Advanced Methods in Biostatistics II

Lecture 15

December 14, 2017

Bayesian statistics

- In Bayesian analysis, one chooses a density $p(\theta)$, the prior, that expresses our beliefs about θ before we see any data.
- Then one chooses a statistical model $p(\mathbf{y}|\theta)$, the likelihood, that reflects our belief about $\mathbf{y}$ given θ.
- After observing $\mathbf{y}$, we update our beliefs and calculate the posterior distribution $p(\theta|\mathbf{y})$.

Bayesian statistics

- Updating is is done as follows:

$$
\begin{aligned}
p(\text{Param}|\text{Data}) &= \frac{p(\text{Param}, \text{Data})}{p(\text{Data})} \\
&\propto p(\text{Data}|\text{Param}) \times p(\text{Param}) \\
&= \text{Likelihood} \times \text{Prior}.
\end{aligned}
$$

- The posterior distribution is used for subsequent inference.

The Normal model

- Today we turn our focus to the Normal model and discuss how to perform posterior inference on the population mean and variance.
- We begin with the case where the mean μ is the parameter of interest, and the variance σ^2 is known.

The Normal model

- Consider the following model:

$$\begin{aligned} y_1 \ldots y_n \mid \mu &\sim N(\mu, \sigma^2) \\ \mu &\sim N(\mu_0, \sigma_0^2) \end{aligned}$$

- Here μ_0 and σ_0^2 are hyperparameters.
- We know seek to compute $p(\mu|\mathbf{y})$.

The Normal model

- Retaining only terms involving μ we can express the log posterior as follows:

$$
\begin{aligned}
\log(p(\mu \mid \mathbf{y})) &\propto \log(p(\mathbf{y} \mid \mu)) + \log(p(\mu)) \\
&\propto -\frac{1}{2\sigma^2}\sum(y_i - \mu)^2 - \frac{1}{2\tau_0^2}(\mu - \mu_0)^2 \\
&\propto \mu n\bar{y}/\sigma^2 - \mu^2 n/(2\sigma^2) - \mu^2/(2\sigma_0^2) + \mu\mu_0/\sigma_0^2 \\
&= \mu\left(\frac{\bar{y}}{\sigma^2/n} + \frac{\mu_0}{\sigma_0^2}\right) - \frac{\mu^2}{2}\left(\frac{1}{\sigma^2/n} + \frac{1}{\sigma_0^2}\right)
\end{aligned}
$$

The Normal model

- This can be recognized as the log density of a normally distributed random variable with variance

$$Var(\mu \mid \mathbf{y}) = \left(\frac{1}{\sigma^2/n} + \frac{1}{\sigma_0^2} \right)^{-1} = \frac{\sigma_0^2 \sigma^2/n}{\sigma^2/n + \sigma_0^2}$$

and mean

$$E[\mu \mid \mathbf{y}] = \left(\frac{1}{\sigma^2/n} + \frac{1}{\sigma_0^2} \right)^{-1} \left(\frac{\bar{y}}{\sigma^2/n} + \frac{\mu_0}{\sigma_0^2} \right).$$

- The Normal distribution is the conjugate prior.

The Normal model

- Note, we can express the expected value as follows:

$$E[\mu \mid \mathbf{y}] = p\bar{y} + (1 - p)\mu_0$$

  where

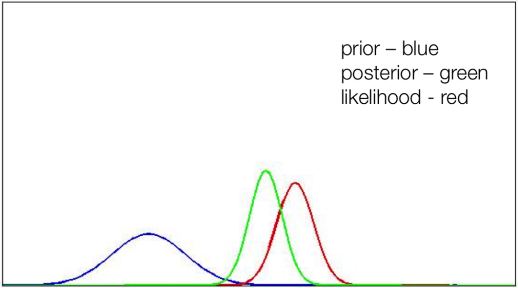
$$p = \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n}.$$

- Thus $E[\mu \mid \mathbf{y}]$ is a mixture of the empirical mean and the prior mean.

