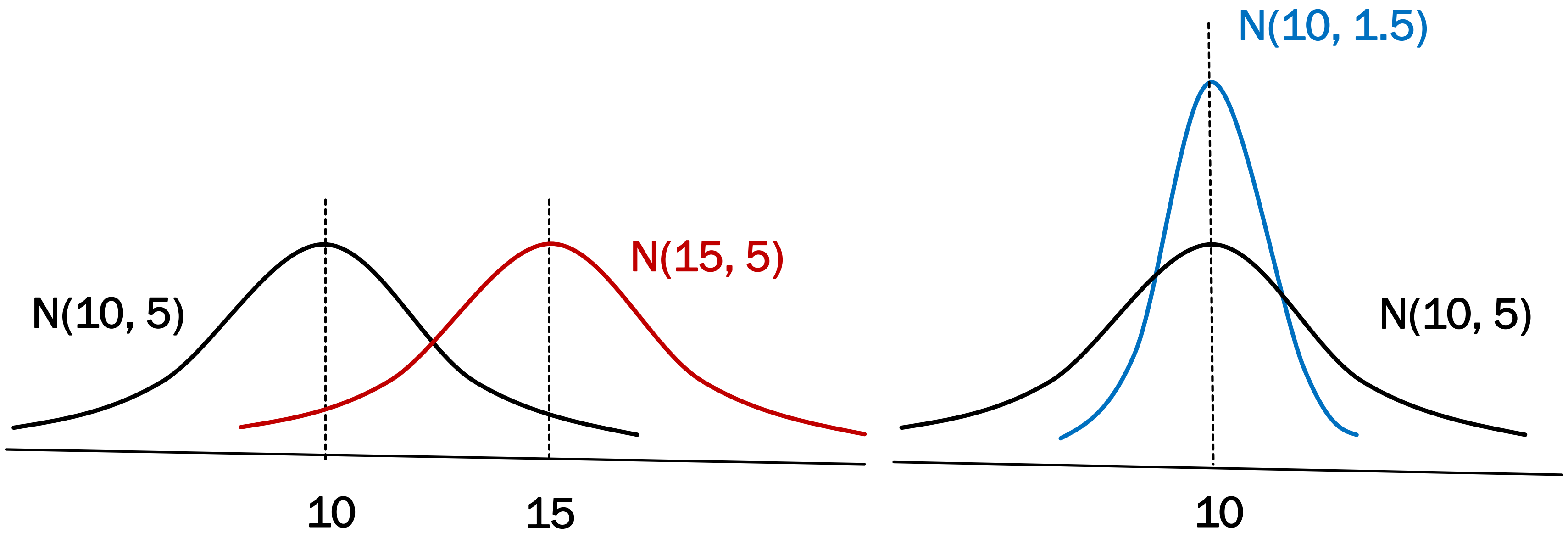
- Most things in the world follow a normal distribution
- Gaussian / Gaussian – Laplace / Normal
- Bell curve



Normal Random Variables (3)

- X is a Normal (Gaussian) R.V. with parameters **μ** dan **σ**^{2}

$$X \sim N(\mu, \sigma^2)$$



Normal Random Variables (4)

Expectation / Mean

$$E[X] = \mu$$

Variance

$$Var(x) = \sigma^2$$

PDF

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

Standard Normal Distribution

- When you hear the word "standard"..
  - Remember standard scores?
- Standard Normal R.V. Z is obtained using:

$$Z = \frac{X - \mu}{\sigma} \qquad X \sim N(\mu, \sigma^2)$$

Standard Normal Distribution

- Standard Normal R.V. Z is obtained using:

$$Z = \frac{X - \mu}{\sigma} \qquad X \sim N(\mu, \sigma^2) \qquad Z \sim N(0,1)$$

Expectation / Mean

$$E[Z] = \frac{E[X] - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$$

Variance

$$Var(Z) = \frac{1}{\sigma^2} Var(X) = 1$$

PDF

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < \infty$$

Standard Normal Distribution (2)

