this time: predictive diagnostics

next time: multivariate θ

read: (as before)

AMS 206 15 Feb 18

(2) typos in Take-Home Test (1)

1: 2(B)(iv) $\Gamma^{-1}\left(\frac{n-1}{2}, n \bar{y} \right)$ not n

and 2(B)(v)(b) likelihood σ should be 1,774 not 1,568

Lecture notes, part 2

informal Cox-Jaynes thm (de Finetti): under simple & reasonable axioms, the Bayesian story is <u>logically internally consistent</u> (coherent): this means that the basic rules of probability ($0 \le P(A) \le 1$, product rule for $\boxed{\text{and}}$, sum rule for $\boxed{\text{or}}$) can never be violated in the Bayesian paradigm

in contrast, there are frequentist examples that are clearly not logically internally consistent:

ex. $\sigma^2 \ge 0$ ← estimating a variance: $\underset{\text{unbiased}}{\hat{\sigma}_u^2} < 0$

However, nothing in pure Bayesian story requires you to pay attention to calibration: Q: how often do your Bayesian methods get the right answer?

This is a relative frequency (frequentist) question

Bayes: good for logical internal consistency

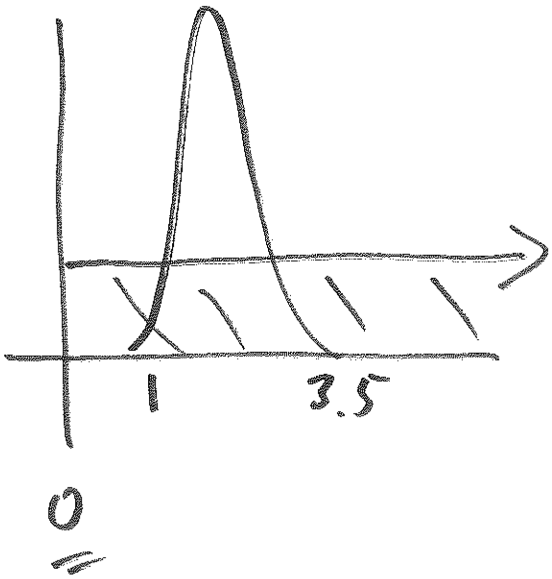
freq: bad ______

freq: good for calibration (logical external consistency)

pure Bayes: bad ______

Bayes-frequentist

goal: fusion: my: reason Bayesianly when formulating inferences, predictions & decisions, AND reason frequentistly when evaluating the calibration quality of the Bayesian answers (well-calibrated Bayesian)



proper

$\lambda \sim U(0, A)$ ⊛

$\underline{\underline{\lambda}} \sim U(0, \infty)$

improper prior (integrates to ∞)

⊛ choose A large enough to avoid inappropriate truncation of the likelihood density

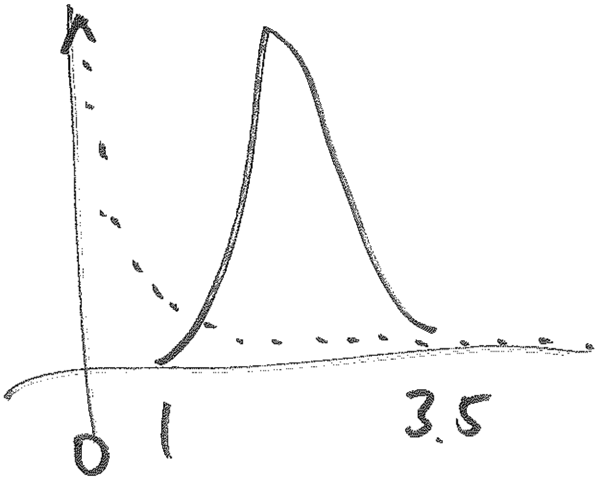
danger with improper priors: they may lead to improper posteriors (to be avoided)

Bayes with training wheels: always use proper priors conj. prior: $\lambda \sim \Gamma(\alpha, \beta)$

$E(\lambda) = \frac{\alpha}{\beta}$

$\Gamma(\epsilon, \epsilon)$ prior

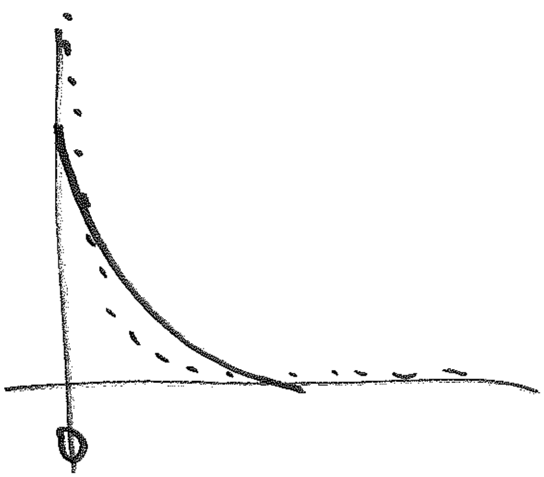
ϵ small positive # like 0.001



prior sample size ϵ

prior mean $\frac{\epsilon}{\epsilon} = 1$

$\Gamma(\epsilon, \epsilon)$ will not be (low info content)



if lik density has a lot of mass near 0

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

predictive distributions & why they're important

$p(y_{n+1} \mid y)$, post. pred. dist.

$p(y_{n+1} \mid y)$ ← hard to think about

$p(y_{n+1} \mid \lambda)$ easy ← sampling dist. for y_{n+1}

$(\lambda) \sim \Gamma(\alpha, \beta)$

$(Y_i \mid \lambda) \overset{iid}{\sim} P(\lambda)$

$$p(y_{n+1} \mid y) \overset{(LTP)}{=} \int_0^\infty p(y_{n+1}, \lambda \mid y)\, d\lambda$$

$$= \int_0^\infty \underbrace{\boxed{p(y_{n+1} \mid \lambda, y)}} \cdot \underbrace{\boxed{p(\lambda \mid y)}}_{\text{post. for } \lambda \text{ given } y} d\lambda$$

$$= \int_0^\infty \underbrace{p(y_{n+1} \mid \lambda)}_{\text{Poisson sampling dist. for } y_{n+1}} \cdot \underbrace{p(\lambda \mid y)}_{\text{post. for } \lambda \;\; \Gamma(\alpha^*, \beta^*)} d\lambda$$

why is prediction important

Prediction Principle: Good models make good predictions of new data; bad models make bad predictions; that's one important way we know if a model is good or bad.

simple 2-way Bayesian cross validation

outcome ↓ ← predictors →

$D =$ n individuals: $Y \quad X_1 \quad \cdots \quad X_k$

Y_{1914} ← $(k+1)$ →

proxy for new data

random partition

M modeling

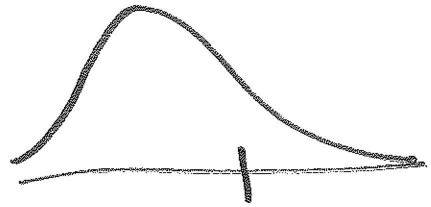
V validation

generate (how?) (later) an ensemble

$\mathcal{M} = \{ M_1, \dots, M_m \}$ of candidate models

& fit all these models to the data

in $\mathcal{M}$



y_{1914}

$p(y_{1914} \mid y_{\mathcal{M}}, M_3)$

$\uparrow$

a typical outcome in

$\Downarrow$

how numerically compare y_{1914} with its predictive dist. under model M_3?

set obs. i

aside:

$\boxed{y = (y_1, \dots, y_n)}$

$y_{-i} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$

omit one obs. at a time: (the jackknife)