# SDS 383C: Statistical Modeling I
# Fall 2022, Module V

**Abhra Sarkar**

Department of Statistics and Data Sciences
The University of Texas at Austin

"All models are wrong, but some are useful."- George E. P. Box

## Normal Model
## Conjugate Prior

$$y_1, \ldots, y_n \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Normal-Inverse-Gamma Prior:** $(\mu, \sigma^2) \sim \text{NIG}\left(\mu_0, \sigma_0^2/\kappa_0, \nu_0, \sigma_0^2\right)$
- **Normal-Inverse-Gamma Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n})$$

$$\propto \left(\sigma^2\right)^{-\left\{\frac{(\nu_0+n)}{2}+1+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2 + (\kappa_0+n)(\mu-\mu_n)^2\right\}\right]$$

$$\equiv \text{NIG}\left(\mu_n, \sigma_n^2/\kappa_n, \nu_n, \sigma_n^2\right), \quad \nu_n = (\nu_0+n), \quad \kappa_n = (\kappa_0+n), \quad \mu_n = (\kappa_0\mu_0 + n\overline{y})/(\kappa_0+n),$$

$$\sigma_n^2 = \frac{1}{\nu_n}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2\right\}$$

(WhiteBoard)

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Normal-Inverse-Gamma Prior:** $(\mu, \sigma^2) \sim \text{NIG}\left(\mu_0, \sigma_0^2/\kappa_0, \nu_0, \sigma_0^2\right)$
- **Normal-Inverse-Gamma Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n})$$

$$\propto \left(\sigma^2\right)^{-\left\{\frac{(\nu_0+n)}{2}+1+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2 + (\kappa_0+n)(\mu-\mu_n)^2\right\}\right]$$

$$\equiv \text{NIG}\left(\mu_n, \sigma_n^2/\kappa_n, \nu_n, \sigma_n^2\right), \quad \nu_n = (\nu_0+n), \quad \kappa_n = (\kappa_0+n), \quad \mu_n = (\kappa_0\mu_0 + n\overline{y})/(\kappa_0+n),$$

$$\sigma_n^2 = \frac{1}{\nu_n}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2\right\}$$

(WhiteBoard)

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

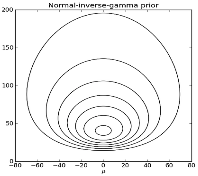
- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Normal-Inverse-Gamma Prior:** $(\mu, \sigma^2) \sim \text{NIG}\left(\mu_0, \sigma_0^2/\kappa_0, \nu_0, \sigma_0^2\right)$

$$p(\mu, \sigma^2) = p(\sigma^2)p(\mu \mid \sigma^2)$$

$$= \text{Inv-Ga}\left(\sigma^2 \mid \nu_0/2, \nu_0\sigma_0^2/2\right) \cdot \text{Normal}(\mu \mid \mu_0, \sigma^2/\kappa_0)$$

$$= \frac{(\nu_0\sigma_0^2)^{\frac{\nu_0}{2}}}{\Gamma(\nu_0/2)\,(\sigma^2)^{(\frac{\nu_0}{2}+1)}} \exp\left(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right) \cdot \frac{\sqrt{\kappa_0}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\kappa_0}{2\sigma^2}(\mu-\mu_0)^2\right]$$

$$\propto \left(\sigma^2\right)^{-\left(\frac{\nu_0}{2}+1+\frac{1}{2}\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + \kappa_0(\mu-\mu_0)^2\right\}\right]$$




- **Normal-Inverse-Gamma Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n})$$

$$\propto \left(\sigma^2\right)^{-\left\{\frac{(\nu_0+n)}{2}+1+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2 + (\kappa_0+n)(\mu-\mu_n)^2\right\}\right]$$

$$\equiv \text{NIG}\left(\mu_n, \sigma_n^2/\kappa_n, \nu_n, \sigma_n^2\right), \quad \nu_n = (\nu_0+n), \quad \kappa_n = (\kappa_0+n), \quad \mu_n = (\kappa_0\mu_0 + n\overline{y})/(\kappa_0+n),$$

$$\sigma_n^2 = \frac{1}{\nu_n}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2\right\}$$

(WhiteBoard)

# Normal Model
## Conjugate Prior

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

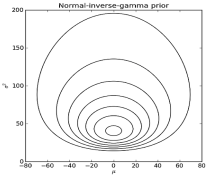
- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Normal-Inverse-Gamma Prior:** $(\mu, \sigma^2) \sim \text{NIG}\left(\mu_0, \sigma_0^2/\kappa_0, \nu_0, \sigma_0^2\right)$

$$p(\mu, \sigma^2) = p(\sigma^2)p(\mu \mid \sigma^2)$$

$$= \text{Inv-Ga}\left(\sigma^2 \mid \nu_0/2, \nu_0\sigma_0^2/2\right) \cdot \text{Normal}(\mu \mid \mu_0, \sigma^2/\kappa_0)$$

$$= \frac{(\nu_0\sigma_0^2)^{\frac{\nu_0}{2}}}{\Gamma(\nu_0/2)\,(\sigma^2)^{(\frac{\nu_0}{2}+1)}} \exp\left(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right) \cdot \frac{\sqrt{\kappa_0}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\kappa_0}{2\sigma^2}(\mu-\mu_0)^2\right]$$

$$\propto \left(\sigma^2\right)^{-\left(\frac{\nu_0}{2}+1+\frac{1}{2}\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + \kappa_0(\mu-\mu_0)^2\right\}\right]$$




- **Normal-Inverse-Gamma Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n})$$

$$\propto \left(\sigma^2\right)^{-\left\{\frac{(\nu_0+n)}{2}+1+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2 + (\kappa_0+n)(\mu-\mu_n)^2\right\}\right]$$

$$\equiv \text{NIG}\left(\mu_n, \sigma_n^2/\kappa_n, \nu_n, \sigma_n^2\right), \quad \nu_n = (\nu_0+n), \quad \kappa_n = (\kappa_0+n), \quad \mu_n = (\kappa_0\mu_0 + n\overline{y})/(\kappa_0+n),$$

$$\sigma_n^2 = \frac{1}{\nu_n}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2\right\}$$

(WhiteBoard)

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Normal-Inverse-Gamma Prior:** $(\mu, \sigma^2) \sim \text{NIG}\left(\mu_0, \sigma_0^2/\kappa_0, \nu_0, \sigma_0^2\right)$

$$p(\mu, \sigma^2) = p(\sigma^2)p(\mu \mid \sigma^2) = \text{Inv-Ga}\left(\sigma^2 \mid \nu_0/2, \nu_0\sigma_0^2/2\right) \cdot \text{Normal}(\mu \mid \mu_0, \sigma^2/\kappa_0)$$
$$\propto \left(\sigma^2\right)^{-\left(\frac{\nu_0}{2}+1+\frac{1}{2}\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + \kappa_0(\mu-\mu_0)^2\right\}\right]$$

- **Normal-Inverse-Gamma Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n})$$
$$\propto \left(\sigma^2\right)^{-\left\{\frac{(\nu_0+n)}{2}+1+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2 + (\kappa_0+n)(\mu-\mu_n)^2\right\}\right]$$
$$\equiv \text{NIG}\left(\mu_n, \sigma_n^2/\kappa_n, \nu_n, \sigma_n^2\right), \quad \nu_n = (\nu_0+n), \quad \kappa_n = (\kappa_0+n), \quad \mu_n = (\kappa_0\mu_0 + n\overline{y})/(\kappa_0+n),$$
$$\sigma_n^2 = \frac{1}{\nu_n}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2\right\}$$

- A-priori and a-posteriori $\mu$ and $\sigma^2$ are dependent but uncorrelated.

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Normal-Inverse-Gamma Prior:** $(\mu, \sigma^2) \sim \text{NIG}\left(\mu_0, \sigma_0^2/\kappa_0, \nu_0, \sigma_0^2\right)$

$$p(\mu, \sigma^2) = p(\sigma^2)p(\mu \mid \sigma^2) = \text{Inv-Ga}\left(\sigma^2 \mid \nu_0/2, \nu_0\sigma_0^2/2\right) \cdot \text{Normal}(\mu \mid \mu_0, \sigma^2/\kappa_0)$$
$$\propto \left(\sigma^2\right)^{-\left(\frac{\nu_0}{2}+1+\frac{1}{2}\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + \kappa_0(\mu-\mu_0)^2\right\}\right]$$

- **Normal-Inverse-Gamma Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n})$$
$$\propto \left(\sigma^2\right)^{-\left\{\frac{(\nu_0+n)}{2}+1+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2 + (\kappa_0+n)(\mu-\mu_n)^2\right\}\right]$$
$$\equiv \text{NIG}\left(\mu_n, \sigma_n^2/\kappa_n, \nu_n, \sigma_n^2\right), \quad \nu_n = (\nu_0+n), \quad \kappa_n = (\kappa_0+n), \quad \mu_n = (\kappa_0\mu_0 + n\overline{y})/(\kappa_0+n),$$
$$\sigma_n^2 = \frac{1}{\nu_n}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2\right\}$$

- A-priori and a-posteriori $\mu$ and $\sigma^2$ are dependent but uncorrelated. **HW.4.1a**

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\bar{y}-\mu)^2\right\}\right]$
- **Normal-Inverse-Gamma Prior:** $(\mu, \sigma^2) \sim \text{NIG}\left(\mu_0, \sigma_0^2/\kappa_0, \nu_0, \sigma_0^2\right)$

$$p(\mu, \sigma^2) = p(\sigma^2)p(\mu \mid \sigma^2) = \text{Inv-Ga}\left(\sigma^2 \mid \nu_0/2, \nu_0\sigma_0^2/2\right) \cdot \text{Normal}(\mu \mid \mu_0, \sigma^2/\kappa_0)$$
$$\propto \left(\sigma^2\right)^{-\left(\frac{\nu_0}{2}+1+\frac{1}{2}\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + \kappa_0(\mu-\mu_0)^2\right\}\right]$$

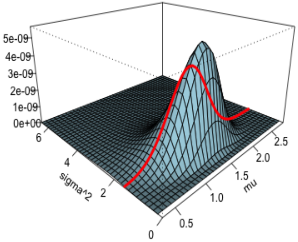
- **Normal-Inverse-Gamma Posterior:**

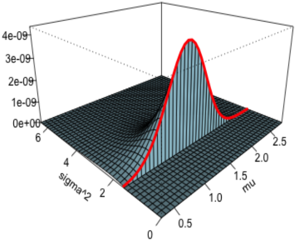
$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n})$$
$$\propto \left(\sigma^2\right)^{-\left\{\frac{(\nu_0+n)}{2}+1+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\bar{y}-\mu_0)^2 + (\kappa_0+n)(\mu-\mu_n)^2\right\}\right]$$
$$\equiv \text{NIG}\left(\mu_n, \sigma_n^2/\kappa_n, \nu_n, \sigma_n^2\right), \quad \nu_n = (\nu_0+n), \quad \kappa_n = (\kappa_0+n), \quad \mu_n = (\kappa_0\mu_0 + n\bar{y})/(\kappa_0+n),$$
$$\sigma_n^2 = \frac{1}{\nu_n}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\bar{y}-\mu_0)^2\right\}$$

- **Conditional Posteriors:**
  - $p(\mu \mid \sigma^2, \mathbf{y}_{1:n}) = \text{Normal}(\mu_n, \sigma^2/\kappa_n)$
  - $p(\sigma^2 \mid \mu, \mathbf{y}_{1:n}) = \text{Inv-Ga}[(\nu_n+1)/2, \{\kappa_n(\mu-\mu_n)^2 + \nu_n\sigma_n^2\}/2]$

**(WB)**

- **Conditional Posterior of $\mu$ for given values of $\sigma^2$:**
  - $p(\mu \mid \sigma^2, \mathbf{y}_{1:n}) = \text{Normal}(\mu_n, \sigma^2/\kappa_n)$




The blue surface shows the joint NIG posterior $p(\mu, \sigma^2 \mid \mathbf{y}_{1:n})$. The sliced red curve shows the conditional Normal posterior $p(\mu \mid \sigma^2, \mathbf{y}_{1:n})$ for $\sigma^2 \approx 1.7$.

