Bernoulli and Binomial Random Variables

A Motivating Example

- Between 1960 and 1980, there were many lawsuits in the South claiming racial bias in jury selection.
- Here's some made up* (but similar) supporting data:
    - 50% of citizens in the local area are African American
    - On an 80 person panel, only 4 were African American
- Can this outcome be explained as the result of pure chance?
- What is the probability of 4 heads in 80 fair coin flips?

*This example was borrowed from *The Cartoon Guide to Statistics*.

Bernoulli Trials

A Bernoulli trial is a random experiment with 2 special properties:

- The result of a Bernoulli trial is binary.
  - Examples: Heads vs. Tails, Healthy vs. Sick, etc.
- The probability of a "success" is some constant p.
  - Example: the probability of heads when you flip a fair coin is always 0.5.

Bernoulli Random Variable

A Bernoulli random variable is a random variable such that:

- The range of possible values is binary.
  - Examples: Heads vs. Tails, Healthy vs. Sick, etc.
- The probability of a "success" is some constant p.
  - Example: the probability of heads is p, and the probability of tails is $1-p$.
  - Example: the probability of healthy is p, and the probability of sick is $1-p$.

Expected Value and Variance

Bernoulli random variable:

- The expected value of a Bernoulli random variable is *p.*
  - E[B] = …
- The variance of a Bernoulli random variable is *p(1-p)*.
  - E[(B - μ)^2] = …
  - E[B^2] - μ^2 = … - p^2 = …

Expected Value and Variance

Bernoulli random variable:

- The expected value of a Bernoulli random variable is *p.*
  - $E[B] = (p)(1) + (1-p)(0) = p$
- The variance of a Bernoulli random variable is *p(1-p)*.
  - $E[(B - \mu)^2] = (p)(1-p)^2 + (1-p)(0-p)^2 = p(1-p)$
  - $E[B^2] - \mu^2 = (p)(1^2) + (1-p)(0^2) - p^2 = p - p^2 = p(1-p)$

Binomial Random Variable

A binomial random variable describes the result of *n* Bernoulli trials:

- The range is the natural numbers, representing the number of successes.
    - Examples: the number of heads, the number of healthy participants, etc.
- The probability of a “success” is constant across all trials.
    - Example: the probability of heads when you flip a fair coin is always 0.5.
- The trials are independent events; earlier trials do not influence later trials.

Examples of Binomial Random Variables

- The number of heads in 10 coin flips.
- The number of times you roll double sixes in 50 rolls of the dice.
- The number of people who quit smoking after treating 100 participants.
- The number of people who report that they prefer Elizabeth Warren to Bernie Sanders in a poll of 500 individuals.
- What else?

The Binomial Distribution

- Here is the formula for the binomial distribution: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- *X* is the binomial random variable, *k* is the number of successes, *n* is the number of trials, and *p* is the probability of success.
- For example, let *X* be the total number of heads in 10 flips of a fair coin. This means *k* ranges from 0 to 10, *n* = 10, and *p* = 0.5.

Aside: Pascal’s Triangle

- Binomial coefficients: “*n* choose 0”, “*n* choose 1”, … “*n* choose *k*”, …, “*n* choose (*n* - 1)”, and “*n* choose *n*”
- “n choose k” = $n! / (k!\,(n - k)!)$

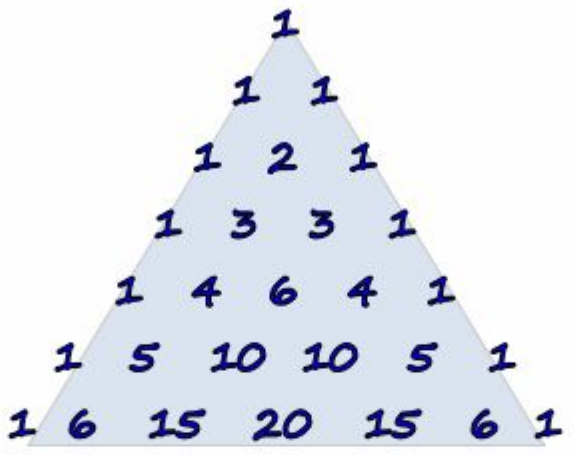


Image source

Calculating Binomial Probabilities in R

- Assume a binomial random variable X with parameters n and p.
  - The random variable can take on values of k ranging from 0 through n.
  - R can help us find $\Pr[X = k]$ for all values of k.

