Bayesian Statistics

more advanced model → more advance computation
MCMC : possibilities of models ↑

① Statistical modelling
statistical model : imitate & approximates
data generating process
– relationship btw variables

objective ① Quantify uncertainty
② Inference
③ Measure support for hypotheses
④ Prediction ← ML focus

process ① Understand the problem
② Plan & collect data
③ Explore data
context
④ Postulate model
↳ complexity vs generazability (Bias-Var tradeoff
⑤ Fit model - estimate parameters
↳ "Bayesian approach"
⑥ Check model
⑦ Iterate
⑧ Use model
*

② heights $n=15$ men
$y_i = \mu + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ $i = 1, \cdots, n \ (=15)$
$y_i \overset{iid}{\sim} N(\mu, \sigma^2)$ → can generate fake instances
: frequentist approach that fits the distribution
↳ change of estimates of μ, σ

Bayesian approach : uncertainty of μ, σ with probability
<prior> ↳ r.v. with distribution

① likelihood : $P(\text{data} \mid \text{unknown parameter}) : P(y \mid \theta)$.
② prior $P(\theta)$ ↖ known
statistical modeling ← ③ posterior $P(\theta \mid y) = \dfrac{P(\theta, y)}{P(y)} = \dfrac{P(\theta, y)}{\int P(\theta, y)\, d\theta} = \dfrac{P(y \mid \theta) P(\theta)}{\int P(y \mid \theta) P(\theta)\, d\theta}$

model hierarchy

$y_i \mid \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$

$p(\mu, \sigma^2) = p(\mu)\, p(\sigma^2)$

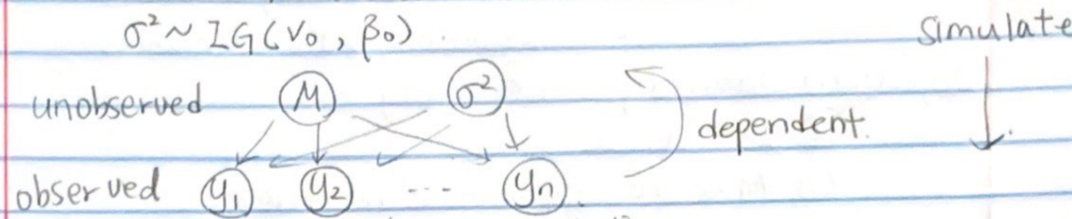
? Conjugate prior?

μ (σ^2 known) $\sim N$, σ^2 (μ known) $\sim IG$

?

$\mu \sim N(\mu_0, \sigma_0^2)$

$\sigma^2 \sim IG(\nu_0, \beta_0)$



graphical representation

→ simulate starting from variables w/o dependency

$y_i \mid \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$

$\mu \mid \sigma^2 \overset{iid}{\sim} N(\mu, \frac{\sigma^2}{w_0})$

$\sigma^2 \overset{iid}{\sim} IG(\nu_0, \beta_0)$

$$p(y_1, \cdots, y_n, \mu, \sigma^2) = p(y_1, \cdots, y_n \mid \mu, \sigma^2)\, p(\mu \mid \sigma^2)\, p(\sigma^2)$$

$$= \prod_{i=1}^{n} \left[N(y_i \mid \mu, \sigma^2) \right] \times N\left(\mu \mid \mu_0, \frac{\sigma^2}{w_0}\right) \times IG(\sigma^2 \mid \nu_0, \beta_0)$$

* Bayes theorem. $p(\theta \mid y) = \dfrac{p(y \mid \theta)\, p(\theta)}{\int p(y \mid \theta)\, p(\theta)\, d\theta} \propto p(y \mid \theta)\, p(\theta)$

$\propto$ ~~$p(\mu, \sigma^2$~~ $p(\mu, \sigma^2 \mid y_1, \cdots, y_n)$

center, scale, df

$n = 10 \quad y_i \mid \mu \overset{iid}{\sim} N(\mu, 1) \qquad \mu \sim t(0, 1, 1)$

$$p(\mu \mid y_1, \cdots, y_n) \propto \prod_{i=1}^{n} \left[N(\mu, 1) \right] t(0, 1, 1)$$

$$= \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i - \mu)^2\right) \right] \frac{1}{\pi(1 + \mu^2)}$$

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

$$\propto \frac{\exp(n(\bar{y}\mu - \mu^2/2))}{1 + \mu^2}$$

hard to compute ↘

: not a form of standard distribution

③ $\theta \sim Ga(a, b) \quad a=2 \quad b=\frac{1}{3}$

$E(\theta) = \int_0^\infty \theta p(\theta) d\theta = \int_0^\infty \theta \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} d\theta = \frac{a}{b}$

$\theta_i^* \quad i=1, \cdots, m$ by LLN, CLT

$\bar{\theta}^* = \frac{1}{m} \sum_{i=1}^{m} \theta_i^*$

$Var(\theta) = \int_0^\infty (\theta - E(\theta))^2 p(\theta) d\theta$

$h(\theta) = \int h(\theta) p(\theta) d\theta = E[h(\theta)] \approx \frac{1}{m} \sum_{i=1}^{m} h(\theta_i^*)$

ex. $h(\theta) = I_{\theta<5}(\theta)$

$E(h(\theta)) = \int_0^\infty I_{\theta<5}(\theta) p(\theta) d\theta$

$= \int_0^5 1 \cdot p(\theta) d\theta = Pr[0 < \theta < 5] \approx \frac{1}{m} \sum_{i=1}^{m} I_{\theta^*<5}(\theta_i^*)$

→ can approximate by drawing many samples θ^*

$\bar{\theta}^* \overset{\cdot}{\sim} N(E(\theta), \frac{Var(\theta)}{m})$

$\widehat{Var}(\theta) = \frac{1}{m} \sum_{i=1}^{m} (\theta_i^* - \bar{\theta}^*)^2 \qquad SE = \sqrt{\widehat{Var}(\theta)/m}$

$y \mid \phi \sim Bin(10, \phi) \qquad$ ~~$p(y, \phi) = p(\phi) p(y \mid \phi)$~~

$\phi \sim Beta(2, 2)$

Simulate : ① ϕ_i^* from Beta

② Given ϕ_i^*, draw $y_i^* \sim Bin(10, \phi_i^*)$

→ (y_i^*, ϕ_i^*)