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Normal-Inverse-Gamma Prior:** $(\mu, \sigma^2) \sim \text{NIG}\left(\mu_0, \sigma_0^2/\kappa_0, \nu_0, \sigma_0^2\right)$

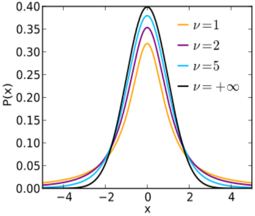
$$p(\mu, \sigma^2) = p(\sigma^2)p(\mu \mid \sigma^2) = \text{Inv-Ga}\left(\sigma^2 \mid \nu_0/2, \nu_0\sigma_0^2/2\right) \cdot \text{Normal}(\mu \mid \mu_0, \sigma^2/\kappa_0)$$
$$\propto \left(\sigma^2\right)^{-\left(\frac{\nu_0}{2}+1+\frac{1}{2}\right)} \exp\left[-\tfrac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + \kappa_0(\mu-\mu_0)^2\right\}\right]$$

- **Normal-Inverse-Gamma Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n})$$
$$\propto \left(\sigma^2\right)^{-\left\{\frac{(\nu_0+n)}{2}+1+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2 + (\kappa_0+n)(\mu-\mu_n)^2\right\}\right]$$
$$\equiv \text{NIG}\left(\mu_n, \sigma_n^2/\kappa_n, \nu_n, \sigma_n^2\right), \quad \nu_n = (\nu_0+n), \quad \kappa_n = (\kappa_0+n), \quad \mu_n = (\kappa_0\mu_0 + n\overline{y})/(\kappa_0+n),$$
$$\sigma_n^2 = \frac{1}{\nu_n}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2\right\}$$

- **Marginal Posteriors:**
  - $p(\mu \mid \mathbf{y}_{1:n}) = t_{\nu_n}(\mu_n, \sigma_n^2/\kappa_n)$
  - $p(\sigma^2 \mid \mathbf{y}_{1:n}) = \text{Inv-Ga}(\nu_n/2, \nu_n\sigma_n^2/2)$ (WB)

# Normal Model
## t-distributions




- **Student's t-distribution:**

$$p(y \mid \nu) = \frac{1}{\sqrt{\nu}\,\mathrm{Beta}\left(\frac{\nu+1}{2}, \frac{1}{2}\right)} \cdot \left(1 + \frac{y^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}$$

$$\mathbb{E}(y) = 0, \quad \nu > 1, \qquad \mathrm{var}(y) = \frac{\nu}{(\nu-2)}, \quad \nu > 2$$

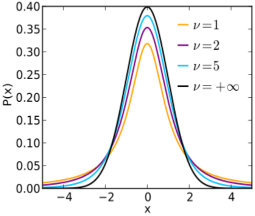
- $p(y \mid \nu) \to \mathrm{Normal}(0, 1)$ as $\nu \to \infty$.

- **Scaled and Shifted t-distribution:**

$$p(y \mid \nu, \mu, \sigma^2) = \frac{1}{\sqrt{\nu}\,\mathrm{Beta}\left(\frac{\nu+1}{2}, \frac{1}{2}\right)\sigma} \cdot \left\{1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right\}^{-\frac{(\nu+1)}{2}}$$

$$\mathbb{E}(y) = \mu, \quad \nu > 1, \qquad \mathrm{var}(y) = \sigma^2 \frac{\nu}{(\nu-2)}, \quad \nu > 2$$

- $p(y \mid \nu, 0, 1) \equiv$ Student's t-distribution.




- **Student's t-distribution:**

$$p(y \mid \nu) = \frac{1}{\sqrt{\nu}\,\text{Beta}\left(\frac{\nu+1}{2}, \frac{1}{2}\right)} \cdot \left(1 + \frac{y^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}$$

$$\mathbb{E}(y) = 0, \quad \nu > 1, \qquad \text{var}(y) = \frac{\nu}{(\nu-2)}, \quad \nu > 2$$

- $p(y \mid \nu) \to \text{Normal}(0,1)$ as $\nu \to \infty$.

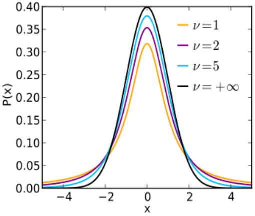
- **Scaled and Shifted t-distribution:**

$$p(y \mid \nu, \mu, \sigma^2) = \frac{1}{\sqrt{\nu}\,\text{Beta}\left(\frac{\nu+1}{2}, \frac{1}{2}\right)\sigma} \cdot \left\{1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right\}^{-\frac{(\nu+1)}{2}}$$

$$\mathbb{E}(y) = \mu, \quad \nu > 1, \qquad \text{var}(y) = \sigma^2 \frac{\nu}{(\nu-2)}, \quad \nu > 2$$

- $p(y \mid \nu, 0, 1) \equiv$ Student's t-distribution.

# Normal Model
## t-distributions




- **Student's t-distribution:**

$$p(y \mid \nu) = \frac{1}{\sqrt{\nu}\,\text{Beta}\left(\frac{\nu+1}{2}, \frac{1}{2}\right)} \cdot \left(1 + \frac{y^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}$$

$$\mathbb{E}(y) = 0, \quad \nu > 1, \qquad \text{var}(y) = \frac{\nu}{(\nu-2)}, \quad \nu > 2$$

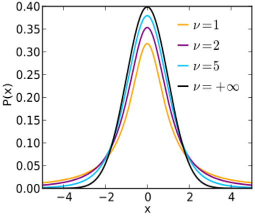
- $p(y \mid \nu) \to \text{Normal}(0,1)$ as $\nu \to \infty$.

- **Scaled and Shifted t-distribution:**

$$p(y \mid \nu, \mu, \sigma^2) = \frac{1}{\sqrt{\nu}\,\text{Beta}\left(\frac{\nu+1}{2}, \frac{1}{2}\right)\,\sigma} \cdot \left\{1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right\}^{-\frac{(\nu+1)}{2}}$$

$$\mathbb{E}(y) = \mu, \quad \nu > 1, \qquad \text{var}(y) = \sigma^2\frac{\nu}{(\nu-2)}, \quad \nu > 2$$

- $p(y \mid \nu, 0, 1) \equiv$ Student's t-distribution.




- **Student's t-distribution:**

$$p(y \mid \nu) = \frac{1}{\sqrt{\nu}\,\mathrm{Beta}\left(\frac{\nu+1}{2}, \frac{1}{2}\right)} \cdot \left(1 + \frac{y^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}$$

$$\mathbb{E}(y) = 0, \quad \nu > 1, \qquad \mathrm{var}(y) = \frac{\nu}{(\nu-2)}, \quad \nu > 2$$

- $p(y \mid \nu) \rightarrow \mathrm{Normal}(0, 1)$ as $\nu \rightarrow \infty$.

- **Scaled and Shifted t-distribution:**

$$p(y \mid \nu, \mu, \sigma^2) = \frac{1}{\sqrt{\nu}\,\mathrm{Beta}\left(\frac{\nu+1}{2}, \frac{1}{2}\right)\ \sigma} \cdot \left\{1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right\}^{-\frac{(\nu+1)}{2}}$$

$$\mathbb{E}(y) = \mu, \quad \nu > 1, \qquad \mathrm{var}(y) = \sigma^2 \frac{\nu}{(\nu-2)}, \quad \nu > 2$$

- $p(y \mid \nu, 0, 1) \equiv$ Student's t-distribution.

$$y_1, \ldots, y_n \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Normal-Inverse-Gamma Prior:** $(\mu, \sigma^2) \sim \text{NIG}\left(\mu_0, \sigma_0^2/\kappa_0, \nu_0, \sigma_0^2\right)$

$$p(\mu, \sigma^2) = p(\sigma^2)p(\mu \mid \sigma^2) = \text{Inv-Ga}\left(\sigma^2 \mid \nu_0/2, \nu_0\sigma_0^2/2\right) \cdot \text{Normal}(\mu \mid \mu_0, \sigma^2/\kappa_0)$$
$$\propto \left(\sigma^2\right)^{-\left(\frac{\nu_0}{2}+1+\frac{1}{2}\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + \kappa_0(\mu-\mu_0)^2\right\}\right]$$

- **Normal-Inverse-Gamma Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n})$$
$$\propto \left(\sigma^2\right)^{-\left\{\frac{(\nu_0+n)}{2}+1+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2 + (\kappa_0+n)(\mu-\mu_n)^2\right\}\right]$$
$$\equiv \text{NIG}\left(\mu_n, \sigma_n^2/\kappa_n, \nu_n, \sigma_n^2\right), \quad \nu_n = (\nu_0+n), \quad \kappa_n = (\kappa_0+n), \quad \mu_n = (\kappa_0\mu_0 + n\overline{y})/(\kappa_0+n),$$
$$\sigma_n^2 = \frac{1}{\nu_n}\left\{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{(n+\kappa_0)}(\overline{y}-\mu_0)^2\right\}$$

- **Predictive Distribution:**
  - $p(y_{new} \mid \mathbf{y}_{1:n}) = t_{\nu_n}\left\{\mu_n, \left(1+\frac{1}{\kappa_n}\right)\sigma_n^2\right\}$ (WB)

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y} - \mu)^2\right\}\right]$
- **Non-informative Improper Prior:** $p(\mu, \sigma^2) \propto 1 \cdot \frac{1}{\sigma^2}$
- **Posterior:** $p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\mu - \overline{y})^2\right\}\right]$
- A-posteriori $\mu$ and $\sigma^2$ are dependent but uncorrelated.

# Normal Model
## Non-informative Improper Prior

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Non-informative Improper Prior:** $p(\mu, \sigma^2) \propto 1 \cdot \dfrac{1}{\sigma^2}$
- **Posterior:** $p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\mu-\overline{y})^2\right\}\right]$
- A-posteriori $\mu$ and $\sigma^2$ are dependent but uncorrelated.

# Normal Model
## Non-informative Improper Prior

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y} - \mu)^2\right\}\right]$
- **Non-informative Improper Prior:** $p(\mu, \sigma^2) \propto 1 \cdot \frac{1}{\sigma^2}$
- **Posterior:** $p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto (\sigma^2)^{-(\frac{n}{2}+1)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\mu - \overline{y})^2\right\}\right]$
- A-posteriori $\mu$ and $\sigma^2$ are dependent but uncorrelated.

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Non-informative Improper Prior:** $p(\mu, \sigma^2) \propto 1 \cdot \frac{1}{\sigma^2}$
- **Posterior:** $p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\mu-\overline{y})^2\right\}\right]$
- A-posteriori $\mu$ and $\sigma^2$ are dependent but uncorrelated. **HW.4.1b**

# Normal Model
## Non-informative Improper Prior

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Non-informative Improper Prior:** $p(\mu, \sigma^2) \propto 1 \cdot \frac{1}{\sigma^2}$
- **Posterior:** $p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\mu-\overline{y})^2\right\}\right]$
- A-posteriori $\mu$ and $\sigma^2$ are dependent but uncorrelated.
- **Conditional Posteriors:**
  - $p(\mu \mid \sigma^2, \mathbf{y}_{1:n}) = \text{Normal}(\overline{y}, \sigma^2/n)$
  - $p(\sigma^2 \mid \mu, \mathbf{y}_{1:n}) = \text{Inv-Ga}\left\{\frac{n}{2}, \frac{(n-1)s^2 + n(\mu-\overline{y})^2}{2}\right\}$
- **Marginal Posteriors:**
  - $p(\mu \mid \mathbf{y}_{1:n}) = t_{n-1}\left(\overline{y}, \frac{s^2}{n}\right)$
  - $p(\sigma^2 \mid \mathbf{y}_{1:n}) = \text{Inv-Ga}\left\{\frac{(n-1)}{2}, \frac{(n-1)s^2}{2}\right\}$
- **Predictive Distribution:** $p(y_{new} \mid \mathbf{y}_{1:n}) = ?$

