New Topic...

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

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

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

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

difficult to deal with!! Some things easy though $\hat\theta_{mmse}, \hat\theta_{map}, \hat\theta_{mae}$?

Recall

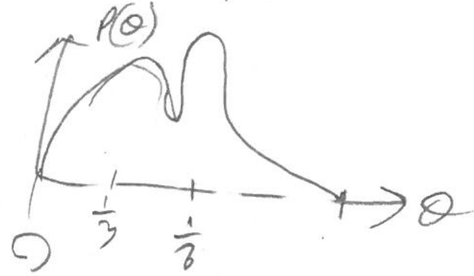
$X \mid \theta \sim \text{Bin}(n, \theta)$

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

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

Beta is the conjugate prior for the binomial likelihood.

But... what if there was no $\text{Beta}(\alpha, \beta)$ which captured your prior belief? e.g.



There is no beta that could do this!!! why?

Can we use this prior? Sure... numerically. How? $P(\theta \mid x) \propto P(x \mid \theta) P(\theta) = k(\theta \mid x)$

$\theta_{min} = 0$, $\theta_{max} = 0$, set $\Delta\theta$... the resolution. Assume you have function form of θ.

Simple... use $G = \langle \theta_{min}, \theta_{min} + \Delta\theta, \theta_{min} + 2\Delta\theta, \ldots, \theta_{max} \rangle$

1. Calculate $k(\theta \mid x) = P(X \mid \theta) P(\theta) \quad \forall \theta \in G$
2. Compute $C \approx \dfrac{1}{\sum_{\theta \in G} k(\theta \mid x)}$
3. Return $P(\theta \mid x) \approx C k(\theta \mid x)$

$\hat\theta_{map} \approx \text{argmax}\, k(\theta \mid x)$, $\hat\theta_{mmse} \approx \sum_{\theta \in G} \theta\, p(\theta \mid x)$ $\Rightarrow$ suffers from some disadvantages

Can we still use conjugacy? Yes...

Imagine $p(\theta)$ as a mixture / compound distribution with conjugate components.

e.g.

$$p(\theta) = \frac{1}{2}\underbrace{Beta(3,3)}_{E(\theta)=\frac{1}{2}} + \frac{1}{2}\underbrace{Beta(2,4)}_{E(\theta)=\frac{1}{3}}$$

this produces something close to our intended prior.

Is $p(\theta)$ "conjugate"? No... but close. Consider

$$p(\theta) = \sum_{m=1}^{M} \gamma_m p_m(\theta)$$

where $\gamma_1 + \gamma_2 + \dots + \gamma_M = 1$ and $p_m(\theta) \sim Beta(\alpha_m, \beta_m)$ i.e. all components are beta with different parameters.

In the case above $M=2$, $\gamma_1 = \frac{1}{2}$, $\gamma_2 = \frac{1}{2}$, $\alpha_1 = 1$, $\beta_1 = 3$, $\alpha_2 = 2$, $\beta_2 = 4$

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

$p(\theta) = \sum \gamma_m p_m(\theta)$

$\theta | X \sim$? Let's see...

$$p(\theta|x) = \frac{p(x|\theta)\,p(\theta)}{p(x)} = \frac{p(x|\theta)\sum \gamma_m p_m(\theta)}{p(x)} = \sum_{m=1}^{M} \gamma_m \frac{p(x|\theta)\,p_m(\theta)}{p(x)} = \sum_{m=1}^{M} \underbrace{\frac{\gamma_m p_m(x)}{p(x)}}_{\gamma_m'} \underbrace{\frac{p(x|\theta)\,p_m(\theta)}{p_m(x)}}_{\substack{Beta(\alpha_m', \beta_m') \\ p_m(\theta|x)}}$$

Posterior has the same form! Thus, is a way it is "conjugate".

Recall

$$P(X) = \int_{\Theta} P(X|\theta)\,P(\theta)\,d\theta = \int_{\Theta} P(X|\theta) \sum_{m=1}^{M} \gamma_m P_m(\theta)\,d\theta = \sum_{m=1}^{M} \gamma_m \overbrace{\int_{\Theta} P(X|\theta) P_m(\theta)\,d\theta}^{P_m(X)}$$

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

If $\gamma_m = \frac{1}{M}\ \forall m$ equal mixing, $P(\theta|X) = \sum_{m=1}^{M} \gamma_m' P_m(\theta|X)$ where $\gamma_m' = \frac{P_m(X)}{\sum_{j=1}^{M} P_j(X)}$

Let's do an example.

Let $P(\theta) = \frac{1}{2}\,Beta(3,3) + \frac{1}{2}\,Beta(2,4)$ as before

$n=10,\ x=$ $\qquad P_m(X) = BetaBinom(n, \alpha_m, \beta_m)$

prior pred. distr!

$P_m(\theta|X) = Beta(\alpha_m', \beta_m')$

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

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

dbetabinom(x, n, α, β) → calc make abbr. dbb

$$= \frac{1}{dbb(5,10,3,3) + dbb(\ ,10,2,4)} \left(dbb(5,10,3,3)\,dbeta(\theta,7,10) + dbb(5,10,2,4)\,dbeta(\theta,6,9) \right)$$

How to ~~plot~~ $P(\theta|X)$. $\theta_{min}=0,\ \theta_{max}=1$ Pick $\Delta\theta$, calc above, graph $\forall \theta \in G$. Plot.

