Hoff Ch. 5 : Normal model

- $\mathbb{R}$-valued random variable Y is <u>normally distributed</u> with mean θ and variance $\sigma^2 > 0$ if

$$p(y \mid \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\theta}{\sigma}\right)^2}.$$

I'll denote this distribution $N(\theta, \sigma^2)$

- Properties.
  - $p(y|\theta,\sigma^2)$ symmetric about θ
  - mean = median = mode = θ
  - $\approx 95\%$ of population within 2 stdevs of θ
  - $X \sim N(\mu,\tau^2)$ and $Y \sim N(\theta,\sigma^2)$, then

X and Y independent $\Longrightarrow aX + bY \sim N(a\mu + b\theta,\ a^2\tau^2 + b^2\sigma^2)$

* Caution: Be careful about whether statistical software uses σ or σ^2 (stdev or variance) when specifying N.

Bayesian model for mean (fixed variance)

- $Y_1, \dots, Y_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$

$$\Rightarrow p(\vec{y}_n \mid \theta, \sigma^2) = \prod_{i=1}^{n} p(y_i \mid \theta, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i - \theta}{\sigma}\right)^2}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{y_i - \theta}{\sigma}\right)^2}$$

From the orange term above:

$$\sum_{i=1}^{n}\left(\frac{y_i-\theta}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_i y_i^2 - 2\theta y_i + \theta^2$$

$$= \frac{1}{\sigma^2}\sum_i y_i^2 - \frac{2\theta}{\sigma^2}\sum_i y_i + \frac{n\theta^2}{\sigma^2}$$

$\Rightarrow \left\{\sum_i^n y_i, \sum_i^n y_i^2\right\}$ a 2-dimensional sufficient statistic

$\Rightarrow \left\{\underbrace{\frac{1}{n}\sum_i^n y}_{\text{sample mean}}, \underbrace{\frac{1}{n-1}\sum_i^n (y_i-\bar{y})^2}_{\text{sample variance}}\right\}$ also sufficient statistic.

Conjugacy (fixed variance)

Gaussian prior for θ + Gaussian data

$\implies$ Gaussian posterior (for θ).

Let's check.

- $Y_1, \ldots, Y_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$
- $p(\theta \mid \sigma^2) \sim N(\mu_0, \tau_0^2)$

$$\Rightarrow p(\theta \mid \sigma^2, \vec{y}_n) = \frac{p(\vec{y}_n \mid \theta, \sigma^2)\, p(\theta \mid \sigma^2)}{p(\vec{y}_n \mid \sigma^2)}$$

$$\propto p(\vec{y}_n \mid \theta, \sigma^2)\, p(\theta \mid \sigma^2)$$

$$\propto \exp\left(-\frac{1}{2}\sum_i \left(\frac{y_i - \theta}{\sigma}\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{\theta - \mu_0}{\tau_0}\right)^2\right)$$

$$= \exp\left(-\frac{1}{2}\left[\left(\frac{\theta - \mu_0}{\tau_0}\right)^2 + \frac{1}{\sigma^2}\sum_i (y_i - \theta)^2\right]\right)$$

$$= \exp\left(-\frac{1}{2}\left[\frac{1}{\tau_0^2}\left(\theta^2 - 2\mu_0\theta + \mu_0^2\right) + \frac{n}{\sigma^2}\theta^2 - \frac{2}{\sigma^2}\sum_i y_i\theta + \frac{1}{\sigma^2}\sum y_i^2\right]\right)$$

$$= \exp\left(-\frac{1}{2}\left[\underbrace{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}_{a}\theta^2 - 2\underbrace{\left(\frac{\mu_0}{\tau_0^2}+\frac{1}{\sigma^2}\sum y_i\right)}_{b}\theta + \underbrace{\frac{\mu_0^2}{\tau_0^2}+\frac{1}{\sigma^2}\sum y_i^2}_{c}\right]\right)$$

