Bayesian inference in simple conjugate families

A) Let $\mathbf{x} := \langle x_1, \ldots, x_N \rangle$, and let k denote the number of successes in $\mathbf{x}$. Then

$$\begin{aligned} p(w|\mathbf{x}) &\propto p(\mathbf{x}|w)p(w) \\ &\propto w^k(1-w)^{N-k}w^{a-1}(1-w)^{b-1} \\ &= w^{a+k-1}(1-w)^{b+(n-k)-1}, \end{aligned}$$

which is the kernel of a Beta distribution with parameters $a+k-1$ and $b+(n-k)-1$. Thus, the posterior distribution is $\text{Beta}\big(a+k-1, b+(n-k)-1\big)$.

B) Let

$$Y_1 = \frac{X_1}{X_1+X_2} \qquad \text{and} \qquad Y_2 = X_1 + X_2.$$

Then, solving for X_1 and X_2 in terms of Y_1 and Y_2, we get

$$X_1 = Y_1Y_2 \qquad \text{and} \qquad X_2 = (1-Y_1)Y_2.$$

Since we want to find the joint density of Y_1 and Y_2, we start by finding the Jacobian of the transformation:

$$\begin{aligned} J &= \begin{vmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 \end{vmatrix} \\ &= \begin{vmatrix} y_2 & y_1 \\ -y_2 & 1-y_1 \end{vmatrix} \\ &= y_2(1-y_1) + y_1y_2 \\ &= y_2. \end{aligned}$$

Note that since X_1 and X_2 are never negative, neither is Y_2, and therefore $|J| = J$. We can now express the joint distribution function in terms of f_{X_1,X_2} as follows:

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &\propto (y_1y_2)^{a_1-1}\cdot e^{-y_1y_2}\big((1-y_1)y_2\big)^{a_2-1}\cdot e^{-(1-y_1)y_2}\cdot y_2 \\ &= \left(y_1^{a_1-1}(1-y_1)^{a_2-1}\right)\left(y_2^{a_1+a_2-1}\cdot e^{-y_2}\right) \end{aligned}$$

where the expression in the first pair of parentheses is the kernel of a Beta distribution and that in the second is the kernel of a Gamma distribution. We therefore see that

$$y_1 \sim \text{Beta}(a_1, a_2) \qquad \text{and} \qquad y_2 \sim \text{Gamma}(a_1+a_2, 1).$$

This provides us with a means to simulate a Beta random variable with parameters a_1 and a_2 from Gamma random variables by simulating $X_1 \sim \text{Gamma}(a_1, 1)$ and $X_2 \sim \text{Gamma}(a_2, 1)$, and then computing the ratio $X_1/(X_1+X_2)$.

C) Let $X_i \sim N(\theta, \sigma^2)$ be independent where the variance is known and $\theta \sim N(m, v)$. Let $\bar{x}$ denote the mean of $x_1, \ldots, x_N$. Then

$$
\begin{aligned}
p(\theta|x_1,\ldots,x_N) &\propto p(x_1,\ldots,x_N|\theta)p(\theta) \\
&= \left(\prod_{i=1}^{N} p(x_i|\theta)\right)p(\theta) \\
&\propto \exp\left(-\sum_{i=1}^{N}\frac{(x_i-\theta)^2}{2\sigma^2}\right)\exp\left(\frac{-(\theta-m)^2}{2v}\right) \\
&\propto \exp\left(\sum_{i=1}^{N}\frac{2\theta x_i-\theta^2}{2\sigma^2}\right)\exp\left(-\frac{\theta^2-2\theta m}{2v}\right) \\
&= \exp\left(\frac{2\theta n\bar{x}-n\theta^2}{2\sigma^2}-\frac{\theta^2-2\theta m}{2v}\right) \\
&= \exp\left(-\left(\frac{n}{2\sigma^2}+\frac{1}{2v}\right)\theta^2+\left(\frac{n\bar{x}}{\sigma^2}+\frac{m}{v}\right)\theta\right)
\end{aligned}
$$

This last expression is complicated, but has the form $e^{-A\theta^2+B\theta}$, with which we can work until it is further simplified. Note that we still only care about things proportional to the expression involving θ.

$$
\begin{aligned}
\exp\left(-A\theta^2+B\theta\right) &= \exp\left(\frac{-\theta^2+(B/A)\theta}{1/A}\right) \\
&\propto \exp\left(\frac{-\left(\theta-(B/2A)\right)^2}{1/A}\right)
\end{aligned}
$$

