Example: $X_1, \cdots, X_n \sim \mathcal{N}(\mu, 1)$, i.i.d., with variance known.

1. $f_{X|\pi}(\mathbf{x}|\theta) = \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}\right) (2\pi)^{-n/2}$

2. Use as prior $\varpi_\mu(\mu) = \exp\left(-\frac{(\mu - 0)^2}{2\tau^2}\right) \tau^{-1/2} (2\pi)^{-1/2}$

3. Use Bayes theorem to get posterior.

a. $f_{\mu|X}(\mu|\mathbf{x})$

\[
= \frac{f_{X|\mu}(\mathbf{x}|\mu) \varpi_\mu(\mu)}{\int_{-\infty}^{\infty} f_{X|\mu}(\mathbf{x}|\mu) \varpi_\mu(\mu) \, d\mu}
\]

\[
= \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right) (2\pi)^{-n/2} \exp\left(-\frac{\mu^2}{2\tau^2}\right) \tau^{-1/2} (2\pi)^{-1/2}}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right) (2\pi)^{-n/2} \exp\left(-\frac{\mu^2}{2\tau^2}\right) \tau^{-1/2} \, d\mu}
\]

\[
= \frac{\exp\left(-\mu^2/(2\tau^2) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right) (2\pi)^{-n/2} \exp\left(-\mu^2/(2\tau^2)\right) \tau^{-1/2} \, d\mu}
\]

\[
= \frac{\exp\left(-\frac{1}{2}\left\{\mu^2[\tau^{-2} + n] - 2n\bar{x}\mu + \sum_{i=1}^{n} x_i^2\right\}\right)}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left\{\mu^2[\tau^{-2} + n] - 2n\bar{x}\mu + \sum_{i=1}^{n} x_i^2\right\}\right) \, d\mu}
\]

\[
= \frac{\exp\left(-\frac{1}{2}\left\{\mu^2[\tau^{-2} + n] - 2n\bar{x}\mu\right\}\right)}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left\{\mu^2[\tau^{-2} + n] - 2n\bar{x}\mu\right\}\right) \, d\mu}
\]

\[
= \frac{\exp\left(-\frac{1}{2}\left[\tau^{-2} + n\right]\left\{\mu^2 - 2n\bar{x}[\tau^{-2} + n]^{-1}\mu\right\}\right)}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left[\tau^{-2} + n\right]\left\{\mu^2 - 2n\bar{x}[\tau^{-2} + n]^{-1}\mu\right\}\right) \, d\mu}
\]
\[ \frac{\exp\left(-\frac{1}{2}[\tau^{-2} + n]\{\mu^2 - 2m\mu + m^2\}\right)}{\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}[\tau^{-2} + n]\{\mu^2 - 2m\mu + m^2\}\right) d\mu} \]

\[ = \exp\left(-\frac{1}{2}\omega^{-2}(\mu - m)^2\right)(2\pi)^{-1/2}\omega^{-1} \]

\[ = \exp\left(-\frac{1}{2}\omega^{-2}(\mu - m)^2\right)(2\pi)^{-1/2}\omega^{-1} \]

for \( m = n\bar{x}[\tau^{-2} + n]^{-1}, \omega = [\tau^{-2} + n]^{-1/2}. \)

4. Hence \( \mu|X_1, \cdots, X_n \sim N(\bar{x}/(1/\tau^2 + 1/n), 1/(1/\tau^2 + 1/n)) \)

5. Hence

a. posterior expectation is a compromise between the sample average and the prior expectation

b. posterior variance

   i. smaller than the frequentist sampling variance of \( \bar{x} \)

   ii. Combines on the inverse scale.

H. Difficulties with Bayesian analyses:

1. Computation of denominator:

   a. The two examples above illustrate \textit{conjugate priors}:

      i. Families of distributions constructed so that the likelihood times the prior was in the same family as the prior

      ii. Integral done by integrating the resulting other member of the
family.

iii. Other examples are: $P(\lambda), \Gamma(\theta, k), \text{NBin}(\pi, l), B(\alpha, \beta)$.

b. Mixtures of conjugate priors: If $\varpi_{1, \theta}(\theta)$ and $\varpi_{2, \theta}(\theta)$ are such that we know $\int_{\Theta} L(\theta) \varpi_{i, \theta}(\theta) \, d\theta$ is known to be $I_i$ for all $i$, and if $\lambda \in [0, 1]$, then

$$\int_{\Theta} (\lambda \varpi_{1, \theta}(\theta) + (1 - \lambda) \varpi_{2, \theta}(\theta)) L(\theta) \, d\theta = \lambda I_1 + (1 - \lambda) I_2.$$ 

c. Numerical Integration:

i. Evaluate $L(\theta) \varpi_{\theta}(\theta)$ on a grid of points, separated by $\Delta$: $\zeta_j = L(\theta_0 + (j - 1)\Delta) \varpi_{\theta}(\theta_0 + (j - 1)\Delta)$ for $j = \{1, \ldots, m\}$.

ii. Approximation to integral is a linear combination of these evaluations

- Trapezoidal rule: $\int_{\Theta} L(\theta) \varpi_{i, \theta}(\theta) \, d\theta \approx \Delta (\omega_1 + 2 \sum_{j=1}^{m-1} \omega_j + \omega_m) / 2$.