$$= \exp\left(-\frac{1}{2}\left[a\theta^2 - 2b\theta + c\right]\right)$$

$$a\theta^2 - 2b\theta = a\left(\theta^2 - \frac{2b}{a}\theta\right)$$

$$= a\left(\theta - \frac{b}{a}\right)^2 - \frac{b^2}{a}$$

$$= \left[\sqrt{a}\left(\theta - \frac{b}{a}\right)\right]^2 - \frac{b^2}{a}$$

$$= \left(\frac{\theta - \frac{b}{a}}{\frac{1}{\sqrt{a}}}\right)^2 - \frac{b^2}{a}$$

$$= \exp\left(-\frac{1}{2}\left[\left(\frac{\theta - \frac{b}{a}}{\frac{1}{\sqrt{a}}}\right)^2 - \frac{b^2}{a} + c\right]\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(\frac{\theta - \frac{b}{a}}{\frac{1}{\sqrt{a}}}\right)^2\right)$$

$$\sim N\left(\frac{b}{a}, \frac{1}{a}\right)$$

where $a = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$ and $b = \frac{\mu_0}{\tau_0^2} + \frac{1}{\sigma^2}\sum_i y_i$.

- Posterior mean:

$$\mu_n = \mathbb{E}[\theta \mid \sigma^2, \vec{y}_n] = \frac{b}{a} = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

- Posterior variance:

$$\tau_n^2 = \mathrm{Var}[\theta \mid \sigma^2, \vec{y}_n] = \frac{1}{a} = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}.$$

Posterior precision.

- <u>Precision</u> is the inverse of variance: $\frac{1}{\text{Var}}$
- Precision = "Amount of information"
- The posterior precision of the above normal model is

$$\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2} n$$

prior precision

data/sampling precision

Posterior mean

$$\mu_n = \frac{1}{1 + \frac{n\tau_0^2}{\sigma^2}} \mu_0 + \frac{1}{1 + \frac{\sigma^2}{n\tau_0^2}} \bar{Y}$$

$$= \underbrace{\frac{\sigma^2}{\sigma^2 + n\tau_0^2}}_{} \mu_0 + \frac{n\tau_0^2}{n\tau_0^2 + \sigma^2} \bar{Y}$$

prior mean

sample mean

Posterior prediction

- $Y_1, \ldots, Y_{n+1} \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$.
- Prior $p(\theta \mid \sigma^2) = N(\mu_0, \tau_0^2)$
- Posterior $p(\theta \mid \sigma^2, \vec{y}_n) = N(\mu_n, \sigma_n^2)$.
- Note that $Y_{n+1} = \theta + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2)$.
- Goal: Find $p(y_{n+1} \mid \sigma^2, \vec{y}_n)$.

- $\mathbb{E}[Y_{n+1} \mid \sigma^2, \vec{y}_n] = \mathbb{E}[\theta \mid \sigma^2, \vec{y}_n] + \mathbb{E}[\varepsilon \mid \sigma^2, \vec{y}_n]$

$$= \mu_n$$

- $\mathrm{Var}[Y_{n+1} \mid \sigma^2, \vec{y}_n] = \mathrm{Var}[\theta \mid \sigma^2, \vec{y}_n] + \mathrm{Var}[\varepsilon \mid \sigma^2, \vec{y}_n]$

$$= \tau_n^2 + \sigma^2.$$

$$\Rightarrow Y_{n+1} \mid \sigma^2, \vec{y}_n \sim N(\mu_n, \tau_n^2 + \sigma^2).$$

Interpretation: Uncertainty about Y_{n+1}

$=$

uncertainty about the population mean (τ_n^2)

$+$

true population variance (σ^2)

As $n \to \infty$, $\mu_n \to$ population mean and

$$\tau_n^2 \longrightarrow 0 \implies \tau_n^2 + \sigma^2 \longrightarrow \sigma^2.$$

Last lecture: Normal w/ fixed variance

- $Y_1, \ldots, Y_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$

- $\theta \mid \sigma^2 \sim N(\mu_0, \tau_0^2)$

- $\theta \mid \sigma^2, \vec{y}_n \sim N(\mu_n, \sigma_n^2)$

    - $\mu_n = \dfrac{\sigma^2}{\sigma^2 + n\tau_0^2} \mu_0 + \dfrac{n\tau_0^2}{\sigma^2 + n\tau_0^2} \overline{Y}$

