
Normal Distribution

a.k.a. Gaussian Distribution

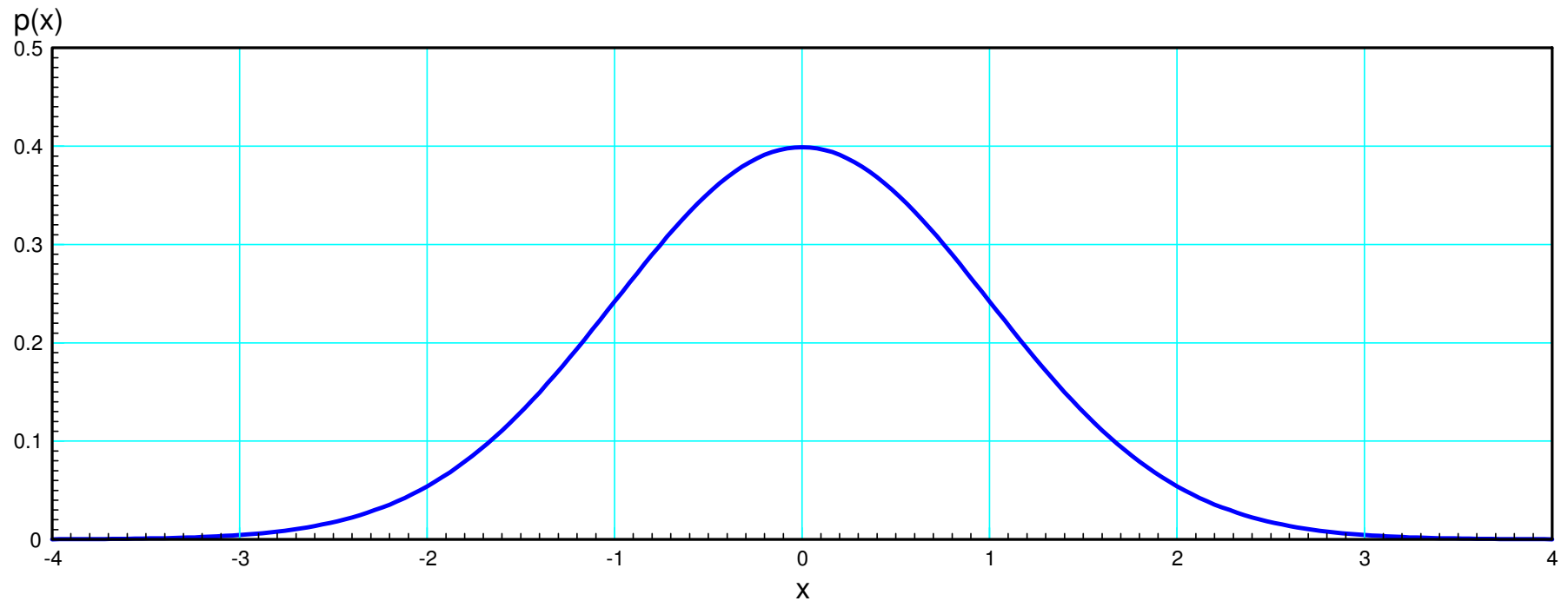
ECE 341: Random Processes

Lecture #17

note: All lecture notes, homework sets, and solutions are posted on www.BisonAcademy.com

Normal Distribution

- a.k.a. Gaussian Distribution
- Probably the most important probability distribution
- Class averages, the height of people, IQ scores, etc.



pdf for a normal distribution with a mean of 0, variance of 1

Central Limit Theorem (coming soon)

- If you sum random variables together (with certain loose restrictions), the resulting distribution converges to a Normal distribution
- If you add to random variables together that have a Normal distribution, the result will have a Normal distribution
- Everything converges to a Normal distribution.
- Once you arrive at a Normal distribution, you're stuck with a Normal distribution.

This is part of the reason Normal distributions are so common.

pdf and mgf:

A Normal distribution is expressed as

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

read as

X has a Normal distribution with mean μ and variance σ^2

The pdf for a Normal distribution is

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

while the moment generating function

$$\psi(s) = \exp\left(\mu s + \frac{\sigma^2 s^2}{2}\right)$$

From the moment generating function you can see that

$$m_0 = \psi(0) = 1$$

This is a valid moment generating function

The mean of a Normal distribution is the 1st moment:

$$m_1 = \psi'(0) = \left((\mu + \sigma^2 s) \exp\left(\mu s + \frac{\sigma^2 s^2}{2}\right) \right)_{s=0} = \mu$$

