MA525 Homework 2

Aaron Cornelius

2020-10-11

Problem 1

Consider data ω_n where $n = 1, \ldots, N$ indexes individual measurements. Each ω_n is obtained through a $\text{Normal}(\mu, v)$ distribution. Assume the variance v has a known value. On μ use a $\text{Normal}(M, V)$ prior where M, V have known values.

(i)

In statistical notation, this problem can be written as:

$$\mu \sim \text{Normal}(M, V)$$
$$\omega_n | \mu \sim \text{Normal}(\mu, v)$$

(ii)

$$p(\mu) = \text{Normal}(\mu; M, V) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{1}{2v}(x-\mu)^2\right)$$

The PDF of the normal distribution, written using the variance, is:

$$p(x) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{1}{2v}(x-\mu)^2\right)$$

Therefore, there are two probability functions. The first explicitly defines the probability density of the distribution of μ:

$$p_\mu(\mu) = \frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{1}{2V}(\mu-M)^2\right)$$

The second defines the probability of an observed value ω conditioned on μ:

$$p_\omega(\omega|\mu) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{1}{2v}(\omega-\mu)^2\right)$$

The likelihood is taken by, the likelihood function for some set of observations ω_n is:

$$p(\omega_{1:N}|\mu;) = \prod_{n=1}^{N} p(\omega_n|\mu)$$

$$= \prod_{n=1}^{N}\left(\frac{1}{\sqrt{2\pi v}}\exp\left(-\frac{1}{2v}(\omega_n-\mu)^2\right)\right)$$

$$= \prod_{n=1}^{N}\left(\frac{1}{\sqrt{2\pi v}}\right)\prod_{n=1}^{N}\left(\exp\left(-\frac{1}{2v}(\omega_n-\mu)^2\right)\right)$$

$$= \left(\frac{1}{\sqrt{2\pi v}}\right)^N \exp\left(-\frac{1}{2v}\sum_{n=1}^{N}(\omega_n-\mu)^2\right)$$

The posterior can be calculated by multiplying the likelihood by the prior distribution:

$$p(\mu|\omega_{1:N}) \propto p(\omega_{1:N}|\mu)p(\mu)$$

$$= \left(\left(\frac{1}{\sqrt{2\pi v}}\right)^N \exp\left(-\frac{1}{2v}\sum_{n=1}^{N}(\omega_n-\mu)^2\right)\right)\left(\frac{1}{\sqrt{2\pi V}}\exp\left(-\frac{1}{2V}(\mu-M)^2\right)\right)$$

$$\propto \exp\left(-\frac{1}{2v}\sum_{n=1}^{N}(\omega_n-\mu)^2\right)\exp\left(-\frac{1}{2V}(\mu-M)^2\right)$$

$$= \exp\left(-\frac{1}{2v}\sum_{n=1}^{N}(\omega_n-\mu)^2-\frac{1}{2V}(M-\mu)^2\right) = \exp\left(-\frac{1}{2}\left(\frac{1}{v}\sum_{n=1}^{N}(\omega_n-\mu)^2+\frac{1}{V}(M-\mu)^2\right)\right)$$

To simplify the equation, define $f = \frac{1}{v}\sum_{n=1}^{N}(\omega_n-\mu)^2+\frac{1}{V}(M-\mu)^2$. Therefore, by defining $\bar{\omega} = \frac{1}{N}\sum_{n=1}^{N}\omega_n$:

$$p(\mu|\omega_{1:N}) \propto \exp\left(-\frac{1}{2}f\right)$$

$$f = \frac{1}{v}\sum_{n=1}^{N}(\omega_n-\mu)^2+\frac{1}{V}(M-\mu)^2$$

$$= \frac{1}{v}\sum_{n=1}^{N}((\omega_n-\bar{\omega})+(\bar{\omega}-\mu))^2+\frac{1}{V}(M-\mu)^2$$

$$= \frac{1}{v}\left(\sum_{n=1}^{N}(\omega_n-\bar{\omega})^2+2\sum_{n=1}^{N}(\omega_n-\bar{\omega})(\bar{\omega}-\mu)+\sum_{n=1}^{N}(\bar{\omega}-\mu)^2\right)+\frac{1}{V}(M-\mu)^2$$

$$= \frac{1}{v}\left(\sum_{n=1}^{N}(\omega_n-\bar{\omega})^2+2(\bar{\omega}-\mu)\sum_{n=1}^{N}(\omega_n-\bar{\omega})+N(\bar{\omega}-\mu)^2\right)+\frac{1}{V}(M-\mu)^2$$

