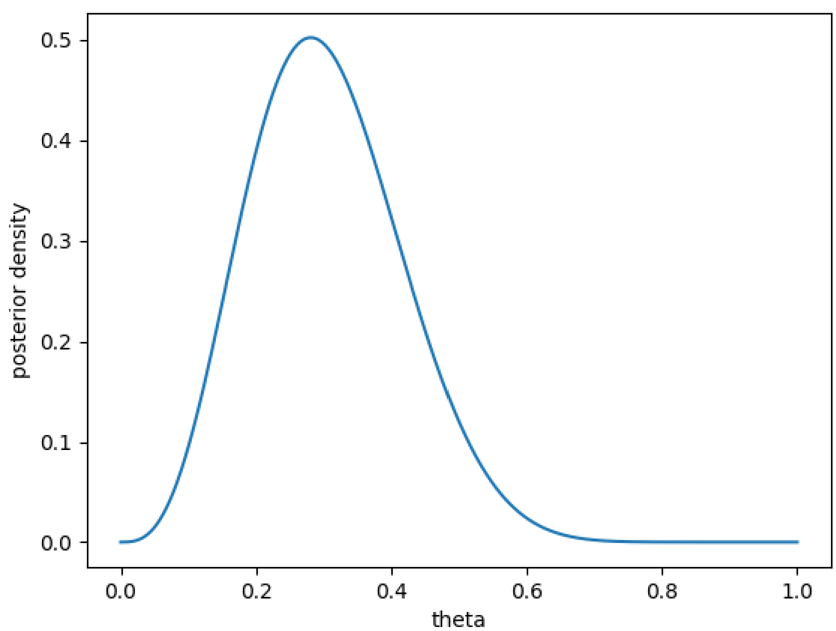
Xinyu Tan

- **1. Posterior inference**

$$\begin{aligned} p(\theta|y<3) &\propto p(\theta)p(y<3|\theta) && (1)\\ &= \text{Beta}(4,4)\sum_{i=0}^{2} p(y=i|\theta) && (2)\\ &= \text{Beta}(4,4)\left((1-\theta)^2 + 10\theta(1-\theta) + 45\theta^2\right)(1-\theta)^8 && (3) \end{aligned}$$

Density plot:



- **2. Predictive distributions**

Denote the result of i-th coin spin x_i, $x_i \in \{T, H\}$. We are given that $x_1 = T$ and $x_2 = T$; for any coin with $p(\text{head}) = p_h$, the probability of the event that until n-th spin a head shows up is:

$$\begin{aligned} p(E) &= p(x_n = H, x_{n-1} = T, \cdots, x_3 = T | x_1 = T, x_2 = T)\\ &= \frac{p(x_n = H, x_{n-1} = T, \cdots)}{p(x_1 = T, x_2 = T)}\\ &= \frac{p_h(1-p_h)^{n-1}}{(1-p_h)^2}\\ &= p_h(1-p_h)^{n-3} \end{aligned}$$

Let's translate the n-th spin to additional spin. Suppose additional spin is m, then total spin is $m+2$, $m \geq 1$. Then $p(E) = p_h(1-p_h)^{m-1}$.

Now, we have two coins with $p(\text{head}|C_1) = 0.6$ and $p(\text{head}|C_2) = 0.4$.

$$\begin{aligned} p(E) &= p(E|C_1)p(C_1) + p(E|C_2)p(C_2) \\ &= 0.6 \times 0.4^{m-1} \times 0.5 + 0.4 \times 0.6^{m-1} \times 0.5 \end{aligned}$$

The expectation of additional spins until a head shows up is:

$$\mathbb{E} = \sum_{m=1}^{\infty} mp(E) = \sum_{m=1}^{\infty} m\left(0.3 \times 0.4^{m-1} + 0.2 \times 0.6^{m-1}\right) = \frac{0.3}{0.6^2} + \frac{0.2}{0.4^2} = 2.08$$

– **5. Posterior distribution as a compromise between prior information and data**

(a) Posterior predictive distribution:

$$\mathbf{Pr}(y = k) = \int_0^1 \mathbf{Pr}(y = k|\theta)d\theta = \int_0^1 \binom{n}{k}\theta^k(1-\theta)^{n-k}d\theta = \frac{1}{n+1} \tag{4}$$

(b) Posterior mean: $\frac{\alpha+y}{\alpha+\beta+n}$

Let's consider one case. If $\frac{\alpha}{\alpha+\beta} > \frac{y}{n} \rightarrow \alpha n > (\alpha+\beta)y$, we then have

$$\frac{\alpha+y}{\alpha+\beta+n} - \frac{y}{n} = \frac{\alpha n - (\alpha+\beta)y}{(\alpha+\beta)(\alpha+\beta+n)} > 0$$

and

$$\frac{\alpha+y}{\alpha+\beta+n} - \frac{\alpha}{\alpha+\beta} = \frac{(\alpha+\beta)y - \alpha n}{(\alpha+\beta)(\alpha+\beta+n)} < 0$$

.

Hence

$$\frac{y}{n} < \frac{\alpha+y}{\alpha+\beta+n} < \frac{\alpha}{\alpha+\beta}$$

.

