# Mathematical Statistics II

STA2212H S LEC9101

Week 9

March 17 2021

Start recording!


link

# Recap



- overview of Bayesian inference
- posterior predictive distribution
- Bayesian computation
- Example 11.9 AoS
- Bayesian hierarchical models

- overview of Bayesian inference
- posterior predictive distribution
- Bayesian computation
- Example 11.9 AoS
- Bayesian hierarchical models

Addendum: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, both parameters unknown, then the conjugate priors for $\mu$ and $\sigma^2$ are

$$\mu \sim N(\mu_0, \tau^2); \qquad \sigma^2 \sim IG(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2})$$

Inverse Gamma distribution $IG(\alpha, \beta)$ has density $f(t; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{x^{\alpha+1}} e^{-\beta/x}$

# Today

Start Recording

1. Notes from March 12 (Friday) still to be posted
2. Friday March 19 – Stein's paradox
3. empirical Bayes
4. introduction to decision theory

B-H proof

- Mar 22 3.00 – 4.00 pm EDT
  Data Science ARES
  Jesse Cisewski Kehe
  University of Wisconsin-Madison
  "Astrostatistics: From Exoplanets to
  the Large-scale Structure of the Universe"

# Bayesian hierarchical model, recap



- $x_i \mid \theta_i \sim N(\theta_i, v_i)$, independent; $\quad \theta_i \mid \mu \sim N(\mu, \sigma^2), i = 1, \ldots, n$ i.i.d; $\quad \mu \sim N(\mu_0, \tau^2)$

# Bayesian hierarchical model, recap

SM Ex. 11.25

- $x_i \mid \theta_i \sim N(\theta_i, v_i)$, independent; $\quad \theta_i \mid \mu \sim N(\mu, \sigma^2), i = 1, \dots, n$ i.i.d; $\quad \mu \sim N(\mu_0, \tau^2)$
- $v_i, i = 1, \dots, n, \sigma^2, \mu_0, \tau^2$ all known

## Bayesian hierarchical model, recap

SM Ex. 11.25

- $x_i \mid \theta_i \sim N(\theta_i, v_i)$, independent; $\quad \theta_i \mid \mu \sim N(\mu, \sigma^2), i = 1, \ldots, n$ i.i.d; $\quad \mu \sim N(\mu_0, \tau^2)$
- $v_i, i = 1, \ldots, n, \sigma^2, \mu_0, \tau^2$ all known
- $$\mathrm{E}(\theta_i \mid \boldsymbol{x}) = x_i \frac{\sigma^2}{\sigma^2 + v_i} + \mathrm{E}(\mu \mid \boldsymbol{x})(1 - \frac{\sigma^2}{\sigma^2 + v_i})$$

- $x_i \mid \theta_i \sim N(\theta_i, v_i)$, independent; $\quad \theta_i \mid \mu \sim N(\mu, \sigma^2), i = 1, \dots, n$ i.i.d; $\quad \mu \sim N(\mu_0, \tau^2)$
- $v_i, i = 1, \dots, n, \sigma^2, \mu_0, \tau^2$ all known
- $$\mathrm{E}(\theta_i \mid \boldsymbol{x}) = x_i \frac{\sigma^2}{\sigma^2 + v_i} + \mathrm{E}(\mu \mid \boldsymbol{x})(1 - \frac{\sigma^2}{\sigma^2 + v_i})$$
- $$\mathrm{E}(\mu \mid \boldsymbol{x}) = \frac{\mu_0/\tau^2 + \sum x_i/(\sigma^2 + v_i)}{1/\tau^2 + \sum 1/(\sigma^2 + v_i)}$$

# Bayesian hierarchical model, recap



- $x_i \mid \theta_i \sim N(\theta_i, v_i)$, independent; $\quad \theta_i \mid \mu \sim N(\mu, \sigma^2), i = 1, \ldots, n$ i.i.d; $\quad \mu \sim N(\mu_0, \tau^2)$
- $v_i, i = 1, \ldots, n, \sigma^2, \mu_0, \tau^2$ all known
- $$\mathrm{E}(\theta_i \mid \boldsymbol{x}) = x_i \frac{\sigma^2}{\sigma^2 + v_i} + \mathrm{E}(\mu \mid \boldsymbol{x})(1 - \frac{\sigma^2}{\sigma^2 + v_i})$$
- $$\mathrm{E}(\mu \mid \boldsymbol{x}) = \frac{\mu_0/\tau^2 + \sum x_i/(\sigma^2 + v_i)}{1/\tau^2 + \sum 1/(\sigma^2 + v_i)}$$
- If $\sigma^2$ unknown, then need to sample from the posterior, no closed form available

# Bayesian hierarchical model, recap

- $x_i \mid \theta_i \sim N(\theta_i, v_i)$, independent; $\theta_i \mid \mu \sim N(\mu, \sigma^2), i = 1, \ldots, n$ i.i.d; $\mu \sim N(\mu_0, \tau^2)$
- $v_i, i = 1, \ldots, n, \sigma^2, \mu_0, \tau^2$ all known
- $$\mathrm{E}(\theta_i \mid \boldsymbol{x}) = x_i \frac{\sigma^2}{\sigma^2 + v_i} + \mathrm{E}(\mu \mid \boldsymbol{x})(1 - \frac{\sigma^2}{\sigma^2 + v_i})$$
- $$\mathrm{E}(\mu \mid \boldsymbol{x}) = \frac{\mu_0/\tau^2 + \sum x_i/(\sigma^2 + v_i)}{1/\tau^2 + \sum 1/(\sigma^2 + v_i)}$$
- If $\sigma^2$ unknown, then need to sample from the posterior, no closed form available
- Figure 11.11 applies similar ideas, plus sampling from the posterior, in logistic regression

