1 Probability distributions

1.1 Beta distribution

1.2 Binomial distribution

2 Bayesian parameter estimation

2.1 Beta-Bernoulli model

2.1.1 Summary

The model

$$X_i \sim \mathrm{Ber}(\theta), \text{for } i \in \{1, \dots, N\} \tag{2.1}$$

$$\mathcal{D} = \{x_1, \dots, x_N\} \tag{2.2}$$

$$N_1 = \sum_{i=1}^{N} \mathbb{I}(x_i = 1) \tag{2.3}$$

$$N_0 = \sum_{i=1}^{N} \mathbb{I}(x_i = 0) \tag{2.4}$$

Likelihood

$$p(\mathcal{D}|\theta) = \theta^{N_1}(1-\theta)^{N_0} \tag{2.5}$$

Prior

$$p(\theta) = \mathrm{Beta}(\theta|a, b) \tag{2.6}$$

Posterior

$$p(\theta|\mathcal{D}) = \mathrm{Beta}(\theta|N_1 + a, N_0 + b) \tag{2.7}$$

Posterior predictive

$$p(x = 1|\mathcal{D}) = \frac{a}{a+b} \tag{2.8}$$

Evidence

2.1.2 Derivations

2.2 Beta-binomial model

2.2.1 Summary

The model

$$N_1 \sim \mathrm{Bin}(N, \theta) \tag{2.9}$$
$$\mathcal{D} = \{N_1, N\} \tag{2.10}$$
$$N_1 = \text{number of successes} \tag{2.11}$$
$$N = \text{total number of trials} \tag{2.12}$$
$$\mathcal{D}' = \{N_1', N'\} \tag{2.13}$$
$$N_1' = \text{number of successes in a new batch of data} \tag{2.14}$$
$$N' = \text{total number of trials in a new batch of data} \tag{2.15}$$

Likelihood

$$p(\mathcal{D}|\theta) = \mathrm{Bin}(N_1|N, \theta) \tag{2.16}$$

Prior

$$p(\theta) = \mathrm{Beta}(\theta|a, b) \tag{2.17}$$

Posterior

$$p(\theta|\mathcal{D}) = \mathrm{Beta}(\theta|N_1 + a, N_0 + b) \tag{2.18}$$

Posterior predictive

$$p(\mathcal{D}'|\mathcal{D}) = \mathrm{Bb}(N_1'; a, b, N') \tag{2.19}$$

Evidence

2.2.2 Derivations

2.3 Dirichlet-categorical model

2.3.1 Summary

The model

$$X_i \sim \mathrm{Cat}\left(\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^T\right), \text{for } i \in \{1, \ldots, N\} \tag{2.20}$$
$$\mathcal{D} = \{x_1, \ldots, x_N\} \tag{2.21}$$
$$n_k = \sum_{i=1}^{N} \mathbb{I}(x_i = k) \tag{2.22}$$

Likelihood

$$p(\mathcal{D}|\theta) = \prod_{k=1}^{K} \theta_k^{n_k} \tag{2.23}$$

Prior

$$p(\theta) = \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha}) \tag{2.24}$$

Posterior

$$p(\theta|\mathcal{D}) = \mathrm{Dir}(\boldsymbol{\theta}; \alpha_1 + n_1, \ldots, \alpha_K + n_K) \tag{2.25}$$

Posterior predictive

$$p(X = j|\mathcal{D}) = \frac{\alpha_j + n_j}{\alpha_0 + N} \tag{2.26}$$

$$\text{where } \alpha_0 = \sum_{k=1}^{K} \alpha_k \tag{2.27}$$

Evidence

2.3.2 Derivations

2.4 Dirichlet-multinomial model

2.4.1 Summary

The model

$$\mathbf{N} \sim \mathrm{Mult}(N, \boldsymbol{\theta}) \in \mathbb{R}^K \tag{2.28}$$

$$\mathcal{D} = \{\mathbf{n} = \text{vector of counts of successes}\} \tag{2.29}$$

$$N = \sum_{i=1}^{K} n_i \tag{2.30}$$

$$\mathcal{D}' = \{\mathbf{n}' = \text{vector of counts of successes in a new batch of data}\} \tag{2.31}$$

$$N' = \sum_{i=1}^{K} n_i' \tag{2.32}$$

Likelihood

$$p(\mathcal{D}|\theta) = \mathrm{Mult}(\mathbf{n}; N, \boldsymbol{\theta}) \tag{2.33}$$

Prior

$$p(\theta) = \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha}) \tag{2.34}$$

Posterior

$$p(\theta|\mathcal{D}) = \mathrm{Dir}(\boldsymbol{\theta}; \alpha_1 + n_1, \ldots, \alpha_K + n_K) \tag{2.35}$$

Posterior predictive

$$p(\mathcal{D}'|\mathcal{D}) = \frac{\Gamma(\alpha_0 + N)}{\Gamma(\alpha_0 + N + N')} \prod_{k=1}^{K} \frac{\Gamma(\alpha_k + n_k + n'_k)}{\Gamma(\alpha_k + n_k)} \tag{2.36}$$

$$\text{where } \alpha_0 = \sum_{k=1}^{K} \alpha_k \tag{2.37}$$

Evidence

$$p(\mathcal{D}|\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + N)} \prod_{k=1}^{K} \frac{\Gamma(\alpha_k + n_k)}{\Gamma(\alpha_k)} \tag{2.38}$$

2.4.2 Derivations