In a similar fashion, we can show that when $\frac{\alpha}{\alpha+\beta} < \frac{y}{n}$, $\frac{\alpha}{\alpha+\beta} < \frac{\alpha+y}{\alpha+\beta+n} < \frac{y}{n}$, and when $\frac{\alpha}{\alpha+\beta} = \frac{y}{n}$, $\frac{\alpha+y}{\alpha+\beta+n} = \frac{\alpha}{\alpha+\beta} = \frac{y}{n}$

– **7. Noninformative prior densities**

(a) Let ϕ be the natural parameter of θ, $\phi = \text{logit}(\theta) = \log\frac{\theta}{1-\theta}$, then $\theta = \frac{e^\phi}{1+e^\phi}$.

$$p(\phi) = p(\theta)\left|\frac{d\theta}{d\phi}\right| \propto \frac{1}{\theta(1-\theta)}\frac{e^\phi}{(1+e^\phi)^2} = \frac{1}{\theta(1-\theta)}\theta(1-\theta) = 1$$

Hence, ϕ has uniform prior.

– **9. Discrete sample spaces**

(a) The likelihood that I observed a car plate numbered 203 is

$$p(y = 203|N) = \begin{cases} 1/N, & N \geq 203 \\ 0, & \text{otherwise} \end{cases}$$

The posterior distribution hence is

$$p(N|y = 203) \propto \begin{cases} 0.99^N N^{-1}, & N \geq 203 \\ 0, & \text{o.w.} \end{cases}$$

(b) First, we need to find the constant that normalize the posterior distribution.

$$1 = \sum_{n=203}^{\infty} c \times 0.99^n n^{-1} \Rightarrow c = 21.47$$

[I didn't think of error analysis first.]

I estimated the c by adding n from 203 to 20000, the error

$$\epsilon = \sum_{n=20001}^{\infty} 0.99^n n^{-1} \leq \frac{1}{20001} \sum_{n=20001}^{\infty} 0.99^n = \frac{1}{20001} \frac{0.99^{20001}}{1-0.99} \approx 2.5 \times 10^{-90}$$

So error is very small.

$$\mathbb{E}(N) = \sum_{n=203}^{\infty} np(n|y) = \sum_{n=203}^{\infty} nc0.99^n n^{-1} = 279.1$$

and

$$\begin{aligned}\textbf{std}(N) &= \sqrt{\sum_{n=203}^{\infty} (n-279.1)^2 \cdot 21.47 \cdot 0.99^n n^{-1}} \\ &\approx 79.97\end{aligned}$$

– **18. Poisson model**

Gamma distribution:

$$p(\theta) = Gamma(\alpha, \beta) \propto \theta^{\alpha-1} e^{-\beta\theta}$$

If

$$p(y|\theta) \propto \theta^{\sum_{i=1}^n y_i} e^{-(\sum_{i=1}^n x_i)\theta}$$

, then the posterior distribution

$$p(\theta|y) \propto p(\theta)p(y|\theta) = \theta^{\sum_{i=1}^n y_i + \alpha - 1} e^{-(\sum_{i=1}^n x_i + \beta)\theta}$$

Hence,

$$p(\theta|y) = Gamma(\sum_{i=1}^n y_i + \alpha, \sum_{i=1}^n x_i + \beta)$$

– **19. Exponential model with conjugate prior distribution**

(a) Given an i.i.d sample of y values and gamma prior distribution, then

$$p(\theta|y) \propto \theta^n e^{-n\bar{y}\theta} \theta^{\alpha-1} e^{-\beta\theta} = \theta^{\alpha+n-1} e^{-(n\bar{y}+\beta)\theta} \sim Gamma(\alpha+n, n\bar{y}+\beta)$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$ Therefore, gamma prior is conjugate for inferences about θ.

(b) Inverse Gamma distribution:

$$p(\phi) \propto \phi^{-(\alpha+1)} e^{-\beta/\phi}$$

Posterior distribution of ϕ:

$$p(\phi|y) \propto p(\phi)p(y|\phi) = \phi^{-(\alpha+n+1)} e^{-(\beta+n\bar{y})/\phi} \sim \text{Inv-}Gamma(\alpha+n, \beta+n\bar{y})$$

(c) The length of life of a light bulb manufactured by a certain process has an exponential distribution with unknown rate θ.

For prior distribution

$$0.5 = \frac{\textbf{std}(\theta)}{\textbf{avg}(\theta)} = \frac{\sqrt{\alpha/\beta^2}}{\alpha/\beta} = \frac{1}{\sqrt{\alpha}} \Rightarrow \alpha = 4$$

Suppose we need n more samples to have the posterior distribution's coefficient of variation reduce to 0.1:

$$0.1 = \frac{1}{\sqrt{\alpha + n}} \Rightarrow n = 96$$

(d) If in part (c), the coefficient of variation refers to ϕ, then

$$0.5 = \frac{\sqrt{\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}}}{\frac{\beta}{\alpha-1}} = \frac{1}{\sqrt{\alpha - 2}} \Rightarrow \alpha = 6$$

For posterior:

$$0.1 = \frac{1}{\sqrt{\alpha + n - 2}} \Rightarrow n = 96$$

– **21. Simple hierarchical modeling**

Check the ipython notebook