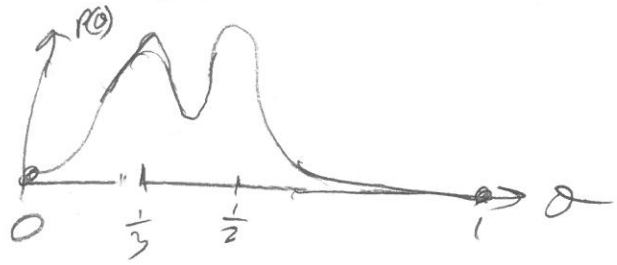
Lecture 19 Math 390.03-02 5/2/16

Recall... $X|\theta \sim Bin(n,\theta)$, $\theta \sim Beta(\alpha,\beta)$

$\Rightarrow \theta|x \sim Beta(x+\alpha, n-x+\beta)$, $X \sim BetaBin(n,\alpha,\beta)$

What if $\theta \not\sim Beta(\alpha,\beta)$ e.g.



Can you still use conjugacy?

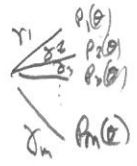
No, but Yes... what if $P(\theta) \approx \sum_{m=1}^{M} \gamma_m P_m(\theta)$ s.t. $P_m(\theta) = Beta(\alpha_m, \beta_m)$ ie $\sum_{m=1}^{M} \gamma_m = 1$

$P(\theta) \approx \frac{1}{2} Beta(3,3) + \frac{1}{2} Beta(2,4)$

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

prior mixing prop. a mixture of

Beta... is this conjugate?



No.. But the math is nice...

Conjugate mixture priors

$$P(\theta|x) = \frac{P(x|\theta)\sum_{m=1}^{M}\gamma_m P_m(\theta)}{P(x)} \longrightarrow \int_\Theta P(x|\theta)\sum_{m=1}^{M}\gamma_m P_m(\theta)\,d\theta = \sum_{m=1}^{M}\gamma_m \underbrace{\int_\Theta P(x|\theta)P_m(\theta)\,d\theta}_{P_m(x)}$$

$$= \frac{\sum \gamma_m P(x|\theta) P_m(\theta)}{P(x)}$$

$$= \frac{\sum \gamma_m \frac{P(x|\theta)P_m(\theta)}{P_m(x)} P_m(x)}{P(x)} = \sum_{m=1}^{M} \frac{\gamma_m P_m(x)}{P(x)} P_m(\theta|x) \propto \sum_{m=1}^{M} \gamma_m P_m(x) P_m(\theta|x)$$

posterior mixing prop.

posterior from conjugate prior

if $\gamma_m = \frac{1}{M}\ \forall m$, $\sum_{m=1}^{M} \frac{P_m(x)}{P_1(x)+\dots+P_M(x)} P_m(\theta|x) \propto \sum P_m(x) P_m(\theta|x)$

e.g. two betas $\theta_1 \sim Beta(\alpha_1,\beta_1)$, $\theta_2 \sim Beta(\alpha_2,\beta_2)$, $\gamma_1 = \gamma_2 = \frac{1}{2}$

$$P(\theta|x) = \frac{1}{BetaBin(n,\alpha_1,\beta_1) + BetaBin(n,\alpha_2,\beta_2)} \left(BetaBin(n,\alpha_1,\beta_1)\, Beta(x+\alpha_1, n-x+\beta_1) + BetaBin(n,\alpha_2,\beta_2)\, Beta(x+\alpha_2, n-x+\beta_2) \right)$$

not a conj. dist!

Real life ex.

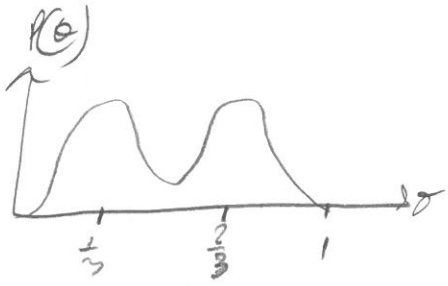
magicians "flip" coin by spinning them. $P(H) = \frac{1}{2}$

BUT... if the side is shaved, the $P(H)$ is biased ... assume bias is $\frac{2}{3}$ or $\frac{1}{3}$ which is not detectable with naked eye

Assume shaved rounds H or rounds T with equal prob...

