What we have covered...
Chapters

- Probability and inference (Ch 1)
- Single-parameter models (Ch 2)
- Introduction to multiparameter models (Ch 3)
- Asymptotics and connections to non-Bayesian approaches (Ch 4)
- Hierarchical models (Ch 5)
- Model checking (Ch 6)
- Model comparison and cross validation (Ch 7)
- Decision theory (Ch 9)
- Bayesian Computation (Ch 10, 11)

Probability and inference

- Three steps of Bayesian data analysis (Sec 1.1)
  - Set up a full probability model: $p(y|\theta)$ and $p(\theta)$
  - Condition on observed data: $p(\theta|y)$
  - Evaluate the fit of the model: $p(y^{rep}|y)$
- Bayesian inference via Bayes' rule (Sec 1.3)
  - Parameter posteriors: $p(\theta|y) \propto p(y|\theta)p(\theta)$
  - Predictions: $p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$
  - Model probabilities $p(M|y) \propto p(y|M)p(M)$ where $p(y|M) = \int p(y|\theta, M)p(\theta|M)d\theta$.
- Interpreting Bayesian probabilities (Sec 1.5)
  - Epistemic probability: my belief
  - Frequency probability: long run percentage
- Computation (Sec 1.9)
  - Inference via simulations
  - example using inverse CDF

Single-parameter models (Ch 2)

General

- Priors
  - Conjugate (Sec 2.4)
- Posteriors
  - Compromise between data and prior (2.2)
  - Point estimation
  - Credible intervals (Sec 2.3)

Specific models

- Binomial (Sec 2.1–2.4)
- Normal, unknown mean (Sec 2.5)
- Normal, unknown variance (Sec 2.6)
- Poisson (Sec 2.6)
- Exponential (Sec 2.6)
- Poisson with exposure (Sec 2.7)

Introduction to multiparameter models (Ch 3)

- Joint posterior

$$p(\theta_1, \ldots, \theta_n | y) \propto p(y | \theta_1, \ldots, \theta_n) p(\theta_1, \ldots, \theta_n)$$

- Marginal posterior

$$p(\theta_1 | y) = \int \cdots \int p(\theta_1, \ldots, \theta_n | y) d\theta_2 \cdots d\theta_n$$

- Conditional posteriors

$$p(\theta_2, \ldots, \theta_n | \theta_1, y) \propto p(\theta_1, \ldots, \theta_n | y)$$

- Posterior decomposition, e.g.

$$p(\theta_1, \ldots, \theta_n | y) = p(\theta_1 | y) \prod_{i=2}^{n} p(\theta_i | \theta_{1:i-1}, y)$$

  where $1 : i - 1 = 1, 2, \ldots, i - 1$.

- Conditional independence, e.g.

$$p(\theta_i | \theta_{1:i-1}, y) = p(\theta_i | \theta_{i-1}, y)$$

Normal model

- Normal model with default prior (Sec 3.2)

$$y_i \overset{iid}{\sim} N(\mu, \sigma^2) \quad p(\mu, \sigma^2) \propto 1/\sigma^2$$

results in

$$p(\mu, \sigma^2|y) = N(\overline{y}, \sigma^2/n)\text{Inv-}\chi^2(n-1, s^2)$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i - \overline{y})^2$.

- Normal model with conjugate prior (Sec 3.3)

$$y \overset{iid}{\sim} N(\mu, \sigma^2) \quad \mu|\sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0) \quad \sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$$

results in

$$p(\mu, \sigma^2|y) = N\left(\frac{\kappa_0\mu_0 + n\overline{y}}{\kappa_0 + n}, \frac{\sigma^2}{\kappa_0 + n}\right)\text{Inv-}\chi^2(\nu_0 + n, \sigma_n^2)$$

where $\sigma_n^2 = \left[\nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\overline{y} - \mu_0)^2\right]/(\nu_0 + n)$.

Data asymptotics (Ch 4)

Consider a model $y_i \stackrel{iid}{\sim} p(y|\theta_0)$ for some true value θ_0.

- Point estimation:

$$\hat{\theta}_{Bayes} \to \hat{\theta}_{MLE} \xrightarrow{p} \theta_0$$

- Limiting distribution:

$$\theta|y \xrightarrow{d} N(\hat{\theta}, \frac{1}{n} I(\hat{\theta})^{-1})$$

Hierarchical models (Ch 5)

- Hierarchical model (Ch 5):

$$p(\theta, \phi | y) \propto p(y|\theta) p(\theta|\phi) p(\phi)$$

- Exchangeability (Sec 5.2)

$$p(y_1, \ldots, y_n) = p(y_{\pi_1}, \ldots, y_{\pi_n})$$

- Hierarchical binomial model (Sec 5.3):

$$y_i \overset{iid}{\sim} Bin(n_i, \theta_i) \quad \theta_i \overset{iid}{\sim} Be(\alpha, \beta)$$

- Hierarchical normal model (Sec 5.4)

$$y_{ij} \overset{iid}{\sim} N(\mu_j, \sigma_j^2) \quad \mu_j \overset{iid}{\sim} N(\eta, \tau^2)$$

- Data replications

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(\theta|y)d\theta$$

- Graphical posterior predictive checks (Sec 6.4)
- Posterior predictive pvalues (Sec 6.3)

$$p_B = P(T(y^{rep},\theta) \geq T(y,\theta)|y)$$

for a test statistic $T(y,\theta)$.

Model comparison (Ch 7)

- Cross validation (leave one out/ k-fold)
- ELPD: Expected log posterior predictive density
- IC: AIC/ BIC/ WAIC ....

In order to make a decision, a utility (or loss) function, i.e. $U(\theta, d) = -L(\theta, d)$, must be set where d is the decision. Then the optimal Bayesian decision is to maximize expected utility (or minimize expected loss), i.e.

$$\text{argmax}_d \int U(\theta, d) p(\theta) d\theta$$

where $p(\theta)$ represents your current state of belief, i.e. it could be a prior or a posterior depending on your perspective.

Example for utility function and point estimate

- Mean minimizes squared error
- Median minimizes absolute error

- Grid sampling
- Direct sampling: transformation/ inverse-CDF
- Rejection sampling
- Importance sampling
- MCMC: Markov chain, Gibbs sampling, Metropolis, Metropolis-Hasting sampling