(2) Cauchy Dist'n: take $\nu = 1$ in Student's $t$.

$$f(x|m,\sigma) = \frac{1}{\sigma\pi}\,\frac{1}{1+\left(\frac{x-m}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

We use median and scale parameter because mean/variance DO NOT EXIST!

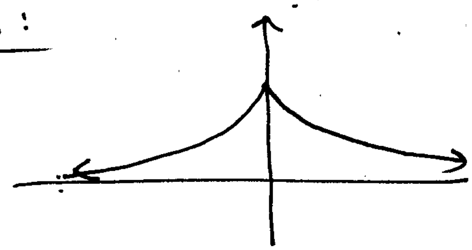
VERY heavy tailed.

(3) Double Exponential (AKA Laplace)

$$f(x|\mu,\sigma^2) = \frac{1}{2\sigma}\exp\left\{-\frac{|x-\mu|}{\sigma}\right\},$$

$$-\infty < x < \infty$$
$$-\infty < \mu < \infty$$
$$\sigma > 0$$

GRAPH:



Here: $\hat{\mu}$ = MLE for $\mu$

= SAMPLE MEDIAN.

Non-Normal Dist'ns ↑

FINITE MIXTURE DISTRIBUTIONS

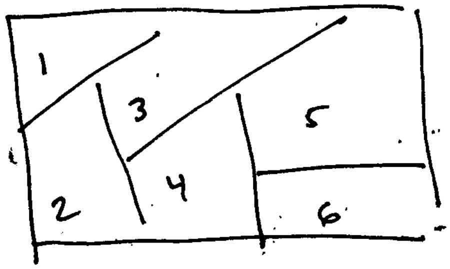
Again, $X_1, \ldots, X_n \sim f(x|\theta)$

But $f$ takes the form

$$f(x|\theta) = \sum_{j=1}^{k} p_j f(x|\theta_j),$$

where $p_j \geq 0$, $\sum_j p_j = 1$

Have $k$ different subpopulations 5/25/11 (2)




The proportion of people in group $j$ is $p_j$

The $j^{th}$ subgroup has dist'n $f(x|\theta_j)$

---

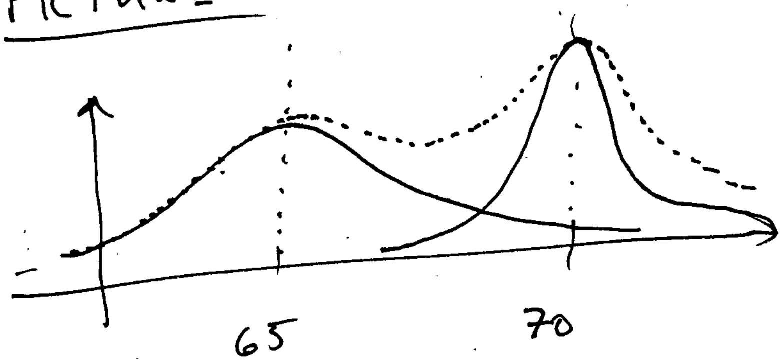
Example: Heights of Students.

Let $X$ = height (in inches).

Male heights $\approx N(70, 4^2)$

Female heights $\approx N(65, 5^2)$

PICTURE:




Then the population density is:

(suppose there are 63% Male
37% Females)

$$f(x|\mu_1, \mu_2, \sigma_1, \sigma_2, p) = p N(\mu_1, \sigma_1^2) + (1-p) N(\mu_2, \sigma_2^2)$$

$$= 0.37 N(65, 5^2) + 0.67 N(75, 4^2)$$

In general, the likelihood looks like

$$L(\underline{\theta}) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{k} p_j f(x_i \mid \theta_j) \right\}$$

$$= \prod_{i=1}^{n} \left( p_1 f(x_i \mid \theta_1) + \cdots + p_k f(x_i \mid \theta_k) \right)$$

this product has $k^n$ terms.

this EXPLODES an $n \to \infty$.

Problem is limited by TIME and RESOURCES.

---

YET ANOTHER MODEL!

Alternative to n Bernoulli trials.

Coin tossing Model:

$X_1, \ldots, X_n$ IID Bernoulli($p$).

where a) trials are independent.

b) $\mathbb{P}(\text{success}) = p$, constant.

BUT...

what if p's change over time?

Maybe there's dependence in sequence.

Suppose we have belief in STREAKY model.

2 states: $p_H \to$ hot state

$p_C \to$ cold state

If you're hot, chances higher that you'll stay hot for next round...

5/25/11 ($)

Make table:

| | | State in trial $i+1$ | |
|---|---|---|---|
| | | Hot | Cold |
| State in trial $i$ | Hot | $0.9a$ | $0.1(1-a)$ |
| | Cold | $0.1$ | $0.9$ |

WE DON'T KNOW:

$p_H$, $p_C$, $a$, states in $n$ trials.