- Simpson’s rule: if $m$ odd, $\int_{\Theta} L(\theta) \varpi_{i, \theta}(\theta) \, d\theta \approx \Delta (\omega_1 + 4 \sum_{j=1, j\text{odd}}^{m-1} \omega_j + 2 \sum_{j=1, j\text{even}}^{m-1} \omega_j + \omega_m) / 3$.

- See Fig. 19/.

d. Laplace’s method

i. We want $\int_{A} \exp(\ell(\theta)) \varpi(\theta) \, d\theta$
ii. Let \( \hat{\theta} \) be the MLE.

iii. Do Taylor series approximation for log likelihood and prior separately:
\[ \int_{A} \exp(\ell(\theta)) \varpi(\theta) \, d\theta \]
\[ \approx \int_{A} \exp(\ell + \frac{\hat{\ell}''(\theta - \hat{\theta})^2}{2} + \frac{\hat{\ell}'''(\theta - \hat{\theta})^3}{6} + \frac{\hat{\ell}''''(\theta - \hat{\theta})^4}{24}) \times \\
(\varpi(\hat{\theta}) + \varpi'(\hat{\theta})(\theta - \hat{\theta}) + \varpi''(\hat{\theta})(\theta - \hat{\theta})^2/2) \, d\theta \]

iv. \( \ell \) with accents represents derivative of \( \ell \) evaluated at \( \hat{\theta} \)

v. Uses \( \hat{\ell} = 0 \).

vi. Often \( \ell'' \approx n\hat{\ell}'' \), \( \ell''' \approx n\hat{\ell}''' \), \( \ell'''' \approx n\hat{\ell}'''' \).

vii. Hence:
\[ \int_{A} \exp(\ell(\theta)) \varpi(\theta) \, d\theta \]
\[ \approx \int_{A} \exp(\frac{\hat{\ell} + n\hat{\ell}''(\theta - \hat{\theta})^2}{2} + \frac{n\hat{\ell}'''(\theta - \hat{\theta})^3}{6} + \frac{n\hat{\ell}''''(\theta - \hat{\theta})^4}{24}) \times \\
(\varpi(\hat{\theta}) + \varpi'(\hat{\theta})(\theta - \hat{\theta}) + \varpi''(\hat{\theta})(\theta - \hat{\theta})^2/2) \, d\theta \]
\[ \approx \sqrt{-n\hat{\ell}} \exp(\hat{\ell}) \int_{A^*} \exp(\frac{-\delta^2}{2} + \frac{\kappa_3\delta^3}{6\sqrt{n}} + \frac{\kappa_4\delta^4}{24n}) \times \\
(\varpi(\hat{\theta}) + \varpi'(\hat{\theta})\delta/\sqrt{-n\hat{\ell}''} + \varpi''(\hat{\theta})/(-n\hat{\ell}''')\delta^2/2) \, d\delta \]
\[ \approx \sqrt{-\hat{\ell}''} \exp(\hat{\ell}) \int_{A^*} \exp(\frac{-\delta^2}{2})(1 + \frac{\kappa_3\delta^3}{6\sqrt{n}} + \frac{\kappa_4\delta^4}{24n} + \frac{\kappa_3\delta^6}{72n}) \times \\
(\varpi(\hat{\theta}) + \varpi'(\hat{\theta})\delta/\sqrt{-n\hat{\ell}''} + \varpi''(\hat{\theta})/(-n\hat{\ell}''')\delta^2/2) \, d\delta \]

viii. \[ \int_{-\infty}^{\infty} \exp(\ell(\theta)) \varpi(\theta) \, d\theta \approx \sqrt{-2\pi\hat{\ell}''} \exp(\hat{\ell}) \varpi(\hat{\theta}) \]

ix. \( \delta = \sqrt{-n\hat{\ell}''} (\theta - \hat{\theta}) \)

x. \( \kappa_3 = \hat{\ell}'''(-\hat{\ell}'')^{-3/2} \), \( \kappa_4 = \hat{\ell}''''(-\hat{\ell}'')^{2} \).
xi. Depends on prior only through value at $\hat{\theta}$, and not on shape, to $O(1/n)$.

xii. Extends to higher-order approximations.

2. Integration scales poorly as dimension of $\theta$ increases
   a. Deterministic integration is replaced by simulation.

I. Bayesian Inference

1. Estimation:
   a. Generally use expectation of posterior distribution.
      i. Minimizes expected squared error loss, analogously to material from lecture 1.
      ii. Ex. $X_1, \ldots, X_n \sim N(\mu, 1)$, $\mu \sim N(0, \tau^2)$: Estimator $\bar{x} / (1/\tau^2 + 1/n)$,
          - shrunk towards prior expectation 0
          - More shrinking if $\tau$ smaller if and only if more confidence in prior.
      iii. Ex. $X \sim \text{Bin}(\pi, n)$, $\pi \sim B(\alpha, \beta)$.
          - $E[\pi | X] = \frac{x + \alpha}{n + \alpha + \beta}$
          - Again, moves toward $x/n$ as prior information becomes less certain, and $\alpha$ and $\beta$ to to 0.
b. Can also use posterior median or mode.

i. Normal example: all coincide.

ii. Binomial:

• median involves inverting incomplete $\beta$ function.

• mode $\frac{\alpha + x - 1}{\alpha + \beta + n - 2}$