STAT 24400
Lecture 6
1/21/16

# Inference

Much of physical and social science is devoted to deducing effects from causes using a model. In constructing such models, we often need to deduce causes (and hence the model) from effects (the data). This is the task of statistical inference.

---

Example: Polygraph Testing

so-called "Lie Detector"

Measure: Respiration Rate
Heart Rate
Blood Pressure
Skin Conductivity (= "sweat")

Combine into a score

Compare score to a threshold

If (score > threshold)
score as liar

Used for screening
" " investigating specific cases.

How effective is it?

Can depend on population being sampled.

The results depend on

$P(\text{"Guilty"})$ for population.

For a specific case, maybe

$P(\text{Deception}) = \frac{1}{2}$ (?)

For screening, maybe

$P(\text{Deception}) = \frac{1}{1000}$ (?)

smaller (?)

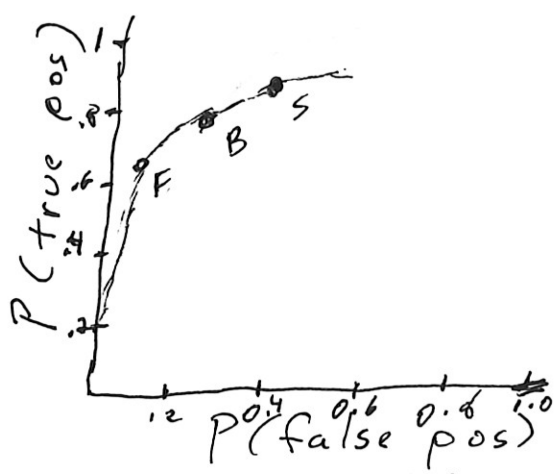
1 in 1000 are spies.

If detection threshold is set to detect the great majority (80%) of spies:

| Test Result | Examinee's True Condition Spy | Non-Spy | Total |
|---|---|---|---|
| Fail test | 8 | 1,598 | 1,606 |
| Pass test | 2 | 8,392 | 8,394 |
| Total | 10 | 9,990 | 10,000 |

If detection threshold is set to greatly reduce false positives:

| Test Result | Spy | Non-spy | Total |
|---|---|---|---|
| Fail test | 2 | 39 | 41 |
| Pass test | 8 | 9,951 | 9,959 |
| Total | 10 | 9,990 | 10,000 |




Illustrative plot of probabilities of responses for various thresholds
F = "Friendly" B = "Balanced" S = "Suspicious"

Inference - reasoning from effect to cause

Deduction - reasoning from cause to effect

Bayesian Inference

Based on far reaching consequence of a deceptively simple relationship.

Recall the definition of conditional probability:

$$P(F|E) = \frac{P(F \cap E)}{P(E)}$$

of course,

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Hence

$$P(F|E)\,P(E) = P(E|F)\,P(F)$$

so

$$P(E/F) = \frac{P(F|E)\,P(E)}{P(F)}$$

# Bayes's Theorem

Hypotheses: There are $n$ possible causes $E_1, E_2, \ldots, E_n$

These are mutually exclusive, so $(E_i \cap E_j) = \phi \quad i \neq j$

They are exhaustive, so $\Omega = \bigcup_{i=1}^{n} E_i$

Given the a priori probabilities $P(E_1)$, $P(E_2)$, ... $P(E_n)$, and the observed effect $F$, and the conditioned probabilities $P(F/E_1)$, $P(F/E_2)$, ..., $P(F/E_n)$

(the prob. of effect given causes)

Find: $P(E_i|F)$ for $i = 1, 2, \ldots n$

a posteriori probability of cause given effect

Ex. 1: Diagnostic Testing for HIV

Patient tested for HIV.

