Bayesian Inference

Sudipto Banerjee

sudiptob@biostat.umn.edu

University of Minnesota

Obtaining Posterior Distributions

- Example: Consider a single data point y from a Normal distribution: $y\ N(y|\theta, \sigma^2)$

$$f(y|\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2\sigma^2}(y-\theta)^2)$$

- Suppose σ is *known*
- Prior on θ: $\theta\ N(\mu, \tau^2)$
- Posterior distribution of θ

$$p(\theta|y) = N\left(\theta|\frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}y, \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau^2}}\right)$$

- Interpret: Posterior mean is a weighted mean of prior mean and data point.
- What is the weight? Think if this makes sense.
- The direct estimate is shrunk towards the prior.
- : What if you had n observations instead of one in the earlier set up?
- Posterior distribution of θ

$$p(\theta|y_1, ..., y_n) = N\left(\theta|\frac{\sigma^2}{\sigma^2 + n\tau^2}\mu + \frac{n\tau^2}{\sigma^2 + n\tau^2}\bar{y}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right)$$

Homework Problem

- Consider the problem of estimating the current weight of a group of people. A sample of 10 people were taken and their average weight was calculated as $\bar{y} = 176$ lbs. Assume that the population standard deviation was known as $\sigma = 3$. Assuming that the data $y_1, ..., y_{10}$ came from a $N(\mu, \sigma^2)$ population perform the following:
- Obtain a 95% confidence interval for μ using classical methods.
- Assume a prior distribution for μ of the form $N(\theta, \tau^2)$. Obtain 95% posterior credible intervals for μ for each of the cases: (a) $\theta = 176, \tau = 8$; (b) $\theta = 176, \tau = 1000$ (c) $\theta = 0, \tau = 1000$. Which case gives results closest to that obtained in the classical method? Note: You may use simulations or closed form analysis.

Homework Problem

- Consider the hypothesis concerning a new physical particle, having prior mean for the mass as $\theta = 82.4$ GeV, and a prior standard deviation of $\tau = 1.1$ GeV. Experiments confirmed the existence of such a particle and yielded an observed mean of 82.1 GeV with standard deviation of 1.7 GeV – based upon a sample of size 7. The physicist wishes to test, on the basis of his prior and the data, whether the mass of this new particle is less than 83 GeV or not. He is willing to assume the following:
- (a) The data comes from a Normal distribution and the prior is also a Normal distribution. (b) The standard deviations are known.
- How would you test the above hypothesis?

A Beta-Binomial model

- Example: Let Y be the number of successes in n independent trials.

$$P(Y = y|\theta) = f(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$$

- Prior on θ: $Beta(\theta|a, b)$
- Prior mean: $\mu = a/(a+b)$; Variance $ab/((a+b)^2(a+b+1))$
- Set $M = (a+b)$ (function of precision).
- Posterior distribution of θ

$$p(\theta|y) = Beta(\theta|a+y, b+n-y)$$

Bayes Factors: A philosophy

- Consider 2 hypotheses: H_0 and H_1. You want to choose one over the other based upon data **y**
- Prior beliefs on the hypotheses: $P(H_0)$ and $P(H_1)$.
- Prior Odds: $P(H_0)/P(H_1)$
- Posterior Odds:

$$\frac{P(H_0|\mathbf{y})}{P(H_1|\mathbf{y})} = \frac{P(H_0)}{P(H_1)} \times \frac{P(\mathbf{y}|H_0)}{P(\mathbf{y}|H_1)}$$

- Posterior Odds = Prior Odds $\times$ *Bayes Factor*
- If prior odds for H_0 and H_1 are same ($P(H_0) = P(H_1) = 0.5$), then Bayes-factor equals Posterior Odds.

Example: Diagnostic testing

- (Spiegelhalter et al. 2004) A new HIV test is claimed to have 95% sensitivity and 98% specificity, and is used in a population with an HIV prevalence of 1/1000. Tabulate expected status of 100,000 individuals in that population who are tested:

	HIV –	HIV +	Marginal
Test –	97,902	5	97,907
Test +	1,998	95	2,093
Marginal	99,900	100	100,000

  Thus of the 2,093 who have test positive, only 95 are truly HIV positive, giving a "predictive value positive" of $95/2093 = 4.5\%$.

- Bayesian language: Let H_0 be hypothesis that individual is truly HIV positive. Let y be observation that an individual tests positive. The disease prevalence is the prior probability $p(H_0) = 0.001$. We are interested in the probability that someone who tests positive is truly HIV positive = posterior probability $P(H_0 \mid y)$. Let H_1 be the hypothesis of truly HIV negative. Then, 95% sensitivity means $P(y \mid H_0) = 0.95$, and 98% specificity means $P(y|H_1) = 0.02$.
  - Prior odds = 1/999; Bayes factor = 0.95/0.02; Posterior odds = 95/1998.
  - These odds correspond to $P(H_0 \mid y) = 95/(95 + 1998) = 0.045$.

Bayes Factor Calibration

Bayes factor range	Strength of evidence in favor of H_0
> 100	Decisive
32 to 100	Very Strong
10 to 32	Strong
3.2 to 10	Substantial
1 to 3.2	Not worth more than a bare mention
	Strength of evidence against H_0
$1/3.2$ to 1	Not worth more than a bare mention
$1/10$ to $1/3.2$	Substantial
$1/32$ to $1/10$	Strong
$1/100$ to $1/32$	Very Strong
$< 1/100$	Decisive

Posterior Predictive Checks

- Assume modelling data as $\mathbf{y} = (y_1, ..., y_n) \sim N(\mu, \sigma^2)$
- Set priors on μ and σ^2
- Run WinBUGS and obtain samples: $\theta_t = \{\mu_t, \sigma_t^2\}$, $t = 1, \ldots, M$
- For each sampled data point θ_t, replicate n data points: $y_{rep,i}^t \sim N(\mu_t, \sigma_t^2)$, $t = 1, \ldots, M$ and $i = 1, \ldots, n$.
- For each sampled value, (μ_t, σ_t^2), we obtain M *replicated data set* $\mathbf{y}_{rep}^t = (y_{rep,1}^t, \ldots, y_{rep,n}^t)$.
- Does our model represent our data adequately? Choose a discrepancy measure, say

$$T(\mathbf{y}; \theta) = \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma^2}$$

Compute $T(\mathbf{y},\ \theta_t)$ and the set of of $T(\mathbf{y}_{rep}^t,\ \theta_t)$ and obtain "Bayesian p-values":

$$P(T(\mathbf{y}_{rep},\ \theta) > T(\mathbf{y},\ \theta) \,|\, \mathbf{y}) = \frac{1}{M} \sum_{t=1}^{M} 1[T(\mathbf{y}_{rep}^t,\ \theta_t) > T(\mathbf{y},\ \theta_t)].$$