20/20

M525: HW2

Dominic Bair

October 16, 2022

1. We consider the random variables w_n, $1 \leqslant n \leqslant N$. We assume these random variables are sampled from an identical distribution, and each sample is independent from any other sample. We presume the random variables are distributed normally with unknown mean μ and known variance v. We assume our unknown mean μ is distributed normally with mean M and variance V.

   (i) We now create the following model:

   $$\begin{aligned} \mu &\sim \text{Normal}(M, V), \quad \mu \in \mathbb{R} \\ w_n|\mu &\sim \text{Normal}(\mu, v), \quad 1 \leqslant n \leqslant N, \quad w_n \in \mathbb{R}. \end{aligned}$$

   (ii) Now that we have a model, we wish to find the posterior distribution of μ given our data, that is $p(\mu|w_{1:N})$. However, we first need to find our prior and likelihood. Using our model from (a), we find our prior to be

   $$\begin{aligned} p(\mu) = \text{Normal}(\mu; M, V) &= \frac{1}{\sqrt{2\pi V}} \exp\left\{-\frac{(\mu - M)^2}{2V}\right\} \\ &\propto \exp\left\{-\frac{(\mu - M)^2}{2V}\right\}. \end{aligned}$$

   Next, we find our likelihood to be

   $$\begin{aligned} p(w_{1:N}|\mu) = \prod_{n=1}^{N} \text{Normal}(w_n; \mu, v) &= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi v}} \exp\left\{-\frac{(w_n - \mu)^2}{2v}\right\} \\ &\propto \exp\left\{-\frac{\sum_{n=1}^{N}(w_n - \mu)^2}{2v}\right\}. \end{aligned}$$

   Using Bayes' Theorem, we know

   $$\begin{aligned} p(\mu|w_{1:N}) \propto p(w_{1:N}|\mu)p(\mu) &\propto \exp\left\{-\frac{\sum_{n=1}^{N}(w_n - \mu)^2}{2v}\right\} \exp\left\{-\frac{(\mu - M)^2}{2V}\right\} \\ &= \exp\left\{-\frac{\sum_{n=1}^{N}(w_n - \mu)^2}{2v} - \frac{(\mu - M)^2}{2V}\right\}. \end{aligned}$$

After some algebra, we let $\alpha = v + VN$, $\beta = 2\left(Mv + V\sum_{n=1}^{N} w_n\right)$, and $\gamma = vV$ and are left with

$$p(\mu|w_{1:N}) \propto \exp\left\{-\frac{\alpha\mu^2 - \beta\mu}{2\gamma}\right\}.$$

We then complete the square of the quadratic in μ and are left with

$$\begin{aligned} p(\mu|w_{1:N}) &\propto \exp\left\{-\frac{\left(\mu - \frac{\beta}{2\alpha}\right)^2 - \left(\frac{\beta}{2\alpha}\right)^2}{2\frac{\gamma}{\alpha}}\right\} \\ &\propto \exp\left\{-\frac{\left(\mu - \frac{\beta}{2\alpha}\right)^2}{2\frac{\gamma}{\alpha}}\right\}. \end{aligned}$$

Knowing this a a probability density, it must be normalized. However, we use the idea of conjugacy to see our posterior is a normal distribution with mean $\frac{\beta}{2\alpha}$ and variance $\frac{\gamma}{\alpha}$. Thus our posterior becomes

$$\begin{aligned} p(\mu|w_{1:N}) &= \text{Normal}\left(\mu; \frac{\beta}{2\alpha}, \frac{\gamma}{\alpha}\right) \\ &= \text{Normal}\left(\mu; \frac{Mv + V\sum_{n=1}^{N} w_n}{v + VN}, \frac{vV}{v + VN}\right). \end{aligned}$$

(iii) Now that we have derived the prior and posterior distributions for μ, we apply the data found in `normal_normal.mat` and plot the prior and posterior density functions. See Figure 1.

(iv) We now use the integral function in MATLAB to determine the probability that μ is negative in both our prior and posterior distributions. For our prior distribution, we get the trivial result that $p(\mu < 0) = 0.5$ by the symmetry of a normal distribution. Using our posterior, however, we get $p(\mu < 0) = 3.4377 \times 10^{-13}$, which means it is incredibly unlikely that the true μ is negative given the data.

(v) Being as our prior and posterior distributions are normal, we know the highest density occurs at the mean of each. Thus, our maximum prior estimate is $\hat{\mu} = M = 0$ and our maximum a posteriori estimate is $\hat{\mu} = \frac{Mv + V\sum_{n=1}^{N} w_n}{v + VN} = 1.4335$.

(vi) Given that each w_n is measured in *adu* units, we know that μ must also be in *adu* since we add w_n and μ. This means M is also in *adu*. Since v and V are each in the denominator of an exponent, they must cancel out the units in the numerator; thus, v and V must be in adu^2.

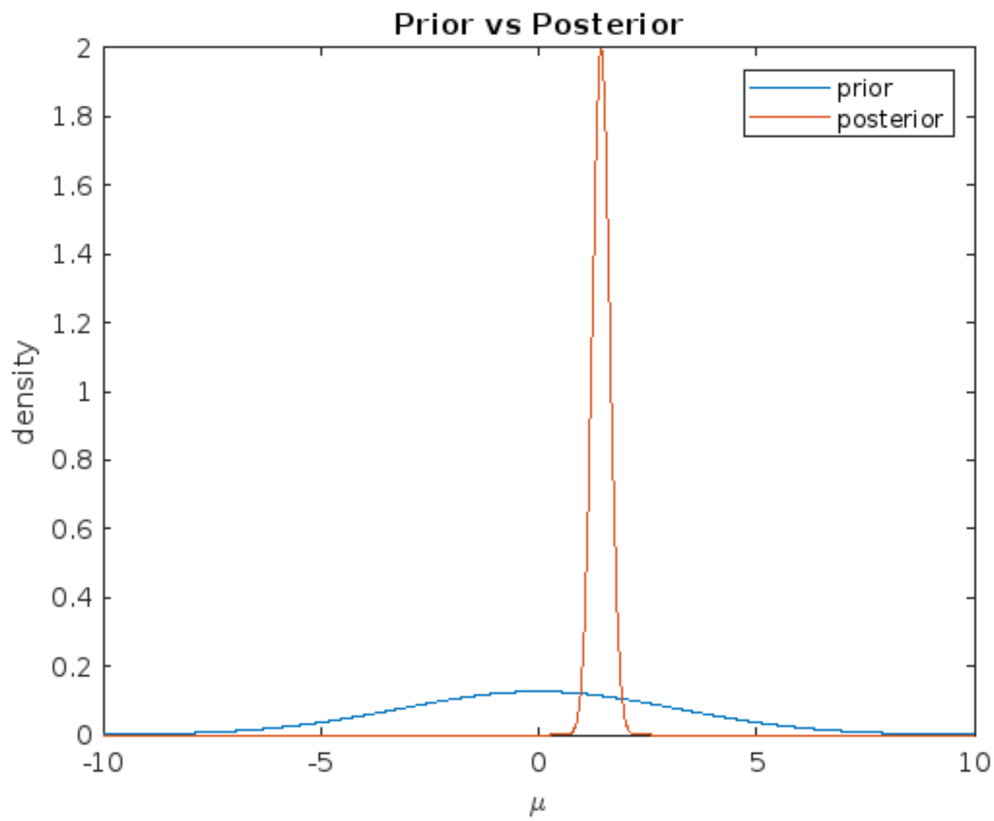

Figure 1: Here we plot the prior and posterior density functions derived in problem 1. Note the posterior is shifted slightly to the right of the prior and has much higher density and much lower variance than the posterior. Our prior mean is 0 and prior variance is 10.

2. Now we consider the random variables w_n, $1 \leqslant n \leqslant N$. These variables are supported on non–negative integers and follow a geometric distribution with unknown parameter π. We assume this π is distributed according to a Beta distribution with known parameters α and β.

   (i) Using this information, we create the following model:

   $$p(\pi) \sim \text{Beta}(\alpha, \beta), \quad \pi \in [0, 1].$$
   $$p(w_n|\pi) \sim \text{Geometric}(\pi), \quad 1 \leqslant n \leqslant N, \quad w_n \in \mathbb{R}^+.$$

   (ii) We now derive the posterior using the same techniques described in 1. (ii). Our prior becomes

   $$\begin{aligned} p(\pi) = \text{Beta}(\pi; \alpha, \beta) &= \frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha, \beta)} \\ &\propto \pi^{\alpha-1}(1-\pi)^{\beta-1}. \end{aligned}$$

   The likelihood of observing our $w_{1:N}$ is

   $$\begin{aligned} p(w_{1:N}|\pi) &= \prod_{n=1}^{N} \text{Geometric}(w_n; \pi) = \prod_{n=1}^{N} \pi(1-\pi)^{w_n} \\ &= \pi^N(1-\pi)^{\sum_{n=1}^N w_n}. \end{aligned}$$

   Now we apply Bayes' Theorem and our posterior becomes

   $$\begin{aligned} p(\pi|w_{1:N}) &\propto \pi^N(1-\pi)^{\sum_{n=1}^N w_n}\pi^{\alpha-1}(1-\pi)^{\beta-1} \\ &= \pi^{N+\alpha-1}(1-\pi)^{\sum_{n=1}^N w_n+\beta-1}. \end{aligned}$$

   Using conjugacy again, we see this is proportional to a Beta density function with parameters $N + \alpha$ and $\displaystyle\sum_{n=1}^{N} w_n + \beta$. Thus, our posterior is

   $$p(\pi|w_{1:N}) = \frac{\pi^{N+\alpha-1}(1-\pi)^{\sum_{n=1}^N w_n+\beta-1}}{B(N+\alpha, \sum_{n=1}^N w_n + \beta).}$$

   (iii) Now that we have the posterior, we can use our data to plot the probability densities of the prior and posterior distributions of our model. We use the parameters $\alpha = \beta = 1$. See Figure 2.

   (iv) Using the integral function in MATLAB, we find that using the prior distribution gives us $p(\pi \geqslant 0.15) = 0.85$, and using the posterior distribution gives us $p(\pi \geqslant 0.15) = 0.0031$. We note that since the prior is just the 1 function, the area under the curve between 0.15 and 1 is just $1 - 0.15 = 0.85$.

   (v) Lastly, w use the principle of maximum a posteriori to find the "best" choice of π.

   $$p(\pi|w_{1:N}) \propto \pi^{N+\alpha-1}(1-\pi)^{\sum_{n=1}^N w_n+\beta-1} = \pi^N(1-\pi)^{\sum_{n=1}^N w_n}.$$
   $$\implies \frac{dp(\pi|w_{1:N})}{d\pi} \propto N(\pi^{N-1}(1-\pi)^{\sum_{n-1}^N w_n} - \pi^N \sum_{n-1}^{N} w_n(1-\pi)^{\sum_{n-1}^N w_n - 1}).$$

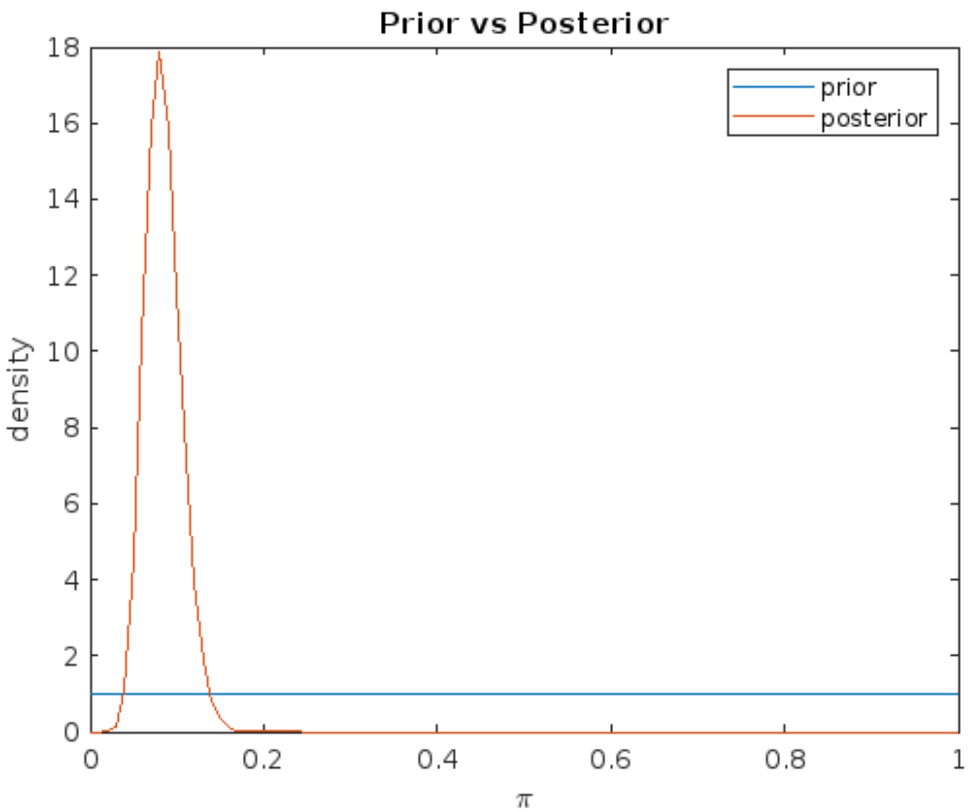

Figure 2: Here we plot the prior and posterior density functions derived in problem 2. Note that the prior is very diffuse, giving no preference to the value of π, but the posterior gives strong preference to values of π near 0.1.

Applying our data, then setting this equal to zero and solving for π, we find $\hat{\pi} = \frac{7}{88}$, which matches our plot of the density very well.

3. Finally, we consider a more exotic model.

$$
\begin{aligned}
(\phi, \psi) &\sim \mathrm{H}(A, B, r, s) \\
w_n | \phi, \psi &\sim \mathrm{Gamma}(\phi, \psi), \quad 1 \leqslant n \leqslant N.
\end{aligned}
$$

Here, $\mathrm{H}(A, B, r, s)$ has a probability density of the form

$$
\mathrm{H}(\phi, \psi; A, B, r, s) \propto \frac{A^{\phi-1} e^{-\frac{B}{\psi}}}{(\Gamma(\phi))^r \, \psi^{\phi s}}, \quad \phi, \psi > 0.
$$

(i) We first derive the posterior of this model. Using Bayes' Theorem we see that

$$
\begin{aligned}
p(\phi, \psi | w_{1:N}) &\propto \mathrm{H}(A, B, r, s) \prod_{n=1}^{N} \mathrm{Gamma}(\phi, \psi) \\
&= \frac{A^{\phi-1} e^{-\frac{B}{\psi}}}{(\Gamma(\phi))^r \, \psi^{\phi s}} \prod_{n=1}^{N} \frac{w_n^{\phi-1} e^{-\frac{w_n}{\psi}}}{\psi^{\phi} \Gamma(\phi)} \\
&= \frac{\left(A \prod_{n=1}^{N}\right)^{\phi-1} e^{\frac{B - \sum_{n=1}^{N} w_n}{\psi}}}{(\Gamma(\phi))^{r+N} \, \psi^{s+N}} \\
\implies p(\phi, \psi | w_{1:N}) &= H\left(\phi, \psi;\ A \prod_{n=1}^{N} w_n,\ B - \sum_{n=1}^{N} w_n,\ r+N,\ s+N\right).
\end{aligned}
$$

(ii) In the last step, we see that our posterior is of the same form as our prior; thus our model is conjugate.