Possible "causes"

$E_1$ = patient has HIV

$E_2$ = patient does not have HIV

If patient selected at random

$P(E_1)$ = prevalence of HIV

$P(E_2) = 1 - P(E_1)$

"effect": $F$ is a positive reaction

suppose that (hypothetical values, correct order of mag)

$$\left.\begin{aligned} P(F|E_1) &= 0.9 \\ P(F|E_2) &= 0.2 \end{aligned}\right] \text{— from empirical trials}$$

Want:

$$P(E_1|F)$$

Back to:

## Bayes's Theorem

We have, from the definitions,

$$P(E_1/F) = \frac{P(F/E_1)\,P(E_1)}{P(F)}$$

Need $P(F)$.

$$F = (F \cap E_1) + (F \cap E_2) + \cdots + (F \cap E_n)$$

$$\begin{aligned} P(F) &= P(F \cap E_1) + P(F \cap E_2) + \cdots + P(F \cap E_n) \\ &= P(F/E_1)P(E_1) + \cdots + P(F/E_n)P(E_n) \\ &= \sum_{i=1}^{n} P(F/E_i)P(E_i) \end{aligned}$$

**Theorem:**

$$P(E_1/F) = \frac{P(F/E_1)\,P(E_1)}{\sum_{i=1}^{n} P(F/E_i)P(E_i)}$$

Ex. 1: Diagnostic Test for HIV

$P(F/E_1) = .9$ true positives

$P(F/E_2) = .2$ false positives

Suppose test is positive.

$$P(E_1|F) = \frac{P(F/E_1)\,P(E_1)}{P(F/E_1)\,P(E_1) + P(F/E_2)\,P(E_2)}$$

$$= \frac{(0.9)\theta}{(0.9)\theta + (0.2)(1-\theta)} \qquad \theta = P(E_1)$$

For randomly selected patient

$\theta$ = prevalence, say $\theta = 0.001$

Then $$P(E_1|F) = \frac{(0.9)(0.001)}{(0.9)(0.001) + (0.2)(0.999)} = \boxed{.0045}$$

If patient seriously at risk, say $\theta = 0.5$

$$P(E_1/F) = \frac{(0.9)(0.5)}{(0.9)(0.5) + (0.1)(0.5)} = \boxed{.8182}$$

Ex. 3: (Ancient history, apocryphal, but great example)

Alan Dershowitz on "Larry King Live": "Only about $\frac{1}{10}$ of 1% of wife-batterers actually murder their wives."

Q: Is $P(OJ \text{ guilty}) = \frac{1}{10}$ of 1% (0.001)?

We condition all probabilities on husband being a batterer.

$G$ = husband guilty of murder

$M$ = Wife murdered

Dershowitz: $P(G) = 0.001$

Want $P(G/M)$

$P(M/G) = 0.9999 \quad (\sim 1)$

$P(M/G^c) = 0.0001$ (US statistics approx)

$$P(G/M) = \frac{P(M/G)\,P(G)}{P(M/G)\,P(G) + P(M/G^c)\,P(G^c)}$$

$$= \frac{(1.0)(0.001)}{1\,(0.001) + (0.0001)(0.999)} = \boxed{.91}$$

# The Beta Distribution(s)

First, need <u>Gamma</u> <u>Function</u>

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$$

generalization of "factorial" ~~to~~ non-integers. For integers

$$\Gamma(n) = (n-1)! = (n-1)(n-2)\cdots 2\cdot 1$$

For non-integers, $\Gamma(a)$ still has the factorial property:

$$\Gamma(a) = (a-1)\Gamma(a-1).$$

Almost always we'll be considering cases where $a$ is a positive integer, so

$$\Gamma(a) = (a-1)!$$

We'll need $\Gamma(a)$ to normalize the $\beta$ distribution

The

<u>Beta</u> <u>Distribution</u> is given by

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$\alpha$ and $\beta$ are positive real parameters. They can be integers, but don't have to be. You could also write $f_X(x)$ or $f_X(x; \alpha, \beta)$.

