→ In a multi-parameter setting, the general concept is the same, but the challenge of multi-parameters is how to simulate the posterior

→ Sometimes we can still derive a nice, closed form posterior w/ only 2 parameters, but probably not w/ 3+ parameters

→ Most of the time, we don't know the form of the posterior (i.e. it is not in a recognized kernel for a named dist). So, for inference, we use posterior samples rather than the exact theoretical result

→ In univariate → $p(\theta|\underset{\sim}{y}) = \frac{p(\underset{\sim}{y}|\theta)\,\pi(\theta)}{p(\underset{\sim}{y})}$

→ In multivariate → $p(\underset{\sim}{\theta}|\underset{\sim}{y}) = \frac{p(\underset{\sim}{y}|\underset{\sim}{\theta})\,\pi(\underset{\sim}{\theta})}{p(\underset{\sim}{y})}$

$\underset{\sim}{\theta}$ → vector

$\propto p(\underset{\sim}{y}|\underset{\sim}{\theta})\,\pi(\underset{\sim}{\theta})$ ①

$: p(\underset{\sim}{y}, \underset{\sim}{\theta})$ ② → use form 1 b/c that is the information we have

Multiparameter models

STA 427/527, Fall 2019, Xin Wang

vector

- In a multiparameter model, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$. The prior distribution is a joint prior for $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta})$ and the posterior is a joint posterior, $p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})$.
- Start from a model with two parameters, θ_1 and θ_2, where θ_1 is the parameter of interest and θ_2 is the nuisance parameter. Need to find the marginal posterior $p(\theta_1|\boldsymbol{y})$

$$\rightarrow p(\theta_1|\underset{\sim}{y}) = \int p(\theta_1, \theta_2|\underset{\sim}{y})\,d\theta_2$$
$$= \int p(\theta_1|\theta_2, \underset{\sim}{y})\, p(\theta_2|\underset{\sim}{y})\,d\theta_2$$

→ How can we draw θ_1 from the posterior dist??

→ Step 1) Draw $\theta_2^{(m)}$ from $p(\theta_2|\underset{\sim}{y})$, $m = 1, 2, \ldots, M$

→ Step 2) Draw $\theta_1^{(m)}$ from $p(\theta_1|\theta_2^{(m)}, \underset{\sim}{y})$, $m = 1, 2, \ldots, M$

⇒ $(\theta_1^{(m)}, \theta_2^{(m)})$ is from $p(\theta_1, \theta_2|\underset{\sim}{y})$

3.1 Normal models with unknown μ and σ^2

- μ: the parameter of interest σ^2: the nuisance parameter
- **Data distribution:** → $y_1, y_2, \ldots, y_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)^T$

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

- **Prior:** $\pi(\mu, \sigma^2) = \pi(\mu)\,\pi(\sigma^2)$ < b/c μ + σ^2 have independent priors > ↪ just assume independence of shape + scale parameters

→ A sensible vague prior density for μ + σ^2 is uniform on (μ, σ^2) or, equivalently

→ $p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$ → which comes from these pieces

→ $\pi(\mu) \propto 1$

→ $\pi(\sigma^2) \propto (\sigma^2)^{-1} \iff \pi(\log(\sigma^2)) \propto 1$

ex) → $f(x) \propto \frac{1}{x}$

→ let $y = g(x) = \ln(x) \Rightarrow x = h(y) = e^y \Rightarrow h'(y) = e^y$ < inverse >

$\Rightarrow f_Y(y) = f_X(h(y))\,|h'(y)|$
$= \frac{1}{e^y} e^y$
$= 1$

- **Posterior:**

$$\rightarrow p(\mu, \sigma^2|\underset{\sim}{y}) \propto p(\underset{\sim}{y}|\mu, \sigma^2)\,\pi(\mu, \sigma^2)$$
$$= \left[(\sigma^2)^{-n/2} \exp\left\{ \frac{-1}{2\sigma^2} \sum (y_i - \mu)^2 \right\} \right] \cdot \left[(\sigma^2)^{-1} \right]$$
$$= (\sigma^2)^{-n/2-1} \exp\left\{ \frac{-1}{2\sigma^2} \sum (y_i - \mu)^2 \right\}$$

→ Need to find $\underbrace{p(\sigma^2|\underset{\sim}{y})}_{\text{marginal posterior}}$ + $\underbrace{p(\mu|\sigma^2, \underset{\sim}{y})}_{\text{conditional posterior}}$ (b/c we need to know σ^2 before getting μ)

