The Normal Distribution

or ...

What Gauss Around, Comes Around

Ryne Estabrook

Department of Psychology
University of Virginia

September 8, 2008
PSYC 771
Previously on...

- Surprisingness, Probability & Distributions.
- Measures of Central Tendency & Spread.
- Histograms & Boxplots.

Today:

We combine all three... and I want questions throughout.
What are Distributions?

- Tim talked about distributions generally and the binomial & Bernoulli distributions on Monday.
- These distributions are relatively easy to understand, as they deal with discrete phenomena with easily estimable probabilities.
- There are a very large number of defined distributions, each with different properties and uses.
What are Distributions?

- Distributions are statements of probability or frequency of a set of numbers.
- We’ve previously discussed the binomial distribution:
  - If \( P(A) = p \) for any trial, then what’s the probability of \( A \) happening \( x \) times in \( n \) trials?
- We can talk about data being distributed a certain way.
  - i.e. what are the frequencies or probabilities associated with each value?
- The definition of *random* is tied to distributions.
  - A random variable is one defined by a distribution.
What are Distributions?

- Empirical or theoretical distributions are described by equations that define these probabilities or densities at specific values of a variable.

- These equations are known as *probability density functions* for continuous variables and *probability mass functions* for categorical variables.

- For instance, the probability mass function for the binomial distribution is:

\[
f(X; p, n) = \frac{n!}{X!(n-X)!} p^X (1 - p)^{(n-X)}\]
The Binomial Distribution

\[ f(X; p, n) = \frac{n!}{X!(n - X)!} p^X (1 - p)^{(n-X)} \]

- There are a couple of important points in this formula.
  - \( X \) is our data. This is a function of \( X \), such that every value of \( X \) yields a probability.
  - \( n \) & \( p \) are parameters. They are values that are fixed for a given sample or experiment.
  - When you change the parameters, you get different versions of the same distribution.
  - Alternatively, you can think of a binomial distribution as a family of distributions with a common set of parameters.
The Binomial Distribution

So instead of saying defining and calculating this every single time we use this distribution,

\[ f(X; p, n) = \frac{n!}{X!(n - X)!} p^X (1 - p)^{n - X}, \]

we can just say that \( X \) is binomially distributed, or distributed binomially, written like this:

\[ X \sim B(n, p) \]
Binomial Distribution

Binomial Distribution
\( p=0.5, n=10 \)

<table>
<thead>
<tr>
<th>X</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
</tr>
<tr>
<td>3</td>
<td>0.10</td>
</tr>
<tr>
<td>4</td>
<td>0.15</td>
</tr>
<tr>
<td>5</td>
<td>0.20</td>
</tr>
<tr>
<td>6</td>
<td>0.25</td>
</tr>
<tr>
<td>7</td>
<td>0.20</td>
</tr>
<tr>
<td>8</td>
<td>0.15</td>
</tr>
<tr>
<td>9</td>
<td>0.05</td>
</tr>
<tr>
<td>10</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Binomial Distribution

Binomial Distribution

\[ p = 0.7, \ n = 10 \]
Binomial Distribution

Binomial Distribution
\( p=1.0, n=10 \)
Other ways of describing distributions

▶ A very closely related quantity to the probability mass & probability density functions are cumulative mass & cumulative density functions.
▶ What do they tell us?
  ▶ PDFs tell us the probability of $X$ equaling some value $P(X=A)$.
  ▶ CDFs tell us the probability that $X$ is less than or equal to some value $P(X\leq A)$.
▶ What do they look like?
Binomial Distribution

Binomial Distribution
p=0.5, n=10

Cumulative Probability

Cumulative Probability

X

0 1 2 3 4 5 6 7 8 9 10

0.0 0.2 0.4 0.6 0.8 1.0
Binomial Distribution

Binomial Distribution

\[ p = 0.7, \ n = 10 \]
Binomial Distribution

Binomial Distribution
p=1.0, n=10

Cumulative Probability

X

Estabrook & Brick
The Normal Distribution
Properties of a Probability Function

- So can any function be a probability mass or probability density function?
- What properties should a function like this have?
Properties of a Probability Function

- So can any function be a probability mass or probability density function?
- What properties should a function like this have?
- It should span all possible values of $X$.
  - In the case of a binomial $B(10,p)$, it should be defined for (whole number) values of $X$ from zero to ten.
Properties of a Probability Function

- So can any function be a probability mass or probability density function?
- What properties should a function like this have?
  - It should span all possible values of $X$.
    - In the case of a binomial $B(10,p)$, it should be defined for (whole number) values of $X$ from zero to ten.
  - All of the probabilities should add up to unity (1).
    - The CDF should range from zero to 1.
    - If it doesn’t, something’s wrong.
What about other distributions?
I don’t study coin flips and ESP.

