5/2/17

$X_1, X_2, \dots X_n | \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$

$\theta \sim N(\mu_0, \tau^2)$

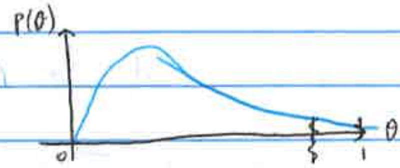
$\sigma^2 \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0 \sigma_0^2}{2}\right)$

$\Rightarrow P(\theta, \sigma^2 | X) = N(\theta_p, \sigma_p^2) K(\sigma^2 | X)$

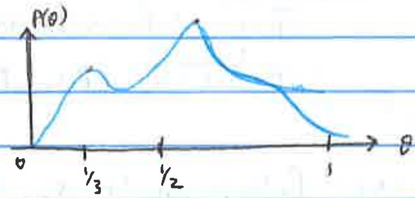
Recall: $X|\theta \sim Bin(n, \theta)$

$\theta \sim Beta(\alpha, \beta)$

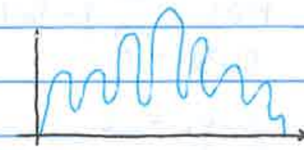
$\theta | X \sim Beta(\alpha + X, \beta + n - x)$



What if you want:



or



But cannot be a Beta!

If you knew the function $P(\theta)$ then you can compute:

$P(\theta|x) \propto P(x|\theta) P(\theta) = K(\theta | X)$ (off by a constant. How do we find the constant. Integrate $\Rightarrow$ Sum.)

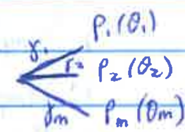
and use a grid search.

$G = \langle \underset{0}{\theta_{min}}, \theta_{min} + \Delta\theta, \theta_{min} + 2\Delta\theta, \dots, \underset{1}{\theta_{max}} \rangle$

Can we still use conjugacy?

Imagine $P(\theta)$ is a mixture / compact distribution of a distinct # of beta components.

$$P(\theta) = \sum_{m=1}^{M} \gamma_m \underbrace{P_m(\theta)}_{Beta(\alpha_m, \beta_m)} \quad \text{s.t. } \gamma_m = 1$$



Say, $X|\theta \sim Bin(n, \theta)$

$P(\theta) = \sum_{m=1}^{M} \gamma_m P_m(\theta)$

$\theta | X \sim ?$

$$P(\theta|X) = \frac{P(X|\theta) P(\theta)}{P(X)} = \frac{P(X|\theta) \sum \gamma_m P_m(\theta)}{P(X)} = \sum_{m=1}^{M} \gamma_m \frac{P(X|\theta) P_m(\theta)}{P(X)} = \sum^{M} \underbrace{\frac{\gamma_m P_m(\cancel{X})}{P(X)}}_{\gamma_m'} \cdot \underbrace{\frac{P(X|\theta) P_m(X)}{P_m(X)}}_{P_m(\theta|X)}$$

$\times \frac{P_m(X)}{P_m(X)}$

$$= \sum_{m=1}^{M} \gamma_m' \underbrace{P_m(\theta|X)}_{Beta(\alpha_m', \beta_m')}$$

$\alpha_m' = \alpha + x$, $\beta_m' = \beta_m + n - x$

Prior Predictive Distribution.

Recall: $P(x) = \int_{\Theta} P(x|\theta)\, P(\theta)\, d\theta \overset{so}{=} \int_{\Theta} P(x|\theta) \sum \gamma_m P_m(\theta)\, d\theta = \sum_{m=1}^{M} \gamma_m \underbrace{\int_{\Theta} P(x|\theta)\, P_m(\theta)\, d\theta}_{P_m(x)}$

$P_m(x) = \text{BetaBinomial}(n, \alpha_m, \beta_m)$

If $\gamma_m = \frac{1}{M} \ \forall m$, $\quad \gamma_m' = \frac{\gamma_m P_m(x)}{P(x)} = \frac{\gamma_m P_m(x)}{\sum \gamma_m P_m(x)} = \frac{P_m(x)}{\sum P_m(x)}$

if $\gamma_m = \frac{1}{M} \ \forall m.$

Again $X|\theta \sim Bin(n,\theta)$

$P(\theta) = \sum_{m=1}^{M} \gamma_m P_m(\theta)$

$\theta|X \sim ?$

$\gamma_1 = \gamma_2 = \frac{1}{2}$

$\alpha_1 = 3, \ \beta_1 = 3$

$\alpha_2 = 2, \ \beta_2 = 4$

$n = 10, \ x = 5$

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

$= \frac{1}{\sum P_m(x)} \sum P_m(x) P_m(\theta|X)$

$= \frac{1}{P_1(5)+P_2(5)} \left(P_1(5)\, P_1(\theta|x=5) + P_2(5)\, P_2(\theta|X=5) \right)$

$P_1(x) = \text{Beta binom}(n, \alpha_1, \beta_1)$

using R

$P_1(5) = \text{dbetabinom}(5,10,3,3) = \cancel{0.112}\ 0.147$ ← should be higher bc $\alpha=3, \beta=3$, centered at 5, splits off evenly.