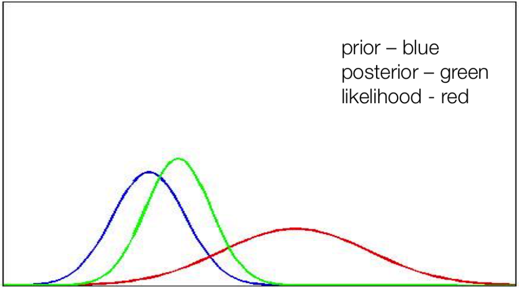
The Normal model

- How much the means are weighted depends on the ratio of the variance of the mean (σ^2/n) and the prior variance (τ_0^2).
- As we collect more data ($n \to \infty$), or if the data is not noisy ($\sigma \to 0$) or we have a lot of prior uncertainty ($\tau_0 \to \infty$) the empirical mean dominates.
- In contrast as we are more certain a priori ($\tau_0 \to 0$) the prior mean dominates.

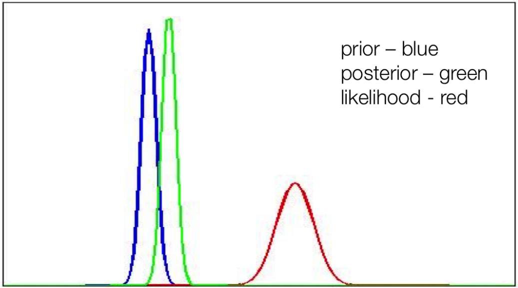
The Normal model



The Normal model



The Normal model



Multi-parameter problems

- Suppose that $\theta = (\theta_1, \ldots \theta_p)$, i.e., we are interested in working with multiple parameters simultaneously.
- The posterior density is now given by

$$p(\theta_1, \ldots \theta_p | y_1, \ldots y_n) \propto p(y_1, \ldots y_n | \theta) p(\theta)$$

- To perform inference on a single parameter we seek the marginal posterior for the parameter of interest.

Multi-parameter problems

- Theoretically, we can compute the marginal posterior as follows:

$$p(\theta_1|y_1,\ldots y_n) = \int \cdots \int p(\theta_1,\ldots\theta_p|y_1,\ldots y_n)d\theta_2\ldots d\theta_p$$

- In practice, it may not be feasible to compute this integral.
- In these cases using Monte Carlo methods is an option.

Two parameter Normal model

- Revisiting the Normal model, in most situations we do not know μ or σ^2.
- Thus, we need to expand our model to allow for joint inference for the mean and variance.
- In the two parameter setting, the form of the joint posterior for μ and σ^2 would look as follows:

$$p(\mu, \sigma^2 | y_1, \ldots y_n) \propto p(y_1, \ldots y_n | \mu, \sigma^2)\ p(\mu, \sigma^2).$$

Two parameter Normal model

- We have now explicitly noted that σ^2 is also an unknown quantity, by including it in the prior distribution.
- Therefore, we now need to specify a joint prior for both μ and σ^2, and not just a prior for μ.

Two parameter Normal model

- Let us begin by defining $\tau = 1/\sigma^2$, which we will work with instead of σ^2 to simplify calculations.
- We can now write the joint prior as follows:

$$p(\mu, \tau) = p(\mu|\tau)p(\tau)$$

- A reasonable prior for μ given τ is from the Normal family.

The Normal model - fixed mean, random variance

- As homework show that the gamma distribution is a conjugate prior for the precision when the mean is known:

$$y_1, \ldots, y_n | \, \mu, \tau \sim N(\mu, 1/\tau)$$

$$\tau \sim \text{Gamma}(\alpha, \beta),$$

then

$$\tau | \, y_1, \ldots, y_n \sim \text{Gamma}\left(\alpha + n/2, \; \beta + \frac{1}{2}\sum (y_i - \mu)^2\right)$$

Two parameter Normal model

- Let us consider the following model:

$$\begin{aligned} y_1, \ldots y_n \mid \mu, \tau &\sim N(\mu, 1/\tau) \\ \mu \mid \tau &\sim N(\mu_0, 1/(\kappa_0 \tau)) \\ \tau &\sim \text{gamma}(\nu_0/2, \nu_0 \sigma_0^2/2) \end{aligned}$$