<u>If</u> $\alpha$ and $\beta$ are integers,

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} = \frac{(\alpha+\beta-2)!(\alpha+\beta-1)}{(\alpha-1)!\,(\beta-1)!}$$

$$= \binom{\alpha+\beta-2}{\alpha-1}(\alpha+\beta-1)$$

Why "Beta" Distribution?

$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$ is called

the "Beta function",

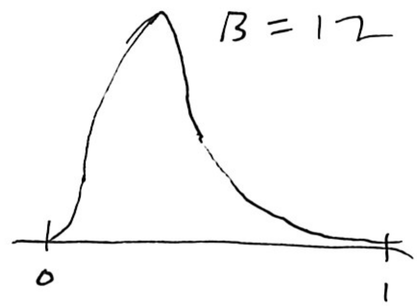
$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

(how Beta distribution is normalized,

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

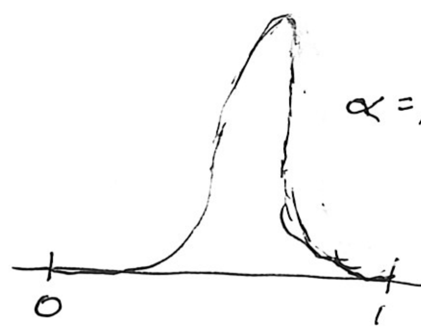
$\alpha=3,\ \beta=2$

$\alpha=5$
$\beta=12$

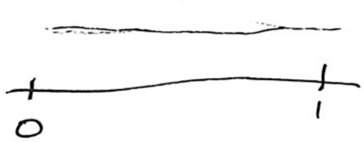


$\alpha=\beta=3$

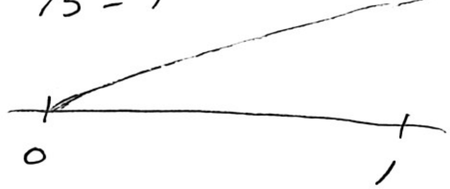
$\alpha=\beta=10$



$\alpha=\beta=1$



$\alpha=2$
$\beta=1$



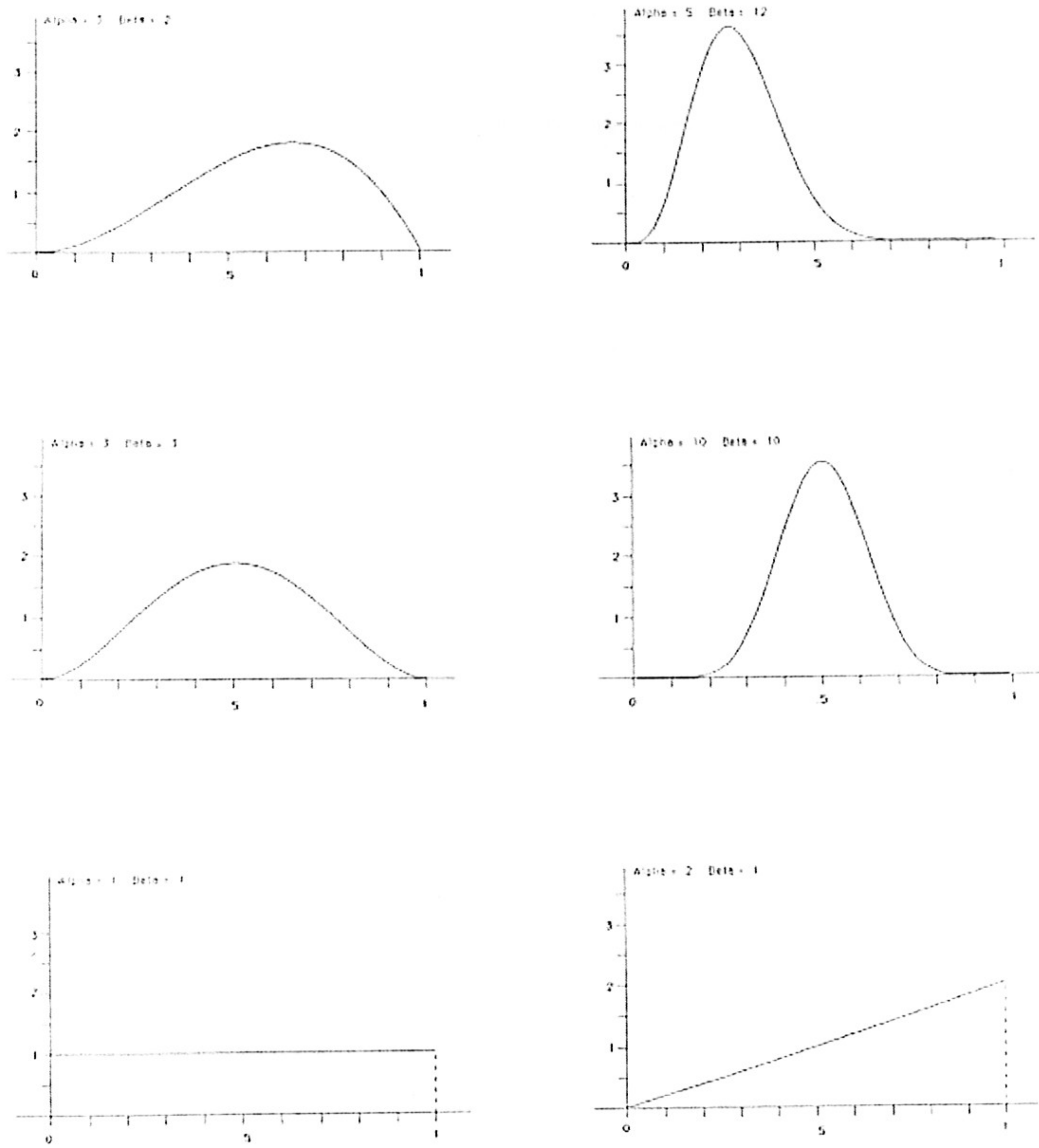



Figure 2.3. Examples of Beta densities.

12A

For the $\beta$ distribution,

$$E(x) = \int_0^1 x \, \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \, x^{\alpha-1} (1-x)^{\beta-1} \, dx$$