$p\left(\mu|\sigma^2, \boldsymbol{y}\right)$ and $p\left(\sigma^2|\boldsymbol{y}\right)$

$\rightarrow p(\mu, \sigma^2 | \underset{\sim}{y}) \propto (\sigma^2)^{-n/2-1} \exp\left\{ -\frac{1}{2\sigma^2} \sum (y_i - \mu)^2 \right\}$ < from above >

$\sum(y_i - \mu)^2 = \sum(\underbrace{y_i - \bar{y}}_{a} + \underbrace{\bar{y} - \mu}_{b})^2$ < add zero $= -\bar{y} + \bar{y}$ to get in the form of a binomial >

$= \sum\left[(y_i - \bar{y})^2 + \underbrace{2(y_i - \bar{y})(\bar{y} - \mu)}_{2ab} + (\bar{y} - \mu)^2 \right]$

$\hookrightarrow = 2\sum[y_i\bar{y} - y_i\mu - (\bar{y})^2 + \bar{y}\mu]$

$= 2[\bar{y}\underbrace{\sum y_i}_{= n\bar{y}} - \mu \underbrace{\sum y_i}_{= n\bar{y}} - n(\bar{y})^2 + n\bar{y}\mu]$

$= 0$

$= \sum\left[(y_i - \bar{y})^2 + (\bar{y} - \mu)^2 \right]$

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

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

$\rightarrow$ where $s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$

$\rightarrow$ ① $p(\mu | \sigma^2, \underset{\sim}{y}) = \dfrac{p(\mu, \sigma^2 | \underset{\sim}{y})}{p(\sigma^2 | \underset{\sim}{y})}$

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

$= (\sigma^2)^{-n/2-1} \exp\left\{ -\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2 \right] \right\}$ < from above >

$\hookrightarrow$ drop b/c no μ

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

$= \exp\left\{ -\dfrac{(\mu - \bar{y})^2}{2(\sigma^2/n)} \right\}$ < rearrange, $(a-b)^2 = (b-a)^2$ >

$\sim N(\bar{y}, \frac{\sigma^2}{n})$ < b/c if $x \sim N(\mu_0, \sigma_0^2) \Rightarrow f(x) = \frac{1}{2\sigma_0^2} \exp\left\{ -\dfrac{(x - \mu_0)^2}{2\sigma_0^2} \right\}$

$\rightarrow$ ② $p(\sigma^2 | \underset{\sim}{y}) = \int p(\mu, \sigma^2 | \underset{\sim}{y})\, d\mu$

$\propto \int (\sigma^2)^{-n/2-1} \exp\left\{ -\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2 \right] \right\}$

$= \underbrace{(\sigma^2)^{-n/2-1} \exp\left\{ -\frac{1}{2\sigma^2} (n-1)s^2 \right\}}_{\hookrightarrow \text{ constant w/ respect to } \mu} \int \underbrace{\exp\left\{ \dfrac{-(\bar{y} - \mu)^2}{2\sigma^2/n} \right\}}\, d\mu$

$\rightarrow \int f(x)\,dx = 1 \Rightarrow 1 = \int \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{\text{normalizing constant}} \underbrace{\exp\left\{ -\frac{1}{2\sigma^2}(y_i - \mu)^2 \right\}}_{\text{Kernel}}$ (normal pdf)

$\Rightarrow \sqrt{2\pi\sigma^2} = \int \exp\left\{ -\frac{1}{2\sigma^2}(y_i - \mu)^2 \right\}$

$= * \sqrt{2\pi \frac{\sigma^2}{n}}$

$\propto \left[(\sigma^2)^{-n/2-1} \exp\left\{ -\frac{1}{2\sigma^2} (n-1)s^2 \right\} \right] (\sigma^2)^{1/2}$

$= (\sigma^2)^{\underbrace{-\left(\frac{n-1}{2}\right)-1}_{\alpha}} \exp\left\{ -\frac{1}{(\sigma^2)} \underbrace{\frac{(n-1)}{2} s^2}_{\beta} \right\}$

< algebra $\rightarrow -\frac{n}{2} - 1 + 1/2 = -\frac{n}{2} - \frac{1}{2} = -\left(\frac{n-1}{2}\right) - 1$ > (for exponent)