    - $\dfrac{1}{\tau_n^2} = \dfrac{1}{\tau_0^2} + \dfrac{n}{\sigma^2}$

- (PPD) $Y_{n+1} \mid \sigma^2, \vec{y}_n \sim N(\mu_n, \tau_n^2 + \sigma^2)$

Example: Fly wing length

- $n = 9$ samples
- Sample mean $\bar{y} = 1.804$ mm
- Sample variance $s^2 = 0.017$
- Assume $Y_1, \ldots, Y_{10} \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$.
- Prior: $p(\theta \mid \sigma^2) \sim N(\mu_0, \tau_0^2)$ where
  - $\mu_0 = 1.9$ (taken from previous studies)
  - $\tau_0 = 0.95$ (weak prior; 2 stdevs below $= 0$)

- Posterior $\theta \mid \sigma^2, \vec{y}_9$ has mean and variance

$$\mu_n = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} = \frac{1.11(1.9) + \left(\frac{9}{\sigma^2}\right)(1.804)}{1.11 + \frac{9}{\sigma^2}}.$$

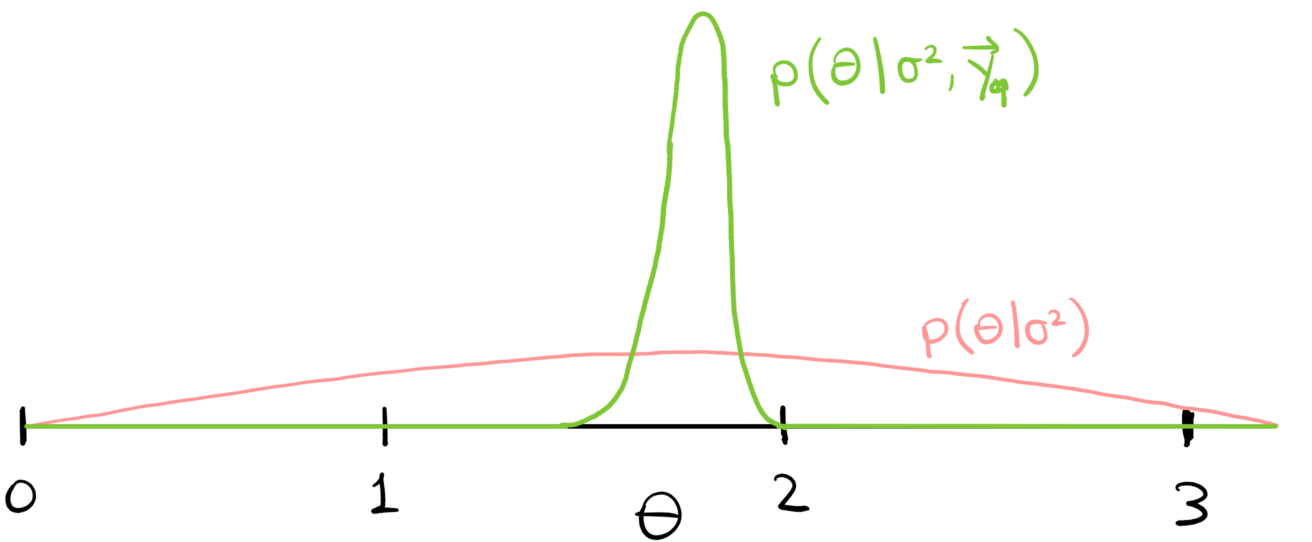
$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} = \frac{1}{1.11 + \frac{9}{\sigma^2}}.$$

If $\sigma^2 = s^2$, then

$$\theta \mid \sigma^2, \vec{y}_n \sim N(1.805, 0.002)$$

95% quantile-based confidence interval for θ is

- $l(\vec{y}_q) \approx \mu_n - 2\tau_n = 1.805 - 2\sqrt{0.002} \approx 1.72$
- $u(\vec{y}_q) \approx \mu_n + 2\tau_n = 1.805 + 2\sqrt{0.002} \approx 1.89$



