Lecture 7

Brian Caffo

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Johns Hopkins University

October 22, 2007

Table of contents

Outline

1. Define the Bernoulli distribution
2. Define Bernoulli likelihoods
3. Define the Binomial distribution
4. Define Binomial likelihoods
5. Define the normal distribution
6. Define normal likelihoods

The Bernoulli distribution

- The **Bernoulli distribution** arises as the result of a binary outcome
- Bernoulli random variables take (only) the values 1 and 0 with a probabilities of (say) p and $1 - p$ respectively
- The PMF for a Bernoulli random variable X is

$$P(X = x) = p^x (1 - p)^{1-x}$$

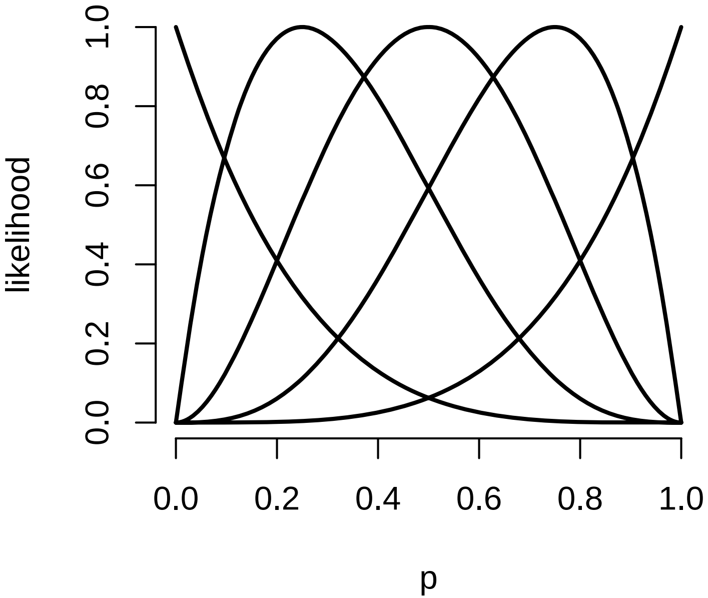
- The mean of a Bernoulli random variable is p and the variance is $p(1 - p)$
- If we let X be a Bernoulli random variable, it is typical to call $X = 1$ as a "success" and $X = 0$ as a "failure"

iid Bernoulli trials

- If several iid Bernoulli observations, say $x_1, \ldots, x_n$, are observed the likelihood is

$$\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

- Notice that the likelihood depends only on the sum of the x_i
- Because n is fixed and assumed known, this implies that the sample proportion $\sum_i x_i/n$ contains all of the relevant information about p
- We can maximize the Bernoulli likelihood over p to obtain that $\hat{p} = \sum_i x_i/n$ is the maximum likelihood estimator for p


likelihood
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
p

Binomial trials

- The **binomial random variables** are obtained as the sum of iid Bernoulli trials
- In specific, let $X_1, \ldots, X_n$ be iid Bernoulli(p); then $X = \sum_{i=1}^n X_i$ is a binomial random variable
- The binomial mass function is

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

for $x = 0, \ldots, n$

- Recall that the notation

$$\left(\begin{array}{c} n \\ x \end{array}\right) = \frac{n!}{x!(n-x)!}$$

(read "n choose x") counts the number of ways of selecting x items out of n without replacement disregarding the order of the items

-

$$\left(\begin{array}{c} n \\ 0 \end{array}\right) = \left(\begin{array}{c} n \\ n \end{array}\right) = 1$$

