Lecture 7

Ciprian Crainiceanu

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

September 8, 2020

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

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

- 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
- Maximizing the Bernoulli likelihood over p: $\hat{p} = \sum_i x_i/n$ is the maximum likelihood estimator for p

Maximizing the Bernoulli likelihood

- The log likelihood is

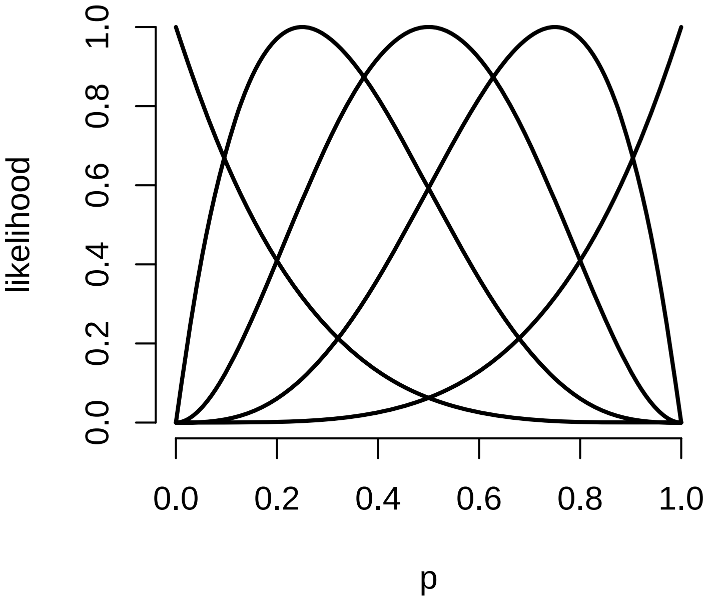
$$2\log \mathcal{L}(p|\mathbf{x}) = \log(p)\sum x_i + \log(1-p)(n-\sum x_i)$$

- The first derivative

$$\begin{aligned}
\frac{\partial}{\partial p}\log \mathcal{L}(p|\mathbf{x}) &= \frac{\sum x_i}{p} - \frac{n-\sum x_i}{1-p} = 0 \\
(1-p)\sum x_i - p(n-\sum x_i) &= 0 \\
p &= \frac{\sum x_i}{n}
\end{aligned}$$

- The second derivative

$$\frac{\partial^2}{\partial p^2} 2\log \mathcal{L}(p|\mathbf{x}) = -\frac{\sum x_i}{p^2} - \frac{n-\sum x_i}{(1-p)^2} < 0$$


likelihood
0.0
0.2
0.4
0.6
0.8
1.0
p

Bernoulli lotteries

- You pay \$$c$ to enter the game
- A coin is flipped with probability p and the outcome is either 0 or 1
- If the outcome is 1 then you win \$$C$
- What is the expected gain from playing the game?
- What is interpretation of the expected gain?
- The stock market is, essentially, a lottery, with trillion of daily bets

Bernoulli lotteries

- You pay \$$c$ to enter the game
- A coin is flipped with probability p and the outcome is either 0 or 1
- If the outcome is 1 then you win \$$C$
- What is the expected gain from playing the game?
- What is interpretation of the expected gain?
- The stock market is, essentially, a lottery, with trillion of daily bets

Bernoulli lotteries

- One coin is flipped: you win \$$C$ with probability p and loose \$$c$ with probability $1-p$
- Show that this is the same lottery
- Would you play if $p = 0.5$, $C = 10,000,000$ and $c = 10,000,000$?
- Would you play if $p = 0.6$, $C = 10,000,000$ and $c = 10,000,000$?
- Would you play if $p = 0.9$, $C = 10,000,000$ and $c = 10,000,000$?
- Would you play if $p = 0.9$, $C = 10$ and $c = 10$?

Bernoulli lotteries

- Expected value and variance of the lottery matter
- Risk (variability matters)
- Calculate the variance of a lottery

The St. Petersburg Paradox

The pot starts at \$1 and is doubled every time a head appears. The first time a tail appears, the game ends and the player wins whatever is in the pot.

Earnings

- Win \$1 if T
- Win \$2 if: HT
- Win \$4 if: HHT

Question: What is a fair price to pay to play such a game?

The St. Petersburg Paradox

Let X be the amount of money in the pot on the last game.

