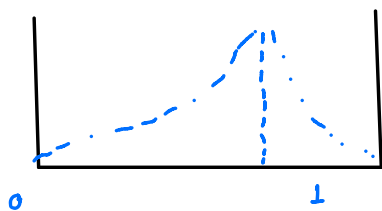


Recall $X|\theta \sim \text{Bin}(n, \theta)$, $\theta \sim \text{Beta}(\alpha, \beta)$



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

and we can make sense of means as weighted averages of prior and likelihood means.

A shorthand for this calculation

$$\begin{aligned} P(\theta|X) &\propto P(X|\theta)P(\theta) \propto \theta^x (1-\theta)^{n-x} \theta^{\alpha-1} (1-\theta)^{\beta-1} \\ &= \theta^{x+\alpha-1} (1-\theta)^{\beta+n-x-1} \sim \text{Beta}(x+\alpha, \beta+n-x) \end{aligned}$$

Post. Predictive Distribution:

(i) $\tilde{X}|\theta \sim \text{Bern}(\theta) \equiv \text{Binom}(1, \theta)$

$$P(\tilde{X}=1|X) = \int_0^1 P(\tilde{X}=1|\theta)P(\theta|X) d\theta = \int_0^1 \theta P(\theta|X) d\theta = \mathbb{E}[\theta|X] = \frac{\alpha+x}{n+\alpha+\beta}$$

(posterior mean)

$$\tilde{X}|X \sim \text{Bern}\left(\frac{\alpha+x}{n+\alpha+\beta}\right)$$

(ii) $\tilde{X}|\theta \sim \text{Binom}(m, \theta)$

$$\begin{aligned} P(\tilde{X}|X) &= \int_0^1 P(\tilde{X}|\theta)P(\theta|X) d\theta = \int_0^1 \binom{m}{\tilde{X}} \theta^{\tilde{X}} (1-\theta)^{m-\tilde{X}} \frac{1}{B(\alpha+x, \beta+n-x)} \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1} d\theta \\ &= \frac{\binom{m}{\tilde{X}}}{B(\alpha+x, \beta+n-x)} \underbrace{\int_0^1 \theta^{\tilde{X}+\alpha+x-1} (1-\theta)^{m-\tilde{X}+\beta+n-x-1} d\theta}_{\text{Beta kernel so}} \end{aligned}$$

$$= \frac{\binom{m}{\tilde{x}} B(\tilde{x} + \alpha + x, m - \tilde{x} + \beta + n - x)}{B(\alpha + x, n - x + \beta)} \quad \text{"Beta-Binomial" with parameters } (m, \alpha + x, n - x + \beta)$$

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

Generally, if we want to simulate $\int_0^1 P(\tilde{X}|\theta) P(\theta|x) d\theta = P(\tilde{X}|x)$

simulate θ then sample $P(\tilde{X}|\theta)$ to get a numerical estimate of $P(\tilde{X}|x)$

$$\boxed{\text{R}} \quad \text{theta} \leftarrow \text{rbeta}(s, \alpha + x, \beta + n - x) \\ \tilde{x} \leftarrow \text{rbinom}(s, m, \text{theta})$$

Def: A family of distributions \mathcal{F} are said to be conjugate for the likelihood $P(x|\theta)$ if $\forall P(\theta) \in \mathcal{F}$ then $P(\theta|x) \in \mathcal{F}$.

Ex: Beta-Binomial

Q: How can we extend this idea?

Ex: $X_1, \dots, X_n | \theta \stackrel{\text{iid}}{\sim} \text{Pois}(\theta)$

first look at likelihood

$$P(x|\theta) = \prod_{i=1}^n P(x_i|\theta) = \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{\theta^{\sum x_i} e^{-n\theta}}{\prod_{i=1}^n x_i!} \quad \text{Want to keep this template}$$

$\theta^{\text{data}} e^{-\text{data} \cdot \theta}$ so $P(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$ needs to be this shape.

$$P(\theta|x) \propto P(x|\theta)P(\theta) \propto \theta^{\sum x_i} e^{-n\theta} \theta^{\alpha-1} e^{-\beta\theta} = \theta^{\sum x_i + \alpha - 1} e^{-(n+\beta)\theta}$$