## Normal Model
## Non-informative Improper Prior

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Non-informative Improper Prior:** $p(\mu, \sigma^2) \propto 1 \cdot \dfrac{1}{\sigma^2}$
- **Posterior:** $p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\mu-\overline{y})^2\right\}\right]$
- A-posteriori $\mu$ and $\sigma^2$ are dependent but uncorrelated.
- **Conditional Posteriors:**
  - $p(\mu \mid \sigma^2, \mathbf{y}_{1:n}) = \text{Normal}(\overline{y}, \sigma^2/n)$
  - $p(\sigma^2 \mid \mu, \mathbf{y}_{1:n}) = \text{Inv-Ga}\left\{\dfrac{n}{2}, \dfrac{(n-1)s^2 + n(\mu-\overline{y})^2}{2}\right\}$
- **Marginal Posteriors:**
  - $p(\mu \mid \mathbf{y}_{1:n}) = t_{n-1}\left(\overline{y}, \dfrac{s^2}{n}\right)$
  - $p(\sigma^2 \mid \mathbf{y}_{1:n}) = \text{Inv-Ga}\left\{\dfrac{(n-1)}{2}, \dfrac{(n-1)s^2}{2}\right\}$
- **Predictive Distribution:** $p(y_{new} \mid \mathbf{y}_{1:n}) = ?$

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Non-informative Improper Prior:** $p(\mu, \sigma^2) \propto 1 \cdot \frac{1}{\sigma^2}$
- **Posterior:** $p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto (\sigma^2)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\mu-\overline{y})^2\right\}\right]$
- A-posteriori $\mu$ and $\sigma^2$ are dependent but uncorrelated.
- **Conditional Posteriors:**
  - $p(\mu \mid \sigma^2, \mathbf{y}_{1:n}) = \text{Normal}(\overline{y}, \sigma^2/n)$
  - $p(\sigma^2 \mid \mu, \mathbf{y}_{1:n}) = \text{Inv-Ga}\left\{\frac{n}{2}, \frac{(n-1)s^2 + n(\mu-\overline{y})^2}{2}\right\}$
- **Marginal Posteriors:**
  - $p(\mu \mid \mathbf{y}_{1:n}) = t_{n-1}\left(\overline{y}, \frac{s^2}{n}\right)$
  - $p(\sigma^2 \mid \mathbf{y}_{1:n}) = \text{Inv-Ga}\left\{\frac{(n-1)}{2}, \frac{(n-1)s^2}{2}\right\}$
- **Predictive Distribution:** $p(y_{new} \mid \mathbf{y}_{1:n}) = ?$

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Non-informative Improper Prior:** $p(\mu, \sigma^2) \propto 1 \cdot \dfrac{1}{\sigma^2}$
- **Posterior:** $p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(\frac{n}{2}+1\right)} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\mu-\overline{y})^2\right\}\right]$
- A-posteriori $\mu$ and $\sigma^2$ are dependent but uncorrelated.
- **Conditional Posteriors:**
  - $p(\mu \mid \sigma^2, \mathbf{y}_{1:n}) = \text{Normal}(\overline{y}, \sigma^2/n)$
  - $p(\sigma^2 \mid \mu, \mathbf{y}_{1:n}) = \text{Inv-Ga}\left\{\dfrac{n}{2}, \dfrac{(n-1)s^2 + n(\mu-\overline{y})^2}{2}\right\}$
- **Marginal Posteriors:**
  - $p(\mu \mid \mathbf{y}_{1:n}) = t_{n-1}\left(\overline{y}, \dfrac{s^2}{n}\right)$
  - $p(\sigma^2 \mid \mathbf{y}_{1:n}) = \text{Inv-Ga}\left\{\dfrac{(n-1)}{2}, \dfrac{(n-1)s^2}{2}\right\}$
- **Predictive Distribution:** $p(y_{new} \mid \mathbf{y}_{1:n}) = ?$ **HW.4.2**

# Normal Model
## Semi-Conjugate Prior

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Independent Normal $\times$ Inverse-Gamma Prior:**

$$(\mu, \sigma^2) \sim \text{Normal}(\mu_0, \sigma_0^2)\ \text{Inv-Ga}(a_0, b_0)$$

$$p(\mu, \sigma^2) = p(\mu)p(\sigma^2) = \text{Normal}(\mu \mid \mu_0, \sigma_0^2) \cdot \text{Inv-Ga}\left(\sigma^2 \mid a_0, b_0\right)$$

$$\propto \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}\left(\sigma^2\right)^{-(a_0+1)} \exp\left(-\frac{b_0}{\sigma^2}\right)$$

- **Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(a_0+\frac{n}{2}+1\right)} \exp\left[-\frac{1}{\sigma^2}\left\{b_0 + \frac{(n-1)s^2}{2} + \frac{n(\overline{y}-\mu)^2}{2}\right\}\right] \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

- **Conditional Posteriors:**

## Normal Model
## Semi-Conjugate Prior

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Independent Normal × Inverse-Gamma Prior:**

$$(\mu, \sigma^2) \sim \text{Normal}(\mu_0, \sigma_0^2)\ \text{Inv-Ga}(a_0, b_0)$$

$$p(\mu, \sigma^2) = p(\mu)p(\sigma^2) = \text{Normal}(\mu \mid \mu_0, \sigma_0^2) \cdot \text{Inv-Ga}\left(\sigma^2 \mid a_0, b_0\right)$$

$$\propto \exp\left\{-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right\}\left(\sigma^2\right)^{-(a_0+1)} \exp\left(-\frac{b_0}{\sigma^2}\right)$$

- **Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(a_0 + \frac{n}{2} + 1\right)} \exp\left[-\frac{1}{\sigma^2}\left\{b_0 + \frac{(n-1)s^2}{2} + \frac{n(\overline{y}-\mu)^2}{2}\right\}\right] \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

- **Conditional Posteriors:**

# Normal Model
## Semi-Conjugate Prior

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Independent Normal $\times$ Inverse-Gamma Prior:**

$$(\mu, \sigma^2) \sim \text{Normal}(\mu_0, \sigma_0^2)\ \text{Inv-Ga}(a_0, b_0)$$

$$p(\mu, \sigma^2) = p(\mu)p(\sigma^2) = \text{Normal}(\mu \mid \mu_0, \sigma_0^2) \cdot \text{Inv-Ga}\left(\sigma^2 \mid a_0, b_0\right)$$

$$\propto \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}\left(\sigma^2\right)^{-(a_0+1)} \exp\left(-\frac{b_0}{\sigma^2}\right)$$

- **Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(a_0+\frac{n}{2}+1\right)} \exp\left[-\frac{1}{\sigma^2}\left\{b_0 + \frac{(n-1)s^2}{2} + \frac{n(\overline{y}-\mu)^2}{2}\right\}\right] \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

# Normal Model
## Semi-Conjugate Prior

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Independent Normal $\times$ Inverse-Gamma Prior:**

$$(\mu, \sigma^2) \sim \text{Normal}(\mu_0, \sigma_0^2)\ \text{Inv-Ga}(a_0, b_0)$$

$$p(\mu, \sigma^2) = p(\mu)p(\sigma^2) = \text{Normal}(\mu \mid \mu_0, \sigma_0^2) \cdot \text{Inv-Ga}\left(\sigma^2 \mid a_0, b_0\right)$$

$$\propto \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}\left(\sigma^2\right)^{-(a_0+1)} \exp\left(-\frac{b_0}{\sigma^2}\right)$$

- **Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(a_0+\frac{n}{2}+1\right)} \exp\left[-\frac{1}{\sigma^2}\left\{b_0 + \frac{(n-1)s^2}{2} + \frac{n(\overline{y}-\mu)^2}{2}\right\}\right] \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

- A-priori $\mu$ and $\sigma^2$ are independent but a-posteriori they are not.

$$y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \quad \text{with} \quad \mu, \sigma^2 \text{ both unknown}$$

- **Normal Likelihood:** $p(\mathbf{y}_{1:n} \mid \mu, \sigma^2) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{(n-1)s^2 + n(\overline{y}-\mu)^2\right\}\right]$
- **Independent Normal $\times$ Inverse-Gamma Prior:**

$$(\mu, \sigma^2) \sim \text{Normal}(\mu_0, \sigma_0^2)\ \text{Inv-Ga}(a_0, b_0)$$

$$p(\mu, \sigma^2) = p(\mu)p(\sigma^2) = \text{Normal}(\mu \mid \mu_0, \sigma_0^2) \cdot \text{Inv-Ga}\left(\sigma^2 \mid a_0, b_0\right)$$

$$\propto \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}\left(\sigma^2\right)^{-(a_0+1)} \exp\left(-\frac{b_0}{\sigma^2}\right)$$

- **Posterior:**

$$p(\mu, \sigma^2 \mid \mathbf{y}_{1:n}) \propto \left(\sigma^2\right)^{-\left(a_0+\frac{n}{2}+1\right)} \exp\left[-\frac{1}{\sigma^2}\left\{b_0 + \frac{(n-1)s^2}{2} + \frac{n(\overline{y}-\mu)^2}{2}\right\}\right] \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

- **Conditional Posteriors:**
  - $p(\mu \mid \sigma^2, \mathbf{y}_{1:n}) = \text{Normal}(\mu_n, \sigma_n^2), \quad \mu_n = \sigma_n^2\left(\frac{\mu_0}{\sigma_0^2} + \frac{n\overline{y}}{\sigma^2}\right), \ \sigma_n^2 = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1},$
  - $p(\sigma^2 \mid \mu, \mathbf{y}_{1:n}) = \text{Inv-Ga}(a_n, b_n), \quad a_n = \left(a_0 + \frac{n}{2}\right), \ b_n = \left\{b_0 + \frac{(n-1)s^2}{2} + \frac{n(\overline{y}-\mu)^2}{2}\right\}.$

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu})\right\}$

$$\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \overline{\mathbf{y}}) + \sum_{i=1}^n (\mathbf{y}_i - \overline{\mathbf{y}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \overline{\mathbf{y}})\right\}\right]$$

$$\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\left[n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \overline{\mathbf{y}}) + \text{trace}\{\boldsymbol{\Sigma}^{-1} \sum_{i=1}^n (\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})^{\mathrm{T}}\}\right]\right)$$

$$\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \overline{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$$

- **Wishart Distribution:** $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ has the pdf

$$p(\mathbf{W}) \propto |\mathbf{W}|^{\frac{n-d-1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{W}\boldsymbol{\Omega}^{-1})\right\}, \quad n > d-1,\ \boldsymbol{\Omega} > 0,\ \mathbf{W} \in \mathbf{M}^+.$$

- **Inverse-Wishart Distribution:** If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ and $\mathbf{V} = \mathbf{W}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Omega}^{-1}$, then $\mathbf{V} \sim \text{IW}_d(n, \boldsymbol{\Lambda})$ and has the pdf

$$p(\mathbf{V}) \propto |\mathbf{V}|^{-\frac{n+d+1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{V}^{-1}\boldsymbol{\Lambda})\right\}, \quad n > d-1,\ \boldsymbol{\Lambda} > 0,\ \mathbf{V} \in \mathbf{M}^+.$$

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Wishart Distribution:** $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ has the pdf
  $p(\mathbf{W}) \propto |\mathbf{W}|^{\frac{n-d-1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{W}\boldsymbol{\Omega}^{-1})\right\}, \quad n > d-1,\ \boldsymbol{\Omega} > 0,\ \mathbf{W} \in \mathbf{M}^+.$
  $\mathbf{M}^+$ is the set of all positive definite matrices.
  - If $z_1, \ldots, z_n \sim \text{Normal}(0, \sigma^2)$, then $\sum_{i=1}^{n} z_i^2 \sim \sigma^2\chi_n^2$.
  - If $\mathbf{z}_1, \ldots, \mathbf{z}_n \sim \text{MVN}_d(\mathbf{0}, \boldsymbol{\Omega})$, then $\mathbf{Z}^{\mathrm{T}}\mathbf{Z} \sim W_d(n, \boldsymbol{\Omega})$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then $\mathbb{E}(\mathbf{W}) = n\boldsymbol{\Omega}$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then, for any $\mathbf{B}^{q\times d}$, $\mathbf{B}\mathbf{W}\mathbf{B}^{\mathrm{T}} \sim W_q(n, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathrm{T}})$.
  - If $\mathbf{W} = ((w_{ij})) \sim W_d(n, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = ((\omega_{ij}))$, then $w_{ii} \sim \omega_{ii}\chi_n^2$.
  - If $\mathbf{W}_j \overset{ind}{\sim} W_d(n_j, \boldsymbol{\Omega})$, then $\sum_{j=1}^{J}\mathbf{W}_j \sim W_d(\sum_{j=1}^{J} n_j, \boldsymbol{\Omega})$.
- **Inverse-Wishart Distribution:** If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ and $\mathbf{V} = \mathbf{W}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Omega}^{-1}$, then $\mathbf{V} \sim \text{IW}_d(n, \boldsymbol{\Lambda})$ and has the pdf
  $p(\mathbf{V}) \propto |\mathbf{V}|^{-\frac{n+d+1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{V}^{-1}\boldsymbol{\Lambda})\right\}, \quad n > d-1,\ \boldsymbol{\Lambda} > 0,\ \mathbf{V} \in \mathbf{M}^+.$

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d \times 1}, \boldsymbol{\Sigma}^{d \times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Wishart Distribution:** $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ has the pdf
  $p(\mathbf{W}) \propto |\mathbf{W}|^{\frac{n-d-1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{W}\boldsymbol{\Omega}^{-1})\right\}, \quad n > d-1,\ \boldsymbol{\Omega} > 0,\ \mathbf{W} \in \mathbf{M}^{+}.$
  - If $z_1, \ldots, z_n \sim \text{Normal}(0, \sigma^2)$, then $\sum_{i=1}^{n} z_i^2 \sim \sigma^2\chi_n^2$.
  - If $\mathbf{z}_1, \ldots, \mathbf{z}_n \sim \text{MVN}_d(\mathbf{0}, \boldsymbol{\Omega})$, then $\mathbf{Z}^{\mathrm{T}}\mathbf{Z} \sim W_d(n, \boldsymbol{\Omega})$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then $\mathbb{E}(\mathbf{W}) = n\boldsymbol{\Omega}$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then, for any $\mathbf{B}^{q \times d}$, $\mathbf{B}\mathbf{W}\mathbf{B}^{\mathrm{T}} \sim W_q(n, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathrm{T}})$.
  - If $\mathbf{W} = ((w_{ij})) \sim W_d(n, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = ((\omega_{ij}))$, then $w_{ii} \sim \omega_{ii}\chi_n^2$.
  - If $\mathbf{W}_j \overset{ind}{\sim} W_d(n_j, \boldsymbol{\Omega})$, then $\sum_{j=1}^{J}\mathbf{W}_j \sim W_d(\sum_{j=1}^{J} n_j, \boldsymbol{\Omega})$.
- **Inverse-Wishart Distribution:** If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ and $\mathbf{V} = \mathbf{W}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Omega}^{-1}$, then $\mathbf{V} \sim \text{IW}_d(n, \boldsymbol{\Lambda})$ and has the pdf
  $p(\mathbf{V}) \propto |\mathbf{V}|^{-\frac{n+d+1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{V}^{-1}\boldsymbol{\Lambda})\right\}, \quad n > d-1,\ \boldsymbol{\Lambda} > 0,\ \mathbf{V} \in \mathbf{M}^{+}.$

# Multivariate Normal Model
## Wishart and Inverse Wishart Distributions

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \bar{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \bar{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Wishart Distribution:** $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ has the pdf
  $p(\mathbf{W}) \propto |\mathbf{W}|^{\frac{n-d-1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{W}\boldsymbol{\Omega}^{-1})\right\}, \quad n > d-1,\ \boldsymbol{\Omega} > 0,\ \mathbf{W} \in \mathbf{M}^+.$
  - If $z_1, \ldots, z_n \sim \text{Normal}(0, \sigma^2)$, then $\sum_{i=1}^{n} z_i^2 \sim \sigma^2\chi_n^2$.
  - If $\mathbf{z}_1, \ldots, \mathbf{z}_n \sim \text{MVN}_d(\mathbf{0}, \boldsymbol{\Omega})$, then $\mathbf{Z}^{\mathrm{T}}\mathbf{Z} \sim W_d(n, \boldsymbol{\Omega})$. Here

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1^{\mathrm{T}} \\ \mathbf{z}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{z}_n^{\mathrm{T}} \end{pmatrix} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1d} \\ z_{21} & z_{22} & \cdots & z_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nd} \end{bmatrix}^{n\times d}$$

$$\mathbf{Z}^{\mathrm{T}}\mathbf{Z} = \sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} = \begin{bmatrix} \sum_{i=1}^{n} z_{i1}^2 & \sum_{i=1}^{n} z_{i1}z_{i2} & \cdots & \sum_{i=1}^{n} z_{i1}z_{id} \\ \sum_{i=1}^{n} z_{i1}z_{i2} & \sum_{i=1}^{n} z_{i2}^2 & \cdots & \sum_{i=1}^{n} z_{i2}z_{id} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} z_{i1}z_{id} & \sum_{i=1}^{n} z_{i2}z_{id} & \cdots & \sum_{i=1}^{n} z_{id}^2 \end{bmatrix}^{d\times d}$$

  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then $\mathbb{E}(\mathbf{W}) = n\boldsymbol{\Omega}$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then, for any $\mathbf{B}^{q\times d}$, $\mathbf{B}\mathbf{W}\mathbf{B}^{\mathrm{T}} \sim W_q(n, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathrm{T}})$.
  - If $\mathbf{W} = ((w_{ij})) \sim W_d(n, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = ((\omega_{ij}))$, then $w_{ii} \sim \omega_{ii}\chi_n^2$.

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \mathrm{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Wishart Distribution:** $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ has the pdf
  $p(\mathbf{W}) \propto |\mathbf{W}|^{\frac{n-d-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{trace}(\mathbf{W}\boldsymbol{\Omega}^{-1})\right\}, \quad n > d-1,\ \boldsymbol{\Omega} > 0,\ \mathbf{W} \in \mathbf{M}^{+}.$
  - If $z_1, \ldots, z_n \sim \mathrm{Normal}(0, \sigma^2)$, then $\sum_{i=1}^{n} z_i^2 \sim \sigma^2\chi_n^2$.
  - If $\mathbf{z}_1, \ldots, \mathbf{z}_n \sim \mathrm{MVN}_d(\mathbf{0}, \boldsymbol{\Omega})$, then $\mathbf{Z}^{\mathrm{T}}\mathbf{Z} \sim W_d(n, \boldsymbol{\Omega})$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then $\mathbb{E}(\mathbf{W}) = n\boldsymbol{\Omega}$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then, for any $\mathbf{B}^{q\times d}$, $\mathbf{BWB}^{\mathrm{T}} \sim W_q(n, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathrm{T}})$.
  - If $\mathbf{W} = ((w_{ij})) \sim W_d(n, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = ((\omega_{ij}))$, then $w_{ii} \sim \omega_{ii}\chi_n^2$.
  - If $\mathbf{W}_j \overset{ind}{\sim} W_d(n_j, \boldsymbol{\Omega})$, then $\sum_{j=1}^{J}\mathbf{W}_j \sim W_d(\sum_{j=1}^{J} n_j, \boldsymbol{\Omega})$.
- **Inverse-Wishart Distribution:** If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ and $\mathbf{V} = \mathbf{W}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Omega}^{-1}$, then $\mathbf{V} \sim \mathrm{IW}_d(n, \boldsymbol{\Lambda})$ and has the pdf
  $p(\mathbf{V}) \propto |\mathbf{V}|^{-\frac{n+d+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{trace}(\mathbf{V}^{-1}\boldsymbol{\Lambda})\right\}, \quad n > d-1,\ \boldsymbol{\Lambda} > 0,\ \mathbf{V} \in \mathbf{M}^{+}.$
  - If $\mathbf{V} \sim \mathrm{IW}_d(n, \boldsymbol{\Lambda})$, then $\mathbb{E}(\mathbf{V}) = \boldsymbol{\Lambda}/(n-d-1)$

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \mathrm{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \bar{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \bar{\mathbf{y}}) + \mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Wishart Distribution:** $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ has the pdf
  $p(\mathbf{W}) \propto |\mathbf{W}|^{\frac{n-d-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{trace}(\mathbf{W}\boldsymbol{\Omega}^{-1})\right\}, \quad n > d-1,\ \boldsymbol{\Omega} > 0,\ \mathbf{W} \in \mathbf{M}^{+}.$
  - If $z_1, \ldots, z_n \sim \mathrm{Normal}(0, \sigma^2)$, then $\sum_{i=1}^{n} z_i^2 \sim \sigma^2\chi_n^2$.
  - If $\mathbf{z}_1, \ldots, \mathbf{z}_n \sim \mathrm{MVN}_d(\mathbf{0}, \boldsymbol{\Omega})$, then $\mathbf{Z}^{\mathrm{T}}\mathbf{Z} \sim W_d(n, \boldsymbol{\Omega})$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then $\mathbb{E}(\mathbf{W}) = n\boldsymbol{\Omega}$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then, for any $\mathbf{B}^{q\times d}$, $\mathbf{B}\mathbf{W}\mathbf{B}^{\mathrm{T}} \sim W_q(n, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathrm{T}})$.
  - If $\mathbf{W} = ((w_{ij})) \sim W_d(n, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = ((\omega_{ij}))$, then $w_{ii} \sim \omega_{ii}\chi_n^2$.
  - If $\mathbf{W}_j \overset{ind}{\sim} W_d(n_j, \boldsymbol{\Omega})$, then $\sum_{j=1}^{J} \mathbf{W}_j \sim W_d(\sum_{j=1}^{J} n_j, \boldsymbol{\Omega})$.
- **Inverse-Wishart Distribution:** If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ and $\mathbf{V} = \mathbf{W}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Omega}^{-1}$, then $\mathbf{V} \sim \mathrm{IW}_d(n, \boldsymbol{\Lambda})$ and has the pdf
  $p(\mathbf{V}) \propto |\mathbf{V}|^{-\frac{n+d+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{trace}(\mathbf{V}^{-1}\boldsymbol{\Lambda})\right\}, \quad n > d-1,\ \boldsymbol{\Lambda} > 0,\ \mathbf{V} \in \mathbf{M}^{+}.$
  - If $\mathbf{V} \sim \mathrm{IW}_d(n, \boldsymbol{\Lambda})$, then $\mathbb{E}(\mathbf{V}) = \boldsymbol{\Lambda}/(n-d-1)$ provided $n > d+1$.

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \overline{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Wishart Distribution:** $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ has the pdf
  $p(\mathbf{W}) \propto |\mathbf{W}|^{\frac{n-d-1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{W}\boldsymbol{\Omega}^{-1})\right\}, \quad n > d-1,\ \boldsymbol{\Omega} > 0,\ \mathbf{W} \in \mathbf{M}^+.$
  - If $z_1, \ldots, z_n \sim \text{Normal}(0, \sigma^2)$, then $\sum_{i=1}^n z_i^2 \sim \sigma^2 \chi_n^2$.
  - If $\mathbf{z}_1, \ldots, \mathbf{z}_n \sim \text{MVN}_d(\mathbf{0}, \boldsymbol{\Omega})$, then $\mathbf{Z}^{\mathrm{T}}\mathbf{Z} \sim W_d(n, \boldsymbol{\Omega})$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then $\mathbb{E}(\mathbf{W}) = n\boldsymbol{\Omega}$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then, for any $\mathbf{B}^{q\times d}$, $\mathbf{B}\mathbf{W}\mathbf{B}^{\mathrm{T}} \sim W_q(n, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathrm{T}})$.
  - If $\mathbf{W} = ((w_{ij})) \sim W_d(n, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = ((\omega_{ij}))$, then $w_{ii} \sim \omega_{ii}\chi_n^2$.
  - If $\mathbf{W}_j \overset{ind}{\sim} W_d(n_j, \boldsymbol{\Omega})$, then $\sum_{j=1}^J \mathbf{W}_j \sim W_d(\sum_{j=1}^J n_j, \boldsymbol{\Omega})$.
- **Inverse-Wishart Distribution:** If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ and $\mathbf{V} = \mathbf{W}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Omega}^{-1}$, then $\mathbf{V} \sim \text{IW}_d(n, \boldsymbol{\Lambda})$ and has the pdf
  $p(\mathbf{V}) \propto |\mathbf{V}|^{-\frac{n+d+1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{V}^{-1}\boldsymbol{\Lambda})\right\}, \quad n > d-1,\ \boldsymbol{\Lambda} > 0,\ \mathbf{V} \in \mathbf{M}^+.$

