→ rearranging Normal pdf for later → if $Y \sim \text{Normal}(\mu, \sigma^2)$ ⇒ $f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-1}{2\sigma^2}(y-\mu)^2\right\}$

$$= \exp\left\{-\frac{1}{2\sigma^2}(y^2 - 2\mu y + \mu^2)\right\}$$

$$= \underbrace{\exp\left\{\frac{-\mu^2}{2\sigma^2}\right\}} \exp\left\{-\frac{1}{2\sigma^2}y^2 + \frac{\mu}{\sigma^2}y\right\}$$

↳ constant w/ respect to y ⇒ dropped w/ $\propto$

Single parameter inference

STA 427/527, Fall 2019, Xin Wang

2.5 Normal models

2.5.1 σ^2 is known μ is the parameter of interest

- **Data:** $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where σ^2 is known. Then the data distribution is $\underset{\sim}{y} = (y_1, \ldots, y_n)$

$$p(\underset{\sim}{y} | \mu) = p(y_1, \ldots, y_n | \mu) = \prod_{i=1}^{n} p(y_i|\mu) = \prod_{i=1}^{n} \cancel{\frac{1}{\sqrt{2\pi\sigma^2}}} \exp\left\{\frac{-1}{2\sigma^2}(y_i - \mu)^2\right\}$$

$$\propto \prod_{i=1}^{n} \exp\left\{\frac{-1}{2\sigma^2}(y_i-\mu)^2\right\}$$ [drop constants w/ respect to μ]

$$= \exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\cancel{y_i^2} - 2\mu y + \mu^2)\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\left(-2\mu \underbrace{\sum_{i=1}^{n} y_i}_{= n(\frac{1}{n}\sum y_i) = n\bar{y}} + n\mu^2\right)\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}(n\mu^2 - 2n\mu\bar{y})\right\}$$

- **Prior:** $\mu \sim N(\mu_0, \tau_0^2)$

$$\pi(\mu) = \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left\{-\frac{1}{2\tau_0^2}(\mu - \mu_0)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2\tau_0^2}(\mu-\mu_0)^2\right\}$$

$$= \exp\left\{-\frac{1}{2\tau_0^2}(\mu^2 - 2\mu_0\mu + \cancel{\mu_0^2})\right\}$$

$$\propto \exp\left\{-\frac{1}{2\tau_0^2}(\mu^2 - 2\mu_0\mu)\right\}$$

- **Posterior:**

→ $p(\mu | \underset{\sim}{y}) \propto p(\underset{\sim}{y}|\mu)\,\pi(\mu)$

$$= \exp\left\{-\frac{1}{2\sigma^2}(n\mu^2 - 2n\mu\bar{y})\right\}\left[\exp\left\{-\frac{1}{2\tau_0^2}(\mu^2 - 2\mu_0\mu)\right\}\right]$$

$$= \exp\left\{-\left(\underbrace{\frac{n}{2\sigma^2} + \frac{1}{2\tau_0^2}}_{=\frac{1}{2\tau_n^2}}\right)\mu^2 + \left(\underbrace{\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}}_{=\frac{\mu_n}{\tau_n^2}}\right)\mu\right\}$$

$$\boxed{\sim \text{Normal}(\mu_n, \tau_n^2)}$$, where $\left\{\begin{array}{l} \frac{1}{\tau_n^2} = \frac{n}{\sigma^2} + \frac{1}{\tau_0^2} \\ \frac{\mu_n}{\tau_n^2} = \frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2} \end{array}\right.$ ⇒ $\tau_n^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right)^{-1}$

$$\mu_n = \tau_n^2\left(\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right) = \left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right)^{-1}\left(\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)$$

(recall algebra from top: Normal $\propto \exp\left\{-\frac{1}{2\sigma^2}y^2 + \frac{\mu}{\sigma^2}y\right\}$)

→ Interpretation of parameters

→ Variance

→ $\tau_n^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right)^{-1} = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}$

↳ variance parameter

$\boxed{\frac{1}{\tau_n^2}} = \boxed{\frac{n}{\sigma^2}} + \boxed{\frac{1}{\tau_0^2}}$ → precision parameter in the prior

↳ precision parameter in the data dist

↳ precision parameter

→ In Frequentist

→ If $Y_1, Y_2, \dots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow \bar{Y} \sim N(\mu, \frac{\sigma^2}{n}) \Rightarrow$ precision parameter is $\frac{n}{\sigma^2}$

→ If $\tau_0^2 \to \infty$ → $\lim_{\tau_0^2 \to \infty} \tau_n^2 = \lim_{\tau_0^2 \to \infty} \left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right) = \frac{n}{\sigma^2}$

→ Mean

→ $\mu_n = \tau_n^2 \left(\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)$

$= \left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right)^{-1}\left(\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)$

$= \bar{y}\left(\frac{n/\sigma^2}{n/\sigma^2 + 1/\tau_0^2}\right) + \mu_0\left(\frac{1/\tau_0^2}{n/\sigma^2 + 1/\tau_0^2}\right)$

⇒ weighted average of sample mean & prior mean

⇒ So if we pick a ≈ kind of non informative prior (there is a difference between a non-informative & a prior just a large variance) ?

