Section 4.4
Means and Variances of Random Variables

Statistics 104

Autumn 2004

Numerical Summaries of Probability Distributions

Data:

Sample mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}x_1 + \frac{1}{n}x_2 + \ldots + \frac{1}{n}x_n$$

Assume that there are k different x's, $x_{(1)}, x_{(2)}, \ldots, x_{(k)}$, where $x_{(i)}$ is observed n_i times.

$$\bar{x} = \frac{n_1}{n}x_{(1)} + \frac{n_2}{n}x_{(2)} + \ldots + \frac{n_k}{n}x_{(k)}$$

where $\frac{n_i}{n}$ is the proportion of times seeing $x_{(i)}$ in the dataset.

This suggests the following summary for the center of a discrete RV.

Mean of a discrete RV

$$\begin{aligned}\mu_x &= x_1p_1 + x_2p_2 + \ldots + x_kp_k \\ &= \sum_{i=1}^{k} x_ip_i\end{aligned}$$

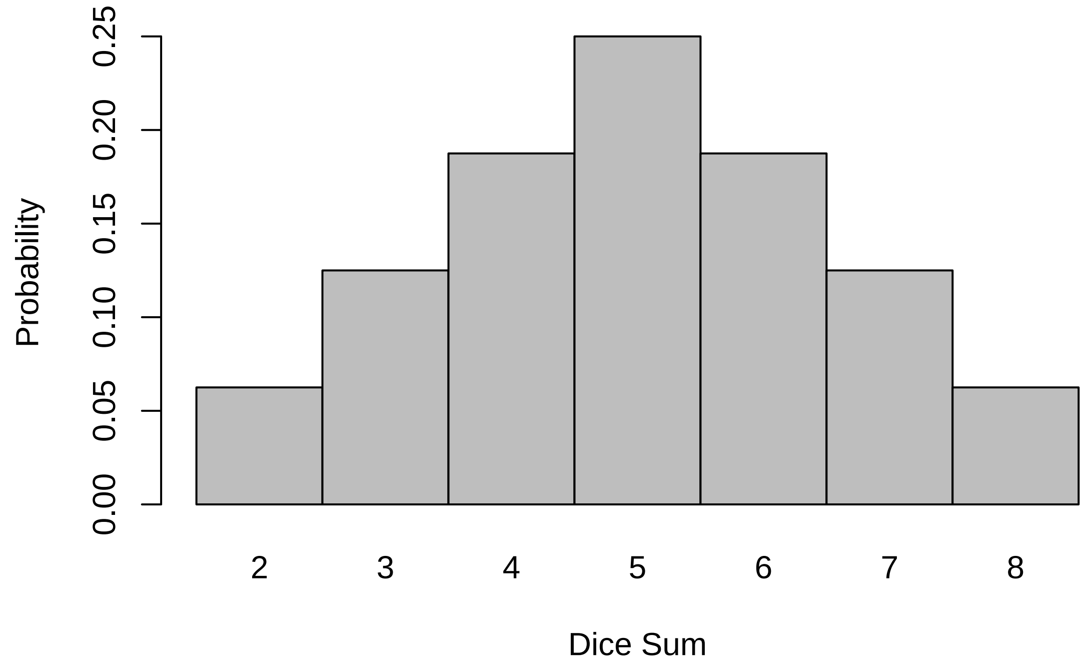
(when the sum is well defined)

μ_x describes the center of a probability distribution like $\bar{x}$ describes the center of a dataset.

Examples:

1) Sum of two 4-sided dice

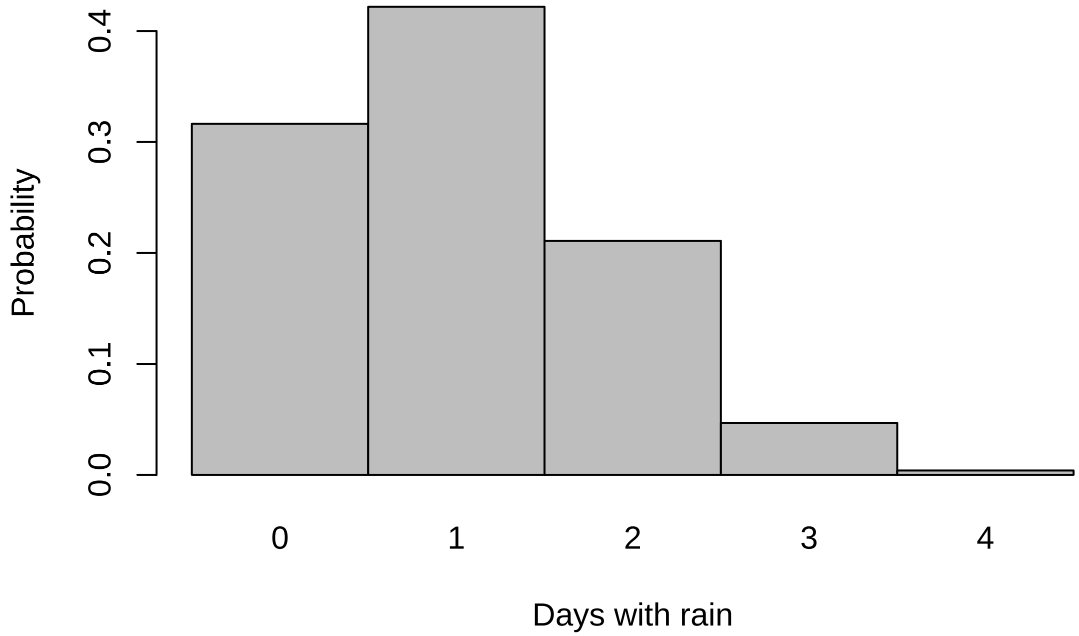
Outcome	Probability
2	0.0625
3	0.1250
4	0.1875
5	0.2500
6	0.1875
7	0.1250
8	0.0625



$$\begin{aligned}\mu_x &= 2 \times \frac{1}{16} + 3 \times \frac{2}{16} + 4 \times \frac{3}{16} + 5 \times \frac{4}{16} \\ &\quad +6 \times \frac{3}{16} + 7 \times \frac{2}{16} + 8 \times \frac{1}{16} = 5\end{aligned}$$

2) Rainfall example

x_i	p_i
0	0.3164
1	0.4219
2	0.2109
3	0.0469
4	0.0039



$$
\begin{aligned}
\mu_x &= 0 \times 0.3164 + 1 \times 0.4219 + 2 \times 0.2109 \\
&\quad +3 \times 0.0469 + 4 \times 0.0039 \\
&= 1
\end{aligned}
$$

Data:

Sample variance

$$
\begin{aligned}
s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\
&= \frac{1}{n-1}(x_1-\bar{x})^2+\frac{1}{n-1}(x_2-\bar{x})^2+\ldots+\frac{1}{n-1}(x_n-\bar{x})^2
\end{aligned}
$$

Similarly to before

$$
\bar{x}=\frac{n_1}{n-1}(x_{(1)}-\bar{x})^2+\frac{n_2}{n-1}(x_{(2)}-\bar{x})^2+\ldots+\frac{n_k}{n-1}(x_{(k)}-\bar{x})^2
$$

Variance of a discrete RV

$$\begin{aligned}\sigma_x^2 &= (x_1-\mu_x)^2 p_1 + (x_2-\mu_x)^2 p_2 + \ldots + (x_k-\mu_x)^2 p_k \\ &= \sum_{i=1}^{k}(x_i-\mu_x)^2 p_i\end{aligned}$$

Standard deviation of a discrete RV

$$\sigma_x = \sqrt{\sigma_x^2}$$

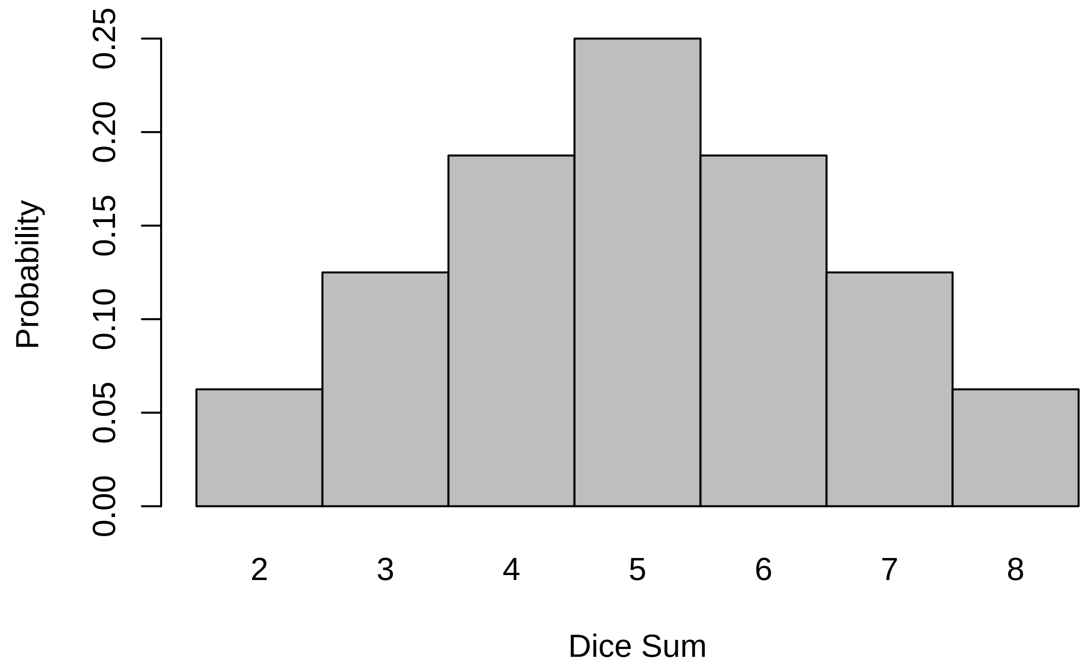
(when the sum is well defined)

