Bayesian Inference for Normal Mean

Al Nosedal.
University of Toronto.

November 18, 2015

Likelihood of Single Observation

The conditional observation distribution of $y|\mu$ is Normal with mean μ and variance σ^2, which is **known**. Its density is

$$f(y|\mu) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right).$$

Likelihood of Single Observation

The part that doesn't depend on the parameter μ can be absorbed into the proportionality constant. Thus the likelihood shape is given by

$$f(y|\mu) \propto exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right).$$

where y is held constant at the observed value and μ is allowed to vary over all possible values.

Likelihood for a Random Sample of Normal Observations

Usually we have a random sample $y_1, y_2, ..., y_n$ of observations instead of a single observation. The observations in a random sample are all independent of each other, so the joint likelihood of the sample is the product of the individual observation likelihoods. This gives

$$f(y_1, ..., y_n|\mu) = f(y_1|\mu) \times f(y_2|\mu) \times ... \times f(y_n|\mu).$$

We are considering the case where the distribution of each observation $y_j|\mu$ is Normal with mean μ and variance σ^2, which is **known**.

Finding the posterior probabilities analyzing the sample all at once

Each observation is Normal, so it has a Normal likelihood. This gives the joint likelihood

$$f(y_1, ..., y_n|\mu) \propto e^{-\frac{1}{2\sigma^2}(y_1-\mu)^2} \times e^{-\frac{1}{2\sigma^2}(y_2-\mu)^2} \times ...e^{-\frac{1}{2\sigma^2}(y_n-\mu)^2}$$

Finding the posterior probabilities analyzing the sample all at once

After "a little bit" of algebra we get

$$f(y_1, ..., y_n|\mu) \propto e^{-\frac{n}{2\sigma^2}(\mu^2 - 2\mu\bar{y} + \bar{y}^2)} \times e^{-\frac{n}{2\sigma^2}\left(\frac{y_1^2 + ... + y_n^2}{n} - \bar{y}^2\right)}$$

When we absorb the part that doesn't involve μ into the proportionality constant we get

$$f(y_1, ..., y_n|\mu) \propto e^{-\frac{1}{2\sigma^2/n}(\bar{y}-\mu)^2}.$$

We recognize that this likelihood has the shape of a Normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. So the joint likelihood of the random sample is proportional to the likelihood of the sample mean, which is

$$f(\bar{y}|\mu) \propto e^{-\frac{1}{2\sigma^2/n}(\bar{y}-\mu)^2}.$$

Flat Prior Density for μ

The flat prior gives each possible value of μ equal weight. It does not favor any value over any other value, $g(\mu) = 1$. The flat prior is not really a proper prior distribution since $-\infty < \mu < \infty$, so it can't integrate to 1. Nevertheless, this **improper** prior works out all right. Even though the prior is improper, the posterior will integrate to 1, so it is proper.

A single Normal observation y

Let y be a Normally distributed observation with mean μ and known variance σ^2. The likelihood

$$f(y|\mu) \propto e^{-\frac{1}{2\sigma^2}(y-\mu)^2},$$

if we ignore the constant of proportionality.

A single Normal observation y (cont.)

Since the prior always equals 1, the posterior is proportional to this. Rewrite it as

$$g(\mu|y) \propto e^{-\frac{1}{2\sigma^2}(y-\mu)^2}.$$

We recognize from this shape that the posterior is a Normal distribution with mean y and variance σ^2.

Normal Prior Density for μ

The observation y is a random variable taken from a Normal distribution with mean μ and variance σ^2 which is assumed **known**. We have a prior distribution that is Normal with mean m and variance s^2. The shape of the prior density is given by

$$g(\mu) \propto e^{-\frac{1}{2s^2}(\mu-m)^2}.$$

Posterior

The prior times the likelihood is

$$g(\mu) \times f(y|\mu) \propto e^{-\frac{1}{2}\left[\frac{(\mu-m)^2}{s^2}+\frac{(y-\mu)^2}{\sigma^2}\right]}.$$

Posterior (cont.)

After a "little bit" of algebra

$$g(\mu) \times f(y|\mu) \propto exp\left(-\frac{1}{2\sigma^2 s^2/(\sigma^2+s^2)}\left[\mu - \frac{(\sigma^2 m + s^2 y)}{\sigma^2+s^2}\right]^2\right).$$

We recognize from this shape that the posterior is a Normal distribution having mean and variance given by $m^{'} = \frac{(\sigma^2 m + s^2 y)}{\sigma^2+s^2}$ and $(s^{'})^2 = \frac{\sigma^2 s^2}{(\sigma^2+s^2)}$ respectively.

Simple updating rule for Normal family

First we introduce the **precision** of a distribution that is the reciprocal of the variance. The posterior precision

$$\frac{1}{(s')^2} = \left(\frac{\sigma^2 s^2}{(\sigma^2 + s^2)}\right)^{-1} = \frac{(\sigma^2 + s^2)}{\sigma^2 s^2} = \frac{1}{s^2} + \frac{1}{\sigma^2}.$$

Thus the posterior precision equals prior precision plus the observation precision.

Simple updating rule for Normal family (cont.)

The posterior mean is given by

$$m^{'} = \frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2} = \frac{\sigma^2}{\sigma^2 + s^2} \times m + \frac{s^2}{\sigma^2 + s^2} \times y$$

This can be simplified to

$$m^{'} = \frac{1/s^2}{1/\sigma^2 + 1/s^2} \times m + \frac{1/\sigma^2}{1/\sigma^2 + 1/s^2} \times y$$

Thus the posterior mean is the weighted average of the prior mean and the observation, where the weights are the proportions of the precisions to the posterior precision.

Simple updating rule for Normal family (cont.)

This updating rule also holds for the flat prior. The flat prior has infinite variance, so it has zero precision. The posterior precision will equal the prior precision

$$\frac{1}{\sigma^2} = 0 + \frac{1}{\sigma^2},$$

and the posterior variance equals the observation variance σ^2. The flat prior doesn't have a well-defined prior mean. It could be anything. We note that

$$\frac{0}{1/\sigma^2} \times \text{anything} + \frac{1/\sigma^2}{1/\sigma^2} \times y = y,$$

so the posterior mean using flat prior equals the observation y.

A random sample $y_1, y_2, ..., y_n$

A random sample $y_1, y_2, ..., y_n$ is taken from a Normal distribution with mean μ and variance σ^2, which is assumed known. We use the likelihood of the sample mean, $\bar{y}$ which is Normally distributed with mean μ and variance $\frac{\sigma^2}{n}$. The precision of $\bar{y}$ is $\frac{n}{\sigma^2}$.

We have reduced the problem to updating given a single Normal observation of $\bar{y}$. Posterior precision equals the prior precision plus the precision of $\bar{y}$.

$$\frac{1}{(s')^2} = \frac{1}{s^2} + \frac{n}{\sigma^2} = \frac{\sigma^2 + ns^2}{\sigma^2 s^2}.$$

The posterior mean equals the weighted average of the prior mean and $\bar{y}$ where the weights are the proportions of the posterior precision:

$$m' = \frac{1/s^2}{n/\sigma^2 + 1/s^2} \times m + \frac{n/\sigma^2}{n/\sigma^2 + 1/s^2} \times \bar{y}$$

Equivalent Prior Sample Size

A useful check on your prior is to consider the "equivalent sample size". Set your prior variance $s^2 = \frac{\sigma^2}{n_{eq}}$ and solve for n_{eq}. This relates your prior precision to the precision from a sample. Your belief is of equal importance to a sample of size n_{eq}.

Specifying Prior Parameters

We already saw that there were many strategies for picking the parameter values for a beta prior to go with a binomial likelihood. Similar approaches work for specifying the parameters of a normal prior for a normal mean. Often we will have some degree of knowledge about where the normal population is centered, so choosing the mean of the prior distribution for μ usually is less difficult than picking the prior variance (or precision). Workable strategies include:

- Graph normal densities with different variances until you find one that matches your prior information.
- Identify an interval which you believe has 95% probability of trapping the true value of μ, and find the normal density that produces it.
- Quantify your degree of certainty about the value of μ in terms of equivalent prior sample size.

Example

Arnie and Barb are going to estimate the mean length of one-year-old rainbow trout in a stream. Previous studies in other streams have shown the length of yearling rainbow trout to be Normally distributed with known standard deviation of 2 cm. Arnie decides his prior mean is 30 cm. He decides that he doesn't believe it is possible for a yearling rainbow to be less than 18 cm or greater than 42 cm. Thus his prior standard deviation is 4 cm. Thus he will use a Normal(30, 4) prior. Barb doesn't know anything about trout, so she decides to use the "flat" prior.

Example (cont.)

They take a random sample of 12 yearling trout from the stream and find the sample mean $\bar{y} = 32$ cm. Arnie and Barb find their posterior distributions using the simple updating rules for the Normal conjugate family.