# ...Bayesian hierarchical model, recap

SM Ex. 11.25



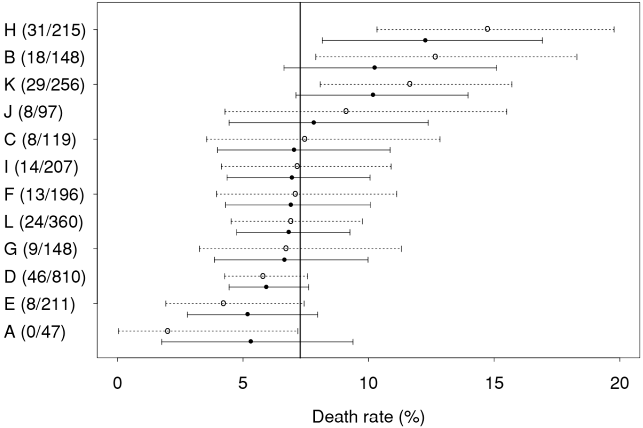


**Figure 11.11** Posterior summaries for mortality rates for cardiac surgery data. Posterior means and 0.95 equitailed credible intervals for separate analyses for each hospital are shown by hollow circles and dotted lines, while blobs and solid lines show the corresponding quantities for a hierarchical model. Note the shrinkage of the estimates for the hierarchical model towards the overall posterior mean rate, shown as the solid vertical line; the hierarchical intervals are slightly shorter than those for the simpler model.

# Empirical Bayes



As above, $x_i \mid \theta_i \sim N(\theta_i, v_i)$, independent; $\quad \theta_i \mid \mu \sim N(\mu, \sigma^2), i = 1, \ldots, n$ i.i.d;

$v_i, i = 1, \ldots, n, \sigma^2, \tau^2$ all known

no hyper-prior now

marginal distribution of $x_i$ is $N(\mu, v_i + \sigma^2)$

suggests

$$\hat{\mu}_{EB} = \frac{\sum x_i/(v_i + \sigma^2)}{\sum 1/(v_i + \sigma^2)}$$

and then

$$\hat{\theta}_{i,EB} = x_i \frac{\sigma^2}{v_i + \sigma^2} + \hat{\mu}_{EB} \frac{v_i}{v_i + \sigma^2}$$

These estimates of $\theta_i$ are biased, but might have smaller mean-squared error, e.g.

"large sets of parallel situations carry their own prior information "

EH, Ch.6

**Table 6.1** *Counts $y_x$ of number of claims $x$ made in a single year by 9461 automobile insurance policy holders. Robbins' formula* (6.7) *estimates the number of claims expected in a succeeding year, for instance* 0.168 *for a customer in the $x = 0$ category. Parametric maximum likelihood analysis based on a gamma prior gives less noisy estimates.*

| Claims $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Counts $y_x$ | 7840 | 1317 | 239 | 42 | 14 | 4 | 4 | 1 |
| Formula (6.7) | .168 | .363 | .527 | 1.33 | 1.43 | 6.00 | 1.75 | |
| Gamma MLE | .164 | .398 | .633 | .87 | 1.10 | 1.34 | 1.57 | |

# Empirical Bayes

# Statistical decision theory

- Loss function
- Risk of an estimator

expected loss

- Mean-squared error
- Bayes risk
- maximum risk

- Loss function

- Risk of an estimator

expected loss

- Mean-squared error

- Bayes risk

- maximum risk

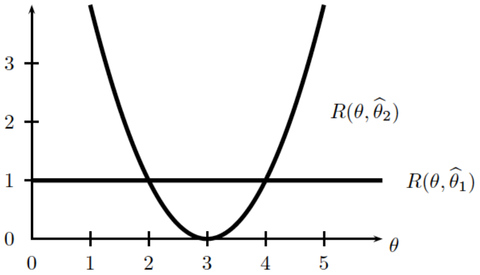


FIGURE 12.1. Comparing two risk functions. Neither risk function dominates the other at all values of $\theta$.

# Examples

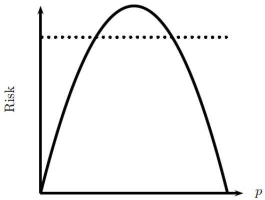


FIGURE 12.2. Risk functions for $\widehat{p}_1$ and $\widehat{p}_2$ in Example 12.3. The solid curve is $R(\widehat{p}_1)$. The dotted line is $R(\widehat{p}_2)$.

# Maximum risk; Bayes risk



$$\hat{p}_1 = \bar{X}; \qquad R(p, \hat{p}_1) = p(1-p)/n$$

$$\hat{p}_2 = \frac{n\bar{X} + \sqrt{n/4}}{n + \sqrt{n}}; \qquad R(p, \hat{p}_2) = \frac{n}{4(n+\sqrt{n})^2}$$

# Bayes rules

# Admissibility