Since $\sum_{n=1}^{N}(\omega_n-\bar{\omega}) = 0$, this further simplifies to:

$$f = \frac{1}{v}\left(\sum_{n=1}^{N}(\omega_n - \bar{\omega})^2 + N(\bar{\omega} - \mu)^2\right) + \frac{1}{V}(M - \mu)^2$$

$$= \frac{1}{v}\sum_{n=1}^{N}(\omega_n - \bar{\omega})^2 + \frac{1}{v}N(\bar{\omega} - \mu)^2 + \frac{1}{V}(M - \mu)^2$$

Substituting back into the original equation:

$$p(\mu|\omega_{1:N}) \propto \exp\left(-\frac{1}{2}\left(\frac{1}{v}\sum_{n=1}^{N}(\omega_n - \bar{\omega})^2 + \frac{1}{v}N(\bar{\omega} - \mu)^2 + \frac{1}{V}(M - \mu)^2\right)\right)$$

$$= \exp\left(-\frac{1}{2v}\sum_{n=1}^{N}(\omega_n - \bar{\omega})^2\right)\exp\left(-\frac{1}{2}\left(\frac{1}{v}N(\bar{\omega} - \mu)^2 + \frac{1}{V}(M - \mu)^2\right)\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(\frac{1}{v}N(\bar{\omega} - \mu)^2 + \frac{1}{V}(M - \mu)^2\right)\right)$$

Now define a quadratic polynomial:

$$g = \frac{N}{v}(\bar{\omega} - \mu)^2 + \frac{1}{V}(M - \mu)^2$$

$$= \frac{N}{v}\mu^2 - \frac{2N\bar{\omega}\mu}{v} + \frac{N\bar{\omega}^2}{v} + \frac{M^2}{V} - \frac{2M\mu}{V} + \frac{\mu^2}{V}$$

$$= \left(\frac{N}{v} + \frac{1}{V}\right)\left(\mu^2 - \frac{2\mu}{\frac{N}{v} + \frac{1}{V}}\left(\frac{N\bar{\omega}}{v} + \frac{M}{V}\right) + \frac{1}{\frac{N}{v} + \frac{1}{V}}\left(\frac{N\bar{\omega}^2}{v} + \frac{M^2}{V}\right)\right)$$

Now some proportionality constant K can be multiplied in, allowing the formula to be written:

$$g = \left(\frac{N}{v} + \frac{1}{V}\right)\left(\mu - \frac{\left(\frac{N\bar{\omega}}{v} + \frac{M}{V}\right)}{\frac{N}{v} + \frac{1}{V}}\right)^2 + K$$

In the original equation, this equals:

$$p(\mu|\omega_{1:N}) \propto \exp\left(-\frac{1}{2}g\right) = \exp\left(-\frac{1}{2}K\right)\exp\left(-\frac{1}{2}\left(\frac{1}{\left(\frac{N}{v} + \frac{1}{V}\right)^{-1}}\right)\left(\left(\mu - \frac{\left(\frac{N\bar{\omega}}{v} + \frac{M}{V}\right)}{\frac{N}{v} + \frac{1}{V}}\right)^2\right)\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\left(\frac{N}{v} + \frac{1}{V}\right)^{-1}}\right)\left(\left(\mu - \frac{\left(\frac{N\bar{\omega}}{v} + \frac{M}{V}\right)}{\frac{N}{v} + \frac{1}{V}}\right)^2\right)\right)$$

This matches the form of the normal distribution $\mathrm{Normal}(\mu^*, v^*)$ where:

$$v^* = \frac{1}{\left(\frac{N}{v} + \frac{1}{V}\right)}, \mu^* = v^*\left(\frac{N\bar{\omega}}{v} + \frac{M}{V}\right)$$