- In R, we can use `dbinom(k, n, p)` to find $\Pr[X = k]$
  - Let's say we flip a coin 10 times. What is the probability we see 3 heads?
  - `dbinom(3, 10, 0.5)` outputs 0.117

- In R, we can use `pbinom(k, n, p)` to find the $\Pr[X \leq k]$
  - If we flip a coin 10 times, what is the probability we see 3 heads, or fewer?
  - `dbinom(0, 10, 0.5) + dbinom(1, 10, 0.5) + dbinom(2, 10, 0.5) + dbinom(3, 10, 0.5)` outputs 0.172
  - `pbinom(3, 10, 0.5)` also outputs 0.172

A Motivating Example

- Between 1960 and 1980, there were many lawsuits in the South claiming racial bias in jury selection.
- Here's some made up[*] (but similar) supporting data:
  - 50% of citizens in the local area are African American
  - On an 80 person panel, only 4 were African American
- Can this outcome be explained as the result of pure chance?
- If $X \sim \text{Binomial}(n = 80, p = 0.5)$, then $P[X = 4] \approx P[X \leq 4] \approx 1 \times 10^{-18}$
- *N.B.*: Statistics can never *prove* anything.
  Still, this outcome is extremely unlikely to be the result of pure chance!

[*]This example was borrowed from *The Cartoon Guide to Statistics*.

Expected Value and Variance

Binomial random variable:

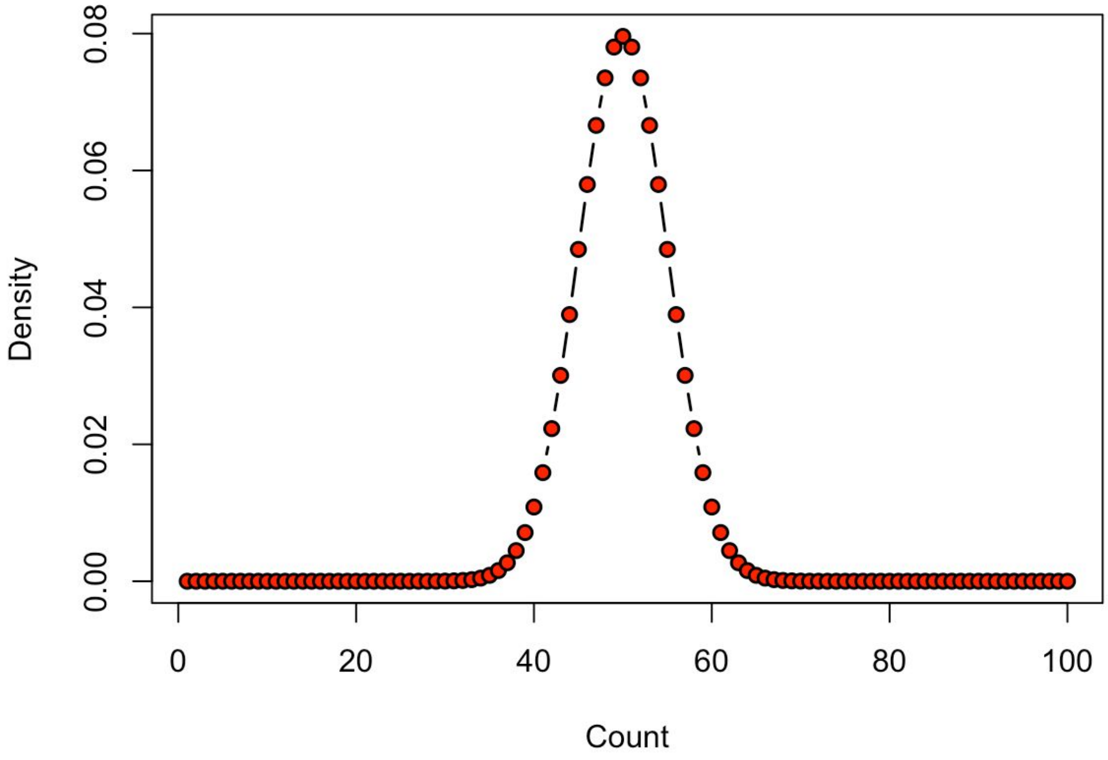
- The expected value of a binomial random variable is *np*.
  - $E[X_1 + X_2 + \ldots + X_{n-1} + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n] = nE[X_1] = np$
  - Linearity of expectations
- The variance of a binomial random variable is *np(1-p)*.
  - A binomial random variable is the sum of is n Bernoulli trials
  - The variance of the sum of independent r.v.s equals the sum of their variances
  - $\mathrm{Var}[X_1 + X_2 + \ldots + X_{n-1} + X_n] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \ldots + \mathrm{Var}[X_n] = n\mathrm{Var}[X_1] = np(1-p)$