$$= \frac{\Gamma(\alpha+\beta+1)\,\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)\,\Gamma(\alpha+1)} \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \, x^{\alpha} (1-x)^{\beta-1} \, dx$$

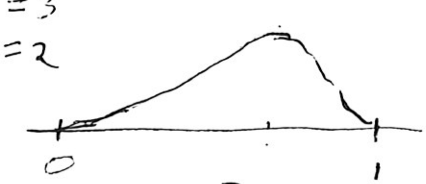
$$= \underbrace{\frac{\Gamma(\alpha+\beta)\,\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)\,\Gamma(\alpha)}}_{=\frac{\alpha}{\alpha+\beta}} \; \underbrace{\int_0^1 \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)} \, x^{\alpha} (1-x)^{\beta-1} \, dx}_{=1}$$

Beta $(\alpha+1, \beta)$

$$\Rightarrow \boxed{E(x) = \frac{\alpha}{\alpha+\beta}}$$

Two of the previous examples:

$\alpha = 3$
$\beta = 2$




$$E(x) = \frac{3}{3+2} = \frac{3}{5} = 0.6$$

$$f(x) = \frac{\Gamma(3+2)}{\Gamma(3)\Gamma(2)} \, x^{3-1} (1-x)^{2-1}$$

$$= \frac{4!}{2!\,1!} \, x^2 (1-x)^1$$

$$= 12\, x^2 (1-x)$$

max at $\frac{2}{3}$

(aka the "mode")

$\alpha = \beta = 3$


$$E(x) = \frac{3}{3+3} = \frac{1}{2}$$

(if $\alpha = \beta$, $E(x) = 0.5$)

max at $\frac{1}{2}$

We'll need the Beta distribution in a little while, but first we need to take another look at densities and at Bayes's Theorem when densities are involved.

