Bayesian Statistics
Lecture 2

"PROBABILITY DOES NOT EXIST"

Subjectivists: De Finetti, Savage, Ramsey

Example: coin tossing

- We want to learn θ from Bern(θ)
- Specifying prior on $[0,1]$

1814 Laplace: uniform prior to express our ignorance. Beta(1,1)

<u>Conjugate prior</u>: given a likelihood $p(x|\theta)$, we choose a family of prior distributions such that the posterior (and also integrals such as the Bayes marginal) can be analytically obtained.

Ex contd: Coin is tossed t times, $z^t \sim \text{Bern}(\theta)$, n_1 ones, n_0 zeros.

$$p(z^t|\theta) = \theta^{n_1}(1-\theta)^{n_0}$$

$\Rightarrow$ posterior $\text{Beta}(1+n_1, 1+n_0)$

$\Rightarrow$ posterior mean: $E(\theta|z) = \frac{n_1+1}{n_1+n_0+2}$

(MCMC)

Exercise 1

$$p(\alpha,\beta|x) \propto \pi(\alpha,\beta)\, p(x|\alpha,\beta)$$

$$= \pi(\alpha)\pi(\beta)\, p(x|\alpha,\beta)$$

$$= (1-\alpha)^{b-1}\alpha^{a-1}(1-\beta)^{q-1}\beta^{p-1}$$

$$\alpha^r(1-\alpha)^{n-r}\left(\prod_{i=1}^{n-r}(1-\beta)^{x_i-1}\right)\beta^{n-r}$$

$$= \underline{\alpha^{r+a-1}(1-\alpha)^{b+n-r-1}}\,\beta^{n+p-r-1}$$

$$(1-\beta)^{\Sigma x_i+r+q-n-1}$$

α: $\text{Beta}(r+a, b+n-r) = \pi(\alpha|x)$

β : Beta$(n+p-r, \sum x_i + r + q - n)$

$= \pi(\beta|x)$

Exercise 2

$$p(x|\vartheta) = \prod_{i=1}^{3} p(x_i|\vartheta) \; P[x_4 > 100] \cdot P[x_5 > 100]$$

$$P[x_j > 100] = \int_{100}^{\infty} \vartheta \exp\{-\vartheta x\} d\cancel{x} = e^{-100\vartheta}$$

$$\rightarrow p(x|\vartheta) = \vartheta^3 \exp\{-449\vartheta\}$$

PDF exponential dist.

$\vartheta e^{-\vartheta x}$

prior

$$p(\vartheta) = Ga(2, 180) \propto \vartheta^{2-1} \exp\{-180\vartheta\}$$

posterior

$$p(\vartheta|x) \propto \vartheta^{2-1} \exp\{-180\vartheta\} \, \vartheta^3 \exp\{-449\vartheta\}$$

$$\propto \vartheta^{5-1} \exp\{-629\vartheta\} \rightarrow Ga(5, 629)$$

mean: $5/629$ Var: $5/629^2$

Posterior mode : $\arg\max_{\vartheta} \pi(\vartheta|x)$

I. J. Good : "There are 46656 varieties of Bayesians." (1971)

Bayesianism

Central to Bayesian Statistics is the interpretation of probability as degrees of belief.

1) Kolmogorov's (1933) axioms
2) Lets S be some statement, then we start with a prior probability $P_{old}(S)$.

Upon acquiring new evidence E, we transform our prior to generate a posterior by conditionalization on E.

$P_{new}(S) = P_{old}(S|E)$

→ Bayes' Rule.

The origin of Priors

Subjectivists (De Finetti, Savage, Ramsey)

They take probability to be the expression of personal opinion.

Priors: no constraints (other than coherence, i.e. Kolmogorov's axioms).

Objectivists (Jeffreys, Jaynes)

Prior probabilities should be rationally constrained.

- no principles exist that uniquely determine rational priors in all cases.

Pragmatists most statisticians.

Choose their priors for pragmatic reasons

- mathematical convenience
- computational "
- special effects.