These can also be simplified:

$$v^* = \frac{1}{\left(\frac{N}{v} + \frac{1}{V}\right)} = \frac{1}{\frac{V}{V}\frac{N}{v} + \frac{v}{v}\frac{1}{V}} = \frac{1}{\frac{\mathrm{VN} + v}{\mathrm{Vv}}} = \frac{\mathrm{Vv}}{\mathrm{VN} + v}$$

$$\mu^* = \frac{\mathrm{Vv}}{\mathrm{VN} + v}\left(\frac{N\bar{\omega}}{v} + \frac{M}{V}\right) = \frac{\mathrm{Vv}}{\mathrm{VN} + v}\left(\frac{V}{V}\frac{N\bar{\omega}}{v} + \frac{v}{v}\frac{M}{V}\right)$$

$$= \frac{\mathrm{Vv}}{\mathrm{VN} + v}\left(\frac{\mathrm{VN}\bar{\omega} + \mathrm{vM}}{\mathrm{Vv}}\right)$$

$$= \frac{\mathrm{VN}\bar{\omega} + \mathrm{vM}}{\mathrm{VN} + v}$$

The equation can thus be written:

$$p(\mu|\omega_{1:N}) \propto \exp\left(-\frac{1}{2v^*}(\mu - \mu^*)^2\right)$$

$$\Rightarrow p(\mu|\omega_{1:N}) = C \cdot \exp\left(-\frac{1}{2v^*}(\mu - \mu^*)^2\right) = p(\mathrm{Normal}(\mu^*, v^*)) = p_\mu(\mu) = \frac{1}{\sqrt{2\pi v^*}}\exp\left(-\frac{1}{2v^*}(\mu - \mu^*)^2\right)$$

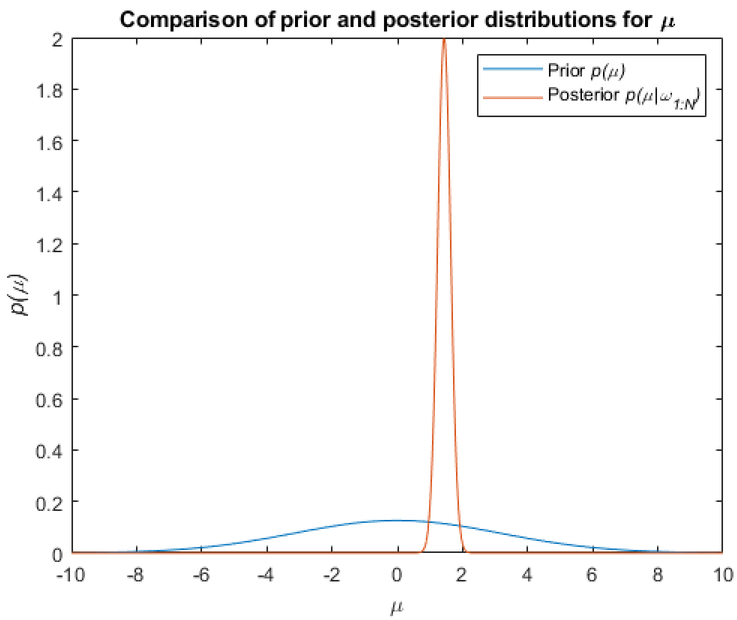
$$\Rightarrow C = \frac{1}{\sqrt{2\pi v^*}}$$

Therefore:

$p(\mu|\omega_{1:N}) = \frac{1}{\sqrt{2\pi v^*}} \cdot \exp\left(-\frac{1}{2v^*}(\mu - \mu^*)^2\right)$, where $v^* = \frac{\mathrm{Vv}}{\mathrm{VN} + v}, \mu^* = \frac{\mathrm{VN}\bar{\omega} + \mathrm{vM}}{\mathrm{VN} + v}$

(iii)

This graph compares the prior and posterior results for the distribution of μ. The prior distribution was very wide, indicating great uncertainty about what the true mean was. After the sample points are added, the posterior has converged much more strongly towards a single value.



(iv)

The probability that $\mu < 0$ can be found by integrating the prior and posterior probabilties from negative infinity to infinity. Since it is not possible to numerically integrate negative infinity, the integrals will instead be performed from -100 to 0. From visual inspection of the graph of the prior and posteriors, it can be see that the probability past -10 is extremely low, and thus there will be little effect on the overall integration.

$$p(\mu) < 0.5 \approx \int_{-100}^{0} p(\mu)\, d\mu = 0.5 = 50\%$$

$$p(\mu|\omega_{1:N}) < 0.5 \approx \int_{-100}^{0} p(\mu|\omega_{1:N})\, d\mu \approx 3.44 \cdot 10^{-13} = 3.44 \cdot 10^{-11}\%$$

These results make intuitive sense: the prior has a mean of 0, so there is a 50% chance that the value will be less than the mean (i.e., negative.) The posterior is centered roughly around $\mu = 1$, with very little probability that $\mu < 0$.

(v)

The maximum a posteriori estimate for μ can be found by differentiating the probability distribution and finding the point where the slope is equal to zero.

$$p(\mu|\omega_{1:N}) = \frac{1}{\sqrt{2\pi v^*}} \cdot \exp\left(-\frac{1}{2v^*}(\mu - \mu^*)^2\right)$$, where $v^* = \frac{\mathrm{Vv}}{\mathrm{VN} + v}$, $\mu^* = \frac{\mathrm{VN}\bar{\omega} + \mathrm{vM}}{\mathrm{VN} + v}$

$$\frac{d}{d\mu} p(\mu|\omega_{1:N}) = \frac{1}{\sqrt{2\pi v^*}} \cdot (f'(g(\mu)) \cdot g'(\mu))$$, where $f(\mu) = \exp(\mu), g(\mu) = -\frac{1}{2v^*}(\mu - \mu^*)^2$

$$= \frac{1}{\sqrt{2\pi v^*}} \cdot \exp\left(-\frac{1}{2v^*}(\mu - \mu^*)^2\right) \cdot \left(-\frac{1}{v^*}(\mu - \mu^*)\right)$$
$$0 = \frac{1}{\sqrt{2\pi v^*}} \cdot \exp\left(-\frac{1}{2v^*}(\mu - \mu^*)^2\right) \cdot \left(-\frac{1}{v^*}(\mu - \mu^*)\right) = \exp\left(-\frac{1}{2v^*}(\mu - \mu^*)^2\right) \cdot \left(-\frac{1}{v^*}(\mu - \mu^*)\right)$$

This is true when either $\exp\left(-\frac{1}{2v^*}(\mu - \mu^*)^2\right) = 0$ or $\left(-\frac{1}{v^*}(\mu - \mu^*)\right) = 0$. However, the exponential function can never be equal to zero. Therefore:

$$\left(-\frac{1}{v^*}(\mu - \mu^*)\right) = 0 = (\mu - \mu^*)$$
$$\Rightarrow \mu = \mu^* = \frac{\mathrm{VN}\bar{\omega} + \mathrm{vM}}{\mathrm{VN} + v} \approx 1.4335$$

This result makes intuitive sense: the peak of a normal distribution is always at its mean value.

(vi)

Each measurement ω_n is in units of $[\mathrm{au}]$. In the $\mathrm{Normal}(\mu, v)$ distribution that the measurements are drawn from, each measurement is subtracted from the mean value μ. Therefore, they must have the same units, so the units of μ must also be $[\mathrm{au}]$. Similarly, in the prior distribution $\mathrm{Normal}(M, V)$, each value of μ is subtracted from the prior mean M, so the units of M must also be $[\mathrm{au}]$. In both cases, the differences between the mean values μ, M and the observed values ω_n, μ are then squared (giving units of $[\mathrm{au}^2]$), and divided by the relevant variances v, V. The result of that division must be unitless for the exponential. Therefore, v, V must have the same units as ω_n^2, μ^2: $[\mathrm{au}^2]$.

