Section 1.3: More Probability and Decisions: Linear Combinations and Continuous Random Variables

Jared S. Murray
The University of Texas at Austin
McCombs School of Business
OpenIntro Statistics, Chapters 2.4.2, 2.4.3, and 3.1-3.2

Introduction

We've seen how the **expected value** (our best prediction) and **variance/standard deviation** (how risky our best prediction is) help us think about uncertainty and make decisions in simple scenarios

We need some more tools for thinking about

1. Multiple random variables (sources of uncertainty)
2. Other kinds of random variables - continuous outcomes

Covariance

- A measure of *dependence* between two random variables...
- It tells us how two unknown quantities tend to move together: Positive $\rightarrow$ One goes up (down), the other tends to go up (down). Negative $\rightarrow$ One goes down (up), the other tends to go up (down).
- If X and Y are independent, $Cov(X, Y) = 0$ **BUT** $Cov(X, Y) = 0$ does not mean X and Y are independent (more on this later).

The Covariance is defined as (for discrete X and Y):

$$Cov(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} Pr(x_i, y_j) \times [x_i - E(X)] \times [y_j - E(Y)]$$

Ford vs. Tesla

- Assume a very simple joint distribution of monthly returns for Ford (F) and Tesla (T):

	t=-7%	t=0%	t=7%	Pr(F=f)
f=-4%	0.06	0.07	0.02	**0.15**
f=0%	0.03	0.62	0.02	**0.67**
f=4%	0.00	0.11	0.07	**0.18**
Pr(T=t)	**0.09**	**0.80**	**0.11**	**1**

Let's summarize this table with some numbers...

Example: Ford vs. Tesla

	t=-7%	t=0%	t=7%	Pr(F=f)
f=-4%	0.06	0.07	0.02	**0.15**
f=0%	0.03	0.62	0.02	**0.67**
f=4%	0.00	0.11	0.07	**0.18**
Pr(T=t)	**0.09**	**0.80**	**0.11**	**1**

- $E(F) = 0.12$, $E(T) = 0.14$
- $Var(F) = 5.25$, $sd(F) = 2.29$, $Var(T) = 9.76$, $sd(T) = 3.12$
- What is the better stock?

Example: Ford vs. Tesla

	t=-7%	t=0%	t=7%	Pr(F=f)
f=-4%	0.06	0.07	0.02	**0.15**
f=0%	0.03	0.62	0.02	**0.67**
f=4%	0.00	0.11	0.07	**0.18**
Pr(T=t)	**0.09**	**0.80**	**0.11**	**1**

$$\begin{aligned} Cov(F,T) =& (-7-0.14)(-4-0.12)0.06 + (-7-0.14)(0-0.12)0.03 + \\ & (-7-0.14)(4-0.12)0.00 + (0-0.14)(-4-0.12)0.07 + \\ & (0-0.14)(0-0.12)0.62 + (0-0.14)(4-0.12)0.11 + \\ & (7-0.14)(-4-0.12)0.02 + (7-0.14)(0-0.12)0.02 + \\ & (7-0.14)(4-0.12)0.07 = 3.063 \end{aligned}$$

Okay, the covariance in positive... makes sense, but can we get a more intuitive number?

Correlation

$$Corr(X, Y) = \frac{Cov(X, Y)}{sd(X)sd(Y)}$$

- What are the units of $Corr(X, Y)$? It doesn't depend on the units of X or Y!
- $-1 \leq Corr(X, Y) \leq 1$

In our Ford vs. Tesla example:

$$Corr(F, T) = \frac{3.063}{2.29 \times 3.12} = 0.428 \text{ (not too strong!)}$$

Linear Combination of Random Variables

Is it better to hold Ford or Tesla? How about half and half?

To answer this question we need to understand the behavior of the weighted sum (linear combinations) of two random variables...

Let X and Y be two random variables:

- $E(aX + bY + c) = aE(X) + bE(Y) + c$
- $Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2ab \times Cov(X, Y)$

Linear Combination of Random Variables

Applying this to the Ford vs. Tesla example...

- $E(0.5F + 0.5T) = 0.5E(F) + 0.5E(T) =$
  $0.5 \times 0.12 + 0.5 \times 0.14 = 0.13$
- $Var(0.5F + 0.5T) =$
  $(0.5)^2 Var(F) + (0.5)^2 Var(T) + 2(0.5)(0.5) \times Cov(F, T) =$
  $(0.5)^2(5.25) + (0.5)^2(9.76) + 2(0.5)(0.5) \times 3.063 = 5.28$
- $sd(0.5F + 0.5T) = 2.297$

so, what is better? Holding Ford, Tesla or the combination?