So $P(\theta) \simeq \frac{1}{2}\,\text{Beta}(10,20) + \frac{1}{2}\,\text{Beta}(20,10)$

these strengths come from prior knowledge about the creators of the story



coin "spun" $n=10$, $x=3$ (3 heads)

$$p(\theta|x) = \frac{1}{\text{bb}(3,10,10,20) + \text{bb}(3,10,20,10)} \left(\begin{array}{l} \text{bb}(3,10,10,20)\ \text{Beta}(13,27) \\ + \text{bb}(3,10,20,10)\ \text{Beta}(23,17) \end{array} \right)$$

$\text{bb}(3,10,10,20) \approx 0.228$ $\quad$ $\text{bb}(3,10,20,10) \approx 0.0277$

How to sample??

draw and compute "Monte Carlo"

Recall

= rbetabinom (n,α,β)

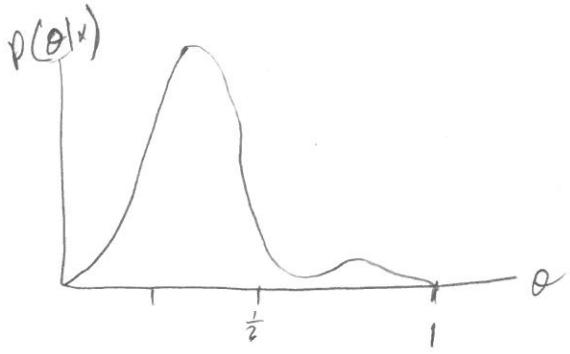
= pbetabinom (x,n,α,β)

= qbetabinom (p,n,α,β)

= dbetabinom (x,n,α,β)

↑ value at $P(X=x)$ if $X \sim \text{BetaBin}(n,\alpha,\beta)$

$$= 0.892\,\text{Beta}(13,27) + 0.108\,\text{Beta}(23,17)$$



$$\hat{\theta}_{MMSE} = 0.892\,\frac{13}{40} + 0.108\,\frac{23}{40} = 0.352$$

$\hat{\theta}_{MAP} = \text{argmax}\{p(\theta|x)\}$ not so simple...

Generally, Previously,

How to get $\hat{\theta}_{MAP}$? ① If you know closed, use Norm approx
② Use grid sampling. Make fine grid and compute on all vals. Take max.

$$\hat{\theta}_{MAP} := \arg\max\{p(\theta|x)\} = \arg\max\{k(\theta|x)\} = \arg\max\left\{\sum_{m=1}^{M} \gamma_m p_m(x)\, p_m(\theta|x)\right\}$$

for the beta-binomial case,

$$\frac{d}{d\theta}\left[\sum \gamma_m \text{beta-bin}(n, \alpha_m, \beta_m)\, \text{Beta}(x+\alpha_m, n-x+\beta_m)\right] \overset{set}{=} 0$$

$$= \frac{d}{d\theta}\left[\sum \frac{\gamma_m}{B(\alpha_m,\beta_m)}\, \theta^{x+\alpha_m-1}(1-\theta)^{n-x+\beta_m-1}\right]$$

$$\frac{d}{d\theta}\left[\sum \gamma_m \left(\binom{n}{x} \frac{B(x+\alpha_m, n-x+\beta_m)}{B(\alpha_m,\beta_m)}\right)\left(\frac{1}{B(x+\alpha_m, n-x+\beta_m)}\, \theta^{x+\alpha_m-1}(1-\theta)^{n-x+\beta_m-1}\right)\right] \overset{set}{=} 0$$

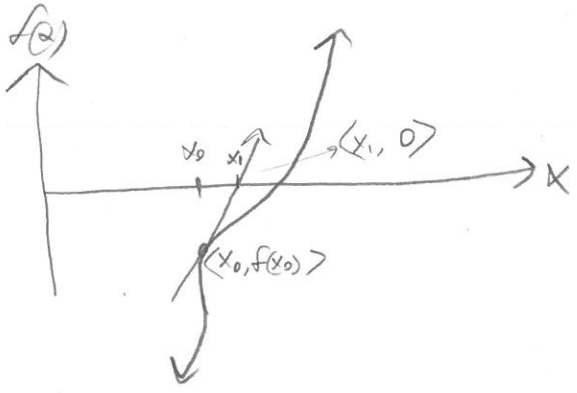
(cancels)

$$\sum \frac{\gamma_m}{B(\alpha_m,\beta_m)} \frac{d}{d\theta}[\;] = \sum \frac{\gamma_m}{B(\alpha_m,\beta_m)}\left((x+\alpha_m-1)\theta^{x+\alpha_m-2}(1-\theta)^{n-x+\beta_m-1} - (n-x+\beta_m-1)\theta^{x+\alpha_m-1}(1-\theta)^{n-x+\beta_m-2}\right) = 0$$

No closed form sol for θ!

