Conjugate prior derivations

Contents

1 Binomial probability

The model is $Y|\theta \sim \text{Binomial}(n, \theta)$ with prior $\theta \sim \text{Beta}(a, b)$.

1.1 Retaining the normalizing constants

In this derivation we retain all terms including those that do not depend on θ. The term in square brackets in the last line is the normalizing constant of the posterior, i.e., the term that does not depend on θ but is required so the posterior integrates to one. The posterior is

$$
\begin{aligned}
p(\theta|Y) &= \frac{f(Y|\theta)\pi(\theta)}{m(Y)} && (1)\\
&= \frac{\left\{\binom{n}{Y}\theta^Y(1-\theta)^{n-Y}\right\}\left\{\frac{\Gamma(a,b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\right\}}{m(Y)} && (2)\\
&= \left[\frac{\binom{n}{Y}\Gamma(a,b)}{\Gamma(a)\Gamma(b)m(Y)}\right]\theta^{Y+a-1}(1-\theta)^{n-Y+b-1} && (3)\\
&= C\theta^{A-1}(1-\theta)^{B-1}, && (4)
\end{aligned}
$$

where $A = Y + a$, $B = n - Y + b$ and C is the terms inside the square brackets. The kernel (i.e., the terms that depend on θ) of this distribution has the form of a Beta(A,B) distribution. Therefore, $\theta|Y \sim \text{Beta}(Y + a, n - Y + b)$.

1.2 Dropping the normalizing constants

In the end, we will ignore the normalizing constant, so to clean up the derivation we ignore terms that do not include θ at each step.

$$
\begin{aligned}
p(\theta|Y) &\propto f(Y|\theta)\pi(\theta) && (5)\\
&\propto \left\{\theta^Y(1-\theta)^{n-Y}\right\}\left\{\theta^{a-1}(1-\theta)^{b-1}\right\} && (6)\\
&\propto \theta^{Y+a-1}(1-\theta)^{n-Y+b-1} && (7)\\
&\propto \theta^{A-1}(1-\theta)^{B-1}. && (8)
\end{aligned}
$$

Therefore, $\theta|Y \sim \text{Beta}(Y + a, n - Y + b)$. Note the every derivation will have the same first line, so below we often start with $p(\theta|Y) \propto \{\}\{\}$ where the expressions in brackets are the likelihood and prior.

2 Poisson rate

The model is $Y_1, ..., Y_n|\theta \overset{iid}{\sim}$ Poisson(θ) with prior $\theta \sim$ Gamma(a, b). The posterior is

$$\begin{aligned} p(\theta|\mathbf{Y}) &\propto \left\{\prod_{i=1}^{n} f(Y_i|\theta)\right\} \pi(\theta) && (9) \\ &\propto \left\{\prod_{i=1}^{n} \theta^{Y_i} \exp(-\theta)\right\} \left\{\theta^{a-1} \exp(-b\theta)\right\} && (10) \\ &\propto \theta^{\sum_{i=1}^{n} Y_i + a - 1} \exp\{-(n+b)\theta\} && (11) \\ &\propto \theta^{A-1} \exp(-B\theta), && (12) \end{aligned}$$

for $A = \sum_{i=1}^{n} Y_i + a$ and $B = n + b$. This is the kernel of a Gamma(A, B) distribution. Therefore, $\theta|\mathbf{Y} \sim$ Gamma$(\sum_{i=1}^{n} Y_i + a, n + b)$.

A related problem is $Y|\theta \sim$ Poisson$(n\theta)$ and $\theta \sim$ Gamma(a, b), which gives posterior $\theta|Y \sim$ Gamma$(Y + a, n + b)$.

