STAT7630: Bayesian Statistics
Lecture Slides # 7

Vague Prior for Normal Data & Bayesian Model for Multivariate Normal

Elvan Ceyhan
Department of Mathematics & Statistics
Auburn University

Fall 2024,
Updated: September, 2024

Outline

Vague Priors for Normal Data

Bayesian Model for MVN Data

Vague Priors with Normal Data

- Conjugate priors often include **subjective** prior information.
- An alternative, more objective approach is to use uninformative or vague priors.
- Consider $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, with both μ and σ^2 unknown.
- We can use vague priors:

$$p(\mu) = 1, \quad -\infty < \mu < \infty \quad \text{and} \quad p(\sigma) = \frac{1}{\sigma}, \quad 0 < \sigma < \infty$$

- These priors are improper (they do not integrate to 1), but this is acceptable as long as the resulting posterior distributions are proper densities.

Vague Priors with Normal Data: Joint Posterior for μ and σ

- The joint posterior for μ and σ is:

$$p(\mu, \sigma|\mathbf{x}) \propto p(\mu)p(\sigma)L(\mu, \sigma|\mathbf{x})$$

- The likelihood is:

$$L(\mu, \sigma \mid \mathbf{x}) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)}$$

$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_i-\bar{x})-(\mu-\bar{x})]^2\right)}$$

$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{\left(-\frac{1}{2\sigma^2}\left(\sum(x_i-\bar{x})^2-2\sum(x_i\mu-x_i\bar{x}-\bar{x}\mu+\bar{x}^2)+n(\bar{x}-\mu)^2\right)\right)}$$

$$\propto \sigma^{-n} e^{\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2+n(\bar{x}-\mu)^2\right]\right)}$$

- Combining with the prior $p(\mu) \propto 1$ and $p(\sigma) \propto \frac{1}{\sigma}$:

$$p(\mu, \sigma|\mathbf{x}) \propto \sigma^{-(n+1)} e^{-\frac{1}{2\sigma^2}\left[(n-1)s^2+n(\mu-\bar{x})^2\right]}$$

Vague Priors with Normal Data: Marginal Posterior for μ

- To get the marginal posterior for μ, integrate out σ:
  $p(\mu \mid \mathbf{x}) = \int_0^\infty p(\mu, \sigma \mid \mathbf{x})\, d\sigma$
- Letting $u^2 = \sigma^2$, $b = n$, and $a = \frac{1}{2}\left[(n-1)s^2 + n(\mu - \bar{x})^2\right]$, we get

$$p(\mu \mid \mathbf{x}) = \int_0^\infty u^{-(b+1)} e^{-\frac{a}{u^2}}\, du \propto \frac{1}{2} a^{-\frac{b}{2}} \Gamma\left(\frac{b}{2}\right)$$

- So, the marginal posterior for μ is:

$$p(\mu \mid \mathbf{x}) \propto \frac{1}{2}\left(\frac{1}{2}\left[(n-1)s^2 + n(\mu - \bar{x})^2\right]\right)^{-\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)$$

$$= \frac{1}{2\left[(n-1)s^2\right]^{n/2}} \left(1 + \frac{n(\mu - \bar{x})^2}{(n-1)s^2}\right)^{-\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)$$

$$\propto \left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right)\left(\frac{n/s^2}{(n-1)}\right)^{-1/2}\left(1 + \frac{1}{n-1}\left(\frac{\mu - \bar{x}}{s/\sqrt{n}}\right)^2\right)^{-\frac{n}{2}}$$

Vague Priors with Normal Data: t-Distribution Transformation

- Make the transformation: $t = \frac{\mu - \bar{x}}{s/\sqrt{n}}$ with the Jacobian: $|J| = \frac{s}{\sqrt{n}}$
- The posterior becomes:

$$p(t|\mathbf{x}) = \frac{\Gamma\left(\frac{n-1+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{[(n-1)\pi]^{1/2}} \left(1 + \frac{t^2}{n-1}\right)^{-\left(\frac{n-1+1}{2}\right)}$$

- This is clearly a t-distribution with $n-1$ degrees of freedom, t_{n-1}.

