Chapter 2.3

Objective priors

Informative versus uninformative priors

- In some cases informative priors are available
- Potential sources include: literature reviews; pilot studies; expert opinions; etc
- **Prior eliciation** is the process of converting expert information to prior distribution
- For example, the expert might not comprehend an inverse gamma PDF, but if they give you an estimate and a spread you can back out a and b.
- There are tools for this, such as `https://jeremy-oakley.shinyapps.io/SHELF-single/`

Informative versus uninformative priors

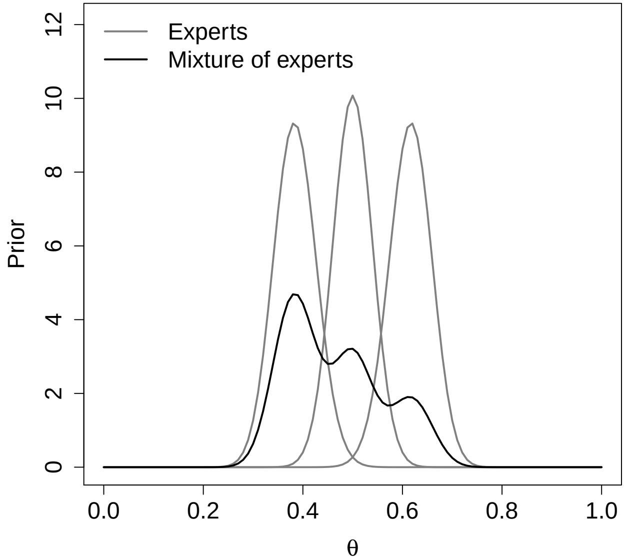
- Strong priors for the parameters of interest can be hard to defend
- Strong priors for nuisance parameters are more common
- For example, say you are doing a Bayesian t-test to study the mean μ, you might use an informative prior for the nuisance parameter σ^2
- Any time informative priors are used you should conduct a **sensitivity analysis**
- That is, compare the posterior for several priors

Prior elicitation

- Prior information can come from a variety of sources
- Say that source j (e.g., journal article, expert, pilot study) suggests prior $\pi_j(\theta)$
- A **mixture of experts** prior combines the J sources into a single prior
- Say source j is given weight $w_j > 0$ with $\sum_{j=1}^{J} w_j = 1$
- The mixture of experts prior is

$$\pi(\theta) = \sum_{j=1}^{J} w_j \pi_j(\theta)$$

Mixture of experts prior ($w_j \in \{0.2, 0.3, 0.5\}$)



Prior elicitation

- Another powerful idea is to elicit prior information using real-life scenarios
- "Do you expect more events on hot days or cold days?"
- "Would you expect more events on a rainy day August or a sunny day in December?"
- Later you can frame these questions using the parameters in your model and back out the prior that best matches the expert responses

Informative versus uninformative priors

- In most cases prior information is not available and so uninformative priors are used
- Other names: “vague”, “weak”, “flat”, “diffuse”, etc.
- These all refer to priors with large variance
- Examples: $\theta \sim \text{Uniform}(0, 1)$ or $\mu \sim \text{Normal}(0, 1000^2)$
- Uninformative priors can be conjugate or not conjugate
- The idea is that the likelihood overwhelms the prior
- You should verify this with a sensitivity analysis

Improper priors

- Extreme case: $\mu \sim \text{Normal}(0, \tau^2)$ and we set $\tau = \infty$
- A "prior" that doesn't integrate to one is called an **improper**
- Example: $\pi(\mu) = 1$ for all $\mu \in \mathcal{R}$
- It's OK to use an improper prior so long as you verify that the posterior integrates to one
- For example, in linear regression an improper prior can be used for the slopes as long as the number of observations exceeds the number of covariates and there are no redundant predictors
- Subjective Bayesian interpretation is tricky

Subjective versus objective priors

- A subjective Bayesian picks a prior that corresponds to their current state of knowledge before collecting data
- Of course, if the reader does not share this prior then they might not accept the analysis, and so uninformative priors and a sensitivity analysis are common
- An **objective analysis** is one that requires no subjective decisions by the analyst
- Subjective decisions include picking the likelihood, treatment of outliers, transformations, ... and prior specification
- A completely objective analysis may be feasible in tightly controlled experiments, but is impossible in many analyses

Objective Bayes

- An objective Bayesian attempts to replace the subjective choice of prior with an algorithm that determines the prior
- There are many approaches: Jeffreys, reference, probability matching, maximum entropy, empirical Bayes, penalized complexity, etc.
- Jeffreys priors are the most common and we'll study these in some detail
- The others we will mention superficially
- Many of these priors are **improper** and so you have to check that the posterior is proper