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \bar{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \bar{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Wishart Distribution:** $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ has the pdf
  $p(\mathbf{W}) \propto |\mathbf{W}|^{\frac{n-d-1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{W}\boldsymbol{\Omega}^{-1})\right\}, \qquad n > d-1,\ \boldsymbol{\Omega} > 0,\ \mathbf{W} \in \mathbf{M}^{+}.$
  - If $z_1, \ldots, z_n \sim \text{Normal}(0, \sigma^2)$, then $\sum_{i=1}^{n} z_i^2 \sim \sigma^2\chi_n^2$.
  - If $\mathbf{z}_1, \ldots, \mathbf{z}_n \sim \text{MVN}_d(\mathbf{0}, \boldsymbol{\Omega})$, then $\mathbf{Z}^{\mathrm{T}}\mathbf{Z} \sim W_d(n, \boldsymbol{\Omega})$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then $\mathbb{E}(\mathbf{W}) = n\boldsymbol{\Omega}$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then, for any $\mathbf{B}^{q\times d}$, $\mathbf{B}\mathbf{W}\mathbf{B}^{\mathrm{T}} \sim W_q(n, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathrm{T}})$.
  - If $\mathbf{W} = ((w_{ij})) \sim W_d(n, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = ((\omega_{ij}))$, then $w_{ii} \sim \omega_{ii}\chi_n^2$.
  - If $\mathbf{W}_j \overset{ind}{\sim} W_d(n_j, \boldsymbol{\Omega})$, then $\sum_{j=1}^{J}\mathbf{W}_j \sim W_d(\sum_{j=1}^{J} n_j, \boldsymbol{\Omega})$.
- **Inverse-Wishart Distribution:** If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ and $\mathbf{V} = \mathbf{W}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Omega}^{-1}$, then $\mathbf{V} \sim \text{IW}_d(n, \boldsymbol{\Lambda})$ and has the pdf
  $p(\mathbf{V}) \propto |\mathbf{V}|^{-\frac{n+d+1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{V}^{-1}\boldsymbol{\Lambda})\right\}, \qquad n > d-1,\ \boldsymbol{\Lambda} > 0,\ \mathbf{V} \in \mathbf{M}^{+}.$

# Multivariate Normal Model
## Wishart and Inverse Wishart Distributions

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \bar{\mathbf{y}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Wishart Distribution:** $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ has the pdf
  $p(\mathbf{W}) \propto |\mathbf{W}|^{\frac{n-d-1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{W}\boldsymbol{\Omega}^{-1})\right\}, \quad n > d-1, \ \boldsymbol{\Omega} > 0, \ \mathbf{W} \in \mathbf{M}^{+}.$
  - If $z_1, \ldots, z_n \sim \text{Normal}(0, \sigma^2)$, then $\sum_{i=1}^n z_i^2 \sim \sigma^2 \chi_n^2$.
  - If $\mathbf{z}_1, \ldots, \mathbf{z}_n \sim \text{MVN}_d(\mathbf{0}, \boldsymbol{\Omega})$, then $\mathbf{Z}^{\mathrm{T}}\mathbf{Z} \sim W_d(n, \boldsymbol{\Omega})$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then $\mathbb{E}(\mathbf{W}) = n\boldsymbol{\Omega}$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then, for any $\mathbf{B}^{q\times d}$, $\mathbf{B}\mathbf{W}\mathbf{B}^{\mathrm{T}} \sim W_q(n, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathrm{T}})$.
  - If $\mathbf{W} = ((w_{ij})) \sim W_d(n, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = ((\omega_{ij}))$, then $w_{ii} \sim \omega_{ii}\chi_n^2$.
  - If $\mathbf{W}_j \overset{ind}{\sim} W_d(n_j, \boldsymbol{\Omega})$, then $\sum_{j=1}^J \mathbf{W}_j \sim W_d(\sum_{j=1}^J n_j, \boldsymbol{\Omega})$.
- **Inverse-Wishart Distribution:** If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ and $\mathbf{V} = \mathbf{W}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Omega}^{-1}$, then $\mathbf{V} \sim \text{IW}_d(n, \boldsymbol{\Lambda})$ and has the pdf
  $p(\mathbf{V}) \propto |\mathbf{V}|^{-\frac{n+d+1}{2}} \exp\left\{-\frac{1}{2}\text{trace}(\mathbf{V}^{-1}\boldsymbol{\Lambda})\right\}, \quad n > d-1, \ \boldsymbol{\Lambda} > 0, \ \mathbf{V} \in \mathbf{M}^{+}.$
  - If $\mathbf{V} \sim \text{IW}_d(n, \boldsymbol{\Lambda})$, then $\mathbb{E}(\mathbf{V}) = \boldsymbol{\Lambda}/(n-d-1)$ provided $n > d+1$.

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \mathrm{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \bar{\mathbf{y}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{y}}) + \mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Wishart Distribution:** $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ has the pdf
  $p(\mathbf{W}) \propto |\mathbf{W}|^{\frac{n-d-1}{2}} \exp\left\{-\frac{1}{2}\mathrm{trace}(\mathbf{W}\boldsymbol{\Omega}^{-1})\right\}, \quad n > d-1,\ \boldsymbol{\Omega} > 0,\ \mathbf{W} \in \mathbf{M}^+.$
  - If $z_1, \ldots, z_n \sim \mathrm{Normal}(0, \sigma^2)$, then $\sum_{i=1}^n z_i^2 \sim \sigma^2 \chi_n^2$.
  - If $\mathbf{z}_1, \ldots, \mathbf{z}_n \sim \mathrm{MVN}_d(\mathbf{0}, \boldsymbol{\Omega})$, then $\mathbf{Z}^{\mathrm{T}}\mathbf{Z} \sim W_d(n, \boldsymbol{\Omega})$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then $\mathbb{E}(\mathbf{W}) = n\boldsymbol{\Omega}$.
  - If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$, then, for any $\mathbf{B}^{q\times d}$, $\mathbf{B}\mathbf{W}\mathbf{B}^{\mathrm{T}} \sim W_q(n, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^{\mathrm{T}})$.
  - If $\mathbf{W} = ((w_{ij})) \sim W_d(n, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = ((\omega_{ij}))$, then $w_{ii} \sim \omega_{ii}\chi_n^2$.
  - If $\mathbf{W}_j \overset{ind}{\sim} W_d(n_j, \boldsymbol{\Omega})$, then $\sum_{j=1}^J \mathbf{W}_j \sim W_d(\sum_{j=1}^J n_j, \boldsymbol{\Omega})$.
- **Inverse-Wishart Distribution:** If $\mathbf{W} \sim W_d(n, \boldsymbol{\Omega})$ and $\mathbf{V} = \mathbf{W}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{\Omega}^{-1}$, then $\mathbf{V} \sim \mathrm{IW}_d(n, \boldsymbol{\Lambda})$ and has the pdf
  $p(\mathbf{V}) \propto |\mathbf{V}|^{-\frac{n+d+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{trace}(\mathbf{V}^{-1}\boldsymbol{\Lambda})\right\}, \quad n > d-1,\ \boldsymbol{\Lambda} > 0,\ \mathbf{V} \in \mathbf{M}^+.$
  - If $\mathbf{V} \sim \mathrm{IW}_d(n, \boldsymbol{\Lambda})$, then $\mathbb{E}(\mathbf{V}) = \boldsymbol{\Lambda}/(n-d-1)$ provided $n > d+1$.

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **MVN-Inverse-Wishart Prior:** $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim \text{NIW}\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0/\kappa_0, \nu_0, \boldsymbol{\Sigma}_0\right)$

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = p(\boldsymbol{\Sigma})p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}) = \text{IW}\left(\boldsymbol{\Sigma} \mid \nu_0, \boldsymbol{\Sigma}_0\right) \cdot \text{MVN}(\boldsymbol{\mu} \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}/\kappa_0)$$
$$\propto |\boldsymbol{\Sigma}|^{-\left(\frac{\nu_0+d+1}{2}+\frac{1}{2}\right)} \exp\left[-\frac{1}{2}\left\{\text{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0) + \kappa_0(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right\}\right]$$

- **MVN-Inverse-Wishart Posterior:**

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_{1:n}) \propto |\boldsymbol{\Sigma}|^{-\left\{\frac{\nu_0+n+d+1}{2}+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2}\left\{\text{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_n) + (\kappa_0 + n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_n)\right\}\right]$$
$$\equiv \text{NIW}\left(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n/\kappa_n, \nu_n, \boldsymbol{\Sigma}_n\right), \quad \nu_n = (\nu_0 + n), \quad \kappa_n = (\kappa_0 + n), \quad \boldsymbol{\mu}_n = (\kappa_0\boldsymbol{\mu}_0 + n\overline{\mathbf{y}})/(\kappa_0 + n),$$
$$\boldsymbol{\Sigma}_n = \left\{\boldsymbol{\Sigma}_0 + \mathbf{S} + \frac{n\kappa_0}{(n + \kappa_0)}(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)^{\mathrm{T}}\right\}, \quad \mathbf{S} = \sum_{i=1}^{n}(\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})^{\mathrm{T}}.$$

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \mathrm{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \bar{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \bar{\mathbf{y}}) + \mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **MVN-Inverse-Wishart Prior:** $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim \mathrm{NIW}\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0/\kappa_0, \nu_0, \boldsymbol{\Sigma}_0\right)$

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = p(\boldsymbol{\Sigma})p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}) = \mathrm{IW}\left(\boldsymbol{\Sigma} \mid \nu_0, \boldsymbol{\Sigma}_0\right) \cdot \mathrm{MVN}(\boldsymbol{\mu} \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}/\kappa_0)$$
$$\propto |\boldsymbol{\Sigma}|^{-\left(\frac{\nu_0+d+1}{2}+\frac{1}{2}\right)} \exp\left[-\frac{1}{2}\left\{\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0) + \kappa_0(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right\}\right]$$

- **MVN-Inverse-Wishart Posterior:**

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_{1:n}) \propto |\boldsymbol{\Sigma}|^{-\left\{\frac{\nu_0+n+d+1}{2}+\frac{1}{2}\right\}}$$
$$\exp\left[-\frac{1}{2}\left\{\mathrm{trace}\left[\boldsymbol{\Sigma}^{-1}\left\{\boldsymbol{\Sigma}_0 + \mathbf{S} + \frac{n\kappa_0}{(n+\kappa_0)}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)^{\mathrm{T}}\right\}\right] + (\kappa_0 + n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^{\mathrm{T}}\right\}\right]$$
$$\propto |\boldsymbol{\Sigma}|^{-\left\{\frac{\nu_0+n+d+1}{2}+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2}\left\{\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_n) + (\kappa_0 + n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_n)\right\}\right]$$
$$\equiv \mathrm{NIW}\left(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n/\kappa_n, \nu_n, \boldsymbol{\Sigma}_n\right), \quad \nu_n = (\nu_0 + n), \quad \kappa_n = (\kappa_0 + n), \quad \boldsymbol{\mu}_n = (\kappa_0\boldsymbol{\mu}_0 + n\bar{\mathbf{y}})/(\kappa_0 + n),$$
$$\boldsymbol{\Sigma}_n = \left\{\boldsymbol{\Sigma}_0 + \mathbf{S} + \frac{n\kappa_0}{(n+\kappa_0)}(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)^{\mathrm{T}}\right\}, \quad \mathbf{S} = \sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^{\mathrm{T}}.$$

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **MVN-Inverse-Wishart Prior:** $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \sim \text{NIW}\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0/\kappa_0, \nu_0, \boldsymbol{\Sigma}_0\right)$

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = p(\boldsymbol{\Sigma})p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}) = \text{IW}\left(\boldsymbol{\Sigma} \mid \nu_0, \boldsymbol{\Sigma}_0\right) \cdot \text{MVN}(\boldsymbol{\mu} \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}/\kappa_0)$$

$$\propto |\boldsymbol{\Sigma}|^{-\left(\frac{\nu_0+d+1}{2}+\frac{1}{2}\right)} \exp\left[-\frac{1}{2}\left\{\text{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0) + \kappa_0(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right\}\right]$$