- PDF

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < \infty$$

- CDF in correlation to PDF

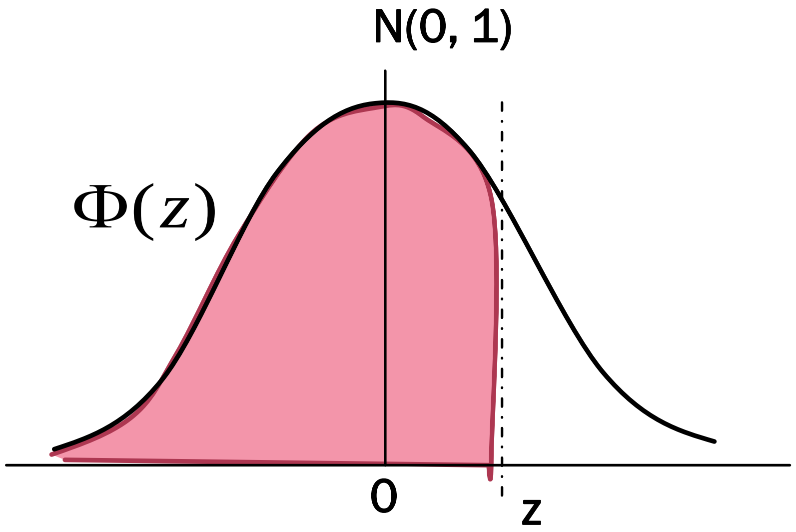
$$F_x(z) = \int_{-\infty}^{z} f(y)dy$$

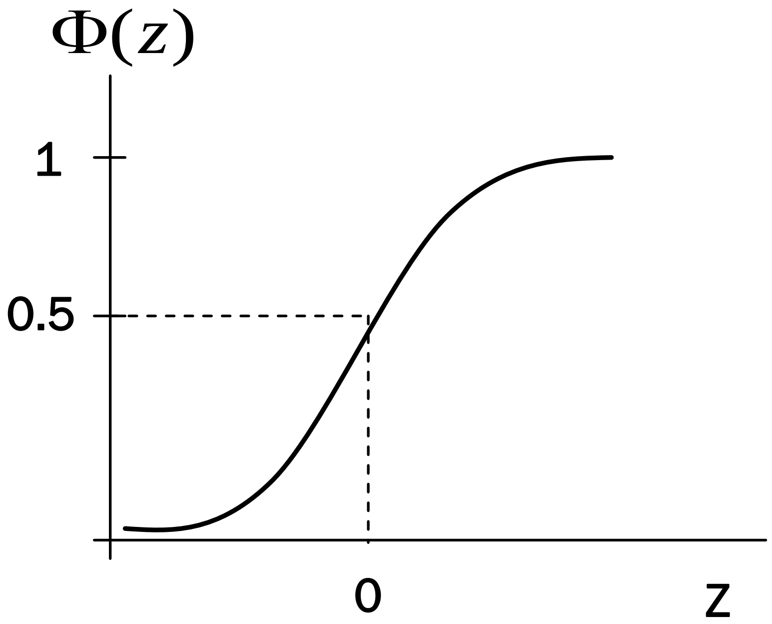
Standard Normal Distribution (3)

- CDF for standard normal distribution:

$$\Phi(z) = \int_{-\infty}^{z} f(y)dy$$

$$\Phi(z) = P(Z \leq z)$$



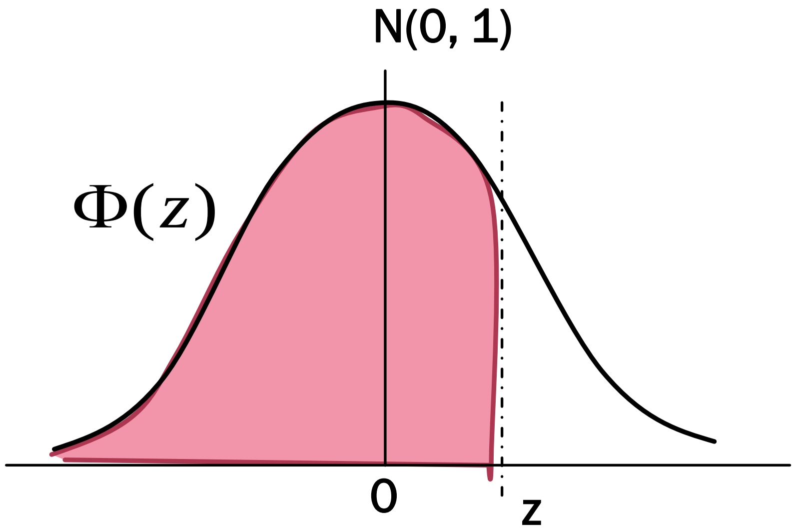


Standard Normal Distribution (4)

- CDF for standard normal distribution:

$$\Phi(z) = \int_{-\infty}^{z} f(y)dy$$

$$\Phi(z) = P(Z \leq z)$$



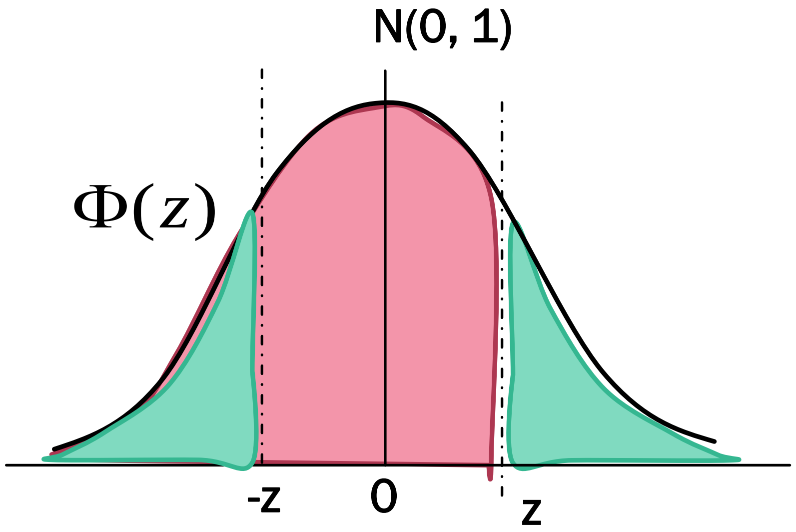
How would I compute the value of $\Phi(z)$?
Area under the curve!

- Integrate $f(y)$
- Use the table!!

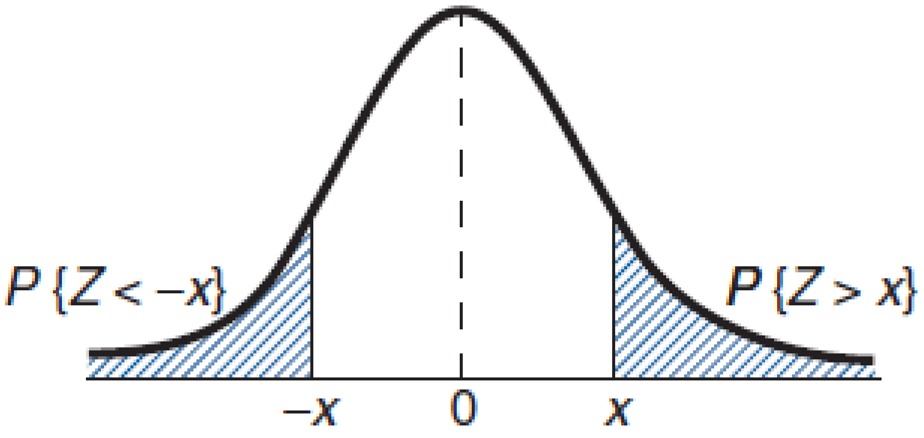
What about the Opposite Side?

