Bayesian modelling of groups and individuals

Empirical and hierarchical Bayesian methods

Christos Dimitrakakis

Frankfurt Institute for Advanced Studies, Goethe University, Germany

12/4/2011

Beliefs, preferences and models

The problem of drawing conclusions from evidence

- How can we test the main assumptions in a behavioural experiment?
- How can we examine multiple hypotheses in a unified framework?
- How can we draw conclusions from experiments in a small group?

Beliefs, preferences and models

The problem of drawing conclusions from evidence

- How can we test the main assumptions in a behavioural experiment?
- How can we examine multiple hypotheses in a unified framework?
- How can we draw conclusions from experiments in a small group?

Assumptions + Evidence → Conclusion

- Bayesian inference.
- Dempster-Shafer theory of evidence.
- Plausibility theory.

Beliefs, preferences and models

The problem of drawing conclusions from evidence

- How can we test the main assumptions in a behavioural experiment?
- How can we examine multiple hypotheses in a unified framework?
- How can we draw conclusions from experiments in a small group?

Assumptions + Evidence → Conclusion

- Bayesian inference.
- Dempster-Shafer theory of evidence.
- Plausibility theory.

Drawing conclusions is not always the same as making reject/accept decisions

Bayesian inference

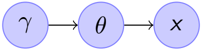


Figure: Graphical model for known prior, single subject.

- A study involving one subject.
- The subject provides us with observations x.
- We assume that the observations are generated $x \mid \theta \sim P(\cdot \mid \theta)$.
- The unknown θ fully characterises the subject with respect to our observations x.
- We assume that $\theta \mid \gamma \sim Q(\cdot \mid \gamma)$.

Bayesian inference

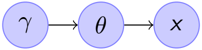


Figure: Graphical model for known prior, single subject.

- A study involving one subject.
- The subject provides us with observations x.
- We assume that the observations are generated $x \mid \theta \sim P(\cdot \mid \theta)$.
- The unknown θ fully characterises the subject with respect to our observations x.
- We assume that $\theta \mid \gamma \sim Q(\cdot \mid \gamma)$.

Known γ, use Bayes' theorem

We only need to condition the distribution of θ to the data of the subject:

$$Q(\theta \mid x, \gamma) = \frac{P(x \mid \theta, \gamma) Q(\theta \mid \gamma)}{\int_\Theta P(x \mid \theta', \gamma) Q(\theta' \mid \gamma) \, d\theta'}$$

Bayesian inference

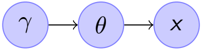


Figure: Graphical model for known prior, single subject.

Example

- $x = x_1, \ldots, x_T$, and $x \in \{0, 1\}$, i.e. $0 =$ failure, $1 =$ success.
- $\theta \in [0, 1]$: probability of success, so: $x_t \mid \theta \sim \mathcal{Bern}(\theta)$.
- Prior for Bernoulli parameters: $\theta \mid \gamma \sim \mathcal{Beta}(\alpha_\gamma, \beta_\gamma)$.
- In this case the posterior is:

$$\theta \mid \gamma, x \sim \mathcal{Beta}(\alpha_\gamma + \sum_t x_t, \beta_\gamma + T - \sum_t x_t)$$