One possible configuration of states

Observe $x = (1,1,1,0,0,1,0,1,1)$

States: $S = (H,H,C,C,H,H,C,H,H)$

(unobserves):

Prob of $x$ for config $s$:

$$p_H \cdot p_H \cdot p_C (1-p_C)(1-p_H) p_H (1-p_C) p_H p_H$$

Then Likelihood looks like

$$L(p_H, p_C, a) = \sum_{\substack{\text{all possible} \\ \text{configs } s}} \mathbb{P}(\text{observe } \underline{x} \mid \text{state is } \underline{s})$$

AKA: "Markov Switching Model"

Number of terms above is: $2^n$.

# Intro to Bayesian Statistics

- Named for ~~Dr~~ Rev. Thomas Bayes (1702–1761)

- Based on theory of subjective probability.

FREQUENTISTS

Bayesians.

→ Central Theme

- $\theta$ is quantity, unknown, think of it as RANDOM.

- Have beliefs/existing knowledge about $\theta$. Represented by PRIOR distribution.

  $\pi(\theta)$, for $\theta \in \Theta$

- $\pi$ is a PDF, nonnegative, integral 1.

WANT TO LEARN about $\theta$.

Do an experiment, observe random variable $X$ which depends on $\theta$.

The conditional distribution of $X$ is $X \sim f(x \mid \theta)$

called "likelihood or sampling density".

Want to UPDATE our beliefs about $\theta$. NEW beliefs represented by POSTERIOR distribution

$\pi(\theta|x)$ = POSTERIOR dist'n of $\theta$.

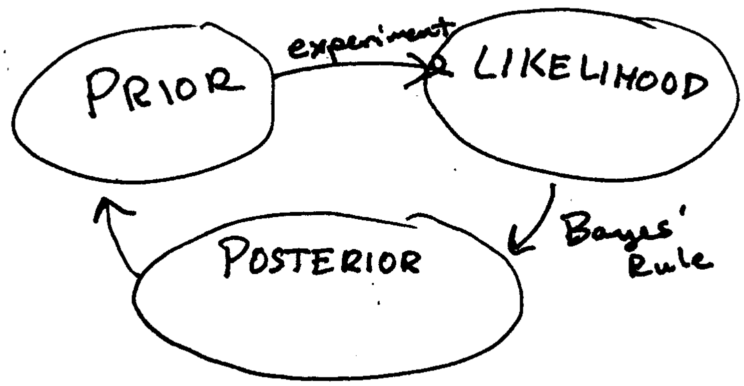
METHOD to find $\pi(\theta|x)$: BAYES' RULE.

Bayes' Rule:

$$\boxed{\pi(\theta|x) = \frac{f(x|\theta)\,\pi(\theta)}{\int f(x|\theta)\,\pi(\theta)\,d\theta}}$$

Remarks:

1) Once beliefs updated, do another experiment, update again.




2) From Bayes' Rule:

$$\pi(\theta|x) = \frac{f(x|\theta)\cdot\pi(\theta)}{\int f(x|\theta)\,\pi(\theta)\,d\theta}$$

this part doesn't depend on $\theta$.

$$= M \cdot f(x|\theta)\,\pi(\theta)$$

So

$$\pi(\theta|x) \propto f(x|\theta) \cdot \pi(\theta)$$

POSTERIOR $\propto$ LIKELIHOOD $\times$ PRIOR

3) Notice difference of interpretation

$f(x|\theta)$ as function of $\theta$ $\longmapsto$ Likelihood.

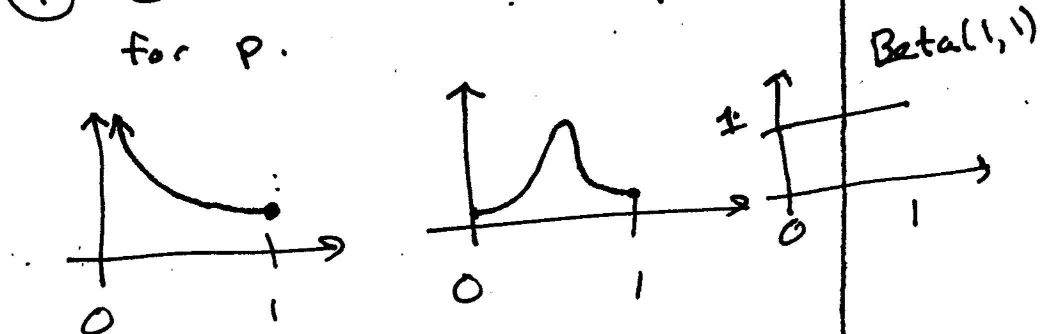
$f(x|\theta) \longrightarrow x \rightarrow$ (sample density)

---

Ex) Want to learn about

$p$ = proportion of goldfish in Lake.