- What about the opposite probability?

$$1-\Phi(z)=P(Z\geq z)=P(Z\leq -z)=\Phi(-z)$$



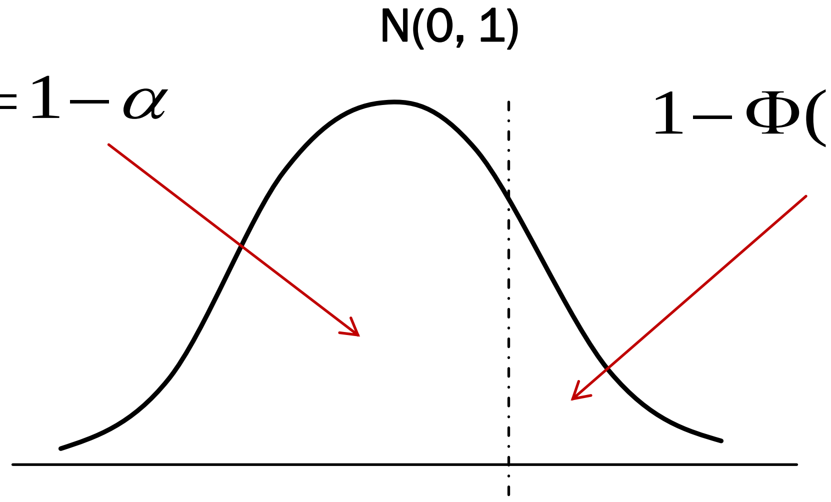
Symmetric property of the normal distribution



Z_α notation

- Introducing the Z_α notation

$$\Phi(z_\alpha) = 1 - \alpha$$



$$1 - \Phi(z_\alpha) = \Phi(-z_\alpha) = \alpha$$

"Probability that a standard normal R.V. is greater than $\mathbf{z_\alpha}$ is equal to $\boldsymbol{\alpha}$."

Computing the Probability of a Standard Normal RV

$$X \sim N(\mu, \sigma^2)$$

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

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right)$$

$$= P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(\mu - c\sigma \le X \le \mu + c\sigma) = P(-c \le Z \le c)$$

$$P(X \le \mu + \sigma z_\alpha) = P(Z \le z_\alpha) = 1-\alpha$$

$$P(\mu - \sigma z_{\alpha/2} \le X \le \mu + c z_{\alpha/2}) = P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha$$

Standard Normal Table $\Phi(z)$

- This table provides Φ(z) for z from 0.00 to 3.49.

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015

Reading the Standard Normal Table $\Phi(z)$

- How to read:
  - to find $\Phi(x.yz)$, go to row $(x.y)$, then in the same row go to the right to find column z.
  - Expecially for $z > 3$ $\Phi(x.y)$, the right column is for y.
  - Recall the symmetry properties of the function of pdf,
    - for example $\Phi(-0.5) = 1 - \Phi(0.5)$

Example

- X is a normal random variable with $\mu = 3$ and variance $\sigma^2 = 16$, find
  - $P(X < 11)$
  - $P(X > -1)$

Example

- X is a normal random variable with $\mu = 3$ and variance $\sigma^2 = 16$, find
  - $P(X < 11)$

$$P(X < 11) = P\left(\frac{X-3}{4} < \frac{11-3}{4}\right) = P(Z < 2) = \Phi(2) = 0.9772$$

  - $P(X > -1)$

$$P(X > (-1)) = P\left(\frac{X-3}{4} < \frac{-1-3}{4}\right) = P\big(Z > (-1)\big) = P(Z < 1) = \Phi(1) = .84134$$

Example

- X is a normal random variable with $\mu = 3$ and variance $\sigma^2 = 16$, find
  - $P(2 < x < 7)$

$$P(2 < x < 7) = P\left(\frac{2-3}{4} < \frac{X-3}{4} < \frac{7-3}{4}\right) = P\left(-\frac{1}{4} < Z < 1\right)$$

$$= P(Z < 1) - P\left(Z < -\frac{1}{4}\right)$$

$$= \Phi(1) - \Phi\left(-\frac{1}{4}\right)$$

$$= \Phi(1) - \left(1 - \Phi\left(\frac{1}{4}\right)\right)$$

$$= 0.8413 - 1 + 0.5987$$

$$= 0.4400$$

Example

- Analogue signals received by a detector is modeled as Gaussian RV N(200, 256) using microvolt unit (μV).