σ describes the spread of a probability distribution like s describes the spread of a dataset.

Examples:

1) Sum of two 4-sided dice

Outcome	Probability
2	0.0625
3	0.1250
4	0.1875
5	0.2500
6	0.1875
7	0.1250
8	0.0625

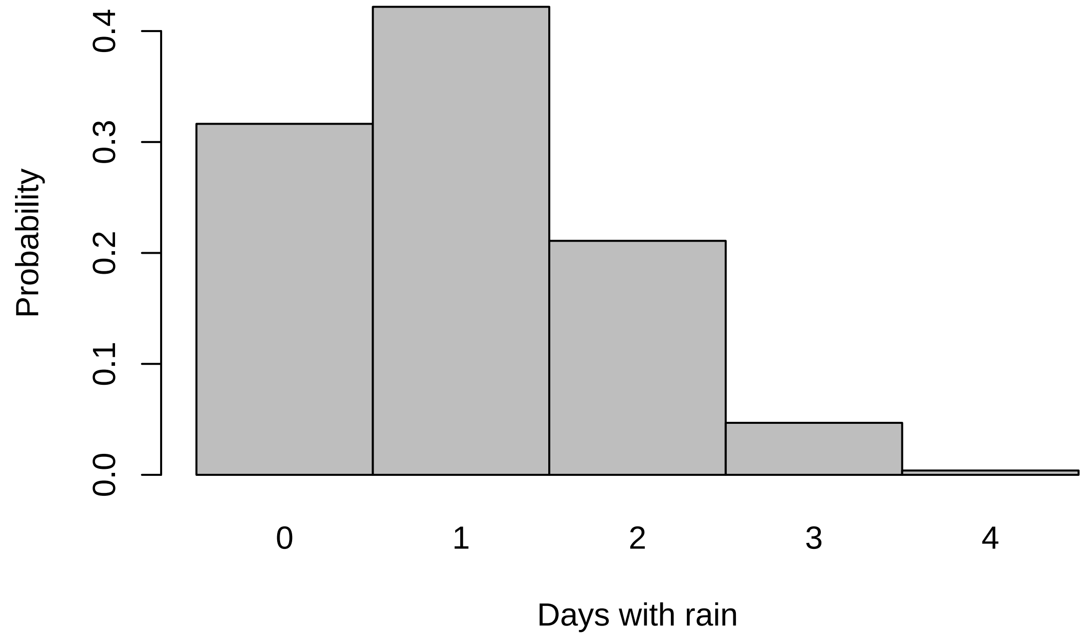


$$\begin{aligned}\sigma_x^2 &= (2-5)^2 \times \frac{1}{16} + (3-5)^2 \times \frac{2}{16} + (4-5)^2 \times \frac{3}{16} + (5-5)^2 \times \frac{4}{16} \\ &\quad +(6-5)^2 \times \frac{3}{16} + (7-5)^2 \times \frac{2}{16} + (8-5)^2 \times \frac{1}{16} = 2.5\end{aligned}$$

$$\sigma_x = \sqrt{2.5} = 1.58$$

2) Rainfall example

x_i	p_i
0	0.3164
1	0.4219
2	0.2109
3	0.0469
4	0.0039



$$
\begin{aligned}
\sigma_x^2 &= (0-1)^2 \times 0.3164 + (1-1)^2 \times 0.4219 + (2-1)^2 \times 0.2109 \\
&\quad +(3-1)^2 \times 0.0469 + (4-1)^2 \times 0.0039 \\
&= 0.75
\end{aligned}
$$

$$\sigma_x = \sqrt{0.75} = 0.866$$

Not surprisingly, μ_x, σ_x, and σ_x^2 are also defined for continuous RVs as follows

$$
\begin{aligned}
\mu_x &= \int x f(x) dx \\
\sigma_x^2 &= \int (x - \mu_x)^2 f(x) dx \\
\sigma_x &= \sqrt{\sigma_x^2}
\end{aligned}
$$

(when the integrals are defined)

Law of Large Numbers

What is the relationship between $\bar{x}$ and μ_x?