The second moment and variance are

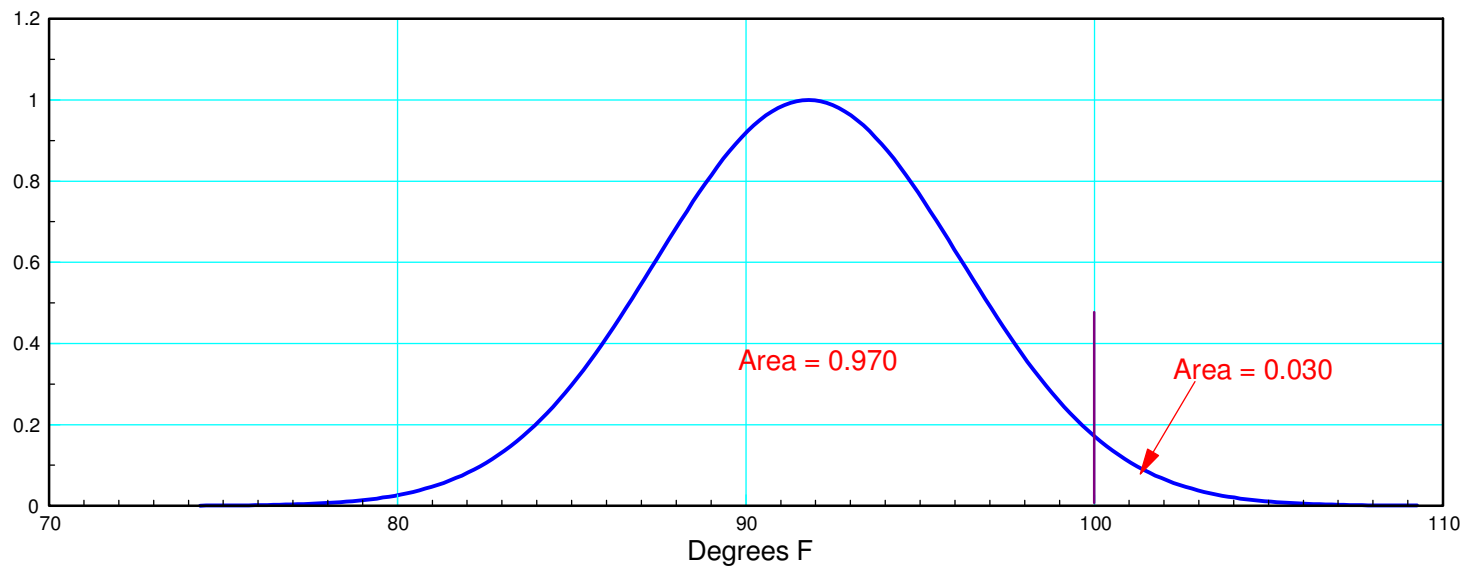
$$\begin{aligned} m_2 &= \psi''(0) \\ &= \left(\sigma^2 \exp\left(\mu s + \frac{\sigma^2 s^2}{2}\right) + (\mu + \sigma^2 s)^2 \exp\left(\mu s + \frac{\sigma^2 s^2}{2}\right) \right)_{s=0} = \sigma^2 + \mu^2 \end{aligned}$$

$$var = m_2 - m_1^2 = \sigma^2$$

Examples of Normal Distributions

Monthly High Temperature:

- June, Fargo, ND (source: Hector Airport)
- mean = 91.7962F
- standard deviation = 4.3657



Normalized pdf for the high for the month of June, Fargo ND

What is the probability that the high will be more than 100F this coming year?

- This is asking what the area of the curve is to the right of 100F. Here, StatTrek is useful:
- $p(\text{high} < 100F) = 0.970$

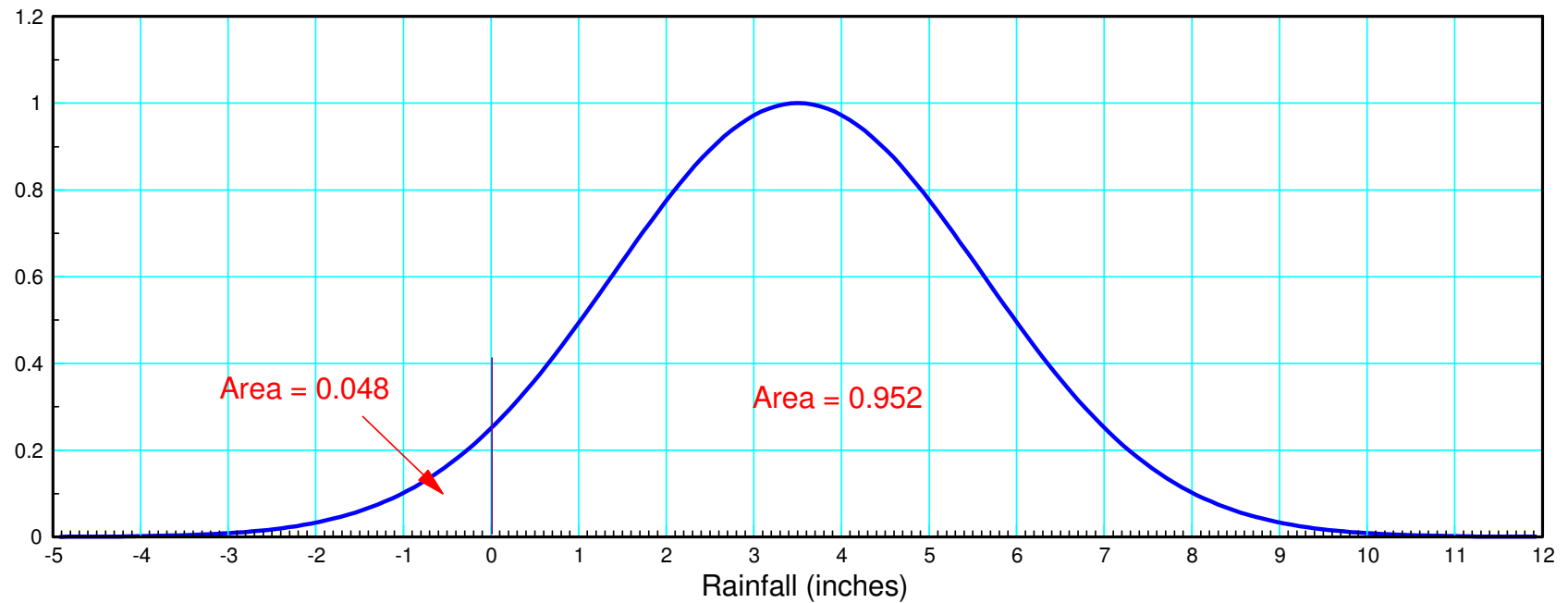
- Enter a value in three of the four text boxes.
- Leave the fourth text box blank.
- Click the **Calculate** button to compute a value for the blank text box.