In summary:

$$\mu : [\mathrm{au}]$$
$$M : [\mathrm{au}]$$
$$v : [\mathrm{au}^2]$$
$$V : [\mathrm{au}^2]$$

Problem 2

Set up a Bayesian model for discrete measurements. Consider data ω_n, where each value $n = 1, \ldots, N$ is an individual measurement. ω_n is drawn from a $\mathrm{Geometric}(\pi)$ distribution with probability mass given by:

$$p(\omega) = \begin{cases} (1-\pi)^{\omega}\pi & \omega \geq 0 \\ 0 & \text{else} \end{cases}$$

The prior of π is defined with a $\mathrm{Beta}(\alpha, \beta)$ distribution where α, β have known values. The probability density of the Beta distribution is:

$$p(\pi) = \begin{cases} \dfrac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha,\beta)} & 0 \le \pi \le 1 \\ \text{else} & 0 \end{cases}$$

where $B(\alpha, \beta)$ is the Beta function:

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, dt$$

(i)

In statistical notation, this problem can be written as:

$\pi \sim \text{Beta}(\alpha, \beta)$
$\omega_n|\pi \sim \text{Geometric}(\pi)$

$$p(\omega|\pi) = (1-\pi)^{\omega}\, \pi$$

$$p(\pi) = \frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha,\beta)}$$

(ii)

The posterior can be calculated as $p(\pi|\omega_{1:N}) = C\ p(\omega_{1:N}|\pi)\ p(\pi)$. $p(\pi)$ is defined from the posterior. The next step is to calculate $p(\omega_{1:N}|\pi)$, which can be done by taking the product of the probability of each individual observation conditioned on π:

$$p(\omega_{1:N}|\pi) = \prod_{n=1}^{N} p(\omega_n|\pi) = \prod_{n=1}^{N} \left((1-\pi)^{\omega_n}\pi\right) = \pi^N (1-\pi)^{\sum_{n=1}^{N}\omega_n}$$

Therefore, the posterior can be written:

$$p(\pi|\omega_{1:N}) \propto p(\omega_{1:N}|\pi)\ p(\pi) = \pi^N (1-\pi)^{\sum_{n=1}^{N}\omega_n} \frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha,\beta)}$$

$$\propto (1-\pi)^{\sum_{n=1}^{N}\omega_n} \pi^{\alpha-1}(1-\pi)^{\beta-1} = \pi^{N+\alpha-1}(1-\pi)^{\beta-1+\sum_{n=1}^{N}\omega_n}$$

It is known that $\int_0^1 p(\pi|\omega_{1:N}) = 1$. The next step is to find the normalization coefficient C. This is simple for this equation: the posterior $p(\pi|\omega_{1:N})$ is proportional to a beta distribution with updated parameters α^*, β^* where:

$\alpha^* = \alpha + N, \beta^* = \beta + \sum_{n=1}^{N} \omega_n$. Therefore:

$$p(\pi|\omega_{1:N}) = C\pi^{\alpha^*-1}(1-\pi)^{\beta^*-1}$$

$$\Rightarrow 1 = \int_0^1 C\pi^{\alpha-1}(1-\pi)^{\beta-1+\sum\limits_{n=1}^{N}\omega_n} d\pi = \int_0^1 \frac{\pi^{\alpha^*-1}(1-\pi)^{\beta^*-1}}{B(\alpha^*, \beta^*)} d\pi$$

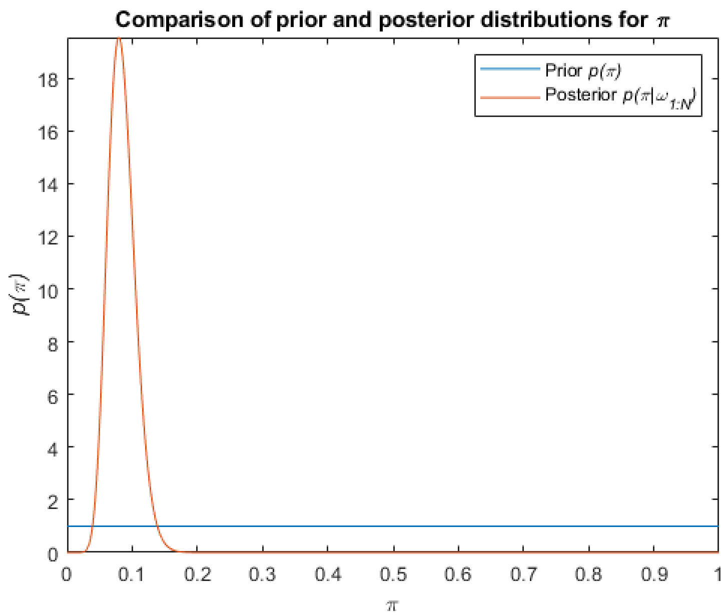
$$= \frac{1}{B(\alpha^*, \beta^*)} \int_0^1 \pi^{\alpha^*-1}(1-\pi)^{\beta^*-1} d\pi$$