- **MVN-Inverse-Wishart Posterior:**

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_{1:n}) \propto |\boldsymbol{\Sigma}|^{-\left\{\frac{\nu_0+n+d+1}{2}+\frac{1}{2}\right\}} \exp\left[-\frac{1}{2}\left\{\text{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_n) + (\kappa_0 + n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_n)\right\}\right]$$

$$\equiv \text{NIW}\left(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n/\kappa_n, \nu_n, \boldsymbol{\Sigma}_n\right), \quad \nu_n = (\nu_0 + n), \quad \kappa_n = (\kappa_0 + n), \quad \boldsymbol{\mu}_n = (\kappa_0\boldsymbol{\mu}_0 + n\overline{\mathbf{y}})/(\kappa_0 + n),$$

$$\boldsymbol{\Sigma}_n = \left\{\boldsymbol{\Sigma}_0 + \mathbf{S} + \frac{n\kappa_0}{(n + \kappa_0)}(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)(\overline{\mathbf{y}} - \boldsymbol{\mu}_0)^{\mathrm{T}}\right\}, \quad \mathbf{S} = \sum_{i=1}^{n}(\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})^{\mathrm{T}}.$$

- **Conditional Posteriors:**
  - $p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathbf{y}_{1:n}) = \text{MVN}_d(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}/\kappa_n)$
  - $p(\boldsymbol{\Sigma} \mid \boldsymbol{\mu}, \mathbf{y}_{1:n}) = \text{IW}\{\nu_n + 1, \boldsymbol{\Sigma}_n + \kappa_n(\boldsymbol{\mu} - \boldsymbol{\mu}_n)(\boldsymbol{\mu} - \boldsymbol{\mu}_n)^{\mathrm{T}}\}$

# Multivariate Normal Model
## Improper Prior

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Improper Prior:** $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{d+1}{2}}$
- **Posterior:**

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_{1:n}) \propto |\boldsymbol{\Sigma}|^{-\left\{\frac{n+d+1}{2}\right\}} \exp\left[-\frac{1}{2}\left\{\text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}})\right\}\right]$$

  $\equiv \text{NIW}(\overline{\mathbf{y}}, \mathbf{S}/n, n-1, \mathbf{S})$
- **Conditional Posteriors:**
  - $p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathbf{y}_{1:n}) = \text{MVN}_d(\overline{\mathbf{y}}, \boldsymbol{\Sigma}/n),$
  - $p(\boldsymbol{\Sigma} \mid \boldsymbol{\mu}, \mathbf{y}_{1:n}) = \text{IW}\{\nu_n, \mathbf{S} + n(\boldsymbol{\mu} - \overline{\mathbf{y}})(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\}, \quad \nu_n = (n-1).$

# Multivariate Normal Model
## Improper Prior

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \mathrm{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$

  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$

- **Improper Prior:** $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{d+1}{2}}$

- **Posterior:**

  $$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_{1:n}) \propto |\boldsymbol{\Sigma}|^{-\left\{\frac{n+d+1}{2}\right\}} \exp\left[-\frac{1}{2}\left\{\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}})\right\}\right]$$

  $\equiv \mathrm{NIW}\left(\overline{\mathbf{y}}, \mathbf{S}/n, n-1, \mathbf{S}\right)$

- **Conditional Posteriors:**
  - $p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathbf{y}_{1:n}) = \mathrm{MVN}_d(\overline{\mathbf{y}}, \boldsymbol{\Sigma}/n)$,
  - $p(\boldsymbol{\Sigma} \mid \boldsymbol{\mu}, \mathbf{y}_{1:n}) = \mathrm{IW}\{\nu_n, \mathbf{S} + n(\boldsymbol{\mu} - \overline{\mathbf{y}})(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\}, \quad \nu_n = (n-1)$.

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Improper Prior:** $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{d+1}{2}}$
- **Posterior:**
  $p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_{1:n}) \propto |\boldsymbol{\Sigma}|^{-\left\{\frac{n+d+1}{2}\right\}} \exp\left[-\frac{1}{2}\left\{\text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}})\right\}\right]$
  $\equiv \text{NIW}\left(\overline{\mathbf{y}}, \mathbf{S}/n, n-1, \mathbf{S}\right)$
- **Conditional Posteriors:**
  - $p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathbf{y}_{1:n}) = \text{MVN}_d(\overline{\mathbf{y}}, \boldsymbol{\Sigma}/n)$,
  - $p(\boldsymbol{\Sigma} \mid \boldsymbol{\mu}, \mathbf{y}_{1:n}) = \text{IW}\{\nu_n, \mathbf{S} + n(\boldsymbol{\mu} - \overline{\mathbf{y}})(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\}, \quad \nu_n = (n-1)$.

# Multivariate Normal Model
## Semi-Conjugate Prior

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \mathrm{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Prior:** $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathrm{MVN}(\boldsymbol{\mu} \mid \boldsymbol{\mu}_0, \boldsymbol{\Omega}_0) \times \mathrm{IW}(\boldsymbol{\Sigma} \mid \nu_0, \boldsymbol{\Sigma}_0)$
- **Posterior:**

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_{1:n}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right\} |\boldsymbol{\Sigma}|^{-\left\{\frac{\nu_0+d}{2}+\frac{1}{2}\right\}} \exp\left\{-\frac{1}{2}\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0)\right\}$$

$$\times \; |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}})\right\}\right].$$

- **Conditional Posteriors:**
  - $p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathbf{y}_{1:n}) = \mathrm{MVN}_d(\boldsymbol{\mu}_n, \boldsymbol{\Omega}_n)$,

  $\boldsymbol{\Omega}_n^{-1} = (\boldsymbol{\Omega}_0^{-1} + n\boldsymbol{\Sigma}^{-1}), \quad \boldsymbol{\mu}_n = \boldsymbol{\Omega}_n(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_0 + n\boldsymbol{\Sigma}^{-1}\overline{\mathbf{y}})$,
  - $p(\boldsymbol{\Sigma} \mid \boldsymbol{\mu}, \mathbf{y}_{1:n}) = \mathrm{IW}\{\nu_n, \boldsymbol{\Sigma}_0 + \mathbf{S} + n(\boldsymbol{\mu} - \overline{\mathbf{y}})(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\}, \quad \nu_n = (n + \nu_0)$.

# Multivariate Normal Model
## Semi-Conjugate Prior

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \stackrel{iid}{\sim} \mathrm{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Prior:** $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathrm{MVN}(\boldsymbol{\mu} \mid \boldsymbol{\mu}_0, \boldsymbol{\Omega}_0) \times \mathrm{IW}(\boldsymbol{\Sigma} \mid \nu_0, \boldsymbol{\Sigma}_0)$
- **Posterior:**

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_{1:n}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right\} |\boldsymbol{\Sigma}|^{-\left\{\frac{\nu_0+d}{2}+\frac{1}{2}\right\}} \exp\left\{-\frac{1}{2}\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0)\right\}$$

$$\times \; |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{\mathrm{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}})\right\}\right].$$

- **Conditional Posteriors:**
  - $p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathbf{y}_{1:n}) = \mathrm{MVN}_d(\boldsymbol{\mu}_n, \boldsymbol{\Omega}_n)$,

  $\boldsymbol{\Omega}_n^{-1} = (\boldsymbol{\Omega}_0^{-1} + n\boldsymbol{\Sigma}^{-1}), \quad \boldsymbol{\mu}_n = \boldsymbol{\Omega}_n(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_0 + n\boldsymbol{\Sigma}^{-1}\overline{\mathbf{y}})$,

  - $p(\boldsymbol{\Sigma} \mid \boldsymbol{\mu}, \mathbf{y}_{1:n}) = \mathrm{IW}\{\nu_n, \boldsymbol{\Sigma}_0 + \mathbf{S} + n(\boldsymbol{\mu} - \overline{\mathbf{y}})(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\}, \quad \nu_n = (n + \nu_0)$.

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu}^{d\times 1}, \boldsymbol{\Sigma}^{d\times d} \text{ both unknown}$$

- **MVN Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}$
  $\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}}) + \text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right\}\right]$
- **Prior:** $p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \text{MVN}(\boldsymbol{\mu} \mid \boldsymbol{\mu}_0, \boldsymbol{\Omega}_0) \times \text{IW}(\boldsymbol{\Sigma} \mid \nu_0, \boldsymbol{\Sigma}_0)$
- **Posterior:**

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_{1:n}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right\} |\boldsymbol{\Sigma}|^{-\left\{\frac{\nu_0+d}{2}+\frac{1}{2}\right\}} \exp\left\{-\frac{1}{2}\text{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0)\right\}$$

$$\times \; |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\left\{\text{trace}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + n(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \overline{\mathbf{y}})\right\}\right].$$

- **Conditional Posteriors:**
  - $p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathbf{y}_{1:n}) = \text{MVN}_d(\boldsymbol{\mu}_n, \boldsymbol{\Omega}_n)$,
    $\boldsymbol{\Omega}_n^{-1} = (\boldsymbol{\Omega}_0^{-1} + n\boldsymbol{\Sigma}^{-1}), \quad \boldsymbol{\mu}_n = \boldsymbol{\Omega}_n(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_0 + n\boldsymbol{\Sigma}^{-1}\overline{\mathbf{y}})$,
  - $p(\boldsymbol{\Sigma} \mid \boldsymbol{\mu}, \mathbf{y}_{1:n}) = \text{IW}\{\nu_n, \boldsymbol{\Sigma}_0 + \mathbf{S} + n(\boldsymbol{\mu} - \overline{\mathbf{y}})(\boldsymbol{\mu} - \overline{\mathbf{y}})^{\mathrm{T}}\}, \quad \nu_n = (n + \nu_0)$.

# Multinomial Model
## Dirichlet Distributions

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \ldots, \pi_K)$$

$$\text{with } \sum_{k=1}^{K} \pi_k = 1 \text{ and } \mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iK})^{K \times 1} \text{ for all } i$$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^{n} \left\{ \frac{m!}{y_{i1}! \ldots y_{iK}!} \pi_1^{y_{i1}} \ldots \pi_K^{y_{iK}} \right\}$
  $\propto \pi_1^{\sum_{i=1}^{n} y_{i1}} \ldots \pi_K^{\sum_{i=1}^{n} y_{iK}} = \pi_1^{S_1} \ldots \pi_K^{S_K}$

- **Dirichlet Distributions:** $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$ has the pdf

$$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \ldots \pi_K^{\alpha_K - 1}, \quad \sum_{i=1}^{K} \pi_k = 1, \quad \alpha_k > 0 \;\forall\, k.$$

# Multinomial Model
## Dirichlet Distributions

$$\mathbf{y}_1, \dots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \dots, \pi_K)$$
$$\text{with } \textstyle\sum_{k=1}^K \pi_k = 1 \text{ and } \mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \dots, y_{iK})^{K \times 1} \text{ for all } i$$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^n \left\{ \frac{m!}{y_{i1}! \dots y_{iK}!} \pi_1^{y_{i1}} \dots \pi_K^{y_{iK}} \right\}$
$$\propto \pi_1^{\sum_{i=1}^n y_{i1}} \dots \pi_K^{\sum_{i=1}^n y_{iK}} = \pi_1^{S_1} \dots \pi_K^{S_K}$$

- **Dirichlet Distributions:** $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$ has the pdf
$$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_K)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \dots \pi_K^{\alpha_K - 1}, \qquad \sum_{i=1}^K \pi_k = 1, \quad \alpha_k > 0 \; \forall \, k.$$

  - $\text{Dir}(\alpha_1, \alpha_2) \equiv \text{Beta}(\alpha_1, \alpha_2)$.
  - $\mathbb{E}(\pi_k) = \frac{\alpha_k}{\alpha}, \qquad \text{var}(\pi_k) = \frac{\alpha_k(\alpha - \alpha_k)}{\alpha^2(\alpha + 1)}, \qquad \alpha = \alpha_1 + \dots + \alpha_K.$
  - If $z_k \overset{ind}{\sim} \text{Ga}(\alpha_k, 1)$ and $\pi_k \propto z_k$, then $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$. $\rightarrow$ **important result!**
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$, then $(\pi_1 + \dots + \pi_j, \pi_{j+1}, \dots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1 + \dots + \alpha_j, \alpha_{j+1}, \dots, \alpha_K)$.
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$, then $\left( \frac{\pi_1}{\sum_{j=1}^J \pi_j}, \dots, \frac{\pi_J}{\sum_{j=1}^J \pi_j} \right)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \dots, \alpha_J)$.