Example (cont.)

For Arnie

$$\frac{1}{(s^{'})^2} = \frac{1}{4^2} + \frac{12}{2^2}$$

Solving for this gives his posterior variance $(s^{'})^2 = 0.3265$. His posterior standard deviation is $s^{'} = 0.5714$. His posterior mean is found by

$$m^{'} = \frac{1/4^2}{\frac{1}{4^2} + \frac{12}{2^2}} \times 30 + \frac{12/2^2}{\frac{1}{4^2} + \frac{12}{2^2}} \times 32 = 31.96$$

Example (cont.)

Barb is using the "flat" prior, so her posterior variance is

$$\frac{1}{(s^{'})^2} = \frac{12}{2^2}$$

and her posterior standard deviation is $s^{'} = 0.5774$. Her posterior mean $m^{'} = 32$, the sample mean.
Both Arnie and Barb have Normal posterior distributions.

Using the Posterior Density to Perform Inference

We have already calculated a Bayesian point estimate of μ, the posterior mean.

$$E(\mu|\bar{y}).$$

Bayesian Credible Interval for Normal mean

Known Variance

Using either a "flat" prior, or a Normal(m, s^2) prior, the posterior distribution of μ given $\bar{y}$ is Normal$(m^{'}, (s^{'})^2)$, where we update according to the rules:

1. Precision is the reciprocal of the variance.

2. Posterior precision equals prior precision plus the precision of sample mean.

3. Posterior mean is weighted sum of prior mean and sample mean, where the weights are the proportions of the precisions to the posterior precision.

Bayesian Credible Interval for Normal mean

Our $(1-\alpha)\times 100\%$ Bayesian Credible Interval for μ is

$$m^{'} \pm z_{\alpha/2} \times s^{'},$$

where the z-value is found in the standard Normal table. Since the posterior distribution is Normal and thus symmetric, the credible interval found is the shortest, as well as having equal tail probabilities.

Bayesian Credible Interval for Normal mean

Unknown Variance

If we don't know the variance, we don't know the precision, so we can't use the updating rules directly. The obvious thing to do is to calculate the sample variance

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

from the data. Then we use our equations to find $(s^{'})^2$ and $m^{'}$ where we use the sample variance $\hat{\sigma}^2$ in place of the unknown variance σ^2.

Bayesian Credible Interval for Normal mean (cont.)

Unknown Variance

There is extra uncertainty here, the uncertainty in estimating σ^2. We should widen the credible interval to account for this added uncertainty. We do this by taking the values from the Student's t table instead of the Standard Normal table. The correct Bayesian credible interval is

$$m^{'} \pm t_{\alpha/2} \times s^{'}.$$

The t value is taken from the row labelled $df = n - 1$ (degrees of freedom equals number of observations minus 1)*.

* The resulting Bayesian credible interval is exactly the same one that we would find if we did the full Bayesian analysis with σ^2 as a nuisance parameter, using the joint prior distribution for μ and σ^2 made up of the same prior for $\mu|\sigma^2$ that we used before ("flat" or Normal(m, s^2)) times the prior for σ^2 given by $g(\sigma^2) \propto (\sigma^2)^{-1}$. We would find the joint posterior by Bayes' Theorem. We would find the marginal posterior distribution of μ by marginalizing out σ^2. We would get the same Bayesian credible interval using Student's t critical values.

Example

Arnie and Barb calculated their 95% credible interval from their respective posterior distributions using

$$m^{'} \pm z_{\alpha/2} \times s^{'} .$$

The R Code to find them is shown in the next slide. Arnie and Barb end up with slightly different credible intervals because they started with different prior beliefs. But the effect of the data was much greater than the effect of their priors and their credible intervals are quite similar.

```
# R Code;

qnorm( c(0.025, 0.975), 31.96, 0.5714 );

# Arnie’s CI;

qnorm( c(0.025, 0.975), 32, 0.5774 );

# Barb’s CI;
```

Predictive Density for next observation

Let y_{n+1} be the next random variable drawn after the random sample $y_1, y_2, ..., y_n$. The predictive density of $y_{n+1}|y_1, y_2, ..., y_n$ is the conditional density

$$f(y_{n+1}|y_1, y_2, ..., y_n).$$

Predictive Density for next observation

The conditional distribution we want is found by integrating μ out of the joint posterior distribution.