- Generate data from a probability model
- Look at $\bar{x}$ after each trial

Law of Large Numbers

Draw independent observations at random from any population with finite mean μ. Decide how accurately you would like to estimate μ. As the number of observations drawn increases, the mean $\bar{x}$ of the observed values eventually approaches the mean μ of the population as closely as you specified and then stays that close.

As the sample size increases, $\bar{x} \to \mu_x$.

Similar to the idea that long-run relative frequencies approach the true probabilities.

To be a bit more precise mathematically, the Law of Large Numbers says than

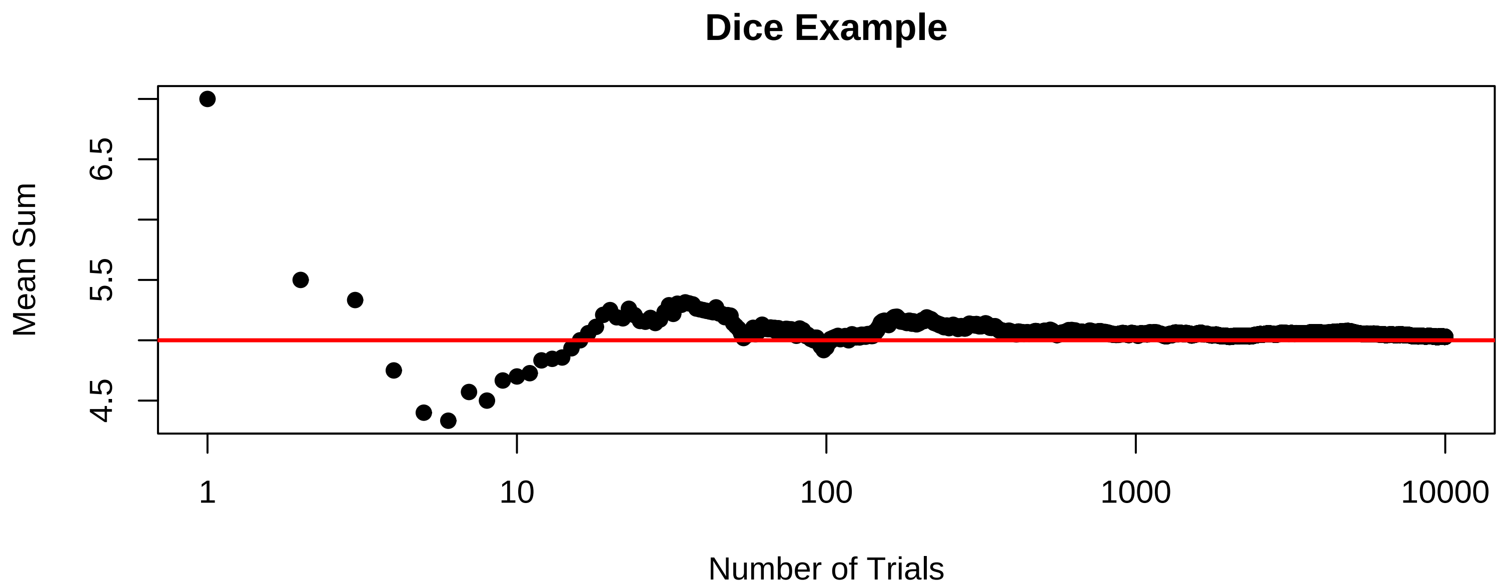
$$P[\bar{X} < \mu_x - c \text{ or } \bar{X} > \mu_x + c] \rightarrow 0$$