## Remarks about Densities

Normal: $\left(\frac{1}{\sqrt{2\pi}}\right) e^{-\frac{x^2}{2}}$

Exponential: $(\theta) e^{-\theta x}$

Beta: $\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right) x^{\alpha-1}(1-x)^{\beta-1}$

---

In each case,

$$f(x) = \underbrace{(\text{constant})}_{\text{needed so } \int f(x)dx = 1} \cdot \underbrace{(\text{part involving } x)}_{\text{shape, dependence on } x}$$

To get actual probabilities, we need both parts. To <u>identify</u> a distribution, need only $2^{nd}$ part.

<u>Ex</u> If we know $f(x) = \text{const} \cdot x^{17}(1-x)^{30}$ we <u>know</u> its Beta(18, 31)

<u>Ex</u> If we know $f(x) = \text{const} \cdot e^{-x^2/2}$ we <u>know</u> its $N(0,1)$

## Recall Bayes's Theorem

$E_i$, $i = 1, 2, 3, \ldots, n$ "causes"
"states of nature"

$F$ "effect" or "data"

$$P(E_i|F) = \frac{P(E_i)\, P(F|E_i)}{P(F)}$$

$$P(F) = \sum_{j=1}^{n} P(E_j)\, P(F/E_j)$$

Equivalent form :

$$\underset{\text{Posterior}}{P(E_i|F)} \propto \underset{\text{Prior}}{P(E_i)}\, \underset{\text{"Liklihood"}}{P(F|E_i)}$$

i.e. $P(E_i|F) = (\text{const.})\, P(E_i)\, P(F|E_i)$

where "const" insures that

$$\sum_{i=1}^{n} P(E_i|F) = 1$$

in fact, $\text{const} = \dfrac{1}{P(F)} = \dfrac{1}{\sum_{j=1}^{n} P(E_j)\, P(F/E_j)}$

# Bayes's Theorem in terms of densities:

Given $f(x|y)$ and $f_Y(y)$,
Find $f(y|x)$

$Y$ - "cause"
$X$ - "effect" or "data

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

also $f(x,y) = f(x|y)\, f_Y(y)$

so $$f(y|x) = \frac{f(x|y)\, f_Y(y)}{f_X(x)}$$

because $\int_{-\infty}^{\infty} f(y|x) = 1$

$$\int_{-\infty}^{\infty} \left( \frac{f(x|y)\, f_Y(y)}{f_X(x)} \right) dy = 1, \text{ or } \int_{-\infty}^{\infty} f(x|y)\, f_Y(y)\, dy = f_X(x)$$

so in the continuous case

$$f(y|x) = \frac{f(x|y)\, f_Y(y)}{\int_{-\infty}^{\infty} f(x|y)\, f_Y(y)\, dy}$$

in summary: Given $x$, so $X$ is fixed, $f(y|x) \propto f(x|y) \cdot f_Y(y)$

**Example:** Polling – a Survey

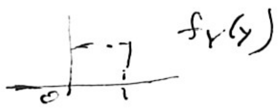
Let $y$ = fraction of all Illinois Democrat voters for Sanders

Sample $n$ Illinois Democrats "at random"

Let $X$ = # in sample who say they are for Sanders

---

**Hypotheses:** $f_Y(y)$ = a priori density of $y$




$\equiv$ uniform on $(0, 1)$, say.

$$p(x|y) = \binom{n}{x} y^x (1-y)^{n-x} \quad \underline{\underline{\text{Binomial}}}$$

**Observe:** $n = 100 \quad x = 40$

what can we infer about $y$?

$$f(y|x=40) \propto \underbrace{\binom{100}{40} y^{40} (1-y)^{60}}_{p(x/y)} \cdot \underbrace{1}_{f_Y(y)} \left(\begin{smallmatrix}\text{for} \\ 0<y<1\end{smallmatrix}\right)$$

A Beta Density

This is a "mixed" case –

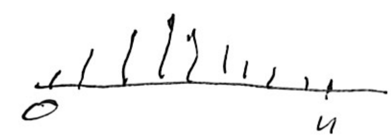
$x$ is discrete, $Y$ continuous 


$Y$ uniform on 0 to 1: $f_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$

<u>Given</u> $Y = y$,

$X$ Binomial $n$ trials, $\theta = y$

$$p(x/y) = \begin{cases} \binom{n}{x} y^x (1-y)^{n-x} & x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

(a prob. func.) 