What to do? Numerical method...

Solve $f(\theta)=0$ for x, where f is diff. real fctn



Step 1 guess an x_0 as the root

Step 2 draw tangent line at $f(x_0)$

Step 3: find its x intercept $= x_1$

Step 4: Repeat steps 1-3 with $x_0 = x_1$, stop this until

$|x_{t+1} - x_t| < \varepsilon$ (tolerance)

where ε is small

(this algorithm is an iterative method)

Sup2 can be translated as follows.. back to 7th grade..

$$y - b = m(x - a)$$

$$y - f(x_0) = f'(x_0)(x - x_0)$$

x_1 is solution for x-intercept i.e. $y = 0$

$$-f(x_0) = f'(x_0)(x_1 - x_0)$$

$$\Rightarrow -\frac{f(x_0)}{f'(x_0)} = x_1 - x_0 \Rightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

1669 - Newton for polynomial
1690 - Raphson
1740 - Simpson (today's version)

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

$$\vdots$$

$$x_{t+1} = x_t - \frac{f(x_t)}{f'(x_t)}$$

} Called the Newton-Raphson Method

Let's get $\hat{\theta}_{MAP}$. $f(x)$

$\theta_0 \stackrel{?}{=} 0.33$

$\theta_1 = 0.3$

$$K(\theta|x) = \underbrace{\frac{1}{B(2,27)}}_{C_1}\theta^{12}(1-\theta)^{26} + \underbrace{\frac{1}{B(23,17)}}_{C_2}\theta^{22}(1-\theta)^{16}$$

$$K'(\theta|x=3) = C_1\left(12\theta^{11}(1-\theta)^{26} + 26\theta^{12}(1-\theta)^{25}\right) + C_2\left(22\theta^{21}(1-\theta)^{16} + 16\theta^{22}(1-\theta)^{15}\right)$$

$$f'(x) = K''(\theta|x=3) = \checkmark$$

N-R gives $\hat{\theta}_{MAP} = 0.31577 \neq 0.352 = \hat{\theta}_{MMSE}$

"Hierarchical Model"

What did we do?

$$p(\theta|x) = \frac{p(x|\theta)\overbrace{\sum \gamma_m p_m(\theta)}}{p(x)}$$

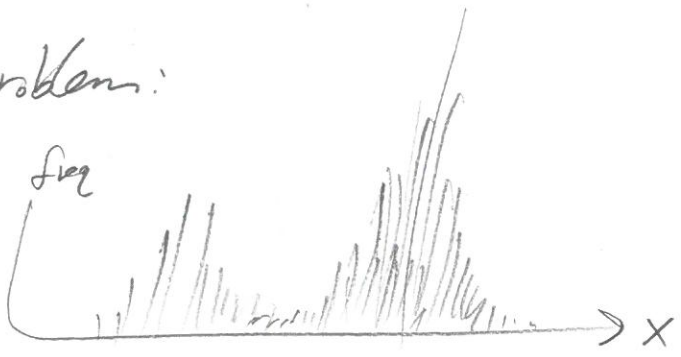
$p(\theta|\gamma_1,\ldots,\gamma_m)\,p(\gamma_1,\ldots,\gamma_m)$

prior (since it is the prior prob of θ) based on dependent other quantities

hyperprior w/ hyperhyperparams

Not observed

New problem:



this looks like... mixture model

$$X_{1,\dots,N} \overset{\text{each}}{\sim} p(x \mid \underbrace{\theta_0, \sigma_0^2, \theta_1, \sigma_1^2, \varrho}_{\text{param vector is 5-dim}}) = \varrho\, N(\theta_0, \sigma_0^2) + (1-\varrho)\, N(\theta_1, \sigma_1^2)$$

what if we just care about a best guess of params? MLE

Set $\nabla \ln \mathcal{L}(\theta_0, \sigma_0^2, \theta_1, \sigma_1^2, \varrho) = 0$

$$\nabla \ln \prod_{i=1}^{N} \varrho \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(x_i-\theta_0)^2} + (1-\varrho)\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x_i-\theta_1)^2}$$

good luck!

what to do?

Grid search $\theta_0 \in [\dots]$, $\theta_1 \in [\dots]$, $\varrho \in (0,1)$, $\sigma_0^2 \in [\dots]$, $\sigma_1^2 \in [\dots]$

In 5 dimension $1000^5 = 1e15$ ... Still okay ... but what if

$$X_{1,\dots,N} \overset{\text{each}}{\sim} \sum_{m=1}^{M} \varrho_m N(\theta_m, \sigma_m^2), \quad M \text{ large... grid search fails...}$$