- satisfying specific criteria

"Frequentist Bayes"

PRIORS: often exhibit both subjective and objective elements

Computation before 1990: conjugate priors

after 1990: MCMC - Markov Chain Monte Carlo

Exercise 3

$X^n := X_1, \dots, X_n \qquad p(\delta) \propto \delta^{-1}$

$$L(x_1, \dots, x_n \mid \theta = 2, \delta) = \prod_{i=1}^{n} 2\delta^2 x = 2^n \delta^{2n} \prod_{i=1}^{n} x_i$$

$$p(\delta \mid x_1, \dots, x_n) \propto \delta^{-1} \cdot \underbrace{2^n \prod_{i=1}^{n} x_i}_{C} \delta^{2n} \propto \underbrace{C\,\delta^{2n-1}}$$

$$\int C\,\delta^{2n-1}\,d\delta = 1 \qquad\qquad 0 < x < \delta^{-1} \to \delta < x_i^{-1} \to \delta < \min_i x_i^{-1} = t$$

$$C \underbrace{\int_0^t \delta^{2n-1}\,d\delta}_{} = 1$$

$$\frac{\delta^{2n}}{2n}\Big|_0^t = \frac{t^{2n}}{2n} \to C = \frac{2n}{t^{2n}} \to p(\delta \mid x_1, \dots, x_n) =$$

$$= \frac{2n\,\delta^{2n-1}}{t^{2n}} \qquad 0 < \delta < t$$

Jeffreys' prior

uninformative priors

Example

$X \sim \text{Bern}(\theta)$

$\rightarrow \pi(\theta) = 1 \qquad 0 \leq \theta \leq 1$

log odds ratio

$$e = \log \frac{\theta}{1-\theta}$$

$$\rightarrow \pi(e) = \frac{e^{e}}{(1+e^{e})^2}$$

'Flat priors' are not parametrization-invariant.

Jeffreys: par.-invariant prior

$$\pi(\theta) \propto \sqrt{|I(\theta)|}$$

$$= \sqrt{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta} \log f(x;\theta)\right)^2 \Big| \theta\right]}$$

if $f(x;\theta)$ is twice differentiable wrt θ and under certain regularity conditions.

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(x;\theta) \Big| \theta\right]$$

Why?

$$\frac{\partial^2}{\partial\theta^2} \log f(\cancel{x};\theta) = \frac{\frac{\partial^2}{\partial\theta^2} f(x;\theta)}{f(x;\theta)} - \left(\frac{\frac{\partial}{\partial\theta} f(x;\theta)}{f(x;\theta)}\right)^2$$

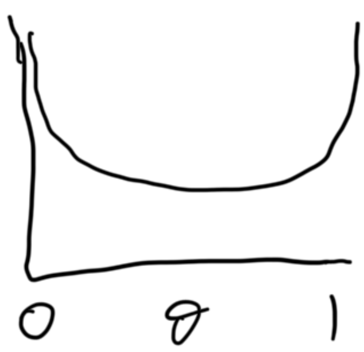
$f(X;\vartheta)$

$$= \frac{\frac{\partial^2}{\partial\vartheta^2} f(X;\vartheta)}{f(X;\vartheta)} - \left(\frac{\partial}{\partial\vartheta}\log f(X;\vartheta)\right)^2$$

and

$$\mathbb{E}\left[\frac{\frac{\partial^2}{\partial\vartheta^2} f(X;\vartheta)}{f(X;\vartheta)} \,\middle|\, \vartheta\right] = \frac{\partial^2}{\partial\vartheta^2}\int_{\mathbb{R}} f(x;\vartheta)\,dx = 0$$

Note: ~~th~~ only holds for 1-dimension.



Jeffreys' prior for Beta-Bernoulli model

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

par. invariant $\rightarrow$ this does not mean that it is 'uninformative'

$\Rightarrow$ A general principle for making non-informative priors turns out not to exist.

Exercise 4

$$p(x|\delta) = \vartheta\,\delta^{\vartheta}\,x$$

$$\log p(x|\delta) = \log\vartheta + \vartheta\log\delta + \log x$$

$$\frac{d\log p}{d\delta} = \frac{\vartheta}{\delta}$$

$$\frac{d^2 \log p}{d\delta^2} = \frac{-\vartheta}{\delta^2}$$

$$I(\delta) = -\int p(x|\delta) \frac{d^2 \log p}{d\delta^2} dx$$

$$= \frac{\vartheta}{\delta^2} \underbrace{\int p(x|\delta)\, dx}_{=1}$$

$$= \frac{\vartheta}{\delta^2}$$

$\Rightarrow$ Jeffreys prior $\propto \sqrt{|I(\delta)|}$

$$\propto \frac{1}{\delta}$$

(since ϑ is known)

Exercise 5

$$p(\theta) \propto \sqrt{|I(\theta)|}$$

$$f(x|n,\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$\log f(x|n,\theta) = \log\binom{n}{x} + x\log\theta + (n-x)\log(1-\theta)$$

$$\frac{d}{d\theta} f(x|\theta) = \frac{x}{\theta} - (n-x)\frac{1}{1-\theta}$$

d^2

$$\frac{d^2}{d\theta^2} f(x|\theta) = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}$$