▶ There are a very large number of distributions and distribution families you can use.
▶ One of the most common and useful distributions is the normal distribution.
  ▶ Also known as the Gaussian distribution, after Carl Fredrich Gauss.
  ▶ What Gauss Around? Heh! Heh? heh...
The Normal Distribution

- The normal distribution has a probability density function as follows:

$$f(X; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

- The normal distribution is defined by mean and variance. We already know those!
Normal Distribution

Normal Distribution
Mean=0, Variance=1

Probability Density Function

Value
Density
Normal Distribution

Normal Distribution
Mean=0, Variance=1

Cumulative Density Function

Value
Density

0.0 0.2 0.4 0.6 0.8 1.0
-2 -1 0 1 2

Estabrook & Brick University of Virginia
Normal Distribution

- The Normal Distribution is symmetric around the mean.
  - The mean is the most value of $X$ with the highest probability.
- The Normal Distribution is scaled by the standard deviation.
  - If you multiply $X$ by 2, both the mean and standard deviation double, and the variance quadruples.
  - 68.2% falls between $\mu \pm \sigma$, 95.4% between $\mu \pm 2\sigma$
- The Normal Distribution is defined from $-\infty$ to $\infty$.
  - Most computer approximations of the normal distribution place a cutoff beyond which the probability of $X$ is zero.
Density, Probability, & Frequency

- Probability density and cumulative density functions don’t exactly give us probabilities.
  - For instance, the density of $\mu$ is 0.3989.
  - Does that mean if we randomly draw from a normal distribution, we’ll pull out the exact mean almost 40% of the time?
  - The probabilities of the binomial added up to one, why doesn’t this?
- How do pdfs tell us probability when the sum of all of the densities is infinite?
  - The *area* under the density function adds up to one.
  - You can figure out the probability of $X$ being between $A$ and $B$ by integrating the pdf from $A$ to $B$. 
Stop.
Did he just say “integrate?”

- Ok, you don’t have to do calculus. Someone did it for you!
  - The cumulative density function is the integral of the probability density function.
  - The probability density function is the first derivative of the cumulative density function.

- How do I use this?
  - The cumulative density function is defined by $P(X \leq A)$.
  - So the probability of $X$ being between $A$ and $B$ ($A$ is bigger), is $P(X \leq A) - P(X \leq B)$.

- What if we wanted to find the probability of $X$ being between $\mu$ and $\mu + \sigma$?
What is the area between $\mu$ and $\mu + \sigma$?
What is the area between $\mu$ and $\mu + \sigma$
What is the area between $\mu$ and $\mu + \sigma$?

![Graph of the Normal Distribution]

- $P(X \leq 0) = 0.500$
- $P(X \leq 1) = 0.841$
What is the area between $\mu$ and $\mu + \sigma$

- The area between $\mu$ and $\mu + \sigma$ is:

$$P(\mu \leq X \leq \mu + \sigma)$$

$$= P(X \leq \mu + \sigma) - P(X \leq \mu)$$

$$= 0.841 - 0.500 = 0.341$$

- The probability of getting a value between $\mu$ and $\mu + \sigma$ is 0.341.

- You can get both of these values using the pnorm() functions in R.
Why Use the Normal Distribution?