$$f(y_{n+1}|y_1, y_2, ..., y_n) = \int f(y_{n+1}|\mu) \times g(\mu|y_1, y_2, ..., y_n) d\mu.$$

Predictive Density for next observation

After a "little bit" of calc and algebra, we have that

$$f(y_{n+1}|y_1, y_2, ..., y_n) \propto exp\left[-\frac{1}{2(\sigma^2 + s_n^2)}(y_{n+1} - m_n)^2\right]$$

We recognize this as a Normal density with mean m_n and variance $\sigma^2 + s_n^2$, where m_n and s_n^2 denote the posterior mean and precision (after observing $y_1, y_2, ..., y_n$). Thus, the predictive mean for the observation y_{n+1} is the posterior mean of μ given the observations $y_1, y_2, ..., y_n$. The predictive variance is the observation variance σ^2 plus the posterior variance of μ given the observations $y_1, y_2, ..., y_n$.

Bayesian One-sided Hypothesis Test about μ

The posterior distribution $g(\mu|y_1, ..., y_n)$ summarizes our entire belief about the parameter, after viewing the data. Sometimes we want to answer a specific question about the parameter. This could be: Given the data, can we conclude the parameter μ is greater than μ_0? The answer to the question can be resolved by testing

$$H_0 : \mu \leq \mu_0 \;\; vs \;\; H_1 : \mu > \mu_0.$$

This is an example of a one-sided hypothesis test.

Bayesian One-sided Hypothesis Test about μ

Testing a one-sided hypothesis in Bayesian statistics is done by calculating the posterior probability of the null hypothesis. When the posterior distribution $g(\mu|y_1, y_2, ..., y_n)$ is Normal$(m^{'}, (s^{'})^2)$ this can easily be found from Standard Normal tables.

$$P(H_0 : \mu \leq \mu_0|y_1, ..., y_n) = P\left(\frac{\mu - m^{'}}{s^{'}} \leq \frac{\mu_0 - m^{'}}{s^{'}}\right) = P\left(Z \leq \frac{\mu_0 - m^{'}}{s^{'}}\right)$$

where Z is a Standard Normal random variable.

Example

Arnie and Barb read in a journal that the mean length of yearling rainbow trout in a typical stream habitat is 31 cm. Then each decide to determine if the mean length of trout in the stream they are researching is greater than that by testing

$$H_0 : \mu \leq 31 \quad vs \quad H_1 : \mu > 31.$$

Solution

Arnie and Barb have Normal posteriors, so they use
$P(H_0 : \mu \leq \mu_0 | y_1, ..., y_n) = P\left(Z \leq \frac{\mu_0 - m^{'}}{s^{'}}\right)$

Arnie's Posterior $N(31.96, 0.5714^2)$.
$P(\mu \leq 31 | y_1, y_2, ..., y_n) = P\left(Z \leq \frac{31-31.96}{0.5714}\right) = 0.0465$

Barb's Posterior $N(32, 0.5774^2)$.
$P(\mu \leq 31 | y_1, y_2, ..., y_n) = P\left(Z \leq \frac{31-32}{0.5774}\right) = 0.0416$

```
# R Code;

pnorm(31, 31.96, 0.5714);

# Arnie's posterior probability of H0;


pnorm(31, 32, 0.5774);

# Barb's posterior probability of H0;
```

Bayesian Two-sided Hypothesis Test about μ

Sometimes the question we want to have answered is: Is the mean for the new population μ, the same as the mean for the standard population which we know equals μ_0? A two-sided hypothesis test attempts to answer this question. We set this up as

$$H_0 : \mu = \mu_0 \quad vs \quad H_a : \mu \neq \mu_0.$$

Bayesian Two-Sided Hypothesis Test about μ

If we wish to test the two-sided hypothesis

$$H_0 : \mu = \mu_0 \quad vs \quad H_a : \mu \neq \mu_0.$$

in a Bayesian manner, and we have a continuous prior, we **can't** calculate the posterior probability of the null hypothesis as we did for the one-sided hypothesis. We know that the probability of any specific value of a continuous random variable always equals 0. The posterior probability of the null hypothesis $H_0 : \mu = \mu_0$ will equal zero.

Bayesian Two-Sided Hypothesis Test about μ

Instead, we calculate a $(1-\alpha)\times 100\%$ credible interval for μ using our posterior distribution. If μ_0 lies inside the credible interval, we conclude that μ_0 still has credibility as a possible value. In that case we will not reject the null hypothesis $H_0 : \mu = \mu_0$.