— OR —

① Sample $\theta_{0,1}$ from Beta(7,10) (use rbeta(7,10))

② Sample $\theta_{0,2}$ from Beta(6,9) (use rbeta(6,9))

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

④ repeat 1-3 many times, plot histogram.

$$\frac{1}{0.112 + 0.147}$$

$$= \underset{\gamma_1'}{.43}\,dbeta() + \underset{\gamma_2'}{.57}\,dbeta()$$

Point Estimator

$$\hat{\theta}_{MMSE} = E[\theta|x] = \int_{\Theta} \theta \sum \gamma'_m \, p_m(\theta|x) \, d\theta$$

$$= \sum \gamma'_m \int_{\Theta} \theta \, p_m(\theta|x) \, d\theta$$

$$= \sum \gamma'_m \, E_m[\theta|x]$$

$$= \sum \gamma'_m \frac{\alpha'_m}{\alpha'_m + \beta'_m}$$

In our example

$$\hat{\theta}_{MMSE} = .43 \frac{7}{17} + .57 \frac{6}{15} = .405$$

$\hat{\theta}_{MAP} = ?$ No closed form. Algorithm?

not closed

$\theta_{min} = 0, \theta_{max} = 1, \Delta\theta = ?$

$\theta_0 = 0, F = 0$

repeat {

$\theta_0 = \theta_0 + \Delta\theta$

$F = F + p(\theta_0|x)\Delta\theta$

if $F \geq 0.5$ return θ_0

}

Same algorithm to get CR's.

Look for e.g. $F \geq 0.025$, $F \geq 0.975$

How to get $\hat{\theta}_{map}$

$$\hat{\theta}_{map} := \operatorname*{argmax}_{\theta} \{ p(\theta \mid x) \} = \operatorname*{argmax}_{\theta} \{ k(\theta \mid x) \}$$

$$p(\theta \mid x) \propto \sum_{m=1}^{M} \gamma_m \, p_m(x) \, p_m(\theta \mid x) = k(\theta \mid x) \propto \sum_{m=1}^{M} \underbrace{p_m(x)}_{\text{betabin}} \, \underbrace{p_m(\theta \mid x)}_{\text{beta}} \quad \text{if } \gamma_m = \frac{1}{M} \; \forall m$$

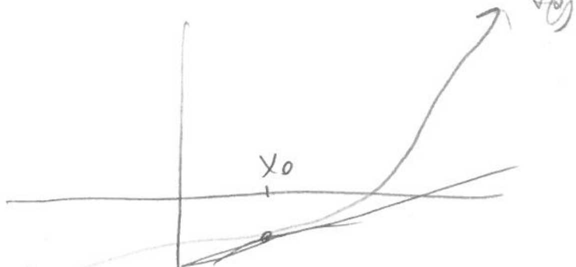
const as $f(m)$

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

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

$\Rightarrow$ No closed form solution

General problem. Find the zero $f(x) = 0$ if f is continuous and has one zero.



Step 1 guess x_0 to be the root

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

Step 3: find the x-intercept of line and call it x_1

Step 4: repeat Steps 1-3 until $|x_{t+1} - x_t| < \varepsilon$, a pre-specified accuracy level (AKA tolerance)

Step 2: $y - b = m(x - a)$

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

Step 3. To find x-intercept, solve for $x = x_1,\ y = 0$

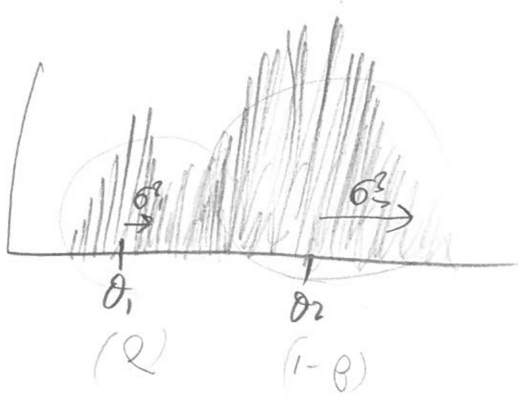
$$-f(x_0) = f'(x_0)(x_1 - x_0) \Rightarrow x_1 - x_0 = -\frac{f(x_0)}{f'(x_0)} \Rightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Step 4: $$x_{t+1} = x_t - \frac{f(x_t)}{f'(x_t)}$$

We just did the case where the prior was a mixture but the likelihood was a single binomial. Now, imagine the prior as single but the likelihood is a mixture! $X_1, \ldots, X_n \mid \ldots \overset{iid}{\sim} \sum_{m=1}^{M} \gamma_m N(\theta_m, \sigma_m^2)$

e.g. $M=2$: $\gamma_1 = \rho,\ \gamma_2 = 1-\rho$

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



$$P(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho \mid X_1, \ldots, X_n) \propto P(X_1, \ldots, X_n \mid \theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho)\, P(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \rho)$$