Continuous Random Variables and the Normal Distribution

Jared S. Murray
The University of Texas at Austin
McCombs School of Business

Continuous Random Variables

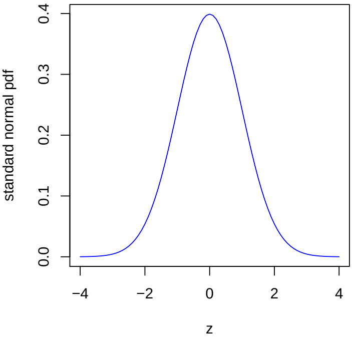
- Suppose we are trying to predict tomorrow's return on the S&P500...
- Question: What is the random variable of interest? What are its possible outcomes? Could you list them?
- Question: How can we describe our uncertainty about tomorrow's outcome?

Continuous Random Variables

- Recall: a random variable is a number about which we're uncertain, but can describe the possible outcomes.
- We can't list all the possible outcomes for continuous random variables, but we can give 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 common distribution for a continuous random variable. Its probability density function (pdf) is symmetric and bell-shaped.

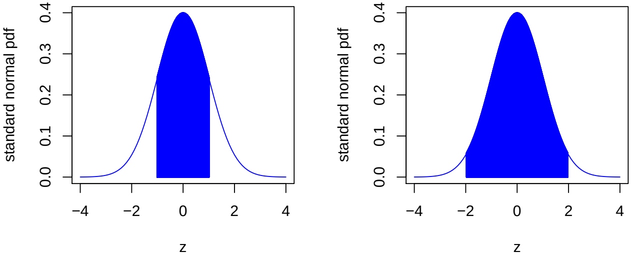


The Normal Distribution

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

$$Pr(-1 < Z < 1) \approx 2/3$$

$$Pr(-2 < Z < 2) \approx 0.95$$



The Normal Distribution

Questions:

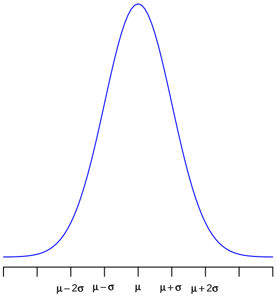
- What is $Pr(Z < 2)$? How about $Pr(Z \leq 2)$?
- What is $Pr(Z < 0)$?

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%. $Pr(\mu - 2\,\sigma < X < \mu + 2\,\sigma) \approx 0.95$



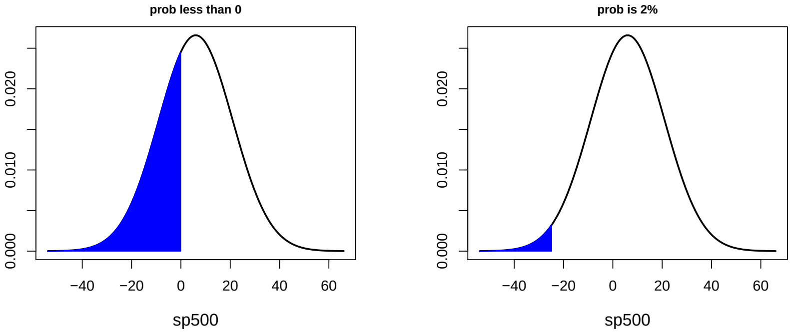
Recall: Mean and Variance of a Random Variable

- 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

- 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: Standardization

Standardization: For **any** random variable,

$$E(aX+b) = aE(X)+b, \quad Var(aX+b) = a^2 Var(X)$$

For normal random variables, if $X \sim N(\mu, \sigma^2)$ then

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

If we take one draw x from a $N(\mu, \sigma^2)$ distribution, then $z = (x-\mu)/\sigma$ tells us how many standard deviations away x is from the mean.

The larger z is in absolute value, the more extreme (unlikely) the value x was to observe.

Standardization – An Example

Since 2000, monthly S&P500 returns (r) have followed (very approximately) a normal distribution mean 0.58% and standard deviation equal to 4.1% How extreme was the October 2008 crash of -16.5%? Standardization helps us interpret these numbers...

$$r \sim N(0.58, 4.1^2)$$

$$z = \frac{r - 0.58}{4.1} \sim N(0, 1)$$

For the crash,

$$z = \frac{-16.5 - 0.58}{4.1} \approx -4.2$$

How extreme is this $z-$score? Over 4 standard deviations away!

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)` returns *n* draws 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:

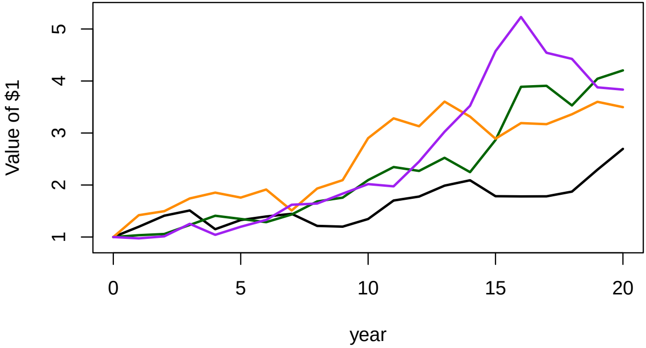
```
for(year in 2:20) {
  val = val*(1 + rnorm(1, 6, 15)/100)
}
print(val)

## [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?

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

## [1] 3.187742
```

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

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

## [1] 0.8083
```

Simulation results

What's the median value?

```r
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: