MATH 640: Bayesian Statistics

Board Examples

A Part I Examples

Example I.1, Slide 26

An insurance company believes that people can be divided into two classes: those who are accident prone and those who are not. The company's statistics show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability 0.4, whereas the probability decreases to 0.2 for a person who not accident prone. Assume that 30% of population is accident prone. Suppose a new policyholder has an accident within a year of purchasing a policy. What is probability that he or she is accident prone?

Solution: Let A denote the event a policy-holder will have an accident within the first year of purchasing the policy and let R denote the event that the policy holder is accident prone. From the information above, we can determine that

$$P(A|R) = 0.4, P(R) = 0.3, P(A|R^c) = 0.2 \text{ and } P(R^c) = 0.7$$

We are interested in determining $P(R|A)$. Via Bayes' Rule, we have

$$P(R|A) = \frac{P(A|R)P(R)}{P(A|R)P(R) + P(A|R^c)P(R^c)} = \frac{(0.4)(0.3)}{(0.4)(0.3) + (0.2)(0.7)} \approx 0.4615$$

Example I.2, Slide 30

Suppose we wish to study the number of ponderosa pine trees in the Black Hills of South Dakota that are infected with the mountain pine beetle (a beetle that ultimately kills ponderosa pines). Let X be the number of infected ponderosa pines in a 10 square mile plot of forested land in the Black Hills. Further, let N be the number of ponderosa pines in that plot and p be the known probability that a randomly selected pine is infected. Finally, note that $X|N \sim Binom(N, p)$ and $N \sim Pois(\Lambda)$. Determine the following:

1. $E(X)$
2. $Var(X)$

Solution:

$$E(X) = E[E(X|N)] = E(Np) = pE(N) = p\Lambda$$

$$\begin{aligned} Var(X) &= E[Var(X|N)] + Var[E(X|N)] = E[Np(1-p)] + Var(Np) \\ &= p(1-p)E(N) + p^2 Var(N) = p(1-p)\Lambda + p^2\Lambda \end{aligned}$$

Example I.3, Slide 35

Suppose $U \sim U(0,1)$ and let $W = -\frac{1}{\lambda}\log(1-U)$. What distribution does W have?
Solution: Begin with the probability of W,

$$\begin{aligned} P(W \le w) &= P[-(1/\lambda)\log(1-U) \le w] = P[\log(1-U) \ge -\lambda w] \\ &= P(1 - U \ge e^{-\lambda w}) = P(-U \ge e^{-\lambda w} - 1) \\ &= P(U \le 1 - e^{-\lambda w}) \end{aligned}$$

But this is the CDF of a standard uniform, so

$$P(U \le 1 - e^{-\lambda w}) = 1 - e^{-\lambda w}$$

Differentiating, we get the density is $\lambda e^{-\lambda w}$, thus $W \sim Exp(\lambda)$

Example I.4, Slide 47

Let $X \sim N(\mu, \sigma^2)$. Show that the family of pdfs where $\boldsymbol{\theta} = (\mu, \sigma), -\infty < \mu < \infty, \sigma > 0$ is an exponential family.
Solution: First note the density:

$$\begin{aligned} p(x|\mu, \sigma^2) &= (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \\ &= (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(x^2 - 2x\mu + \mu^2)\right] \\ &= (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left[-\frac{1}{2\sigma^2}(x^2 - 2x\mu)\right] \\ &= (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left[-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x)\right] \end{aligned}$$

Then, $h(x) = 1$, $C(\boldsymbol{\theta}) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{\mu^2}{2\sigma^2}\right)$, $w_1(\boldsymbol{\theta}) = \frac{1}{\sigma^2}$, $w_2(\boldsymbol{\theta}) = \frac{\mu}{\sigma^2}$, $t_1(x) = -\frac{x^2}{2}$, and $t_2(x) = x$.

Example I.5, Slide 49

Let $\theta \sim Gamma(\alpha, \beta)$. Using Kernel recognition, show that $E(\theta) = \alpha/\beta$.
Solution: Note that the density of θ is $p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$. Begin with the definition of the expectation,

$$\begin{aligned} E(\theta) &= \int_0^\infty \theta \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}d\theta \\ &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha+1-1}e^{-\beta\theta}d\theta \end{aligned}$$

Now note that $\theta^{\alpha+1-1}e^{-\beta\theta}$ is the kernel of a $Gamma(\alpha+1, \beta)$ distribution. Rearranging,

$$\begin{aligned} E(\theta) &= \frac{1}{\beta\Gamma(\alpha)} \int_0^\infty \beta^{\alpha+1}\theta^{\alpha+1-1}e^{-\beta\theta}d\theta \\ &= \frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)} \int_0^\infty \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{\alpha+1-1}e^{-\beta\theta}d\theta \end{aligned}$$

The integrand is now the pdf of a $Gamma(\alpha+1, \beta)$ distribution integrated over its support and is thus 1, leaving

$$E(\theta) = \frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)} = \frac{\alpha\Gamma(\alpha)}{\beta\Gamma(\alpha)} = \frac{\alpha}{\beta}$$

recalling the property of the Gamma function that $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$.

Example I.6, Slide 52

Let $X_1, \ldots, X_n \sim Pois(\lambda)$. Determine the likelihood function. If, instead, $X_1, \ldots, X_n \sim Pois(\lambda_i)$, what would the likelihood look like?
Solution: Note the mass function of X_i is $p(x_i) = \frac{\lambda^{x_i}e^{-\lambda}}{x_i!}$. The likelihood is then

$$\mathcal{L}(X_1, \ldots, X_n|\lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}e^{-\lambda}}{x_i!} = \left(\prod_{i=1}^n \frac{1}{x_i!}\right)\lambda^{\sum x_i}e^{-n\lambda}$$

If, instead, $X_1, \ldots, X_n \sim Pois(\lambda_i)$, the likelihood would look like

$$\mathcal{L}(X_1, \ldots, X_n|\lambda) = \prod_{i=1}^n \frac{\lambda_i^{x_i}e^{-\lambda_i}}{x_i!}$$

which cannot be simplified.

Example I.9, Slide 95

Suppose we observe y, a single data point, from a normal distribution parameterized by a mean θ and variance σ^2. Further assume that (however unlikely) that σ^2 is known and therefore fixed. Let's build a basic model for θ using a conjugate prior.
First note the form of the likelihood:

$$\begin{aligned}\mathcal{L}(\theta) &= (2\sigma^2\pi)^{-1/2}\exp\left[-\frac{1}{2\sigma^2}(y-\theta)^2\right]\\ &\propto \exp\left[-\frac{1}{2\sigma^2}(y-\theta)^2\right]\end{aligned}$$

The kernel of a Normal distribution. Thus, via conjugacy, if we take our prior on θ to be normal, we will have a normal posterior. Our prior is then

$$\begin{aligned}\pi(\theta) &= (2\pi\tau_0^2)^{-1/2}\exp\left[-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right]\\ &\propto \exp\left[-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right]\end{aligned}$$

Therefore our posterior is

$$\begin{aligned}p(\theta|y) &\propto \exp\left[-\frac{1}{2\sigma^2}(y-\theta)^2\right]\exp\left[-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right]\\ &= \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}(y-\theta)^2+\frac{1}{\tau_0^2}(\theta-\mu_0)^2\right)\right]\end{aligned}$$

we are only concerned with finding the kernel of the density, so we can expand this out and rearrange terms:

$$\begin{aligned}&= \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}(y^2-2\theta y+\theta^2)+\frac{1}{\tau_0^2}(\theta^2-2\theta\mu_0+\mu_0^2)\right)\right]\\ &= \exp\left[-\frac{1}{2}\left(\theta^2\left\{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}\right\}-2\theta\left\{\frac{\mu_0}{\tau_0^2}+\frac{y}{\sigma^2}\right\}+\frac{\mu_0^2}{\tau_0^2}+\frac{y^2}{\sigma^2}\right)\right]\\ &= \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}\right)\left(\theta^2-2\theta\left\{\frac{\frac{\mu_0}{\tau_0^2}+\frac{y}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}\right\}+\frac{\frac{\mu_0^2}{\tau_0^2}+\frac{y^2}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}\right)\right]\end{aligned}$$

We then complete the square to obtain:

$$\begin{aligned}\mathcal{L}(\theta|X)p(\theta) &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}\right)\left(\left\{\theta-\frac{\frac{\mu_0}{\tau_0^2}+\frac{y}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}\right\}^2+\frac{\frac{\mu_0^2}{\tau_0^2}+\frac{y^2}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}-\left\{\frac{\frac{\mu_0}{\tau_0^2}+\frac{y}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}\right\}^2\right)\right]\\ &= \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}\right)\left(\theta-\frac{\frac{\mu_0}{\tau_0^2}+\frac{y}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}\right)^2\right]\exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}\right)\left(\frac{\frac{\mu_0^2}{\tau_0^2}+\frac{y^2}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}-\left\{\frac{\frac{\mu_0}{\tau_0^2}+\frac{y}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}\right\}^2\right)\right]\end{aligned}$$

Note that the term in the second exponential is not dependent upon θ. In fact, it is only a function of the data and the hyperparameters from the prior distribution. We can consider this part of the integrating constant since, given the data and the hyperparameters it is fixed. Denote this second term as $h(X)$. Then,

$$\begin{aligned}p(\theta|X) &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}\right)\left(\theta-\frac{\frac{\mu_0}{\tau_0^2}+\frac{y}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}\right)^2\right]h(X)\\ &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}\right)\left(\theta-\frac{\frac{\mu_0}{\tau_0^2}+\frac{y}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}\right)^2\right]\end{aligned}$$

which we recognize as the kernel of a $N(\mu_1, \tau_1^2)$ distribution where

$$\mu_1 = \frac{\frac{\mu_0}{\tau_0^2}+\frac{y}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}} \text{ and } \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2}+\frac{1}{\sigma^2}$$

Example I.10, Slide 99

Suppose instead we observe y_i, $i = 1, \ldots, n$, from a normal distribution parameterized by a known mean θ and unknown variance σ^2. In this case, we can consider θ. Once again, let's build a basic model for θ using a conjugate prior.
The likelihood is then

$$\begin{aligned}\mathcal{L}(y|\theta) &= (2\sigma^2\pi)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right] \\ &\propto (\sigma^2)^{-n/2} \exp\left[-\frac{nv}{2\sigma^2}\right]\end{aligned}$$

where $v = \frac{1}{n}\sum_{i=1}^{n}(y_i-\theta)^2$, a sufficient statistic. Notice that this is the kernel of a scaled inverse-χ^2. Thus we take this distribution as the prior for σ^2 and say that $\sigma^2 \sim Inv\text{-}\chi^2(\nu_0, s_0^2)$ giving the following prior density:

$$\pi(\sigma^2) \propto \left(\frac{s_0^2}{\sigma^2}\right)^{\nu_0/2+1} \exp\left(-\frac{\nu_0 s_0^2}{2\sigma^2}\right)$$

Combining with the likelihood, we get the posterior:

$$\begin{aligned}p(\sigma^2|y) &\propto \left(\frac{s_0^2}{\sigma^2}\right)^{\nu_0/2+1} \exp\left(-\frac{\nu_0 s_0^2}{2\sigma^2}\right)(\sigma^2)^{-n/2}\exp\left[-\frac{n}{2}\frac{v}{\sigma^2}\right] \\ &\propto (\sigma^2)^{-((n+\nu_0)/2+1)} \exp\left[-\frac{1}{2\sigma^2}(\nu_0 s_0^2 + nv)\right] \\ &\propto (\sigma^2)^{-((n+\nu_0)/2+1)} \exp\left[-\frac{(n+\nu_0)}{2\sigma^2}\frac{(\nu_0 s_0^2+nv)}{(n+\nu_0)}\right]\end{aligned}$$

Thus, in terms of the scaled inverse-chi-square described on slide 95, we take $\nu_1 = n + \nu_0$ and $s_1^2 = \frac{(\nu_0 s_0^2+nv)}{(n+\nu_0)}$. Thus,

$$p(\sigma^2|y) \propto (\sigma^2)^{-(\nu_1/2+1)} \exp\left[-\frac{\nu_1 s_1^2}{2\sigma^2}\right]$$

and we can see that $\sigma^2|y \sim Inv\text{-}\chi^2(\nu_1, s_1^2)$ or $\sigma^2|y \sim IG\left(\frac{\nu_1}{2}, \frac{\nu_1}{2}s_1^2\right)$

Example I.12, Slide 104

Consider again the example where y_i, $i = 1, \ldots, n$ is Normal with known mean θ and unknown mean σ^2. Suppose we take the prior $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$.

1. Is the prior proper or improper?
   The prior is improper as it has an infinite integral over the range $(0, \infty)$.

2. Will it lead to a proper posterior?
   To find out, first consider the likelihood

   $$\begin{aligned}\mathcal{L}(y|\theta) &= (2\sigma^2\pi)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right] \\ &\propto (\sigma^2)^{-n/2} \exp\left[-\frac{nv}{2\sigma^2}\right]\end{aligned}$$

   Using the prior in the prompt, we know that the posterior has the following form:

   $$\begin{aligned}p(\sigma^2|y) &\propto (\sigma^2)^{-1}(\sigma^2)^{-n/2} \exp\left[-\frac{nv}{2\sigma^2}\right] \\ &\propto (\sigma^2)^{-(n/2+1)} \exp\left[-\frac{nv}{2\sigma^2}\right]\end{aligned}$$

   which we recognize as being an scaled inverse-chi-square with $\nu = n$ and $s^2 = v$ or inverse-gamma with $\alpha = \frac{n}{2}$ and $\beta = \frac{n}{2}v$. In other words, it has the effect of using the conjugate prior from Example 1.10 with $\nu_0 = 0$.

Example I.13, Slide 110

Consider again our investigation of the infection of ponderosa pines in the Black Hills of South Dakota with the mountain pine beetle. As before, we assume $X|\theta \sim Bin(n, \theta)$. Determine a noninformative prior for θ using Jeffreys' prior.
Note first the form of the likelihood:

$$\mathcal{L}(X|\theta) = \binom{n}{X} \theta^X (1-\theta)^{n-X}$$

The log-likelihood is then

$$\ell(X|\theta) = \log \binom{n}{X} + X \log(\theta) + (n-X) \log(1-\theta)$$

Differentiating, we get

$$\frac{\partial}{\partial \theta} \ell(X|\theta) = 0 + \frac{X}{\theta} - \frac{(n-X)}{1-\theta}$$

Differentiating for a second time gives

$$\frac{\partial^2}{\partial \theta^2} \ell(X|\theta) = -\frac{X}{\theta^2} - \frac{(n-X)}{(1-\theta)^2}$$

Now we take the negative expectation (i.e. finding the Fisher information):

$$\begin{aligned} -E\left[\frac{\partial^2}{\partial \theta^2} \ell(X|\theta)\right] &= -E\left[-\frac{X}{\theta^2} - \frac{(n-X)}{(1-\theta)^2}\right] = \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} \\ &= \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n(1-\theta)+n\theta}{\theta(1-\theta)} \\ &= \frac{n}{\theta(1-\theta)} \end{aligned}$$

Thus, Jeffreys' prior has the form $[J(\theta)]^{1/2}$, thus a noninformative prior for θ is

$$\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}.$$

We can re-express this prior as

$$\pi(\theta) \propto \theta^{\frac{1}{2}-1}(1-\theta)^{\frac{1}{2}-1}$$

which we should recognize as the kernel of a Beta distribution with parameters $\alpha = 1/2$ and $\beta = 1/2$.

Example I.14, Slide 111

Now find the posterior using Jeffreys' prior for θ given that $X|\theta \sim Bin(n, \theta)$. Compare the theoretical results of this posterior to what we found in before.
First, find the posterior that results by taking the prior on θ to be $Beta(1/2, 1/2)$. Multiplying the likelihood and the prior gives us:

$$\begin{aligned} p(\theta|X) &\propto \theta^X (1-\theta)^{n-X} \theta^{\frac{1}{2}-1}(1-\theta)^{\frac{1}{2}-1} \\ &\propto \theta^{X+\frac{1}{2}-1}(1-\theta)^{n-X+\frac{1}{2}-1} \end{aligned}$$

Note that this is the kernel of a Beta distribution with parameters $\alpha = X + \frac{1}{2}$ and $\beta = n - X + \frac{1}{2}$. When we assumed $\theta \sim U(0,1)$, the resulting posterior was $Beta(X+1, n-X+1)$. When using Jeffreys' prior, the posterior mean, variance, and mode are

$$E(\theta|X) = \frac{X+\frac{1}{2}}{n+1}, Var(\theta|X) = \frac{(X+\frac{1}{2})(n-X+\frac{1}{2})}{(n+2)^2(n+2)}, \text{ and } Mode(\theta) = \frac{X-\frac{1}{2}}{n-1}$$

Example I.17, Slides 124-133

Suppose y_i, $i = 1, \ldots, n$ are iid $N(\mu, \sigma^2)$ where both parameters are unknown. The model we wish to estimate is then

$$p(\mu, \sigma^2 | y) \propto \mathcal{L}(y|\mu, \sigma^2)\pi(\mu, \sigma^2)$$

Suppose we select the noninformative prior

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$$

What does our joint posterior look like?
Note that the likelihood is

$$\mathcal{L}(y|\mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

Thus, the posterior is

$$\begin{aligned} p(\mu, \sigma^2|y) &\propto (\sigma^2)^{-n/2-1} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right] \\ &= (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right] \end{aligned}$$

Recall that to use the approach of averaging over nuisance parameters, we need the posterior to factor into the form:

$$p(\theta|\gamma, X)p(\gamma|X)$$

In the context of this example, we're looking to factor the posterior like so:

$$p(\mu|\sigma^2, y)p(\sigma^2|y)$$

How do we accomplish this?
First, we must break apart the summation so that we have a term that is dependent on μ as well as one that is *not* dependent on μ. One way to do this is to add and subtract $\bar{y}$, a sufficient statistic for μ, to the inside of the squared term in the summation:

$$p(\mu, \sigma^2|y) \propto (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \bar{y} + \bar{y} - \mu)^2\right]$$

We now expand the quadratic treating $(y_i - \bar{y})$ and $(\bar{y} - \mu)$ as our terms in the FOIL:

$$p(\mu, \sigma^2|y) \propto (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left\{(y_i - \bar{y})^2 + 2(y_i - \bar{y})(\bar{y} - \mu) + (\bar{y} - \mu)^2\right\}\right]$$

Next, distribute the summation:

$$p(\mu, \sigma^2|y) \propto (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n}(y_i - \bar{y})^2 + \sum_{i=1}^{n}2(y_i - \bar{y})(\bar{y} - \mu) + \sum_{i=1}^{n}(\bar{y} - \mu)^2\right\}\right]$$

The first term cannot be reduced, but not the second term:

$$\sum_{i=1}^{n}2(y_i - \bar{y})(\bar{y} - \mu) = 2(\bar{y} - \mu)\sum_{i=1}^{n}(y_i - \bar{y}) = 2(\bar{y} - \mu)\left(\sum_{i=1}^{n}y_i - \sum_{i=1}^{n}\bar{y}\right) = 2(\bar{y} - \mu)(n\bar{y} - n\bar{y}) = 0$$

Also note the third term is not dependent on i, thus

$$\sum_{i=1}^{n}(\bar{y} - \mu)^2 = n(\bar{y} - \mu)^2$$

Thus the form of the posterior can be simplified to:

$$p(\mu, \sigma^2|y) \propto (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n}(y_i-\bar{y})^2 + n(\bar{y}-\mu)^2\right\}\right]$$

Note that a sufficient statistic for σ^2 is the sample variance, i.e.

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

By substituting $(n-1)s^2$ for the summation term, we have the most simplified form of the posterior:

$$p(\mu, \sigma^2|y) \propto (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\bar{y}-\mu)^2\right\}\right]$$

Further note that we can now factor the posterior into

$$p(\mu, \sigma^2|y) \propto (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \exp\left[-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right]$$

Conditional Posterior of μ

1 Which term or terms represent the conditional posterior of μ given σ^2 and y?

The conditional posterior distribution is

$$\begin{aligned} p(\mu|\sigma^2, y) &\propto \exp\left[-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right] \\ &= \exp\left[-\frac{1}{2(\sigma^2/n)}(\mu-\bar{y})^2\right] \end{aligned}$$

which is the kernel of a normal distribution with mean $\bar{y}$ and variance σ^2/n. Thus, $\mu|\sigma^2, y \sim N(\bar{y}, \sigma^2/n)$.

Marginal Posterior of σ^2

2 Which term or terms are need to find the marginal posterior of σ^2 given y?

All of them are needed. To find the marginal posterior distribution, we must average the joint distribution over μ:

$$p(\sigma^2|y) \propto \int (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \exp\left[-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right] d\mu$$

Note that, with respect to μ, the first two terms are constants and can thus be pulled out of the integrand:

$$p(\sigma^2|y) \propto (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \int \exp\left[-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right] d\mu$$

Next note that the remaining integral is integrating over the kernel of a normal. By multiplying and dividing by the appropriate integrating constant, we can evaluate the integral:

$$\begin{aligned} \int \exp\left[\frac{1}{2(\sigma^2/n)}(\mu-\bar{y})^2\right] d\mu &= \int \frac{\sqrt{2\pi\sigma^2/n}}{\sqrt{2\pi\sigma^2/n}} \exp\left[-\frac{1}{2(\sigma^2/n)}(\mu-\bar{y})^2\right] d\mu \\ &= \sqrt{2\pi\sigma^2/n} \int (2\pi\sigma^2/n)^{-1/2} \exp\left[-\frac{1}{2(\sigma^2/n)}(\mu-\bar{y})^2\right] d\mu \\ &= \sqrt{2\pi\sigma^2/n} \end{aligned}$$

Thus, the marginal of σ^2 is

$$p(\sigma^2|y) \propto (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \sqrt{2\pi\sigma^2/n}$$

Note that $\sqrt{2\pi/n}$ is a constant that can be dropped leaving us with a $(\sigma^2)^{1/2}$. Combining this with the leading term gives us

$$p(\sigma^2|y) \propto (\sigma^2)^{-(n+1)/2} \exp\left[-\frac{(n-1)}{2\sigma^2}s^2\right]$$

which is the kernel of a scaled inverse chi-squared with $n-1$ degrees of freedom and scale parameter s^2. Thus,

$$\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2) \text{ or } \sigma^2|y \sim IG\left(\frac{n-1}{2}, \frac{n-1}{2}s^2\right)$$

Marginal of μ

As we are typically interested in inference on μ, we must find the posterior marginal of μ given y We must integrate σ^2 out of the joint posterior:

$$\begin{aligned} p(\mu|y) &= \int_0^\infty p(\mu, \sigma^2|y)d\sigma^2 \\ &\propto \int_0^\infty (\sigma^2)^{-(n+2)/2} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\bar{y}-\mu)^2\right\}\right]d\sigma^2 \end{aligned}$$

For ease of notation, let's let $\theta = \sigma^2$ and $\beta = \frac{(n-1)s^2+n(\bar{y}-\mu)^2}{2}$. The integral can then be written as

$$\begin{aligned} p(\mu|y) &\propto \int_0^\infty \theta^{-(n+2)/2}e^{-\beta/\theta}d\theta \\ &\propto \int_0^\infty \theta^{-(n/2+1)}e^{-\beta/\theta}d\theta \end{aligned}$$

This should remind us of the kernel of an inverse-gamma density with parameters $n/2$ and β. Thus we should multiply and divide by the necessary normalizing constants:

$$p(\mu|y) \propto \frac{\Gamma(n/2)}{\beta^{n/2}} \int_0^\infty \frac{\beta^{n/2}}{\Gamma(n/2)}\theta^{-(n/2+1)}e^{-\beta/\theta}d\theta$$

The integrand is now the fully specified density of an $IG(n/2, \beta)$ integrated over the support of θ. Thus the integral is 1. Further note that $\Gamma(n/2)$ is a constant as n is fixed, thus

$$\begin{aligned} p(\mu|y) &\propto \beta^{-n/2} \\ &= \left[\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2}\right]^{-n/2} \end{aligned}$$

But here $(1/2)^{-n/2}$ is also a constant, thus

$$\begin{aligned} p(\mu|y) &\propto \left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]^{-n/2} \\ &= [(n-1)s^2]^{-n/2}\left[1 + \frac{n(\bar{y}-\mu)^2}{(n-1)s^2}\right]^{-n/2} \end{aligned}$$

The leading term is, once again, a constant, thus

$$p(\mu|y) \propto \left[1 + \frac{n(\bar{y}-\mu)^2}{(n-1)s^2}\right]^{-n/2}$$

Rearranging the final pieces we get

$$p(\mu|y) \propto \left[1 + \frac{1}{n-1}\frac{(\mu-\bar{y})^2}{s^2/n}\right]^{-(n-1+1)/2}$$

which is the kernel of a general t distribution with $n-1$ degrees of freedom, location $\bar{y}$, and scale s^2/n.

Example I.22, Slide 153

Suppose we want to study the polls from 2017 French Presidential Election. There were many parties, but suppose we restrict to three candidates: Le Pen (National Front), Fillon (the Republicans), and Marcon (En Marche!). We'll also consider a fourth category: other (including all other candidates as well as those who abstained). Let's build a multinomial model for this data. Also determine a noninformative prior and find the form of the posterior.

First, we begin by letting $\theta_1, \ldots, \theta_4$ denote the true proportion of voters who prefer each candidate and let $y_1, \ldots, y_4$ denote the observed counts from the poll. Further, define $\mathbf{y} = [y_1 \cdots y_4]$ as the vector containing the data and let $\boldsymbol{\theta} = [\theta_1 \cdots \theta_4]$. The likelihood is then

$$\begin{aligned}\mathcal{L}(\mathbf{y}|\boldsymbol{\theta}) &\propto \prod_{j=1}^{4} \theta_j^{y_j} \\ &= \theta_1^{y_1}\theta_2^{y_2}\theta_3^{y_3}\theta_4^{y_4}\end{aligned}$$

Assuming a conjugate prior, the prior on $\boldsymbol{\theta}$ is

$$\begin{aligned}\pi(\boldsymbol{\theta}) &\propto \prod_{j=1}^{4} \theta_j^{\alpha_j - 1} \\ &= \theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\theta_3^{\alpha_3-1}\theta_4^{\alpha_4-1}\end{aligned}$$

Combining, these two, we get the posterior:

$$\begin{aligned}p(\boldsymbol{\theta}) &\propto \theta_1^{y_1}\theta_2^{y_2}\theta_3^{y_3}\theta_4^{y_4}\theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\theta_3^{\alpha_3-1}\theta_4^{\alpha_4-1} \\ &= \theta_1^{y_1+\alpha_1-1}\theta_2^{y_2+\alpha_2-1}\theta_3^{y_3+\alpha_3-1}\theta_4^{y_4+\alpha_4-1}\end{aligned}$$

which we recognize as the kernel of a Dirichlet with parameters $y_1 + \alpha_1, y_2 + \alpha_2, y_3 + \alpha_3$, and $y_4 + \alpha_4$. Thus,

$$\boldsymbol{\theta}|\mathbf{y} \sim Dir(y_1 + \alpha_1, y_2 + \alpha_2, y_3 + \alpha_3, y_4 + \alpha_4)$$

To find a noninformative prior, we must first consider what one is for the Dirichlet prior. Recall that for the Beta distribution, taking $\alpha = \beta = 1$ results in the standard uniform distribution. Since the Dirichlet is the generalization of the Beta to multiple dimensions, taking $\alpha_1 = \ldots = \alpha_k = 1$ results in a multivariate uniform distribution defined over the k-dimensional hyper-cube. Using the noninformative prior, our posterior is then

$$\boldsymbol{\theta}|\mathbf{y} \sim Dir(y_1 + 1, y_2 + 1, y_3 + 1, y_4 + 1)$$

Example I.23, Slide 171

Using the likelihood and prior from slide 169, write out the model and derive the conditional of $\beta|\sigma^2$ and marginal of σ^2.

Note we assume our outcome vector $Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}'$ has the distribution

$$Y \sim MVN\left(X\boldsymbol{\beta}, \sigma^2 I_{n\times n}\right)$$

where $I_{n\times n}$ is the $n \times n$ identity matrix. The likelihood is then

$$\begin{aligned}\mathcal{L}(Y|X, \boldsymbol{\beta}, \sigma^2) &\propto \left|\sigma^2 I_{n\times n}\right|^{-1/2} \exp\left[-\frac{1}{2}(Y - X\boldsymbol{\beta})'(\sigma^2 I_{n\times n})^{-1}(Y - X\boldsymbol{\beta})\right] \\ &= \left(\sigma^2\right)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(Y - X\boldsymbol{\beta})'(Y - X\boldsymbol{\beta})\right]\end{aligned}$$

Our joint prior is

$$p(\boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-1}$$

which gives a joint posterior of

$$P(\boldsymbol{\beta}, \sigma^2|Y, X,) \propto \left(\sigma^2\right)^{-n/2-1} \exp\left[-\frac{1}{2\sigma^2}(Y - X\boldsymbol{\beta})'(Y - X\boldsymbol{\beta})\right].$$

We now factor the posterior into two components: one is the conditional posterior of $\boldsymbol{\beta}|\sigma^2$, the other is a function of σ^2 alone. Taking a cue from Example 1.17, we add and subtract an estimate of the y, which in this case is $\hat{Y} = X\hat{\boldsymbol{\beta}} = X(X'X)^{-1}X'y$, the Best Linear Unbiased Predictor for Y. Our joint posterior is then

$$P(\boldsymbol{\beta}, \sigma^2|Y, X,) \propto (\sigma^2)^{-n/2-1} \exp\left[-\frac{1}{2\sigma^2}(Y - X\hat{\boldsymbol{\beta}} + X\hat{\boldsymbol{\beta}} - X\boldsymbol{\beta})'(y - X\hat{\boldsymbol{\beta}} + X\hat{\boldsymbol{\beta}} - X\boldsymbol{\beta})\right].$$

We note that $X\hat{\boldsymbol{\beta}} - X\boldsymbol{\beta} = X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and rewrite as

$$P(\boldsymbol{\beta}, \sigma^2|Y, X,) \propto (\sigma^2)^{-n/2-1} \exp\left[-\frac{1}{2\sigma^2}\left\{Y - X\hat{\boldsymbol{\beta}} + X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\}'\left\{(y - X\hat{\boldsymbol{\beta}} + X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\}\right].$$

Now we FOIL treating $Y - X\hat{\boldsymbol{\beta}}$ and $X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ each as a whole:

$$P(\boldsymbol{\beta}, \sigma^2|Y, X,) \propto (\sigma^2)^{-n/2-1} \exp\left[-\frac{1}{2\sigma^2}\left\{(Y - X\hat{\boldsymbol{\beta}})'(Y - X\hat{\boldsymbol{\beta}}) + 2(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'X'(Y - X\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'X'X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\}\right].$$

Note that $(Y - X\hat{\boldsymbol{\beta}})$ is the residuals and that by the property of residuals, $X'(Y - X\hat{\boldsymbol{\beta}}) = 0$ (we'll leave this to your regression class), so

$$P(\boldsymbol{\beta}, \sigma^2|Y, X,) \propto (\sigma^2)^{-n/2-1} \exp\left[-\frac{1}{2\sigma^2}\left\{(Y - X\hat{\boldsymbol{\beta}})'(Y - X\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'X'X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\}\right].$$

We can now factor. The conditional distribution of $\boldsymbol{\beta}|\sigma^2$ is

$$P(\boldsymbol{\beta}|\sigma^2, Y, X) \propto \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]$$

which is multivariate normal with mean $\hat{\boldsymbol{\beta}}$ and variance $\sigma^2(X'X)^{-1}$. Integrating out $\boldsymbol{\beta}$ from the joint posterior, we get

$$\begin{aligned}
P(\sigma^2|X, Y) &\propto \int (\sigma^2)^{-n/2-1} \exp\left[-\frac{1}{2\sigma^2}(Y - X\hat{\boldsymbol{\beta}})'(Y - X\hat{\boldsymbol{\beta}})\right] \exp\left[-\frac{1}{2\sigma^2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'X'X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right] d\boldsymbol{\beta} \\
&\propto (\sigma^2)^{-n/2-1} \exp\left[-\frac{1}{2\sigma^2}(Y - X\hat{\boldsymbol{\beta}})'(Y - X\hat{\boldsymbol{\beta}})\right] (\sigma^2)^{k/2} \\
&= (\sigma^2)^{-n/2+k/2-1} \exp\left[-\frac{1}{2\sigma^2}(Y - X\hat{\boldsymbol{\beta}})'(Y - X\hat{\boldsymbol{\beta}})\right]
\end{aligned}$$

which we recognize as an inverse-gamma with parameters $\frac{n-k}{2}$ and $\frac{1}{2}(Y - X\hat{\boldsymbol{\beta}})'(Y - X\hat{\boldsymbol{\beta}})$.

Example I.24, Slides 176-177

As with the univariate case, we begin by examining the multivariate normal where the covariance is known and fixed. Suppose $\mathbf{y} = [\ y_1 \quad y_2 \quad \cdots \quad y_d\ ]'$ is a vector of normally distributed random variables with unknown mean, $\boldsymbol{\mu}$, and fixed variance, $\boldsymbol{\Sigma}$. Determine the form of the likelihood.

Note the form of likelihood for a single observed y:

$$\mathcal{L}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})\right]$$

If we have a sample of $\mathbf{y}_i$ for $i = 1, \ldots, n$, that are iid $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the likelihood has the form

$$\begin{aligned}
\mathcal{L}(\mathbf{y}_1, \ldots, \mathbf{y}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right] \\
&= |\boldsymbol{\Sigma}|^{-n/2} \exp\left[-\frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}S_0)\right]
\end{aligned}$$

where $S_0 = \sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})'$.

(slide 177)

Given the quadratic form of the likelihood, having a prior that is also of quadratic form would make sense. Thus the conjugate prior is $\boldsymbol{\mu} \sim MVN(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$ which has the form

$$\pi(\boldsymbol{\mu}) \propto |\boldsymbol{\Lambda}_0|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\boldsymbol{\Lambda}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)\right]$$

What is the form of the resulting posterior distribution?

