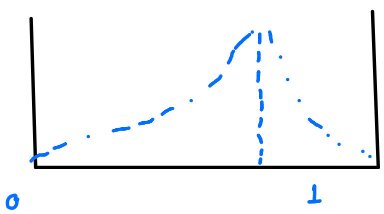
Recall $X|\theta \sim Bin(n,\theta)$, $\theta \sim Beta(\alpha,\beta)$



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

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

A shorthand for this calculation

$$\mathbb{P}(\theta|X) \propto \mathbb{P}(x|\theta)\mathbb{P}(\theta) \propto \theta^{x}(1-\theta)^{n-x}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= \theta^{x+\alpha-1}(1-\theta)^{\beta+n-x-1} \sim Beta(x+\alpha, \beta+n-x)$$

Post. Predictive Distribution:

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

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

(posterior mean)

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

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

$$\mathbb{P}(\tilde{x}|X) = \int_0^1 \mathbb{P}(\tilde{x}|\theta)\mathbb{P}(\theta|X)\,d\theta = \int_0^1 \binom{m}{\tilde{x}}\theta^{\tilde{x}}(1-\theta)^{m-\tilde{\theta}}\frac{1}{B(\alpha+x,\beta+\alpha+x)}\theta^{\alpha+x-1}(1-\theta)^{\beta+\alpha+x-1}\,d\theta$$

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

$$= \frac{\binom{m}{\tilde{x}} B(\tilde{x}+\alpha+x, m-\tilde{x}+\beta+n-x)}{B(\alpha+x, n-x+\beta)}$$ "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 $\int_0^1 \mathbb{P}(\tilde{x}|\theta)\mathbb{P}(\theta|x)d\theta = \mathbb{P}(\tilde{x}|x)$

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

$\boxed{R}$ theta ← rbeta(s, α+x, β+n−x)

$\tilde{x}$ ← rbinom(s, m, theta)

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

if $\forall\ \mathbb{P}(\theta) \in \mathcal{F}$ then $\mathbb{P}(\theta|x) \in \mathcal{F}$.

Ex: Beta–Binomial

Q: How can we extend this idea?

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

first look at likelihood

$$\mathbb{P}(x|\theta) = \prod_{i=1}^{n} \mathbb{P}(x_i|\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{\boxed{\theta^{\Sigma x_i} e^{-n\theta}}}{\prod_{i=1} x_i!}$$ want to keep this template

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

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

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

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

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

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

$$\mathbb{P}(\tilde{X}|X) = \int_0^\infty \mathbb{P}(\tilde{X}|\theta)\mathbb{P}(\theta|X)\,d\theta = \int_0^\infty \frac{\theta^{\tilde{x}} e^{-\theta}}{\tilde{x}\,!} \frac{(\beta+n)^{\alpha+\sum x_i - 1}}{\Gamma(\alpha+\sum x_i)} \theta^{\alpha+\sum_i x_i - 1} e^{-(\beta+n)\theta}\,d\theta$$

$$= \frac{1}{\tilde{x}\,!} \frac{(\beta+n)^{\alpha+\sum x_i - 1}}{\Gamma(\alpha+\sum x_i)} \int_0^\infty \theta^{\tilde{x}+\alpha+\sum x_i - 1} e^{-\theta-(\beta+n)\theta}\,d\theta$$

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

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

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

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

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

Conjugacy is guaranteed in exponential families.

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

If $X_1, \dots, X_n | \theta \sim_{iid} \mathbb{P}(x|\theta)$

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

$\boxed{\text{Mimic and replace parameters when data is.}}$

$$\mathbb{P}(\theta) \propto \exp\left\{g(\theta)\tau - r\, b(\theta)\right\}$$

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

So the conclusion is we achieve conjugacy with

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

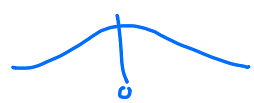
<u>Rmk</u>: While we use noninformative priors they pose certain issues

- usually **not** invariant to reparametrization

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

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

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



not at all uninformative prior.

<u>Solution</u>: Jeffrey's Prior