or equivalently

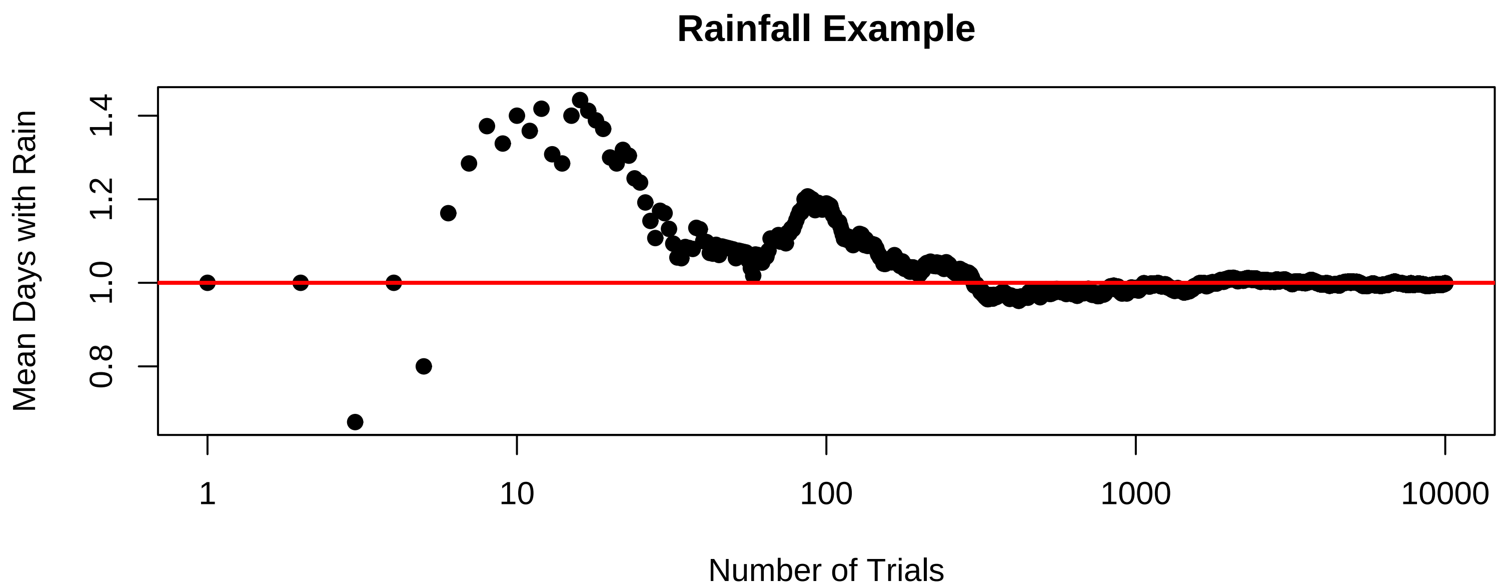
$$P[-c < \bar{X} - \mu_x < c] \rightarrow 1$$

as the sample size n goes to ∞ for any value of c. (This is the weak law of large numbers. There is also the strong law of large numbers.)

How big does n need to be for $\bar{x}$ to be close to μ? It depends on the problem. With a bigger σ, you need a bigger sample to guarantee you will get close to μ. We'll discuss this relationship in chapter 5.



Dice Example
Mean Sum
6.5
5.5
4.5
1
10
100
1000
10000
Number of Trials



Rainfall Example
Mean Days with Rain
1.4
1.2
1.0
0.8
1
10
100
1000
10000
Number of Trials

Linear Transformations

e.g. °F →°C or $Z = \frac{X-\mu}{\sigma}$

In general, $Y = aX + b$

As we saw when dealing with real data

$$\begin{aligned} \bar{y} &= a\bar{x} + b \\ s_y &= |a| s_x \end{aligned}$$

Similar relations hold for probability distributions

$$\begin{aligned} \mu_y &= a\mu_x + b \\ \sigma_y &= |a| \sigma_x \end{aligned}$$

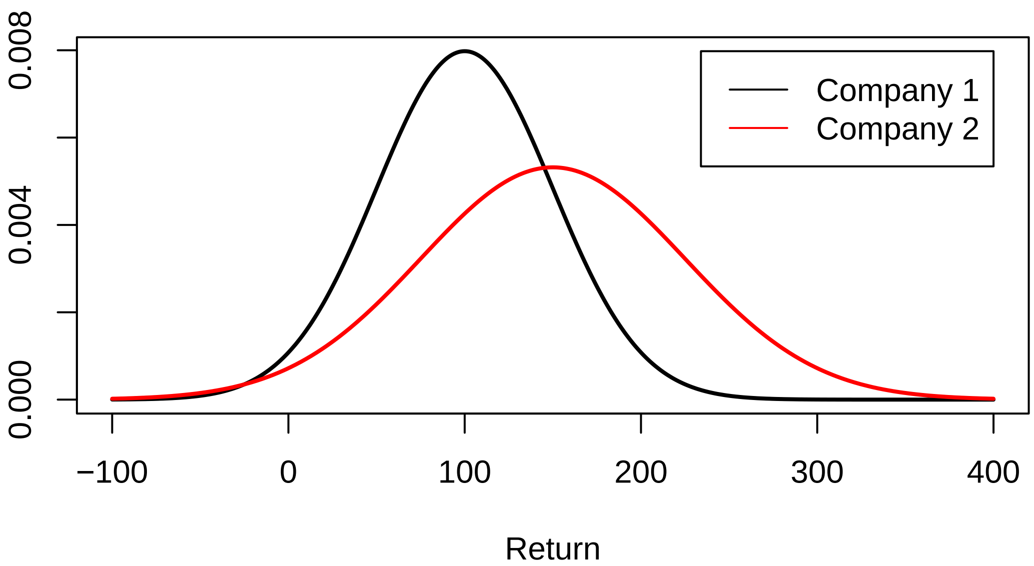
Note that these relationships hold for discrete and continuous distributions.

