SDSC2102
Statistical Methods and Data Analysis

Topic 1. Basic Probability and Statistics Theory

Special Continuous Distributions

- Exponential distribution
- Gama distribution
- Uniform distribution
- Normal distribution

Exponential Distribution

- A continuous r.v. X with an exponentially descreasing p.d.f.
  - Often models the time until an event occurs
  - The p.d.f. and c.d.f. depend on parameter $\lambda > 0$

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

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

Class Problems on Continuous Distributions

1. The lifetime of a single-cell organism is exponentially distributed with parameter λ=0.1 per hour. What is the probability that this organism will live for more than 20 hours? What is the probability that it will die within 5 hours?

Let X be the lifetime of the organism

$X \sim \text{Exponential}\ (\lambda)$

$$P(X > 20) = 1 - P(X \leq 20)$$

$$= 1 - \left(1 - e^{-0.1\times 20}\right) = e^{-2}$$

$$P(X < 5) = F(5) = 1 - e^{-0.1\times 5} = 1 - e^{-0.5}$$

Gamma Distribution

- The gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$
  - $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
  - $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
  - For a positive integer n: $\Gamma(n) = (n - 1)!$
- Example

$$\Gamma\left(\frac{5}{2}\right) = \left(\frac{3}{2}\right)\Gamma\left(\frac{3}{2}\right) = \left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right) = \left(\frac{3}{4}\right)\sqrt{\pi}$$

Gamma Distribution

- Generalization of the Exponential distribution
  - Often models the time required for multiple occurrences of a specific event
  - The p.d.f. depends on parameters $\alpha > 0$ and $\beta > 0$

$$f(x) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

- For $\alpha = 1$, we get the Exponential distribution.
- Called Erlong distribution if α is an integer.

Gamma Distribution

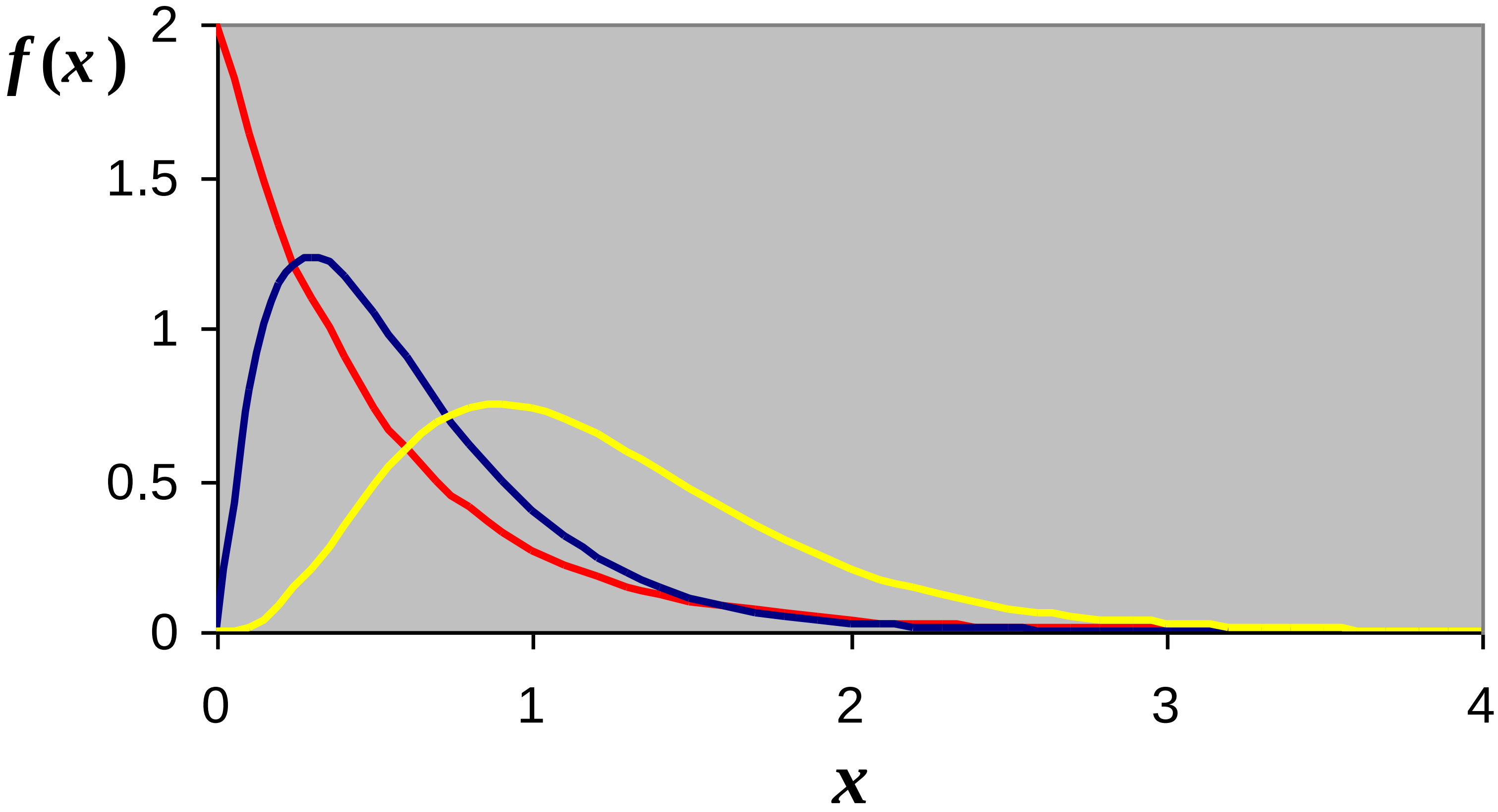
- The CDF does not exist in closed form unless α is a positive integer (Erlong distribution).
- In this case, $\Gamma(\alpha) = (\alpha - 1)!$