---

Marginal Distribution of $X$?

$$p(x, y) = f_Y(y)\, p(x/y)$$

$$= \begin{cases} \binom{n}{x} y^x (1-y)^{n-x} & 0 < y < 1 \\ & x = 0, 1, \ldots n \\ 0 & \text{otherwise} \end{cases}$$

(prob. func. in $X$, density in $y$)

$$p_X(x) = \int_0^1 \binom{n}{x} y^x (1-y)^{n-x}\, dy$$

$$= \binom{n}{x} \int_0^1 y^x (1-y)^{n-x}\, dy$$

marg dist of X (continued)

$$p_X(x) = \binom{n}{x} \int_0^1 y^x (1-y)^{n-x}\, dy$$

The is the Beta function, which we saw earlier in normalizing the Beta dist.

Explanatory Digression

$$\left[\; B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx \right.$$

$$= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \frac{1}{\binom{\alpha+\beta-2}{\alpha-1}(\alpha+\beta-1)}$$

which you remember — right? ]

$$p_X(x) = \binom{n}{x} \cdot \frac{1}{\binom{n}{x}(n+1)}$$

$$= \frac{1}{n+1}$$

<u>Uniform</u>

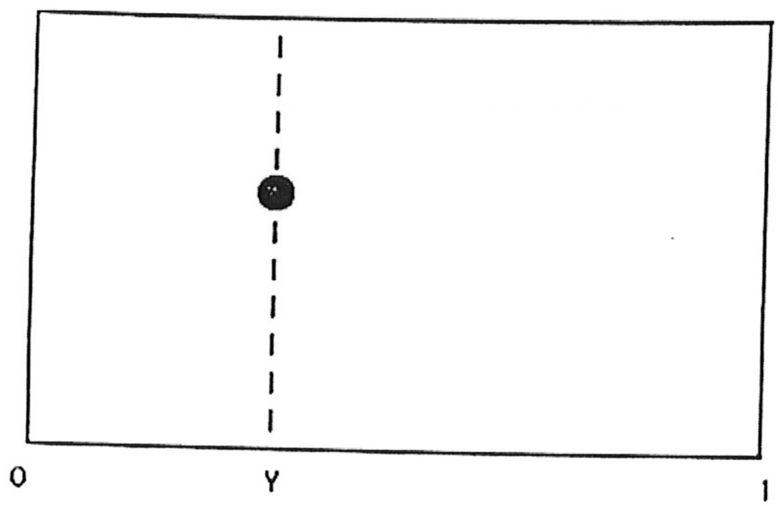



Figure 3.7. Bayes's billiard table.

