Generative Bayesian Hierarchical Models for Individual Observations

Wagner L. Truppel

March 2000

1 Assumptions

- $N \geq 1$ individuals, indexed by i, $1 \leq i \leq N$;
- m_i observations $\{x_{ij}\}$ for the i-th individual, indexed by j, $1 \leq j \leq m_i$;
- For fixed i, the observations $\{x_{ij}\}$ are realizations of independently distributed random variables X_i, with probability distribution $p(X_i|\vec{\theta}_i)$;
- The parameters $\{\vec{\theta}_i\}$ are realizations of independently and identically distributed random (vector) variables $\vec{\Theta}_i$, with probability distribution $p(\vec{\Theta}|\vec{\Psi})$;
- The prior distribution of the parameters $\vec{\Psi}$ is assumed to be uninformative.

2 General results

The likelihood of the data $D \equiv \{x_{ij}, 1 \leq i \leq N, 1 \leq j \leq m_i\}$, given all the parameters, is

$$L(D|\{\vec{\theta}_i\}, \vec{\Psi}) = p(D|\{\vec{\theta}_i\}, \vec{\Psi}) = \prod_{i=1}^{N} \prod_{j=1}^{m_i} p(X_i = x_{ij}|\vec{\theta}_i)\,, \tag{1}$$

so the posterior probability of the parameters, given the data, is

$$p(\{\vec{\theta}_i\}, \vec{\Psi}|D) = \frac{p(D|\{\vec{\theta}_i\}, \vec{\Psi})\, p(\{\vec{\theta}_i\}|\vec{\Psi})\, p(\vec{\Psi})}{p(D)} = \frac{p(\vec{\Psi})}{p(D)} \prod_{i=1}^{N} \left\{ p(\vec{\Theta} = \vec{\theta}_i|\vec{\Psi}) \prod_{j=1}^{m_i} p(X_i = x_{ij}|\vec{\theta}_i) \right\}. \tag{2}$$

3 Specific model: Fixed-variance, Normally-distributed observations, with a conjugate prior for the mean

Assume that the observations for a given individual are normally distributed around some mean value μ_i, with some variance σ_i^2, characteristic of that individual. Thus, $\vec{\theta}_i \equiv \{\mu_i, \sigma_i^2\}$ and

$$p(X_i = x_{ij}|\vec{\theta}_i) \equiv N(x_{ij}|\mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\Big[-\frac{1}{2\sigma_i^2}(x_{ij} - \mu_i)^2\Big], \tag{3}$$

with $\sigma_i^2 > 0$.

We must now choose a probability distribution for the μ_i's and σ_i^2's. Since the Normal distribution is its own conjugate, we'll **assume** that the μ_i's are independently and identically distributed according to a Normal distribution of mean α and variance β^2. We'll also **assume** that the μ_i's are independent from σ_i^2 and that all variances are equal, $\sigma_i^2 = \sigma_0^2$. So, $\vec{\Psi} \equiv \{\alpha, \beta^2\}$, and

$$p(\vec{\Theta} = \vec{\theta}_i|\vec{\Psi}) \equiv \delta(\sigma_i - \sigma_0)\, N(\mu_i|\alpha, \beta^2) = \frac{\delta(\sigma_i - \sigma_0)}{\sqrt{2\pi\beta^2}} \exp\Big[-\frac{1}{2\beta^2}(\mu_i - \alpha)^2\Big], \tag{4}$$

with $\beta^2 > 0$. The delta function accounts for the fact that every σ_i^2 is fixed at σ_0^2.

Combining the results above, the posterior joint probability of the various parameters, given the data, is then

$$p(\{\mu_i, \sigma_i^2\}, \alpha, \beta^2|D) = \frac{p(\alpha, \beta^2)}{p(D)} \prod_{i=1}^{N} \delta(\sigma_i - \sigma_0)\, N(\mu_i|\alpha, \beta^2) \prod_{j=1}^{m_i} N(x_{ij}|\mu_i, \sigma_i^2)\,. \tag{5}$$

Since we're fixing the individual variances, we may as well integrate over all the σ_i's to obtain

$$\begin{aligned}
p(\{\mu_i\}, \alpha, \beta^2|D) &= \frac{p(\alpha, \beta^2)}{p(D)} \prod_{i=1}^{N} N(\mu_i|\alpha, \beta^2) \prod_{j=1}^{m_i} N(x_{ij}|\mu_i, \sigma_0^2) \\
&= \frac{p(\alpha, \beta^2)}{p(D)} \prod_{i=1}^{N} \frac{\tau_i}{\beta} (2\pi\sigma_0^2)^{-m_i/2} \exp\left[-\frac{m_i\, \tau_i^2}{2\beta^2\sigma_0^2}(\alpha - \bar{\mu}_i)^2 - \frac{m_i\, \bar{\sigma}_i^2}{2\sigma_0^2}\right] N(\mu_i|\lambda_i, \tau_i^2) \\
&= \frac{p(\alpha, \beta^2)}{p(D)} \exp\left\{-\frac{1}{2\beta^2\sigma_0^2} \sum_{i=1}^{N} m_i \Big[\tau_i^2\, (\alpha - \bar{\mu}_i)^2 + \beta^2\, \bar{\sigma}_i^2\Big]\right\} \times \\
&\quad \times \prod_{i=1}^{N} \frac{\tau_i}{\beta} (2\pi\sigma_0^2)^{-m_i/2}\, N(\mu_i|\lambda_i, \tau_i^2)\,,
\end{aligned} \tag{6}$$

where

$$\lambda_i \equiv \tau_i^2 \left[\frac{\alpha}{\beta^2} + \frac{\bar{\mu}_i}{(\sigma_0^2/m_i)} \right], \quad \tau_i^2 \equiv \left[\frac{1}{\beta^2} + \frac{1}{(\sigma_0^2/m_i)} \right]^{-1}, \tag{7}$$
$$\bar{\mu}_i \equiv \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ij}\,, \quad \text{and} \quad \bar{\sigma}_i^2 \equiv \frac{1}{m_i} \sum_{j=1}^{m_i} (x_{ij} - \bar{\mu}_i)^2\,.$$

Note that $\bar{\mu}_i$ and $\bar{\sigma}_i^2$ are the sample mean and sample variance of the observations for the i-th individual. Now,

$$\sum_{i=1}^{N} m_i \Big[\tau_i^2 \,(\alpha - \bar{\mu}_i)^2 + \beta^2 \,\bar{\sigma}_i^2 \Big] = Z \Big[(\alpha - Z_{\bar{\mu}})^2 + Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2 \Big] + \beta^2 \sum_{i=1}^{N} m_i \,\bar{\sigma}_i^2\,, \tag{8}$$

where

$$Z \equiv \sum_{i=1}^{N} m_i \,\tau_i^2\,, \quad Z_{\bar{\mu}} \equiv \frac{1}{Z} \sum_{i=1}^{N} m_i \,\tau_i^2 \,\bar{\mu}_i\,, \quad \text{and} \quad Z_{\bar{\mu}^2} \equiv \frac{1}{Z} \sum_{i=1}^{N} m_i \,\tau_i^2 \,\bar{\mu}_i^2 \tag{9}$$

are all functions of β^2, through the various τ_i's, but are not functions of α. Combining all the results obtained so far, we find

$$p(\{\mu_i\}, \alpha, \beta^2 | D) = \frac{p(\alpha, \beta^2)}{p(D)} \sqrt{2\pi \,\frac{\beta^2 \sigma_0^2}{Z}} \,\exp\left\{ - \frac{Z}{2\beta^2 \sigma_0^2} \Big[Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2 \Big] - \frac{1}{2\sigma_0^2} \sum_{i=1}^{N} m_i \,\bar{\sigma}_i^2 \right\} \times \tag{10}$$
$$\times \left[\prod_{i=1}^{N} \frac{\tau_i}{\beta} \,(2\pi\sigma_0^2)^{-m_i/2} \right] N(\alpha | Z_{\bar{\mu}}, \frac{\beta^2 \sigma_0^2}{Z}) \prod_{i=1}^{N} N(\mu_i | \lambda_i, \tau_i^2)\,.$$