the weight on the sample data dominates the posterior

→ If $\tau_0^2 \to \infty$ $\lim_{\tau_0^2 \to \infty} \mu_n = \bar{y}$

→ Together → if $\tau_0^2 \to \infty \Rightarrow$ Posterior $\mu \mid \underline{y} \sim N(\bar{y}, \frac{\sigma^2}{n})$, which is same as frequentist $\bar{Y}$ dist

$\hat{\mu}_{MLE} = \bar{Y}$

- **Prediction:** $p(\tilde{y}|\boldsymbol{y}) = \int p(\tilde{y}|\mu)\, p(\mu|\boldsymbol{y})\, d\mu$ $\longrightarrow$ $\mu|\underset{\sim}{y} \sim N(\mu_n, \tau_n^2)$ from above

$\rightarrow$ $p(\mu|\underset{\sim}{y}) = \frac{1}{\sqrt{2\pi\tau_n^2}} \exp\left\{-\frac{1}{2\tau_n^2}(\mu-\mu_n)^2\right\}$

$p(\tilde{y}|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\tilde{y}-\mu)^2\right\}$

$\rightarrow$ from the definition $\rightarrow$ $p(\tilde{y}|\underset{\sim}{y}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\tau_n^2}} \exp\left\{-\frac{1}{2\tau_n^2}(\mu-\mu_n)^2\right\} * \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\tilde{y}-\mu)^2\right\} d\mu$

$=$ < ... properties of bivariate normal dist ... >

$\tilde{y}|\underset{\sim}{y} \sim N(\tilde{\mu}, \tilde{\tau}_n^2)$

$\rightarrow$ we know normal, now to find what the new $\tilde{\mu}$ + $\tilde{\tau}_n^2$ are

$\rightarrow$ $E(\tilde{y}|\underset{\sim}{y}) = E[E(\tilde{y}|\mu)|\underset{\sim}{y}] = \underbrace{E[\mu|\underset{\sim}{y}]}_{\text{posterior mean}} = \mu_n$

$\rightarrow$ $V(\tilde{y}|\underset{\sim}{y}) = E[V(\tilde{y}|\mu)|\underset{\sim}{y}] + V[E(\tilde{y}|\mu)|\underset{\sim}{y}]$

$= E[\sigma^2|\underset{\sim}{y}] + \underbrace{V[\mu|\underset{\sim}{y}]}_{\text{posterior variance}}$

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

Double expectation theorems

$\rightarrow$ $E(X) = E[E(X|Y)]$

$\rightarrow$ $V(X) = E[V(X|Y)] + V[E(X|Y)]$

$\rightarrow$ just now, everything is also conditional on data $\underset{\sim}{y}$

$\Longrightarrow$ $\boxed{\tilde{y}|\underset{\sim}{y} \sim N(\mu_n, \sigma^2 + \tau_n^2)}$

$\hookrightarrow$ Extra term b/c of prediction

2.5.2 Normal model with known mean

- **Data distribution:** $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where μ is known and σ^2 is the parameter of interest.

$p(\underset{\sim}{y}|\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-1}{2\sigma^2}(y_i-\mu)^2\right\}$

$\propto (\sigma^2)^{-n/2} \exp\left\{\frac{-1}{\sigma^2}\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}\right\}$

$\rightarrow$ In practice, we don't see this scenario often, but it is included here so we can see that inverse gamma is a conjugate prior for normal σ^2

- **Prior:** $\sigma^2 \sim \text{Inverse-gamma}(\alpha, \beta)$, that is $\pi(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} e^{-\frac{\beta}{\sigma^2}}$

$\rightarrow$ If $X \sim \text{Gamma}(\alpha, \beta) \Longrightarrow \frac{1}{X} \sim \text{Inv-gamma}(\alpha, \beta)$

[$\rightarrow$ note we can derive this density w/ gamma pdf + transformation $\frac{1}{x}$, not showing though]

- **Posterior:**

$\rightarrow$ $p(\sigma^2|\underset{\sim}{y}) \propto p(\underset{\sim}{y}|\sigma^2)\, \pi(\sigma^2)$

$\propto (\sigma^2)^{-n/2} \exp\left\{\frac{-1}{\sigma^2}\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}\right\} * \left[(\sigma^2)^{-(\alpha+1)} \exp\left\{-\beta/\sigma^2\right\}\right]$

$= \sigma^{2^{-((\frac{n}{2}+\alpha)+1)}} \exp\left\{-\frac{1}{\sigma^2}\left(\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2} + \beta\right)\right\}$

$\boxed{\sim \text{Inv-gamma}\left(\frac{n}{2} + \alpha,\ \frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2 + \beta\right)}$

$\hookrightarrow$ Conjugate $\Rightarrow$ prior + posterior both $\sim$ Inv-Gamma

→ From above example, $\sigma^2 | \underset{\sim}{y} \sim \text{Inv-Gamma}\left(\frac{n}{2} + \alpha, \frac{1}{2}\sum(y_i - \mu)^2 + \beta\right)$

↳ these are hyperparameters b/c they are not of interest but are needed to analyze the data