$$F(x) = 1 - e^{-\frac{x}{\beta}} - \frac{x}{\beta} e^{-\frac{x}{\beta}} - \frac{1}{2!}\left(\frac{x}{\beta}\right)^2 e^{-\frac{x}{\beta}} - \cdots$$
$$- \frac{1}{(\alpha-1)!}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}}$$
$$\text{for } x \geq 0$$

$$F(x) = 0 \quad \text{for } x < 0$$

Exponential and Gamma

- Examples: Exp(0.5), G(2, 0.3), G(4, 0.3)



Class Problems on Continuous Distributions

2. The time for a message to reach www.microsoft.com from CityU is distributed according to an gamma distribution with $\alpha = 3$ and $\beta = 100$ ms. What is the probability that a message will reach within 600 ms? What is the probability it will take more than 900 ms?

Let X be the time for a message to reach the website

$$X \sim \text{Gamma}\ (\alpha = 3, \beta = 100)$$

$$P(X \le 600) = F(600)$$

$$= 1 - e^{-\frac{600}{100}} - \frac{600}{100} e^{-\frac{600}{100}} - \frac{1}{2!}\left(\frac{600}{100}\right)^2 e^{-\frac{600}{100}}$$

$$= 1 - e^{-6} - 6e^{-6} - 18e^{-6}$$

$$= 1 - 25e^{-6}$$

Class Problems on Continuous Distributions

$X \sim \text{Gamma}\,(\alpha = 3, \beta = 100)$

$$P(X > 900) = 1 - P(X \leq 900) = 1 - F(900)$$

$$= 1 - \left(1 - e^{-9} - 9e^{-9} - 40.5e^{-9}\right)$$

$$= 50.5e^{-6}$$

Class Problems on Continuous Distributions

3. The lifetime of a machine tool is according to a Gamma distribution with mean 12 hours and standard deviation 6 hours. What is the probability that a machine tool would live longer than 24 hours?