From this, it's easy to obtain the joint posterior distribution of α and β^2, given the data, by integrating over all the μ_i's:

$$p(\alpha, \beta^2 | D) = \frac{p(\alpha, \beta^2)}{p(D)} \sqrt{2\pi \,\frac{\beta^2 \sigma_0^2}{Z}} \,\exp\left\{ - \frac{Z}{2\beta^2 \sigma_0^2} \Big[Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2 \Big] - \frac{1}{2\sigma_0^2} \sum_{i=1}^{N} m_i \,\bar{\sigma}_i^2 \right\} \times \tag{11}$$
$$\times \; N(\alpha | Z_{\bar{\mu}}, \frac{\beta^2 \sigma_0^2}{Z}) \prod_{i=1}^{N} \frac{\tau_i}{\beta} \,(2\pi\sigma_0^2)^{-m_i/2}\,.$$

Alternatively, we may integrate Eq. (**??**) over all the μ_i's but one, say, μ_k, and isolate all the α dependence in one term, to obtain

$$p(\mu_k, \alpha, \beta^2 | D) = \frac{p(\alpha, \beta^2)}{p(D)} \sqrt{2\pi \,\frac{\beta^2 \sigma_0^2}{Z}} \,\exp\left\{ - \frac{Z}{2\beta^2 \sigma_0^2} \Big[Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2 \Big] - \frac{1}{2\sigma_0^2} \sum_{i=1}^{N} m_i \,\bar{\sigma}_i^2 \right\} \times \tag{12}$$
$$\times \left[\prod_{i=1}^{N} \frac{\tau_i}{\beta} \,(2\pi\sigma_0^2)^{-m_i/2} \right] N(\alpha | \tilde{\alpha}_k, \frac{1}{\delta_k} \frac{\beta^2 \sigma_0^2}{Z}) \, N(\mu_k | \tilde{\mu}_k, \delta_k \tau_k^2)\,,$$

where

$$
\begin{aligned}
\tilde{\alpha}_k \;&\equiv\; \frac{1}{\delta_k}\Big[Z_{\bar{\mu}} + \frac{\sigma_0^2}{Z}\Big(\mu_k - \frac{\tau_k^2}{(\sigma_0^2/m_k)}\,\bar{\mu}_k\Big)\Big]\,, \\
\tilde{\mu}_k \;&\equiv\; \tau_k^2\Big[\frac{\bar{\mu}_k}{(\sigma_0^2/m_k)} + \frac{Z_{\bar{\mu}}}{\beta^2}\Big]\,, \quad \text{and} \quad \delta_k \equiv 1 + \frac{\sigma_0^2}{Z}\frac{\tau_k^2}{\beta^2}\,.
\end{aligned}
\tag{13}
$$

Next, **assuming** that α and β are independent, $p(\alpha,\beta^2) = p(\alpha)\,p(\beta^2)$, with a flat $p(\alpha)$[1], we integrate Eq. (**??**), over α, to find

$$
\begin{aligned}
p(\mu_k,\beta^2|D) \;&=\; \frac{p(\beta^2)}{p(D)}\sqrt{2\pi\,\frac{\beta^2\sigma_0^2}{Z}}\;\exp\left\{-\,\frac{Z}{2\beta^2\sigma_0^2}\Big[Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2\Big] - \frac{1}{2\sigma_0^2}\sum_{i=1}^{N} m_i\,\bar{\sigma}_i^2\right\}\times \\
&\times\;\left[\prod_{i=1}^{N}\frac{\tau_i}{\beta}\,(2\pi\sigma_0^2)^{-m_i/2}\right]N(\mu_k|\tilde{\mu}_k,\delta_k\tau_k^2)\,.
\end{aligned}
\tag{14}
$$

Finally, from either $p(\alpha,\beta^2|D)$ or $p(\mu_k,\beta^2|D)$, we may obtain the posterior distribution of β^2, given the data:

$$
p(\beta^2|D) = \frac{p(\beta^2)}{p(D)}\sqrt{2\pi\,\frac{\beta^2\sigma_0^2}{Z}}\;\exp\left\{-\,\frac{Z}{2\beta^2\sigma_0^2}\Big[Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2\Big] - \frac{1}{2\sigma_0^2}\sum_{i=1}^{N} m_i\,\bar{\sigma}_i^2\right\}\prod_{i=1}^{N}\frac{\tau_i}{\beta}\,(2\pi\sigma_0^2)^{-m_i/2}\,. \tag{15}
$$

We can summarize the results above, under the **assumptions** of $p(\alpha,\beta^2) = p(\alpha)\,p(\beta^2)$ and a flat $p(\alpha)$, as follows:

$$
\begin{aligned}
p(\{\mu_i\},\alpha,\beta^2|D) \;&=\; p(\beta^2|D)N(\alpha|Z_{\bar{\mu}},\frac{\beta^2\sigma_0^2}{Z})\prod_{i=1}^{N}N(\mu_i|\lambda_i,\tau_i^2)\,, \\
p(\mu_k,\alpha,\beta^2|D) \;&=\; p(\beta^2|D)\,N(\alpha|Z_{\bar{\mu}},\frac{\beta^2\sigma_0^2}{Z})\,N(\mu_k|\lambda_k,\tau_k^2) \\
&=\; p(\beta^2|D)\,N(\alpha|\tilde{\alpha}_k,\frac{1}{\delta_k}\frac{\beta^2\sigma_0^2}{Z})\,N(\mu_k|\tilde{\mu}_k,\delta_k\tau_k^2)\,, \\
p(\mu_k,\beta^2|D) \;&=\; p(\beta^2|D)\,N(\mu_k|\tilde{\mu}_k,\delta_k\tau_k^2)\,, \\
p(\alpha,\beta^2|D) \;&=\; p(\beta^2|D)\,N(\alpha|Z_{\bar{\mu}},\frac{\beta^2\sigma_0^2}{Z})\,.
\end{aligned}
\tag{16}
$$

[1]It would seem that a flat $p(\alpha)$ is not appropriate, since α takes values over an infinite range, thus preventing the proper normalization of a flat $p(\alpha)$. Still...

Note that λ_k depends on α and β, while $\tilde{\alpha}_k$ depends on μ_k and β. Hence, to obtain $p(\mu_k, \beta^2|D)$ in the simplest possible way, we must use the *second* form of $p(\mu_k, \alpha, \beta^2|D)$, whereas to obtain $p(\alpha, \beta^2|D)$ in the simplest possible way, we must use the *first* form.

