# Bayesian Statistics

1. Bayes' Theorem:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

2. Law of Total Probability:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

3. Bayes' Theorem in continuous case:

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)p(\theta)}{f(\boldsymbol{x})} \propto f(\boldsymbol{x}|\theta)p(\theta)$$

Likelihood: $f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$

$$f(\boldsymbol{x}) = \int_{\Theta} f(\boldsymbol{x}|\theta)p(\theta)d\theta$$

4. Non-informative prior: bad when transformation

5. Jeffrey's prior:

$$p(\theta) \propto \sqrt{I(\theta|\boldsymbol{x})} \propto \sqrt{I(\theta|x)}$$

$$I(\theta|\boldsymbol{x}) = -\mathbb{E}\left[\frac{d^2 \log f(\boldsymbol{x}|\theta)}{d\theta^2}\right]$$

6. Transformation of variables:

$$g(y) = f(x)\left|\frac{dx}{dy}\right|$$

7. HDPI, CI: choose prior backwards

8. Bayes estimate w.r.t. the quadratic loss function $L(\theta, a) = (\theta - a)^2$ is the MEAN of the posterior distribution.

9. Bayes estimate w.r.t. the absolute error loss function $L(\theta, a) = |\theta - a|$ is the MEDIAN of the posterior distribution.

10. Bayes factor for $H_0$ against $H_1$ is

$$B_{01} = \frac{\mathbb{P}(\theta \in \Theta_0|\boldsymbol{x})/\mathbb{P}(\theta \in \Theta_1|\boldsymbol{x})}{p_0/p_1}$$

$B_{01} < 3$: No evidence of $H_0$ over $H_1$
$B_{01} > 3$: Positive evidence for $H_0$
$B_{01} > 20$: Strong evidence for $H_0$
$B_{01} > 150$: Very strong evidence for $H_0$

11. $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$:

$$B_{01} = \frac{f(\boldsymbol{x}|\theta_0)}{f(\boldsymbol{x}|\theta_1)}$$

12. $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$:

$$B_{01} = \frac{f(\boldsymbol{x}|\theta_0)}{\int_{-\infty}^{\infty} f(\boldsymbol{x}|\theta)p_1(\theta)d\theta}$$

13. $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$:

$$B_{01} = \frac{\int_{\Theta_0} f(\boldsymbol{x}|\theta)p_0(\theta)d\theta}{\int_{\Theta_1} f(\boldsymbol{x}|\theta)p_1(\theta)d\theta}$$

14. Prior predictive distribution:

$$f(\boldsymbol{x}) = \int_{\theta\in\Theta} f(\boldsymbol{x}|\theta)p(\theta)d\theta$$

15. Posterior predictive distribution:

$$f(\boldsymbol{y}|\boldsymbol{x}) = \int_{\Theta} f(\boldsymbol{y}|\theta)\pi(\theta|\boldsymbol{x})d\theta$$

# Bayesian Computation

16. Monte Carlo integration:

$$\mathbb{E}_\pi(\theta) \approx \frac{1}{n}\sum_{i=1}^{n}\theta^i,\ \mathbb{E}_\pi\left(f(\theta)\right) \approx \frac{1}{n}\sum_{i=1}^{n} f(\theta^i)$$

$$\begin{aligned}\mathrm{Var}_\pi(\theta) &\approx \frac{1}{n-1}\sum_{i=1}^{n}\left(\theta^i - \mathbb{E}_\pi(\theta^i)\right)^2 \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\theta^i)^2 - \frac{1}{n}\left(\sum_{i=1}^{n}\theta^i\right)^2\right]\end{aligned}$$

17. Method of Inversion:

- Calculate cdf $F(x)$
- Let $F(x) = u$, calculate $x = F^{-1}(u)$

Algorithm:
Step 1: Simulate $U \sim U[0, 1]$
Step 2: Set $X = F^{-1}(U)$
Discrete case: (check WS4 Q2)

18. Rejection Sampling:

- Calculate $M = \sup_p \left(\frac{\pi(\theta|\boldsymbol{x})}{h(\theta)}\right)$

Algorithm:
Step 1: Simulate $\theta \sim h(\theta)$
Step 2: Generate $Y \sim U[0, Mh(\theta)]$
Step 3: Accept $\theta$ iff $Y \leq \pi(\theta|\boldsymbol{x})$

19. Importance Sampling:
    Step 1: Simulate $\theta \sim g(\theta)$
    Step 2: Calculate weight
    $$w_i = \frac{\pi(\theta^i|\boldsymbol{x})}{g(\theta^i)}$$
    Step 3: Calculate estimator
    $$\hat{\theta}_g = \frac{1}{n}\sum_{i=1}^{n} w(\theta^i)\theta^i$$

20. Sampling Importance Resampling (SIR):
    Step 1: Simulate $\theta \sim g(\theta)$
    Step 2: Calculate weight
    $$w_i = \frac{\pi(\theta^i|\boldsymbol{x})}{g(\theta^i)}$$
    Step 3: Calculate updated weight
    $$w_i^* = \frac{w_i}{\sum_{j=1}^{n} w_j}$$
    Step 4: Resample these values from these weight with replacement, with resampling probability $w_i^*$

21. Gibbs Sampler:
    Step 1: Set initial parameter value for $\theta_1$ and $\theta_2$ denoted by $\theta_1^0$ and $\theta_2^0$
    Step 2: Generate $\theta_1^{t+1} \sim \pi(\theta_1^{t+1}|\boldsymbol{x}, \theta_2^t)$
    Step 3: Generate $\theta_2^{t+1} \sim \pi(\theta_2^{t+1}|\boldsymbol{x}, \theta_1^{t+1})$
    Step 4: Increase $t$ by one and return to Step 2
    Step 5: Discard $\boldsymbol{\theta}^0, \ldots, \boldsymbol{\theta}^N$ as burn-in

22. Metropolis-Hastings:
    Step 1: Set initial parameter value for $\boldsymbol{\theta}$ denoted by $\boldsymbol{\theta}^0$
    Step 2: Given the current position, $\boldsymbol{\theta}^t = \boldsymbol{\theta}$, generate a new value $\boldsymbol{\theta}'$ from the distribution $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$
    Step 3: Calculate
    $$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left(1, \frac{\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}'|\boldsymbol{\theta})}{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}|\boldsymbol{\theta}')}\right)$$
    Step 4: With probability $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}')$, set $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}'$, else set $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}$
    Step 5: Discard $\boldsymbol{\theta}^0, \ldots, \boldsymbol{\theta}^N$ as burn-in