Simple Univariate Inference for Common Situations

As you have probably seen in your previous classes and in your experience, many data analyses begin with very simple univariate analyses, using models such as the normal (for continuous data), the binomial (for dichotomous data), the Poisson (for count data) and the multinomial (for multicategorical data).

Here we will see how analyses typically proceeds for these simple models from a Bayesian viewpoint.

As usual in Bayesian analyses, aside from a data model (by which I mean the likelihood function), we need a prior distribution over all unknown parameters in the model. Thus, here we consider "standard" likelihood-prior combinations for these simple situations.

To begin, here is a summary chart of what we will see:

Data Type (summary)	Model (likelihood function)	Conjugate Prior	Posterior Density
Continuous $(\overline{x}, n)$	normal(μ, σ^2)	normal(θ, τ^2)	normal$\left(\frac{\frac{\theta}{\tau^2}+\frac{\overline{x}}{\sigma^2/n}}{\frac{1}{\tau^2}+\frac{1}{\sigma^2/n}}, \left[\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right]^{-1}\right)$
Dichotomous (x, n)	binomial(θ, n)	beta(α, β)	beta$(\alpha + x, \beta + (n - x))$
Count (x)	Poisson(λ)	gamma(α, β)	gamma$(\alpha + x, \beta + 1)$
Multicat $(x_1, x_2, \ldots, x_m)$	multinom$(p_1, p_2, \ldots, p_m)$	dirich$(\alpha_1, \alpha_2, \ldots, \alpha_m)$	dirich$(\alpha_1 + x_1, \alpha_2 + x_2, \ldots, \alpha_p + x_m)$

We will now look at each of these four cases in detail.

Bayesian Inference For A Single Normal Mean

EXAMPLE: Consider the situation where we are trying to estimate the mean diastolic blood pressure of Americans living in the United States from a sample of 27 patients. The data are:

76, 71, 82, 63, 76, 64, 64, 74, 70, 64, 75, 81, 75, 78, 66, 62, 79, 82, 78, 62, 72, 83, 79, 41, 80, 77, 67.

[Note: These are in fact real data obtained from an experiment designed to estimate the effects of calcium supplementation on blood pressure. These are the baseline data for 27 subjects from the study, whose reference is: Lyle, R.M., Melby, C.L., Hyner, G.C., Edmonson, J.W., Miller, J.Z., and Weinberger, M.H. (1987). Blood pressure and metabolic effects of calcium supplementation in normotensive white and black men. *Journal of the American Medical Association*, **257**, 1772–1776.]

From this data, we find $\overline{x} = 71.89$, and $s^2 = 85.18$, so that $s = \sqrt{85.18} = 9.22$

Let us assume the following:

1. The standard deviation is known *a priori* to be 9 mm Hg.
2. The observations come from a Normal distribution, i.e.,

$$x_i \sim N(\mu, \sigma^2 = 9^2), \quad \text{for } i = 1, 2, \ldots, 27.$$

We will again follow the three usual steps used in Bayesian analyses:

1. Write down the likelihood function for the data.
2. Write down the prior distribution for the unknown parameter, in this case μ.
3. Use Bayes theorem to derive the posterior distribution. Use this posterior distribution, or summaries of it like 95% credible intervals for statistical inferences.

Step 1: The likelihood function for the data is based on the Normal distribution, i.e.,

$$f(x_1, x_2, \ldots, x_n|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x_i - \mu)^2}{2\sigma^2}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp(-\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}).$$

Step 2: Suppose that we have *a priori* information that the random parameter μ is likely to be in the interval (60,80). That is, we think that the mean diastolic blood pressure should be about 70, but would not be too surprised if it were as low as perhaps 60, or as high as about 80. We will represent this prior distribution as a second Normal distribution (not to be confused with the fact that the data are also assumed to follow a Normal density). The Normal prior density is chosen here for the same reason as the Beta distribution is chosen when we looked at the binomial distribution: it makes the solution of Bayes Theorem very easy. We can therefore approximate our prior knowledge as:

$$\mu \sim N(\theta, \tau^2) = N(70, 5^2 = 25). \tag{1}$$

In general, this choice for a prior is based on any information that may be available at the time of the experiment. In this case, the prior distribution was chosen to have a somewhat large standard deviation ($\tau = 5$) to reflect that we have very little expertise in blood pressures of average Americans. A clinician with experience in this area may elect to choose a much smaller value for τ. The prior is centered around $\mu = 70$, our best guess.

We now wish to combine this prior density with the information in the data to derive the posterior distribution. This combination is again carried out by a version of Bayes Theorem.

$$\text{posterior distribution} = \frac{\text{prior distribution} \times \text{likelihood of the data}}{\text{a normalizing constant}}$$

The precise formula is

$$f(\mu|x_1, \ldots, x_n) = \frac{f(\mu) \times f(x_1, \ldots, x_n|\mu)}{\int_{-\infty}^{+\infty} f(\mu) \times f(x_1, \ldots, x_n|\mu)\; d\mu} \tag{2}$$

In our case, the prior is given by the Normal density discussed above, and the likelihood function was the product of Normal densities given in Step 1.

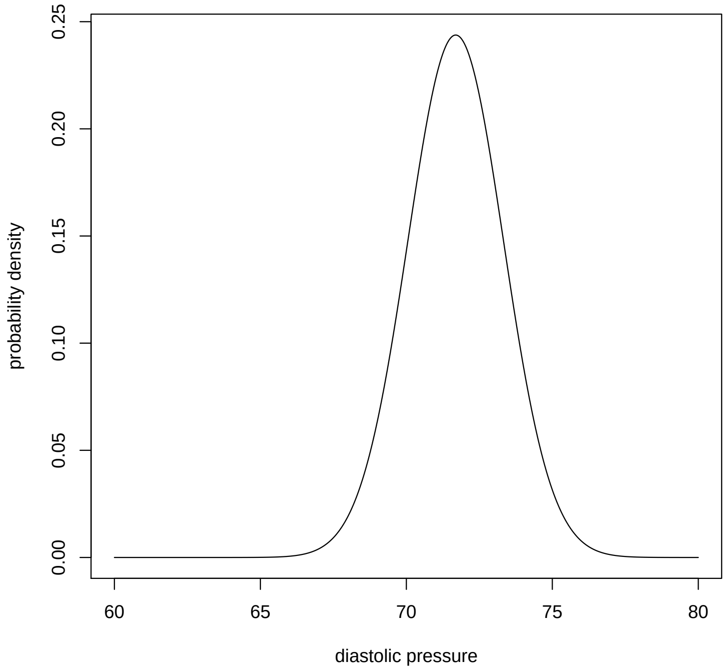
Using Bayes Theorem, we multiply the likelihood by the prior, so that after some algebra, the posterior distribution is given by:

$$\text{Posterior of } \mu \sim N\left(A \times \theta + B \times \overline{x}, \frac{\tau^2\sigma^2}{n\tau^2 + \sigma^2}\right)$$

where

$A = \frac{\sigma^2/n}{\tau^2+\sigma^2/n} = 0.107$
$B = \frac{\tau^2}{\tau^2+\sigma^2/n}$=.893
$n = 27$
$\sigma = 9$
$\tau = \sqrt{25} = 5$
$\theta = 70$, and
$\overline{x} = 71.89$

Hence $\mu \sim N(71.69, 2.68)$, so that graphically, the prior and posterior distributions are:



The mean value depends on both the prior mean, θ, and the observed mean, $\overline{x}$.

Again, the posterior distribution is interpreted as the actual probability density of μ given the prior information and the data, so that we can calculate the probabilities of being in any interval we like. These calculations can be done in the usual way, using normal tables. For example, a 95% credible interval is given by (68.5, 74.9).

Bayesian Inference For Binomial Proportion

Suppose that in a given experiment x successes are observed in N independent Bernoulli trials. Let θ denote the true but unknown probability of success, and suppose that the problem is to find an interval that covers the most likely locations for θ given the data.

The Bayesian solution to this problem follows the usual pattern, as outlines in the previous handout on "Elements of Bayesian Inference". Here we consider only the first five steps, so that we ignore the decision analysis aspects. Hence the steps of interest can be summarized as:

1. Write down the likelihood function for the data.
2. Write down the prior distribution for the data.
3. Use Bayes theorem to derive the posterior distribution. Use this posterior distribution, or summaries of it like 95% credible intervals for statistical inferences.

For the case of a single binomial parameter, these steps are realized by:

1. The likelihood is the usual binomial probability formula, the same one used in the frequentist analysis,

$$L(\theta|x) = Pr\{x \text{ successes in } N \text{ trials}\} = \frac{N!}{(N-x)!\ x!}\theta^x(1-\theta)^{(N-x)}.$$

   In fact, all one needs to specify is that

$$L(\theta|x) = Pr\{x \text{ successes in } N \text{ trials}\} \propto \theta^x(1-\theta)^{(N-x)},$$

   since $\frac{N!}{(N-x)!\ x!}$ is simply a constant that does not involve θ. In other words, inference will be the same whether one uses this constant or ignores it.