$\sim$ Inv-Gamma $\left(\alpha = \frac{n-1}{2}, \beta = \frac{n-1}{2} s^2\right)$

< If $x \sim$ Inv Gamma $(\alpha, \beta) \Rightarrow f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\{ -\beta/x \}$

$\Rightarrow \sigma^2$ is our variable (like x)

$\rightarrow$ Now, draw samples from the joint posterior (but really the marginal posterior & then the conditional posterior b/c way easier)

$\rightarrow$ Step 0 $\rightarrow$ Calculate $\bar{y}$ & s^2

$\rightarrow$ Step 1 $\rightarrow$ Draw $(\sigma^2)^{(m)}$ from Inv-gamma $\left(\frac{n-1}{2}, \frac{(n-1)}{2} s^2\right)$ for $m = 1, 2, \dots, M$

$\rightarrow$ Step 2 $\rightarrow$ Draw $\mu^{(m)}$ from $N\left(\bar{y}, \frac{\sigma^{2(m)}}{n}\right)$

$\Rightarrow$ Thus we have a pair $(\mu^{(m)}, (\sigma^2)^{(m)})$

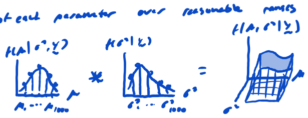
$\rightarrow$ note $\rightarrow$ to find density values of the joint posterior, just multiply

$\underbrace{p(\mu | \sigma^2, \underset{\sim}{y})} * \underbrace{p(\sigma^2 | \underset{\sim}{y})}$, which of course $= p(\mu, \sigma^2 | \underset{\sim}{y})$

$\hookrightarrow$ These turned out to have a known form $\Rightarrow$ so can just use dnorm() & dinvgamma() in R to get the respective functional values

$\rightarrow$ Just have to discretize sample space of each parameter over reasonable ranges for each parameter

$\Rightarrow$ ie we are trying to get



→ once we have a sample from the posterior, we can do inference

→ e.g.) ① posterior mean → $E(\mu) = \frac{1}{M}\sum_{m=1}^{M} \mu^{(m)}$

② posterior prediction → dist → $\tilde{y}^{(m)} \sim N(\mu^{(m)}, (\sigma^2)^{(m)})$, $m = 1, \dots, M$

→ mean → $E(\tilde{y}) = \frac{1}{M}\sum_{m=1}^{M} \tilde{y}^{(m)}$

} → simply estimating the quantities of interest

⇒ this differs from the analytical approach where we ① derive $E(\mu) = \int \mu\, p(\mu, \sigma^2 | y)\, d\sigma^2$

OR ② derive $p(\tilde{y}|y) = \iint p(\tilde{y}, \mu, \sigma^2 | y)\, d\mu\, d\sigma^2$

$= \iint p(\tilde{y} | \mu, \sigma^2, y)\, p(\mu, \sigma^2 | y)\, d\mu\, d\sigma^2$

+ sample directly from this result

- Comparsion to the normal models in frequentist

→ parameter	Frequentist	Bayesian
μ	$\frac{\bar{y}-\mu}{s/\sqrt{n}} \mid \mu, \sigma^2 \sim t_{n-1}$	$\frac{\mu-\bar{y}}{s/\sqrt{n}} \mid y \sim t_{n-1}$
σ^2	$\frac{(n-1)}{\sigma^2} s^2 \mid \mu, \sigma^2 \sim \chi^2_{n-1}$	$\frac{(n-1)}{\sigma^2} s^2 \mid y \sim \chi^2_{n-1}$
	→ This is the foundation of building confidence intervals for μ → For both μ + σ^2 in frequentist, the randomness comes from $\bar{y}$ (i.e. only the data, parameters are fixed unknown constants)	→ In Bayesian, the randomness comes from the parameters (i.e. b/c of μ) → once we condition on the data, $\bar{y}$ + s_n^2 are fixed values

3.2 Informative priors in Normal models

$\theta | y \propto p(y|\theta)\, \pi(\theta)$

- Use historical data to define priors
- **Power prior in Ibrahim and Chen (2000):** Weigh historical data as prior input for analysis of current data, but with a discount.
- $Y_1^0, Y_2^0, \dots, Y_m^0 \sim N(\mu, \sigma^2)$: historical data
- $Y_1, Y_2, \dots, Y_n \sim N(\mu, \sigma^2)$: current data. ✩ Notation → [ ] = distribution of ...
- Unknown parameters: $\boldsymbol{\theta} = (\mu, \sigma^2)$
- **Update prior:** $\left[\boldsymbol{\theta}|y_1^0, \dots, y_m^0\right] \propto [\boldsymbol{\theta}] \cdot \left[y_1^0, \dots, y_m^0|\boldsymbol{\theta}\right]^{a_0}$, where $0 \leq a_0 \leq 1$ ⇒ This is like a "posterior" of θ given the historical data

