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

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

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

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

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

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

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

2.1.2 Derivations

2.2 Beta-binomial model

2.2.1 Summary

The model

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

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

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

2.2.2 Derivations

2.3 Dirichlet-categorical model

2.3.1 Summary

The model

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

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

$$p(\theta|\mathcal{D}) = \mathrm{Dir}(\boldsymbol{\theta}; \alpha_1 + N_1, \dots, \alpha_K + N_K) \tag{2.25}$$

Posterior predictive

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

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

2.3.2 Derivations