Central Limit Theorem

The Binomial Distribution

The red distribution is binomial(100, 0.5)

For i = 1:100, Pr[X = i]

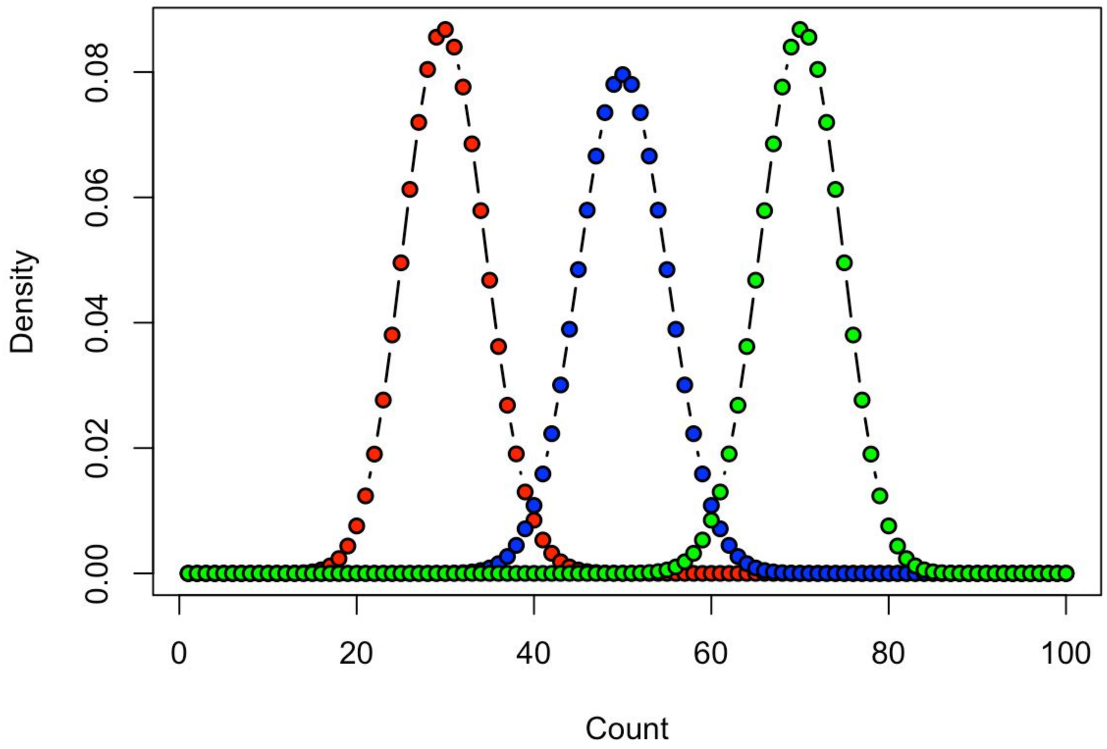


The Binomial Distribution

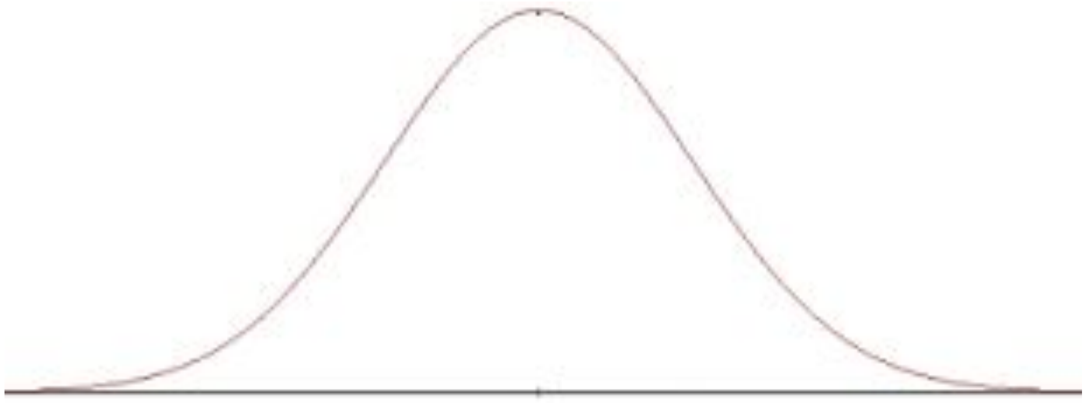
The blue distribution is binomial(100, 0.3)

The red distribution is binomial(100, 0.5)

The green distribution is binomial(100, 0.7)



Does this shape ring a bell?



The Normal Distribution

- A bell-shaped curve, whose shape depends on two things
  - Mean, Median, and Mode: the center of the curve
  - Variance: the spread (i.e., the height and width) of the curve