- It’s easy to parameterize.
  - It only requires two parameters (mean & variance).
  - These parameters are so easy to understand, they’re used to describe lots of data.
- A number of statistical techniques are based on it.
  - All of the statistical procedures in this course assume some level of normality.
- A lot of data tends to be normally or approximately normally distributed.
- The normal distribution is the basis for many other distributions.
The Central Limit Theorem

Care of Hayes “Statistics for Psychologists," 1963

The Central Limit Theorem states:

If a population has a finite variance $\sigma^2$ and mean $\mu$ on some trait or variable $x$, then the distribution of sample means from samples of $N$ independent observations approaches a normal distribution with variance $\frac{\sigma^2}{N}$ and mean $\mu$ as sample $N$ increases. When $N$ is very large, the sampling distribution of the mean ($\bar{x}$) is approximately normal.
The Central Limit Theorem

- The Central Limit Theorem states that if you have a population and you’re curious about the mean, your sampling distribution has a mean of $\mu$ and a standard deviation of $\frac{\sigma^2}{N}$.

- Absolutely nothing is said about the form of the population distribution.

- Regardless of the distribution of your data, the sampling distribution of the mean of that variable is unbiased and approaches normality.
  - The only restriction is that the variance of the distribution you’re sampling from is finite, which usually isn’t a problem.
How does the Central Limit Theorem work?

- Let's consider a distribution that is decidedly not normal.
- If we just randomly sampled people from this population, our sample would mirror that population.
- However, the central limit theorem states that if we repeatedly sample this distribution, we can build a distribution of the means of those samples.
- This *sampling distribution* looks more and more like a normal distribution the more people go into our sample.
How does the Central Limit Theorem work?

This is our population.
How does the Central Limit Theorem work?
What would the distribution of samples means look like?

Sampling Distribution of the Mean
N=2
How does the Central Limit Theorem work?

What would the distribution of samples means look like?

Sampling Distribution of the Mean

N=3

Frequency

-4 -2 0 2 4

0 100 200 300 400 500

Estabrook & Brick University of Virginia

The Normal Distribution
How does the Central Limit Theorem work?
What would the distribution of samples means look like?

Sampling Distribution of the Mean
N=4
How does the Central Limit Theorem work?
What would the distribution of samples means look like?
How does the Central Limit Theorem work?

What would the distribution of samples means look like?

Sampling Distribution of the Mean

N=6

Frequency

0 100 200 300 400 500 600
How does the Central Limit Theorem work?
What would the distribution of samples means look like?
How does the Central Limit Theorem work?
What would the distribution of samples means look like?

**Sampling Distribution of the Mean**

N=10

Frequency

0

100

200

300

400

0

-2

0

2

Estabrook & Brick University of Virginia

The Normal Distribution
How does the Central Limit Theorem work?
What would the distribution of samples means look like?

Sampling Distribution of the Mean
N=15

Frequency
0 100 200 300 400 500

The Normal Distribution
How does the Central Limit Theorem work?
What would the distribution of samples means look like?
How does the Central Limit Theorem work?
What would the distribution of samples means look like?

Sampling Distribution of the Mean
N=50

Frequency

0 100 200 300 400
Central Limit Theorem

- So the central limit theorem works at the population level.
  - The sampling distribution of the mean approaches normality in the limit, making the mean an unbiased estimator of the population mean.
  - We also know the variance of that sampling distribution.
- But that doesn’t explain why the distributions of samples look normal so often.
  - The distribution of the people in a sample should mirror the distribution of the population.
Applying the CLT to Error.

- We can also apply the central limit theorem to measurement error.
- What is “error" or “measurement error?"
Applying the CLT to Error.

- We can also apply the central limit theorem to measurement error.
- What is “error" or “measurement error?"
- It’s all the stuff we don’t care about.
  - A better conceptualization for error is unique variance.
  - This construct consists of any number of things that aren’t related to the trait of interest.
Applying the CLT to Error.

▶ For example, let's consider a math test.

**Table**: Things A Math Test Might Measure

<table>
<thead>
<tr>
<th>Computational Ability</th>
<th>Reasoning Ability</th>
<th>gF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Comprehension</td>
<td>Language Ability</td>
<td>Memory</td>
</tr>
<tr>
<td>Cultural Factors</td>
<td>Test Anxiety</td>
<td>Education</td>
</tr>
</tbody>
</table>
Applying the CLT to Error.

- Error can be viewed as the sum of any number of factors you don’t care about.
- The mean is really nothing more than a scaled sum.
- While the central limit theorem describes repeated sampling from the same population, it has been extended to describe sums from different distributions.
  - Lyapunov condition, Lindeberg condition for independent distributions.
  - M-dependent and Martingale CLTs for dependent distributions.
- If each observation includes the sum of a variety of unique or error factors, the central limit theorem still holds.
Applying the CLT to Error.
Applying the CLT to Error.