Let X be the lifetime of a machine tool

$$X \sim \text{Gamma}\,(\alpha, \beta)$$

$$E(X) = \alpha\beta = 12, Var(X) = \alpha\beta^2 = 36$$

$$\Rightarrow \alpha = 4, \beta = 3$$

$$P(X > 24) = 1 - F(24)$$

$$= 1 - e^{-8} - 8e^{-8} - \frac{1}{2!}8^2e^{-8} - \frac{1}{3!}8^3e^{-8}$$

$$= e^{-8}\left(41 + \frac{512}{6}\right)$$

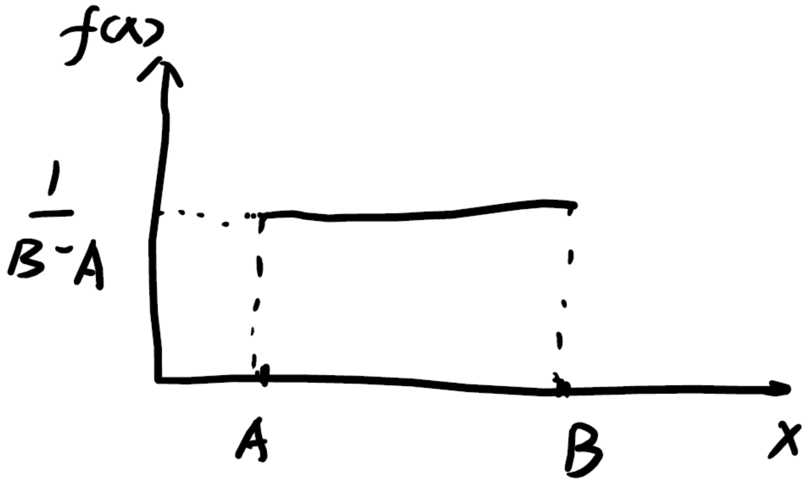
Uniform Distribution

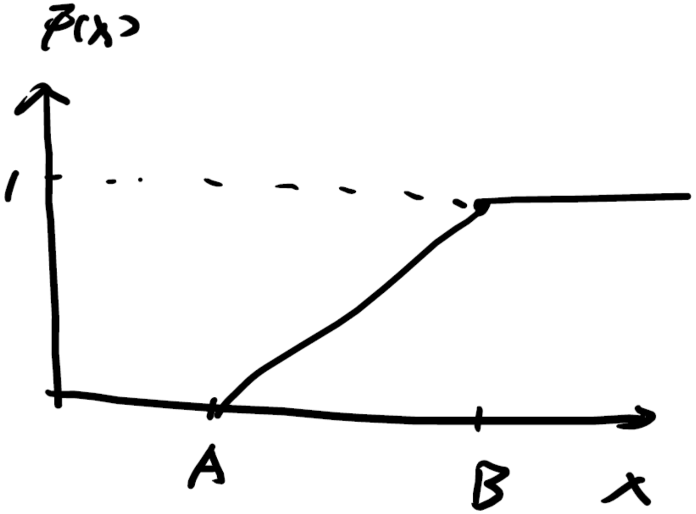
- A continuous r.v. X with a "rectangular" p.d.f. over a specified range $[A, B]$
  - The p.d.f. and c.d.f depend on parameters A and B

$$f(x) = \begin{cases} 0, & x < A \\ \dfrac{1}{B - A}, & A \leq x \leq B \\ 0, & x > B \end{cases}$$

$$F(x) = \begin{cases} 0, & x < A \\ \dfrac{x - A}{B - A}, & A \leq x \leq B \\ 1, & x > B \end{cases}$$

Uniform Distribution





Class Problems on Continuous Distributions

4. The amount of time you have to wait for a bus is uniformly distributed between 0 and 15 minutes. What is the probability you will wait for less than 10 minutes for a bus?

Let X be the waiting time for a bus

$$X \sim \text{Uniform}\ (A = 0, B = 15)$$

$$P(X < 10) = F(10) = \frac{10 - 0}{15 - 0} = \frac{2}{3}$$

Normal Distribution

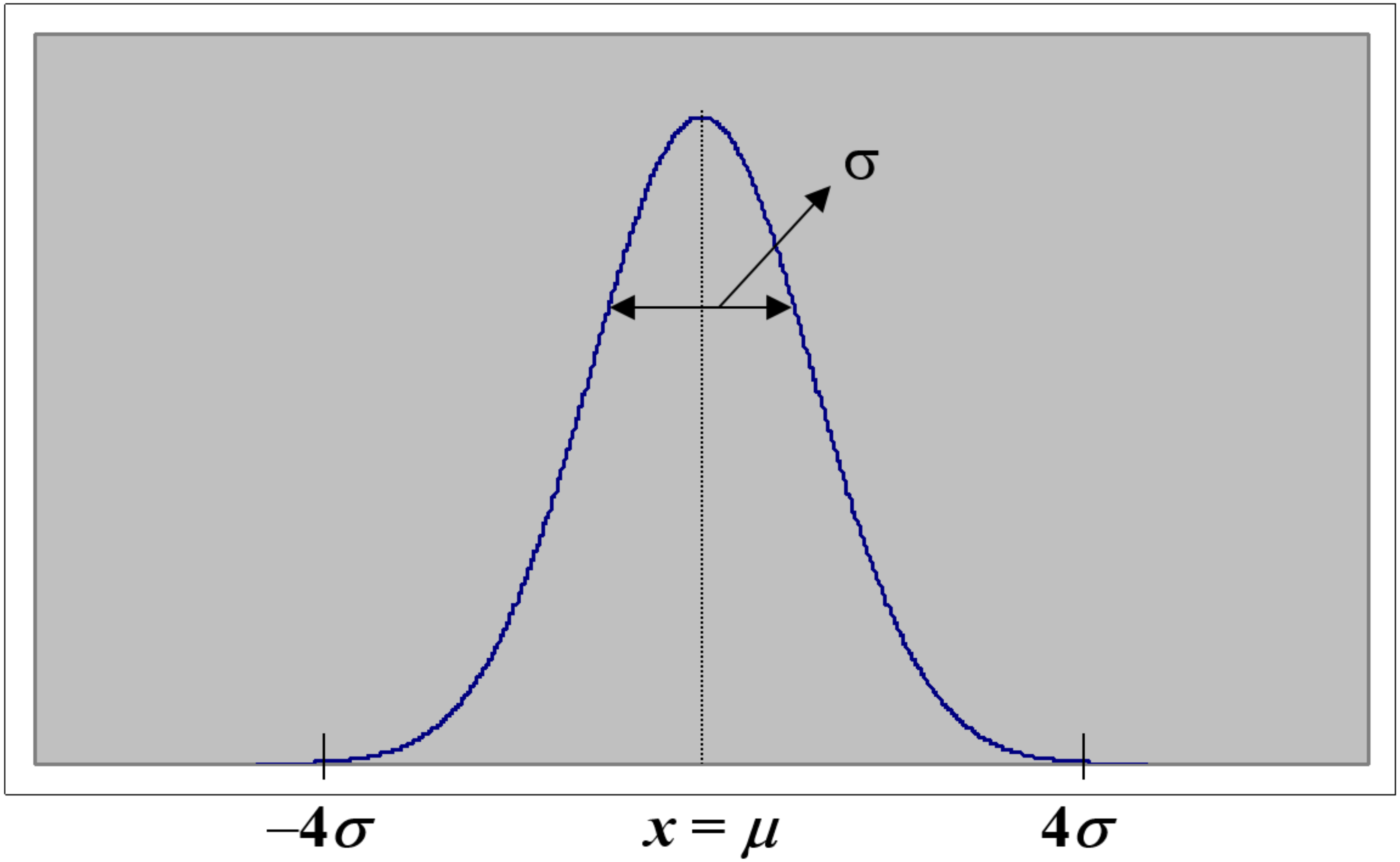
- A continuous r.v. *X* with a "bell curve" PDF
  - The most important distribution in statistics
  - The PDF depends on mean μ and variance σ^2