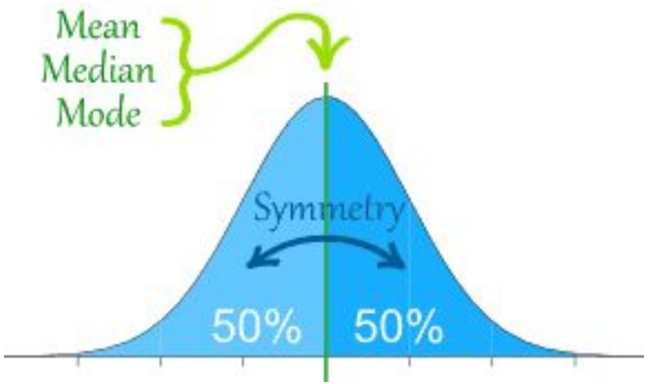


Image source

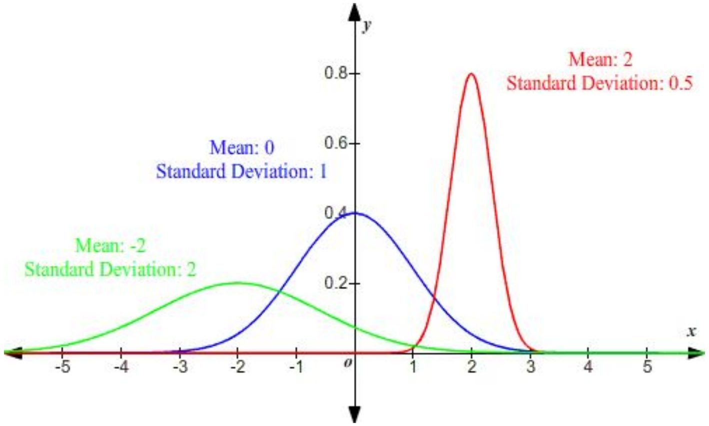


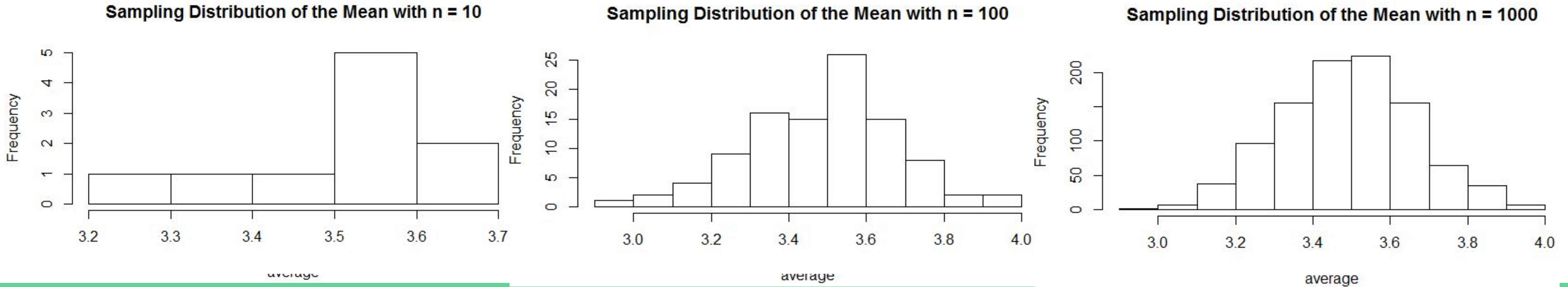
Image source

Central Limit Theorem

- Recall the expected value of one roll of a die: 3.5
- Let’s consider rolling a die, say, 100 times, and computing the mean roll

```
num_trials <- 100
sample_data <- sample(1:6, num_trials, replace = TRUE)
mean(sample_data)
```

- And let’s repeat this process, first 10, then 100, and then 1000 times



Central Limit Theorem (cont'd)

- The distribution of sample means is called the sampling distribution.
- As we collect more and more sample means, the sampling distribution looks more and more like the normal distribution, *even though the distribution that we were sampling from was uniform* (not normal).
- Remarkable fact: This is true regardless of the underlying distribution.
  (It was also true when we repeatedly sampled from a Bernoulli distribution.)

In the limit (meaning, as the number of experiments grows to infinity),
the sample mean is normally distributed around the true mean,
regardless of the underlying distribution.

Standard Error

The mean of this sampling distribution is the true mean.

And what is the standard deviation (or the variance) of the sampling distribution?