Joint inference for mean and variance

As before, assume $Y_1, \ldots, Y_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$.

$$p(\theta, \sigma^2 \mid \vec{y}_n) = \frac{p(\vec{y}_n \mid \theta, \sigma^2)\, p(\theta, \sigma^2)}{p(\vec{y}_n)}$$

$$\propto p(\vec{y}_n \mid \theta, \sigma^2)\, p(\theta, \sigma^2).$$

where $p(\theta, \sigma^2) = p(\theta \mid \sigma^2)\, p(\sigma^2).$

Choosing priors

- Priors $p(\theta|\sigma^2)$ and $p(\sigma^2)$
- Saw that $p(\theta|\sigma^2) = N(\mu_0, \tau_0^2)$ conjugate when σ^2 is known/fixed.
- For σ^2, need family of distributions on $(0, \infty)$.
  - Turns out Gamma is not conjugate for σ^2
  - However, Gamma <u>is</u> conjugate for $\frac{1}{\sigma^2}$ (precision)

Inverse gamma distribution

$\frac{1}{\sigma^2} \sim \text{Gamma}(a,b) \implies \sigma^2 \sim IG(a,b).$

Set $a = \frac{\nu_0}{2}$, $b = \frac{\nu_0}{2}\sigma_0^2 = a\sigma_0^2$. Then

- $\mathbb{E}[\sigma^2] = \dfrac{\frac{\nu_0}{2}}{\frac{\nu_0}{2} - 1}\sigma_0^2$ (note this requires $\nu_0 > 2$).
- $\text{Mode}[\sigma^2] = \dfrac{\frac{\nu_0}{2}}{\frac{\nu_0}{2} + 1}\sigma_0^2 \leq \sigma_0^2 \leq \mathbb{E}[\sigma^2]$
- $\text{Var}[\sigma^2]$ decreases in ν_0.

Posterior inference.

- Priors:
  - $\frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2\right)$
  - $\theta \mid \sigma^2 \sim N\left(\mu_0, \frac{1}{\kappa_0}\sigma^2\right)$

- Likelihood:

$$Y_1, \ldots, Y_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2).$$

- Posterior :

$$p(\theta, \sigma^2 \mid \vec{y}_n) = p(\theta \mid \sigma^2, \vec{y}_n)\, p(\sigma^2 \mid \vec{y}_n)$$

Let's look at right-hand terms separately :

- $p(\theta \mid \sigma^2, \vec{y}_n) = N(\mu_n, \tau_n^2)$ where

* Define $\tau_0^2 = \frac{\sigma^2}{\kappa_0}$

$$\mu_n = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} = \frac{\frac{\kappa_0 \mu_0}{\sigma^2} + \frac{n}{\sigma^2}\bar{y}}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}$$

$$= \frac{\kappa_0 \mu_0 + n\bar{y}}{\kappa_0 + n}$$

$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} = \frac{\sigma^2}{\kappa_0 + n}.$$

Setting $\kappa_n = \kappa_0 + n$ gives

$$\boxed{p(\theta \mid \sigma^2, \vec{y}_n) = N\left(\mu_n, \frac{1}{\kappa_n}\sigma^2\right) = N\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_n}, \frac{\sigma^2}{\kappa_n}\right)}$$

$$\mathbb{E}[\theta \mid \sigma^2, \vec{y}_n] = \left(\frac{\kappa_0}{\kappa_n}\right)\mu_0 + \left(\frac{n}{\kappa_n}\right)\bar{Y}$$

prior mean

sample mean

$$\mathrm{Var}[\theta \mid \sigma^2, \vec{y}_n] = \frac{\sigma^2}{\kappa_0 + n}$$