Risk Adjustment: Sharpe Ratio

The Sharpe ratio is a unitless quantity used to compare investments:

$$\frac{\text{(average return) - (return on a risk-free investment)}}{\text{standard deviation of returns}}$$

Idea: Standardize the average excess return by the amount of risk. ("Risk adjusted returns")

Ignoring the risk-free investment, what are the Sharpe ratios for Ford, Tesla, and the 50-50 portfolio?

Linear Combination of Random Variables

More generally...

- $E(w_1X_1 + w_2X_2 + ...w_pX_p + c) = w_1E(X_1) + w_2E(X_2) + ... + w_pE(X_p) + c = \sum_{i=1}^{p} w_iE(X_i) + c$
- $Var(w_1X_1 + w_2X_2 + ...w_pX_p + c) = w_1^2Var(X_1) + w_2^2Var(X_2) + ... + w_p^2Var(X_p) + 2w_1w_2 \times Cov(X_1, X_2) + 2w_1w_3Cov(X_1, X_3) + ... = \sum_{i=1}^{p} w_i^2Var(X_i) + \sum_{i=1}^{p}\sum_{j\neq i} w_iw_jCov(X_i, X_j)$

where $w_1, w_2, \ldots, w_p$ and c are constants

Continuous Random Variables

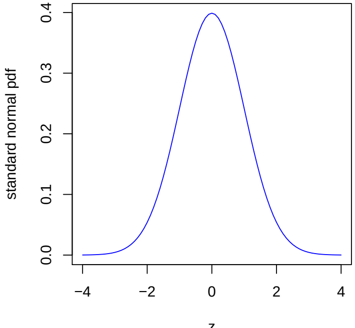
- Suppose we are trying to predict tomorrow's return on the S&P500 (Or on a real Ford/Tesla portfolio)...
- Question: What is the random variable of interest? What are its possible outcomes? Could you list them?
- Question: How can we describe our uncertainty about tomorrow's outcome?

Continuous Random Variables

- Recall: a random variable is a number about which we're uncertain, but can describe the possible outcomes.
- Listing all possible values isn't possible for continuous random variables, we have to use intervals.
- The probability the r.v. falls in an interval is given by the **area** under the **probability density function.** For a continuous r.v., the probability assigned to any single value is zero!

The Normal Distribution

- The Normal distribution is the most used probability distribution to describe a continuous random variable. Its probability density function (pdf) is symmetric and bell-shaped.
- The probability the number ends up in an interval is given by the area under the pdf.

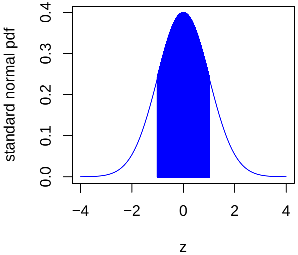


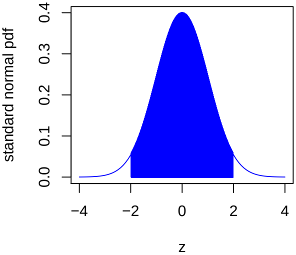
The Normal Distribution

- The standard Normal distribution has mean 0 and has variance 1.
- Notation: If $Z \sim N(0,1)$ (Z is the random variable)

$$Pr(-1 < Z < 1) = 0.68$$

$$Pr(-1.96 < Z < 1.96) = 0.95$$





The Normal Distribution

Note:

For simplicity we will often use $P(-2 < Z < 2) \approx 0.95$

Questions:

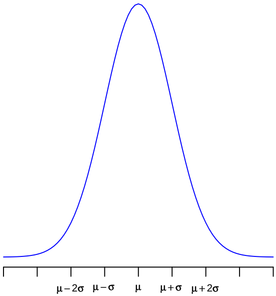
- What is $Pr(Z < 2)$? How about $Pr(Z \leq 2)$?
- What is $Pr(Z < 0)$?

The Normal Distribution

- The standard normal is not that useful by itself. When we say “the normal distribution”, we really mean a family of distributions.
- We obtain pdfs in the normal family by shifting the bell curve around and spreading it out (or tightening it up).

The Normal Distribution

- We write $X \sim N(\mu, \sigma^2)$. "X has a Normal distribution with mean μ and variance σ^2.
- The parameter μ determines where the curve is. The center of the curve is μ.
- The parameter σ determines how spread out the curve is. The area under the curve in the interval $(\mu - 2\sigma, \mu + 2\sigma)$ is 95%. $Pr(\mu - 2\,\sigma < X < \mu + 2\,\sigma) \approx 0.95$

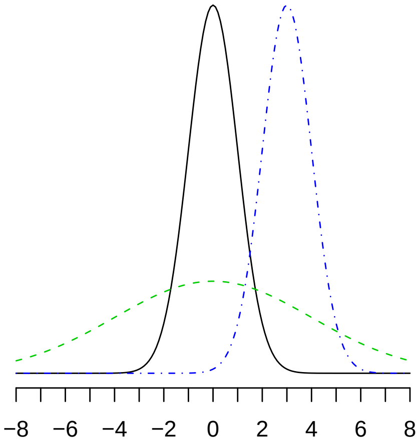


Recall: Mean and Variance of a Random Variable

- For the normal family of distributions we can see that the parameter μ determines "where" the distribution is *located* or *centered*.
- The expected value μ is usually our best guess for a *prediction*.
- The parameter σ (the standard deviation) indicates how *spread out* the distribution is. This gives us and indication about how *uncertain* or how *risky* our prediction is.

