MSF100
Statistical Inference Principles - Lectures 7

Rebecka Jörnsten
Mathematical Statistics
University of Gothenburg/Chalmers University of Technology

February 12, 2012

1 Bayesian inference

1.1 Philosophy

Frequentist:

- Parameters are fixed, unknown constants
- It's all about long-run frequency behaviors - what would happen if we could repeat the experiment under the exact same conditions again and again. E.g. a 95% confidence interval covers the true θ in 95% such experiments.

Bayesian:

- probability - belief, not a long-run frequency
- can make probability statements on parameters
- inference based on probability distribution for θ

When to be Bayesian?

- Depends on who you ask! But here's what I think
- when freq methods are too uncertain - need to regularize estimation (e.g. small sample size, complex model)
- prior knowledge exists (e.g. population incidence rates of disease)
- logical to do so - inference that we want is of the form $Pr(\theta|Data)$

1.2 Setup

$L(\theta|\tilde{x}) = \prod_{i=1}^{n} f(x_i|\theta)$ is the likelihood. We assume a prior $\pi(\theta)$ for the parameter. This distribution can also depend on parameters that can be assumed to come from a distribution (hyperprior, hyperparameters).

The posterior distribution is

$$\pi(\theta|\tilde{x}) = \frac{L(\theta|\tilde{x})\pi(\theta)}{\int L(\theta|\tilde{x})\pi(\theta)d\theta}$$

For large n, the likelihood dominates this expression. Frequently, we need only concern ourselves with the numerator above, i.e. the product of the likelihood and the prior. The denominator is normalizing constant.

How do we use this setup? We can construct point estimates from the posterior, frequently just the posterior mean is used

$$\hat{\theta}^B = E(\theta|\tilde{x}) = \int \theta\pi(\theta|\tilde{x})d\theta$$

We can also construct interval estimates from the posterior. Find the shortest interval $[a, b]$ with a probability mass exceeding $1 - \alpha$:

$$\int_a^b \pi(\theta|\tilde{x})d\theta = 1 - \alpha$$

is a $1 - \alpha$ *credible region* for θ.

However, it is really the entire posterior distribution that is the Bayesian result.

Example

- $X_i, i = 1, \cdots, n$ is iid $Be(p)$. Assume prior for p to be $U[0, 1]$
- the posterior is $\pi(p|tix) \propto \pi(p)L(p|\tilde{x}) = 1p^{\sum x_i}(1-p)^{n-\sum x_i}$
- Alternatively, let $\pi(p)$ be $Beta(\alpha, \beta)$
- That is, $\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}$
- The prior mean is $\frac{\alpha}{\alpha+\beta}$
- The posterior is
$$\pi(p|\tilde{x}) = p^{\alpha-1}p^{\sum x}(1-p)^{\beta-1}(1-p)^{n-\sum x}c(\sum x, \alpha, \beta)$$
- That is
$$\pi(p|\tilde{x}) = p^{\alpha+\sum x-1}(1-p)^{n-\sum x+\beta-1}c(\sum x, \alpha, \beta)$$
which is a $Beta(\sum x + \alpha, n - \sum x + \beta)$ distribution (all we need to do is match the normalizing constant)
- The posterior mean is thus
$$\hat{p}^B = \frac{\sum x + \alpha}{n + \alpha + \beta} = \bar{x}(\frac{n}{n + \alpha + \beta}) + (\frac{\alpha}{\alpha + \beta})(1 - \frac{n}{n + \alpha + \beta})$$
That is, a weighted average of the freq estimate $\bar{x}$ and the prior mean. The larger n is, the closer this estimates gets to just the MLE $\bar{x}$. The larger $\alpha + \beta$ is, the closer the estimate is to the prior mean. You can think of $\alpha + \beta$ as representing "extra" data that has mean $\alpha/(\alpha + \beta)$.
- If we choose $\alpha = \beta = 1$ we call this a "flat prior" with prior mean $p = 1/2$. If we choose $\beta > \alpha$, our prior belief is that p is small.

The above is an example of using a *conjugate prior*. If we carefully match the data generative distribution f_θ and the prior π, we obtain a posterior of the same family as the prior. Other examples include: Poisson - Gamma, Normal-Normal, Binomial-Beta.