Sampling Distribution for Sum of Distributions.

This is the sampling distribution of the mean for one contribution from each of the six distributions.
Applying the CLT to Error.

Sampling Distribution for Sum of Distributions.

This is the sampling distribution of the mean for five contributions from each of the six distributions.
Applying the CLT to Individual Measurements

- The earlier demonstrations can also apply to the distribution of measurements (true scores, or traits, or whatever).
- In the case of sum scores, the application is easy.
  - Many psychological measures are the sum of categorical items.
  - This is not an endorsement of sum scores/
- Just like error, whenever we can view a measurement as encompassing several components, the central limit theorem applies.
  - The aggregation is simply occurring prior to measurement.
Central Limit Theorem

Summary

- The Central Limit Theorem makes the normal distribution incredibly powerful.
  - The sampling distribution of the mean is unbiased & asymptotically normal regardless.
  - The sample itself may be normally distributed, depending on the characteristics of the trait itself and its associated measurement error.

- Data can also be normally distributed just by itself.
Central Limit Theorem

Pause for Questions.

▶ Ok, pausing.
The Normal and Other Distributions

- The Normal Distribution is also the basis of several other empirical distributions.
  - Log-normal (If \( X \) is normally distributed, and \( Y = e^X \), then \( Y \) is log-normal).
  - Cauchy (\( Y = \frac{X_1}{X_2} \)), Levy and Rayleigh distributions (this may be the only time you ever hear of these).
  - Many distributions (Binomial, Poisson, etc) can be approximated by the normal.
    - Abraham de Moivre

- The most common and important extension of the normal distribution is the \( \chi^2 \) distribution.
The $\chi^2$ Distribution

- The $\chi^2$ distribution is the square of the normal distribution.
- The version of the normal distribution used in the $\chi^2$ is the standard normal, with $\mu=0$ and $\sigma^2=1$.
  \[ X \sim N(0, 1) \]
  \[ \chi_1^2 = X^2 \]
- This version of the $\chi^2$ distribution has 1 degree of freedom.
The $\chi^2$ Distribution

- The $\chi^2$ distribution with $k$ degrees of freedom is actually the sum of $k$ independent squared standard normal distributions.

$$X_1 \sim \mathcal{N}(0, 1)$$
$$X_2 \sim \mathcal{N}(0, 1)$$

$X_1$ and $X_2$ are independent.

$$\chi^2_k = X_1^2 + X_2^2$$

- $\chi^3_3$ is the sum of three independent squared standard normals, etc.
The $\chi^2$ Distribution

- The $\chi^2$ increasingly approaches the normal distribution as the degrees of freedom ($k$) increases.
  - The $\chi^2$ is a sum, so the Central Limit Theorem still applies!
- There’s also one more important distribution based on these distributions.
The $F$ Distribution

- The $F$ distribution is an important distribution for analyses of variance.
- The $F$ distribution with $(d_1, d_2)$ degrees of freedom is the ratio of two $\chi^2$ distributions, divided by their degrees of freedom.

$$F(d_1, d_2) = \frac{\chi^2_{d_1}}{d_1} \frac{\chi^2_{d_2}}{d_2}$$

- By extension, the $F$ distribution is an extension of the normal.
Transformations of the Normal Distribution

Transformations are functions applied to variables, changing them from their original values to a new transformed value.

- Linear Transformations do not change the order or relative differences between observations.
- They can be useful for reformatting a poorly understood variable into an already understood scale.

Linear transformations take the following form:

\[ f(X) = aX + b \]
Transformations of the Normal Distribution

▶ One important application of the normal distribution is the \( z \) score.
   ▶ \( z \) scores are defined as \( z = \frac{X - \mu}{\sigma} \)
   ▶ \( z \) scores thus have a mean of zero and a standard deviation of one.

▶ Another application of the normal distribution is the \( T \) score.
   ▶ \( T \) scores have a mean of fifty and a standard deviation of ten.
Is My Data Normal?