3 Negative-binomial probability

The model is $Y|\theta \sim$ NegBinomial(r, θ) with prior $\theta \sim$ Beta(a, b). Then

$$\begin{aligned} p(\theta|Y) &\propto \left\{\theta^{Y}(1-\theta)^{r}\right\}\left\{\theta^{a-1}(1-\theta)^{b-1}\right\} && (13) \\ &\propto \theta^{Y+a-1}(1-\theta)^{r+b-1} && (14) \\ &\propto \theta^{A-1}(1-\theta)^{B-1}. && (15) \end{aligned}$$

Therefore, $\theta|Y \sim$ Beta$(Y + a, r + b)$.

4 Multinomial probabilities

Say $\mathbf{Y} = (Y_1, ..., Y_p)$ is a vector of counts with Y_j equal to the number observations in group j. The vector of unknown probabilities is $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$. The model is $\mathbf{Y}|\boldsymbol{\theta} \sim$ Multinomial$(\theta_1, ..., \theta_p)$ with prior $\boldsymbol{\theta} \sim$ Dirichlet$(\alpha_1, ..., \alpha_p)$. Then

$$p(\boldsymbol{\theta}|\mathbf{Y}) \propto \left\{\prod_{j=1}^{p} \theta_j^{Y_j}\right\} \left\{\prod_{j=1}^{p} \theta_j^{\alpha_j - 1}\right\} \propto \prod_{j=1}^{p} \theta_j^{Y_j + \alpha_j - 1} \propto \prod_{j=1}^{p} \theta_j^{A_j - 1} \tag{16}$$

for $A_j = Y_j + \alpha_j$, so $\boldsymbol{\theta}|\mathbf{Y} \sim$ Dirichlet$(Y_1 + \alpha_1, ..., Y_p + \alpha_p)$.

5 Normal mean

The model is $Y_i|\mu \overset{iid}{\sim} \text{Normal}(\mu, \sigma)^2$ with σ^2 known. Below we derive the posterior of the mean parameter μ three times. First we start with improper prior so the math is cleaner. In this case, we complete the square in the exponent to obtain the posterior. We then repeat this derivation without completing the square by working backwards from the Gaussian PDF. Finally, we add a prior for μ as is customary.

5.1 Improper prior

Assume prior $\pi(\mu) = 1$ for all μ

5.1.1 Completing the square

$$
\begin{aligned}
p(\mu|\mathbf{Y}) &\propto \left\{\prod_{i=1}^{n} \exp\left[-\frac{1}{2\sigma^2}(Y_i - \mu)^2\right]\right\}\{1\} && (17)\\
&\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i^2 - 2Y_i\mu + \mu^2)\right] && (18)\\
&\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2Y_i\mu + \mu^2)\right] && (19)\\
&\propto \exp\left[-\frac{1}{2\sigma^2}(-2n\bar{Y}\mu + n\mu^2)\right] && (20)\\
&\propto \exp\left[-\frac{n}{2\sigma^2}(-2\bar{Y}\mu + \mu^2)\right] && (21)\\
&\propto \exp\left[-\frac{n}{2\sigma^2}(-\bar{Y}^2 + \bar{Y}^2 - 2\bar{Y}\mu + \mu^2)\right] && (22)\\
&\propto \exp\left[-\frac{n}{2\sigma^2}(\bar{Y}^2 - 2\bar{Y}\mu + \mu^2)\right] && (23)\\
&\propto \exp\left[-\frac{1}{2(\sigma^2/n)}(\mu - \bar{Y})^2\right]. && (24)
\end{aligned}
$$

Therefore, $\mu|\mathbf{Y} \sim \text{Normal}(\bar{Y}, \sigma^2/n)$.