"prior sample size" sample size

- $p(\sigma^2 \mid \vec{y}_n) \propto p(\vec{y}_n \mid \sigma^2)\, p(\sigma^2)$

$$= p(\sigma^2) \int_{-\infty}^{\infty} p(\vec{y}_n \mid \theta, \sigma^2)\, p(\theta \mid \sigma^2)\, d\theta.$$

... tedious calculations ...

$$\Rightarrow \boxed{\frac{1}{\sigma^2} \,\Big|\, \vec{y}_n \sim \text{Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n}{2}\sigma_n^2\right) \quad \text{where}}$$

"prior sample size"

- $\nu_n = \nu_0 + n$

- $\sigma_n^2 = \dfrac{1}{\nu_n}\left[\nu_0\sigma_0^2 + (n-1)s^2 + \dfrac{\kappa_0 n}{\kappa_n}(\bar{y}-\mu_0)^2\right]$

"prior sample variance"

sample variance

Note: $(n-1)s^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2$ (sum of squares)

In summary:

- Priors $\theta \mid \sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right)$ and $\sigma^2 \sim IG\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2\right)$

- Likelihood $Y_1, \ldots, Y_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$

- Posterior

$$p(\theta, \sigma^2 \mid \vec{y}_n) = p(\theta \mid \sigma^2, \vec{y}_n)\, p(\sigma^2 \mid \vec{y}_n)$$

$$= N\left(\mu_n, \frac{\sigma^2}{\kappa_n}\right) IG\left(\frac{\nu_n}{2}, \frac{\nu_n}{2}\sigma_n^2\right)$$

- $\mu_n = \dfrac{\kappa_0 \mu_0 + n\bar{y}}{\kappa_0 + n}$

- $\kappa_n = \kappa_0 + n$

- $\nu_n = \nu_0 + n$

- $\sigma_n^2 = \dfrac{1}{\nu_n}\left[\nu_0 \sigma_0^2 + (n-1)s^2 + \dfrac{\kappa_0 n}{\kappa_n}(\bar{y} - \mu_0)^2\right]$

Interpretation of prior parameters

- μ_0 = prior guess for θ
- σ_0^2 = prior guess for σ^2
- κ_0 = prior sample size for θ
- ν_0 = prior sample size for σ^2

Example: Fly wing length continued

- $n = 9$ samples; $\bar{y} = 1.804$; $s^2 = 0.0169$

- Based on previous studies set $\mu_0 = 1.9$, $\sigma_0^2 = 0.01$

- "Weakly" informative priors: $\kappa_0 = \nu_0 = 1$.

$$\Rightarrow \quad \mu_9 = \frac{1.9 + 9(1.804)}{1 + 9} = 1.814$$

$$\sigma_9^2 = \frac{1}{10}\left[\frac{1}{100} + 8(0.0169) + \frac{9}{10}(1.804 - 1.9)^2\right]$$

$$\approx 0.015$$

Monte Carlo sampling.

- Interest often in estimating things about θ, eg.

$$\mathbb{P}(\theta_2 > \theta_1 \mid y_{11}, \dots, y_{n1}, y_{12}, \dots, y_{n2}).$$

However, only know $p(\theta, \sigma^2 \mid \vec{y}_n)$. What we need is marginal samples of θ from $p(\theta \mid \vec{y}_n)$.

for $1 \le i \le S$:

$$\sigma_i^2 \overset{iid}{\sim} IG\left(\frac{\nu_n}{2}, \frac{\nu_n}{2}\sigma_n^2\right)$$

$$\theta_i \overset{iid}{\sim} N\left(\mu_n, \frac{1}{\kappa_n}\sigma_i^2\right)$$

Poor notation on my part: σ_n^2 refers to the posterior quantity we computed, while σ_i^2 is a MC sample.

Output: $\{(\sigma_1^2, \theta_1), \ldots, (\sigma_S^2, \theta_S)\}$.

These are samples from joint posterior $p(\theta, \sigma^2 | \vec{y}_n)$.

$\{\theta_1, \ldots, \theta_S\}$ are samples from <u>marginal posterior</u>

$$p(\theta | \vec{y}_n)$$

and can be used for MC approximations.