$$\Rightarrow C = \frac{1}{B(\alpha^*, \beta^*)} = \frac{1}{B\left(\alpha + N, \beta + \sum\limits_{n=1}^{N}\omega_n\right)}$$

$$\Rightarrow p(\pi|\omega_{1:N}) = \frac{\pi^{\alpha+N-1}(1-\pi)^{\beta+\sum\limits_{n=1}^{N}\omega_n - 1}}{B\left(\alpha + N, \beta + \sum\limits_{n=1}^{N}\omega_n\right)}$$

(iii)

The following graph compares the prior and posterior distributions of π when $\alpha = \beta = 1$ and updating using the provided observations $\omega_{1:N}$. It can be seen that the prior distribution is entirely uniform over $0 \leq \pi \leq 1$. This makes sense since for the case $\alpha = \beta = 1$, the prior distribution PDF simplifies to $p(\pi) = \frac{\pi^0(1-\pi)^0}{B(1,1)} = \frac{1}{B(1,1)}$ when $0 \leq \pi \leq 1$.



(iv)

The probability that $\pi > 0.15$ can be determined by integrating the prior and posterior distributions from 0.15 to 1. This is done numerically. Therefore:

$$p(\pi) > 0.15 = \int_{0.15}^{1} \frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha,\beta)} d\pi = 0.85 = 85\%$$

$$p(\pi|\omega_{1:N}) > 0.15 = \int_{0.15}^{1} \frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha,\beta)} d\pi \approx 0.0034 \approx 0.34\%$$

This matches up with visual inspection from the graph above: the posterior is uniform and thus the probability will be proportional to the percentage of the range integrated over (i.e., $100\% - 15\% = 85\%$), but the posterior is extremely confident that the value of π will be just below 0.1, with very little probability where $\pi > 0.15$.

(v)

The maximum value of the posterior can be analytically determined by differentiating the posterior probability distribution $p(\pi|\omega_{1:N})$ and finding the zero-slope point:

$$\frac{d}{d\pi}p(\pi|\omega_{1:N}) = \frac{1}{B(\alpha^*,\beta^*)}\left(\left(\frac{d}{d\pi}\pi^{\alpha+N-1}\right)(1-\pi)^{\beta+\sum_{n=1}^{N}\omega_n-1} + \left(\frac{d}{d\pi}(1-\pi)^{\beta+\sum_{n=1}^{N}\omega_n-1}\right)\pi^{\alpha+N-1}\right)$$

$$= \frac{1}{B(\alpha^*,\beta^*)}\left((\alpha+N-1)\pi^{a+N-2}(1-\pi)^{\beta+\sum_{n=1}^{N}\omega_n-1} + \left(-\beta+\sum_{n=1}^{N}\omega_n-1\right)(1-\pi)^{\beta+\sum_{n=1}^{N}\omega_n-1}\pi^{\alpha+N-1}\right)$$

$$0 = \frac{1}{B(\alpha^*,\beta^*)}\left((\alpha+N-1)\pi^{a+N-2}(1-\pi)^{\beta+\sum_{n=1}^{N}\omega_n-1} + \left(-\beta+\sum_{n=1}^{N}\omega_n-1\right)(1-\pi)^{\beta+\sum_{n=1}^{N}\omega_n-1}\pi^{\alpha+N-1}\right)$$