  ↗ ≈ weight

  - If $a_0 = 0$, $\left[\boldsymbol{\theta}|y_1^0, \dots, Y_m^0\right] \propto [\boldsymbol{\theta}]$ ⇒ no weight (i.e. no consideration) for the historical data ⇒ This is just the standard prior we have been using
  - If $a_0 = 1$, $\left[\boldsymbol{\theta}|y_1^0, \dots, y_m^0\right] \propto [\boldsymbol{\theta}] \cdot \left[y_1^0, \dots, y_m^0|\boldsymbol{\theta}\right]$ ⇒ posterior distribution based on the historical data

    ↳ we are still using θ as the prior! Then we make the posterior off the current data like usual

ex) Normal model

→ $\theta = (\mu, \sigma^2)$

→ $\pi(\theta) \propto (\sigma^2)^{-1}$ (from earlier)

→ $\underbrace{p(y_1^0, \dots, y_m^0 | \mu, \sigma^2)}_{\text{historical data dist}} = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i^0-\mu)^2}{2\sigma^2}\right\}$

→ update prior → $\pi(\mu, \sigma^2 | y_1^0, \dots, y_m^0) \propto \pi(\mu, \sigma^2)\left[p(y_1^0, \dots, y_m^0 | \mu, \sigma^2)\right]^{a_0}$

$= (\sigma^2)^{-1}\left[(\sigma^2)^{-m/2} \exp\left\{-\frac{\sum_{i=1}^{m}(y_i^0-\mu)^2}{2\sigma^2}\right\}\right]^{a_0}$

- **Posterior:**

→ $p(\mu, \sigma^2 | \underbrace{y_1, \dots, y_n, y_1^0, \dots, y_m^0}_{\text{→ now based on both sets of data}}) \propto p(y_1, \dots, y_n | \mu, \sigma^2) \cdot \pi(\mu, \sigma^2 | y_1^0, \dots, y_m^0)$

$= \left[(\sigma^2)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}\right\}\right] \cdot \left[(\sigma^2)^{-1}\left[(\sigma^2)^{-m/2} \exp\left\{-\frac{\sum_{i=1}^{m}(y_i^0-\mu)^2}{2\sigma^2}\right\}\right]^{a_0}\right]$

⟨ → continue analysis like usual ... ⟩

3.3 Conjugate priors

- Data distribution: $p(\boldsymbol{y}|\mu, \sigma^2) \propto \left(\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\mu\right)^2\right\}$
- Conjuate priors are more complicated in multiparameter case.

$\pi(\mu, \sigma^2) = \pi(\mu | \sigma^2)\, \pi(\sigma^2)$

$= \text{Normal} * \text{Inv Gamma}$ (for this example)

$$\sigma^2 \sim \text{Inverse Gamma}\,(\alpha, \beta)$$

$$\mu|\sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{K_0}\right)$$

where the density function of Inverse Gamma(α, β) is

$$f\left(x\right) = \frac{\beta^{\alpha}}{\Gamma\left(\alpha\right)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right)$$

- **Joint Posterior** (μ, σ^2)

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

$= \quad \downarrow \quad \cdot\, \pi(\mu | \sigma^2) \cdot \pi(\sigma^2)$ → kind of making the prior sequentially

$\propto \left[(\sigma^2)^{-n/2} \exp\left\{ -\frac{\sum_{i=1}^{n} (y_i - \mu)^2}{2\sigma^2} \right\} \right] \cdot \left[(\sigma^2)^{-1/2} \exp\left\{ -\frac{(\mu - \mu_0)^2}{2\,\sigma^2/k_0} \right\} \right] \cdot \left[(\sigma^2)^{-\alpha - 1} \exp\left\{ -\beta/\sigma^2 \right\} \right]$ < dropped all/parts of normalizing constants in $\pi(\mu|\sigma^2)$ + $\pi(\sigma^2)$ >