- $X =$\$1 with probability $1/2$
- $X =$\$2 with probability $1/2^2$
- $X =$\$4 with probability $1/2^3$
- $P(X = 2^k) = 1/2^{k+1}$, $k \geq 0$
- The expectation of wins is

$$E[X] = 1\frac{1}{2} + 2\frac{1}{4} + 4\frac{1}{8} + \ldots = \infty$$

- If the expected win is what it counts then a fair price is infinite!

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

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

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

-

$$\binom{n}{0} = \binom{n}{n} = 1$$

Binomial in R

```
# Simulate 15 independent Binomial(10,.2)
rbinom(15,size=10,prob=0.2)

# Simulate 15 independent Binomial(10,p)
# First p=0.1, last p=0.9
rbinom(15,size=10,prob=seq(0.1,0.9,length=15))

# Simulate 15 independent Binomial(n,p)
# First (n,p)=(1,0.1), last (n,p)=(15,0.9)
rbinom(15,size=1:15,prob=seq(0.1,0.9,length=15))
```

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

Example

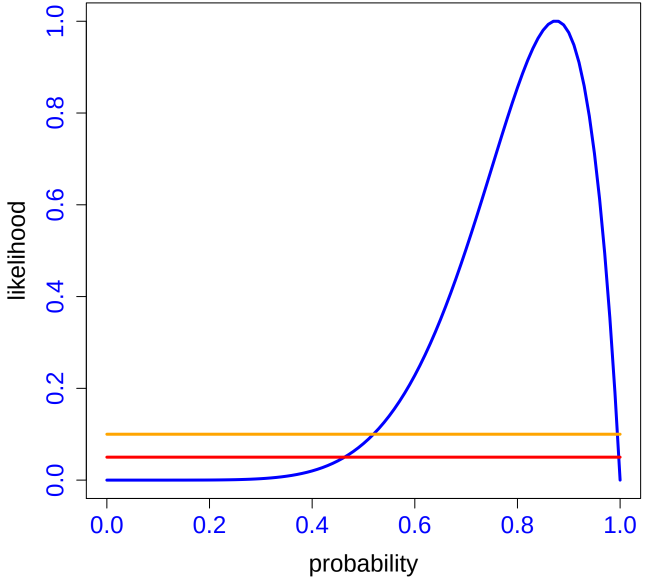
- Suppose that we do not know what is the probability p of having a girl
- Observe: FFMFFFFF
- Likelihood

$$\mathcal{L}(p|\mathbf{x}) = \left(\begin{array}{c} 8 \\ 7 \end{array} \right) p^7(1-p) = \frac{8!}{7!1!} p^7(1-p) = 8p^7(1-p)$$

- Suppose that you have a working hypothesis that $H_0 : p_0 = 0.5$
- What is the probability of observing at least 7F among 8 children?
- Answer: the p-value=0.04
- If $p_0 = 0.5$ then 4% of the families with eight children will have at least 7 girls

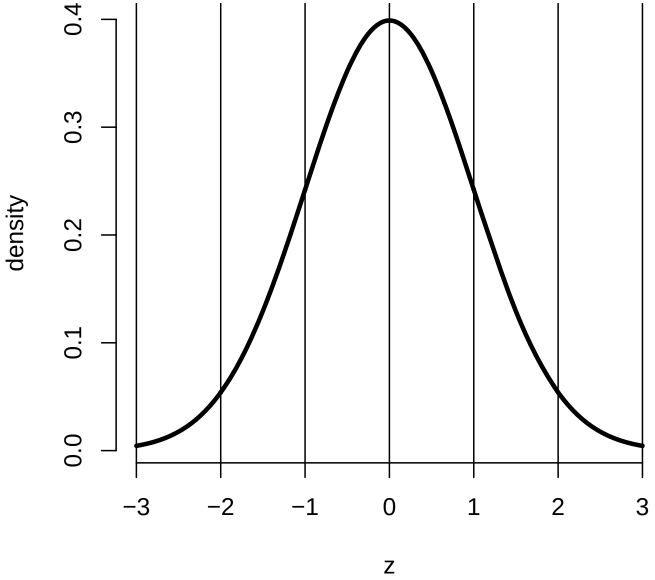
Example: R

```
p=seq(0,1,length=101)
like=8*p^7*(1-p)
plot(p,like/max(like),type="l",col="blue",lwd=3)
lines(p,rep(0.05,101),col="red",lwd=3)
lines(p,rep(0.1,101),col="orange",lwd=3)
```


likelihood
0.0
0.2
0.4
0.6
0.8
1.0
probability
0.0
0.2
0.4
0.6
0.8
1.0

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
- Standardizing RVs is referred to as the Z-score transform