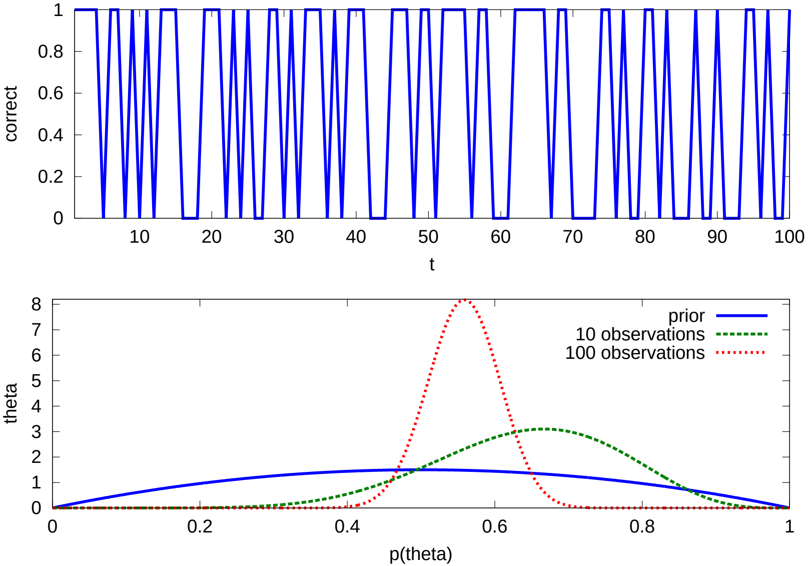


Figure: An example

Overview

- A study involving n subjects.
- The k-th subject provides us with observations x_k.
- We assume that each observation is generated as $x_k \mid \theta_k \sim P(\cdot \mid \theta_k)$.
- The unknown θ_k fully characterise each subject with respect to the study.
- We assume that $\theta_k \mid \gamma \sim Q(\cdot \mid \gamma)$.

Known γ

We only need to condition the distribution of θ_k to the data of the k-th subject:

$$Q(\theta_k \mid x_k, \gamma)$$

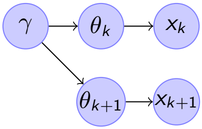


Figure: Graphical model for known case

A general model

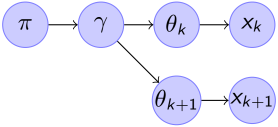


Figure: Graphical model for unknown case

Let $\theta = \theta_1, \ldots, \theta_n$ and $x = x_1, \ldots, x_n$. Our model is as follows:

$$\gamma \sim \pi \tag{3.1}$$

$$\theta_k \mid \gamma \sim Q(\cdot \mid \gamma), \qquad \forall k \in \{1, \ldots, n\} \tag{3.2}$$

$$x_k \mid \theta_k \sim P(\cdot \mid \theta_k), \qquad \forall k \in \{1, \ldots, n\} \tag{3.3}$$

Known gamma: Use Bayes' theorem directly

$$\pi(\theta \mid x, \gamma) = \frac{\pi(x \mid \theta, \gamma)\, \pi(\theta \mid \gamma)}{\pi(x \mid \gamma)}$$

A general model

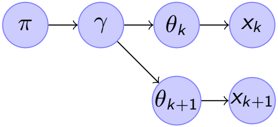


Figure: Graphical model for unknown case

Known gamma: Use Bayes' theorem directly

$$\pi(\theta \mid x, \gamma) = \frac{\pi(x \mid \theta, \gamma)\, \pi(\theta \mid \gamma)}{\pi(x \mid \gamma)}$$

This is fully factorisable:

$$\pi(x \mid \theta, \gamma) = \prod_k P(x_k \mid \theta_k), \quad \pi(\theta \mid \gamma) = \prod_k Q(\theta_k \mid \gamma), \quad \pi(x \mid \gamma) = \prod_k \pi(x_k \mid \gamma).$$

A general model

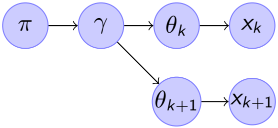


Figure: Graphical model for unknown case

Known gamma: Use Bayes' theorem directly

$$\pi(\theta \mid x, \gamma) = \frac{\pi(x \mid \theta, \gamma)\, \pi(\theta \mid \gamma)}{\pi(x \mid \gamma)} = \prod_k \pi(\theta_k \mid \gamma, x_k).$$

This is fully factorisable:

$$\pi(x \mid \theta, \gamma) = \prod_k P(x_k \mid \theta_k), \quad \pi(\theta \mid \gamma) = \prod_k Q(\theta_k \mid \gamma), \quad \pi(x \mid \gamma) = \prod_k \pi(x_k \mid \gamma).$$

A general model

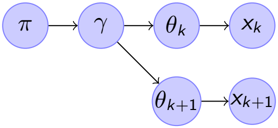


Figure: Graphical model for unknown case

Known gamma: Use Bayes' theorem directly

$$\pi(\theta \mid x, \gamma) = \frac{\pi(x \mid \theta, \gamma)\, \pi(\theta \mid \gamma)}{\pi(x \mid \gamma)} = \prod_k \pi(\theta_k \mid \gamma, x_k).$$

Unknown gamma.