$= (\sigma^2)^{-(\frac{n}{2} + \alpha + 3/2)} \exp\left\{ -\frac{1}{\sigma^2} \left[\frac{\sum_{i=1}^{n} (y_i - \mu)^2}{2} + \frac{k_0(\mu - \mu_0)^2}{2} + \beta \right] \right\}$

→ Now, we want: ① $p\left(\mu | \sigma^2, \boldsymbol{y}\right)$ and ② $p\left(\sigma^2 | \boldsymbol{y}\right)$

↳ conditional posterior ↳ marginal posterior

Always true w/ Bayes: $\Rightarrow p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{\text{Joint}}{\text{marginal}} \propto$ joint ☆☆

↳ always factors to $= p(B|A) \cdot p(A)$ switched

$\rightarrow$ ① $p(\mu | \sigma^2, \underset{\sim}{y}) = \frac{p(\mu, \sigma^2, \underset{\sim}{y})}{p(\sigma^2, \underset{\sim}{y})}$ < conditional probability >

$= \frac{p(\underset{\sim}{y} | \mu, \sigma^2) \cdot \pi(\mu, \sigma^2)}{p(\sigma^2, \underset{\sim}{y})}$

$\propto p(\underset{\sim}{y} | \mu, \sigma^2) \cdot \pi(\mu, \sigma^2)$ < = joint posterior of (μ, σ^2) $\Longrightarrow$ Technically also $\propto p(\mu, \sigma^2, \underset{\sim}{y})$ >

$\propto \exp\left\{ -\frac{1}{\sigma^2} \left[\frac{\sum (y_i - \mu)^2}{2} + \frac{k_0(\mu - \mu_0)^2}{2} \right] \right\}$ < substitute + σ^2 is fixed now $\Longrightarrow$ dropped first term >

$\sum (y_i - \mu)^2 = n\mu^2 - 2\mu \underbrace{\sum y_i}_{= n\bar{y}} + \cancel{\sum y_i^2}$ ↳ drop (no μ)

$k_0(\mu - \mu_0)^2 = k\mu^2 - 2k_0\mu\mu_0 + \cancel{\mu_0^2 k_0}$ ↳ drop (no μ)

$\propto \exp\left\{ -\frac{1}{2\sigma^2} \left[(n + k_0)\mu^2 - 2(n\bar{y} + k_0\mu_0)\mu \right] \right\}$

$\left\{ \rightarrow \text{if } x \sim N(\mu_0, \sigma^2) \Rightarrow f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} (x - \mu_0)^2 \right\} \quad \downarrow \quad = \quad \downarrow \quad \exp\left\{ -\frac{1}{2\sigma^2} (x^2 - 2\mu_0 x + \mu_0^2) \right\} \right.$

$\sim N(\mu_n, \sigma_n^2)$

$\left\{ \begin{array}{l} \rightarrow \frac{1}{\sigma_n^2} = \frac{n + k_0}{\sigma^2} \\ \rightarrow \frac{\mu_n}{\sigma_n^2} = \frac{n\bar{y} + k_0\mu_0}{\sigma^2} \end{array} \right. \Longrightarrow \left\{ \begin{array}{l} \rightarrow \sigma_n^2 = \frac{\sigma^2}{n + k_0} \\ \rightarrow \mu_n = \frac{n\bar{y} + k_0\mu_0}{\sigma^2} \times \frac{\sigma^2}{n + k_0} \\ \quad = \frac{n\bar{y} + k_0\mu_0}{n + k_0} \end{array} \right.$

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

→ rearranging parameters → $\mu_n = \frac{n\bar{y} + k_0\mu_0}{n + k_0} = \left(\frac{n}{n + k_0}\right)\bar{y} + \left(\frac{k_0}{n + k_0}\right)\mu_0$

$\Longrightarrow$ weighted average of sample mean + prior mean

→ situational results → As $k_0 \rightarrow 0$, $\mu_n \rightarrow \bar{y}$ + $\frac{\sigma^2}{k_0} \rightarrow +\infty$

→ recall prior → $\mu | \sigma^2 \sim N(\mu_0, \frac{\sigma^2}{k_0})$

$\Rightarrow \pi(\mu | \sigma^2) \propto 1$ (flat, improper prior)

$\Rightarrow$ posterior $\rightarrow p(\mu | \sigma^2, \underset{\sim}{y}) \propto p(\underset{\sim}{y} | \mu, \sigma^2)\, \pi(\mu, \sigma^2)$ < restating from above >