- What is the probability the signal received exceeds 240 μV ?
- What is the probability the signal received exceeds 240 μV if the signal is known to be larger than 210 μV ?

Example

- Analogue signals received by a detector is modeled as Gaussian RV N(200, 256) using microvolt unit (μV).

- What is the probability the signal received exceeds 240 μV ?
- Let **X** : analogue signals received by a detector, **X ~ N(200, 256)**

See the Z-table !

$$P(X > 240) = 1 - P(X \le 240)$$
$$= 1 - \Phi\left(\frac{240 - 200}{16}\right) = 1 - \Phi(2.5) = 0.0062$$

Example

- Analogue signals received by a detector is modeled as Gaussian RV N(200, 256) using microvolt unit (μV).

- What is the probability the signal received exceeds 240 μV if the signal is known larger than 210 μV ?

$$P(X > 240 \mid X > 210) = \frac{P(X > 240)}{P(X > 210)}$$

$$= \frac{1 - P(X \le 240)}{1 - P(X \le 210)} = \frac{1 - \Phi(2.5)}{1 - \Phi\left(\frac{210 - 200}{16}\right)} = \frac{1 - \Phi(2.5)}{1 - \Phi(0.625)}$$

$$= 0.02335$$

Exercise

- A car battery has an average life of 3 years, with a standard deviation of 0.5 years. If the battery life is assumed to follow a normal distribution, find:

- The probability the battery lives less than 4 years
- The probability the battery lives less than 2.3 years
- The probability the battery lives more than 3.5 years
- The probability the battery lives between 2.5 to 3.5 years

Linear Combination of Normal Random Variables

- Sum of **independent** normal random variables is also a normal random variable

$$X \sim N(\mu, \sigma^2)$$
$$\Rightarrow Y = aX + b$$
$$\Rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$$

a and **b** are constants

$$X_1 \sim N(\mu_1, \sigma_1^{\;2})$$
$$X_2 \sim N(\mu_2, \sigma_2^{\;2})$$
$$\Rightarrow Y = X_1 + X_2$$
$$\Rightarrow Y \sim N(\mu_1 + \mu_2, \sigma_1^{\;2} + \sigma_2^{\;2})$$

X_1 and X_2 are **independent** !

Linear Combination of Normal Random Variables

$$X_i \sim N(\mu_i, \sigma_i^{\,2})$$

for **1 ≤ i ≤ n**, X_i are **independent**

$$\Rightarrow Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b$$

a_i and **b** are constants

$$\Rightarrow Y \sim N\left(b + \sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^{\,2} \sigma_i^{\,2}\right)$$

Example

- The yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches.
- Find the probability that the total precipitation during the next 2 years will exceed 25 inches !

Example 1

- The yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches.
- Find the probability that the total precipitation during the next 2 years will exceed 25 inches!

$$X_i \sim N(12.08, (3.1)^2)$$

- We know that X1 + X2 is also normal random variable.

Example 1

- The yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches.
- Find the probability that the total precipitation during the next 2 years will exceed 25 inches!

$$X_i \sim N(12.08, (3.1)^2)$$

- We know that X1 + X2 is also normal random variable.

$$X_i \sim N(12.08, (3.1)^2)$$

$$\begin{aligned} X_1 + X_2 &\sim N(12.08 + 12.08,\ 2(3.1)^2) \\ &\sim N(24.16, 19.22) \end{aligned}$$

$$P(X_1 + X_2 > 25) = P\left(\frac{X_1 + X_2 - 24.16}{\sqrt{19.22}} > \frac{25 - 24.16}{\sqrt{19.22}}\right)$$

$$= P(Z > 0.19) = P(Z < -0.19) = \Phi(-0.19) = 0.42465$$

Example 2

- The yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches.
- Find the probability that next year's precipitation will exceed that of the following year by more than 3 inches ?

$$X_i \sim N(12.08, (3.1)^2)$$

**Assume that the precipitation totals for the next 2 years are independent.

Example 2

- The yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches.
- Find the probability that next year's precipitation will exceed that of the following year by more than 3 inches ?