5.1.2 Matching components

If $\mu \sim \text{Normal}(M, V)$ then

$$p(\mu) \propto \exp\left[-\frac{1}{2V}(\mu - M)^2\right] \quad (25)$$

$$\propto \exp\left[-\frac{1}{2}\left(-2\frac{M}{V}\mu - \frac{1}{V}\mu^2\right)\right]. \quad (26)$$

So if we get to the point where the posterior has the form

$$p(\mu|\mathbf{Y}) \propto \exp\left[-\frac{1}{2}\left(-2A\mu - B\mu^2\right)\right]. \quad (27)$$

we can solve the equations $A = M/V$ and $B = 1/V$ for the posterior mean M and variance V. This gives $V = 1/B$ and $M = A/B$. So picking up in the middle of the derivation above

$$p(\mu|\mathbf{Y}) \propto \exp\left[-\frac{1}{2\sigma^2}\left(-2n\bar{Y}\mu + n\mu^2\right)\right] \quad (28)$$

$$\propto \exp\left[-\frac{1}{2}\left(-2A\mu + B\mu^2\right)\right] \quad (29)$$

with $A = n\bar{Y}/\sigma^2$ and $B = n/\sigma^2$, so $M = A/B = \bar{Y}$ and $V = 1/B = \sigma^2/n$ and thus as before

$$\mu|\mathbf{Y} \sim \text{Normal}(\bar{Y}, \sigma^2/n).$$

5.2 Proper prior

Now assume prior $\mu \sim \text{Normal}(\mu_0, \tau^2)$. Then

$$p(\mu|\mathbf{Y}) \propto \left\{\prod_{i=1}^{n} \exp\left[\frac{-1}{2\sigma^2}(Y_i-\mu)^2\right]\right\}\left\{\exp\left[-\frac{1}{\tau^2}(\mu-\mu_0)^2\right]\right\} \tag{30}$$

$$\propto \exp\left[-\frac{1}{2}\left(-2\frac{n\bar{Y}}{\sigma^2}\mu + \frac{n}{\sigma^2}\mu^2 - 2\frac{\mu_0}{\tau^2}\mu + \frac{1}{\tau^2}\mu^2\right)\right] \tag{31}$$

$$\propto \exp\left[-\frac{1}{2}\left(-2A\mu + B\mu^2\right)\right], \tag{32}$$

where $A = n\bar{Y}/\sigma^2 + \mu_0/\tau^2$ and $B = n/\sigma^2 + 1/\tau^2$. Therefore,

$$\mu|\mathbf{Y} \sim \text{Normal}\left(\frac{n\bar{Y}/\sigma^2 + \mu_0/\tau^2}{n/\sigma^2 + 1/\tau^2}, \frac{1}{n/\sigma^2 + 1/\tau^2}\right).$$

6 Normal variance

The model is $Y_i|\theta \stackrel{iid}{\sim} \text{Normal}(\mu_i, \sigma^2)$ with all μ_i known (e.g., in linear regression $\mu_i = \mathbf{X}_i\boldsymbol{\beta}$) and prior $\sigma^2 \sim \text{InvGamma}(a, b)$. Then

$$p(\sigma^2|\mathbf{Y}) \propto \left\{\prod_{i=1}^n (\sigma^2)^{-1/2} \exp\left[\frac{-1}{2\sigma^2}(Y_i - \mu_i)^2\right]\right\} \left\{(\sigma^2)^{-a-1} \exp\left(-\frac{b}{\sigma^2}\right)\right\} \quad (33)$$

$$\propto (\sigma^2)^{-n/2-a-1} \exp\left[-\frac{\sum_{i=1}^n (Y_i - \mu_i)^2/2 + b}{\sigma^2}\right]. \quad (34)$$

Therefore, $\sigma^2|\mathbf{Y} \sim \text{InvGamma}\left(n/2 + a, \sum_{i=1}^n (Y_i - \mu_i)^2/2 + b\right)$.