Working from the case of observing a single $\mathbf{y}$, the posterior as the form:

$$p(\boldsymbol{\mu}|\boldsymbol{\Sigma},\mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right] |\boldsymbol{\Lambda}_0|^{-d/2} \exp\left[-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\boldsymbol{\Lambda}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)\right]$$

But both $|\boldsymbol{\Sigma}|^{-n/2}$ and $|\boldsymbol{\Lambda}_0|^{-d/2}$ are fixed and known, thus

$$\begin{aligned} p(\boldsymbol{\mu}|\boldsymbol{\Sigma},\mathbf{y}) &\propto \exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right] \exp\left[-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\boldsymbol{\Lambda}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)\right] \\ &= \exp\left[-\frac{1}{2}\left\{(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}) + (\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\boldsymbol{\Lambda}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)\right\}\right] \end{aligned}$$

Expanding, we have

$$\begin{aligned} p(\boldsymbol{\mu}|\boldsymbol{\Sigma},\mathbf{y}) &\propto \exp\left[-\frac{1}{2}\left(\mathbf{y}'\boldsymbol{\Sigma}^{-1}\mathbf{y} - \mathbf{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}'\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0'\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}_0'\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)\right] \\ &= \exp\left[-\frac{1}{2}\left(\mathbf{y}'\boldsymbol{\Sigma}^{-1}\mathbf{y} - 2\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}'\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}'\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0'\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)\right] \end{aligned}$$

Further note that $\mathbf{y}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}_0$, and $\boldsymbol{\Lambda}_0$ are all known. Thus,

$$\begin{aligned} p(\boldsymbol{\mu}|\boldsymbol{\Sigma},\mathbf{y}) &\propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\mu}'\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}'\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)\right] \\ &= \exp\left[-\frac{1}{2}\left(\boldsymbol{\mu}'\left\{\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1}\right\}\boldsymbol{\mu} - 2\boldsymbol{\mu}'\left\{\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right\}\right)\right] \\ &= \exp\left[-\frac{1}{2}\left\{\boldsymbol{\mu}'(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})\boldsymbol{\mu} - 2\boldsymbol{\mu}'\left(\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)\right\}\right] \\ &= \exp\left[-\frac{1}{2}\left\{\boldsymbol{\mu}'(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})\boldsymbol{\mu} - 2\boldsymbol{\mu}'(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})^{-1}\left(\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)\right\}\right] \end{aligned}$$

From here we complete the square by adding and subtracting the following term:

$$\left[(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})^{-1}\left(\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)\right]'(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})\left[(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})^{-1}\left(\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)\right]$$

Thus, the posterior is

$$p(\boldsymbol{\mu}|\boldsymbol{\Sigma},\mathbf{y}) \propto \exp\left[-\frac{1}{2}\left\{\boldsymbol{\mu} - (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})^{-1}\left(\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)\right\}'(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})\left\{\boldsymbol{\mu} - (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})^{-1}\left(\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)\right\}\right]$$

which is the kernel of a multivariate normal distribution with mean

$$E(\boldsymbol{\mu}|\mathbf{y}) = (\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})^{-1}\left(\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)$$

and varianace

$$(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})^{-1}$$

Looking back, we see that the key line is

$$p(\boldsymbol{\mu}|\boldsymbol{\Sigma},\mathbf{y}) \propto \exp\left[-\frac{1}{2}\left\{\boldsymbol{\mu}'(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Lambda}_0^{-1})\boldsymbol{\mu} - 2\boldsymbol{\mu}'\left(\boldsymbol{\Sigma}^{-1}\mathbf{y} + \boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0\right)\right\}\right]$$

where the middle term of the quadratic form on $\boldsymbol{\mu}$ is the posterior variance and the last piece of the second term, when multiplied by the variance, becomes the posterior mean.

Example I.27, Slide 195

Let $y_i \sim Pois(\lambda)$ and assume the noninformative prior $\pi(\lambda) \propto \lambda^{-1}$. Find the posterior distribution, the log-posterior, posterior mode, and the resulting information. What form of the normal distribution approximates the posterior resulting from this model?

Let's begin with the likelihood which is given by

$$\mathcal{L}(y_i|\lambda) \propto \lambda^{\sum y_i} e^{-n\lambda}.$$

Multiplying by the prior yields

$$p(\lambda|y_i) \propto \lambda^{\sum y_i - 1} e^{-n\lambda}.$$

The log posterior is then

$$\log\left[p(\lambda|y_i)\right] = C + \left(\sum y_i - 1\right) \log(\lambda) - n\lambda$$

where C is a constant term wrt λ, i.e. $C = -\sum_{i=1}^{n} \log(y_i!)$. To find the posterior mode, we first differentiate the log posterior and then set to zero and solve for λ:

$$\frac{\partial}{\partial\lambda} \log\left[p(\lambda|y_i)\right] = \frac{\sum y_i - 1}{\lambda} - n \overset{\text{set}}{=} 0 \implies \hat{\lambda} = \frac{\sum y_i - 1}{n}.$$

The information is the negative second derivative evaluated at $\lambda = \hat{\lambda}$, thus we have

$$\frac{\partial^2}{\partial\lambda^2} \log\left[p(\lambda|y_i)\right] = -\frac{\sum y_i - 1}{\lambda^2}$$

for the second derivative and

$$I(\lambda)|_{\lambda=\hat{\lambda}} = \frac{\sum y_i - 1}{\hat{\lambda}^2} = \left(\sum y_i - 1\right) \Big/ \frac{\left(\sum y_i - 1\right)^2}{n^2}.$$

Thus the normal approximation of this posterior is given by

$$\lambda|y_i \dot{\sim} N\left(\frac{\sum y_i - 1}{n}, \frac{\sum y_i - 1}{n^2}\right)$$

since the variance is equal to the inverse of the information, i.e. $I(\lambda)|_{\lambda=\hat{\lambda}}^{-1}$.