$$X_i \sim N(12.08, (3.1)^2)$$

- We know that X1 - X2 is also a normal random variable

Example 2

- The yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches.
- Find the probability that next year's precipitation will exceed that of the following year by more than 3 inches ?

$$X_i \sim N(12.08, (3.1)^2)$$

- We know that X1 - X2 is also a normal random variable

$$E[X_1 - X_2] = E[X_1] - E[X_2] = 0$$

$$Var(X_1 - X_2) = Var(X_1) + (-1)^2 Var(X_2)$$

$$= 2.(3.1)^2 = 19.22$$

Example 2

- The yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches.
- Find the probability that next year's precipitation will exceed that of the following year by more than 3 inches ?

$$X_i \sim N(12.08, (3.1)^2)$$

- We know that X1 - X2 is also a normal random variable

$$\begin{aligned} P(X_1 > X_2 + 3) &= P(X_1 - X_2 > 3) \\ &= P\left(\frac{X_1 - X_2 - 0}{\sqrt{19.22}} > \frac{3-0}{\sqrt{19.22}}\right) \\ &= P(Z > 0.6843) \\ &= 0.2469 \end{aligned}$$

Normal Approximation to The Binomial Distribution

- Besides being approximated by a Poisson RV, Binomial RV can also be approximated by the Normal RV.

$$X \sim B(n, p) \underset{approx}{\rightarrow} X \underset{approx}{\sim} N(np, np(1-p))$$

- This approximation works well as long as $np \geq 5$ and $n(1-p) \geq 5$.

Normal Approximation to The Binomial Distribution (2)

- We need **correction** because Normal RV is continuous RV.

$$X \sim B(n, p) \underset{approx}{\rightarrow} Y \sim N(np, np(1-p))$$

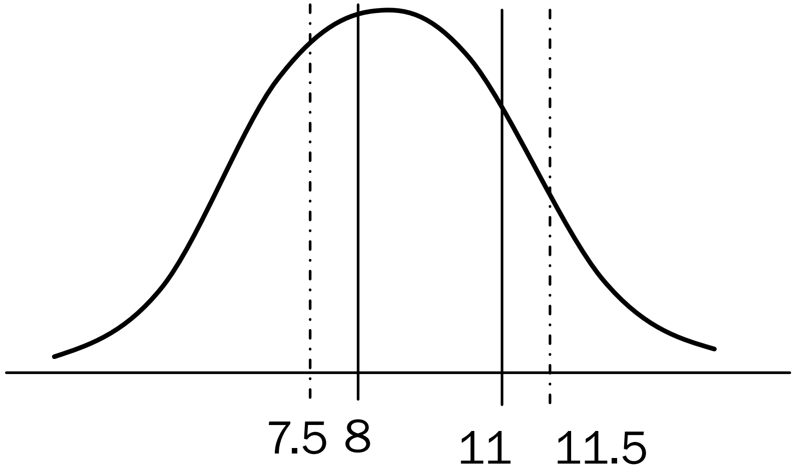
$$P(X \leq x) \approx P(Y < x + 0.5) = \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

$$P(X < x) \approx P(Y < x - 0.5) = \Phi\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

$$P(X \geq x) \approx P(Y \geq x - 0.5) = 1 - \Phi\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

$$P(X > x) \approx P(Y > x + 0.5) = 1 - \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Normal Approximation to The Binomial Distribution (3)