7 Normal precision

The precision is the inverse variance, i.e., $\tau = 1/\sigma^2$. With this parameterization, the model is $Y_i|\theta \stackrel{iid}{\sim} \text{Normal}(\mu_i, \tau^{-1})$ with all μ_i known and prior $\tau \sim \text{Gamma}(a, b)$. Then

$$p(\tau|\mathbf{Y}) \propto \left\{\prod_{i=1}^n \tau^{1/2} \exp\left[-\frac{\tau}{2}(Y_i - \mu_i)^2\right]\right\} \left\{\tau^{a+1} \exp\left(-b\tau\right)\right\} \quad (35)$$

$$\propto \tau^{n/2+a+1} \exp\left\{-\left[\sum_{i=1}^n (Y_i - \mu_i)^2/2 + b\right]\tau\right\}. \quad (36)$$

Therefore, $\tau|\mathbf{Y} \sim \text{Gamma}\left(n/2 + a, \sum_{i=1}^n (Y_i - \mu_i)^2/2 + b\right)$.

8 Gaussian linear regression

The model is $\mathbf{Y}|\boldsymbol{\beta} \sim \text{Normal}(\mathbf{X}\boldsymbol{\beta}, \Sigma)$ for $\mathbf{Y} = (Y_1, ..., Y_n)^T$, $\mathbf{X}$ is a known $n \times p$ matrix, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ is an unknown vector of regression coefficients and Σ is a known $n \times n$ covariance matrix with inverse $\Omega = \Sigma^{-1}$. The prior is $\boldsymbol{\beta} \sim \text{Normal}(\boldsymbol{\beta}_0, \Delta^{-1})$.

Before starting, note that if $\boldsymbol{\beta}$ has posterior mean vector $\mathbf{M}$ and covariance $\mathbf{V}$, then

$$p(\boldsymbol{\beta}|\mathbf{Y}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{M})^T\mathbf{V}^{-1}(\boldsymbol{\beta} - \mathbf{M})\right] \propto \exp\left[-\frac{1}{2}(-2\mathbf{A}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{B}\boldsymbol{\beta})\right]$$

for $\mathbf{A} = \mathbf{M}^T\mathbf{V}^{-1}$ and $\mathbf{B} = \mathbf{V}^{-1}$. Solving gives $\mathbf{V} = \mathbf{B}^{-1}$ and $\mathbf{M} = \mathbf{B}^{-1}\mathbf{A}$.

The posterior of $\boldsymbol{\beta}$ is

$$p(\boldsymbol{\beta}|\mathbf{Y}) \propto \left\{\exp\left[-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\Omega(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right]\right\}\left\{\exp\left[-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T\Delta(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right]\right\} \quad (37)$$

$$\propto \exp\left\{-\frac{1}{2}\left[-2(\mathbf{Y}^T\Omega\mathbf{X} + \beta_0^T\Delta)\boldsymbol{\beta} + \boldsymbol{\beta}^T(\mathbf{X}^T\Omega\mathbf{X} + \Delta)\boldsymbol{\beta}\right]\right\}. \quad (38)$$

Letting $\mathbf{A} = \mathbf{Y}^T\Omega\mathbf{X} + \boldsymbol{\beta}_0^T\Delta$ and $\mathbf{B} = \mathbf{X}^T\Omega\mathbf{X} + \Delta$ gives

$$\boldsymbol{\beta}|\mathbf{Y} \sim \text{Normal}\left[(\mathbf{X}^T\Omega\mathbf{X} + \Delta)^{-1}(\mathbf{X}^T\Omega\mathbf{Y} + \Delta\boldsymbol{\beta}_0), (\mathbf{X}^T\Omega\mathbf{X} + \Delta)^{-1}\right].$$

9 Multivariate normal precision matrix

The model for vectors $\mathbf{Y}_i = (Y_{i1}, ..., Y_{ip})^T$ is $\mathbf{Y}_i|\Omega \stackrel{indep}{\sim} \text{Normal}(\boldsymbol{\mu}_i, \Omega^{-1})$ for known mean vectors $\boldsymbol{\mu}_i$ and unknown $p \times p$ precision (inverse covariance) matrix Ω. The prior is $\Omega \sim \text{Wishart}(\nu, \mathbf{V})$ for scalar $\nu > p - 1$ and symmetric positive definite matrix $\mathbf{V}$. Defining $\mathbf{R}_i = \mathbf{Y}_i - \boldsymbol{\mu}_i$, the posterior is