The Normal Distribution

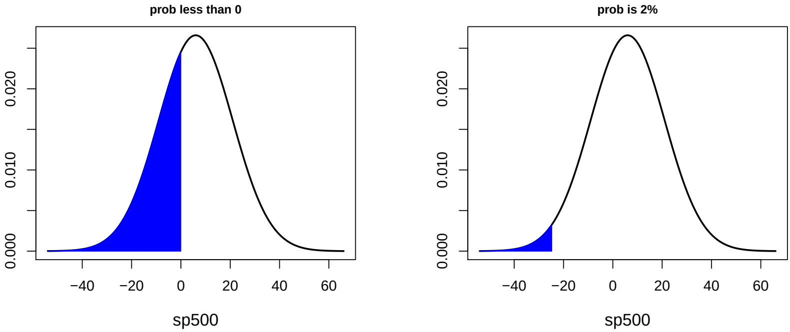
- Example: Below are the pdfs of $X_1 \sim N(0,1)$, $X_2 \sim N(3,1)$, and $X_3 \sim N(0,16)$.
- Which pdf goes with which X?



The Normal Distribution – Example

- Assume the annual returns on the SP500 are normally distributed with mean 6% and standard deviation 15%. SP500 $\sim N(6, 225)$. (Notice: $15^2 = 225$).
- Two questions: (i) What is the chance of losing money in a given year? (ii) What is the value such that there's only a 2% chance of losing that or more?
- Lloyd Blankfein: *"I spend 98% of my time thinking about .02 probability events!"*
- (i) $Pr(SP500 < 0)$ and (ii) $Pr(SP500 <?) = 0.02$

The Normal Distribution – Example



- (i) $Pr(SP500 < 0) = 0.35$ and (ii) $Pr(SP500 < -25) = 0.02$

The Normal Distribution in R

In R, calculations with the normal distribution are easy!

(Remember to use SD, not Var)

To compute $Pr(SP500 < 0) = ?$:

```
pnorm(0, mean = 6, sd = 15)

## [1] 0.3445783
```

To solve $Pr(SP500 < ?) = 0.02$:

```
qnorm(0.02, mean = 6, sd = 15)

## [1] -24.80623
```

The Normal Distribution

1. Note: In

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

   μ is the mean and σ^2 is the variance.

2. Standardization: if $X \sim N(\mu, \sigma^2)$ then

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

3. Summary:

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

   μ: where the curve is

   σ: how spread out the curve is

   95% chance $X \in \mu \pm 2\sigma$.

The Normal Distribution – Another Example

Prior to the 1987 crash, monthly S&P500 returns (r) followed (approximately) a normal with mean 0.012 and standard deviation equal to 0.043. How extreme was the crash of -0.2176? The standardization helps us interpret these numbers...

$$r \sim N(0.012, 0.043^2)$$

$$z = \frac{r - 0.012}{0.043} \sim N(0,1)$$

For the crash,

$$z = \frac{-0.2176 - 0.012}{0.043} = -5.27$$

How extreme is this $z$value? 5 standard deviations away!!

Portfolios, once again...

- As before, let's assume that the annual returns on the SP500 are normally distributed with mean 6% and standard deviation of 15%, i.e., $SP500 \sim N(6, 15^2)$
- Let's also assume that annual returns on bonds are normally distributed with mean 2% and standard deviation 5%, i.e., $Bonds \sim N(2, 5^2)$
- What is the best investment?
- What else do I need to know if I want to consider a portfolio of SP500 and bonds?

Portfolios once again...

- Additionally, let's assume the correlation between the returns on SP500 and the returns on bonds is -0.2.
- How does this information impact our evaluation of the best available investment?

Recall that for two random variables X and Y:

- $E(aX + bY) = aE(X) + bE(Y)$
- $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab \times Cov(X, Y)$

- One more very useful property... sum of normal random variables is a new normal random variable!

Portfolios once again...