This is the kernel of a normal distribution with mean and variance as follows:

$$
\mu_{\text{post}} = \frac{B}{2A} = \frac{\dfrac{n\bar{x}}{\sigma^2}+\dfrac{m}{v}}{\dfrac{n}{\sigma^2}+\dfrac{1}{v}}
$$

$$
\sigma^2_{\text{post}} = \frac{1}{2A} = \frac{1}{\dfrac{n}{\sigma^2}+\dfrac{1}{v}}
$$

D) Let $X_i \sim N(\theta, \sigma^2)$ be independent where the mean θ is known and $w := 1/\sigma^2 \sim \text{Gamma}(a, b)$. Let $\hat{\sigma}^2$ denote the standard deviation of $x_1, \ldots, x_N$. Then

$$\begin{aligned}
p(w|x_1, \ldots, x_N) &\propto p(x_1, \ldots, x_N|w)p(w) \\
&= \left(\prod_{i=1}^{N} p(x_i|w)\right) p(w) \\
&\propto w^{1/2} \exp\left(\frac{-w}{2} \sum_{i=1}^{N} (x_i - \theta)^2\right) w^{a-1} \exp(-bw) \\
&= w^{a-1/2} \exp\big(-w \cdot N\hat{\sigma}^2/2\big) \exp(-bw) \\
&= w^{a-1/2} \exp\big(-(N\hat{\sigma}^2/2 + b)w\big).
\end{aligned}$$

This is the kernel of a Gamma distribution, namely $\text{Gamma}\big(a + 1/2, b + N\hat{\sigma}^2/2\big)$. Thus, the posterior distribution for σ^2 is $\text{IG}\big(a + 1/2, b + N\hat{\sigma}^2/2\big)$

E) Let $X_i \sim N(\theta, \sigma_i^2)$ be independent where each σ_i^2 is known and $\theta \sim N(m, v)$. Then

$$\begin{aligned}
p(\theta|x_1, \ldots, x_N) &\propto p(x_1, \ldots, x_N|\theta)p(\theta) \\
&= \left(\prod_{i=1}^{N} p(x_i|\theta)\right) p(\theta) \\
&\propto \exp\left(-\sum_{i=1}^{N} \frac{(x_i - \theta)^2}{2\sigma_i^2}\right) \exp\left(\frac{-(\theta - m)^2}{2v}\right) \\
&\propto \exp\left(-\sum_{i=1}^{N} \frac{\theta^2 - 2x_i\theta}{\sigma_i^2}\right) \exp\left(\frac{-(\theta^2 - 2m\theta)}{2v}\right) \\
&= \exp\left(-\frac{\theta^2 - 2m\theta}{2v} - \sum_{i=1}^{N} \frac{\theta^2 - 2x_i\theta}{\sigma_i^2}\right) \\
&= \exp\left(-\left(\frac{1}{2v} + \sum_{i=1}^{N} \frac{1}{2\sigma_i^2}\right)\theta^2 + \left(\frac{m}{v} + \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}\right)\theta\right)
\end{aligned}$$

As in (C) above, this has the form $e^{-A\theta^2 + B\theta}$, and we solve this in the same way. Thereby we find that $\theta|x_1, \ldots, x_N$ is normally distributed with mean and variance given by

$$\mu_{\text{post}} = \frac{\dfrac{m}{v} + \displaystyle\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\dfrac{1}{v} + \displaystyle\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

$$\sigma^2_{\text{post}} = \frac{1}{\dfrac{1}{v} + \displaystyle\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

F)

The multivariate normal distribution

A)

$$
\begin{aligned}
\operatorname{cov}(x) &:= E\big((x-\mu)(x-\mu)^T\big) \\
&= E\big((x-\mu)(x^T-\mu^T)\big) \\
&= E\big(xx^T - \mu x^T - x\mu^T + \mu\mu^T\big) \\
&= E\big(xx^T\big) - E\big(\mu x^T\big) - E\big(x\mu^T\big) + E\big(\mu\mu^T\big) \\
&= E\big(xx^T\big) - \mu E\big(x^T\big) - E\big(x\big)\mu^T + \mu\mu^T E(I_N) \\
&= E\big(xx^T\big) - \mu\mu^T - \mu\mu^T + \mu\mu^T \\
&= E\big(xx^T\big) - \mu\mu^T
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{cov}(Ax+b) &:= E\Big(\big(Ax+b-E(Ax+b)\big)\big(Ax+b-E(Ax+b)\big)^T\Big) \\
&= E\Big(\big(Ax-E(Ax)\big)\big(Ax-E(Ax)\big)^T\Big) \\
&= E\Big(A(x-\mu)\big(A(x-\mu)\big)^T\Big) \\
&= E\Big(A(x-\mu)(x-\mu)^T A^T\Big) \\
&= AE\Big((x-\mu)(x-\mu)^T\Big)A^T \\
&= A\operatorname{cov}(x)A^T
\end{aligned}
$$