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

for $-\infty < x < \infty$

Normal Distribution

- $N(\mu, \sigma^2)$



Standard Normal Distribution

- Let X be distributed $N(\mu, \sigma^2)$
- Then $Z = \dfrac{X - \mu}{\sigma}$ is distributed $N(0, 1)$.
  - The PDF for Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad \text{for } -\infty < z < \infty$$

  - The CDF for Z is tabulated in a table

$$\Phi(z) = P[Z \leq z] \quad \text{for } -\infty < z < \infty$$

Table of CDF for Standard Normal Distribution

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt$$

Cumulative Distribution Function Values for Standard Normal

z	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55966	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774

How to Read the Table

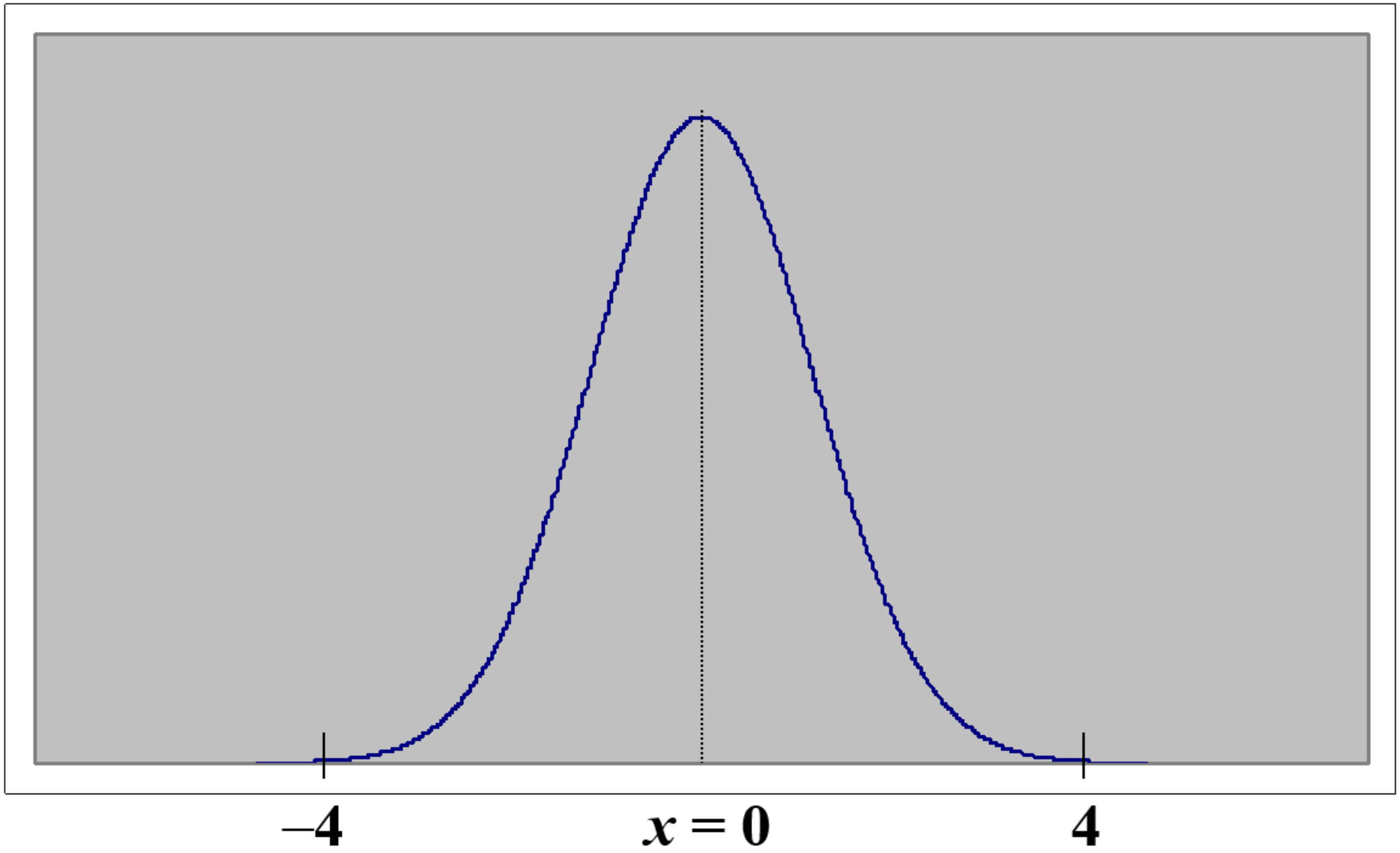
z	0.00	0.01	0.02	0.03
0.0	0.5000	0.5040	0.5080	0.5120
0.1	0.5398	0.5438	0.5478	0.5517
$\vdots$				
1.2	0.8849	0.8869	0.8888	0.8907
1.3	0.9032	0.9049	0.9066	0.9082

$$P[Z \leq 1.32] = \Phi(1.32) = 0.9066$$

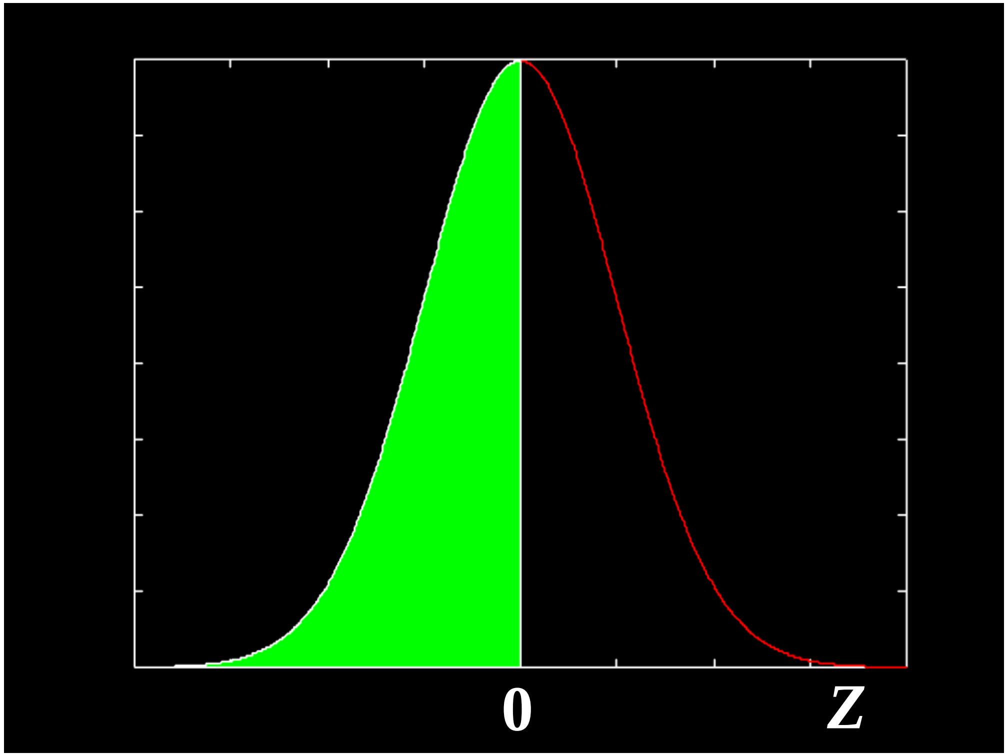
$$P[0.13 \leq Z \leq 1.31] = \Phi(1.31) - \Phi(0.13) = 0.9049 - 0.5517$$