3.1 Parameter estimators

Given the joint posterior distribution $p(x, z|D)$, let

$$p(z|D) \equiv \int p(x,z|D)\,dx \qquad \text{and} \qquad p(x|D) \equiv \int p(x,z|D)\,dz \tag{17}$$

be the corresponding marginal posteriors. Also, define

$$\begin{aligned} \hat{x}(z) &\equiv \arg\max_x p(x,z|D)\,, \qquad \bar{x}(z) \equiv \frac{\int x\,p(x,z|D)\,dx}{\int p(x,z|D)\,dx} = \frac{\int x\,p(x,z|D)\,dx}{p(z|D)}\,, \\ \hat{z} &\equiv (z)_{\text{MAP}} \equiv \arg\max_z p(z|D)\,, \quad \text{and} \quad \bar{z} \equiv (z)_{\text{Bayes}} \equiv (z)_{\text{avg}} \equiv \frac{\int z\,p(z|D)\,dz}{\int p(z|D)\,dz}\,. \end{aligned} \tag{18}$$

We may then define several different estimators for the parameter x, such as:

$$\begin{aligned} \langle x\rangle_1 &= \hat{x}(\hat{z})\,, \qquad \langle x\rangle_2 = \hat{x}(\bar{z})\,, \qquad \langle x\rangle_3 = \bar{x}(\hat{z})\,, \qquad \langle x\rangle_4 = \bar{x}(\bar{z})\,, \\ \langle x\rangle_5 &= \max_z \hat{x}(z)\,, \qquad \langle x\rangle_6 = \max_z \bar{x}(z)\,, \qquad \langle x\rangle_7 = \Big[\hat{x}(z)\Big]_{\text{avg}}\,, \qquad \langle x\rangle_8 = \Big[\bar{x}(z)\Big]_{\text{avg}}\,, \\ \langle x\rangle_9 &= (x)_{\text{MAP}}\,, \qquad \langle x\rangle_{10} = (x)_{\text{Bayes}}\,. \end{aligned} \tag{19}$$

In our case, x represents either α or μ_k, and z represents β^2. In practice, we also have the **standard estimators**

$$\begin{aligned} (\alpha)_{\text{std}} &= \frac{1}{N}\sum_{i=1}^{N} \bar{\mu}_i\,, \\ (\beta^2)_{\text{std}} &= \frac{1}{N}\sum_{i=1}^{N} \Big[\bar{\mu}_i - (\hat{\alpha})_{\text{std}}\Big]^2 = \left(\frac{1}{N}\sum_{i=1}^{N} \bar{\mu}_i^2\right) - (\hat{\alpha})^2_{\text{std}}\,, \\ (\mu_k)_{\text{std}} &= \bar{\mu}_k\,. \end{aligned} \tag{20}$$

3.1.1 Estimators for β^2

The **maximum a-posteriori estimator** of β^2, $\hat{\beta^2} = (\beta^2)_{\text{MAP}} \equiv \arg\max_{\beta^2} p(\beta^2|D)$, can be found by solving for $\hat{\beta^2}$ in the equation

$$\frac{\partial}{\partial\beta^2}\Big[2\ln p(\beta^2|D)\Big]\bigg|_{\beta^2=\hat{\beta^2}} = 0\,. \tag{21}$$

The argument of the derivative, **assuming** a flat $p(\beta^2)$[2] and neglecting terms which are independent of β^2, is found to be

$$g(\beta^2) \equiv 2\ln p(\beta^2|D) = \frac{\Big(\sum_{i=1}^{N}\xi_i\,\bar{\mu}_i\Big)^2}{\Big(\sum_{i=1}^{N}\xi_i\Big)} - \sum_{i=1}^{N}\xi_i\,\bar{\mu}_i^2 + \sum_{i=1}^{N}\ln\xi_i - \ln\Big(\sum_{i=1}^{N}\xi_i\Big)\,, \tag{22}$$

with $\xi_i \equiv \Big(\beta^2 + \frac{\sigma_0^2}{m_i}\Big)^{-1}$, and it then follows that $\hat{\beta^2}$ is the solution (hopefully unique and positive) of

$$\begin{aligned}\Big(\sum_{i=1}^{N}\xi_i\Big)^2\frac{\partial g}{\partial\beta^2}\bigg|_{\beta^2=\hat{\beta^2}} = \bigg[&-2\Big(\sum_{i=1}^{N}\xi_i\bar{\mu}_i\Big)\Big(\sum_{i=1}^{N}\bar{\mu}_i\xi_i^2\Big)\Big(\sum_{i=1}^{N}\xi_i\Big) + \Big(\sum_{i=1}^{N}\xi_i\bar{\mu}_i\Big)^2\Big(\sum_{i=1}^{N}\xi_i^2\Big)\\ &+\Big(\sum_{i=1}^{N}\xi_i^2\bar{\mu}_i^2\Big)\Big(\sum_{i=1}^{N}\xi_i\Big)^2 - \Big(\sum_{i=1}^{N}\xi_i\Big)^3 + \Big(\sum_{i=1}^{N}\xi_i^2\Big)\Big(\sum_{i=1}^{N}\xi_i\Big)\bigg]\bigg|_{\beta^2=\hat{\beta^2}} = 0\,.\end{aligned} \tag{23}$$

Since, in the general case of arbitrary N and arbitrary m_i's, the integrations over β^2 involved in computing the average value of β^2 cannot be carried out analytically, an analytical expression for that average cannot be written down in general.

3.1.2 Estimators for α

In general, the marginal posterior distribution $p(\alpha|D)$ cannot be written down analytically. However, all of the estimators defined previously, with the exception of $\langle\alpha\rangle_9 = (\alpha)_{\text{MAP}}$, can still be derived without

[2] It would seem that $p(\beta^2)$, like $p(\alpha)$, must not be flat, for β^2 also takes values over an infinite range.

an explicit form for $p(\alpha|D)$. Starting with

$$\hat{\alpha}(\beta^2) \equiv \arg\max_{\alpha} p(\alpha, \beta^2|D) = Z_{\bar{\mu}}(\beta^2)\,,$$

$$\bar{\alpha}(\beta^2) \equiv \frac{\int_{-\infty}^{+\infty} \alpha\, p(\alpha, \beta^2|D)\, d\alpha}{\int_{-\infty}^{+\infty} p(\alpha, \beta^2|D)\, d\alpha} = \frac{\int_{-\infty}^{+\infty} \alpha\, p(\alpha, \beta^2|D)\, d\alpha}{p(\beta^2|D)} = \int_{-\infty}^{+\infty} \alpha\, N(\alpha|Z_{\bar{\mu}}, \frac{\beta^2\sigma_0^2}{Z})\, d\alpha = Z_{\bar{\mu}}(\beta^2)\,, \tag{24}$$

we obtain

$$\begin{aligned} \langle\alpha\rangle_1 &= \langle\alpha\rangle_3 = Z_{\bar{\mu}}(\hat{\beta^2})\,, \qquad \langle\alpha\rangle_2 = \langle\alpha\rangle_4 = Z_{\bar{\mu}}(\bar{\beta^2})\,, \\ \langle\alpha\rangle_5 &= \langle\alpha\rangle_6 = \max_{\beta^2} Z_{\bar{\mu}}(\beta^2)\,, \qquad \langle\alpha\rangle_7 = \langle\alpha\rangle_8 = \langle\alpha\rangle_{10} = \Big[Z_{\bar{\mu}}(\beta^2)\Big]_{\text{avg}}\,. \end{aligned} \tag{25}$$