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

Example: R

```
# Probability of at most 1σ deviation
pnorm(1)-pnorm(-1)
# Probability of at most 2σ deviation
pnorm(2)-pnorm(-2)
# Probability of at most 3σ deviation
pnorm(3)-pnorm(-3)
```

If $X \sim N(\mu, \sigma^2)$ then

$$P\{X \in (\mu - 1.96\sigma, \mu + 1.96\sigma)\} = 0.95$$

Facts about the normal density

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

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

- The non-standard normal density is

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

Facts about the normal density

- If Z is standard normal $X = \mu + \sigma Z$
- $E[X] = E[\mu + \sigma Z] = \mu + \sigma E[Z] = \mu$
- $\mathrm{Var}[X] = \mathrm{Var}[\mu + \sigma Z] = \sigma^2 \mathrm{Var}[Z] = \sigma^2$

Skewness

- If X is a random variable the skewness is

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

- Calculate the skewness of a normal distribution
- Assume that X is a symmetric distribution, what is its skewness?

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

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

Joint distribution for μ, σ

- If $X_1 = x_1, \ldots, X_n = x_n$ are n iid samples from $N(\mu, \sigma^2)$ then what does the joint likelihood look like?
- The joint likelihood is proportional to

$$\mathcal{L}(\mu, \sigma|\mathbf{x}) = \exp\left\{-\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2} - n\log(\sigma)\right\}$$

- We need two ingredients:
  - A grid of values for μ and σ
  - Plotting of 3D images

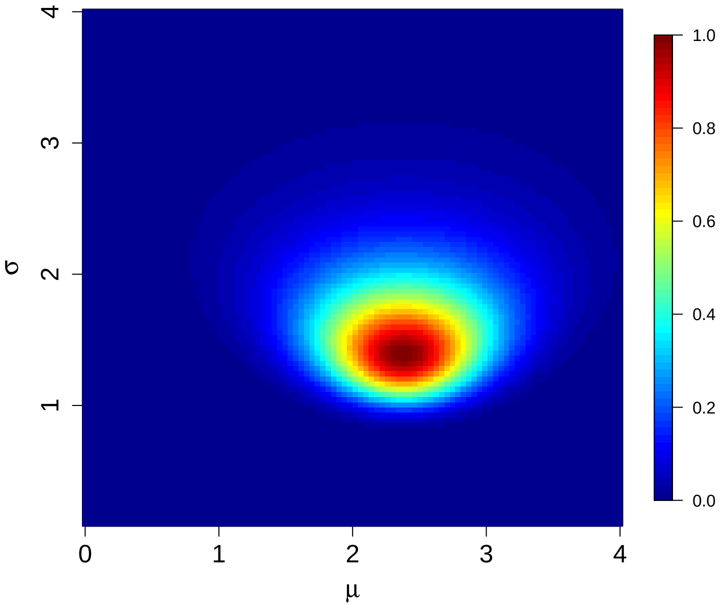
Joint distribution in R

```
x=rnorm(10,3,2)
n=length(x)
mu=seq(0,4,length=100)
sigma=seq(0.1,4,length=100)
like=matrix(rep(0,10000),ncol=100)

for (i in 1:100)
{for (j in 1:100)
   {like[i,j]=exp(-sum((x-mu[i])^2)/
   (2*sigma[j]^2)-n*log(sigma[j]))}}
like=like/max(like)
```