Standard Normal Distribution

- $N(0, 1)$



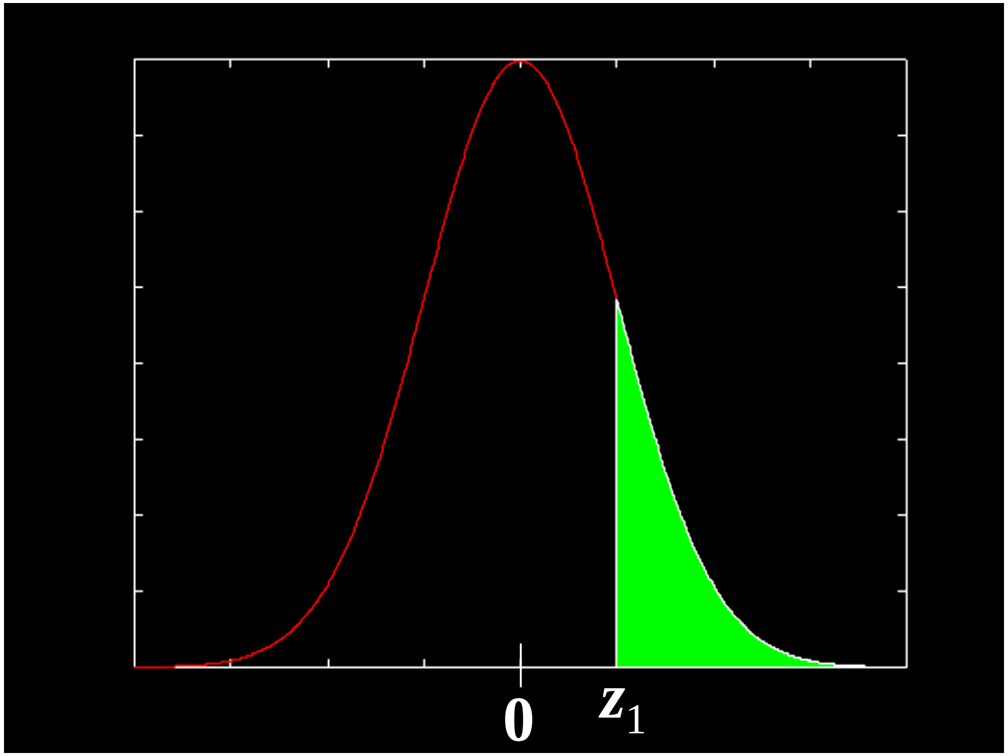
Properties of Standard Normal



$$P(Z \leq 0) = P(Z \geq 0) = 0.5$$

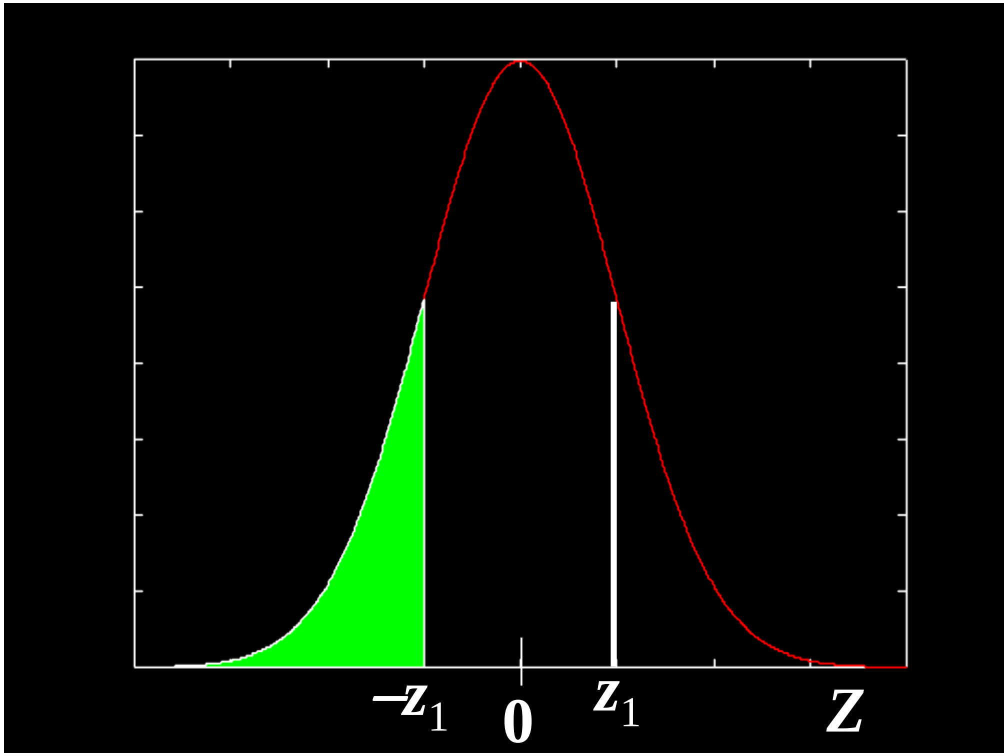
$$\Phi(0) = 0.5$$

Properties of Standard Normal



$$P(Z > z_1) = 1 - P(Z \le z_1)$$
$$= 1 - \Phi(z_1)$$

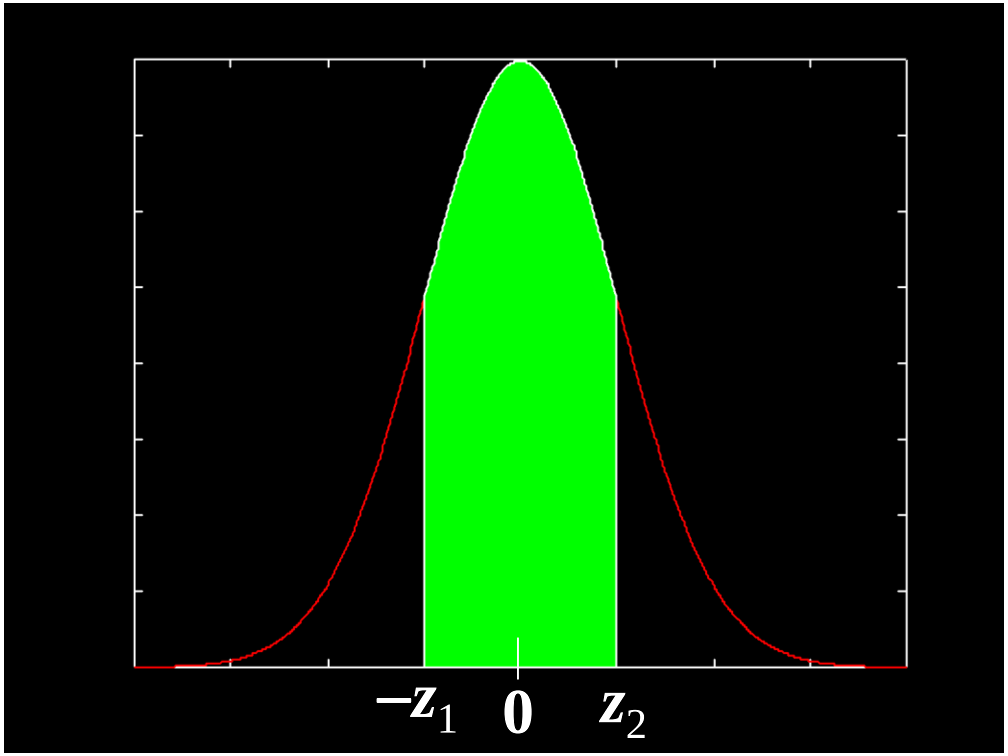
Properties of Standard Normal



$$P(Z \le -z_1) = P(Z > z_1) = 1 - P(Z \le z_1)$$

$$\Phi(-z_1) \quad = \quad 1 - \Phi(z_1)$$

Properties of Standard Normal



$$
\begin{aligned}
P(-z_1 \le Z \le z_2) &= P(Z \le z_2) - P(Z \le -z_1) \\
&= P(Z \le z_2) - (1 - P(Z \le z_1) \\
&= \Phi(z_2) \qquad - (1 - \quad \Phi(z_1)\ )
\end{aligned}
$$

Normal Distribution Probabilities

- Let X be distributed $N(\mu, \sigma^2)$

1) Standardize $X \rightarrow Z = \dfrac{X-\mu}{\sigma}$

2) Convert range for X

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

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

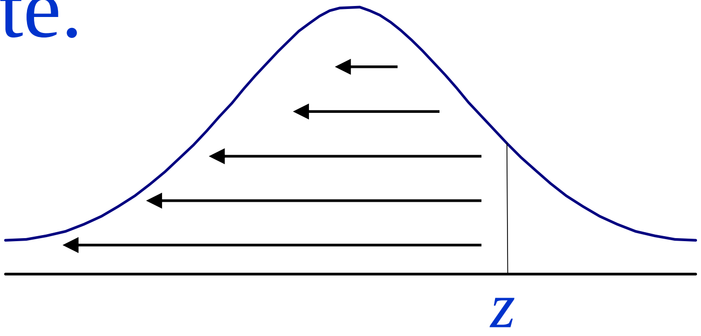
Normal Distribution Probabilities

- Let X be distributed $N(\mu, \sigma^2)$
  3) Rewrite using standard normal CDF

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

  4) Look up CDF probabilities for z_1 and z_2 in the table, then calculate.

$$z_1 = \frac{a-\mu}{\sigma}; \quad z_2 = \frac{b-\mu}{\sigma}$$



Class Problems on Normal Distribution

5. For a normal random variable X with μ = 30 and σ = 10, find the probability that X takes a value between 15 and 35.

$$X \sim N(\mu = 30, \sigma = 10)$$

