1 Probability distributions

1.1 Uniform distribution

1.2 Beta distribution

1.3 Bernoulli distribution

1.4 Binomial distribution

1.5 Beta-binomial distribution

1.6 Categorical distribution

1.7 Dirichlet distribution

1.8 Multinomial distribution

1.9 Pareto 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

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

Posterior predictive

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

Evidence

2.1.2 Derivations

2.2 Beta-binomial model

2.2.1 Summary

The model

$$N_1 \sim \text{Bin}(N, \theta) \tag{2.9}$$
$$\mathcal{D} = \{N_1, N\} \tag{2.10}$$
$$N_1 = \text{number of successes} \tag{2.11}$$
$$N = \text{total number of trials} \tag{2.12}$$
$$\tilde{\mathcal{D}} = \{\tilde{N}_1, \tilde{N}\} \tag{2.13}$$
$$\tilde{N}_1 = \text{number of successes in a new batch of data} \tag{2.14}$$
$$\tilde{N} = \text{total number of trials in a new batch of data} \tag{2.15}$$

Likelihood

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

Prior

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

Posterior

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

Posterior predictive

$$p(\tilde{\mathcal{D}}|\mathcal{D}) = \text{Bb}(\tilde{N}_1; a', b', \tilde{N}) \tag{2.19}$$

Evidence

2.2.2 Derivations

2.3 Dirichlet-categorical model

2.3.1 Summary

The model

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

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

$$p(\theta|\mathcal{D}) = \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha}' = \boldsymbol{\alpha} + (n_1, \dots, n_K)^T) \tag{2.25}$$

Posterior predictive

$$\begin{aligned} p(\tilde{X} = j|\mathcal{D}) &= \frac{\alpha'_j}{\sum_{k=1}^{K} \alpha'_i} && (2.26) \\ &= \frac{\alpha_j + n_j}{\alpha_0 + N} && (2.27) \\ \text{where } \alpha_0 &= \sum_{k=1}^{K} \alpha_k && (2.28) \end{aligned}$$

Evidence

2.3.2 Derivations

2.4 Dirichlet-multinomial model

2.4.1 Summary

The model

$$\begin{aligned} &\mathbf{N} \sim \mathrm{Mult}(N, \boldsymbol{\theta}) \in \mathbb{R}^K && (2.29) \\ &\mathcal{D} = \{\mathbf{n} = \text{vector of counts of successes}\} && (2.30) \\ &N = \sum_{i=1}^{K} n_i && (2.31) \\ &\tilde{\mathcal{D}} = \{\tilde{\mathbf{n}} = \text{vector of counts of successes in a new batch of data}\} && (2.32) \\ &\tilde{N} = \sum_{i=1}^{K} \tilde{n}_i && (2.33) \end{aligned}$$

Likelihood

$$p(\mathcal{D}|\theta) = \mathrm{Mult}(\mathbf{n}; N, \boldsymbol{\theta}) \tag{2.34}$$

Prior

$$p(\theta) = \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha}) \tag{2.35}$$

Posterior

$$p(\theta|\mathcal{D}) = \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha}' = \boldsymbol{\alpha} + (n_1, \dots, n_K)^T) \tag{2.36}$$

Posterior predictive

$$p(\tilde{\mathcal{D}}|\mathcal{D}) = \frac{\Gamma(\alpha_0 + N)}{\Gamma(\alpha_0 + N + \tilde{N})} \prod_{k=1}^{K} \frac{\Gamma(\alpha_k + n_k + \tilde{n}_k)}{\Gamma(\alpha_k + n_k)} \tag{2.37}$$

$$\text{where } \alpha_0 = \sum_{k=1}^{K} \alpha_k \tag{2.38}$$

Evidence

$$p(\mathcal{D}|\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + N)} \prod_{k=1}^{K} \frac{\Gamma(\alpha_k + n_k)}{\Gamma(\alpha_k)} \tag{2.39}$$

2.4.2 Derivations

2.5 Poisson-gamma model

2.5.1 Summary

The model

$$x \sim \mathrm{Poi}(\lambda) \tag{2.40}$$

$$\mathcal{D} = \{x_1, \dots, x_N\} \tag{2.41}$$

Likelihood

$$p(\mathcal{D}|\lambda) = \prod_{i=1}^{N} \frac{\lambda^{x_i}}{x_i!} \exp(-\lambda) \tag{2.42}$$

Prior

$$p(\lambda) = \mathrm{Gamma}(\lambda; a, b) \tag{2.43}$$

Posterior

$$p(\lambda|\mathcal{D}) = \mathrm{Gamma}\left(\lambda; a' = a + \sum_{i=1}^{N} x_i, b' = b + N\right) \tag{2.44}$$

Posterior predictive

$$p(\tilde{x}|\mathcal{D}) = \mathrm{NB}(\tilde{x}|a', \frac{1}{1+b'}) \tag{2.45}$$

Evidence

$$p(\mathcal{D}) = \tag{2.46}$$

2.5.2 Derivations

3 Sampling algorithms