Beta - Bernoulli

$$X_i \sim Ber(\theta)$$

$$N = N_1 + N_0$$

$$p(D|\theta) = \theta^{N_1}(1-\theta)^{N_0}$$

$$N_1 = \sum_{i=1}^{N} I(X_i = 1) \; ; \quad N_0 = \sum_{i=1}^{N} I(X_i = 0)$$

$$p(\theta) = Beta(\theta|a,b) \propto \theta^{a-1}(1-\theta)^{b-1}$$

$$p(\theta|D) \propto p(D|\theta)\cdot p(\theta)$$

$$\Rightarrow p(\theta|D) \propto \theta^{N_1}(1-\theta)^{N_0}\,\theta^{a-1}(1-\theta)^{b-1}$$

$$\Rightarrow p(\theta|D) \propto \theta^{N_1+a-1}(1-\theta)^{N_0+b-1}$$

$$\Rightarrow p(\theta|D) \sim Beta(N_1+a, N_0+b)$$

$$\beta(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$$

normalizing constant for Beta distribution

Marginal likelihood

$$p(y|\alpha) = \int p(y,\theta|\alpha)\,d\theta$$

$$= \int p(y|\theta)\cdot p(\theta|\alpha)\,d\theta$$

$$= \int \frac{\boxed{Ber(\theta)\cdot Beta(a,b)\,d\theta}}{\beta(a,b)}$$

equal to normalizing constant for Posterior

$$= \frac{\beta(N_1+a, N_0+b)}{\beta(a,b)}$$

$$\left(\frac{\text{normalizing constant for Posterior}}{\text{normalizing constant for prior}}\right)$$

1.3 Posterior mean

$$E_\theta[\theta_i] = \int \theta \cdot p(\theta \mid y, \alpha) \cdot d\theta$$

$$= \int \theta \cdot p(\theta \mid D, a, b) \cdot d\theta$$

$$= \int \theta \cdot \frac{\left[\theta^{N_1+a-1}(1-\theta)^{N_0+b-1}\right]}{B(a+N_1, b+N_0)} d\theta$$

$$\beta(a,b) = \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)}$$

$$\Rightarrow E_\theta[\theta_i] = \frac{\int \theta^{N_1+a}(1-\theta)^{N_0+b-1} \, d\theta}{B(a+N_1, b+N_0)}$$

$$= \frac{B(a+N_1+1, b+N_0)}{B(a+N_1, b+N_0)}$$

$$= \left[\frac{\Gamma(a+N_1+1)\,\Gamma(b+N_0)}{\Gamma(a+b+N+1)}\right]\left[\frac{\Gamma(a+b+N)}{\Gamma(a+N_1)\,\Gamma(b+N_0)}\right]$$

$$\Gamma(a+1) = a\,\Gamma(a)$$

$$\Rightarrow E_\theta[\theta_i] = \frac{a+N_1}{a+b+N} \quad //$$

1.4 Posterior Predictive

$$P(\hat{y} \mid y, \alpha) = \int p(\hat{y}, \theta \mid y, \alpha) \cdot d\theta$$

$$= \int p(\hat{y}, \theta \mid D, a, b) \cdot d\theta$$

$$= \int p(\hat{y} \mid \theta, D, a, b) \cdot p(\theta \mid D, a, b) \cdot d\theta$$

$$= \int \frac{\theta^{I(\hat{y}=1)} \cdot (1-\theta)^{I(\hat{y}=0)} \cdot \theta^{N_1+a-1} \cdot (1-\theta)^{N_0+b-1}}{\beta(N_1+a, N_0+b)} \cdot d\theta$$

$$p(\hat{y} \mid D, a, b) = \frac{1}{\beta(N_1+a, N_0+b)} \int \theta^{I(\hat{y}=1)+N_1+a-1} \cdot (1-\theta)^{N_0+b-1+I(\hat{y}=0)} \cdot d\theta$$

$$p(\hat{y}=1 \mid D, a, b) = \frac{1}{\beta(N_1+a, N_0+b)} \int \theta \cdot \theta^{N_1+a-1} (1-\theta)^{N_0+b-1} \cdot d\theta$$

$$= E[\theta]$$

$$\Rightarrow \quad p(\hat{y}=1 \mid D, a, b) = \frac{N_1+a}{N+a+b}$$