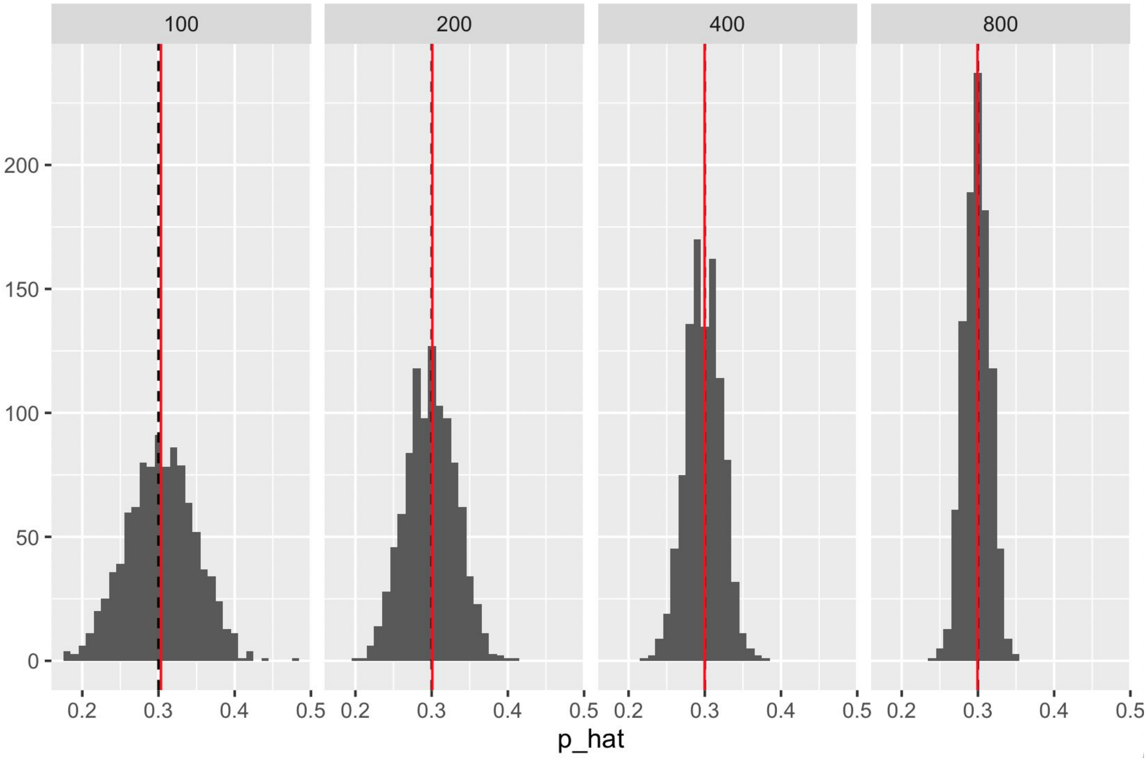
The standard deviation of the sampling distribution is called standard error (SE).

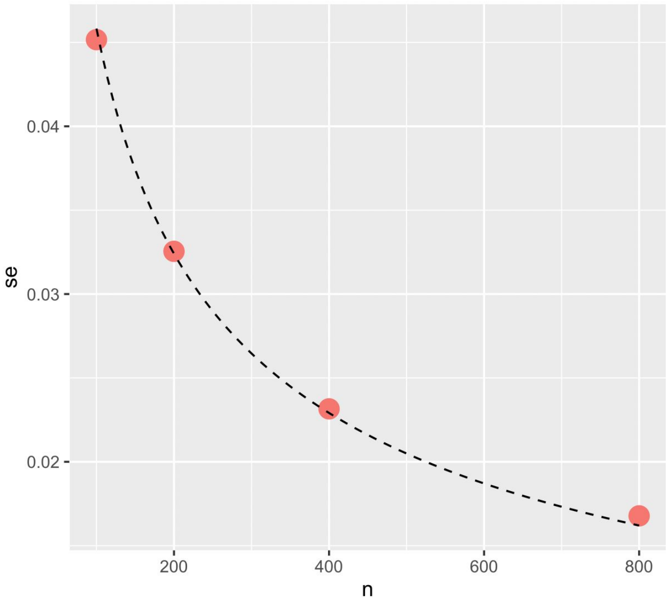
- Standard deviation measures variation in a distribution,
  meaning how individual measurements differ from the mean.
- Standard error measures how sample means differ from the true mean.

Theorem If the variance of a random variable is σ^2, then the variance of the sample mean (i.e., of the sampling distribution of that random variable) is σ^2/n.

So the formula for the standard error, which is the standard deviation of the sample mean, is $\sigma/\sqrt{n}$.