$P_2(5) = \text{dbetabinom}(5,10,2,4) = \cancel{0.147}\ 0.112$

$= \frac{1}{\text{dbetabinom}(5,10,3,3) + \text{dbb}(5,10,2,4)} \left(\text{dbb}(5,10,3,3) \cdot \text{dbeta}(\theta, \underset{\alpha_1+x}{8}, \underset{\beta_1+n-x}{8}) + \text{dbb}(5,10,2,4) \cdot \text{dbeta}(\theta, \underset{\alpha_2+x}{7}, \underset{\beta_2+n-x}{9}) \right)$

$= 0.57\ \text{dbeta}(\ ) + 0.43\ \text{dbeta}(\ )$

→ graph:

Sample from $P(\theta|X)$

1. Sample $\theta_{0,1}$ from Beta(8,8) use rbeta(8,8)
2. Sample $\theta_{0,2}$ from Beta(7,9) use rbeta(7,9)

} Pull a sample from Beta

3. Retain $\theta_0 = \gamma_1' \theta_{0,1} + \gamma_2' \theta_{0,2}$
4. Repeat 1–3 many times.

Point Estimation

$\hat{\theta}_{MMSE} = E[\theta|X] = \int_{\Theta} \theta \sum \gamma_m' P_m(\theta|x)\, d\theta = \sum \gamma_m' \int_{\Theta} \theta\, P_m(\theta|X)\, d\theta$

$= \sum \gamma_m' E_m(\theta|X) = \sum_{m=1}^{M} \gamma_m' \frac{\alpha_m'}{\alpha_m' + \beta_m'}$

For last Ex $= 0.57 \left(\frac{8}{16}\right) + 0.43 \left(\frac{7}{16}\right)$

(Not on Exam)
Not covered but : (med) $\hat{\theta}_{MAE} = \cdots$ Sample median.
(mode) $\hat{\theta}_{MAP} = \text{argmax}\{P(\theta|X)\}$

$\hat{\theta} = \text{argmax}\{K(\theta|X)\}$

$P(\theta|X) \propto \sum \gamma_m P_m(x) P_m(\theta|x) = K(\theta|X)$

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

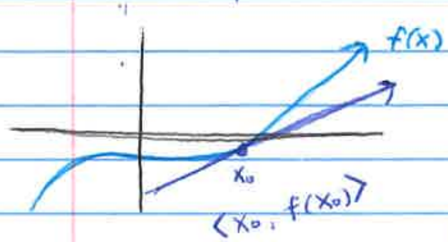
Set $0 = \frac{d}{d\theta}[\ ]$

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

Doesn't matter $= 0$

Can't be solved.

Assume $f(x)$ is continuous and has one zero on X.
Want x^* s.t. $f(x^*) = 0$



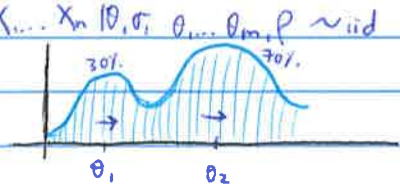
Step 1: Given $x_0 = x^*$
Step 2: Draw tangent line
Step 3: Set x_1 = x-intercept of tangent line
Step 4: Repeat until $|x_{t+1} - x_t| < \varepsilon$ by setting $x_0 = x_t$, ε is your accuracy / tolerance level.

Step 2: $y - b = m(x-a)$
: Eq. of tangent line
$y - f(x_0) = f'(x_0)(x - x_0)$

Step 3: Solve for x-intercept (x_1)
$-f(x_0) = f'(x_0)(x_1 - x_0)$
$\Rightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$
$\therefore \ \vdots$
$x_{t+1} = x_t - \frac{f(x_t)}{f'(x_t)}$

Gibb-Sampling
if Prior is a known mixture.
What if likelihood model is a mixture?
$X_1, \ldots, X_n | \theta \overset{iid}{\sim} \sum_{m=1}^{m} \gamma_m P_m(x|\theta)$

example of this : $X_1 \ldots X_n | \theta, \sigma_1, \theta_1 \ldots \theta_m, \rho \sim \text{iid} \ \underset{\gamma_1}{\rho} N(\theta_1, \sigma_1^2) + \underset{\gamma_2}{(1-\rho)} N(\theta_2, \sigma_2^2)$



ρ — $N(\theta_1, \sigma_1^2)$
$1-\rho$ — $N(\theta_2, \sigma_2^2)$

Goal: Get posterior or function of posterior

$$p(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho \mid x) \propto \left(\prod_{i=1}^{n} p(x_i \mid \theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho)\right) p(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho)$$