Vague Priors with Normal Data: Marginal Distribution of σ^2

- To get the marginal distribution of σ^2, note that:

$$\begin{aligned} p(\sigma \mid \mathbf{x}) &= \int_{-\infty}^{\infty} p(\mu, \sigma \mid \mathbf{x})\, d\mu \\ &\propto \sigma^{-(n+1)} e^{\left(-\frac{1}{2\sigma^2}(n-1)s^2\right)} \int_{-\infty}^{\infty} e^{\left(-\frac{1}{2\sigma^2} n(\mu-\bar{x})^2\right)} d\mu \\ &= \sigma^{-(n+1)} e^{\left(-\frac{1}{2\sigma^2}(n-1)s^2\right)} \sqrt{\frac{2\pi\sigma^2}{n}} \end{aligned}$$

- Including the Jacobian of the transformation from σ to σ^2, we get:

$$\begin{aligned} p(\sigma^2 \mid \mathbf{x}) &\propto (\sigma^2)^{-\frac{n+1}{2}} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \sigma \left|\frac{1}{2\sigma}\right| \\ &\propto (\sigma^2)^{-\left(\frac{n-1}{2}+1\right)} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \end{aligned}$$

- Thus, $\sigma^2|\mathbf{x} \sim \text{Inverse Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$.

Vague Priors with Normal Data: Summary

- Both posterior distributions (for μ and σ^2) are proper.
- Compared to posteriors from conjugate analyses, these posteriors are more diffuse (spread out).
- The increased diffuseness is due to the vague prior information used.
- For large sample sizes, there is little difference between results from conjugate analysis and "uninformative" analysis.
- **Example 1(a):** Midge data revisited.

Outline

Vague Priors for Normal Data

Bayesian Model for MVN Data

Bayesian Model for Multivariate Normal Distribution

- **Setup**: Each individual has q variables observed, forming q-dimensional random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$.
- **Assumption**: These vectors are i.i.d. multivariate normal:

$$\mathbf{X}_i \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad \text{both } \boldsymbol{\mu}, \boldsymbol{\Sigma} \text{ are unknown}$$

  where:
  - $\boldsymbol{\mu}$ is the q-dimensional mean vector,
  - $\boldsymbol{\Sigma}$ is the $q \times q$ variance-covariance matrix.
- **Prior Distributions** for the parameters:

$$\boldsymbol{\mu} \mid \boldsymbol{\Sigma} \sim \text{MVN}\left(\boldsymbol{\delta}, \frac{1}{n_0}\boldsymbol{\Sigma}\right) \text{ and } \boldsymbol{\Sigma}^{-1} \sim \text{Wishart}(\nu_0, \boldsymbol{S}_0)$$

  where:
  - $\boldsymbol{\delta}$ is the prior mean for $\boldsymbol{\mu}$,
  - n_0 is a scalar reflecting the strength of prior information,
  - ν_0 and $\boldsymbol{S}_0$ are the degrees of freedom and scale matrix of the Wishart prior.

Posterior Distributions in Bayesian Multivariate Normal Model

- **Posterior for the Mean Vector**:

$$\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \mathbf{X} \sim \text{MVN}\left(\frac{n_0\delta + n\bar{x}}{n_0 + n}, \frac{1}{n_0 + n}\boldsymbol{\Sigma}\right)$$

  - The posterior mean is a weighted average of the prior mean (δ) and the sample mean ($\bar{x}$).
  - The posterior covariance decreases with increasing sample size (n), reducing uncertainty about $\boldsymbol{\mu}$.
- **Posterior for the Covariance Matrix**:

$$\boldsymbol{\Sigma}^{-1} \mid \mathbf{X} \sim \text{Wishart}(\nu_0 + n, \boldsymbol{S}_0 + \boldsymbol{S}_x)$$

  where $\boldsymbol{S}_x$ being the sample covariance.
- **Model Flexibility**:
  - The model incorporates prior beliefs about both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
  - As the sample size increases, the posterior distribution relies more heavily on the observed data.