$= \quad \downarrow \quad \cdot\, \underbrace{\pi(\mu | \sigma^2)}_{\propto 1} \cdot \underbrace{\cancel{\pi(\sigma^2)}}_{\text{constant w/ respect to } \mu}$

$\propto p(\underset{\sim}{y} | \mu, \sigma^2)$ → = likelihood $L(\mu | \underset{\sim}{y})$ < recall likelihood = joint density, but from a different perspective >

$\sim N(\bar{y}, \frac{\sigma^2}{n})$ $\Big\}$ $\propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \right\}$

→ As $k_0 \rightarrow +\infty$, $\mu_n \rightarrow \mu_0$ + $\sigma_n^2 \rightarrow 0$

☆ Key conceptual takeaway

→ Every posterior is the joint distribution of the parameters + data viewed as a function of the unknown quantities

→ ② $p(\sigma^2 \mid \underset{\sim}{y}) = \int p(\mu, \sigma^2 \mid \underset{\sim}{y})\, d\mu$

< substitute from above + rearrange >

$$= \int (\sigma^2)^{-(\frac{n}{2} + \alpha + 3/2)} \exp\left\{ -\beta/\sigma^2 \right\} \exp\left\{ -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} (y_i - \mu)^2 + k_0 (\mu - \mu_0)^2 \right] \right\} d\mu$$

$$\sim \text{Inv Gamma}\left(\frac{n}{2} + \alpha,\ \beta + \frac{(n-1)}{2} s^2 + \frac{n k_0}{2(k_0 + n)} (\mu_0 - \bar{y}^2) \right)$$

↳ $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$

- Sampling procedure → Bayes update → $p(\mu, \sigma^2 \mid \underset{\sim}{y}) \propto p(\underset{\sim}{y} \mid \mu, \sigma^2)\, \pi(\mu, \sigma^2)$ < Tells you how to construct joint posterior from model >

  Joint posterior

  → Probability factorization → $p(\mu, \sigma^2 \mid \underset{\sim}{y}) = \underbrace{p(\mu \mid \sigma^2, \underset{\sim}{y})}_{\text{conditional posterior}} \cdot \underbrace{p(\sigma^2 \mid \underset{\sim}{y})}_{\text{marginal posterior}}$ < tells you how that joint posterior can be factored into pieces >

  SAME THING!!

  → Steps to sample → 1) Draw $(\sigma^2)^{(m)}$ from $\text{Inv Gamma}\left(\frac{n}{2} + \alpha,\ \beta + \frac{(n-1)}{2} s^2 + \frac{n k_0}{2(k_0 + n)} (\mu_0 - \bar{y}^2) \right)$

  → 2) Draw $\mu^{(m)}$ from $N(\mu_n, \sigma_n^2)$, $\mu_n = \frac{n\bar{y} + k_0 \mu_0}{n + k_0}$, $\sigma_n^2 = \frac{(\sigma^2)^{(m)}}{n + k_0}$

  → 3) posterior predictive → $\tilde{y}^{(m)} \sim N(\mu^{(m)}, \sigma^{2\,(m)})$

3.4 Multinomial Model

→ We only need to estimate $J-1$ parameters b/c if we know $J-1$ of the θ_j, then w/ the constraint we know what the remaining θ_j must be. In other words, only $J-1$ unknown parameters.

- An extension of binomial distribution. → see below.
- There are J catergories, $Y_1, \ldots, Y_J | \theta_1, \theta_2, \ldots, \theta_J \sim \text{Multinomial}(n, \theta_1, \theta_2, \ldots, \theta_J)$, where $\sum_{j=1}^{J} Y_j = n$ and $\sum_{j=1}^{J} \theta_j = 1$.

  ↳ count for each category ↳ probability of each category ↳ This is a constraint

$$p(\boldsymbol{y}) = P(y_1, y_2, \ldots, y_J) = P(Y_1 = y_1, \ldots, Y_J = y_J) = \frac{n!}{y_1! \cdots y_J!} \theta_1^{y_1} \cdots \theta_J^{y_J}$$

- **Conjugate prior:** Dirichlet priors for $\theta_1, \ldots, \theta_J$

$$f(\theta_1, \ldots, \theta_J | \alpha_1, \ldots, \alpha_J) = \frac{\prod_{j=1}^{J} \Gamma(\alpha_i)}{\Gamma\left(\sum_{j=1}^{J} \alpha_i\right)} \theta_1^{\alpha_1 - 1} \cdots \theta_J^{\alpha_J - 1}, \quad \sum_{j=1}^{J} \theta_j = 1, \quad 0 < \theta_j < 1$$

no constraints on α_j

→ The Dirichlet is a good choice for conditional data (i.e. probabilities) b/c it naturally has constraints