2. Although any prior distribution can be used, a convenient prior family is the Beta family, since it is the conjugate prior distribution for a binomial experiment. A random variable, θ, has a distribution that belongs to the Beta family if it has a probability density given by

$$f(\theta) = \begin{cases} \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, & 0 \le \theta \le 1,\ \alpha,\beta > 0, \text{ and} \\ 0, & \text{otherwise}, \end{cases} .$$

[$B(\alpha,\beta)$ represents the Beta function evaluated at (α,β). It is simply the normalizing constant that is necessary to make the density integrate to one, that is, $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$.] The mean of the Beta distribution is given by

$$\mu = \frac{\alpha}{\alpha+\beta},$$

and the standard deviation is given by

$$\sigma = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} \; .$$

Therefore, at this step, one needs only to specify α and β values, which can be done by finding the α and β values that give the correct prior mean and standard deviation values. This involves solving two equations in two unknowns. The solution is

$$\alpha = -\frac{\mu\,(\sigma^2+\mu^2-\mu)}{\sigma^2}$$

and

$$\beta = \frac{(\mu-1)\,(\sigma^2+\mu^2-\mu)}{\sigma^2}$$

3. As always, Bayes Theorem says

   posterior distribution $\propto$ prior distribution $\times$ likelihood function.

   In this case, it can be shown (by relatively simple algebra) that if the prior distribution is $Beta(\alpha,\beta)$, and the data is x successes in N trials, then the posterior distribution is $Beta(\alpha+x, \beta+N-x)$.

Example: Suppose that a new diagnostic test for a certain disease is being investigated. Suppose that 100 persons with confirmed disease are tested, and that 80 of these persons test positively.

(a) What is the posterior distribution of the sensitivity of the test if a Uniform $Beta(\alpha = 1, \beta = 1)$ prior is used? What is the posterior mean and standard deviation of this distribution?

(b) What is the posterior distribution of the sensitivity of the test if a $Beta(\alpha = 27, \beta = 3)$ prior is used? What is the posterior mean and standard deviation of this distribution?

(c) Draw a sketch of the prior and posterior distributions from both (a) and (b).

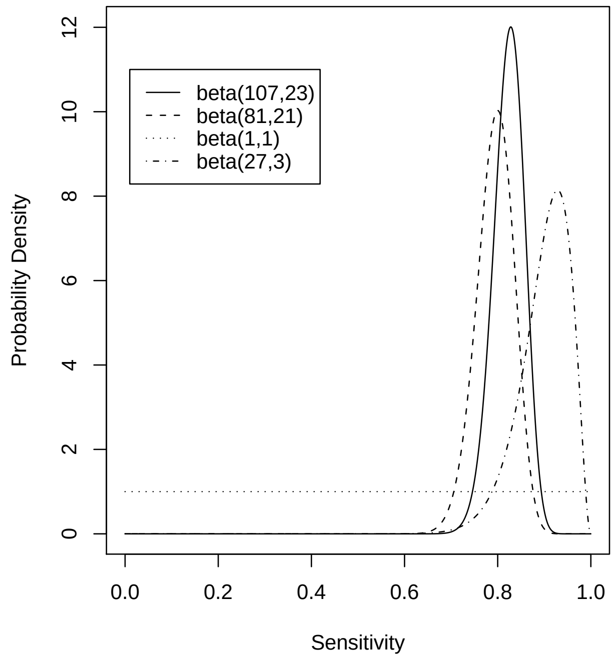
(d) Derive the 95% posterior credible intervals from the two posterior distributions given above, and compare it to the usual frequentist confidence interval for the data. Clearly distinguish the two different interpretations given to confidence intervals and credible intervals.

Solution:

(a) According to the result given above, the posterior distribution is again a Beta, with parameters $\alpha = 1 + 80 = 81$, $\beta = 1 + 20 = 21$. The mean of this distribution is $81/(81 + 21) = 0.794$, and the standard deviation is 0.0398.

(b) Again the posterior distribution is a Beta, with parameters $\alpha = 27 + 80 = 107$, $\beta = 3 + 20 = 23$. The mean of this distribution is $107/(107 + 23) = 0.823$, and the standard deviation is 0.0333.

(c) See Below.



(d) From tables of the beta density (contained in many books of statistical tables) or software that includes Bayesian analysis, the 95% credible intervals are (0.71, 0.86) from the Beta(81,21) posterior density, and (0.75, 0.88) from the Beta(107,23) posterior density. The frequentist 95% confidence interval is (0.71, 0.87).

