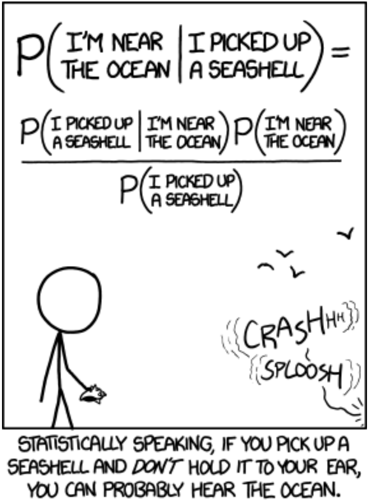
Bayesian Updating: Discrete Priors: 18.05 Spring 2014

Jeremy Orloff and Jonathan Bloom



http://xkcd.com/1236/

Three 'phases'

- Data Collection:
  Informal Investigation / Observational Study / Formal Experiment
- Descriptive statistics
- Inferential statistics

> *To consult a statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of.*

R.A. Fisher

What is a statistic?

Definition. A *statistic* is anything that can be computed from the collected data.

- *Point statistic*: a single value computed from data, e.g sample average $\overline{x}_n$ or sample standard deviation s_n.

- *Interval or range statistics*: an interval $[a, b]$ computed from the data. (Just a pair of point statistics.) Often written as $\overline{x} \pm s$.

Notation

Big letters X, Y, X_i are random variables.

Little letters x, y, x_i are data (values) generated by the random variables.

Example. Experiment: 10 flips of a coin:

X_i is the random variable for the i^{th} flip: either 0 or 1.

x_i is the actual result (data) from the i^{th} flip.
e.g. $x_1, \ldots, x_{10} = 1, 1, 1, 0, 0, 0, 0, 0, 1, 0$.

Reminder of Bayes' theorem

Bayes's theorem is the key to our view of statistics.
(Much more next week!)

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}.$$

$$P(\text{hypothesis}|\text{data}) = \frac{P(\text{data}|\text{hypothesis})P(\text{hypothesis})}{P(\text{data})}$$

Estimating a parameter

Example. Suppose we want to know the percentage p of people for whom cilantro tastes like soap.

Experiment: Ask n random people to taste cilantro.

Model:
$X_i \sim \text{Bernoulli}(p)$ is whether the i^{th} person says it tastes like soap.

Data: $x_1, \ldots, x_n$ are the results of the experiment

Inference: Estimate p from the data.

Maximum likelihood estimate

The maximum likelihood estimate (MLE) is a way to estimate values of a *parameter of interest*.

Example. You ask 100 people to taste cilantro and 55 say it tastes like soap. Use this data to estimate p.

Likelihood

For a given value of p the probability of getting 55 'successes' is the binomial probability

$$P(55 \text{ soap}|p) = \binom{100}{55} p^{55}(1-p)^{45}.$$

Definition:

$$\text{The likelihood } P(\text{data}|p) = \binom{100}{55} p^{55}(1-p)^{45}.$$

MLE

The MLE is the value of p for which the observed data is most likely.

That is, the MLE is the value of p that *maximizes* the likelihood.

Calculus: To find the MLE, solve $\frac{d}{dp}P(\text{data} \mid p) = 0$ for p.

Note: Sometimes the derivative is never 0 and the MLE is at an endpoint of the allowable range. We should also check that the critical point is a maximum.

$$\frac{dP(\text{data} \mid p)}{dp} = \binom{100}{55}\left(55p^{54}(1-p)^{45} - 45p^{55}(1-p)^{44}\right) = 0$$

$\Rightarrow\ 55p^{54}(1-p)^{45} = 45p^{55}(1-p)^{44} \ \Rightarrow\ 55(1-p) = 45p \ \Rightarrow\ 55 = 100p$
$\Rightarrow$ the MLE $\hat{p} = \frac{55}{100}$.

Log likelihood

Often convenient to use log likelihood.

$$\text{log likelihood} = \ln(\text{likelihood}) = \ln(P(\text{data} \mid p)).$$

Example.

$$\text{Likelihood}\ \ P(\text{data}|p) = \binom{100}{55} p^{55}(1-p)^{45}$$

$$\text{Log likelihood} \quad = \ln\left(\binom{100}{55}\right) + 55\ln(p) + 45\ln(1-p).$$

(Note first term is just a constant.)

Board Question: Coins

A coin is taken from a box containing three coins, which give heads with probability $p = 1/3$, $1/2$, and $2/3$. The mystery coin is tossed 80 times, resulting in 49 heads and 31 tails.

(a) What is the likelihood of this data for each type on coin? Which coin gives the maximum likelihood?