So prior goes from $\theta^{\alpha-1} e^{-\beta\theta} \rightarrow \theta^{\sum x_i + \alpha - 1} e^{-(n+\beta)\theta}$

$$P(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \quad \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \quad \text{Gamma dist } \overset{\text{shape}}{\text{Gamma}}(\alpha, \overset{\text{rate}}{\beta})$$

Prior: $\theta \sim \text{Gamma}(\alpha, \beta)$ Post: $\theta | X \sim \text{Gamma}(\alpha + \sum x_i, \beta + n)$

Q: What is predictive distribution of $\bar{x} | X$ for $X | \theta \sim \text{Po}(\theta)$

$$\begin{aligned} P(\bar{x} | X) &= \int_0^\infty P(\bar{x} | \theta) P(\theta | X) d\theta = \int_0^\infty \frac{\theta^{\bar{x}} e^{-\theta}}{\bar{x}!} \frac{(\beta+n)^{\alpha+\sum x_i-1}}{\Gamma(\alpha+\sum x_i)} \theta^{\alpha+\sum x_i-1} e^{-(\beta+n)\theta} d\theta \\ &= \frac{1}{\bar{x}!} \frac{(\beta+n)^{\alpha+\sum x_i-1}}{\Gamma(\alpha+\sum x_i)} \int_0^\infty \theta^{\bar{x}+\alpha+\sum x_i-1} e^{-\theta-(\beta+n)\theta} d\theta \end{aligned}$$

Ex: $X | \mu \sim N(\mu, \sigma^2)$, σ^2 known

$$P(X | \mu) \propto \exp\left(-\frac{1}{2\sigma^2} (X - \mu)^2\right) \quad \text{this is where the data is} \Rightarrow \text{match}$$

$$P(\mu) \propto \exp\left(-\frac{1}{2\sigma^2} (\mu_0 - \mu)^2\right), \mu \sim N(\mu_0, \sigma^2)$$

$$P(\mu | X) \propto P(X | \mu) P(\mu) \propto \exp\left\{-\frac{1}{2\sigma^2} [(X - \mu)^2 + (\mu - \mu_0)^2]\right\} \propto \exp\left\{-\frac{1}{2\sigma^2/2} \left(\mu - \frac{X + \mu_0}{2}\right)^2\right\}$$

$$\mu | X \sim N\left(\frac{X + \mu_0}{2}, \frac{\sigma^2}{2}\right)$$

Conjugacy is guaranteed in exponential families.

$$P(X | \theta) = \exp\left\{ \underbrace{g(\theta)}_{\text{natural parameter}} \underbrace{t(x)}_{\text{sufficient stat.}} - b(\theta) + c(x) \right\}$$

If $X_1, \dots, X_n | \theta \sim \text{iid } P(X | \theta)$

$$P(X_1, \dots, X_n | \theta) = \exp\left\{ \underbrace{g(\theta) \sum_{i=1}^n t(x_i)}_{\text{data}} - \underbrace{nb(\theta)}_{\text{data}} + \sum_{i=1}^n c(x_i) \right\}$$

Mimic and replace parameters when data is.

$$P(\theta) \propto \exp\{g(\theta)\tau - r b(\theta)\}$$

$$P(\theta|X) \propto \exp\{g(\theta)\left(\sum_i t(x_i) + \tau\right) - (n+r)b(\theta)\}$$

So the conclusion is we achieve conjugacy with

$$\theta \sim f(\tau, r) \quad \theta|X \sim f\left(\tau + \sum_i t(x_i), r+n\right)$$

Prnk: While we use noninformative priors they pose certain issues

— usually not invariant to reparameterization

Ex: $X|\theta \sim \text{Binom}(n, \theta)$, $\theta \sim \text{Unif}(0,1) \equiv \text{Beta}(1,1)$

Say we want to model $\lambda = \log \frac{\theta}{1-\theta}$

$$P(\lambda) = \left| \frac{d\theta}{d\lambda} \right| P(\lambda^{-1}/\theta) = \frac{e^\lambda}{(1+e^\lambda)^2} \Rightarrow \lambda \sim \text{Logistic}(0,1)$$



not at all uninformative prior.

Solution: Jeffreys' Prior