Adding and Subtracting Random Variables

Example:

X: return for stock 1 — $X \sim N(\mu_x = 100, \sigma_x = 50)$

Y: return for stock 2 — $Y \sim N(\mu_y = 150, \sigma_y = 75)$



What is the return from both stocks?

$$Z = X + Y$$

What are μ_z and σ_z?

The first is easy

$$\mu_z = \mu_x + \mu_y$$

So for this example, $\mu_z = 100 + 150 = 250$.

However there isn't enough information to get the standard deviation.

Suppose stock 1 is Motorola and stock 2 is Verizon. Since both companies have large involvement in the cellular phone industries, it wouldn't be surprising to for stocks to do well together or to do poorly together.

However suppose that stock 2 was Getty Oil. Motorola and Getty stock returns should have much less association.

So to get σ_z, we also need to know the correlation between X and Y (call it ρ). ρ is the probability analogue to the sample correlation r and has the same properties as r. Then

$$\begin{aligned} \sigma_z^2 &= \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y \\ \sigma_z &= \sqrt{\sigma_z^2} = \sqrt{\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y} \end{aligned}$$

Suppose that $\rho = 0.5$ (Motorola and Verizon), then

$$\begin{aligned}
\sigma_z^2 &= 50^2 + 75^2 + 2 \times 0.5 \times 50 \times 75 \\
&= 11875 \\
\sigma_z &= \sqrt{11875} = 108.972
\end{aligned}$$

Instead, now suppose that $\rho = 0$ (Motorola and Getty), then

$$\begin{aligned}
\sigma_z^2 &= 50^2 + 75^2 + 2 \times 0 \times 50 \times 75 \\
&= 8125 \\
\sigma_z &= \sqrt{8125} = 90.139
\end{aligned}$$

Now suppose that we are interested in the difference in the returns between the two stocks (which one has the better payoff).

$$D = X - Y$$

What are μ_D and σ_D?

$$\mu_D = \mu_x - \mu_y$$

so for the example, $\mu_D = 100 - 150 = -50$ (stock 2 is expected to be better by 50).

Like for the sum, we need the correlation between the two stocks to get the variance and the standard deviation of the difference.