$$P(15 < X < 35) = P\left(\frac{15-30}{10} < \frac{X-30}{10} < \frac{35-30}{10}\right)$$

$$= P(-1.5 < Z < 0.5)$$

$$= \Phi(0.5) - \Phi(-1.5)$$

$$= \Phi(0.5) - [1 - \Phi(1.5)]$$

$$= 0.69146 - [1 - 0.93319]$$

$$= 0.62465$$

Class Problems on Normal Distribution

6. The time for you to perform a task for a customer is a normal random variable X with $\mu = 60$ minutes and $\sigma = 15$ minutes.

   a. What time would you quote to the customer to be 90% sure of completing the task by then?
   b. The customer wants it earlier and is willing to take a chance to come x minutes from now even if there is a 66% chance of not completing the task by then. What is the value of x?

Let X be the time to perform a task

$$X \sim N(\mu = 60,\ \sigma = 15)$$

a. $P(X \leq x) = 0.9$

$$P\left(\frac{X-60}{15} \leq \frac{x-60}{15}\right) = 0.9 \Rightarrow \Phi\left(\frac{x-60}{15}\right) = 0.9$$

Class Problems on Normal Distribution

z	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55966	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147

$$\Phi(1.285) = 0.9 \Rightarrow \frac{x-60}{15} = 1.285$$

$$x = 1.285 \times 15 + 60$$

Class Problems on Normal Distribution

6. The time for you to perform a task for a customer is a normal random variable X with $\mu = 60$ minutes and $\sigma = 15$ minutes.

   a. What time would you quote to the customer to be 90% sure of completing the task by then?

   b. The customer wants it earlier and is willing to take a chance to come x minutes from now even if there is a 66% chance of not completing the task by then. What is the value of x?

