this time: foundations,

next time: multiparameter problems

read: DSch.8

AMS 132

today I'll finish covering all the material in hwk 1

16 Feb 17

(1)

eCommons site now ready for you to upload your PDF of hwk 1

drop-dead due date for hwk 1 at eCommons set to 23 Mar, but please get your assignment posted to eCommons by next Tue, when hwk 2 will be assigned

next Mon is a holiday, so no discussion sections next week

Ingredients in the full Bayesian paradigm

problem $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$

questions

context

I will regard this mapping as unique

$\downarrow$

$(\theta, \mathbb{D}, \mathcal{B})$

unknown of principal interest

data set

background T/F propositions (all T)

In the Bayesian story, to draw good inferences, make good predictions, and help people make good decisions under uncertainty, you need to build a <u>probability model</u> that relates unknown quantities (θ) to known quantities (D, B).

Inference & prediction:

- $p(\theta \mid B)$ ✓ prior dist
- $p(D \mid \theta\, B)$ ✓ sampling dist

$$\downarrow$$

$$\ell(\theta \mid D\, B) = \underset{\text{positive constant}}{c} \cdot p(\theta \mid D\, B)$$

model for inf. &/or pred.: $M = \{ p(\theta \mid B),\ \ell(\theta \mid D\, B) \}$

Decision

- $(\mathcal{A} \mid \mathcal{B})$ = action space $= \{a_1, a_2, \ldots\}$

possible actions

- $U(a, \theta \mid \mathcal{B})$ = trades off costs & benefits of choosing action a if unknown really were θ

takes values on real # line (also by u = letter)

ex. how fast should I drive on I5? to LA

optimal pick action to maximize

$$E_{(\theta \mid D, \mathcal{B})} U(a, \theta \mid \mathcal{B})$$

The hard part in applied statistical work:

The mapping from $\mathbb{P}$ to $M = \{ p(\theta|B), \ell(\theta|D,B) \}$

or $M_d = \{ p(\theta|B), \ell(\theta|D,B), (a|B), u(a,\theta|B) \}$

is rarely unique – typically there's a range of plausible sampling dist & a range of plausible priors – this is <u>model uncertainty</u>.

---

objectivity in uncertainty quantification

<u>Venn wanted:</u>

if scientists A, B have same data set D, then A, B draw same conclusions: to Venn this is <u>objectivity</u>.

However, A & B may have different amounts of info. external to D, & even if they have the same external info. they may reasonably disagree on appropriate assumptions & judgments about how to use D to quantify uncertainty about θ.

The kind of objectivity Van wanted is both impossible & undesirable.

Bayes with high-dimensional $\underset{\sim}{\theta}$

LHS

$p(\underset{\sim}{\theta} | D B) = c \cdot p(\underset{\sim}{\theta} | B) \cdot \ell(\underset{\sim}{\theta} | D B)$

$\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)$ p — positive integer (could be big)

---

to normalize LHS, need to compute

$$\int_{\theta_1} \int_{\theta_2} \cdots \int_{\theta_p} p(\theta | B) \cdot \ell(\theta | DB) \, d\theta_p \cdots d\theta_1$$

$p = 1{,}000{,}000$: fun? nooooo

Technical challenge | Approximating high-dimensional integrals

---

This was only solved in the 1980s & since. [end of pg. for week 2

Inference with multi-parameter problems

Ex. New case study: NB10 (7)

search string
draper clay google 2013

We'll demonstrate in R that the data vector

$y = (409, 400, \ldots, 404)$

does not look like it can be modeled well by a normal (Gaussian) sampling distribution, but in phase I of the analysis we're going to use the Gaussian sampling model anyway, to get practice with it.