Objective Bayes

- One subjective decision is the parameterization to use
- For example, should we select $\sigma \sim \text{Uniform}(0, 10)$ or $\sigma^2 \sim \text{Uniform}(0, 100)$?
- These are quite different; $\sigma \sim \text{Uniform} \to p(\sigma^2) \propto 1/\sigma$
- Any objective prior must give the same results, e.g., posterior mean for σ, for any parameterization
- The Jeffreys prior is invariant to transformations

Jeffreys prior

- The Jeffreys prior for parameter θ is $p(\theta) = \sqrt{I(\theta)}$
- The Fisher information is $I(\theta) = -\mathrm{E}_{Y|\theta}\left[\frac{d^2}{d\theta^2}\log p(Y|\theta)\right]$
- Once you have specified the likelihood the Jeffreys prior is determined with no additional input
- Therefore you do not have to make a subjective decision about the prior (other than "subjective" decision to use a Jeffreys prior)

Examples of Jeffreys priors

Likelihood	Jeffreys prior
$Y \sim \text{Binomial}(n, \theta)$	$\theta \sim \text{Beta}(1/2, 1/2)$
$Y \sim \text{N}(\mu, 1)$	$p(\mu) \propto 1$
$Y \sim \text{N}(0, \sigma^2)$	$p(\sigma) \propto 1/\sigma$

- See "Jeffreys" in the online derivations
- The first example is derived on the next slides

Jeffreys priors for the model $Y \sim \text{Binomial}(n, \theta)$

- The log likelihood is

$$l(\theta) = \log[f(Y|\theta)] = \log \binom{n}{Y} + Y \log(\theta) + (n - Y) \log(1 - \theta)$$

- The first derivative is

$$l'(\theta) = \frac{Y}{\theta} - \frac{n - Y}{1 - \theta}$$

- The second derivative is

$$l''(\theta) = -\frac{Y}{\theta^2} - \frac{n - Y}{(1 - \theta)^2}$$

Jeffreys priors for the model $Y \sim \text{Binomial}(n, \theta)$

- Recalling that $E(Y) = n\theta$,

$$I(\theta) = -E\left[l''(\theta)\right] = \frac{n\theta}{\theta^2} - \frac{n - n\theta}{(1-\theta)^2} = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)} = n\theta^{-1}(1-\theta)^{-1}$$

- Finally, the prior is

$$\pi(\theta) \propto I(\theta)^{1/2} \propto \theta^{-1/2}(1-\theta)^{-1/2} \propto \theta^{1/2-1}(1-\theta)^{1/2-1}$$

- This the kernel of a Beta(1/2,1/2) PDF

Reference priors

- These priors try to formally be “uninformative”
- As with Jeffreys they are objective
- They are defined as the prior that gives the maximum expected “distance” between the prior and posterior
- For univariate models they give Jeffreys priors
- For multivariate models they are different
- Reference priors are harder to compute than Jeffreys

Probability matching priors (PMP)

- The PMP is designed so that posterior credible intervals have correct frequentist coverage
- There are only a few cases where this can be done exactly
- For example, if $Y_i|\mu \sim \text{Normal}(\mu, 1)$, the PMP is $p(\mu) = 1$
- The posterior is $\mu|\mathbf{Y} \sim \text{Normal}(\bar{Y}, 1/n)$
- In this case, credible sets have correct frequentist coverage
- Approximations are available for medium-complexity cases

Empirical Bayes

- Empirical Bayes is also objective
- Here you pick the priors based on the data
- Example: maybe σ^2 has prior mean s^2
- Example: σ^2 is fixed at s^2 in the Bayesian analysis of μ
- More formally, you can use marginal maximum likelihood to fix nuisance parameters
- The analysis then proceeds as a usual Bayesian analysis
- Often criticized for “using the data twice”

Penalized complexity priors

- A PCP prior begins with a simple base model, e.g., linear regression with all the slopes equal zero
- The full model, e.g., regression with non-zero slopes, is shrunk towards to the base model
- The “distance” of the full model to the base model has exponential prior to penalized the more complex model from deviating from the base model
- Requires picking the parameter in the exponential prior and setting priors for the parameters in the base model
- So technically this is not purely objective

Maximum entropy priors

- In a maximum entropy prior you fix a few quantities of the prior distribution, e.g., $E(\theta) = 0.5$

- The prior is taken to be the distribution with the maximum entropy of these that satisfy the known constraints

- "Entropy" is a measure of uncertainty, e.g., the entropy of the PMF $f(x)$ is
$$-\sum_{x \in \mathcal{S}} f(x) \log[f(x)]$$

- If θ has support $\mathcal{R}$ and the mean and variance are known, the maximum entropy prior is Gaussian

- Not purely objective because you have to set the constraints