Standard Error (cont'd)





Deriving Standard Error

Standard Error

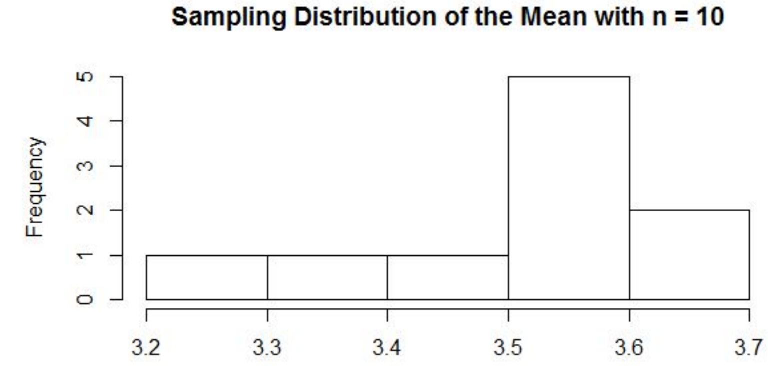
- The sample mean is the average of all the sample values.
- The sample variance is the average of the squared deviations from the mean.
- The sample standard deviation is the square root of the sample variance.
- The standard error is the standard deviation of the sampling distribution.

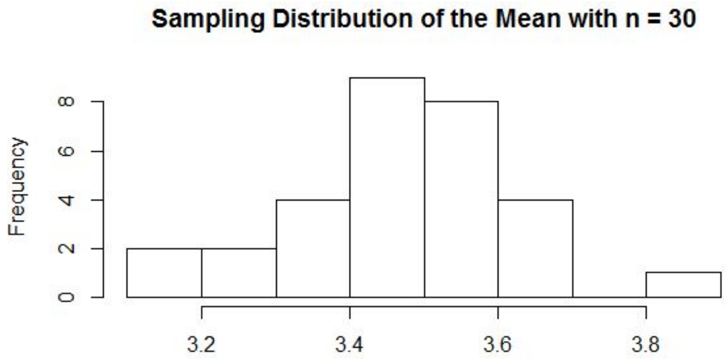
The Standard Error of the Sample Mean

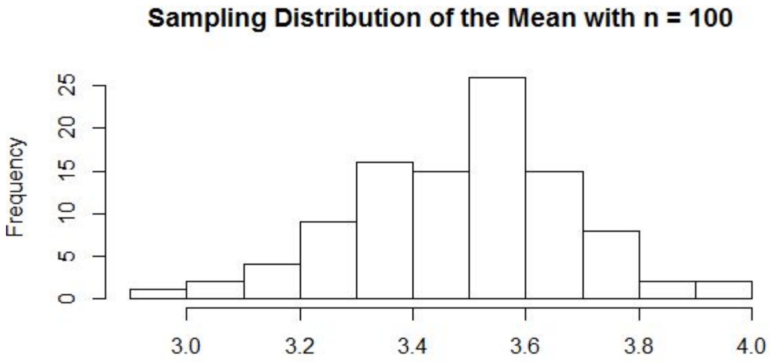
- Let X_M represent the sample mean:
  - $X_M = (X_1 + \ldots + X_N)/n$
- First, we need the variance of the sample mean:
  - X_i are normally distributed with variance σ^2
  - $\mathrm{Var}[X_M] = \mathrm{Var}[(X_1 + \ldots + X_N)/n] = (1/n^2)\ \mathrm{Var}[X_1 + \ldots + X_N] = (1/n^2)\ (\mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \ldots + \mathrm{Var}[X_N]) = (1/n^2)\ (n)\ \mathrm{Var}[X_1] = \sigma^2/n$
- We now have a formula for the standard error of the sample mean:
  - $\mathrm{SE}[X_M] = \sigma/\sqrt{n}$
- If X is a binomial random variable, sample mean is called sample proportion.
- The standard error of the sample proportion is:
  - $\mathrm{SE}[X_M] = \sqrt{np(1-p)}/\sqrt{n} = \sqrt{p(1-p)}$