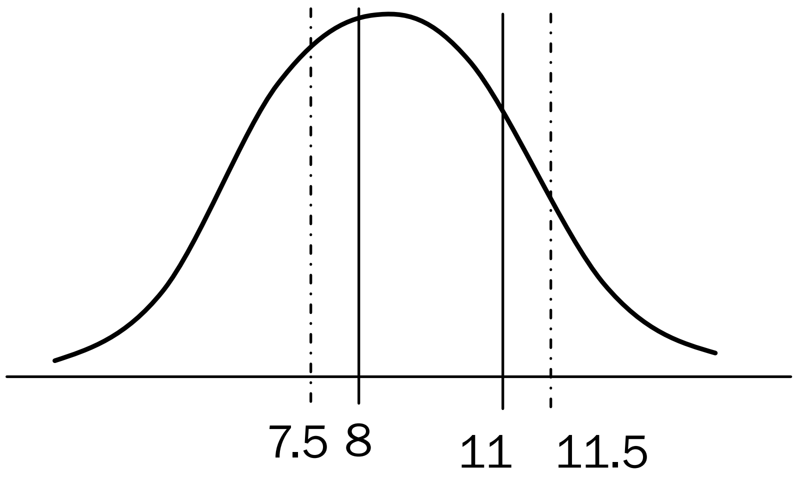


$$X \sim B(16, 0.5)$$

$$Y \sim N(8, 4)$$

- $P(8 \leq X \leq 11) =$
- $P(7.5 \leq Y \leq 11.5) =$

Normal Approximation to The Binomial Distribution (3)



$$X \sim B(16, 0.5)$$

$$Y \sim N(8, 4)$$

- $P(8 \leq X \leq 11) =$

$$\sum_{x=8}^{11} \binom{16}{x} (0.5)^x (0.5)^{16-x} = 0.5598$$

- $P(7.5 \leq Y \leq 11.5) =$

$$= \Phi\left(\frac{11.5 - \mu}{\sigma}\right) - \Phi\left(\frac{7.5 - \mu}{\sigma}\right)$$

$$= \Phi\left(\frac{11.5 - 8}{2}\right) - \Phi\left(\frac{7.5 - 8}{2}\right)$$

$$= \Phi(1.75) - \Phi(-0.25) = 0.5586$$

Example

- The polls are often conducted with a random selection from a number of people from several places. For instance, it is selected 400 people from a city inhabited by men and the women assumed in the same amount. What is the probability that within the selected people there are 190 women at most?
- This is a Binomial problem with n = 400 and p = 0.5.
- Find $P(\mathrm{X} \leq 190)$. Try defining with binomial PMF!

Example

- The polls are often conducted with a random selection from a number of people from several places. For instance, it is selected 400 people from a city inhabited by men and the women assumed in the same amount. What is the probability that within the selected people there are 190 women at most?
- This is a Binomial problem with n = 400 and p = 0.5 → X~Bin(400, 0.5)
- Find $P(X \leq 190)$. Try defining with binomial PMF!
- Manual calculations with a 400! is impossible. While using Normal RV approximation, we obtain: $Y \sim N$ (200, 100)

$$P(X \leq 190) \approx P(Y \leq 190 + 0.5) = \Phi\left(\frac{190 + 0.5 - 200}{\sqrt{100}}\right) = \Phi(-0.95) = 0.1711$$

Exercise

- The probability that an oyster produces a pearl is 0.6.
- How many oyster does an oyster farmer need to farm in order to be 99% confident of having at least 1000 pearls ?

Exercise

- The probability that an oyster produces a pearl is 0.6.
- How many oyster does an oyster farmer need to farm in order to be 99% confident of having at least 1000 pearls ?
  - X: the number of pearls
  - X ~ B(n, 0.6) => we approximate using Y ~ N(0.6n, 0.24n)

$$P(X \geq 1000) \approx P(Y \geq 999.5) = 1 - \Phi\left(\frac{999.5 - 0.6n}{\sqrt{0.24n}}\right) = 0.99$$

$$\Phi\left(\frac{999.5 - 0.6n}{\sqrt{0.24n}}\right) = 1 - 0.99 = 0.01$$

$$\frac{999.5 - 0.6n}{\sqrt{0.24n}} = -2.33 \Leftrightarrow 999.5\text{-}0.6\text{n} = \text{-}2.3\sqrt{0.24\text{n}}$$

$$0.36n^2 - 1200.7n + 999000.25 = 0 \Leftrightarrow n = \frac{1200.7 + \sqrt{3127.18}}{0.72} = 1745.31 \approx 1746$$