STAT 431 — Applied Bayesian Analysis — Course Notes

Improper Priors

Fall 2022

Improper Densities

An **improper density** $f(\boldsymbol{x})$ satisfies

- $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$
- $\sum_{\boldsymbol{x}} f(\boldsymbol{x}) \ = \ \infty$ (discrete) or
  $\int f(\boldsymbol{x})\, d\boldsymbol{x} \ = \ \infty$ (continuous)
  (i.e. it cannot be *normalized* to a **proper** density)

When an improper density is used as a ("noninformative") prior density, it is called an **improper prior**.

Eg: Population proportion θ (binomial model)

$$\pi(\theta) \;\; = \;\; \frac{1}{\theta(1-\theta)} \qquad 0 < \theta < 1$$

(like a beta density with "$\alpha = 0$" and "$\beta = 0$")

Improper because

$$\int_0^1 \pi(\theta)\, d\theta \;\; = \;\; \infty$$

[Draw density ...]

Warning: Improper priors may lead to improper posteriors!

Eg: (continued)

If $Y \sim \mathrm{Binomial}(n, \theta)$ has prior

$$\pi(\theta) = \frac{1}{\theta(1-\theta)} \qquad 0 < \theta < 1$$

and either $y = 0$ or $y = n$, then can show that $p(\theta \mid y)$ is improper!

Alternative: Use “vague” priors — proper, but close to improper.

Eg: Use $\mathrm{Beta}(\alpha, \beta)$ with α and β “near” zero.

Example: Mean-Only Normal Sample

$$Y_1, \ldots, Y_n \mid \mu \;\sim\; iid\ \text{Normal}\big(\mu, \sigma^2\big) \qquad \sigma^2 \text{ known}$$

Recall likelihood

$$\propto \; e^{-\frac{n}{2\sigma^2}\,(\mu - \bar{y})^2}$$

Try improper “flat” prior

$$\pi(\mu) \;\propto\; 1 \qquad -\infty < \mu < \infty$$

We must check that the posterior is proper ...

$$p(\mu \mid \boldsymbol{y}) \quad \propto \quad f(\boldsymbol{y} \mid \mu) \underbrace{\pi(\mu)}_{\propto 1} \quad \propto \quad e^{-\frac{n}{2\sigma^2}(\mu-\bar{y})^2}$$

Recognize as the kernel of $\mathrm{Normal}(\bar{y},\, \sigma^2/n)$ (why?), so the posterior is indeed proper:

$$\mu \mid \boldsymbol{y} \;\sim\; \mathrm{Normal}\big(\bar{y},\, \sigma^2/n\big)$$

Note: The posterior mean is $\bar{y}$ and the posterior standard deviation is the (frequentist) standard error of $\bar{y}$.

Can show that, under this improper prior,

- credible intervals for μ are the same as confidence intervals

- the posterior probability of a *one-sided* H_0 is the same as a p-value

  (Not true for the two-sided case.)

- this posterior is the limit as $\tau_0^2 \to 0$ (equiv. $m \to 0$) of the posterior when the $\mathrm{Normal}(\mu_0,\, 1/\tau_0^2)$ prior is used

Notation: For a flat prior, write, e.g.

$$\mu \sim 1\,d\mu$$

Eg: Jevons's Coin Data

Posterior under flat prior:

$$\mu \mid \boldsymbol{y} \quad \sim \quad \mathrm{Normal}(\mu_1 = \bar{y} = 7.8730,\ \sigma_1^2 = \sigma^2/n \approx 0.0001194)$$

(so $\sigma_1 \approx 0.01093$)

[Draw density w/ probability limits ...]

So an approx. 95% credible interval is

$$\mu_1 \pm 1.96\,\sigma_1 \quad \approx \quad (7.8516,\ 7.8944)$$

It excludes all values meeting the min. legal weight (7.9379).

Indeed, for $\qquad H_0 :\ \mu \geq 7.9379 \qquad$ we find

$$\mathrm{Prob}(H_0 \mid \boldsymbol{y}) \;=\; 1 - \Phi\left(\frac{7.9379 - \mu_1}{\sigma_1}\right) \;\approx\; 10^{-9}$$

(same as a p-value, in this case)

Now consider randomly selecting another coin of the same kind (minted before 1830). Its (random) weight will be Y^*.

Using the flat prior and previous formulas for the posterior predictive distribution,

$$Y^* \mid \boldsymbol{y} \quad \sim \quad \mathrm{Normal}\big(7.8730,\, 0.0001194 + (0.05353)^2\big)$$

The posterior predictive standard deviation works out to be about 0.05463.

The posterior predictive prob. that *this coin* is of legal weight:

$$\begin{aligned}\mathrm{Prob}(Y^* \geq 7.9379 \mid \boldsymbol{y}) \quad &\approx \quad 1 \,-\, \Phi\!\left(\frac{7.9379 - 7.8730}{0.05463}\right) \\ &\approx \quad 0.1174\end{aligned}$$

Remarks

- An improper prior may be hard to interpret subjectively, since it can't produce probabilities for $\boldsymbol{\theta}$.

  Perhaps interpret as equivalent to a series of proper priors whose posteriors converge to the posterior produced by the improper prior.

  (e.g. as $\sigma_0^2 \to \infty$, $\mathrm{Normal}(0, \sigma_0^2)$ priors produce posteriors converging to the flat prior's posterior)

- Representing the condition of "no prior information" often requires considering improper priors ("noninformative" priors — next).