2.5.3 Hyperparameters

- We need to specify values of parameters of the prior distributions:

→ we need to specify values of the prior distribution parameters

→ there is no special way to pick these

⇒ can look in literature, or if we have knowledge of what the mean + variance of the prior should be, we can choose values of the hyperparameters to achieve that

→ e.g.) can figure out how to set α + β if we know $E(\sigma^2)$ + maybe even the ranges/spread

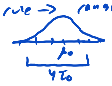
→ ex 1 If $X \sim \text{Inv-Gamma}(\alpha, \beta)$ → $E(X) = \frac{\beta}{\alpha - 1}$ $v(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$

$$\Rightarrow \alpha = 2 + \frac{(E(X))^2}{v(X)} \quad + \quad \beta = E(X)(\alpha - 1)$$

→ ex 2 In the normal model w/ σ^2 known + prior $\mu \sim N(\mu_0, \tau_0^2)$

→ we know the range of the data (b/c σ^2 is known + general values of a + b ├─┤ a b)
so how can we specify τ_0^2?

e.g.) Based on empirical rule → range $\approx 4\tau_0$ ⇒ $\tau_0 \approx \frac{\text{range}}{4}$ + range is (a, b) ⇒ $\tau_0 = \frac{b-a}{4}$



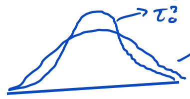
$a = \mu_0 - 2\tau_0$

$b = \mu_0 + 2\tau_0$

$\Rightarrow \mu_0 = \frac{a+b}{2}$

2.5.4 Noninformative prior for the normal model with known σ^2.

→ $\mu \sim N(\mu_0, \tau_0^2)$



→ If $\tau_0^2 \to \infty$, would it just be a "flat" distribution??

→ If flat prior or improper prior

$\mu \propto c$ but $c = ?$ doesn't exist for the integral below
$\downarrow \propto 1$

↳ $\int_{-\infty}^{\infty} c \, d\mu \neq 1$

⇒ consider Jeffrey's prior for μ / σ^2 is known

→ $p(\underset{\sim}{y} | \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{(y_i - \mu)^2}{2\sigma^2}\right\}$

$\downarrow \quad \propto \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right\}$

→ $\ell(\mu) = \ln[p(\underset{\sim}{y} | \mu)] = \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2$

→ $\frac{\partial \ell(\mu)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu) = \frac{\sum y_i - n\mu}{\sigma^2}$

→ $\frac{\partial^2 \ell(\mu)}{\partial \mu^2} = -\frac{n}{\sigma^2}$

→ $I(\mu) = -E\left[\frac{\partial^2 \ell(\mu)}{\partial \mu^2}\right]$

$= -E_y\left[-\frac{n}{\sigma^2}\right]$

$= n/\sigma^2$

⇒ $\mu \propto \sqrt{I(\mu)}$

$= \sqrt{n/\sigma^2}$

$\propto 1$ (improper prior)

→ open question → what is the Jeffrey's prior for normal models when σ^2 is unknown + μ is known?

→ $p(\underset{\sim}{y} | \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{(y_i - \mu)^2}{2\sigma^2}\right\}$

$\downarrow \quad \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{\sigma^2} \frac{\sum_{i=1}^{n} (y_i - \mu)^2}{2}\right\}$

→ $\ell(\sigma^2) = \ln[p(\underset{\sim}{y} | \sigma^2)] = -\frac{n}{2} \ln(\sigma^2) - \frac{1}{\sigma^2} \frac{\sum_{i=1}^{n} (y_i - \mu)^2}{2}$

→ $\frac{\partial \ell(\sigma^2)}{\partial \sigma^2} = \frac{-n/2}{\sigma^2} + \frac{1}{(\sigma^2)^2} \frac{\sum_{i=1}^{n} (y_i - \mu)^2}{2}$

→ $\frac{\partial^2 \ell(\sigma^2)}{\partial \sigma^2} = \frac{n/2}{(\sigma^2)^2} - \frac{\cancel{2}}{(\sigma^2)^3} \frac{\sum_{i=1}^{n} (y_i - \mu)^2}{\cancel{2}}$

$I(\sigma^2) = -E\left[\frac{\partial^2 \ell(\sigma^2)}{\partial \sigma^2}\right]$

$= -E_y\left[\frac{n/2}{(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^{n} (y_i - \mu)^2\right]$

$= \frac{-n}{2(\sigma^2)^2} + \frac{\sum_{i=1}^{n} E(y_i - \mu)^2}{(\sigma^2)^3}$

$= \frac{-n}{2(\sigma^2)^2} + \frac{n\, v(y_i)}{(\sigma^2)^3}$ → $= \sigma^2$

$= \frac{-n}{2(\sigma^2)^2} + \frac{n}{(\sigma^2)^2}$

$= \frac{n}{2(\sigma^2)^2}$

⇒ $\sigma^2 \propto \sqrt{I(\sigma^2)}$

$= \sqrt{\frac{n}{2(\sigma^2)^2}}$

$= \frac{1}{\sigma^2} \sqrt{\frac{n}{2}}$

$\propto \frac{1}{\sigma^2}$