Intro to Likelihood

Gov 2001 Section

February 2, 2012

Outline

Replication Paper

- Read "How to Write a Publishable Paper" on Gary's website and "Publication, Publication".
- Find a partner.
- Find a set of papers you would be interested in replicating.
  1. Recently published (in the last two years).
  2. From a good journal.
  3. Use methods at least as sophisticated as in this class.
- E-mail us (Gary, Jen, and Molly) to get our opinion.
- Find the data.

Outline

An R Note on the Homework

- How would we find the expected value of a distribution without simulating in R?
- For example, $Y \sim Normal(\mu, \sigma^2)$, where $\mu = 6, \sigma^2 = 3$.
- In math, we want to integrate

$$\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

- Plugging in for μ and σ

$$\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2 * 3\pi}} e^{-\frac{(x-6)^2}{2*3}} dx$$

An R Note on the Homework cont

1. First, we would write a function of what we want to integrate out:

```
ex.normal <- function(x){
x*1/(sqrt(6*pi))*exp(-(x-6)^2/6)
}
```

2. Use integrate to get the expected value.

```
integrate(ex.normal, lower=-Inf, upper=Inf)

6 with absolute error < 0.00016
```

Outline

Why become familiar with probability distributions?

- You can fit models to a variety of data.

What do you have to do to use probability distributions?

- You have to recognize what data you are working with.
- What's the best way to learn the distributions? Learn the "stories" behind them.

Outline

The Bernoulli Distribution

- Takes value 1 with success probability π and value 0 with failure probability $1 - \pi$.
- Ideal for modelling **one-time** yes/no (or success/failure) events.
- The best example is **one** coin flip – if your data resemble a single coin flip, then you have a Bernoulli distribution.
- ex) one voter voting yes/no
- ex) one person being either a man/woman
- ex) the Patriots winning/losing the Super Bowl

The Bernoulli Distribution

$Y \sim \text{Bernoulli}(\pi)$

$y = 0, 1$

probability of success: $\pi \in [0, 1]$

$p(y|\pi) = \pi^y (1 - \pi)^{(1-y)}$

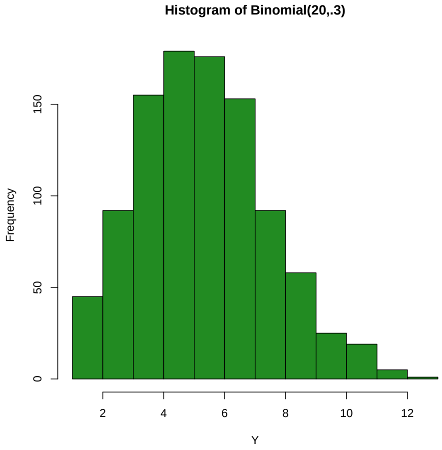
$E(Y) = \pi$

$\text{Var}(Y) = \pi(1 - \pi)$

The Binomial Distribution

- The Binomial distribution is the total of a bunch of Bernoulli trials.
- You flip a coin three times and count the total number of heads you got. (The order doesn't matter.)
- The number of women in a group of 10 Harvard students
- The number of rainy days in the seven-day week

The Binomial Distribution



$$Y \sim \text{Binomial}(n, \pi)$$

$$y = 0, 1, \ldots, n$$

number of trials: $n \in \{1, 2, \ldots\}$

probability of success: $\pi \in [0, 1]$

$$p(y|\pi) = \binom{n}{y} \pi^y (1-\pi)^{(n-y)}$$

$$E(Y) = n\pi$$

$$\text{Var}(Y) = n\pi(1-\pi)$$

The Multinomial Distribution

- Suppose you had more than just two outcomes – e.g., vote for Republican, Democrat, or Independent. Can you use a binomial?
- We can't use a binomial, because a binomial requires two outcomes(yes/no, 1/0, etc.). Instead, we use the multinomial.
- Multinomial lets you work with several mutually exclusive outcomes.

  For example:

  - you toss a die 15 times and get outcomes 1-6
  - ten undergraduate students are classified freshmen, sophomores, juniors, or seniors
  - Gov graduate students divided into either American, Comparative, Theory, or IR

The Multinomial Distribution

$Y \sim \text{Multinomial}(n, \pi_1, \dots, \pi_k)$

$y_j = 0, 1, \dots, n; \;\; \sum_{j=1}^{k} y_j = n$

number of trials: $n \in \{1, 2, \dots\}$

probability of success for j: $\pi_j \in [0, 1]; \;\; \sum_{j=1}^{k} \pi_j = 1$

$p(\mathbf{y}|n, \boldsymbol{\pi}) = \frac{n!}{y_1! y_2! \dots y_k!} \pi_1^{y_1} \pi_2^{y_2} \dots \pi_k^{y_k}$

$E(Y_j) = n\pi_j$

$\text{Var}(Y_j) = n\pi_j(1 - \pi_j)$

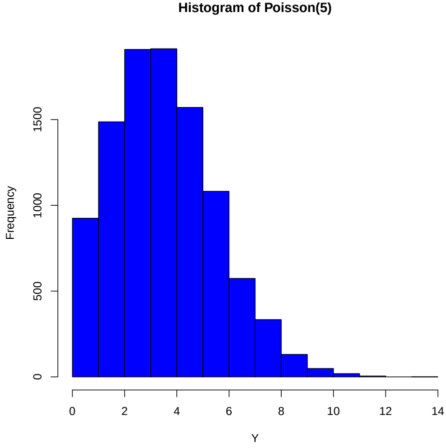
The Poisson Distribution

- Represents the number of events occurring in a fixed period of time.
- Can also be used for the number of events in other specified intervals such as distance, area, or volume.
- Can never be negative – so, good for modeling events.

  For example:

  - The number Prussian solders who died each year by being kicked in the head by a horse (Bortkiewicz, 1898)
  - The of number shark attacks in Australia per month
  - The number of search warrant requests a federal judge hears in one year

The Poisson Distribution



$Y \sim \text{Poisson}(\lambda)$

$y = 0, 1, \ldots$

expected number of occurrences: $\lambda > 0$

$$p(y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$$

$$E(Y) = \lambda$$

$$\text{Var}(Y) = \lambda$$

Outline

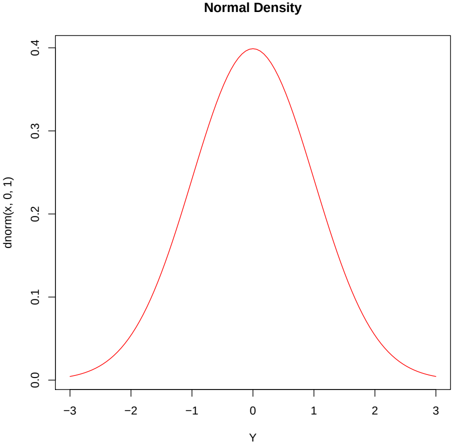
The Univariate Normal Distribution

- Describes data that cluster in a bell curve around the mean.

  For example:

  - the heights of male students in our class
  - high school students' SAT scores

The Univariate Normal Distribution



$$Y \sim \text{Normal}(\mu, \sigma^2)$$

$$y \in \mathbb{R}$$

mean: $\mu \in \mathbb{R}$

variance: $\sigma^2 > 0$

$$p(y|\mu, \sigma^2) = \frac{\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}$$

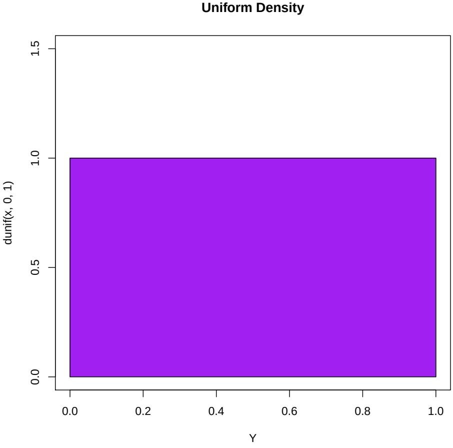
$$E(Y) = \mu$$

$$\text{Var}(Y) = \sigma^2$$

The Uniform Distribution

- Any number in the interval you chose is equally probable. Intuitively easy to understand, but often examples are discrete
    - the number of a person who comes in first in a race (discrete)
    - the lottery tumblers out of which a person draws one ball with a number on it (also discrete)

The Uniform Distribution



$$Y \sim \text{Uniform}(\alpha, \beta)$$

$$y \in [\alpha, \beta]$$

Interval: $[\alpha, \beta]; \quad \beta > \alpha$

$$p(y|\alpha, \beta) = \frac{1}{\beta - \alpha}$$

$$E(Y) = \frac{\alpha + \beta}{2}$$

$$\text{Var}(Y) = \frac{(\beta - \alpha)^2}{12}$$

Quiz: Test Your Knowledge of Discrete Distributions

Are the following Bernoulli (coin flip), Binomial(several coin flips), Multinomial (Rep, Dem, Indep), Poisson (Prussian soldier deaths), Normal (SAT scores), or Uniform (race numbers)?