Bias, variance, and mean squared error (MSE)

- Let θ_* be a true (unknown) population parameter.

- An <u>estimator</u> of θ_* is a function $\hat{\theta}(Y)$ estimates θ_* from data Y

- An estimator $\hat{\theta}$ of θ_* is <u>unbiased</u> if

$$\mathbb{E}[\hat{\theta}(Y)] = \theta_* .$$

and is <u>biased</u> otherwise.

- Mean squared error (how far is $\hat{\theta}(Y)$ from θ_*?)

Dropping Y from notation and setting $m = \mathbb{E}[\hat{\theta}]$,

$$\mathbb{E}\left[(\hat{\theta} - \theta_*)^2\right] = \mathbb{E}\left[(\hat{\theta} - m + m - \theta_*)^2\right]$$

$$= \mathbb{E}\left[(\hat{\theta} - m)^2 + 2(\hat{\theta} - m)(m - \theta_*) + (m - \theta_*)^2\right]$$

$$= \mathrm{Var}[\hat{\theta}] + 2(m - \theta_*)\cancelto{0}{\mathbb{E}[\hat{\theta} - m]} + (m - \theta_*)^2$$

$$= \mathrm{Var}[\hat{\theta}] + \mathrm{Bias}(\hat{\theta})^2$$

$\mathrm{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta_*$

Interpretation:

Expected distance between $\hat{\theta}$ and θ

=

How spread out $\hat{\theta}$ is (its variance)

+

How far center of distribution of $\hat{\theta}$ is from θ_* (bias)

- <u>Example.</u> Sample mean is an unbiased estimator of the population mean. Specifically, assume $Y_1, \ldots, Y_n$ iid copies of a r.v. Y w/ mean θ_*, variance σ_*^2. Sample mean $\hat{\theta}(\vec{Y}_n) = \frac{1}{n} \sum_i Y_i$ satisfies

$$\mathbb{E}\left[\hat{\theta}(\vec{Y}_n)\right] = \frac{1}{n} \sum_i^n \mathbb{E}[Y_i] = \theta_* .$$

$$\mathrm{Var}\left[\hat{\theta}(\vec{Y}_n)\right] = \frac{1}{n^2} \sum_i^n \sigma_*^2 = \frac{\sigma_*^2}{n}$$

$$\mathrm{MSE}(\hat{\theta}) = \frac{\sigma_*^2}{n} + \left(\mathbb{E}(\hat{\theta}) - \theta_*\right) = \frac{\sigma_*^2}{n}.$$

- Now consider case of normal model
  - $Y_1, \ldots, Y_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$
  - $\theta \mid \sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{K_0}\right)$, $\sigma^2 \sim IG\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2\right)$.

Posterior mean is

$$\hat{\theta}_b(\vec{y}_n) = \mathbb{E}\left[\theta \mid \vec{y}_n\right] = \frac{K_0}{K_0+n}\mu_0 + \frac{n}{K_0+n}\bar{Y}$$

Taking $\hat{\theta}_b$ as our estimator:

$$\Rightarrow \mathbb{E}[\hat{\theta}_b] = \frac{K_0}{K_0+n}\mu_0 + \frac{n}{K_0+n}\theta_* \quad \text{(biased)}$$

$$\mathrm{Var}[\hat{\theta}_b] = \left(\frac{n}{K_0+n}\right)^2 \frac{\sigma_*^2}{n} < \frac{\sigma_*^2}{n}$$

$$\mathrm{MSE}(\hat{\theta}_b) = \left(\frac{n}{K_0+n}\right)^2 \frac{\sigma_*^2}{n} + \left(\mathbb{E}[\hat{\theta}_b] - \theta_*\right)^2$$

Can show that

$$\mathrm{MSE}(\hat{\theta}_b) < \mathrm{MSE}(\hat{\theta}) = \frac{\sigma_*^2}{n} \quad \text{whenever}$$

$$(\mu_0 - \theta_*)^2 < \left(\frac{1}{n} + \frac{2}{K_0}\right) \sigma^2$$