Summarizing the posterior

1. Laplace approximation
   - fit a Gaussian to the mode of the posterior

Derivation

We have $p(\theta|y)$. If $p(\theta|y)$ continuous then the post. mode $\hat{\theta}$ is given by $\frac{d}{d\theta} p(\theta|y) = 0$ (if locally concave).

To report the reliability of $\hat{\theta}$, we approximate $p(\theta|y)$ locally using Taylor series expansion about $\hat{\theta}$:

Let $L(\theta) = \log p(\theta|y)$

$$L(\theta) \approx \underbrace{L(\hat{\theta})}_{\substack{\text{constant} \\ \text{"A"}}} + \underbrace{\frac{d}{d\theta} L(\theta)\Big|_{\hat{\theta}} (\theta - \hat{\theta})}_{0 \text{ by def'n.}} + \frac{\frac{d^2}{d\theta^2} L(\theta)\Big|_{\hat{\theta}} (\theta - \hat{\theta})^2}{2!}$$

so

$$p(\theta|y) = \exp\{L(\theta)\} \approx \underbrace{\exp\{A\}}_{C} \exp\left\{\frac{1}{2} \frac{d^2 L}{d\theta^2}\Big|_{\hat{\theta}} (\theta - \hat{\theta})^2\right\}$$

Notice this is a Gaussian w/ mean $\hat{\theta}$ & $\sigma^2 = \frac{-1}{L''(\hat{\theta})}$ where $L''(\hat{\theta})$ is necessarily $-$ due to locally concave req: $L''(\hat{\theta} \pm \varepsilon) < 0$.

Thus Laplace approximation approximates posterior w/

$$N\left(\hat{\theta}, \frac{-1}{L''(\hat{\theta})}\right) \quad \text{where } \hat{\theta} = \text{MAP estimate.}$$

Example: Binomial coin flip. w/ $p(\theta) = 1$ prior.

$$p(\theta|y) \propto \theta^y (1-\theta)^{n-y}$$

$$L(\theta) = C + y \log\theta + (n-y)\log(1-\theta)$$

$$L'(\theta) = \frac{y}{\theta} - \frac{n-y}{1-\theta}$$

$$L''(\theta) = \frac{-y}{\theta^2} - \frac{n-y}{(1-\theta)^2}$$

$p(\theta|y) \approx N\left(\hat{\theta}, \frac{\hat{\theta}(1-\hat{\theta})}{N}\right)$

using $y = \hat{\theta} \cdot n$

2. Confidence regions (Bayesian)

After observing the data $Y=y$:

$p(l(y) < \theta < u(y) \mid y) = 1-\alpha$

$l(y)$: lower bound, $u(y)$: upper

This **is** a probability statement about θ.

$\theta \in (l(y), u(y))$ with $(1-\alpha)$ probability

In practice, we'll use a quantile based interval.

To construct a $100\times(1-\alpha)\%$ quantile-based interval, chop off the top & bottom $\frac{\alpha}{2}\%$ of the posterior.

Example: beta-binomial posterior

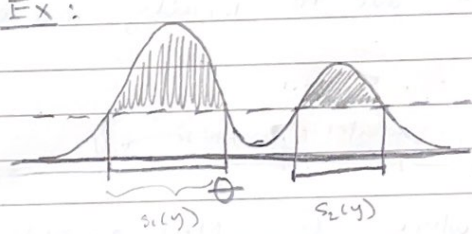
$p(\theta \mid y) \sim beta(y+a, n-y+b)$

qbeta(c(0.025, 0.975), a+y, n-y+b)

3. High posterior density

- same interpretation: $p(\theta \in s(y) \mid y) = 1-\alpha$
  - * but $s(y)$ need not be an interval

* Ex:



$s(y) = s_1(y) \cup s_2(y)$

Root finding problem: $\int_{-\infty}^{+\infty} p(\theta \mid x)\, \mathbb{1}_{\{p(\theta) \geq c\}}\, d\theta - (1-\alpha) = 0$

we can choose diff't c and see sometimes >0 sometimes <0.

Note

It's good practice to report both a point estimate
ex: posterior mean
posterior mode

together w/ a measure of reliability: CI, HPD region, $\hat{\sigma}$ from Laplace approx.

Exponential families

Ex:
- binomial (w/ fixed n.) & Poisson models
- beta & Gamma
- normal

& more

A single parameter exponential family model is any w/ density

$$p(y|\phi) = h(y)\, c(\phi) \exp\{\phi\, t(y)\}$$

ϕ: parameter

$t(y) :=$ sufficient statistic

"sufficient" because no other function of the data can provide add'l info about θ.

Notice $\underbrace{L(\phi)}_{\log p(y|\phi)} = \log h(y) + \log c(\phi) + \phi t(y)$

so when finding, e.g. MLE or MAP the constant $h(y)$ is irrelevant.

(†) Suppose $p(\phi|n_0, t_0) = K(n_0, t_0)\, c(\phi)^{n_0} \exp\{n_0 t_0 \phi\}$

Let $y = \{y_1, \dots, y_n\}$

then
$$p(\phi|y) \propto p(y|\phi)p(\phi)$$
$$\propto \underbrace{\prod_{i=1}^{n} c(\phi) \exp\{\phi t(y)\}}_{\text{cond'l iid}} \cdot c(\phi)^{n_0} \exp\{n_0 t_0 \phi\}$$

$$\propto c(\phi)^{n+n_0} \exp\{\phi \cdot (\sum_{i=1}^{n} t(y_i) + n_0 t_0)\}$$
$$\propto p(\phi \mid n+n_0, \frac{\sum t(y_i) + n_0 t_0}{n_0+n}) \quad \text{by} \quad (†)$$

n_0: prior sample size

t_0: prior mean

exp family example:

$Y \sim \text{binomial}(\theta)$ single y; no const. $\binom{n}{y}$

$p(y|\theta) = \theta^y (1-\theta)^{1-y}$ Need to look like $h(y)\,c(\phi)\,e^{\phi t(y)}$

$= \left(\frac{\theta}{1-\theta}\right)^y (1-\theta)$

no exp{ } so.. I need to exponentiate:

$\exp\{\underbrace{y\log\left(\frac{\theta}{1-\theta}\right)} + \log(1-\theta)\}$

$\exp\{t(y)\phi\} \cdot \underbrace{\exp\{\log(1-\theta)\}}$ Don't need another exp{ }

$t(y) := y$ $\exp\{\log \underbrace{\frac{1+e^\phi}{1+e^\phi} - \frac{e^\phi}{1+e^\phi}}$

$\phi := \log\frac{\theta}{1-\theta}$

$\star \Rightarrow e^\phi = \frac{\theta}{1-\theta}$ $\star \log\frac{1}{1+e^\phi}$

$\Rightarrow e^\phi - e^\phi\theta = \theta$ $(1+e^\phi)^{-1}$

$\Rightarrow e^\phi = \theta(1+e^\phi)$

$\star \Rightarrow \theta = \frac{e^\phi}{1+e^\phi}$

$c(\phi) = (1+e^\phi)^{-1}$

Now we know conj. prior

$p(\phi|n_0,t_0) \propto (1+e^\phi)^{-n_0} e^{n_0 t_0 \phi}$