# Multinomial Model
## Dirichlet Distributions

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \ldots, \pi_K)$$
$$\text{with } \sum_{k=1}^{K} \pi_k = 1 \text{ and } \mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iK})^{K \times 1} \text{ for all } i$$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^{n} \left\{ \frac{m!}{y_{i1}! \ldots y_{iK}!} \pi_1^{y_{i1}} \ldots \pi_K^{y_{iK}} \right\}$
$$\propto \pi_1^{\sum_{i=1}^{n} y_{i1}} \ldots \pi_K^{\sum_{i=1}^{n} y_{iK}} = \pi_1^{S_1} \ldots \pi_K^{S_K}$$
- **Dirichlet Distributions:** $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$ has the pdf
$$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \ldots \pi_K^{\alpha_K - 1}, \qquad \sum_{i=1}^{K} \pi_k = 1, \quad \alpha_k > 0 \; \forall \; k.$$
  - $\text{Dir}(\alpha_1, \alpha_2) \equiv \text{Beta}(\alpha_1, \alpha_2)$.
  - $\mathbb{E}(\pi_k) = \dfrac{\alpha_k}{\alpha}, \qquad \text{var}(\pi_k) = \dfrac{\alpha_k(\alpha - \alpha_k)}{\alpha^2(\alpha + 1)}, \qquad \alpha = \alpha_1 + \cdots + \alpha_K.$
  - If $z_k \overset{ind}{\sim} \text{Ga}(\alpha_k, 1)$ and $\pi_k \propto z_k$, then $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$. → **important result!**
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$, then $(\pi_1 + \cdots + \pi_j, \pi_{j+1}, \ldots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1 + \cdots + \alpha_j, \alpha_{j+1}, \ldots, \alpha_K)$.
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$, then $\left( \frac{\pi_1}{\sum_{j=1}^{J} \pi_j}, \ldots, \frac{\pi_J}{\sum_{j=1}^{J} \pi_j} \right)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \ldots, \alpha_J)$.

## Multinomial Model
## Dirichlet Distributions

$$\mathbf{y}_1,\dots,\mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m,\pi_1,\dots,\pi_K)$$
$$\text{with } \textstyle\sum_{k=1}^K \pi_k = 1 \text{ and } \mathbf{y}_i^{\mathrm{T}} = (y_{i1},\dots,y_{iK})^{K\times 1} \text{ for all } i$$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^n \left\{ \frac{m!}{y_{i1}!\dots y_{iK}!} \pi_1^{y_{i1}} \dots \pi_K^{y_{iK}} \right\}$
  $$\propto \pi_1^{\sum_{i=1}^n y_{i1}} \dots \pi_K^{\sum_{i=1}^n y_{iK}} = \pi_1^{S_1} \dots \pi_K^{S_K}$$
- **Dirichlet Distributions:** $\boldsymbol{\pi} = (\pi_1,\dots,\pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1,\dots,\alpha_K)$ has the pdf
  $$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1+\cdots+\alpha_K)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_K)} \pi_1^{\alpha_1-1} \dots \pi_K^{\alpha_K-1}, \qquad \textstyle\sum_{i=1}^K \pi_k = 1, \quad \alpha_k > 0\ \forall\, k.$$
  - $\text{Dir}(\alpha_1,\alpha_2) \equiv \text{Beta}(\alpha_1,\alpha_2)$.
  - $\mathbb{E}(\pi_k) = \dfrac{\alpha_k}{\alpha}, \qquad \text{var}(\pi_k) = \dfrac{\alpha_k(\alpha-\alpha_k)}{\alpha^2(\alpha+1)}, \qquad \alpha = \alpha_1 + \cdots + \alpha_K.$
  - If $z_k \overset{ind}{\sim} \text{Ga}(\alpha_k, 1)$ and $\pi_k \propto z_k$, then $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1,\dots,\alpha_K)$. → **important result!**
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1,\dots,\alpha_K)$, then $(\pi_1+\cdots+\pi_j, \pi_{j+1},\dots,\pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1+\cdots+\alpha_j, \alpha_{j+1},\dots,\alpha_K)$.
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1,\dots,\alpha_K)$, then $\left(\frac{\pi_1}{\sum_{j=1}^J \pi_j},\dots,\frac{\pi_J}{\sum_{j=1}^J \pi_j}\right)^{\mathrm{T}} \sim \text{Dir}(\alpha_1,\dots,\alpha_J)$.

# Multinomial Model
## Dirichlet Distributions

$$\mathbf{y}_1, \dots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \dots, \pi_K)$$
$$\text{with } \sum_{k=1}^{K} \pi_k = 1 \text{ and } \mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \dots, y_{iK})^{K \times 1} \text{ for all } i$$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^{n} \left\{ \frac{m!}{y_{i1}! \dots y_{iK}!} \pi_1^{y_{i1}} \dots \pi_K^{y_{iK}} \right\}$
$$\propto \pi_1^{\sum_{i=1}^{n} y_{i1}} \dots \pi_K^{\sum_{i=1}^{n} y_{iK}} = \pi_1^{S_1} \dots \pi_K^{S_K}$$
- **Dirichlet Distributions:** $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$ has the pdf
$$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_K)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \dots \pi_K^{\alpha_K - 1}, \qquad \sum_{i=1}^{K} \pi_k = 1, \quad \alpha_k > 0 \; \forall \, k.$$
  - $\text{Dir}(\alpha_1, \alpha_2) \equiv \text{Beta}(\alpha_1, \alpha_2)$.
  - $\mathbb{E}(\pi_k) = \dfrac{\alpha_k}{\alpha}, \qquad \text{var}(\pi_k) = \dfrac{\alpha_k(\alpha - \alpha_k)}{\alpha^2(\alpha + 1)}, \qquad \alpha = \alpha_1 + \dots + \alpha_K.$
  - If $z_k \overset{ind}{\sim} \text{Ga}(\alpha_k, 1)$ and $\pi_k \propto z_k$, then $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$. $\rightarrow$ **important result!**
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$, then $(\pi_1 + \dots + \pi_j, \pi_{j+1}, \dots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1 + \dots + \alpha_j, \alpha_{j+1}, \dots, \alpha_K)$.
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$, then $\left( \frac{\pi_1}{\sum_{j=1}^{J} \pi_j}, \dots, \frac{\pi_J}{\sum_{j=1}^{J} \pi_j} \right)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \dots, \alpha_J)$.

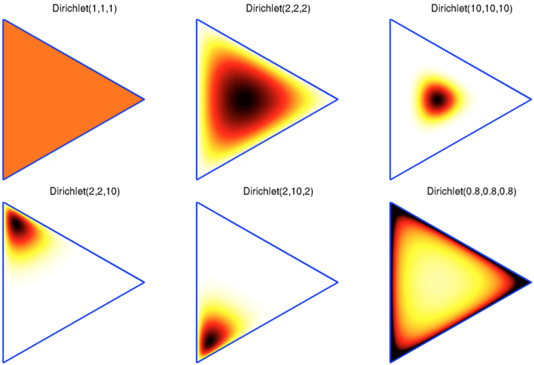
# Multinomial Model
## Dirichlet Distributions

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \ldots, \pi_K)$$
$$\text{with } \textstyle\sum_{k=1}^K \pi_k = 1 \text{ and } \mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iK})^{K\times 1} \text{ for all } i$$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^n \left\{ \frac{m!}{y_{i1}! \ldots y_{iK}!} \pi_1^{y_{i1}} \ldots \pi_K^{y_{iK}} \right\}$
$$\propto \pi_1^{\sum_{i=1}^n y_{i1}} \ldots \pi_K^{\sum_{i=1}^n y_{iK}} = \pi_1^{S_1} \ldots \pi_K^{S_K}$$
- **Dirichlet Distributions:** $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$ has the pdf
$$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \ldots \pi_K^{\alpha_K - 1}, \qquad \sum_{i=1}^K \pi_k = 1, \quad \alpha_k > 0 \; \forall\, k.$$
  - $\text{Dir}(\alpha_1, \alpha_2) \equiv \text{Beta}(\alpha_1, \alpha_2)$.
  - $\mathbb{E}(\pi_k) = \dfrac{\alpha_k}{\alpha}, \qquad \text{var}(\pi_k) = \dfrac{\alpha_k(\alpha - \alpha_k)}{\alpha^2(\alpha+1)}, \qquad \alpha = \alpha_1 + \cdots + \alpha_K.$
  - If $z_k \overset{ind}{\sim} \text{Ga}(\alpha_k, 1)$ and $\pi_k \propto z_k$, then $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$. $\rightarrow$ **important result!**
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$, then $(\pi_1 + \cdots + \pi_j, \pi_{j+1}, \ldots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1 + \cdots + \alpha_j, \alpha_{j+1}, \ldots, \alpha_K)$.
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$, then $\left( \frac{\pi_1}{\sum_{j=1}^J \pi_j}, \ldots, \frac{\pi_J}{\sum_{j=1}^J \pi_j} \right)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \ldots, \alpha_J)$.

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \ldots, \pi_K)$$
$$\text{with } \textstyle\sum_{k=1}^K \pi_k = 1 \text{ and } \mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iK})^{K\times 1} \text{ for all } i$$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^n \left\{ \frac{m!}{y_{i1}!\ldots y_{iK}!} \pi_1^{y_{i1}} \ldots \pi_K^{y_{iK}} \right\}$
$$\propto \pi_1^{\sum_{i=1}^n y_{i1}} \ldots \pi_K^{\sum_{i=1}^n y_{iK}} = \pi_1^{S_1} \ldots \pi_K^{S_K}$$
- **Dirichlet Distributions:** $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$ has the pdf
$$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \ldots \pi_K^{\alpha_K - 1}, \qquad \textstyle\sum_{i=1}^K \pi_k = 1, \quad \alpha_k > 0 \,\forall\, k.$$
  - $\text{Dir}(\alpha_1, \alpha_2) \equiv \text{Beta}(\alpha_1, \alpha_2)$.
  - $\mathbb{E}(\pi_k) = \dfrac{\alpha_k}{\alpha}, \qquad \text{var}(\pi_k) = \dfrac{\alpha_k(\alpha - \alpha_k)}{\alpha^2(\alpha + 1)}, \qquad \alpha = \alpha_1 + \cdots + \alpha_K.$
  - If $z_k \overset{ind}{\sim} \text{Ga}(\alpha_k, 1)$ and $\pi_k \propto z_k$, then $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$. $\rightarrow$ **important result!**
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$, then $(\pi_1 + \cdots + \pi_j, \pi_{j+1}, \ldots, \pi_K)^{\mathrm{T}} \sim \text{Dir}(\alpha_1 + \cdots + \alpha_j, \alpha_{j+1}, \ldots, \alpha_K)$.
  - If $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$, then $\left( \frac{\pi_1}{\sum_{j=1}^J \pi_j}, \ldots, \frac{\pi_J}{\sum_{j=1}^J \pi_j} \right)^{\mathrm{T}} \sim \text{Dir}(\alpha_1, \ldots, \alpha_J)$.


Dirichlet(1,1,1)
Dirichlet(2,2,2)
Dirichlet(10,10,10)
Dirichlet(2,2,10)
Dirichlet(2,10,2)
Dirichlet(0.8,0.8,0.8)

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \ldots, \pi_K)$$
$$\text{with } \pi_k \geq 0 \; \forall k, \; \sum_{k=1}^K \pi_k = 1 \text{ and } \mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iK})^{K \times 1} \text{ for all } i$$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^n \left\{ \frac{m!}{y_{i1}! \ldots y_{iK}!} \pi_1^{y_{i1}} \ldots \pi_K^{y_{iK}} \right\}$
$$\propto \pi_1^{\sum_{i=1}^n y_{i1}} \ldots \pi_K^{\sum_{i=1}^n y_{iK}} = \pi_1^{S_1} \ldots \pi_K^{S_K}$$
- **Dirichlet Prior:** $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$ has the pdf
$$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \ldots \pi_K^{\alpha_K - 1}, \qquad \pi_k \geq 0 \forall k, \; \sum_{i=1}^K \pi_k = 1, \; \alpha_k > 0 \; \forall \; k.$$
- **Posterior:** $p(\boldsymbol{\pi} \mid \mathbf{y}_{1:n}) \propto \pi_1^{\alpha_1 + S_1 - 1} \ldots \pi_K^{\alpha_K + S_K - 1} \equiv \text{Dir}\left(\alpha_1 + S_1, \ldots, \alpha_K + S_K\right)$
- **Jeffreys' NIP:** $\boldsymbol{\pi} \sim \text{Dir}(1/2, \ldots, 1/2)$ so that $p(\boldsymbol{\pi}) \propto \prod_{k=1}^K \pi_k^{-1/2}$.
- **Uniform NIP:** $\boldsymbol{\pi} \sim \text{Dir}(1, \ldots, 1)$

$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \ldots, \pi_K)$

with $\pi_k \geq 0 \; \forall k,\; \sum_{k=1}^{K} \pi_k = 1$ and $\mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iK})^{K \times 1}$ for all $i$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^{n} \left\{ \frac{m!}{y_{i1}! \ldots y_{iK}!} \pi_1^{y_{i1}} \ldots \pi_K^{y_{iK}} \right\}$

$$\propto \pi_1^{\sum_{i=1}^{n} y_{i1}} \ldots \pi_K^{\sum_{i=1}^{n} y_{iK}} = \pi_1^{S_1} \ldots \pi_K^{S_K}$$

- **Dirichlet Prior:** $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$ has the pdf

$$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \ldots \pi_K^{\alpha_K - 1}, \qquad \pi_k \geq 0 \forall k,\; \sum_{i=1}^{K} \pi_k = 1,\; \alpha_k > 0 \; \forall \; k.$$

- **Posterior:** $p(\boldsymbol{\pi} \mid \mathbf{y}_{1:n}) \propto \pi_1^{\alpha_1 + S_1 - 1} \ldots \pi_K^{\alpha_K + S_K - 1} \equiv \text{Dir}\left(\alpha_1 + S_1, \ldots, \alpha_K + S_K\right)$

- **Jeffreys' NIP:** $\boldsymbol{\pi} \sim \text{Dir}(1/2, \ldots, 1/2)$ so that $p(\boldsymbol{\pi}) \propto \prod_{k=1}^{K} \pi_k^{-1/2}$.