$$= \left((\alpha+N-1)\pi^{a+N-2}(1-\pi)^{\beta+\sum_{n=1}^{N}\omega_n-1} + \left(-\beta+\sum_{n=1}^{N}\omega_n-1\right)(1-\pi)^{\beta+\sum_{n=1}^{N}\omega_n-2}\pi^{\alpha+N-1}\right)$$

$$\Rightarrow (\alpha+N-1)\pi^{a+N-2}(1-\pi)^{\beta+\sum_{n=1}^{N}\omega_n-1} = \left(\beta+\sum_{n=1}^{N}\omega_n-1\right)(1-\pi)^{\beta+\sum_{n=1}^{N}\omega_n-2}\pi^{\alpha+N-1}$$

$$\Rightarrow (\alpha+N-1)(1-\pi) = \left(\beta+\sum_{n=1}^{N}\omega_n-1\right)\pi$$

$$\Rightarrow \frac{(1-\pi)}{\pi} = \frac{\left(\beta+\sum_{n=1}^{N}\omega_n-1\right)}{(a+N-1)} = \frac{1}{\pi}-1$$

$$\sum_{n=1}^{N}\omega_n = 162, N=14, \alpha=1, \beta=1$$

$$\Rightarrow \frac{1}{\pi}-1 = \frac{(1+162-1)}{(1+14-1)} = \frac{162}{14}$$

$$\Rightarrow \frac{1}{\pi} = \frac{176}{14}$$

$$\Rightarrow \pi = \frac{14}{176} \sim 0.0795$$

This matches up with visual inspection of the graph above: the peak is slightly below $\pi = 0.1$. This can also be verified by evaluating $p(\pi|\omega_{1:N})$ at points slightly above and below $\frac{14}{176}$ and making sure that both points are lower than $p\left(\frac{14}{176}\middle|\omega_{1:N}\right)$:

$$p\left(\frac{13}{176}\middle|\omega_{1:N}\right) \approx 17.7754$$

$$p\left(\frac{14}{176}\middle|\omega_{1:N}\right) \approx 19.5531$$

$$p\left(\frac{15}{176}\middle|\omega_{1:N}\right) \approx 18.8391$$

Problem 3

Determine if the models in the previous problems belong to the exponential family. For this to be true, with x as the parameters and y as the measurements, the function must be able to be written as:

$$p(x) \propto (G(x))^{\eta} \exp(\phi(x)\nu)$$
$$p(y|x) = F(y)G(x) \exp(\phi(x)U(y))$$

(i)

For the conjugate model to belong to the exponential family, the following must be true:

$$p(\mu) \propto G(\mu)^{\eta} \exp(\phi(\mu)\,\nu)$$
$$p(\omega|\mu) = F(\omega)G(\mu) \exp(\phi(\mu)U(\omega))$$

Start by looking at the probability of ω conditioned on μ first:

$$p(\omega|\mu) = \left(\frac{1}{\sqrt{2\pi v}}\right)\exp\left(-\frac{1}{2v}(\omega-\mu)^2\right)$$
$$= \left(\frac{1}{\sqrt{2\pi v}}\right)\exp\left(-\frac{1}{2v}(\omega^2-2\omega\mu)+\mu^2\right)$$
$$= \left(\frac{1}{\sqrt{2\pi v}}\right)\exp\left(-\frac{\omega^2}{2v}\right)\exp\left(\frac{\omega\mu}{v}-\frac{\mu^2}{2v}\right)$$

Define $F(\omega) = \exp\left(-\frac{\omega^2}{2v}\right)$. Therefore:

$$p(\omega|\mu) = F(\omega)\left(\frac{1}{\sqrt{2\pi v}}\right)\exp\left(\omega\frac{\mu}{v}-\frac{\mu^2}{2v}\right)$$