Normal random variable (x)	<input type="text" value="100"/>
Cumulative probability: $P(X \leq 100)$	<input type="text" value="0.970"/>
Mean	<input type="text" value="91.79"/>
Standard deviation	<input type="text" value="4.36"/>

<https://www.stattrek.com/online-calculator/normal.aspx>

Rain Fall: (actually a Poisson distribution)

- mean = 3.5025 inches
- standard deviation = 2.1054



pdf for the rainfall in June

What is the chance we will receive no rain this coming June?

- Sticking with the Normal distribution, the area to the left of 0" is 0.048. There is 4.8% chance we will get no rain this coming June.

- Enter a value in three of the four text boxes.
- Leave the fourth text box blank.
- Click the **Calculate** button to compute a value for the blank text box.

Normal random variable (x)	0
Cumulative probability: $P(X \leq 0)$	0.048
Mean	3.5025
Standard deviation	2.105

<https://www.stattek.com/online-calculator/normal.aspx>

Addition of Normal Distributions:

If you add to Normal distributions, the result is a Normal distribution with

- mean = mean(a) + mean(b)
- variance = variance(a) + variance(b)

Proof: Multiply the moment generating functions

$$\psi_a(s) = \exp\left(\mu_a s + \frac{\sigma_a^2 s^2}{2}\right) \quad \psi_b(s) = \exp\left(\mu_b s + \frac{\sigma_b^2 s^2}{2}\right)$$

$$\psi_a(s)\psi_b(s) = \exp\left(\mu_a s + \frac{\sigma_a^2 s^2}{2}\right) \exp\left(\mu_b s + \frac{\sigma_b^2 s^2}{2}\right)$$

$$= \exp\left(\left(\mu_a s + \frac{\sigma_a^2 s^2}{2}\right) + \left(\mu_b s + \frac{\sigma_b^2 s^2}{2}\right)\right)$$

$$= \exp\left((\mu_a + \mu_b)s + \frac{(\sigma_a^2 + \sigma_b^2)s^2}{2}\right)$$

Example: Rainfall in Fargo

The rainfall for the months of June, July, and August have the following statistics:

- What is the pdf for the entire summer?
- Jule + July + August?

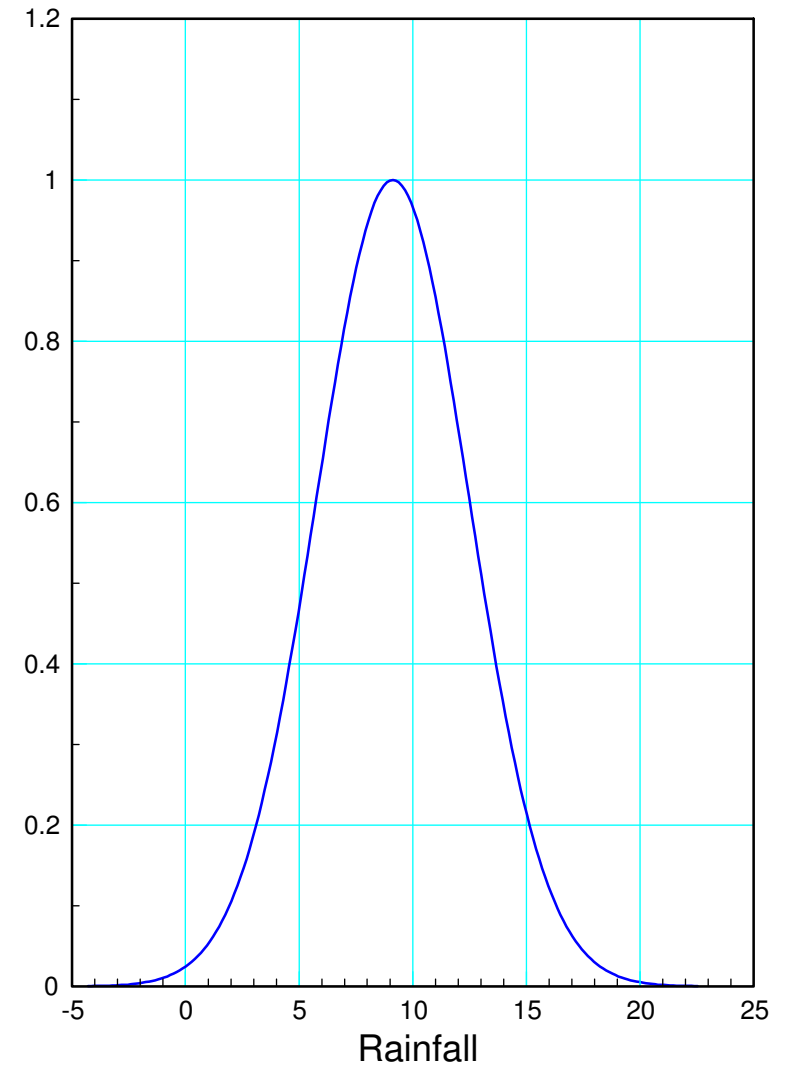
Month	Mean	Variance
June	3.5025	4.4327
July	2.9668	3.8044
August	2.6529	3.0063
Summer Total	?	?

