Lecture 17:

5/2/17

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

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

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

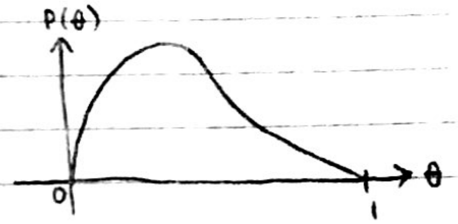
$\Rightarrow P(\theta, \sigma^2 | X) = N(\theta_p, \sigma_p^2) K(\sigma^2 | X)$

Recall:

$X|\theta \sim \text{Bin}(n, \theta)$

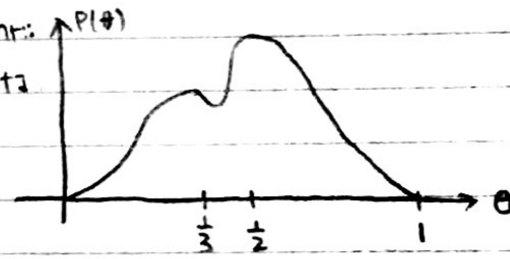
$\theta \sim \text{Beta}(\alpha, \beta)$

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



What if you want:

↳ cannot be a beta



$$P(\theta) = \underbrace{\tfrac{1}{2}}_{\gamma_1} \text{Beta}(\underbrace{3}_{\alpha_1}, \underbrace{3}_{\beta_1}) + \underbrace{\tfrac{1}{2}}_{\gamma_2} \text{Beta}(\underbrace{2}_{\alpha_2}, \underbrace{4}_{\beta_2})$$

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

$P(\theta | X) \propto P(X|\theta) P(\theta) = K(\theta | X)$

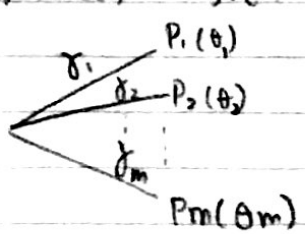
and use a grid search

$$G = \langle \underbrace{\theta_{min}}_{=0}, \theta_{min} + \Delta\theta, \theta_{min} + 2\Delta\theta, \ldots, \underbrace{\theta_{max}}_{=1} \rangle$$

Can we still use conjugacy?

Imagine $P(\theta)$ is a mixture (compound distribution of a discrete # of beta components

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



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

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

$\theta|X \sim ?$

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

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

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

$P_m(\theta|X) = Beta(\alpha_m', \beta_m')$, $\alpha_m' = \alpha_m + x$, $\beta_m' = \beta_m + n - x$

Recall:

$$P(x) = \int_\Theta P(X|\theta)P(\theta)\,d\theta = \int_\Theta P(X|\theta)\sum \gamma_m P_m(\theta)\,d\theta$$

$$= \sum_{m=1}^{M} \gamma_m \overbrace{\int_\Theta P(X|\theta)P_m(\theta)\,d\theta}^{P_m(x) = \text{Beta Binomial}(n, \alpha_m, \beta_m)}$$

If $\gamma_m = \frac{1}{M} \; \forall m$

$$\gamma_m' = \frac{\gamma_m P_m(x)}{P(X)} = \frac{\gamma_m P_m(x)}{\sum \gamma_m P_m(x)} = \frac{P_m(X)}{\sum P_m(X)} \text{ if } \gamma_m = \frac{1}{M} \; \forall m$$

Example:

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

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

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

$n = 10, \; x = 5$

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

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

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

centered at 5

$P_1(5) = \text{dbetabinom}(5, 10, 3, 3) = .147$

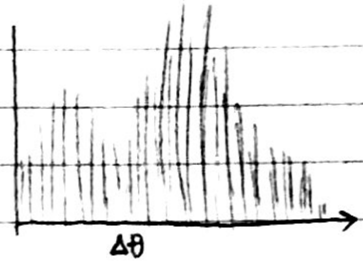
$P_1(X) = \text{BetaBinom}(n, \alpha, \beta) = \text{dbetabinum}(x, 10, 3, 3)$, $n = 10$, $\alpha = 3$, $\beta = 3$

$P_2(5) = \text{dbetabinum}(5, 10, \underbrace{2, 4}_{\text{not centered at 5.}}) = .112$

dbb = dbetabinomial

$$= \frac{1}{dbb(5,10,3,3) + dbb(5,10,2,4)} \left(dbb(5,10,3,3)\, dbeta(\theta,8,8) + dbb(5,10,2,4)\, dbeta(\theta,7,9) \right)$$

$$= .57\ dbeta(\quad) + .43\ dbeta(\quad)$$



<u>Sample from $P(\theta|X)$</u>

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)

③ return $\theta_0 = \gamma_1' \theta_{0,1} + \gamma_2' \theta_{0,2}$

④ Repeat 1-3 many times.