Define $G(\mu) = \left(\frac{1}{\sqrt{2\pi v}}\right)\exp\left(-\frac{\mu^2}{2v}\right)$. Therefore:

$$p(\omega|\mu) = F(\omega)G(\mu)\exp\left(\frac{\omega\mu}{v}\right)$$

Define $\phi(\mu) = \frac{\mu}{v}, U(\omega) = \omega$:

$$p(\omega|\mu) = F(\omega)G(\mu)\exp(\phi(\mu)U(\omega))$$

Now, the prior $p(\mu)$ must be defined using this same set of equations:

$$p(\mu) = \frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{1}{2V}(\mu - M)^2\right) \propto G(\mu)^\eta \exp(\phi(\mu)\,\nu)$$

$$= \left(\left(\frac{1}{\sqrt{2\pi v}}\right) \exp\left(-\frac{\mu^2}{2v}\right)\right)^\eta \exp(\nu\,\mu)$$

$$= \left(\frac{1}{\sqrt{2\pi v}}\right)^\eta \exp\left(\nu\,\mu - \eta\frac{\mu^2}{2v}\right)$$

If $\eta = \frac{v}{V}$ and $\nu = \frac{M}{V}$:

$$p(\mu) \propto \left(\frac{1}{\sqrt{2\pi v}}\right)^{\frac{v}{V}} \exp\left(\frac{M}{V}\mu - \frac{v}{V}\frac{\mu^2}{2v}\right)$$

$$\propto \exp\left(\frac{M\mu}{V} - \frac{\mu^2}{2V}\right)$$

$$\propto \exp\left(\frac{M\mu}{V} - \frac{\mu^2}{2V}\right) \exp\left(-\frac{1}{2}\frac{M^2}{V}\right)$$

$$= \exp\left(\frac{M\mu}{V} - \frac{\mu^2}{2V} - \frac{M^2}{2V}\right)$$

$$= \exp\left(-\frac{1}{2V}(\mu^2 - M\mu + M^2)\right)$$

$$= \exp\left(-\frac{1}{2V}(M - \mu)^2\right)$$

This is the correct form for the prior, with $p(\mu) = C \cdot \exp\left(-\frac{1}{2V}(M - \mu)^2\right)$. Therefore, the normal conjugate model from problem 1 is part of the exponential family. While this examination only looked at the case where there was a single observation (i.e., $N = 1$), this scales to any number of examples, since the posterior resulting from updating with all samples simultaneously is equal to the posterior obtained by chaining the samples by applying a sample, taking the posterior, and treating it as the prior for the next sample, throughout all the samples.

(ii)

For the conjugate model to belong to the exponential family, the following must be true:

$$p(\pi) \propto G(\pi)^\eta \exp(\phi(\pi)\,\nu)$$
$$p(\omega|\pi) = F(\omega)G(\pi)\exp(\phi(\pi)U(\omega))$$

The prior is:

$p(\pi) = \frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha,\beta)} \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}$, and the likelihood of observing some sample given μ is:

$$p(\omega|\pi) = \begin{cases} (1-\pi)^\omega \pi & \omega \geq 0 \\ 0 & \text{else} \end{cases}$$

First, the probability of ω conditioned on π for $0 \leq \pi$ will be examined:

$$p(\omega|\pi) = (1-\pi)^{\omega}\pi$$
$$= \pi \exp(\ln(1-\pi)\omega)$$

Let $G(\pi) = \pi, \phi(\pi) = \ln(1-\pi), F(\omega) = 1, U(\omega) = \omega$:

$$p(\omega|\pi) = G(\pi)\exp(\phi(\pi)U(\omega))$$

Now the prior $p(\pi)$ must be constructed using these equations:

$$p(\pi) = \pi^{\alpha-1}(1-\pi)^{\beta-1} \propto G(\pi)^{\eta}\exp(\phi(\pi)\,\nu) = \pi^{\eta}\exp(\ln(1-\pi)\,\nu)$$

Assume $\eta = \alpha - 1$:

$$p(\pi) \propto \pi^{\alpha-1}\exp(\ln(1-\pi)\,\nu)$$
$$= \pi^{\alpha-1}(1-\pi)^{\nu}$$

Assume $\nu = \beta - 1$:

$$p(\pi) \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}$$

This is of the correct form for the input beta distribution, with $p(\pi) = C \cdot \pi^{\alpha-1}(1-\pi)^{\beta-1}$. Therefore the beta conjugate model from problem 1 is part of the exponential family. As before, while only a single sample was examined for the likelihood function, this is equivilent to evaluating the likelihood for an arbitrary number of samples.