Rainfall in Fargo:

Solution:

- The means add
- The variance adds

Month	Mean	Variance
June	3.5025	4.4327
July	2.9668	3.8044
August	2.6529	3.0063
Summer Total	9.1221	11.2434



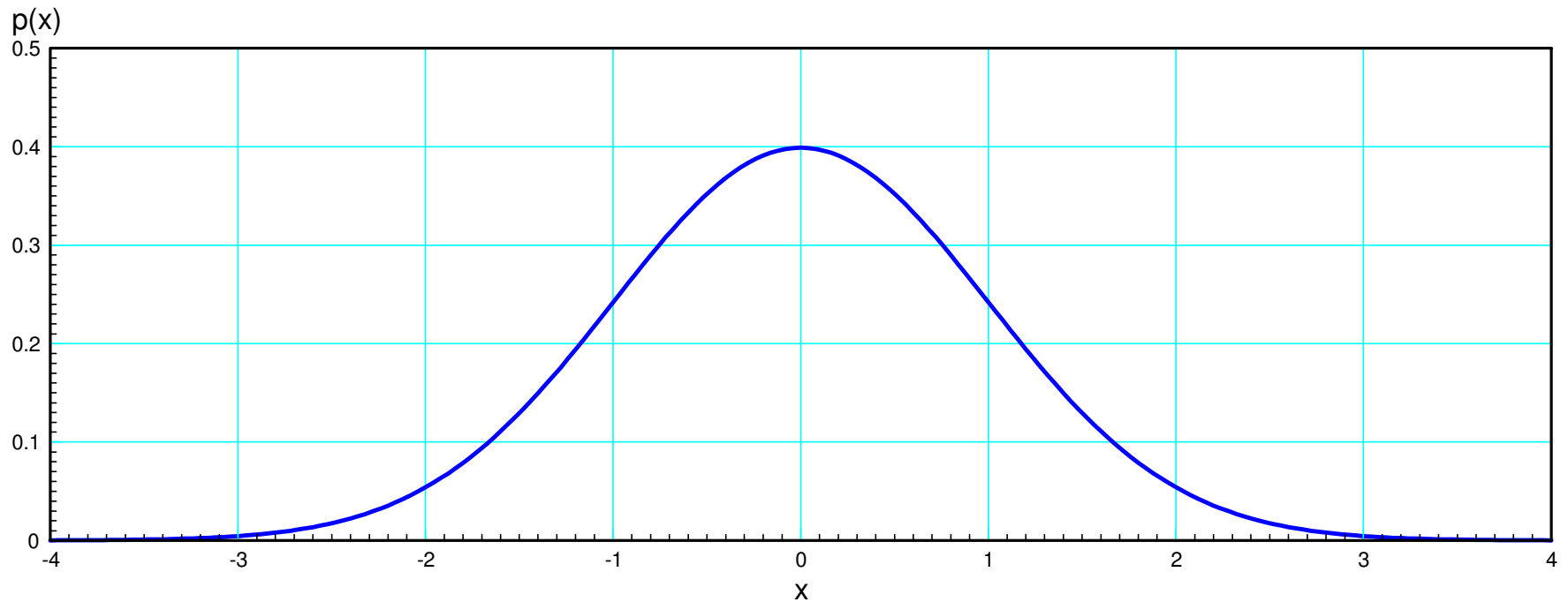
Standard Normal Distribution

- The Normal distribution is extremely common and extremely useful. Unfortunately, it is difficult to integrate.
 - To get around this, tables showing the area under the curve as you move away from the mean could be used. Unfortunately, there are an infinite number of possible means and standard deviations.
 - To get around this, the Standard Normal Distribution is used. This is a Normal distribution with
 - mean = 0
 - standard deviation = 1
 - With this stipulation, you can come up with a table showing the area under the curve as you move away from the mean. Once known, you can easily convert this to the distribution you care about.
-

The standard normal distribution has the following pdf:

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

It's shape is a nice bell curve:



pdf for a Normal distribution with mean = 0, standard deviation = 1 (a.k.a. a Standard Normal Distribution)

The area under the tail as you move away from the mean is then

z	0	-1	-2	-3	-4	-5
p(x<z)	0.5	0.1587	0.0227	0.001349	3.167 10 ⁻⁵	2.866 10 ⁻⁷

z	-1.28	-1.64	-1.96	-2.24	-2.33	-2.58
p(x<z)	0.1	0.05	0.025	0.0125	0.01	0.005

The z-score is how far your value is from the mean in terms of standard deviations

$$z = \left(\frac{x - \mu}{\sigma} \right)$$

This converts your data (x) into a standard Normal distribution (z).