Note that, as it turns out,

$$\arg\max_{\beta^2} Z_{\bar{\mu}}(\beta^2) = \text{value of } \beta^2 \text{ such that } \sum_{i=1}^{N} m_i\, \tau_i^4\, (\bar{\mu}_i - Z_{\bar{\mu}}) = 0\,. \tag{26}$$

3.1.3 Estimators for μ_k

Similarly, the marginal posterior distribution $p(\mu_k|D)$ cannot be written down analytically in the general case of arbitrary N and arbitrary m_i's. However, the situation follows closely that of estimating α:

$$\hat{\mu}_k(\beta^2) \equiv \arg\max_{\mu_k} p(\mu_k, \beta^2|D) = \tilde{\mu}_k(\beta^2) = \frac{\beta^2\bar{\mu}_k + (\sigma_0^2/m_k)\, Z_{\bar{\mu}}(\beta^2)}{\beta^2 + (\sigma_0^2/m_k)}\,,$$

$$\bar{\mu}_k(\beta^2) \equiv \frac{\int_{-\infty}^{+\infty} \mu_k\, p(\mu_k, \beta^2|D)\, d\mu_k}{\int_{-\infty}^{+\infty} p(\mu_k, \beta^2|D)\, d\mu_k} = \frac{\int_{-\infty}^{+\infty} \mu_k\, p(\mu_k, \beta^2|D)\, d\mu_k}{p(\beta^2|D)} = \int_{-\infty}^{+\infty} \mu_k\, N(\mu_k|\tilde{\mu}_k, \delta_k\tau_k^2)\, d\mu_k = \tilde{\mu}_k(\beta^2)\,, \tag{27}$$

from which

$$\begin{aligned} \langle\mu_k\rangle_1 &= \langle\mu_k\rangle_3 = \tilde{\mu}_k(\hat{\beta^2})\,, \qquad \langle\mu_k\rangle_2 = \langle\mu_k\rangle_4 = \tilde{\mu}_k(\bar{\beta^2})\,, \\ \langle\mu_k\rangle_5 &= \langle\mu_k\rangle_6 = \max_{\beta^2} \tilde{\mu}_k(\beta^2)\,, \qquad \langle\mu_k\rangle_7 = \langle\mu_k\rangle_8 = \langle\mu_k\rangle_{10} = \Big[\tilde{\mu}_k(\beta^2)\Big]_{\text{avg}}\,. \end{aligned} \tag{28}$$

Note that $\langle\mu_k\rangle_9 = (\mu_k)_{\text{MAP}}$ cannot be computed analytically. Also,

$$\arg\max_{\beta^2} \tilde{\mu}_k(\beta^2) = \text{value of } \beta^2 \text{ such that } (\beta^2 + \frac{\sigma_0^2}{m_k})\, \frac{\partial Z_{\bar{\mu}}}{\partial \beta^2} = Z_{\bar{\mu}} - \bar{\mu}_k\,. \tag{29}$$

3.2 Special case: a single individual

When $N = 1$, the various Z's reduce to $Z = m\,\tau^2$, $Z_{\bar\mu} = \bar\mu$, and $Z_{\bar\mu^2} = \bar\mu^2$, and we find:

$$\begin{aligned}
p(\mu,\alpha,\beta^2|D) &= p(\beta^2|D)\,N(\alpha|\bar\mu,\beta^2+\frac{\sigma_0^2}{m})\,N(\mu|\lambda,\tau^2)\,,\\
&= p(\beta^2|D)\,N(\alpha|\mu,\beta^2)\,N(\mu|\bar\mu,\frac{\sigma_0^2}{m})\,,\\
p(\alpha,\beta^2|D) &= p(\beta^2|D)\,N(\alpha|\bar\mu,\beta^2+\frac{\sigma_0^2}{m})\,,\\
p(\mu,\beta^2|D) &= p(\beta^2|D)\,N(\mu|\bar\mu,\frac{\sigma_0^2}{m})\,,\\
p(\beta^2|D) &= f(D)\,p(\beta^2)\,,
\end{aligned} \tag{30}$$

with

$$\begin{aligned}
f(D) &\equiv \frac{1}{p(D)}\,(2\pi\sigma_0^2)^{-m/2}\sqrt{2\pi\frac{\sigma_0^2}{m}}\,\exp\Big[-\frac{m\,\bar\sigma^2}{2\sigma_0^2}\Big]\,,\\
\lambda &\equiv \tau^2\left(\frac{\alpha}{\beta^2}+\frac{m\,\bar\mu}{\sigma_0^2}\right),\quad \tau^2 \equiv \Big(\frac{1}{\beta^2}+\frac{m}{\sigma_0^2}\Big)^{-1},\\
\bar\mu &\equiv \frac{1}{m}\sum_{j=1}^{m} x_j\,,\quad\text{and}\quad \bar\sigma^2 \equiv \frac{1}{m}\sum_{j=1}^{m}(x_j-\bar\mu)^2\,.
\end{aligned} \tag{31}$$

With $p(\beta^2)$ normalized to 1, it's then easy to show that

$$\begin{aligned}
p(\mu|D) &\equiv \int_0^\infty p(\mu,\beta^2|D)\,d\beta^2 = f(D)\,N(\mu|\bar\mu,\frac{\sigma_0^2}{m})\,,\\
p(\alpha|D) &\equiv \int_0^\infty p(\alpha,\beta^2|D)\,d\beta^2 = \frac{f(D)}{2\sqrt{\pi}}\,|\alpha-\bar\mu|\int_0^z \frac{e^{-t}\,dt}{t^{3/2}}\,p(\beta^2)\Big|_{\beta^2=\frac{\sigma_0^2}{m}(\frac{z}{t}-1)}\,,\quad\text{where}\\
z &\equiv \frac{1}{2}\,\frac{(\alpha-\bar\mu)^2}{\sigma_0^2/m}\,.
\end{aligned} \tag{32}$$

Note that $p(\alpha|D)$ is symmetric around $\alpha = \bar\mu$. Also, since the integral $\int_0^z t^r e^{-t}\,dt$ does not converge unless $z > 0$ and $\mathrm{Re}(r) > -1$, we see that $p(\beta^2)$ cannot be flat. But convergence of the integral in $p(\alpha|D)$ is not sufficient; we also need $p(\beta^2)$ to be normalizable to 1.

Given a properly defined $p(\beta^2)$, it then follows that *all* the estimators defined in Eq. (**??**), when computed for $x = \mu$, agree with $(\mu)_{\rm std} = \bar{\mu}$. Similarly, $\langle\alpha\rangle_1 \ldots \langle\alpha\rangle_8$ and $\langle\alpha\rangle_{10} = (\alpha)_{\rm Bayes}$ also *all* agree with $(\alpha)_{\rm std}$ which, for $N = 1$, also equals $\bar{\mu}$. $\langle\alpha\rangle_9 = (\alpha)_{\rm MAP}$, however, depends on the choice of $p(\beta^2)$. Estimating β^2, of course, makes no sense when $N = 1$.

