this time: Poisson inference

next time: vector θ

read: (same + ⊛)

drop-dead due date for Take-Home test 1 changed: now it's ① 11.59pm on Tue 20 Feb

⊛ read G ch. 4

A note on conditional probability

A, B, C propositions

(B and C)

$$p(A \mid B\,C) = \begin{cases} \dfrac{p(A\,B\,C)}{p(B\,C)} & \text{if } p(B\,C) > 0 \\ \underline{\text{undefined}} & \text{if } p(B\,C) = 0 \end{cases}$$

if either $p(B = 0)$ or $p(C) = 0$, $p(A \mid B\,C) =$ undefined

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \xrightarrow{\text{pre-math} \;\vdots\; \text{math}} (\theta, \mathcal{D}, \mathcal{B})$$

$\mathbb{P}$: problem

$\mathbb{Q}$: question(s)

$\mathbb{C}$: context

θ: unknown

$\mathcal{D}$: data

$\mathcal{B}$: background info.

$$M = \{ p(\underline{\theta \mid \mathcal{B}}),\ p(\underline{\mathcal{D} \mid \theta\,\mathcal{B}}) \}$$

← inference, prediction

← your statistical model

$M^* = \{ p(\theta|B), p(D|\theta B), (\mathcal{A}|B), u(a,\theta|B) \}$

inference, prediction, decision

---

thm 1: If can uniquely specify M, result is optimal inference & prediction

---

thm 2: If ____________ M^*,

____________,

____________ & decision

It would be great if $C + B$ always uniquely specifies $\{ p(\theta|B), p(D|\theta B), (\mathcal{A}|B), u(a,\theta|B) \}$,

but it's actually somewhat rare for unique choice of these 4 things to arise from $C + B$.

ex. AMI case study:

$$\left\{ \begin{array}{c} (\mathcal{Y}_i \mid \theta\, \mathcal{B}) \stackrel{IID}{\sim} \text{Bernoulli}(\theta) \\ (i = 1, \dots, n) \end{array} \right\}$$

this is uniquely determined by

$\mathcal{C}$: (de Finetti's thm): before

$D = (y_1, \dots, y_n)$ arrives, Your uncertainty

about the $\mathcal{Y}_i$ is <u>exchangeable</u>:

$\rightarrow \stackrel{IID}{\text{Bernoulli}}(\theta)$ but $p(\theta \mid \mathcal{B})$ was

not unique in this case study

In general we have uncertainty about $(\mathcal{M}, \mathcal{M}^*)$: <u>model</u> <u>uncertainty</u>.

length-of-stay case study | lecture notes part 3 pp. 77 →

MLE ($\lambda > 0$)

P ← poisson assumption

$(Y_i \mid \lambda B) \overset{IID}{\sim} \text{Poisson}(\lambda)$

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

① joint sampling dist:

and

$$p(Y_1 = y_1, \dots, Y_n = y_n \mid \lambda) = \prod_{i=1}^{n} p(Y_i = y_i \mid \lambda)$$

$$= \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} = \frac{\lambda^{s} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!} = *$$

$s = \sum_{i=1}^{n} y_i$

② $\ell(\lambda \mid y) = c \; *$

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

$$= \cancel{c}\, \lambda^{s} e^{-n\lambda}$$

$$* \quad \ell\ell(\lambda \mid y) = s \log \lambda - n\lambda$$

(5)

(3) $\frac{d}{d\lambda} \ell\ell(\lambda|y) = 0 = \frac{s}{\lambda} - n \to$

$\hat{\lambda}_{MLE} = \frac{s}{n} = \bar{y}$

$\frac{d^2}{d\lambda^2} \ell\ell(\lambda|y) = -\frac{s}{\lambda^2}$

$$\left[\frac{d^2}{d\lambda^2} \ell\ell(\lambda|y)\right]_{\lambda = \hat{\lambda}_{MLE}} = -\frac{s}{\left(\frac{s}{n}\right)^2} = \frac{-n^2}{s} = -n\left(\frac{n}{s}\right) = -\frac{n}{\bar{y}}$$

$< 0 \therefore$ local max.

(4) $\hat{I}(\hat{\lambda}_{MLE}) = -\left[\frac{d^2}{d\lambda^2} \ell\ell(\lambda|y)\right]_{\lambda = \hat{\lambda}_{MLE}}$

$= \frac{n}{\bar{y}} = \underline{O}(n)$ ✓

$\hat{V}_{RS}(\hat{\lambda}_{MLE}) = \frac{1}{\hat{I}(\hat{\lambda}_{MLE})} = \frac{\bar{y}}{n} = \underline{O}\left(\frac{1}{n}\right)$ ✓

$\widehat{SE}_{RS}(\hat{\lambda}_{MLE}) = ?$

$$\hat{SE}_{RS}(\hat{\lambda}_{MLE}) = \sqrt{\hat{V}_{RS}(\hat{\lambda}_{MLE})} = \frac{\sqrt{\bar{y}}}{\sqrt{n}}$$

$= O\left(\frac{1}{\sqrt{n}}\right)$ ✓

(n large) CLT

approx. 95% CI for

$$\lambda : \quad \hat{\lambda}_{MLE} \pm \underbrace{1.96}_{\Phi^{-1}(1-\frac{\alpha}{2})} \hat{SE}(\hat{\lambda}_{MLE})$$

$100(1-\alpha)\%$

$\boxed{\text{Bayes}}$

$$\left\{ \begin{array}{c} (\lambda) \sim p(\lambda) \\ (Y_i \mid \lambda) \stackrel{IID}{\sim} \text{Poisson}(\lambda) \\ (i = 1, \dots, n) \end{array} \right\}$$

$$\ell(\lambda \mid y) = c\, \lambda^{s} e^{-n\lambda} = \ell(\lambda \mid s)$$

$\boxed{s \text{ is suff. for } \lambda}$

conj. prior is $\Gamma(\alpha, \beta)$:

$$p(\lambda) = c\, \lambda^{\alpha - 1} e^{-\beta\lambda}$$

$$p(\lambda|s) = c\,\underbrace{\left[\lambda^{\alpha-1} e^{-\beta\lambda}\right]}_{\text{prior}}\underbrace{\left[\lambda^{s} e^{-n\lambda}\right]}_{\text{lik.}}$$ (7)

post.

$$= c\,\lambda^{(\alpha+s)-1} e^{-(\beta+n)\lambda}$$

$$= \Gamma(\alpha+s, \beta+n)$$

$$(\lambda \mid \underset{\text{prior}}{\text{conj.}}\, \mathcal{B}) \sim \Gamma(\alpha, \beta)$$

$$(Y_i \mid \lambda\, \mathcal{P}\, \mathcal{B}) \overset{\text{IID}}{\sim} \text{Poisson}(\lambda)$$

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

$$\rightarrow (\lambda \mid s\, \mathcal{P} \left(\underset{\text{prior}}{\text{conj}}\right) \mathcal{B}) \sim \Gamma(\alpha+s, \beta+n),$$

$$s = \sum_{i=1}^{n} y_i$$

Principle of stable

Estimation (Edwards, Savage, Lindeman)

Any prior that is close to flat ⑧

in the region in which the likelihood

is appreciable will have low

information content