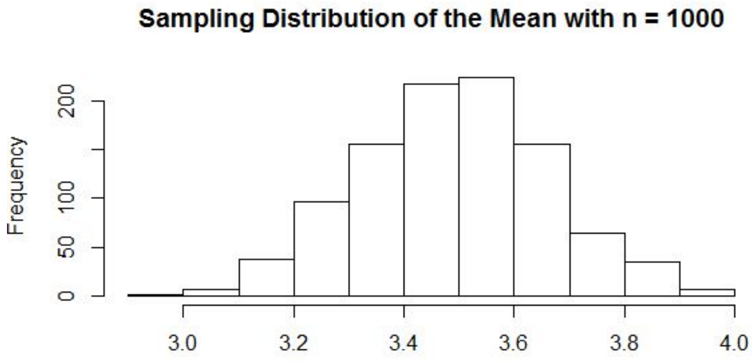
Student t-Distribution

When does the CLT kick in?









What is a large enough sample size?

- The distribution looks roughly normal with a sample size of 30.
- So, as a rule of thumb, people often say that 30 is a large enough sample size for the central limit theorem to apply.
- However, the histogram becomes more and more bell-shaped as the sample size increases.
- So all things being equal, larger sample sizes are always better than smaller ones!

What if the sample size is not large enough?

- The *(Student) t*-distribution can be used to approximate the normal, when the sample size is not large enough.
  - The aforementioned student was one William S. Gosset.
  - He discovered this (family of) distribution(s) in 1908, while employed as a statistician by the Guinness brewing company, who forbid him from publishing under his own name.
  - He wrote under the pen name "Student" instead.
- The *t*-distribution allows us to perform statistical inference even when the sample size is not large enough to apply the central limit theorem.

The *t*-Distribution

- The *t*-distribution has 1 parameter: the degrees of freedom.
- As *n* goes to infinity, the *t*-distribution converges to the normal.
- Thus, using the *t*-distribution when the sample size is small is consistent with using the normal distribution when the sample size is large.

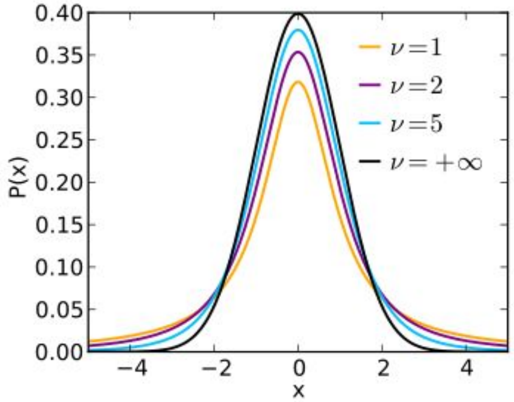


Image source

Extras

Geometric Distribution

- The geometric distribution can be used to model how many trials we need until we have a success.
  - If we have trials that occur with probability *p*, then what is the likelihood we will have the first success on trial *k*?
  - An example: If we go to a slot machine with a 0.0001 probability (*p*) of a jackpot. What is the probability we win by trial 1000 *(k)*?
- Mathematically, P(success on trial *k*) = $p(1-p)^{(k-1)}$
  - In R, we can use `dgeom(k, p)` to find this for any given trial.
  - We can use `pgeom(k, p)` to find this for any trial before a given trial.
- In the case of the slot machine, `pgeom(1000, 0.0001)` = ~10%.

Poisson Distribution

- The Poisson distribution can be used to model how many events take place at a fixed time period.
  - An example: At a certain stoplight, there are typically 5 stopped cars. What is the probability that there are 7 cars at the light today?
- Mathematically, Pr[7 cars] = 5^7 {exp(-5)} / 7!
  - In R, we can use `dpois(7, 5)` to find this probability.
- There is a ~10% chance that there will be exactly 7 cars at the stoplight.

Mathematical Aside

- Here's a fun fact you might recall about histograms:
  The total area under all bars is 1.
- Here's a fun fact about probability distributions:
  The total area under the curve is 1.
- The normal distribution takes on continuous values
  - There are no jumps or holes along the x-axis
- Integrating this curve from -∞ to +∞ yields 1!