b. $P(X > x) = 0.66$

$$P\left(\frac{X-60}{15} > \frac{x-60}{15}\right) = 0.66$$

$$\Rightarrow P\left(Z > \frac{x-60}{15}\right) = 0.66$$

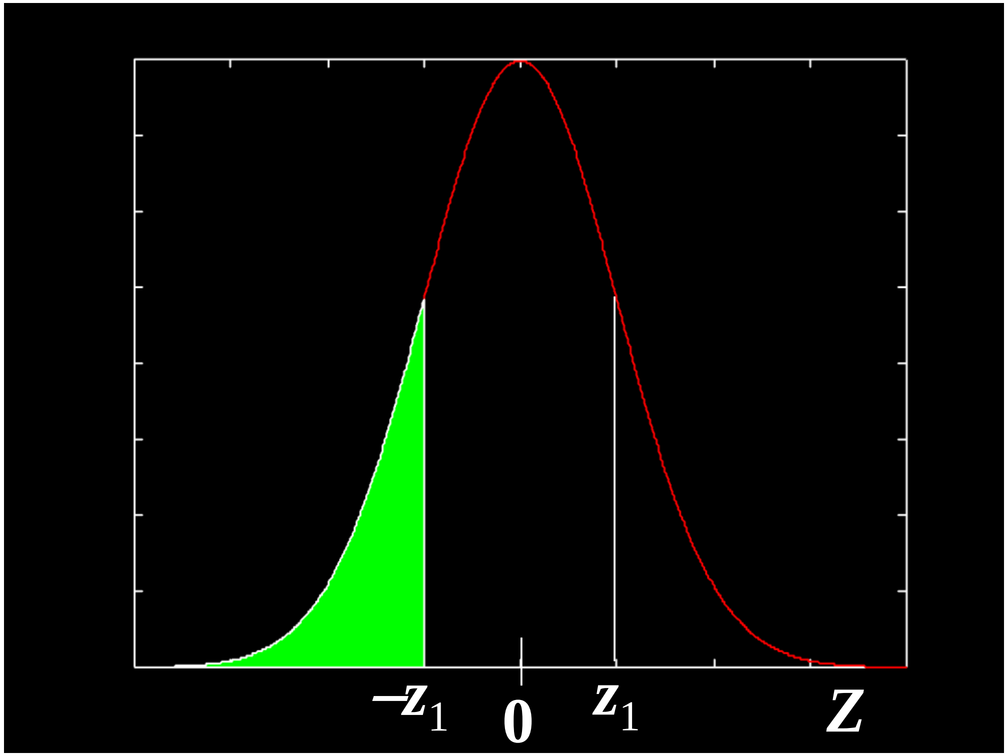
$$\Rightarrow 1 - \Phi\left(\frac{x-60}{15}\right) = 0.66 \Rightarrow \Phi\left(\frac{x-60}{15}\right) = 0.34$$

Class Problems on Normal Distribution

z	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55966	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147

$$\Phi(?) = 0.34$$

Class Problems



$$\Phi\left(\frac{x-60}{15}\right) = 0.34 \Rightarrow \Phi\left(-\frac{x-60}{15}\right) = 1 - 0.34 = 0.66$$

Class Problems

z	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55966	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793

$$\Phi(0.415) = 0.66$$

$$\Rightarrow -\frac{x - 60}{15} = 0.415$$

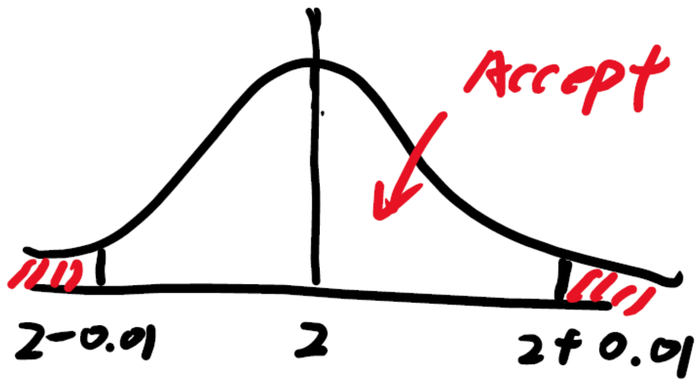
$$x = -0.415 \times 15 + 60$$

Class Problems

7. An industry manufactures pistons for IC engines. The buyer sets specifications on the diameter to be 2.0 +/- 0.01 inches. The implication is that no part falling outside these specifications will be accepted. It is known that in the process, the diameter of a piston has a normal distribution with mean 2.00 and standard deviation 0.005 inches. On an average how many pistons will be scrapped?

Let X be the diameter of a piston

$$X \sim N(\mu = 2, \sigma = 0.005)$$



$$P(scrapped) = 1 - P(accept)$$

$$= 1 - P(1.99 < X < 2.01)$$

$$= 1 - P\left(\frac{1.99 - 2}{0.005} < Z < \frac{2.01 - 2}{0.005}\right)$$

$$= 1 - [\Phi(2) - \Phi(-2)] = 1 - [\Phi(2) - 1 + \Phi(2)]$$

$$= 2[1 - \Phi(2)] = 2(1 - 0.97725) = 0.0455$$

Class Problems

8. For the above problem (number 7), what should the buyer's specs be (2.0 +/- d inches) so that 95% of the manufactured pistons are accepted (not scrapped)?

$$P(2-d<X<2+d)=0.95$$

$$\Rightarrow P\left(\frac{2-d-2}{0.005}<Z<\frac{2+d-2}{0.005}\right)=0.95$$

$$\Rightarrow P\left(-\frac{d}{0.005}<Z<\frac{d}{0.005}\right)=0.95$$

$$\Rightarrow \Phi\left(\frac{d}{0.005}\right)-\left[1-\Phi\left(\frac{d}{0.005}\right)\right]=0.95$$

$$\Rightarrow \Phi\left(\frac{d}{0.005}\right)=0.975 \Rightarrow \frac{d}{0.005}=1.96 \Rightarrow d=\cdots$$