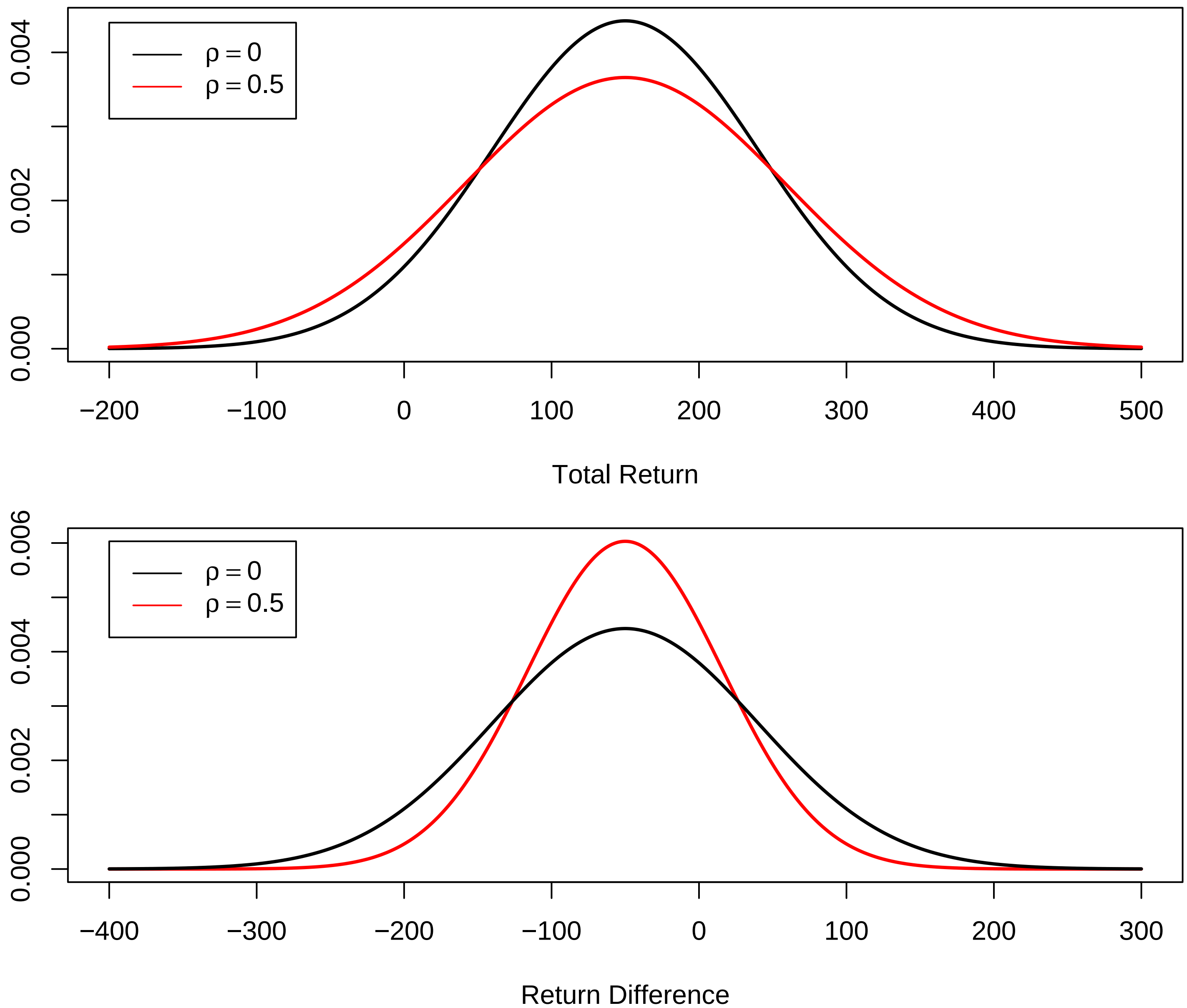
$$\begin{aligned} \sigma_D^2 &= \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y \\ \sigma_D &= \sqrt{\sigma_D^2} = \sqrt{\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y} \end{aligned}$$