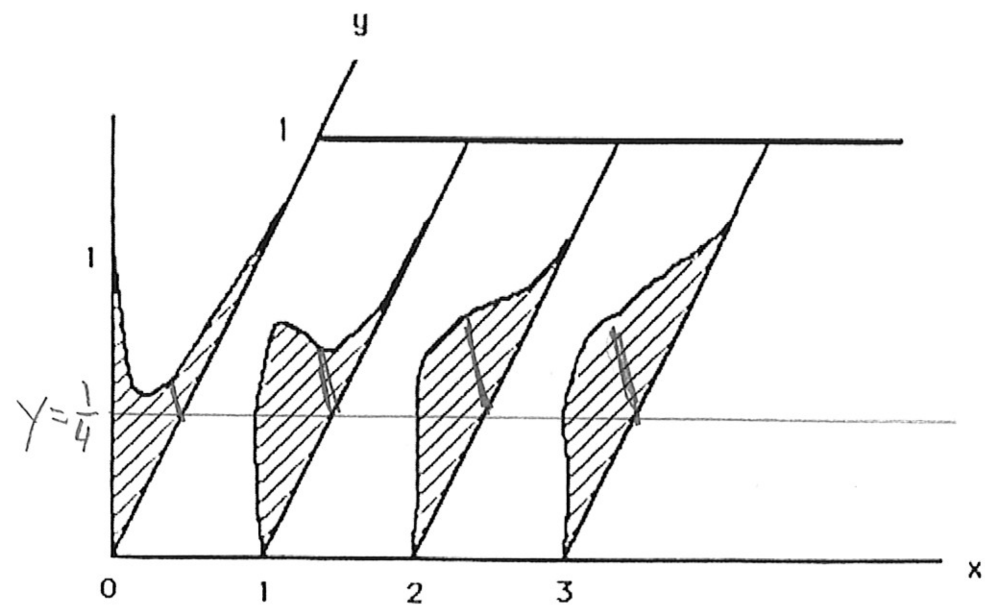
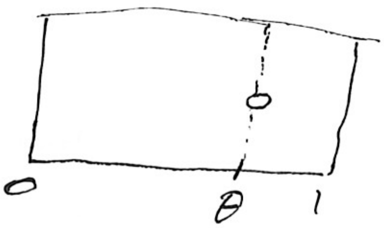



Figure 3.8

$R_x(x)$

# How to think about prior?

## A. Billiard Table




Roll a ball. It is equally likely to stop at any point. Ball stops distance of $\Theta$ from the end. Remove ball. Roll 100 more, one at a time. Count $X =$ the number that land to the left of $\theta$.

$$f_\theta(\theta) = \begin{cases} 1 & 0 < \theta < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$p(x|\theta) = \binom{100}{x} \theta^x (1-\theta)^{100-x} \quad x = 0, 1, \dots, n$$

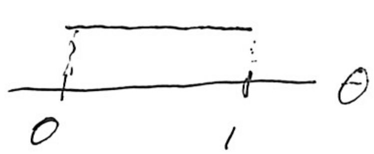
What if we are given $X = 40$?

What about $\theta$?

$$f(\theta | x) \propto \binom{100}{40} \theta^{40} (1-\theta)^{60}$$

(Beta, as previously)

Before




After

# Comparison of the two examples

They are the same mathematically!

Conceptually? Is the status of $f_Y(y)$ different from $f_\Theta(\theta)$?

$f_\Theta(\theta)$ represents a tangible model

$f_Y(y)$ represents... what?
uncertainty
subjective opinion
prior knowledge

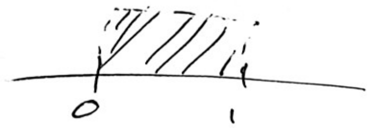
**But**

(1) others may disagree with your assessment of $f_Y$

(2) Uniform $f_Y$ may not do the job.

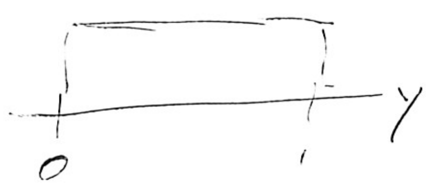
So: <u>Hypotheses</u> were:

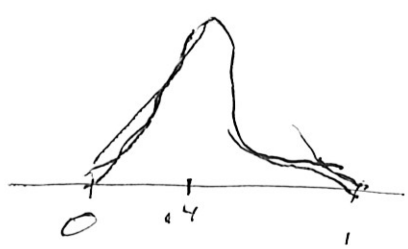
(1) $y$ uniform on $(0,1)$ a priori




(2) Given $y$, $X$ is binomial (generated by random sampling)

Conclusion: a posteori — after observing $x = 40$ $y$ can be viewed as having a particular Beta distribution




BEFORE + Data = AFTER

BAYESIAN INFERENCE