- Once you have the z-score, you can find the probability using Matlab (or StatTrek)

```
>> z = -1;  
>> tail = (1 - erf(-z/sqrt(2)))/2  
tail =    0.158655253931457
```

```
>> z = -2;  
>> tail = (1 - erf(-z/sqrt(2)))/2  
tail =    0.022750131948179
```

Example: Recall that the high for the month of June has

- mean = 91.7962F
- standard deviation = 4.3657

Determine the probability that it will break 100F this coming June.

Solution: Determine the z-score corresponding to 100F

$$z = \left(\frac{100F - 91.7962}{4.3658} \right) = 1.8791$$

From the table at the end, 1.8751 has an area ($p(x < z)$) of 0.0301

You can also use Matlab

```
>> z = -1.8791;  
>> tail = (1 - erf(-z/sqrt(2)))/2  
  
tail = 0.030115420076187
```

You can also use StatTrek:

- Enter a value in three of the four text boxes.
- Leave the fourth text box blank.
- Click the **Calculate** button to compute a value for the blank text box.

Standard score (z)

Cumulative probability: $P(Z \leq -1.8791)$

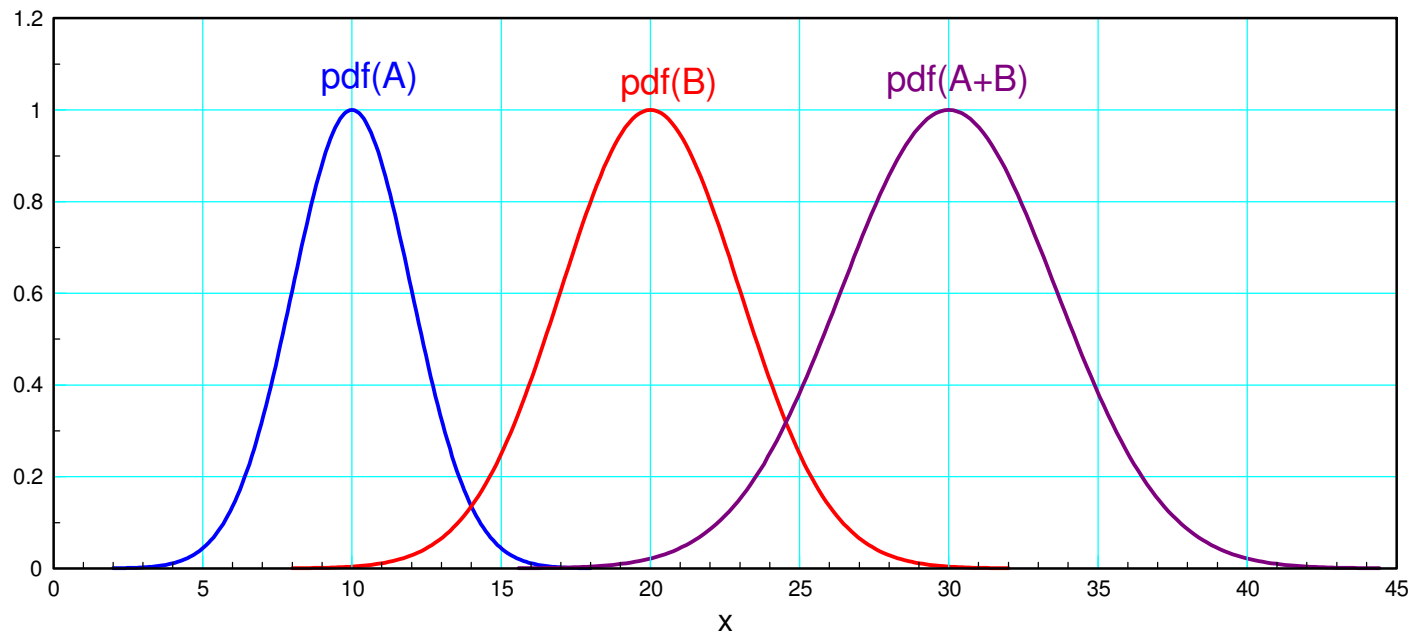
Mean

Standard deviation

Adding of Normal Distributions:

When you add normal distributions,

- The result is a normal distribution,
- The means add, and
- The variance adds

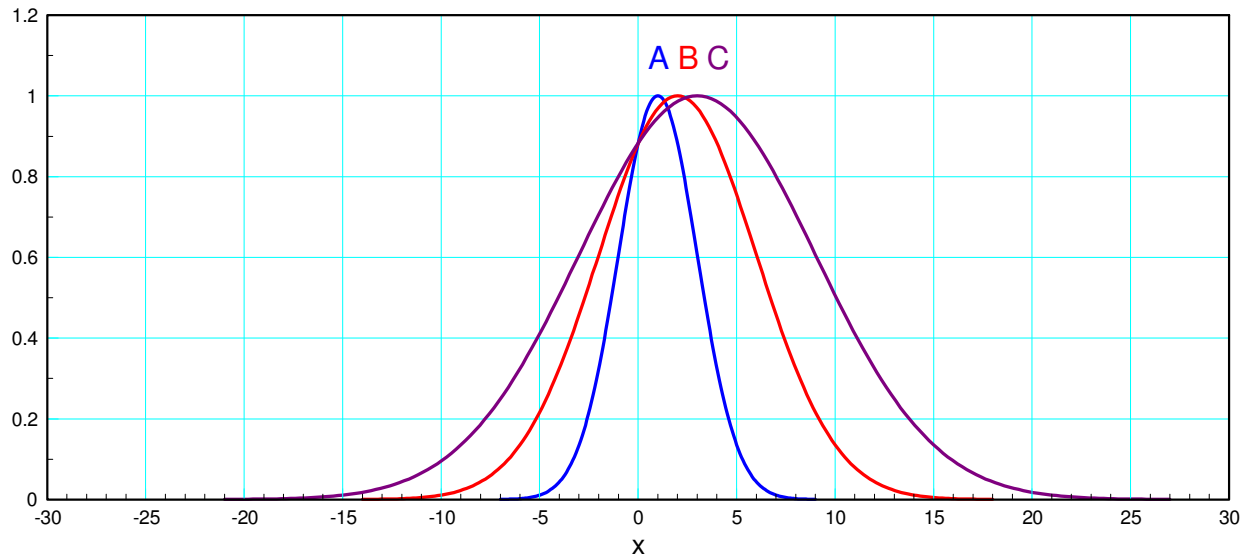


Problem: Let

- a be a normal distribution with a mean of 1 and standard deviation of 2
- b be a normal distribution with a mean of 2 and standard deviation of 4
- c be a normal distribution with a mean of 3 and standard deviation of 6

Find

- The distribution of $y = a + b + c$
- The probability that $y > 10$



Distribution of y

y will have a normal distribution.

Mean:

$$\mu_y = \mu_a + \mu_b + \mu_c$$

$$\mu_y = 1 + 2 + 3 = 6$$

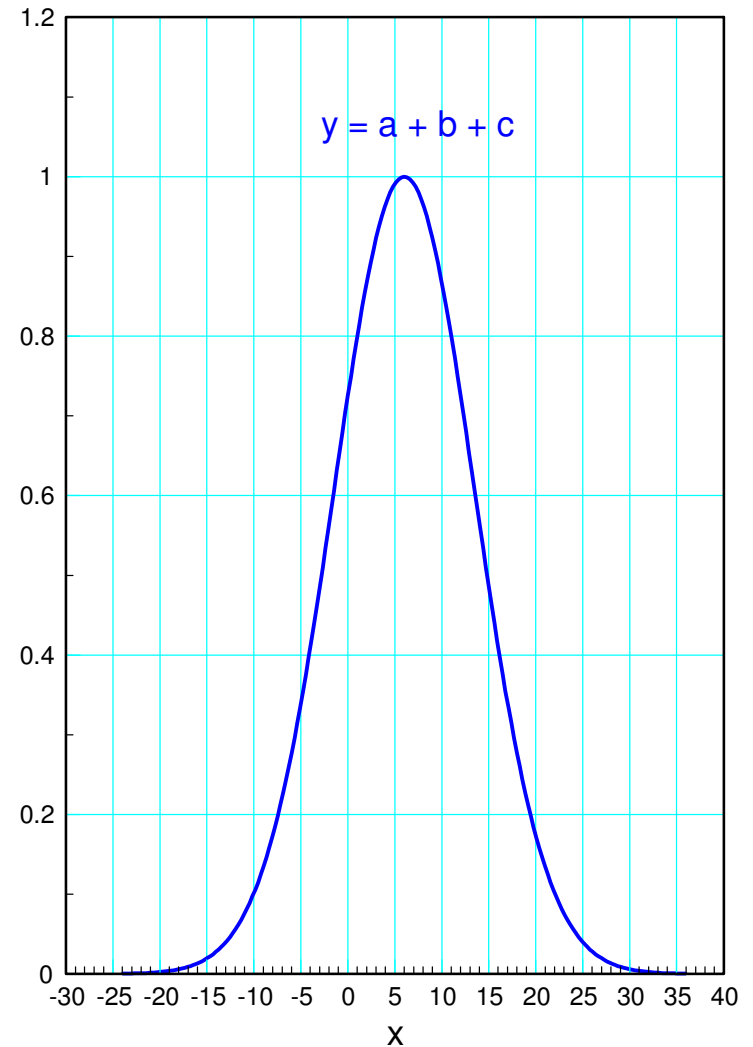
Variance:

$$\sigma_y^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2$$

$$\sigma_y^2 = 2^2 + 4^2 + 6^2 = 56$$

Standard Deviation:

$$\sigma_y = \sqrt{56}$$



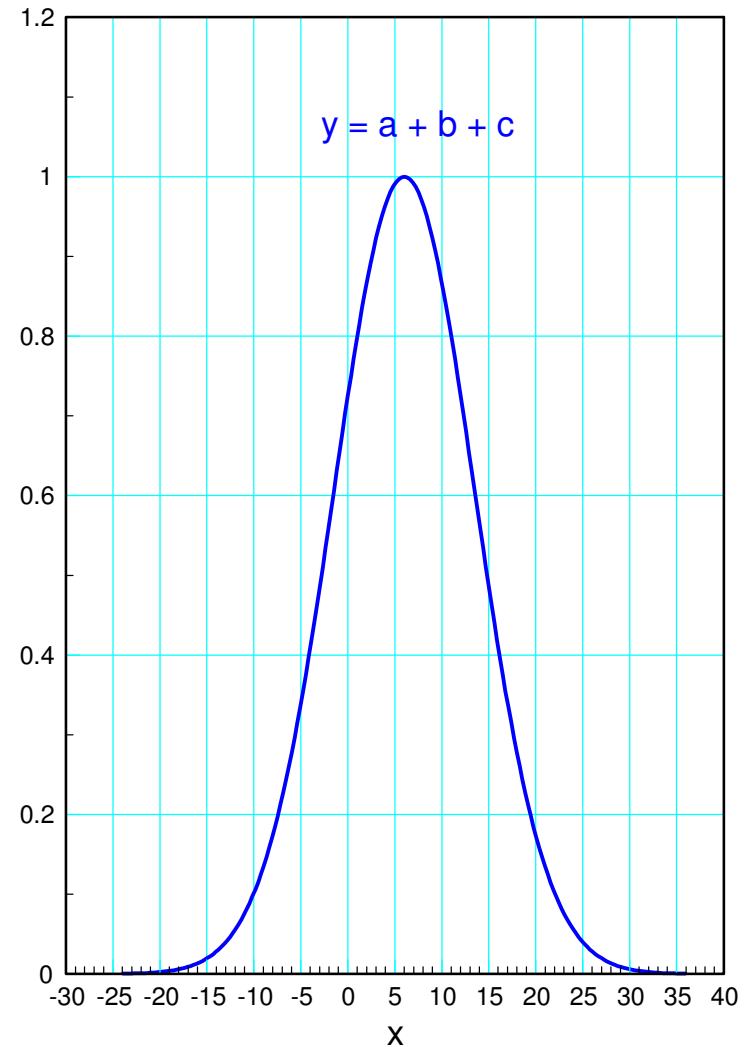
Checking with a Monte-Carlo simulation

```
n = 1e5;  
y = zeros(n,1);  
for i=1:n  
    a = 2*randn + 1;  
    b = 4*randn + 2;  
    c = 6*randn + 3;  
    y(i) = a + b + c;  
end  
disp('    mean    variance    std')  
disp([mean(y), var(y), std(y)])
```

mean	variance	std
6.0141	56.0709	7.4881

Check:

- The mean adds
- The variance adds



Probability $y > 10$:

Find the z-score

$$z = \left(\frac{10 - \mu_y}{\sigma_y} \right)$$

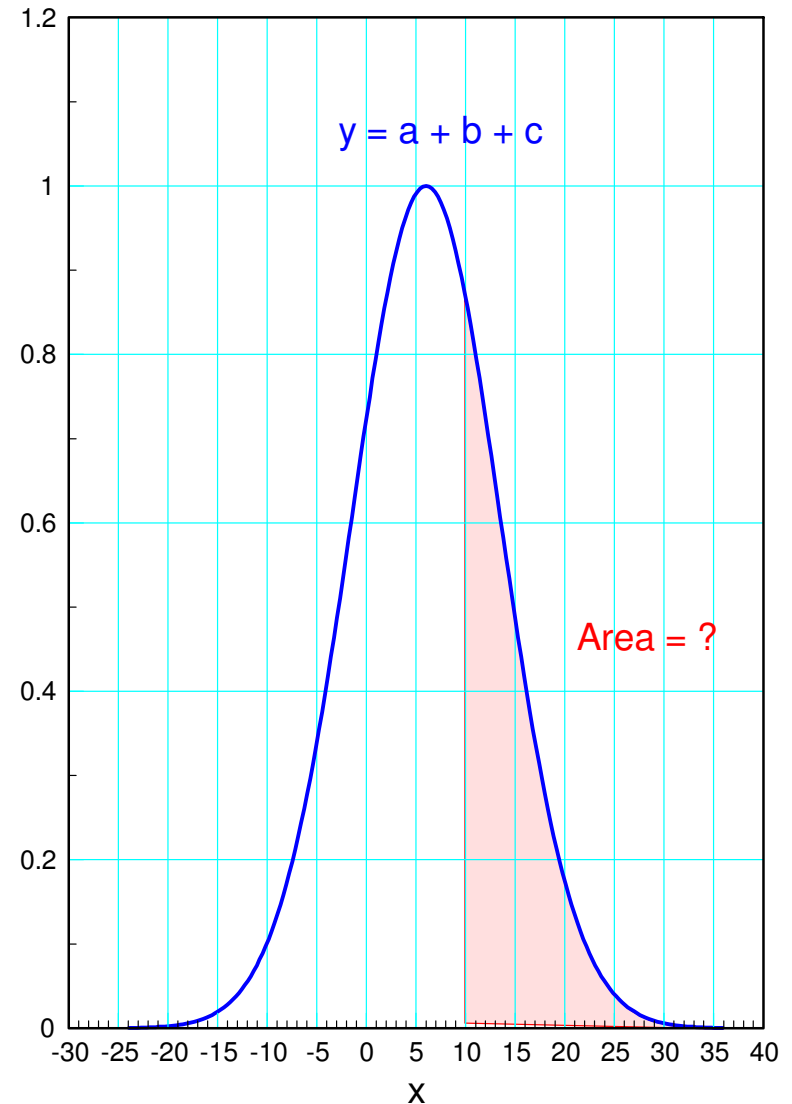
$$z = \left(\frac{10 - 6}{\sqrt{56}} \right) = 0.5345$$