① Construct continuous prior for $p$.




Let $p$ have a Beta$(\alpha, \beta)$ density.

$$p \sim \pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \quad 0<p<1.$$

Some Properties:

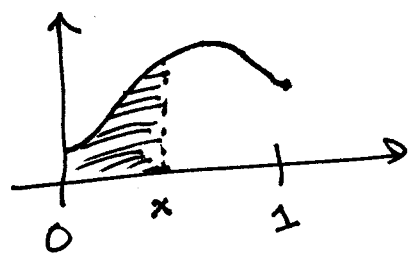
A) $\mathbb{E}p = \dfrac{\alpha}{\alpha+\beta} = \eta$

B) $\mathrm{Var}(p) = \dfrac{\eta(1-\eta)}{\alpha+\beta+1} = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Think of $\eta$ as "prior guess" about $p$.

Think of $K = \alpha + \beta$ as "precision of belief".

c) $$\mathbb{P}(p \le x) = \underbrace{\int_0^x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\,dp}_{\text{"Incomplete Beta Function"}}$$




(2) Want to learn about $p$:

GO FISHING!

We catch $n$ fish, and let

$Y$ = number of goldfish caught.

$= X_1 + X_2 + \cdots + X_n$, where

$$X_i = \begin{cases} 1, & \text{if } i^{th} \text{ goldfish} \\ 0, & \text{otherwise.} \end{cases}$$

So $X_i$ are IID Bernoulli($p$)

Then $Y \sim$ Binomial $(n, p)$

with
$$\underbrace{f(y|p) = \binom{n}{y} p^y (1-p)^{n-y},}_{\text{this is Likelihood.}} \quad y = 0, 1, 2, \ldots, n.$$

(3) Update beliefs with Bayes' RULE!

POST. $\propto$ LIKELIHOOD $\times$ PRIOR.

$$\pi(p|y) \propto f(y|p) \cdot \pi(p)$$

$$= \binom{n}{y} p^{y}(1-p)^{n-y} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

$$\pi(p|y) = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot p^{(\alpha+y)-1}(1-p)^{(n+\beta-y)-1}$$

$$= M \cdot p^{\alpha+y-1}(1-p)^{n-y+\beta-1},$$

where $M$ is chosen so that

$$\int \pi(p|y)\,dp = 1.$$

That is $p \sim \text{Beta}(\alpha+y, n-y+\beta)$

By inspection, we can conclude

$$M = \frac{\Gamma((\alpha+y)+(n-y+\beta))}{\Gamma(\alpha+y)\cdot\Gamma(n-y+\beta)}$$

$$\boxed{= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+y)\Gamma(n-y+\beta)}}$$

# Conclusion:



PRIOR: $p \sim \text{Beta}(\alpha, \beta)$

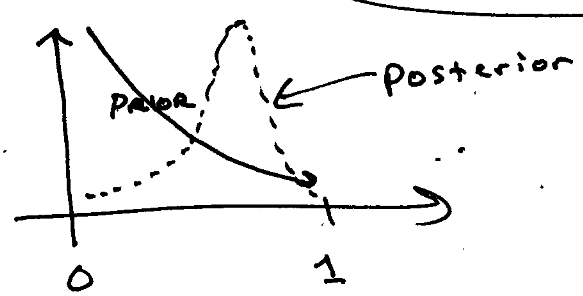
did experiment, observe $y$.

POSTERIOR: $p|y \sim \text{Beta}(\alpha+y, \beta+n-y)$.

---

**Bayesian Statistics:** draw all inference from posterior distribution.

$$\text{Beta}(\alpha+y, \beta+n-y)$$




NEW GUESS at $p$:

$$\eta^* = \frac{\alpha^*}{\alpha^*+\beta^*} = \frac{\alpha+y}{(\alpha+y)+(\beta+n-y)}$$

$$= \frac{\alpha+y}{\alpha+\beta+n}$$

$$= \frac{\alpha}{\alpha+\beta+n} \cdot \frac{(\alpha+\beta)}{(\alpha+\beta)} + \frac{y}{\alpha+\beta+n} \cdot \frac{n}{n}$$

$$= \frac{\alpha}{\alpha+\beta} \cdot \frac{\alpha+\beta}{\alpha+\beta+n} + \frac{y}{n} \cdot \frac{n}{\alpha+\beta+n}$$

$$= \eta \cdot \frac{K}{K+n} + \hat{p} \cdot \frac{n}{K+n}$$

PRIOR MEAN ($\eta$) MLE ($\hat{p}$)