- The heights of trees on campus?
- The number of airplane crashes in one year?
- A yes or no vote cast by Senator Brown?
- The number of parking tickets Cambridge PD gives out in one month?
- The poll your Facebook friends took to choose their favorite sport out of football, basketball, and soccer
- The time until a country adopts a treaty?

Outline

Likelihood

The whole point of likelihood is to leverage information about the data generating process into our inferences.

Here are the basic steps:

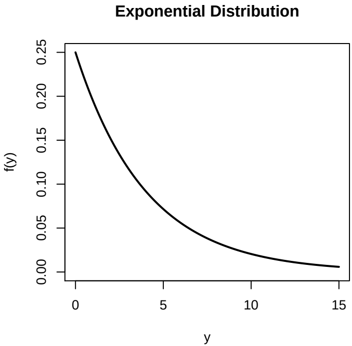
- Think about your data generating process. (What do the data look like? Use your substantive knowledge.)
- Find a distribution that you think explains the data. (Poisson, Binomial, Normal? Something else?)
- Derive the likelihood.
- Maximize the likelihood to get the MLE.

Note: This is the case in the univariate context. We'll be introducing covariates later on in the term.

Likelihood: An Example

Ex. Waiting for the Redline – How long will it take for the next T to get here?

Likelihood: Waiting for the Redline



Y is a Exponential random variable with parameter $\lambda = .25$.

$$f(y) = \lambda e^{-\lambda y} = .25e^{-.25y}$$

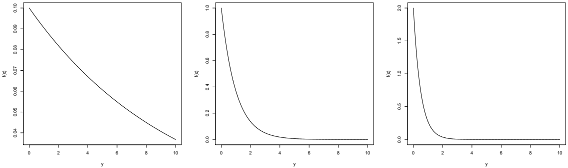
Likelihood: Waiting for the Redline

Last week we assumed λ to get the probability of waiting for the redline for X mins.

- $\lambda = .25 \rightarrow data.$
- $p(y|\lambda = .25) = .25e^{-.25y} \rightarrow p(2 < y < 10|\lambda) = .525$

This week we will observe the data to get the probability of λ.

- $data \rightarrow \lambda.$
- $p(\lambda|y) = ?$



Likelihood: Waiting for the Redline

- From Bayes' Rule:

$$p(\lambda|y) = \frac{p(y|\lambda)p(\lambda)}{p(y)}$$

- Let

$$k(y) = \frac{p(\lambda)}{p(y)}$$

(Note that the λ in $k(y)$ is the true λ, a constant that doesn't vary. So $k(y)$ is just a function of y.)

- Define $L(\lambda|y) = p(y|\lambda)k(y)$
- $\rightarrow L(\lambda|y) \propto p(y|\lambda)$

Monday Data

$$\begin{aligned} L(\lambda|y_1) &\propto p(y_1|\lambda) \\ &= \lambda e^{-\lambda * y_1} \\ &= \lambda e^{-\lambda * 12} \end{aligned}$$

Plotting the likelihood

First, note that we can take advantage of a lot of pre-packaged R functions

- `rbinom`, `rpoisson`, `rnorm`, `runif` → gives random values from that distribution
- `pbinom`, `ppoisson`, `pnorm`, `punif`→ gives the cumulative distribution (the probability of that value or less)
- `dbinom`, `dpoisson`, `dnorm`, `dunif`→ gives the density (i.e., height of the PDF – useful for drawing)
- `qbinom`, `qpoisson`, `qnorm`, `qunif`→ gives the quantile function (given quantile, tells you the value)