3.3 Special case: $N = 2$ individuals

When $N = 2$ individuals, Eq. (**??**) can be easily solved analytically to produce

$$\hat{\beta}^2 = (\beta^2)_{\rm MAP} = \frac{1}{2}\,(\bar{\mu}_1 - \bar{\mu}_2)^2 - \frac{(m_1 + m_2)}{2}\,\frac{\sigma_0^2}{m_1 m_2} = 2\,(\beta^2)_{\rm std} - \frac{(m_1 + m_2)}{2}\,\frac{\sigma_0^2}{m_1 m_2}\,. \tag{33}$$

However, it **assumes** a flat prior for β^2, in which case $p(\alpha|D)$ and $p(\mu_k|D)$ cannot be computed, because the relevant integrations do not converge. Nevertheless, some of the estimators defined in Eq. (**??**) can still be computed. First, note that

$$Z_{\bar{\mu}}(\beta^2) = \frac{\beta^2\left(\dfrac{\bar{\mu}_1 + \bar{\mu}_2}{2}\right) + \dfrac{\sigma_0^2}{m_1 m_2}\left(\dfrac{m_1\bar{\mu}_1 + m_2\bar{\mu}_2}{2}\right)}{\beta^2 + \dfrac{\sigma_0^2}{m_1 m_2}\left(\dfrac{m_1 + m_2}{2}\right)}\,, \tag{34}$$

from which we obtain

$$\begin{aligned}
\langle\alpha\rangle_1 = \langle\alpha\rangle_3 = Z_{\bar{\mu}}(\hat{\beta}^2) &= \left(\frac{\bar{\mu}_1 + \bar{\mu}_2}{2}\right) + \frac{\sigma_0^2}{2\,m_1 m_2}\left(\frac{m_2 - m_1}{\bar{\mu}_2 - \bar{\mu}_1}\right) = (\alpha)_{\rm std} + \frac{\sigma_0^2}{2\,m_1 m_2}\left(\frac{m_2 - m_1}{\bar{\mu}_2 - \bar{\mu}_1}\right)\\
\langle\mu_k\rangle_1 = \langle\mu_k\rangle_3 = \tilde{\mu}_k(\hat{\beta}^2) &= \bar{\mu}_k + \frac{(-1)^{k+1}}{(\bar{\mu}_2 - \bar{\mu}_1)}\left(\frac{\sigma_0^2}{m_k}\right) = (\mu_k)_{\rm std} + \frac{(-1)^{k+1}}{(\bar{\mu}_2 - \bar{\mu}_1)}\left(\frac{\sigma_0^2}{m_k}\right).
\end{aligned} \tag{35}$$

The remaining estimators for α and μ_k are undetermined, due to the improper nature of the priors. Note that the estimators for μ_1 and μ_2 lie on opposite sides of their respective standard estimators.

3.4 Special case: arbitrary N, but with a common number of observations per individual

When $m_i = m$, we also have $\tau_i = \tau$, $Z_{\bar{\mu}}$ and $Z_{\bar{\mu}^2}$ are now independent of β^2, and Z is proportional to τ^2. Much simplification then results, and:

$$
\begin{aligned}
p(\{\mu_i\}, \alpha, \beta^2|D) &= p(\beta^2|D)\, N(\alpha|Z_{\bar{\mu}}, \frac{\beta^2}{N} + \frac{\sigma_0^2}{Nm}) \prod_{i=1}^{N} N(\mu_i|\lambda_i, \tau^2)\,, \\
p(\mu_k, \alpha, \beta^2|D) &= p(\beta^2|D)\, N(\alpha|Z_{\bar{\mu}}, \frac{\beta^2}{N} + \frac{\sigma_0^2}{Nm})\, N(\mu_k|\lambda_k, \tau^2) \\
&= p(\beta^2|D)\, N(\alpha|\tilde{\alpha}_k, \frac{\beta^2\sigma_0^2}{Nm\,\tau^2\delta})\, N(\mu_k|\tilde{\mu}_k, \delta\,\tau^2)\,, \\
p(\alpha, \beta^2|D) &= p(\beta^2|D)\, N(\alpha|Z_{\bar{\mu}}, \frac{\beta^2}{N} + \frac{\sigma_0^2}{Nm})\,, \\
p(\mu_k, \beta^2|D) &= p(\beta^2|D)\, N\Big(\mu_k|\tilde{\mu}_k, \frac{\sigma_0^2}{Nm}\Big[\frac{Nm\,\beta^2 + \sigma_0^2}{m\,\beta^2 + \sigma_0^2}\Big]\Big)\,, \\
p(\beta^2|D) &= f(D)\, p(\beta^2)\, h\Big(\frac{\tau^2}{\beta^2}\Big)\,,
\end{aligned}
\tag{36}
$$

where

$$
\begin{aligned}
f(D) &\equiv \frac{(2\pi\sigma_0^2)^{-Nm/2}}{p(D)} \sqrt{2\pi\frac{\sigma_0^2}{Nm}}\, \exp\left[-\frac{m}{2\sigma_0^2}\sum_{i=1}^{N}\bar{\sigma}_i^2\right], \\
h(x) &\equiv x^{(N-1)/2} \exp\Big[-\frac{\xi(D)}{2}\,x\Big], \qquad \xi(D) \equiv \frac{Nm}{\sigma_0^2}\Big[Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2\Big], \\
\lambda_i &\equiv \tau^2\left[\frac{\alpha}{\beta^2} + \frac{\bar{\mu}_i}{(\sigma_0^2/m)}\right], \qquad \tau^2 \equiv \left[\frac{1}{\beta^2} + \frac{1}{(\sigma_0^2/m)}\right]^{-1}, \qquad \delta \equiv 1 + \frac{\sigma_0^2}{Nm\,\beta^2}\,, \\
\tilde{\alpha}_k &\equiv \frac{1}{\delta}\Big[Z_{\bar{\mu}} - \frac{\bar{\mu}_k}{N} + \frac{\sigma_0^2\,\mu_k}{Nm\,\tau^2}\Big], \qquad \tilde{\mu}_k \equiv \tau^2\Big[\frac{Z_{\bar{\mu}}}{\beta^2} + \frac{\bar{\mu}_k}{(\sigma_0^2/m)}\Big] = \lambda_k(\alpha = Z_{\bar{\mu}})\,, \\
\bar{\mu}_i &\equiv \frac{1}{m}\sum_{j=1}^{m} x_{ij}\,, \qquad \bar{\sigma}_i^2 \equiv \frac{1}{m}\sum_{j=1}^{m}(x_{ij} - \bar{\mu}_i)^2\,, \\
Z &\equiv Nm\,\tau^2\,, \qquad Z_{\bar{\mu}} \equiv \frac{1}{N}\sum_{i=1}^{N}\bar{\mu}_i\,, \quad \text{and} \qquad Z_{\bar{\mu}^2} \equiv \frac{1}{N}\sum_{i=1}^{N}\bar{\mu}_i^2\,.
\end{aligned}
\tag{37}
$$

The **maximum a-posteriori estimator** for β^2 is then:

$$\hat{\beta^2} = (\beta^2)_{\text{MAP}} = \frac{\sigma_0^2}{m}\Big[\frac{\xi(D)}{N-1} - 1\Big] = \frac{N}{N-1}\Big[Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2\Big] - \frac{\sigma_0^2}{m} = \frac{N}{N-1}\,(\beta^2)_{\text{std}} - \frac{\sigma_0^2}{m}\,. \tag{38}$$