Example justification of the binomial likelihood

- Consider the probability of getting 6 heads out of 10 coin flips from a coin with success probability p
- The probability of getting 6 heads and 4 tails in any specific order is

$$p^6(1-p)^4$$

- There are

$$\left(\begin{array}{c}10\\6\end{array}\right)$$

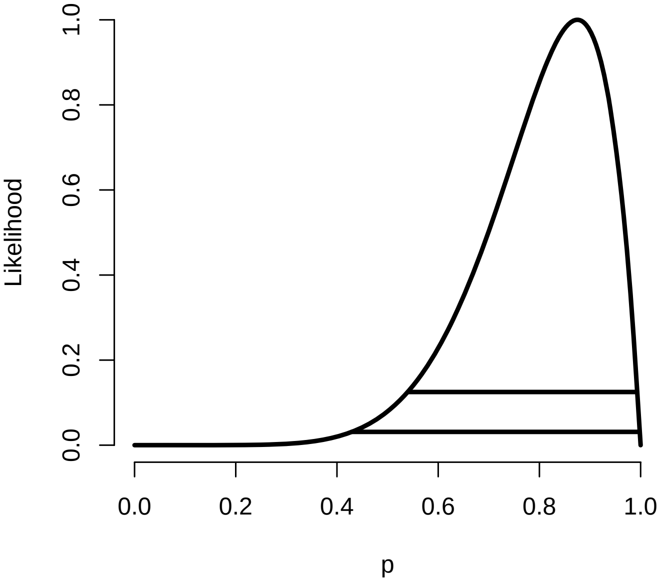
possible orders of 6 heads and 4 tails

Example

- Suppose a friend has 8 children, 7 of which are girls and none are twins
- If each gender has an independent 50% probability for each birth, what's the probability of getting 7 or more girls out of 8 births?

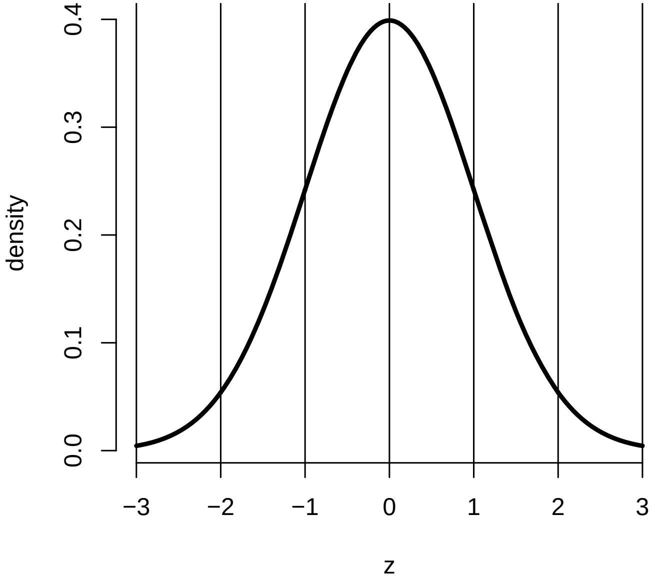
$$\binom{8}{7} .5^7(1-.5)^1 + \binom{8}{8} .5^8(1-.5)^0 \approx 0.04$$

- This calculation is an example of a $P - value$ - the probability under a null hypothesis of getting a result as extreme or more extreme than the one actually obtained


Likelihood
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
p

The normal distribution

- A random variable is said to follow a **normal** or **Gaussian** distribution with mean μ and variance σ^2 if the associated density is
$$(2\pi\sigma^2)^{-1/2}e^{-(x-\mu)^2/2\sigma^2}$$
If X a RV with this density then $E[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$
- We write $X \sim \mathrm{N}(\mu, \sigma^2)$
- When $\mu = 0$ and $\sigma = 1$ the resulting distribution is called **the standard normal distribution**
- The standard normal density function is labeled ϕ
- Standard normal RVs are often labeled Z


density
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3
z

Facts about the normal density

- If $X \sim \mathrm{N}(\mu, \sigma^2)$ the $Z = \frac{X-\mu}{\sigma}$ is standard normal
- If Z is standard normal

$$X = \mu + \sigma Z \sim \mathrm{N}(\mu, \sigma^2)$$

- The non-standard normal density is

$$\phi\{(x-\mu)/\sigma\}/\sigma$$

More facts about the normal density

1. Approximately 68%, 95% and 99% of the normal density lies within 1, 2 and 3 standard deviations from the mean, respectively
2. -1.28, -1.645, -1.96 and -2.33 are the 10^{th}, 5^{th}, 2.5^{th} and 1^{st} percentiles of the standard normal distribution respectively
3. By symmetry, 1.28, 1.645, 1.96 and 2.33 are the 90^{th}, 95^{th}, 97.5^{th} and 99^{th} percentiles of the standard normal distribution respectively