<u>Point Estimate</u>

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

$$= \sum \gamma_m' \int_{\Theta} \theta P_m(\theta|X)\, d\theta$$

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

in our example $= .57 \frac{8}{16} + .43 \frac{7}{16}$

not on exam

$\hat{\theta}_{MAE}$ = sample median

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

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

$$= \sum \gamma_m \underbrace{\left(\binom{n}{x}\frac{\cancel{B(x+\alpha_m, n-x+\beta_m)}}{B(\alpha_m, \beta_m)}\right)}_{\text{PMF of BetaBinom}} \underbrace{\left(\frac{1}{\cancel{B(x+\alpha_m, n-x+\beta_m)}} \theta^{x+\alpha_m-1}(1-\theta)^{n-x+\beta_m-1}\right)}_{\text{PPF of posterior}}$$

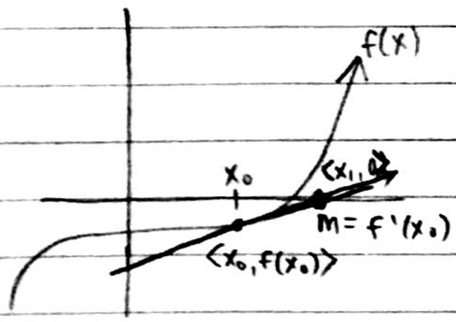
$$= \frac{d}{d\theta}\left[\quad\right] \overset{set}{=} 0$$

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

$$= \sum \frac{1}{B(\alpha_m, \beta_m)}\left((1-\theta)^{n-x+\beta_m-1}(x+\alpha_m-1)\theta^{x+\alpha_m-2} - \theta^{x+\alpha_m-1}(n-x+\beta_m-1)(1-\theta)^{n-x+\beta_m-2}\right)$$

$= 0 \Rightarrow$ can't be solved

Assume $f(x)$ is continuous & differentiable and has one zero on $\mathcal{X}$. Want x^* s.t. $f(x^*) = 0$



Newton's Method:

Step 1: guess $x_0 = x^*$

Step 2: draw tangent line $\Rightarrow y - b = m(x-a) \Rightarrow y - f(x_0) = f'(x_0)(x - x_0)$

Step 3: set $x_1 =$ x-intercept of tangent line

Step 4: repeat by setting $x_0 = x_m$ until $|x_{t+1} - x_t| < \varepsilon$

↳ ε is your accuracy/tolerance level

Solve for x-intercept (x_1):

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

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

$\vdots$

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

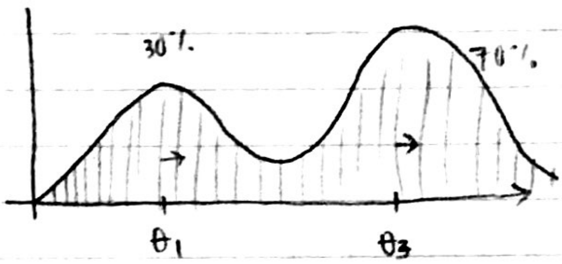
Prior is a known mixture.

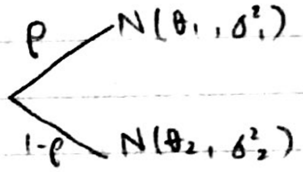
What if likelihood model is a mixture?

$$X_1, \ldots, X_m \mid \theta \overset{iid}{\sim} \sum_{m=1}^{M} \gamma_m P_m(x \mid \theta)$$

$\gamma_1, \ldots, \gamma_m$

ex: $X_1, \ldots, X_n \mid \theta_1, \sigma_1^2, \theta_2, \sigma^2, \rho \overset{iid}{\sim} \underset{\gamma_1}{\rho} N(\theta_1, \sigma_1^2) + \underset{\gamma_2}{(1-\rho)} N(\theta_2, \sigma_2^2)$





Goal - get posterior or function of posterior

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