Note that <u>numerically</u>, the frequentist confidence interval is nearly identical to the Bayesian credible interval starting from a Uniform prior. <u>However</u>, their interpretations are very different. Credible intervals are interpreted <u>directly</u> as the posterior probability that θ is in the interval, given the data and the prior distribution. No references to long run frequencies or other experiments are required. On the other hand, confidence intervals have the interpretation that if such procedures are used repeatedly, then $100(1-\alpha)\%$ of all such sets would in the long run contain the true parameter of interest. Notice that there can be nothing said about what happened in this particular case, the only inference is to the long run. To infer anything about the particular case from a frequentist analysis involves a "leap of faith."

Bayesian Inference For Multinomial Proportions

Recall that the multinomial distribution, given by

$$f(x_1, x_2, \ldots, x_m; p_1, p_2, \ldots, p_m) = \left(\begin{array}{c} n \\ x_1, x_2, \ldots, x_m \end{array} \right) p_1^{x_1} p_2^{x_2} \ldots p_m^{x_m}$$

is simply a multivariate version of the binomial distribution. While the binomial accommodates only two possible outcomes (yes/no, success/failure, male/female, etc.), the multinomial allows for analyses of categorical data with more than two categories (yes/no/maybe, Liberal/Conservative/NDP, items rated on a 1 to 5 scale, etc.).

Similarly, the Dirichlet distribution is simply a multivariate version of the beta density, given by

$$f(p_1, p_2, \ldots, p_m; \alpha_1, \alpha_2, \ldots, \alpha_m) = \frac{\Gamma(\alpha_1 + \alpha_2 + \ldots + \alpha_m)}{\Gamma(\alpha_1)\Gamma(\alpha_2) \times \ldots \times \Gamma(\alpha_m)} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} \times \ldots \times p_m^{\alpha_m - 1}$$

All two-dimensional marginal densities of Dirichlet distributions are beta densities. The mean probability in the i^{th} category (similar again to the beta) is $\frac{\alpha_i}{\sum_{j=1}^{m} \alpha_j}$.

As indicated in the chart, similar to the result for a beta distribution, if we begin with a prior distribution that is Dirichlet, and we have multinomial data, our posterior distribution is also a Dirichlet distribution. As indicated in the chart, and similar to the case of a beta-binomial model, in the Dirichlet-multinomial model, again, we can simply add the prior and data values together, category by category, to derive the final posterior distribution.

Example: Suppose we track slightly overweight Canadians for five years, to see if they remain slightly overweight, become more overweight (obese), or become normal weight (these categories are usually defined by BMI values, i.e., body mass index). Suppose out of 100 slightly overweight persons tracked, 75 remain slightly overweight, 10 become obese, and 15 become normal weight. Suppose you believe that about 50% would stay in their original weight category, and half of the rest would move to each adjacent category. Suppose also that you believe your prior information to be the equivalent of 10 observations. Provide the prior and posterior distributions. Also, provide the marginal distribution for the probability of remaining slightly overweight.

Solution: Prior would be Dirichlet(2.5, 5, 2.5). Posterior would be Dirichlet(2.5 + 15, 5+ 75, 2.5 + 10) = Dirichlet(17.5, 80, 12.5). Marginal posterior for middle category is beta(80, 30).

Bayesian Inference For Poisson Count Data

Recall that the Poisson distribution is given by:

$$p(y|\lambda) = \frac{\lambda^y \exp(-\lambda)}{y!}$$

where y is an observed count, and λ is the rate of the Poisson distribution. Recall that both the mean and variance of the Poisson distribution is given by the rate λ.

Further, recall (see math background section of course) that the Gamma distribution is given by:

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \exp(-\beta\lambda)\lambda^{\alpha-1}, \; for \; \lambda > 0 \; .$$

Although one is a discrete distribution and the other is continuous, note the similarity of form between the Poisson and Gamma distributions: Both have a parameter raised to a power, and both have a simple exponential term. The Gamma can be shown to be a conjugate prior distribution for the Poisson likelihood function.

Suppose count data y follow a Poisson distribution with parameter λ, and the prior distribution put on λ is Gamma(α, β), where α and β are known constants. Then it can be shown that the posterior distribution for λ is again a Gamma density, in particular Gamma$(\alpha + y, \beta + 1)$. The prior parameters α and β can be thought of as having observed a total count of $\alpha - 1$ events in β prior observations.