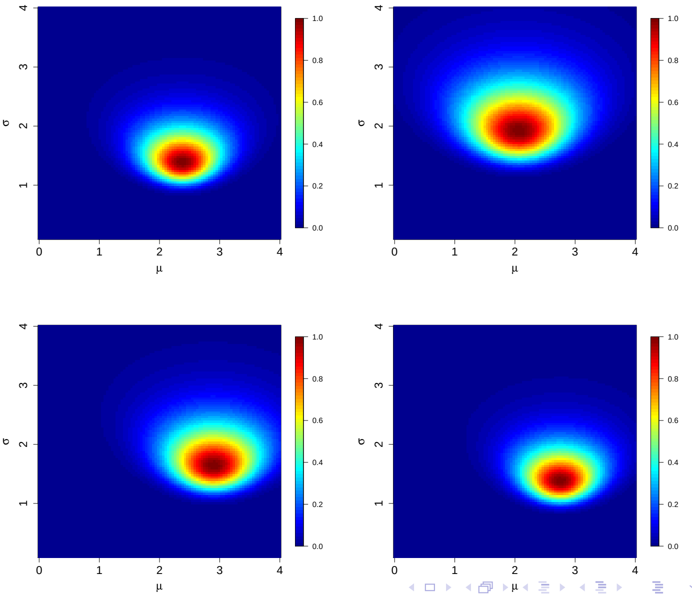
Joint distribution in R

```
library(fields)
image.plot(mu,sigma,like,xlab=expression(mu),
ylab=expression(sigma),cex.lab=1.5,cex.axis=1.5)
```

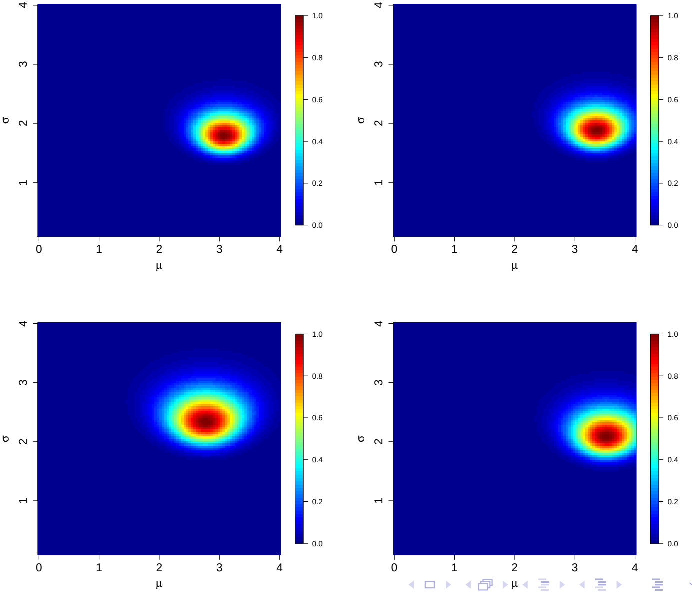


4
3
2
1
σ
0
1
2
3
4
μ
1.0
0.8
0.6
0.4
0.2
0.0

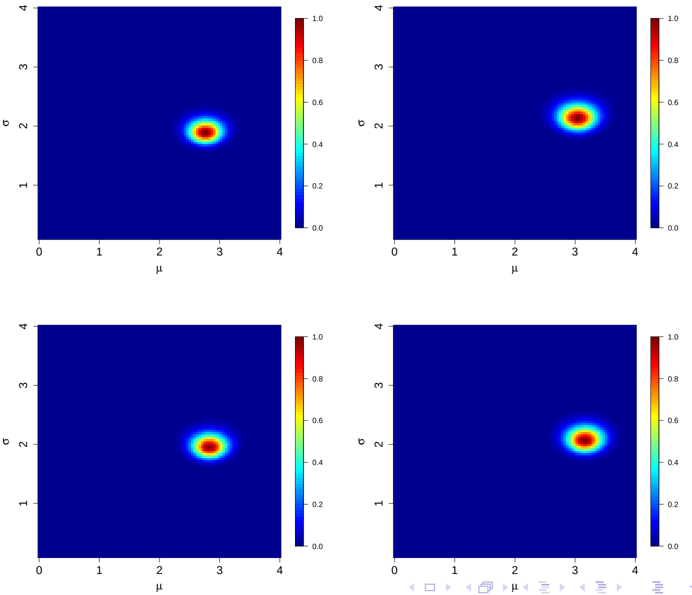
Joint distribution: $n = 10$



Joint distribution: $n = 30$



Joint distribution: $n = 100$



Some observations

- MLEs get closer to their true values when $n \uparrow \infty$
- Likelihood concentration $\uparrow$ around the MLEs as $n \uparrow \infty$
- Marginal likelihoods become more "bell-shaped" as $n \uparrow \infty$
- Confidence about the true value of the parameter increases with $n \uparrow \infty$
- Confidence about the values is not necessarily symmetric around the MLE
- Variability around the MLE goes to 0 but is never 0
- Likelihood depends on the observed sample