- **Posterior distribution**

→ $p(\theta_1, \ldots, \theta_J \mid y_1, \ldots, y_J) \propto \theta_1^{y_1} \cdots \theta_J^{y_J} \cdot \theta_1^{\alpha_1 - 1} \cdots \theta_J^{\alpha_J - 1}$

$= \theta_1^{y_1 + \alpha_1 - 1} \cdots \theta_J^{y_J + \alpha_J - 1}$

⇒ $\theta_1, \ldots, \theta_J \mid y_1, \ldots, y_J \sim \text{Dirichlet}(y_1 + \alpha_1, \ldots, y_J + \alpha_J)$

→ Note → $E(\theta_j \mid y_1, \ldots, y_J) = \frac{y_j + \alpha_j}{\sum_{j=1}^{J} (y_j + \alpha_j)}$

→ Extension of Binomial Cont...

→ If $J = 2$ → $Y_1, Y_2 \mid \theta_1, \theta_2 \sim \text{Multinomial}(n, \theta_1, \theta_2)$, $n = Y_1 + Y_2$, $1 = \theta_1 + \theta_2$

$= \text{Binomial}(n, \theta_1)$

→ If we think of Y_1 as # of successes + Y_2 as # of failures, then θ_1 = prob of success + $\theta_2 = 1 - \theta_1$ = prob of failure

⇒ $p(y_1, y_2) = \frac{n!}{y_1!\, y_2!} \theta_1^{y_1} \theta_2^{y_2}$

→ let $y_2 = n - y_1$

$= \frac{n!}{y_1!\,(n - y_1)!} \theta_1^{y} (1 - \theta_1)^{n - y_1}$ < familiar binomial pdf >

→ For multivariate, we only showed derivations for normal + multinomial models b/c we can find a conjugate prior, so the resulting posteriors have a nice form

→ But usually the integrals are too hard to calculate, so we just use simulation pretty much all of the time

→ univariate case → $y_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$
(identically distributed, not necessarily ⫫ b/c each y_i represents only one observation)

3.5 Multivariate Normal model

- $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \overset{iid}{\sim} MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $E(\boldsymbol{Y}_i) = \boldsymbol{\mu}$ and $Cov(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}$, with

$\underset{p\times 1}{\underset{\sim}{y_i}} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ ↓ p×1 vector ↘ p×p matrix

$$p(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right)$$

↳ Joint distribution of $\underset{\sim}{y_i}$

↙ so for each subject, we record numerous things (e.g. height, weight, age, etc.) rather than one variable at a time

→ ex) if $p = 2 \Rightarrow y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \end{bmatrix}$, this is the joint dist of y_{i1} + y_{i2} (same subject, different RV)

id	y_{i1} weight	y_{i2} hight
1	⋮	⋮
2		
⋮	⋮	⋮
n		

→ Here, instead of studying only one RV, we study the joint dist of several RVs

→ If all RVs are ⫫, then the joint distribution $\underset{\sim}{y_i}$ is just the product of all $y_{i1}, \ldots, y_{ip}$ densities. But if not all ⫫, they will have some covariance

→ $\Sigma = Cov(\underset{\sim}{y_i}) = Cov\left(\begin{bmatrix} y_{i1} \\ y_{i2} \end{bmatrix}\right) = \begin{bmatrix} V(y_{i1}) & Cov(y_{i1}, y_{i2}) \\ Cov(y_{i1}, y_{i2}) & V(y_{i2}) \end{bmatrix}$

- $\boldsymbol{\Sigma}$ is known and $\boldsymbol{\mu}$ is the parameter of interest
  - **Data distribution**

→ notation note → By default, $\underset{\sim}{y_i}$'s are columns of values (ie dimensions $1 \times p$) so we transpose them so can do matrix multiplication with $\underset{p\times p}{\Sigma^{-1}}$

→ $p(\underset{\sim}{y_1}, \ldots, \underset{\sim}{y_n} | \underset{\sim}{\mu}) \propto \prod_{i=1}^{n} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\underset{\sim}{y_i} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{y_i} - \underset{\sim}{\mu})\right\}$