We can also compute an average β^2, according to Eq. (**??**), as follows:

$$\bar{\beta^2} = (\hat{\beta^2})_{\text{Bayes}} = (\hat{\beta^2})_{\text{avg}} = \frac{\displaystyle\int_0^\infty \beta^2\, p(\beta^2|D)\, d\beta^2}{\displaystyle\int_0^\infty p(\beta^2|D)\, d\beta^2}\,. \tag{39}$$

Consider the more general integral,

$$\begin{aligned} I(r) \;&\equiv \int_0^\infty (\beta^2)^r\, p(\beta^2|D)\, d\beta^2 = f(D)\int_0^\infty d\beta^2\,(\beta^2)^r\, h\Big(\frac{\tau^2}{\beta^2}\Big)\, p(\beta^2) \\ &= f(D)\Big(\frac{\sigma_0^2}{m}\Big)^{r+1}\int_0^1 ds\,(1-s)^r s^{(N-2r-5)/2}\,\exp\Big[-\frac{\xi(D)}{2}\,s\Big]\, p(\beta^2)\Big|_{\beta^2 = \frac{\sigma_0^2}{m}(\frac{1}{s}-1)}\,. \end{aligned} \tag{40}$$

When $p(\beta^2)$ is flat, $(N-2r) > 3$, and $r > -1$, this integral can be written in terms of the Kümmel Hypergeometric function ${}_1F_1(a,b,z)$, which is defined by

$${}_1F_1(a,b,z) \equiv \frac{\Gamma(b)}{\Gamma(b-a)\,\Gamma(a)}\int_0^1 dt\, e^{zt}\, t^{a-1}\,(1-t)^{b-a-1}\,, \tag{41}$$

as

$$I(r) = f(D)\Big(\frac{\sigma_0^2}{m}\Big)^{r+1}\Gamma(r+1)\,\frac{\Gamma\Big(\frac{N-3}{2} - r\Big)}{\Gamma\Big(\frac{N-1}{2}\Big)}\,{}_1F_1\Big[\frac{N-3}{2} - r, \frac{N-1}{2}, -\frac{\xi(D)}{2}\Big]\,. \tag{42}$$

Then,

$$\bar{\beta^2} = (\beta^2)_{\text{Bayes}} = (\beta^2)_{\text{avg}} = \frac{I(1)}{I(0)} = \Big(\frac{\sigma_0^2}{m}\Big)\Big(\frac{2}{N-5}\Big)\frac{{}_1F_1\Big[\frac{N-5}{2}, \frac{N-1}{2}, -\frac{\xi(D)}{2}\Big]}{{}_1F_1\Big[\frac{N-3}{2}, \frac{N-1}{2}, -\frac{\xi(D)}{2}\Big]}\,, \tag{43}$$

provided that $N > 5$.

We next compute $p(\alpha|D)$,

$$
\begin{aligned}
p(\alpha|D) &\equiv \int_0^\infty p(\alpha,\beta^2|D)\,d\beta^2 = f(D)\int_0^\infty d\beta^2\,p(\beta^2)\,h\Big(\frac{\tau^2}{\beta^2}\Big)\,N\Big(\alpha|Z_{\bar\mu},\frac{\beta^2}{N}+\frac{\sigma_0^2}{Nm}\Big)\\
&= f(D)\Big(\frac{N}{2\pi}\Big)^{1/2}\Big(\frac{\sigma_0^2}{m}\Big)^{1/2}\int_0^1 dt\,t^{(N-4)/2}\,\exp(-A\,t)\,p(\beta^2)\Big|_{\beta^2=\frac{\sigma_0^2}{m}(\frac{1}{t}-1)}\\
&= f(D)\Big(\frac{N}{2\pi}\Big)^{1/2}\Big(\frac{\sigma_0^2}{m}\Big)^{1/2}A^{-(N-2)/2}\Big[\Gamma\Big(\frac{N-2}{2}\Big)-\Gamma\Big(\frac{N-2}{2},A\Big)\Big],\quad\text{where}\\
A &\equiv \frac{1}{2}\Big[\frac{(\alpha-Z_{\bar\mu})^2+Z_{\bar\mu^2}-(Z_{\bar\mu})^2}{(\sigma_0^2/Nm)}\Big]\quad\text{and}\quad\Gamma(a,z)\equiv\int_z^\infty t^{a-1}e^{-t}dt\,.
\end{aligned}
\tag{44}
$$

Note that the last line of this result is valid only if $N>2$ and **assumes** a flat $p(\beta^2)$. Even without this explicit result, however, it follows that *all* the estimators for α defined in Eq. (**??**) and Eq. (**??**) equal the standard estimator, $\langle\alpha\rangle_1=\ldots=\langle\alpha\rangle_{10}=(\alpha)_{\rm std}=\ Z_{\bar\mu}$. Similarly,

$$
\begin{aligned}
p(\mu_k|D) &\equiv \int_0^\infty p(\mu_k,\beta^2|D)\,d\beta^2 = f(D)\int_0^\infty d\beta^2\,p(\beta^2)\,h\Big(\frac{\tau^2}{\beta^2}\Big)\,N\Big(\mu_k|\tilde\mu_k,\frac{\sigma_0^2}{Nm}\Big[\frac{Nm\,\beta^2+\sigma_0^2}{m\,\beta^2+\sigma_0^2}\Big]\Big)\\
&= f(D)\Big(\frac{\sigma_0^2}{m}\Big)\int_0^1\frac{dt\,h(t)}{t^2}\,N\bigg(\mu_k|\bar\mu_k+(Z_{\bar\mu}-\bar\mu_k)t,\frac{\sigma_0^2}{m}\Big[1-\frac{(N-1)}{N}\,t\Big]\bigg)p(\beta^2)\Big|_{\beta^2=\frac{\sigma_0^2}{m}(\frac{1}{t}-1)}\\
&= \frac{f(D)}{\sqrt{2\pi}}\Big(\frac{\sigma_0^2}{m}\Big)^{1/2}\int_0^1\frac{t^{(N-5)/2}}{\sqrt{1-\big(\frac{N-1}{N}\big)t}}\,\exp\left\{-\frac{\Big[(\mu_k-\bar\mu_k)-(Z_{\bar\mu}-\bar\mu_k)t\Big]^2}{2\,\big(\sigma_0^2/m\big)\Big[1-\big(\frac{N-1}{N}\big)t\Big]}-\frac{\xi(D)\,t}{2}\right\}dt\,.
\end{aligned}
\tag{45}
$$

From the estimator definitions, we find

$$
\begin{aligned}
\langle\mu_k\rangle_1=\langle\mu_k\rangle_3=\tilde\mu_k(\hat{\beta^2}) &= \bar\mu_k \qquad + \Big(\frac{N-1}{N}\Big)\frac{(\sigma_0^2/m)}{\Big[Z_{\bar\mu^2}-(Z_{\bar\mu})^2\Big]}(Z_{\bar\mu}-\bar\mu_k)\\
&= (\mu_k)_{\rm std} + \Big(\frac{N-1}{N}\Big)\frac{(\sigma_0^2/m)}{\Big[Z_{\bar\mu^2}-(Z_{\bar\mu})^2\Big]}(Z_{\bar\mu}-\bar\mu_k)\,.
\end{aligned}
\tag{46}
$$

The remaining estimators, except for $\langle\mu_k\rangle_5$, $\langle\mu_k\rangle_6$, and $\langle\mu_k\rangle_9$, are too complicated to compute analytically but can be computed numerically. $\langle\mu_k\rangle_5=\langle\mu_k\rangle_6$ are actually not defined, and $\langle\mu_k\rangle_9$ depends on having an analytical expression for the integral above.

