Handout 5: Conjugate Prior Distributions

STA 131B

Summary of Common Conjugate Priors:

Model $f(x\|\theta)$	Conjugate Prior $\xi(\theta)$	Posterior Distribution $\xi(\theta\|\mathbf{X})$
Bernoulli$(n, \theta), n$ known	Beta(α, β)	Beta$(\sum_{i=1}^n X_i + \alpha, n - \sum_{i=1}^n X_i + \beta)$
Negative Binomial$(r, \theta), r$ known	Beta(α, β)	Beta$(\alpha + nr, \beta + \sum_{i=1}^n X_i)$
Poisson(θ)	Gamma$(\alpha, \beta), \alpha =$ shape	Gamma$(\alpha + \sum_{i=1}^n X_i, \beta + n)$
Normal$(\theta, \sigma^2), \sigma^2$ known	Normal(τ, ν^2)	Normal$(\frac{\sigma^2\tau + n\nu^2\overline{X}}{\sigma^2 + n\nu^2}, \frac{\sigma^2\nu^2}{\sigma^2+n\nu^2})$
Exponential(θ)	Gamma(α, β)	Gamma$(\alpha + n, \beta + \sum_{i=1}^n X_i)$
Uniform$(0, \theta)$	Pareto(λ, r)	Pareto$(\max\{\lambda, X_{(n)}\}, n + r)$

* Note that the p.d.f. of the Pareto(λ, r) distribution is given by

$$f(x|\lambda, r) = \frac{r\lambda^r}{x^{r+1}}, \text{ for } x > \lambda.$$

* Also, the negative binomial distribution with parameters r and θ describes the number of failures before the r^{th} success, where the probability of success is θ.

Example 1:

Suppose that in a given population the probability of catching a cold is p. A sample $X_1, \ldots, X_n$ of the population is taken, with

$$X_i = \begin{cases} 1 \text{ if the } i^{th} \text{ person catches a cold} \\ 0 \text{ otherwise} \end{cases}$$

This implies that $X_1, \ldots, X_n|\theta \sim$ Binomial$(n = 1, \theta = p) \equiv$ Bernoulli$(\theta = p)$, and assume the parameter space is $\Theta = [0, 1]$. Then we have

$$f_{\mathbf{X}|\theta}(\mathbf{x}|\theta) = \theta^{\sum_{i=1}^n X_i}(1-\theta)^{n-\sum_{i=1}^n X_i}.$$

If the prior distribution of θ is Beta(α, β), i.e.

$$\xi(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \text{ for } \alpha, \beta > 0,$$

then

$$\begin{aligned} f_{\mathbf{X},\theta}(\mathbf{x}, \theta) &= \theta^{\sum_{i=1}^n X_i}(1-\theta)^{n-\sum_{i=1}^n X_i} \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &\propto \theta^{\sum_{i=1}^n X_i+\alpha-1}(1-\theta)^{n-\sum_{i=1}^n X_i+\beta-1} \\ &\sim \text{Beta}(\sum_{i=1}^n X_i + \alpha, n - \sum_{i=1}^n X_i + \beta). \end{aligned}$$

Therefore $\xi(\theta|\mathbf{X}) \sim \text{Beta}(\sum_{i=1}^n X_i + \alpha, n - \sum_{i=1}^n X_i + \beta)$.

Notice here that the posterior distribution is also a Beta distribution. We say "the Beta distribution is a conjugate prior family of prior distributions for samples from a Bernoulli distribution." Another way to put this is that the family of Beta distributions is closed under sampling from a Bernoulli distribution (both prior and posterior distributions are from the Beta family).

In this scenario, α and β are called "prior hyperparameters", while $(\sum_{i=1}^n X_i + \alpha, n - \sum_{i=1}^n X_i + \beta)$ are called "posterior hyperparameters".