Plotting the example

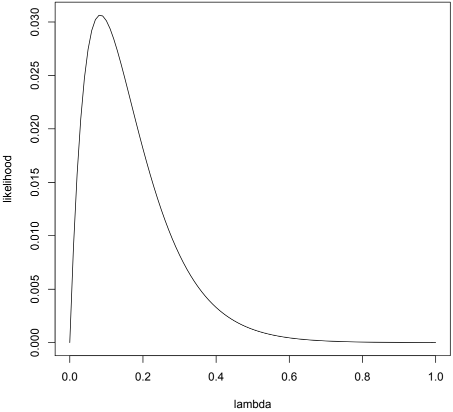
We want to plot $L(\lambda|y) \propto \lambda e^{-\lambda * 12}$

```
dexp(x, rate, log=FALSE)
```

```
e.g. dexp(12, .25)
[1] 0.01244677
```

```
curve(dexp(12, rate = x),
  xlim =c(0,1), xlab ="lambda", ylab = "likelihood")
```

Plotting the example



What do you think the maximum likelihood estimate will be?

Solving Using R

1. Write a function.

```
expon <- function(lambda,data) {
-lambda*exp(-lambda*data)
}
```

2. Optimize.

```
optimize(f=expon, data=12, lower=0, upper=100)
```

3. Output

```
$minimum
[1] 0.0833248

$objective
[1] -0.03065662
```

Where are we going with this?

- What if we have two or more data points that we believe come from the same model?
- We can derive a likelihood for the combined data by multiplying the independent likelihoods together.

Tuesday Data

$$
\begin{aligned}
L(\lambda|y_2) &\propto p(y_2|\lambda) \\
&= \lambda e^{-\lambda * y_2} \\
&= \lambda e^{-\lambda * 7}
\end{aligned}
$$

Likelihood for Monday and Tuesday

Remember that for independent events:

$$P(A, B) = P(A)P(B)$$

$$\begin{aligned} L(\lambda | y_1, y_2) &= \lambda e^{-\lambda * y_1} * \lambda e^{-\lambda * y_2} \\ &= \lambda e^{-\lambda * 12} * \lambda e^{-\lambda * 7} \end{aligned}$$

A Whole Week of Data

$$
\begin{aligned}
L(\lambda | y_1 \dots y_5) &= \prod_{i=1}^{5} \lambda e^{-\lambda * y_i} \\
&= \lambda e^{-\lambda * y_1} * \lambda e^{-\lambda * y_2} * \lambda e^{-\lambda * y_3} * \lambda e^{-\lambda * y_4} * \lambda e^{-\lambda * y_5} \\
&= \lambda e^{-\lambda * 12} * \lambda e^{-\lambda * 7} * \lambda e^{-\lambda * 4} * \lambda e^{-\lambda * 19} * \lambda e^{-\lambda * 2}
\end{aligned}
$$

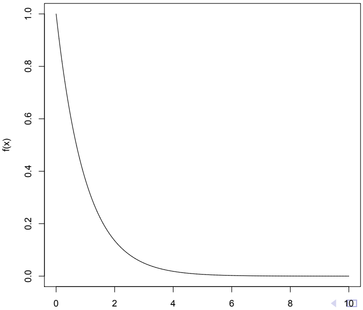
Outline

Transforming Distributions

- $X \sim p(x|\theta)$
- $y = g(x)$
- How is y distributed?

For example, if $X \sim Exponential(\lambda = 1)$ and $y = log(x)$

$$y \sim ?$$



Transforming Distributions

- It is NOT true that $p(y|\theta) \sim g(p(x|\theta))$. Why?

Transforming Distributions

The Rule

- $X \sim p_x(x|\theta)$
- $y = g(x)$

$$p_y(y) = p_x(g^{-1}(y)) \left| \frac{dg^{-1}}{dy} \right|$$

- What is $g^{-1}(y)$?
- What is $\left| \frac{dg^{-1}}{dy} \right|$? The Jacobian.

Transforming Distributions – the log-Normal Example

For example,

- $X \sim Normal(x|\mu = 0, \sigma = 1)$
- $y = g(x) = e^x$
- what is $g^{-1}(y)$?

$$g^{-1}(y) = x = log(y)$$

- What is $\frac{dg^{-1}}{dy}$?

$$\frac{d(log(y))}{dy} = \frac{1}{y}$$

Transforming Distributions – the log-Normal Example

- Put it all together

$$p_y(y) \quad = \quad p_x\left(log(y)\right)\left|\frac{1}{y}\right|$$

- Notice we don't need the absolute value because $y > 0$.

$$p_y(y) \quad = \quad \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(log(y))^2} \frac{1}{y}$$

- $Y \sim$ log-Normal$(0, 1)$
- Challenge: derive the chi-squared distribution.