$= |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (\underset{\sim}{y_i} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{y_i} - \underset{\sim}{\mu})\right\}$

  - **Conjugate prior for $\boldsymbol{\mu}$, $\boldsymbol{\mu} \sim MVN(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$**

↳ multivariate normal → $\pi(\underset{\sim}{\mu}) \propto \exp\left\{-(\underset{\sim}{\mu} - \underset{\sim}{\mu_0})^T \Lambda_0^{-1} (\underset{\sim}{\mu} - \underset{\sim}{\mu_0})\right\}$

  - **Posterior distribution**

→ $p(\underset{\sim}{\mu} | \underset{\sim}{y_1}, \ldots, \underset{\sim}{y_n}) \propto p(\underset{\sim}{y_1}, \ldots, \underset{\sim}{y_n} | \underset{\sim}{\mu}) \cdot \pi(\underset{\sim}{\mu})$

$\propto \exp\left\{-\sum_{i=1}^{n} (\underset{\sim}{y_i} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{y_i} - \underset{\sim}{\mu})\right\} \exp\left\{-(\underset{\sim}{\mu} - \underset{\sim}{\mu_0})^T \Lambda_0^{-1} (\underset{\sim}{\mu} - \underset{\sim}{\mu_0})\right\}$

$\sim MVN(\mu_n, \Lambda_n)$, where $\Lambda_n = (n\Sigma^{-1} + \Lambda_0^{-1})^{-1}$

$\mu_n = (n\Sigma^{-1} + \Lambda_0^{-1})^{-1}(\Lambda_0^{-1}\underset{\sim}{\mu_0} + n\Sigma^{-1}\underset{\sim}{\bar{y}})$

→ Extension of univariate Normal

→ similar to what we have seen before, just extended to matrix form

⇒ (μ, σ^2) becomes $(\underset{\sim}{\mu}, \Sigma)$

- $\mathbf{\Sigma}$ is unknown
  - Conjuate priors: $\underset{\sim}{\mu} \mid \Sigma$ [redacted] $\sim MVN\,(\boldsymbol{\mu}_0, \mathbf{\Sigma}/K_0)$ and $\mathbf{\Sigma} \sim$ inverse Wishart$(\gamma_0, \Lambda_0^{-1})$

→ ≈ kind of like covariance

(capital lambda)

↳ extension of invGamma, used to describe distributions of positive definite matrices

→ recall positive definite matrix def → $\underset{\sim}{x}^T \Sigma \underset{\sim}{x} > 0$ for any $\underset{\sim}{x} \neq \underset{\sim}{0}$

→ Note that this derivation for posterior $\underset{\sim}{\mu}$ & Σ is long & hard, but the process is similar to what we have done before

→ $p(\underset{\sim}{\mu} \mid \underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_n) \propto p(\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_n \mid \underset{\sim}{\mu}, \Sigma)\, \pi(\underset{\sim}{\mu}, \Sigma)$

↓ = ↓ $* \; \pi(\mu \mid \Sigma)\, \pi(\Sigma)$

→ $\underset{\sim}{\mu} \mid \Sigma, \underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_n \sim MVN(\mu_n, \Sigma_n)$

→ $\Sigma_n = \dfrac{\Sigma}{K_0 + n}$, $\underset{\sim}{\mu}_n = \dfrac{K_0}{K_0 + n} \underset{\sim}{\mu}_0 + \dfrac{n}{K_0 + n} \underset{\sim}{\bar{y}}$ < some weighted avg form as univariate case, except w/ vectors >

→ $\Sigma \mid \underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_n \sim$ Inverse Wishart $(\gamma_n, \Lambda_n^{-1})$

→ $\gamma_n = \gamma_0 + n$, $\Lambda_n = \Lambda_0 + S + \dfrac{K_0 n}{K_0 + n} (\underset{\sim}{\bar{y}} - \underset{\sim}{\mu}_0)(\underset{\sim}{\bar{y}} - \underset{\sim}{\mu}_0)^T$

↳ $= \sum_{i=1}^{n} (\underset{\sim}{y}_i - \underset{\sim}{\bar{y}})(\underset{\sim}{y}_i - \underset{\sim}{\bar{y}})^T$

→ Sampling procedure

→ step 1 → sample $\Sigma^{(m)}$ from inverse Wishart

→ step 2 → sample $\underset{\sim}{\mu}^{(m)}$ from MVN