Large-sample likelihood analysis

phase I sampling model

$$(Y_i \mid \mu, \sigma^2) \overset{IID}{\sim} N(\mu, \sigma^2) \quad (i = 1, \ldots, n)$$

$$\underset{\sim}{y} = (y_1, \ldots, y_n)$$

mean variance

vector of unknowns here is $\underset{\sim}{\theta} = (\mu, \sigma^2)$: dimension $p = 2$

Fisher's algorithm

step 0 marginal sampling distribution of Y_i

Y_i is continuous, so this is a density (pdf):

$$p(y_i \mid \mu, \sigma^2, B) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y_i-\mu)^2\right]$$

step 1 Joint sampling dist. is product of marginals:

$$p(\underset{\sim}{y} \mid \overbrace{\mu, \sigma^2}^{\underset{\sim}{\theta}}, B) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y_i-\mu)^2\right]$$

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

step 2 Likelihood function:

$$\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, B) = \ell(\mu, \sigma^2 \mid \underset{\sim}{y}, B)$$

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

Step 2.5 log likelihood function

$$\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}, \mathcal{B}) = \ell(\mu, \sigma \mid \underset{\sim}{y}, \mathcal{B})$$

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

side bar

$$\theta \sim \text{Beta}(\overset{\alpha}{\tfrac{1}{2}}, \overset{\beta}{\tfrac{1}{2}})$$

$$(Y_i \mid \theta, \mathcal{B}) \overset{IID}{\sim} \text{Bernoulli}(\theta)$$

$$(i = 1, \ldots, n)$$

$\theta \sim \text{Beta}(\alpha, \beta)$

$$\rightarrow \quad E(\theta) = \frac{\alpha}{\alpha + \beta}$$

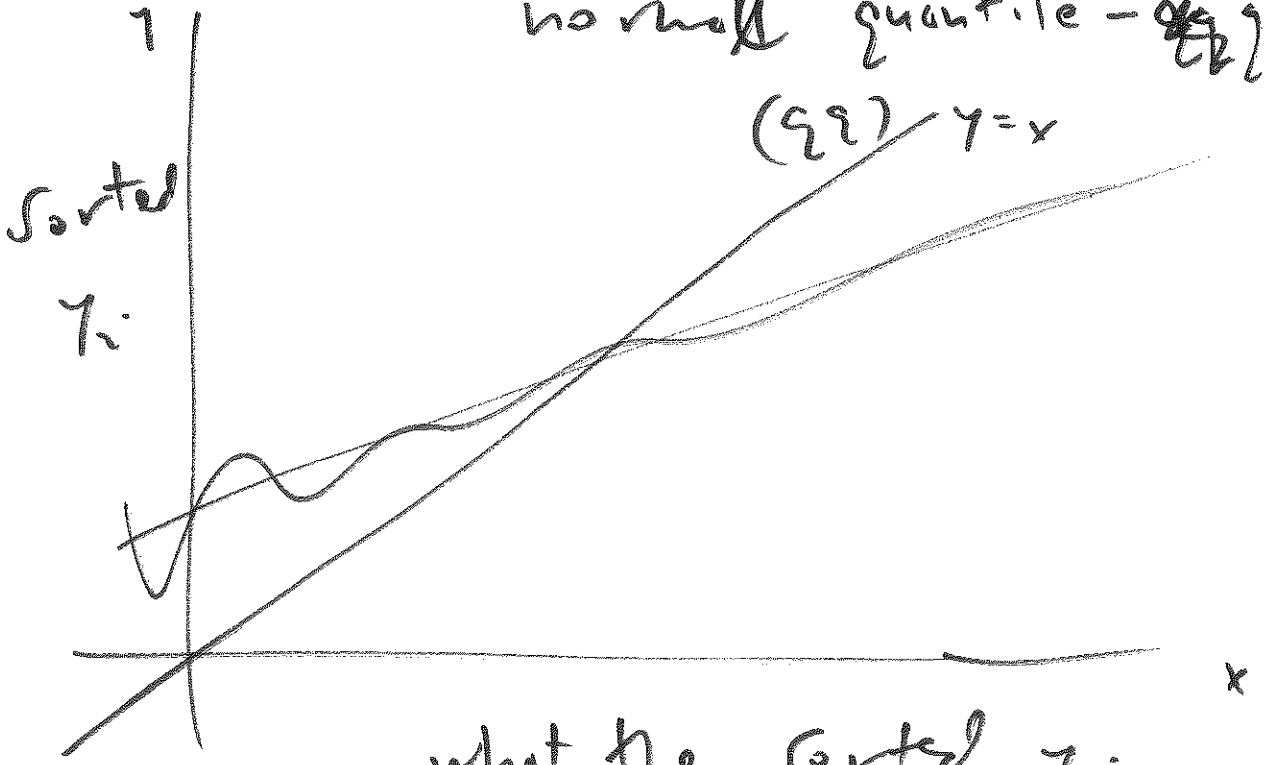
$$(\theta \mid s, \mathcal{B}) \sim \text{Beta}(\alpha + s, \beta + n - s)$$

$$E(\theta \mid \underset{\sim}{y}, \mathcal{B}) = \frac{s + \frac{1}{2}}{(\frac{1}{2} + s) + (n - s - \frac{1}{2})} = \frac{s + \frac{1}{2}}{n + 1}$$