Question

- What is the 95^{th} percentile of a $N(\mu, \sigma^2)$ distribution?
- We want the point x_0 so that $P(X \leq x_0) = .95$

$$
\begin{aligned}
P(X \leq x_0) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{x_0-\mu}{\sigma}\right) \\
&= P\left(Z \leq \frac{x_0-\mu}{\sigma}\right) = .95
\end{aligned}
$$

- Therefore
$$\frac{x_0-\mu}{\sigma} = 1.645$$
or $x_0 = \mu + \sigma 1.645$
- In general $x_0 = \mu + \sigma z_0$ where z_0 is the appropriate standard normal quantile

Question

- What is the probability that a $\mathrm{N}(\mu, \sigma^2)$ RV is 2 standard deviations above the mean?
- We want to know

$$
\begin{aligned}
P(X > \mu + 2\sigma) &= P\left(\frac{X - \mu}{\sigma} > \frac{\mu + 2\sigma - \mu}{\sigma}\right) \\
&= P(Z \geq 2) \\
&\approx 2.5\%
\end{aligned}
$$

Other properties

1. The normal distribution is symmetric and peaked about its mean (therefore the mean, median and mode are all equal)
2. A constant times a normally distributed random variable is also normally distributed (what is the mean and variance?)
3. Sums of normally distributed random variables are again normally distributed even if the variables are dependent (what is the mean and variance?)
4. Sample means of normally distributed random variables are again normally distributed (with what mean and variance?)
5. The square of a *standard normal* random variable follows what is called **chi-squared** distribution
6. The exponent of a normally distributed random variables follows what is called the **log-normal** distribution
7. As we will see later, many random variables, properly normalized, *limit* to a normal distribution

Question

If X_i are iid $\mathrm{N}(\mu, \sigma^2)$ with a known variance, what is the likelihood for μ?

$$
\begin{aligned}
\mathcal{L}(\mu) &= \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2}\ exp\left\{-(x_i - \mu)^2/2\sigma^2\right\} \\
&\propto \exp\left\{-\sum_{i=1}^{n}(x_i - \mu)^2/2\sigma^2\right\} \\
&= \exp\left\{-\sum_{i=1}^{n} x_i^2/2\sigma^2 + \mu\sum_{i=1}^{n} X_i/\sigma^2 - n\mu^2/2\sigma^2\right\} \\
&\propto \exp\left\{\mu n\bar{x}/\sigma^2 - n\mu^2/2\sigma^2\right\}
\end{aligned}
$$

Later we will discuss methods for handling the unknown variance

Question

- If X_i are iid $\mathrm{N}(\mu, \sigma^2)$, with known variance what's the ML estimate of μ?
- We calculated the likelihood for μ on the previous page, the log likelihood is

$$\mu n\bar{x}/\sigma^2 - n\mu^2/2\sigma^2$$

- The derivative with respect to μ is

$$n\bar{x}/\sigma^2 - n\mu/\sigma^2 = 0$$

- This yields that $\bar{x}$ is the ml estimate of μ
- Since this doesn't depend on σ it is also the ML estimate with σ unknown

Final thoughts on normal likelihoods

- The maximum likelihood estimate for σ^2 is

$$\frac{\sum_{i=1}^{n}(X_i - \bar X)^2}{n}$$

  Which is the biased version of the sample variance
- The ML estimate of σ is simply the square root of this estimate
- To do likelihood inference, the bivariate likelihood of (μ, σ) is difficult to visualize
- Later, we will discuss methods for constructing likelihoods for one parameter at a time