4 Specific model: Normally-distributed observations, with conjugate priors for both the mean and the variance

Recall that the posterior probability of the parameters, given the data, is

$$p(\{\vec{\theta}_i\}, \vec{\Psi}|D) = \frac{p(\vec{\Psi})}{p(D)} \prod_{i=1}^{N} \left\{ p(\vec{\Theta} = \vec{\theta}_i|\vec{\Psi}) \prod_{j=1}^{m_i} p(X_i = x_{ij}|\vec{\theta}_i) \right\}. \tag{47}$$

Once again, we'll assume that the observations for any single individual are normally distributed, with mean μ_i and variance σ_i^2 characteristic of the individual in question. To keep the notation manageable, we'll define $s_i \equiv \sigma_i^2$. Thus, $\vec{\theta}_i \equiv \{\mu_i, s_i\}$ and

$$p(X_i = x_{ij}|\vec{\theta}_i) \equiv N(x_{ij}|\mu_i, s_i) = (2\pi s_i)^{-1/2} \exp\Big[-\frac{1}{2s_i}(x_{ij} - \mu_i)^2 \Big], \qquad \text{with } s_i > 0. \tag{48}$$

Now, however, the variances are not assumed to be fixed but, like the means, are themselves drawn from some distribution. In other words, we need to choose the two probability distributions on the right-hand-side of the equation below:

$$p(\vec{\Theta} = \vec{\theta}_i|\vec{\Psi}) = p(\mu_i, s_i|\vec{\Psi}) = p(\mu_i|s_i, \vec{\Psi})\, p(s_i|\vec{\Psi}). \tag{49}$$

It turns out to be convenient to choose an inverse chi-squared distribution for $p(s_i|\vec{\Psi})$,

$$p(s_i|\vec{\Psi}) = \frac{(S_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)}\, s_i^{-\nu_0/2-1} \exp(-\frac{1}{2}\frac{S_0}{s_i}), \qquad \text{with } S_0 > 0 \text{ and } \nu_0 \geq 0, \tag{50}$$

and $p(\mu_i|s_i, \vec{\Psi})$ to be Normal, with mean θ_0 and variance s_i/n_0 ($n_0 > 0$),

$$p(\mu_i|s_i, \vec{\Psi}) \equiv N(\mu_i|\theta_0, s_i/n_0) = (2\pi s_i/n_0)^{-1/2} \exp\Big[-\frac{n_0}{2s_i}(\mu_i - \theta_0)^2 \Big]. \tag{51}$$

Notice that these choices imply $\vec{\Psi} = \{\nu_0, S_0, \theta_0, n_0\}$.

Collecting these results, we obtain

$$
\begin{aligned}
p(\{\mu_i, s_i\}, \vec{\Psi}|D) \;&=\; (2\pi)^{-\Big(\dfrac{N+\Sigma_i\, m_i}{2}\Big)} \times n_0^{N/2} \times \left[\frac{(S_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)}\right]^N \times \frac{p(\nu_0, S_0, \theta_0, n_0)}{p(D)} \times \\
&\prod_{i=1}^{N} s_i^{-(\nu_0+m_i+3)/2} \exp\Big\{ -\frac{1}{2s_i}\Big[S_0 + n_0(\mu_i-\theta_0)^2 + \sum_{j=1}^{m_i}(x_{ij}-\mu_i)^2\Big]\Big\} \\
&=\; (2\pi)^{-\Big(\dfrac{N+\Sigma_i\, m_i}{2}\Big)} \times n_0^{N/2} \times \left[\frac{(S_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)}\right]^N \times \frac{p(\nu_0, S_0, \theta_0, n_0)}{p(D)} \times \\
&\prod_{i=1}^{N} s_i^{-(\nu_0+m_i+3)/2} \exp\left[-\frac{(m_i+n_0)}{2s_i}\Big[\mu_i - \Big(\frac{m_i\,\bar{\mu}_i + n_0\,\theta_0}{m_i+n_0}\Big)\Big]^2\right] \times \\
&\exp\left[-\frac{1}{2s_i}\Big(\frac{m_i\, n_0}{m_i+n_0}\Big)(\theta_0-\bar{\mu}_i)^2\right] \times \exp\left[-\frac{(S_0 + m_i\,\bar{s}_i)}{2s_i}\right].
\end{aligned}
\tag{52}
$$

where $\bar{\mu}_i$ and $\bar{s}_i$ are the sample mean and sample variance, respectively, defined in Eq. (**??**).

This result can be simplified by expressing two of the exponentials in terms of Normal distributions, as follows:

$$
\begin{aligned}
p(\{\mu_i, s_i\}, \vec{\Psi}|D) \;&=\; (2\pi)^{-\Big(\dfrac{N+\Sigma_i\, m_i}{2}\Big)} \times n_0^{N/2} \times \left[\frac{(S_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)}\right]^N \times \frac{p(\nu_0, S_0, \theta_0, n_0)}{p(D)} \times \\
&\prod_{i=1}^{N} s_i^{-(\nu_0+m_i+3)/2} \times \Big[2\pi\,(\frac{s_i}{m_i+n_0})\Big]^{1/2} N\Big(\mu_i\,|\,\frac{m_i\,\bar{\mu}_i + n_0\,\theta_0}{m_i+n_0}, \frac{s_i}{m_i+n_0}\Big) \times \\
&\Big[2\pi\, s_i\,\frac{(m_i+n_0)}{m_i\, n_0}\Big]^{1/2} N\Big(\theta_0\,|\,\bar{\mu}_i, s_i\,\frac{(m_i+n_0)}{m_i\, n_0}\Big) \times \exp\left[-\frac{(S_0 + m_i\,\bar{s}_i)}{2s_i}\right].
\end{aligned}
\tag{53}
$$

Further simplification yields

$$
\begin{aligned}
p(\{\mu_i, s_i\}, \vec{\Psi}|D) \;&=\; (2\pi)^{(N-\Sigma_i\, m_i)/2} \times \left[\frac{(S_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)}\right]^N \times \frac{p(\nu_0, S_0, \theta_0, n_0)}{p(D)} \times \\
&\prod_{i=1}^{N} \Big\{ m_i^{-1/2}\, s_i^{-(\nu_0+m_i+1)/2} \exp\left[-\frac{(S_0 + m_i\,\bar{s}_i)}{2s_i}\right] \times \\
&\qquad N\Big(\mu_i\,|\,\frac{m_i\,\bar{\mu}_i + n_0\,\theta_0}{m_i+n_0}, \frac{s_i}{m_i+n_0}\Big) \times N\Big(\theta_0\,|\,\bar{\mu}_i, s_i\,\frac{(m_i+n_0)}{m_i\, n_0}\Big) \Big\}.
\end{aligned}
\tag{54}
$$