- **Uniform NIP:** $\boldsymbol{\pi} \sim \text{Dir}(1, \ldots, 1)$ so that $p(\boldsymbol{\pi}) \propto 1$.

# Multinomial Model
## Conjugate Prior

$$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \ldots, \pi_K)$$

$$\text{with } \pi_k \geq 0 \ \forall k, \ \textstyle\sum_{k=1}^{K} \pi_k = 1 \text{ and } \mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iK})^{K \times 1} \text{ for all } i$$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^{n} \left\{ \frac{m!}{y_{i1}! \ldots y_{iK}!} \pi_1^{y_{i1}} \ldots \pi_K^{y_{iK}} \right\}$

  $$\propto \pi_1^{\sum_{i=1}^{n} y_{i1}} \ldots \pi_K^{\sum_{i=1}^{n} y_{iK}} = \pi_1^{S_1} \ldots \pi_K^{S_K}$$

- **Dirichlet Prior:** $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$ has the pdf

  $$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \ldots \pi_K^{\alpha_K - 1}, \qquad \pi_k \geq 0 \forall k, \ \textstyle\sum_{i=1}^{K} \pi_k = 1, \ \alpha_k > 0 \ \forall \ k.$$

- **Posterior:** $p(\boldsymbol{\pi} \mid \mathbf{y}_{1:n}) \propto \pi_1^{\alpha_1 + S_1 - 1} \ldots \pi_K^{\alpha_K + S_K - 1} \equiv \text{Dir}\left(\alpha_1 + S_1, \ldots, \alpha_K + S_K\right)$

- **Jeffreys' NIP:** $\boldsymbol{\pi} \sim \text{Dir}(1/2, \ldots, 1/2)$ so that $p(\boldsymbol{\pi}) \propto \prod_{k=1}^{K} \pi_k^{-1/2}$.

$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \ldots, \pi_K)$

with $\pi_k \geq 0 \ \forall k, \ \sum_{k=1}^K \pi_k = 1$ and $\mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iK})^{K \times 1}$ for all $i$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^n \left\{ \frac{m!}{y_{i1}! \ldots y_{iK}!} \pi_1^{y_{i1}} \ldots \pi_K^{y_{iK}} \right\}$

$$\propto \pi_1^{\sum_{i=1}^n y_{i1}} \ldots \pi_K^{\sum_{i=1}^n y_{iK}} = \pi_1^{S_1} \ldots \pi_K^{S_K}$$

- **Dirichlet Prior:** $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$ has the pdf

$$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \ldots \pi_K^{\alpha_K - 1}, \qquad \pi_k \geq 0 \forall k, \ \sum_{i=1}^K \pi_k = 1, \ \alpha_k > 0 \ \forall \ k.$$

- **Posterior:** $p(\boldsymbol{\pi} \mid \mathbf{y}_{1:n}) \propto \pi_1^{\alpha_1 + S_1 - 1} \ldots \pi_K^{\alpha_K + S_K - 1} \equiv \text{Dir}\left(\alpha_1 + S_1, \ldots, \alpha_K + S_K\right)$
- **Jeffreys' NIP:** $\boldsymbol{\pi} \sim \text{Dir}(1/2, \ldots, 1/2)$ so that $p(\boldsymbol{\pi}) \propto \prod_{k=1}^K \pi_k^{-1/2}$. **HW.4.3**

# Multinomial Model
## Conjugate Prior

$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \ldots, \pi_K)$

with $\pi_k \geq 0 \; \forall k, \; \sum_{k=1}^K \pi_k = 1$ and $\mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iK})^{K \times 1}$ for all $i$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^n \left\{ \frac{m!}{y_{i1}! \ldots y_{iK}!} \pi_1^{y_{i1}} \ldots \pi_K^{y_{iK}} \right\}$

  $$\propto \pi_1^{\sum_{i=1}^n y_{i1}} \ldots \pi_K^{\sum_{i=1}^n y_{iK}} = \pi_1^{S_1} \ldots \pi_K^{S_K}$$

- **Dirichlet Prior:** $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$ has the pdf

  $$p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \ldots \pi_K^{\alpha_K - 1}, \qquad \pi_k \geq 0 \forall k, \; \sum_{i=1}^K \pi_k = 1, \; \alpha_k > 0 \; \forall \; k.$$

- **Posterior:** $p(\boldsymbol{\pi} \mid \mathbf{y}_{1:n}) \propto \pi_1^{\alpha_1 + S_1 - 1} \ldots \pi_K^{\alpha_K + S_K - 1} \equiv \text{Dir}\left(\alpha_1 + S_1, \ldots, \alpha_K + S_K\right)$
- **Jeffreys' NIP:** $\boldsymbol{\pi} \sim \text{Dir}(1/2, \ldots, 1/2)$ so that $p(\boldsymbol{\pi}) \propto \prod_{k=1}^K \pi_k^{-1/2}$.
- **Uniform NIP:** $\boldsymbol{\pi} \sim \text{Dir}(1, \ldots, 1)$ so that $p(\boldsymbol{\pi}) \propto 1$.
  - In this case, $\widehat{\boldsymbol{\pi}}_{MAP} = \widehat{\boldsymbol{\pi}}_{MLE}$.

$\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{iid}{\sim} \text{Mult}(m, \pi_1, \ldots, \pi_K)$

with $\pi_k \geq 0\ \forall k,\ \sum_{k=1}^{K} \pi_k = 1$ and $\mathbf{y}_i^{\mathrm{T}} = (y_{i1}, \ldots, y_{iK})^{K \times 1}$ for all $i$

- **Multinomial Likelihood:** $p(\mathbf{y}_{1:n} \mid \boldsymbol{\pi}) = \prod_{i=1}^{n} \left\{ \frac{m!}{y_{i1}! \ldots y_{iK}!} \pi_1^{y_{i1}} \ldots \pi_K^{y_{iK}} \right\}$
  $\propto \pi_1^{\sum_{i=1}^{n} y_{i1}} \ldots \pi_K^{\sum_{i=1}^{n} y_{iK}} = \pi_1^{S_1} \ldots \pi_K^{S_K}$
- **Dirichlet Prior:** $\boldsymbol{\pi} \sim \text{Dir}(\alpha_1, \ldots, \alpha_K)$ has the pdf
  $p(\boldsymbol{\pi}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)} \pi_1^{\alpha_1 - 1} \ldots \pi_K^{\alpha_K - 1}, \qquad \pi_k \geq 0 \forall k,\ \sum_{i=1}^{K} \pi_k = 1,\ \alpha_k > 0\ \forall\ k.$
- **Posterior:** $p(\boldsymbol{\pi} \mid \mathbf{y}_{1:n}) \propto \pi_1^{\alpha_1 + S_1 - 1} \ldots \pi_K^{\alpha_K + S_K - 1} \equiv \text{Dir}\left(\alpha_1 + S_1, \ldots, \alpha_K + S_K\right)$
- **Jeffreys' NIP:** $\boldsymbol{\pi} \sim \text{Dir}(1/2, \ldots, 1/2)$ so that $p(\boldsymbol{\pi}) \propto \prod_{k=1}^{K} \pi_k^{-1/2}$.
- **Uniform NIP:** $\boldsymbol{\pi} \sim \text{Dir}(1, \ldots, 1)$ so that $p(\boldsymbol{\pi}) \propto 1$.
  - In this case, $\widehat{\boldsymbol{\pi}}_{MAP} = \widehat{\boldsymbol{\pi}}_{MLE}$.

# Module V Summary

- Normal model
  - Normal-inverse-Gamma prior
  - Improper prior
  - Independent Normal-inverse-Gamma prior
- Multivariate normal model
  - Normal-inverse-Wishart prior
  - Improper prior
  - Independent Normal-inverse-Wishart prior
- Multinomial model
  - Dirichlet prior

# Module V Summary

- Normal model
  - Normal-inverse-Gamma prior
  - Improper prior
  - Independent Normal-inverse-Gamma prior
- Multivariate normal model
  - Normal-inverse-Wishart prior
  - Improper prior
  - Independent Normal-inverse-Wishart prior
- Multinomial model
  - Dirichlet prior

# Module V Summary

- Normal model
  - Normal-inverse-Gamma prior
  - Improper prior
  - Independent Normal-inverse-Gamma prior
- Multivariate normal model
  - Normal-inverse-Wishart prior
  - Improper prior
  - Independent Normal-inverse-Wishart prior
- Multinomial model
  - Dirichlet prior

# Module V Summary

- Normal model
  - Normal-inverse-Gamma prior
  - Improper prior
  - Independent Normal-inverse-Gamma prior
- Multivariate normal model
  - Normal-inverse-Wishart prior
  - Improper prior
  - Independent Normal-inverse-Wishart prior
- Multinomial model
  - Dirichlet prior

# Module V Summary

- Normal model
  - Normal-inverse-Gamma prior
  - Improper prior
  - Independent Normal-inverse-Gamma prior
- Multivariate normal model
  - Normal-inverse-Wishart prior
  - Improper prior
  - Independent Normal-inverse-Wishart prior
- Multinomial model
  - Dirichlet prior

# Module V Summary

- Normal model
  - Normal-inverse-Gamma prior
  - Improper prior
  - Independent Normal-inverse-Gamma prior
- Multivariate normal model
  - Normal-inverse-Wishart prior
  - Improper prior
  - Independent Normal-inverse-Wishart prior
- Multinomial model
  - Dirichlet prior

## Module V Summary

- Normal model
  - Normal-inverse-Gamma prior
  - Improper prior
  - Independent Normal-inverse-Gamma prior
- Multivariate normal model
  - Normal-inverse-Wishart prior
  - Improper prior
  - Independent Normal-inverse-Wishart prior
- Multinomial model
  - Dirichlet prior

# Module V Summary

- Normal model
  - Normal-inverse-Gamma prior
  - Improper prior
  - Independent Normal-inverse-Gamma prior
- Multivariate normal model
  - Normal-inverse-Wishart prior
  - Improper prior
  - Independent Normal-inverse-Wishart prior
- Multinomial model
  - Dirichlet prior

## Module V Summary

- Normal model
    - Normal-inverse-Gamma prior
    - Improper prior
    - Independent Normal-inverse-Gamma prior
- Multivariate normal model
    - Normal-inverse-Wishart prior
    - Improper prior
    - Independent Normal-inverse-Wishart prior
- Multinomial model
    - Dirichlet prior

# Module V Summary

- Normal model
  - Normal-inverse-Gamma prior
  - Improper prior
  - Independent Normal-inverse-Gamma prior
- Multivariate normal model
  - Normal-inverse-Wishart prior
  - Improper prior
  - Independent Normal-inverse-Wishart prior
- Multinomial model
  - Dirichlet prior