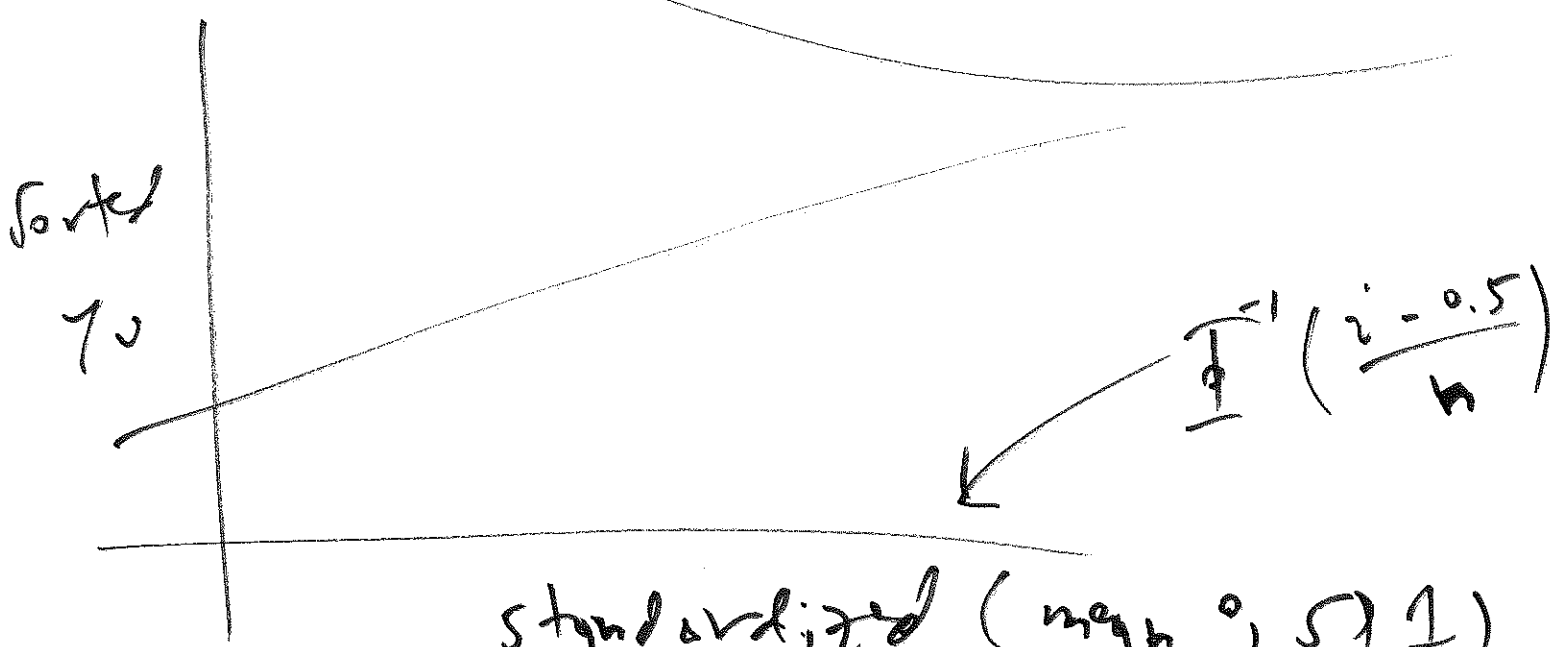
small-sample "nudge"

$\text{Beta}(\frac{1}{2}, \frac{1}{2})$: Jeffreys prior

normal quantile-quantile (10) plot



what the sorted y_i should look like if normality is an ok assumption



standardized (mean 0, sd 1) expected normal values