Convert to a probability using a standard normal table or Matlab

```
>> z = -0.5345;  
>> tail = (1 - erf(-z/sqrt(2)))/2
```

```
tail = 0.296497824732055
```

There is a 29.65% chance that the sum will be more than 10.0



Checking with a Monte-Carlo simulation

```
n = 1e5;  
y = zeros(n,1);  
for i=1:n  
    a = 2*randn + 1;  
    b = 4*randn + 2;  
    c = 6*randn + 3;  
    y(i) = a + b + c;  
end  
disp('    mean    variance    std')  
disp([mean(y), var(y), std(y)])
```

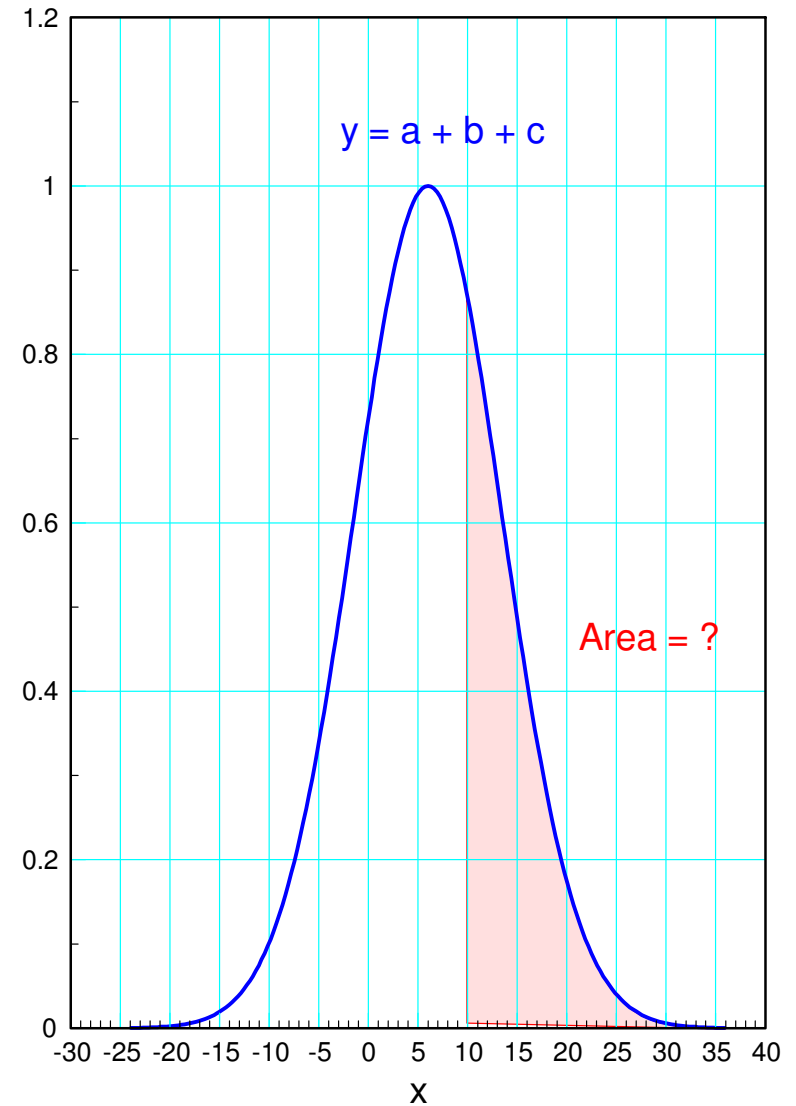
mean	variance	std
6.0141	56.0709	7.4881

```
sum(y > 10)
```

```
ans =    29818
```

$p(y > 10) = 29.818\%$

- 29.65% chance calculated



Summary

Normal distributions are fairly easy to work with

- Normal + Normal = Normal
- The means add
- The variance adds

Probabilities are defined by the distance to the mean

- z-score
- Matlab function `erf()`

$$\text{tail} = (1 - \text{erf}(-z/\text{sqrt}(2)))/2$$

With a normal table, you can determine

- The area of a tail (one-sided test), or
 - The area of two tails (two-sided test)
-

Standard Normal Table: z-score

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.025	0.0244	0.0239	0.0233
-1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.063	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1094	0.1075	0.1057	0.1038	0.102	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.123	0.121	0.119	0.117
-1	0.1587	0.1563	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.166	0.1635	0.1611
-0.8	0.2119	0.209	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.242	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2207	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2644	0.2611	0.2579	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.305	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.281	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.33	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.352	0.3483
-0.2	0.4207	0.4168	0.4129	0.4091	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0	0.5	0.496	0.492	0.488	0.4841	0.4801	0.4761	0.4721	0.4681	0.4641