(b) Now suppose that we have a single coin with unknown probability p of landing heads. Find the likelihood and log likelihood functions given the same data. What is the maximum likelihood estimate for p?

See next slide.

Solution

answer: (a) The data D is 49 heads in 80 tosses.
We have three hypotheses: the coin has probability $p = 1/3$, $p = 1/2$, $p = 2/3$. So the likelihood function $P(D|p)$ takes 3 values:

$$P(D|p = 1/3) = \binom{80}{49}\left(\frac{1}{3}\right)^{49}\left(\frac{2}{3}\right)^{31} = 6.24 \cdot 10^{-7}$$

$$P(D|p = 1/2) = \binom{80}{49}\left(\frac{1}{2}\right)^{49}\left(\frac{1}{2}\right)^{31} = 0.024$$

$$P(D|p = 2/3) = \binom{80}{49}\left(\frac{2}{3}\right)^{49}\left(\frac{1}{3}\right)^{31} = 0.082$$

The maximum likelihood is when $p = 2/3$ so this our maximum likelihood estimate is that $p = 2/3$.

Answer to part (b) is on the next slide

Solution to part (b)

(b) Our hypotheses now allow p to be any value between 0 and 1. So our likelihood function is

$$P(D|p) = \binom{80}{49} p^{49}(1-p)^{31}$$

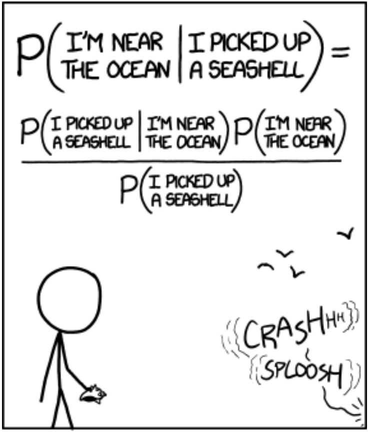
To compute the maximum likelihood over all p, we set the derivative of the log likelihood to 0 and solve for p:

$$\begin{aligned}
\frac{d}{dp}\ln(P(D|p)) = \frac{d}{dp}\left(\ln\left(\binom{80}{49}\right) + 49\ln(p) + 31\ln(1-p)\right) &= 0 \\
\Rightarrow\ \frac{49}{p} - \frac{31}{1-p} &= 0 \\
\Rightarrow\ p &= \frac{49}{80}
\end{aligned}$$

So our MLE is $\hat{p} = 49/80$.

Bayesian Updating: Discrete Priors: 18.05 Spring 2014

Jeremy Orloff and Jonathan Bloom



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

http://xkcd.com/1236/

Board Question: Dice

Five dice: 4-sided, 6-sided, 8-sided, 12-sided, 20-sided.

Jon picks one at random and, without showing it to you, rolls it and reports a 13.

1. Make a table and compute the posterior probabilities that the chosen die is each of the five dice.

2. Same question if he rolls a 5.

3. Same question if he rolls a 9.

(Keep the tables for 5 and 9 handy! Do not erase!)

Tabular solution

$\mathcal{D}$ = 'rolled a 13'

hypothesis	prior	likelihood	unnormalized posterior	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|\mathcal{D})$
$\mathcal{H}_4$	$1/5$	0	0	0
$\mathcal{H}_6$	$1/5$	0	0	0
$\mathcal{H}_8$	$1/5$	0	0	0
$\mathcal{H}_{12}$	$1/5$	0	0	0
$\mathcal{H}_{20}$	$1/5$	$1/20$	$1/100$	1
total	1		$1/100$	1

Tabular solution

$\mathcal{D}$ = 'rolled a 5'

hypothesis	prior	likelihood	unnormalized posterior	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|\mathcal{D})$
$\mathcal{H}_4$	$1/5$	0	0	0
$\mathcal{H}_6$	$1/5$	$1/6$	$1/30$	0.392
$\mathcal{H}_8$	$1/5$	$1/8$	$1/40$	0.294
$\mathcal{H}_{12}$	$1/5$	$1/12$	$1/60$	0.196
$\mathcal{H}_{20}$	$1/5$	$1/20$	$1/100$	0.118
total	1		.085	1

Tabular solution

$\mathcal{D}$ = 'rolled a 9'

hypothesis	prior	likelihood	unnormalized posterior	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|\mathcal{D})$
$\mathcal{H}_4$	$1/5$	0	0	0
$\mathcal{H}_6$	$1/5$	0	0	0
$\mathcal{H}_8$	$1/5$	0	0	0
$\mathcal{H}_{12}$	$1/5$	$1/12$	$1/60$	0.625
$\mathcal{H}_{20}$	$1/5$	$1/20$	$1/100$	0.375
total	1		.0267	1