- Here the joint prior is referred to as normal-gamma distribution, i.e.

$$\begin{aligned} p(\mu, \tau) &\sim NG\left(\mu_0, \kappa_0, \nu_0, \sigma_0^2\right) \\ &\equiv N(\mu_0, 1/(\kappa_0 \tau)) \times \text{gamma}(\nu_0/2, \nu_0 \sigma_0^2/2) \end{aligned}$$

- There are four hyperparameters: μ_0, κ_0, ν_0, and σ_0^2.

Two parameter Normal model

- It turns out this is a conjugate prior for the two parameter Normal model.
- The joint posterior can be expressed as follows:

$$p(\mu, \tau \mid y_1, \ldots y_n) \quad \sim \quad NG\left(\mu, \tau \mid \mu_n, \kappa_n, \nu_n, \sigma_n^2\right)$$

where

$$\begin{aligned}
\mu_n &= \frac{\kappa_0 \mu_0 + n\bar{y}}{\kappa_0 + n} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\nu_n \sigma_n^2 &= \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2
\end{aligned}$$

Two parameter Normal model

- Note, the term μ_n corresponds to a weighted average as before.
- The term $\nu_n \sigma_n^2$ combines the sample sum of squares and the prior sum of squares, as well as an additional uncertainty term that arises due to the difference between the sample and prior mean.

Marginal posterior

- We would like to make inferences about the marginal distributions $p(\mu|y)$ and $p(\sigma^2|y)$ rather than the joint distribution $p(\mu, \sigma^2|y)$.

- The marginal posterior for τ is given by

$$\tau \mid y_1, \ldots y_n \sim \text{Gamma}(\nu_n/2, \nu_n\sigma_n^2/2).$$

- The marginal posterior for μ is given by

$$\mu \mid y_1, \ldots y_n \sim t_{\nu_n}(\mu_n, \sigma_n^2/\kappa_n).$$

  which is a kernel of a t-distribution with ν_n degrees of freedom, centered at μ_n and with scale parameter σ_n^2/κ_n.

Bayesian Linear Models

- Finally, let us revisit our standard linear model, where

$$\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}).$$

- There are many ways to formulate priors for this model.
- Consider the following three common prior specifications:

1. $\boldsymbol{\beta} \mid \sigma^2 \sim N(\boldsymbol{\beta}_0, \sigma^2\Sigma_0)$ and $\sigma^{-2} \sim \text{Gamma}(\alpha_0, \tau_0^{-1})$.
2. $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \Sigma_0)$ and $\sigma^{-2} \sim \text{Gamma}(\alpha_0, \tau_0^{-1})$.
3. $(\boldsymbol{\beta}, \sigma^2) \sim \sigma^{-2}$.

Bayesian Linear Models

- The distinction between the first case and the second is the inclusion of σ^2 in the prior specification for $\boldsymbol{\beta}$.
- This is useful for making all posterior distributions tractable, including that of $\boldsymbol{\beta}$ integrated over σ^2.
- However, it may or may not reflect the desired prior distribution.

Bayesian Linear Models

- The second specification has tractable full conditionals.
- That is, we can easily figure out $\boldsymbol{\beta} \mid \sigma^2, \mathbf{y}, \mathbf{X}$ and $\sigma^2 \mid \boldsymbol{\beta}, \mathbf{y}, \mathbf{X}$.
- However, the posterior marginals of the parameters ($\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}$ in particular) are not tractable.
- This posterior is often explored using Monte Carlo.

Bayesian Linear Models

- The final prior specification is not a proper density, as it doesn't have a defined integral for the elements of β from $-\infty$ to ∞ and for $0 \leq \sigma^2 < \infty$.
- However, proceeding as if it were a proper density yields a proper distribution for the posterior.
- Such "improper" priors are often used to specify putatively uninformative distributions that yield valid posteriors.
- In this case, the posterior has the property of the posterior mode being centered around $\hat{\beta}$.