this time: multivariate θ

next time: MCMC

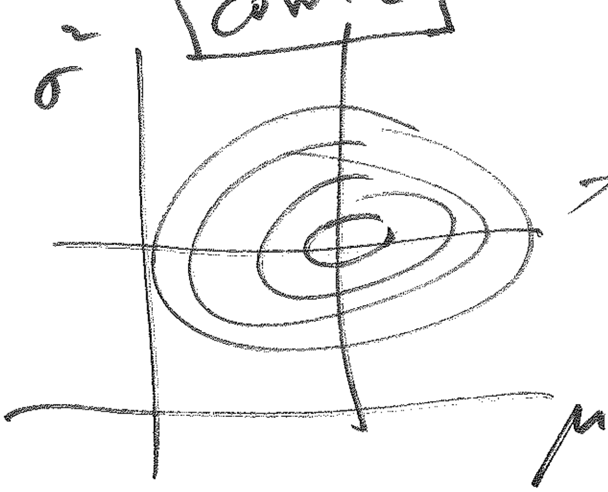
read: (or before)

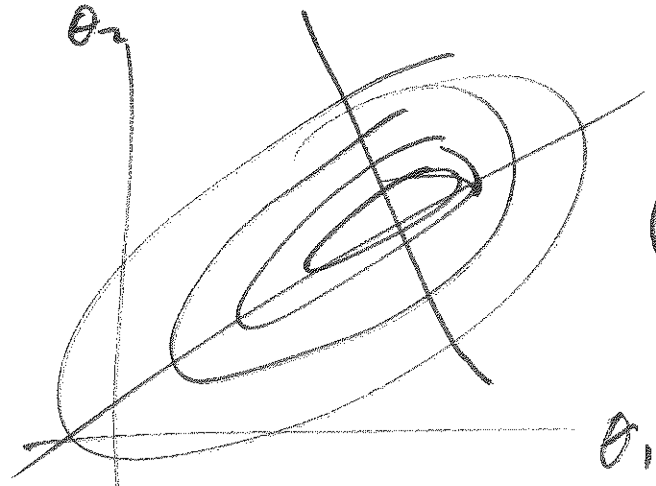
AMS 206

22 Feb 18

Quiz 4: do everything in terms of $\theta = \sigma^2$ ①

read: ⓒ ch. 10, 11

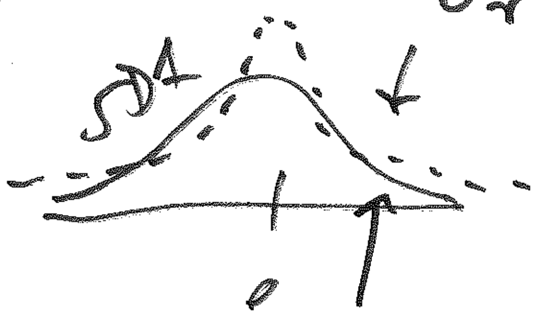




positive corr.

$(\nu > 0)$ degrees of freedom

t_ν, ν small

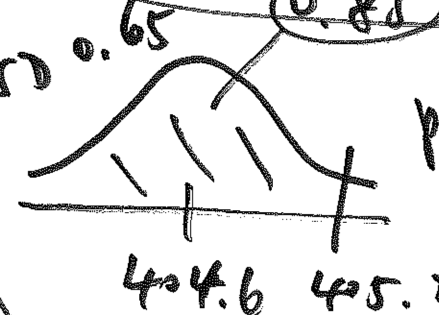


standard normal

$$P_B(\mu < 405.25 \mid y, GB(\text{diffuse prior})) = 0.85$$

fixed

$$P_F(\mu < 405.25 \mid GB) = \text{undefined}$$



$p(\mu \mid y \ldots)$

[Q:] Predictive diagnostics reveal that ②

$M_{current}$ is not good enough; what now?
(model)

[A:] Do model expansion: embed

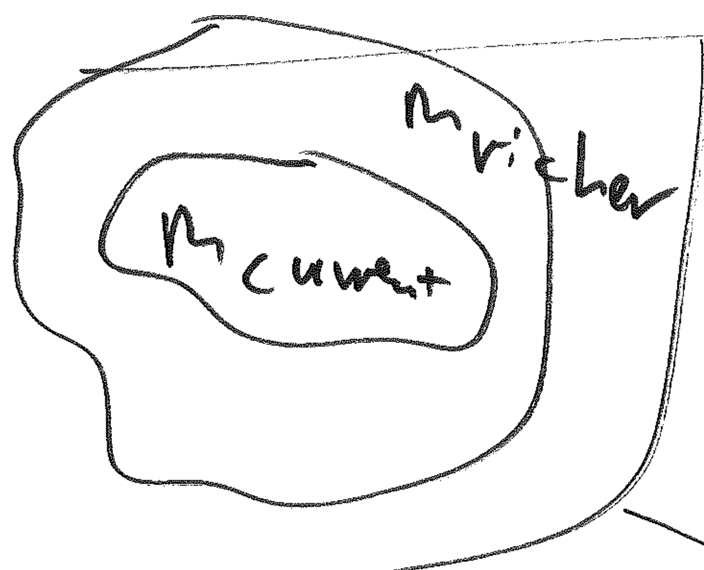
$M_{current}$ in a richer model class,

of which it's a special case, with

the "direction" of "richer" determined

by what the predictive diagnostics

showed you is wrong with $M_{current}$



here, deficiency of

$M_{current}$: $\begin{cases} (\mu, \sigma^2 \mid \chi^{-2}) \sim \overset{\text{diffuse}}{\chi^{-2}} \\ (Y_i \mid \mu\, \sigma^2\, \mathcal{G}\, \mathcal{B}) \overset{IID}{\sim} N(\mu, \sigma^2) \\ (i = 1, \dots, n) \end{cases}$

weak point →

M_{richer} should have heavier tails than $M_{current}$: $\mathcal{G}$

M_{richer}: $\begin{cases} (\mu, \sigma^2, \nu) \sim p(\boxed{\mu, \sigma^2}, \nu) & \text{diffuse; small } \nu \\ (Y_i \mid \mu\, \sigma^2\, \nu\, (t_\nu)\, B) \overset{IID}{\sim} t_\nu(\mu, \sigma^2) \\ (i = 1, \dots, n) \end{cases}$ (3)

---

t_ν has no conjugate prior; **Q:** what now?

A: we need a new, more general, Bayesian computing method: MCMC

A₂: standard numerical integration (only works for small k)

$$\underset{\sim}{\theta} = (\theta_1, \dots, \theta_k)$$

$\downarrow$ (k large)

Bayes's Thm:

$y = (y_1, \dots, y_n)$

$$p(\underset{\sim}{\theta} \mid \underset{\sim}{y}) = c\, p(\underset{\sim}{\theta})\, \ell(\underset{\sim}{\theta} \mid \underset{\sim}{y})$$

---

ex. multiple linear regression model

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + e_i \quad (i = 1, \dots, n), \quad e_i \overset{IID}{\sim} N(0, \sigma^2)$$

$(y = f(x) + e)$ (f: linear)

what if $k > n$?

θ

x_{ij} : j^{th} feature (ML)

j^{th} predictor variable (stat)

$$\underset{\sim}{\theta} = (\beta_0, \beta_1, \dots, \beta_k, \sigma^2) = \underline{\underline{O}}(k)$$

what if ($k >> n$) ?

$\beta_j = 0 \leftrightarrow$ feature j does not help to predict y

($j = 1, \dots, k$)

$$\begin{array}{c|c|cccc} & & & k & & \\ n & y & x_1 & x_2 & \cdots & x_k \end{array}$$

disease genes

ex. $p(\underset{\sim}{\theta}|y) = c\, p(\theta)\, \ell(\theta|y)$

only at most $k_1 < n$ of the β_j are nonzero

($n = 10{,}000$, $k = 32{,}000$)

Icelandic genome study

$$\log p(\underset{\sim}{\theta} \mid \underset{\sim}{y}) = \underbrace{\log p(\underset{\sim}{\theta})} + \boxed{\ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y})}$$

maximum a posteriori (MAP) estimate of $\underset{\sim}{\beta}$

regularizer (complexity penalty)

$-\lambda \, \| \underset{\sim}{\beta} - \underset{\sim}{0} \|^2$ (L^2)

same for β, Max. Lik.

maximize modified likelihood (criterion function)

$p(\underset{\sim}{\theta}) = e^{-\lambda \|\underset{\sim}{\beta} - 0\|}$

= Gaussian