- The normal distribution can be a very useful model of one's data.
- However, there is no guarantee that one's data is normal.
- How can you tell?
  - Statistical Techniques
  - Graphical Techniques
Sample Moments

- Last time, we briefly discussed sample moments and moments around the mean.
- Sample moments are the expected values of $X$ raised to various powers.
- For instance, the first sample moment is the expected value of $X$, which is the mean:
  
  $$E(X) = \text{mean}$$

- The $2^{nd}$ sample moment is the expected value of $E(X^2)$, $3^{rd}$ is $E(X^3)$, and so on.
Other Sample Moments

- Sample moments are centered at zero (i.e. the origin), but it usually makes sense to center them around the mean of your data.
- The second moment around the mean is then variance:
  \[ E([X - E(X)]^2) = \sigma^2 \]
- The normal distribution is defined by the first two moments (mean and variance).
- All other moments should be zero!
The third and fourth moments about the mean are skewness and kurtosis.

They are typically standardized, as shown below.

Formulas:

\[
S = \frac{\mu^3}{\sigma^3} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^3
\]

\[
K = \frac{\mu^4}{\sigma^4} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^4
\]
Using Sample Moments

- We can test the normality of a set of data by testing whether skewness and kurtosis are equal to zero.
- This test is known as the Jarque-Berra test, and is specified as follows:

\[
JB = \frac{n}{6}(S^2 + \frac{(K - 3)^2}{4})
\]

- Kurtosis as defined before has an expected value of 3; most statistics packages subtract 3 from all calculations of kurtosis.
- For normal distributions, this statistic is approximately chi-square distributed with two degrees of freedom.
Problems with Statistical Normality Tests

- This (and many other normality tests) have some potential problems.
  - Normality test statistics are sensitive to sample size.
  - In small samples, very large deviations from normality will still look normal.
  - In large samples, much smaller deviations will "fail" this test.
  - There’s no value for a "big" deviation from normality.

- While this test was a good segue, graphical techniques are an easier and better understood way to check normality.
Graphical Normality Tests

- Graphical normality tests simply plot the data in some way and rely on your eye to decide what’s normal and what’s not.
- We could just compare a histogram, pdf or cdf of our data to a theoretically defined version and see if they look the same.
Graphical Normality Tests

Binomial Distribution
\( p=0.5, n=10 \)

Normal Distribution
Mean=0, Variance=1

Probability Density Function
Graphical Normality Tests

Normal Distribution
Mean=0, Variance=1
Probability Density Function
Value
Density

-4 -3 -2 -1 0 1 2 3 4
0.0 0.1 0.2 0.3 0.4
A Better Way

- We can combine all of that information into one plot, called a Q-Q or Quantile-Quantile plot.
  - On one axis is our data.
  - On the other axis is an empirical normal distribution with the same mean and variance as our data.
  - We plot each observation in our data against its parallel quantile in that theoretical normal distribution.

- This approach has tremendous benefits:
  - All of the information is in one graph.
  - It’s simple and clear.
  - It depends on something we’re good at: straight lines.
QQPlot

Data is Normal.
QQPlot

Data is Binomial - the Normal is a Decent Approximation.

Histogram
n=100, p=.5
rbinom(1000, 100, 0.5)
Frequency
35 40 45 50 55 60 65
0 50 100 150

Q-Q Plot
n=100, p=.5
Theoretical Quantiles
Sample Quantiles
35 40 45 50 55 60 65
-3 -2 -1 0 1 2 3
QQPlot

Data is Not Normal

c(n(500, -3, 1), n(500, 3, 1))

Frequency

-6 -4 -2 0 2 4 6
0 50 100 150

Transformations Evaluation

Q−Q Plot

Theoretical Quantiles

Sample Quantiles

Hist

Q−Q Plot
Summary

- Distributions are just ways of describing probability and frequency, given that a set of numbers can be described by a given set of parameters.
- The normal distribution has a number of benefits, many of them related to the central limit theorem.
- A number of other distributions common in psychology are either based on the normal distribution or can be approximated by the normal distribution.
- While a number of statistics exist for normality evaluation, Q-Q plots are the easiest and most useful tool for evaluating distributional qualities.
Next Time

- Sampling, in all its forms.
- Dealing with imprecision.
- Resampling.
- Developing tests, research questions and the like.