Suppose that $\rho = 0.5$ (Motorola and Verizon), then

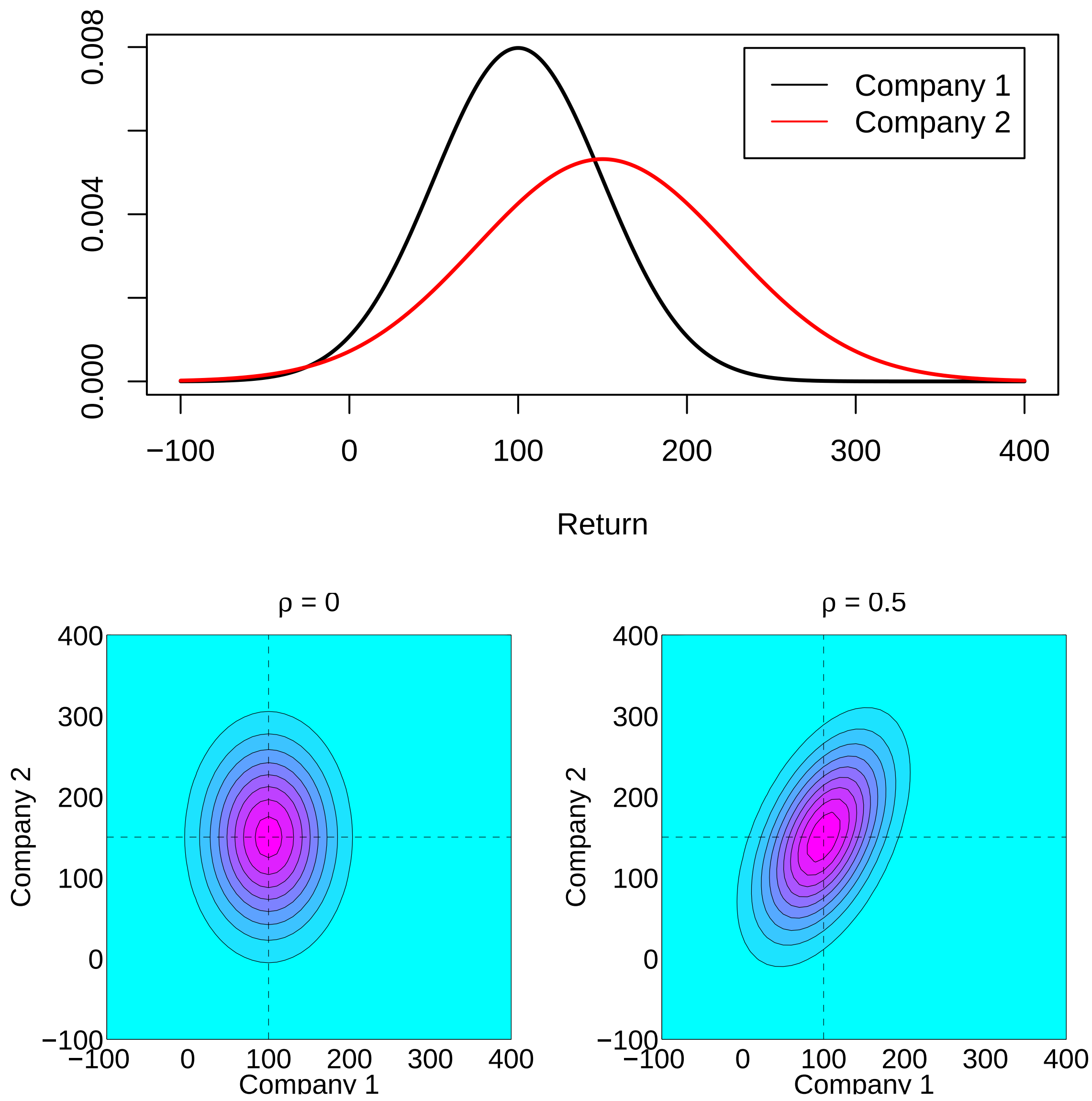
$$\begin{aligned}
\sigma_D^2 &= 50^2 + 75^2 - 2 \times 0.5 \times 50 \times 75 \\
&= 4375 \\
\sigma_D &= \sqrt{4375} = 66.144
\end{aligned}$$

Instead, now suppose that $\rho = 0$ (Motorola and Getty), then

$$\begin{aligned}
\sigma_D^2 &= 50^2 + 75^2 - 2 \times 0 \times 50 \times 75 \\
&= 8125 \\
\sigma_D &= \sqrt{8125} = 90.139
\end{aligned}$$


ρ = 0
ρ = 0.5
0.000
0.002
0.004
−200
−100
0
100
200
300
400
500
Total Return
ρ = 0
ρ = 0.5
0.000
0.002
0.004
0.006
−400
−300
−200
−100
0
100
200
300
Return Difference

Why does the correlation make a difference in σ of a sum or difference?



When $\rho > 0$, there is a positive association, so values greater than the mean for one variable tend to go with values greater than the mean in the other variable. Similarly, smaller than the mean tends to go with smaller than the mean.

When $\rho \approx 0$, a value greater than the mean on the first variable should get matched roughly half the time with a value greater than the mean on the second variable. The other half of the time, it should get matched with a value less than the mean.

So when two positively associated variables are added, both values will tend to be on the same size of the mean, so adding them together will push the sum far from the sum of the means.

And when two positively associated variables are differenced, you will get some cancellation, pulling the difference back towards the differences of the means.

When the two variables are negatively associated ($\rho < 0$), you get the opposite effects. (Think about when you will get cancellation and when you will get effects to combine in the same direction.)

The rule about means holds regardless of ρ. However ρ is needed to get the variances and standard deviations.

Also note that the rules for means and variances don't depend on what the distributions are. The normality assumptions made in the example were only there so the plots could be created easily.

Relationship between ρ and independence

If two random variables are independent, then $\rho = 0$.

However $\rho = 0$ does not imply independence. Like r in the data case, ρ describes the linear relationship between two random variables. There could be a non-linear relationship between two random variables and ρ could be 0.

If $\rho = 0$, the two variance formulas reduce to

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 \qquad \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$$
$$\sigma_D^2 = \sigma_x^2 + \sigma_y^2 \qquad \sigma_D = \sqrt{\sigma_x^2 + \sigma_y^2}$$

Remember: Variances add, not standard deviations.

Math note: The formula for the variance of the sum of two independent random variables is effectively Pythagoras theorem and the dependent case is related to the Law of Cosines.