Defining the Z values for this case as

$$Z \equiv \sum_{i=1}^{N} \frac{m_i\, n_0}{(m_i + n_0)}\, \frac{1}{s_i}\,, \quad Z_{\bar{\mu}} \equiv \frac{1}{Z} \sum_{i=1}^{N} \frac{m_i\, n_0}{(m_i + n_0)}\, \frac{\bar{\mu}_i}{s_i}\,, \quad \text{and} \quad Z_{\bar{\mu}^2} \equiv \frac{1}{Z} \sum_{i=1}^{N} \frac{m_i\, n_0}{(m_i + n_0)}\, \frac{\bar{\mu}_i^2}{s_i}\,, \tag{55}$$

we can then write

$$\begin{aligned} p(\{\mu_i, s_i\}, \vec{\Psi} | D) \;=\; & (2\pi)^{-(\Sigma_i\, m_i - 1)/2} \times \left[\frac{(S_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} \right]^N \times \frac{p(\nu_0, S_0, \theta_0, n_0)}{p(D)} \times \\ & Z^{-1/2}\, \exp\left\{ -\, \frac{Z}{2} \left[Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2 \right] \right\} \times \\ & \left\{ \prod_{i=1}^{N} \Big(\frac{n_0}{m_i + n_0} \Big)^{1/2} s_i^{-(\nu_0 + m_i)/2 - 1}\, \exp\left[-\, \frac{(S_0 + m_i\, \bar{s}_i)}{2 s_i} \right] \right\} \times \\ & \left\{ \prod_{i=1}^{N} N\Big(\mu_i \,|\, \frac{m_i\, \bar{\mu}_i + n_0\, \theta_0}{m_i + n_0}, \frac{s_i}{m_i + n_0} \Big) \right\} \times N\Big(\theta_0 \,|\, Z_{\bar{\mu}}, Z^{-1} \Big)\,. \end{aligned} \tag{56}$$

Integrating over all μ_i's but one, say, μ_k, we obtain

$$\begin{aligned} p(\mu_k, \{s_i\}, \vec{\Psi} | D) \;=\; & (2\pi)^{-(\Sigma_i\, m_i - 1)/2} \times \left[\frac{(S_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} \right]^N \times \frac{p(\nu_0, S_0, \theta_0, n_0)}{p(D)} \times \\ & Z^{-1/2}\, \exp\left\{ -\, \frac{Z}{2} \left[Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2 \right] \right\} \times \\ & \left\{ \prod_{i=1}^{N} \Big(\frac{n_0}{m_i + n_0} \Big)^{1/2} s_i^{-(\nu_0 + m_i)/2 - 1}\, \exp\left[-\, \frac{(S_0 + m_i\, \bar{s}_i)}{2 s_i} \right] \right\} \times \\ & N\Big(\mu_k \,|\, \frac{m_k\, \bar{\mu}_k + n_0\, \theta_0}{m_k + n_0}, \frac{s_k}{m_k + n_0} \Big) \times N\Big(\theta_0 \,|\, Z_{\bar{\mu}}, Z^{-1} \Big)\,. \end{aligned} \tag{57}$$

Other integrations that are feasible at this point are with respect to μ_k, θ_0, and S_0, the last two assuming a flat $p(\nu_0, S_0, \theta_0, n_0)$.

4.1 Special case: arbitrary N, but with a common number of observations per individual

In this approximation, $m_i = m$ for $1 \leq i \leq N$, and only Z depends on n_0,

$$Z \equiv \frac{m\, n_0}{(m + n_0)} \sum_{i=1}^{N} \frac{1}{s_i}\,, \quad Z_{\bar{\mu}} \equiv \sum_{i=1}^{N} \bar{\mu}_i\, s_i^{-1} \Bigg/ \sum_{i=1}^{N} s_i^{-1}\,, \quad \text{and} \quad Z_{\bar{\mu}^2} \equiv \sum_{i=1}^{N} \bar{\mu}_i^2\, s_i^{-1} \Bigg/ \sum_{i=1}^{N} s_i^{-1}\,. \tag{58}$$

Then,

$$
\begin{aligned}
p(\mu_k, \{s_i\}, \vec{\Psi}|D) \;=\; & (2\pi)^{-(Nm-1)/2} \times \left[\frac{(S_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)}\right]^N \times \frac{p(\nu_0, S_0, \theta_0, n_0)}{p(D)} \times \\
& Z^{-1/2}\, \exp\left\{-\frac{Z}{2}\left[Z_{\bar{\mu}^2} - (Z_{\bar{\mu}})^2\right]\right\} \times \Big(\frac{n_0}{m+n_0}\Big)^{N/2} \times \\
& \left\{\prod_{i=1}^{N} s_i^{-(\nu_0+m)/2-1}\, \exp\left[-\frac{(S_0 + m\,\bar{s}_i)}{2s_i}\right]\right\} \times \\
& N\Big(\mu_k \,|\, \frac{m\,\bar{\mu}_k + n_0\,\theta_0}{m+n_0}, \frac{s_k}{m+n_0}\Big) \times N\Big(\theta_0 \,|\, Z_{\bar{\mu}}, Z^{-1}\Big)\,.
\end{aligned}
\tag{59}
$$

4.2 Special case: a single individual ($N = 1$) for which we have m observations

The above results further simplify to

$$
Z \equiv \frac{m\,n_0}{(m+n_0)}\,\frac{1}{s}\,, \quad Z_{\bar{\mu}} \equiv \bar{\mu}\,, \quad Z_{\bar{\mu}^2} \equiv \bar{\mu}^2\,, \tag{60}
$$

and

$$
\begin{aligned}
p(\mu, s, \nu_0, S_0, \theta_0, n_0|D) \;=\; & (2\pi)^{-(m-1)/2} \times m^{-1/2} \times \frac{\Gamma[(\nu_0+m-1)/2]}{\Gamma(\nu_0/2)} \times \\
& \frac{(S_0/2)^{\nu_0/2}}{[(S_0 + m\,\bar{s})/2]^{(\nu_0+m-1)/2}} \times \frac{p(\nu_0, S_0, \theta_0, n_0)}{p(D)} \times \\
& \frac{[(S_0 + m\,\bar{s})/2]^{(\nu_0+m-1)/2}}{\Gamma[(\nu_0+m-1)/2]}\; s^{-(\nu_0+m+1)/2}\, \exp\left[-\frac{(S_0 + m\,\bar{s})}{2\,s}\right] \times \\
& N\Big(\mu \,|\, \frac{m\,\bar{\mu} + n_0\,\theta_0}{m+n_0}, \frac{s}{m+n_0}\Big) \times N\Big(\theta_0 \,|\, \bar{\mu}, \frac{(m+n_0)s}{m\,n_0}\Big)\,.
\end{aligned}
\tag{61}
$$

Integrating successively over μ, θ_0 (assuming a flat prior), then s yields

$$
\begin{aligned}
p(\nu_0, S_0, n_0|D) \;=\; & (2\pi)^{-(m-1)/2} \times m^{-1/2} \times \frac{\Gamma[(\nu_0+m-1)/2]}{\Gamma(\nu_0/2)} \times \\
& \frac{(S_0/2)^{\nu_0/2}}{[(S_0 + m\,\bar{s})/2]^{(\nu_0+m-1)/2}} \times \frac{p(\nu_0, S_0, n_0)}{p(D)}\,.
\end{aligned}
\tag{62}
$$

■