$$
\begin{aligned}
p(\Omega|\mathbf{Y}) &\propto \left\{\prod_{i=1}^n |\Omega|^{1/2} \exp\left[-\frac{1}{2}\mathbf{R}_i^T\Omega\mathbf{R}_i\right]\right\}\left\{|\Omega|^{(\nu-p-1)/2}\exp\left[-\frac{1}{2}\text{tr}(\mathbf{V}^{-1}\Omega)\right]\right\} && (39)\\
&\propto |\Omega|^{(n+\nu-p-1)/2}\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n \mathbf{R}_i^T\Omega\mathbf{R}_i + \text{tr}(\mathbf{V}^{-1}\Omega)\right]\right\} && (40)\\
&\propto |\Omega|^{(n+\nu-p-1)/2}\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n \text{tr}[\mathbf{R}_i^T\Omega\mathbf{R}_i] + \text{tr}(\mathbf{V}^{-1}\Omega)\right]\right\} && (41)\\
&\propto |\Omega|^{(n+\nu-p-1)/2}\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n \text{tr}[\Omega\mathbf{R}_i\mathbf{R}_i^T] + \text{tr}(\mathbf{V}^{-1}\Omega)\right]\right\} && (42)\\
&\propto |\Omega|^{(n+\nu-p-1)/2}\exp\left\{-\frac{1}{2}\left[\text{tr}[\Omega\sum_{i=1}^n \mathbf{R}_i\mathbf{R}_i^T] + \text{tr}(\Omega\mathbf{V}^{-1})\right]\right\} && (43)\\
&\propto |\Omega|^{(n+\nu-p-1)/2}\exp\left\{-\frac{1}{2}\text{tr}\left[\Omega\left(\sum_{i=1}^n \mathbf{R}_i\mathbf{R}_i^T + \mathbf{V}^{-1}\right)\right]\right\}. && (44)
\end{aligned}
$$

Therefore,

$$
\Omega|\mathbf{Y} \sim \text{Wishart}\left\{n+\nu, \left[\sum_{i=1}^n \mathbf{R}_i\mathbf{R}_i^T + \mathbf{V}^{-1}\right]^{-1}\right\}.
$$

10 Multivariate normal covariance matrix

The model for vectors $\mathbf{Y}_i = (Y_{i1}, ..., Y_{ip})^T$ is $\mathbf{Y}_i|\Sigma \stackrel{indep}{\sim} \text{Normal}(\boldsymbol{\mu}_i, \Sigma)$ for known mean vectors $\boldsymbol{\mu}_i$ and unknown $p \times p$ covariance matrix Ω. The prior is $\Sigma \sim \text{InvWishart}(\nu, \mathbf{U})$ for scalar $\nu > p - 1$ and symmetric positive definite matrix $\mathbf{U}$. Defining $\mathbf{R}_i = \mathbf{Y}_i - \boldsymbol{\mu}_i$, the posterior is (applying many of the steps from the precision derivation above)

$$p(\Sigma|\mathbf{Y}) \propto \left\{\prod_{i=1}^{n} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{R}_i^T\Sigma^{-1}\mathbf{R}_i\right]\right\}\left\{|\Sigma|^{-(\nu+p+1)/2} \exp\left[-\frac{1}{2}\text{trace}(\mathbf{U}\Sigma^{-1})\right]\right\} \quad (45)$$

$$\propto |\Sigma|^{-(n+\nu+p+1)/2} \exp\left\{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}\left(\sum_{i=1}^{n}\mathbf{R}_i\mathbf{R}_i^T + \mathbf{U}\right)\right]\right\} \quad (46)$$

Therefore, $\Sigma|\mathbf{Y} \sim \text{InvWishart}[n + \nu, \sum_{i=1}^{n} \mathbf{R}_i\mathbf{R}_i^T + \mathbf{U}]$.