- What is the behavior of the returns of a portfolio with 70% in the SP500 and 30% in Bonds?
- $E(0.7SP500 + 0.3Bonds) = 0.7E(SP500) + 0.3E(Bonds) = 0.7 \times 6 + 0.3 \times 2 = 4.8$
- $Var(0.7SP500 + 0.3Bonds) = (0.7)^2 Var(SP500) + (0.3)^2 Var(Bonds) + 2(0.7)(0.3) \times Corr(SP500, Bonds) \times sd(SP500) \times sd(Bonds) = (0.7)^2(15^2) + (0.3)^2(5^2) + 2(0.7)(0.3) \times -0.2 \times 15 \times 5 = 106.2$
- $Portfolio \sim N(4.8, 10.3^2)$
- What do you think about this portfolio? Is there a better set of weights?

Simulating Normal Random Variables

- Imagine you invest $1 in the SP500 today and want to know how much money you are going to have in 20 years. We can assume, once again, that the returns on the SP500 on a given year follow $N(6, 15^2)$
- Let's also assume returns are independent year after year...
- Are my total returns just the sum of returns over 20 years? Not quite... compounding gets in the way.

Let's simulate potential "futures"

Simulating one normal r.v.

At the end of the first year I have \$$(1 \times (1 + \text{pct return}/100))$.

```
val = 1 + rnorm(1, 6, 15)/100
print(val)

## [1] 0.9660319
```

`rnorm(n, mu, sigma)` draws *n* samples from a normal distribution with mean μ and standard deviation σ.

Simulating compounding

We reinvest our earnings in year 2, and every year after that:

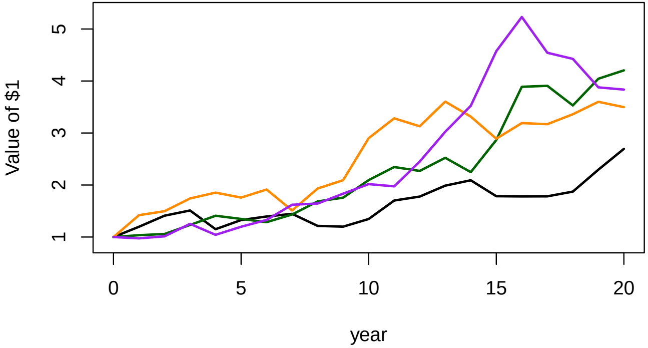
```
for(year in 2:20) {
  val = val*(1 + rnorm(1, 6, 15)/100)
}
print(val)

## [1] 4.631522
```

Simulating a few more "futures"

We did pretty well - our \$1 has grown to \$4.63, but is that typical?

Let's do a few more simulations:



More efficient simulations

Let's simulate 10,000 futures under this model. Recall the value of my investment at time T is

$$\prod_{t=1}^{T}(1 + r_t/100)$$

where r_t is the percent return in year t

```
library(mosaic)
num.sim = 10000
num.years = 20
values = do(num.sim) * {
  prod(1 + rnorm(num.years, 6, 15)/100)
}
```

Simulation results

Now we can answer all kinds of questions:

What is the mean value of our investment after 20 years?

```r
vals = values$result
mean(vals)

## [1] 3.187742
```

What's the probability we beat a fixed-income investment (say at 2%)?

```r
sum(vals > 1.02^20)/num.sim

## [1] 0.8083
```

Simulation results

What's the median value?

```
median(vals)

## [1] 2.627745
```

(Recall: The median of a probability distribution (say m) is the point such that $\Pr(X \leq m) = 0.5$ and $\Pr(X > m) = 0.5$ when X has the given distribution).

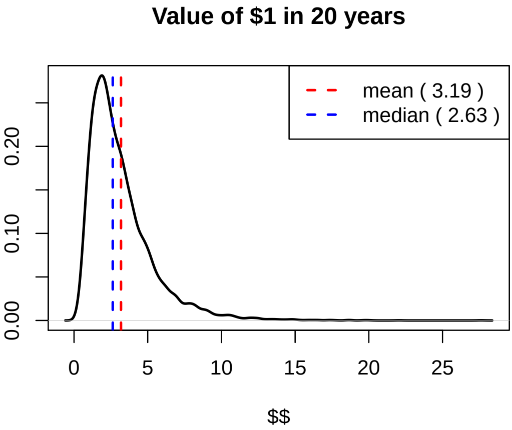
Remember the mean of our simulated values was 3.19...

Median and skewness

- For symmetric distributions, the expected value (mean) and the median are the same... look at all of our normal distribution examples.
- But sometimes, distributions are skewed, i.e., not symmetric. In those cases the median becomes another helpful summary!

Probability density function of our wealth at $T = 20$

We see the estimated distribution is skewed to the right if we use the simulations to estimate the pdf:



What's next?

What's mising from this picture?

Where did SP500's 6% returns with an SD of 15% come from?

Up next: Learning parameters from data (statistics!), and uncertainty in parameters