- Empirical Bayes: Find best γ in a restricted class, according to some criterion.
- Hierarchical Bayes: Estimate full joint distribution $\pi(\theta, \gamma \mid x)$.

A test with known γ and a glimpse of Empirical Bayes.

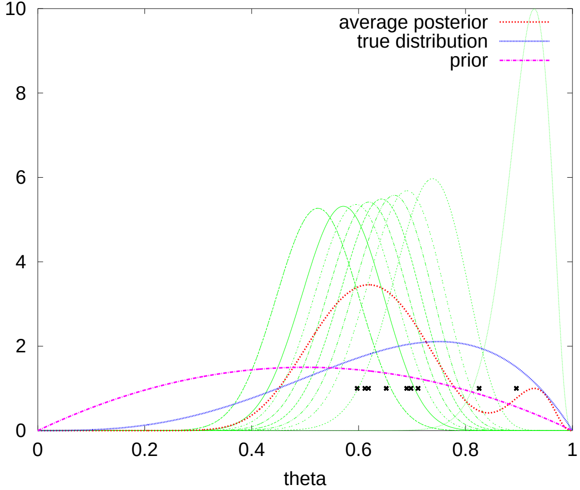


Figure: Samples from 10 subjects, 40 trials each.

Hierarchical Bayes and Gibbs samplers

Theorem

Let a joint distribution $P(x, y)$. The following Markov chain, $Q_t(x, y)$:

$$x^{(t)} \sim P(x \mid y^{(t-1)}), \qquad y^{(t)} \sim P(y \mid x^{(t)})$$

converges to $P(x, y)$, under suitable conditions:

$$\lim_{t \to \infty} \|P(x, y) - Q_t(x, y)\| = 0.$$

Thus, we can estimate $P(x, y)$ by sampling alternately from $P(x \mid y)$ and $P(y \mid x)$.

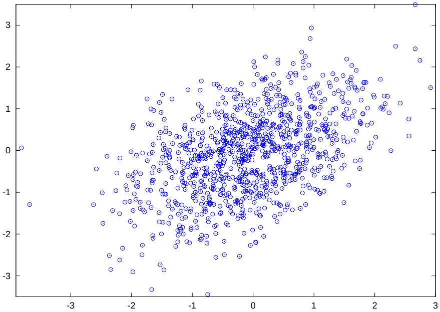
Bi-variate normal density

$$f(x, y; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp(-\frac{1}{2(1-\rho^2)} \cdot (x^2 + y^2 - 2\rho xy))$$

To generate samples from the joint distribution:

$$x^{(t)} \sim \mathcal{N}(\rho y^{(t-1)}, \sqrt{1-\rho^2}) \tag{4.1}$$

$$y^{(t)} \sim \mathcal{N}(\rho x^{(t)}, \sqrt{1-\rho^2}). \tag{4.2}$$



A Gibbs sampler for population data

$$\theta_k^{(t)} \sim \pi(\theta_k \mid \gamma^{(t)}, x_k), \tag{4.3}$$

$$\gamma^{(t+1)} \sim \pi(\gamma \mid \theta^{(t)}). \tag{4.4}$$

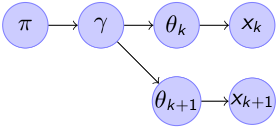


Figure: Graphical population model

A simple example

$$x_k \mid \theta_k \sim \mathcal{B}ern(\cdot \mid \theta_k) \tag{4.5}$$

$$\theta_k \mid \gamma \sim \mathcal{B}eta(\cdot \mid \gamma) \tag{4.6}$$

$$\gamma \sim \mathcal{E}xp(\cdot \mid 1). \tag{4.7}$$

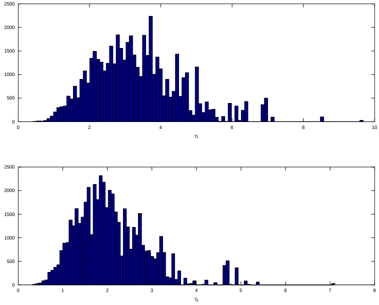


Figure: Samples from $\pi(\gamma \mid x)$ when $\gamma = (4, 2)$.

A simple example

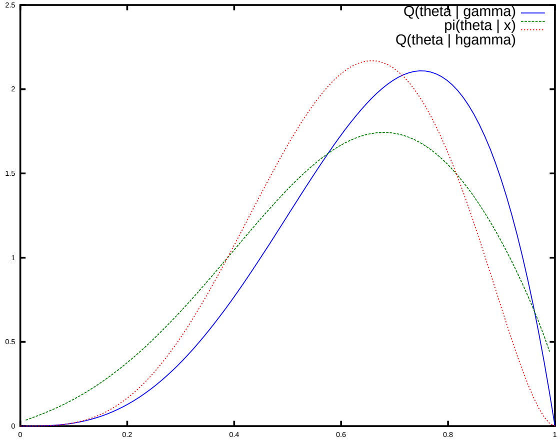


Figure: Marginal posterior $\pi(\theta \mid x)$.

Going further

- Given two groups A, B we can analyse the posteriors $\pi(\gamma \mid x_A), \pi(\gamma \mid x_B)$.
- We can also do a Bayesian hypothesis test:

$$H_0 = \{\gamma_A, \gamma_B : \|\gamma_A - \gamma_B\| \leq \epsilon\}, \qquad H_1 = \{\gamma_A, \gamma_B : \|\gamma_A - \gamma_B\| > \epsilon\}. \tag{4.5}$$

Going further

- Given two groups A, B we can analyse the posteriors $\pi(\gamma \mid x_A), \pi(\gamma \mid x_B)$.
- We can also do a Bayesian hypothesis test:

$$H_0 = \{\gamma_A, \gamma_B : \|\gamma_A - \gamma_B\| \leq \epsilon\}, \qquad H_1 = \{\gamma_A, \gamma_B : \|\gamma_A - \gamma_B\| > \epsilon\}. \tag{4.5}$$

$$\pi(H_0 \mid x) = \frac{\pi(x \mid H_0)\,\pi(H_0)}{\sum_i \pi(x \mid H_i)\,\pi(H_i)}, \tag{4.6}$$

$$\pi(x \mid H_i) = \int_{H_i} \pi(\gamma_A, \gamma_b \mid x_A, x_B)\,\mathrm{d}(\gamma_A, \gamma_B) \tag{4.7}$$

Going further

- Given two groups A, B we can analyse the posteriors $\pi(\gamma \mid x_A), \pi(\gamma \mid x_B)$.
- We can also do a Bayesian hypothesis test:

$$H_0 = \{\gamma_A, \gamma_B : \|\gamma_A - \gamma_B\| \leq \epsilon\}, \qquad H_1 = \{\gamma_A, \gamma_B : \|\gamma_A - \gamma_B\| > \epsilon\}. \tag{4.5}$$

- Any models could be used, depending on the nature of the experimental data.

Further material

Books

- Optimal statistical decisions.
- Bayesian data analysis.
- Statistical decision theory and Bayesian analysis.
- Monte Carlo statistical methods.
- Introducing Monte Carlo methods with R.
- Bayesian computation with R.

Sofware

- R
- BUGS