Summarising posterior inference

- If feasible, the most informative is to report the whole posterior distribution $p(\theta|y)$.
For the "genetic status" example, we reported
$p(\theta=1|y)$ (and $p(\theta=0|y)$ implicitly), i.e. the whole distribution
- For single-parameter models, a posterior density (or histogram) plot is informative
- However, in many situations it is practical to report some sort of posterior estimates accompanied by posterior uncertainty

"Bayes estimate" (Material loosely based on NLH exercise 2)

Consider a general Bayesian model
- Sampling distribution $y \sim p(y|\theta)$, $y \in \mathcal{Y}$
- Prior distribution $\theta \sim p_0(\theta)$, $\theta \in \Theta$

sep 8-13:15

- A statistical decision function $\hat{a}$, is a function of the data, i.e. $\hat{a}(y)$
  - The decision might be a point estimate of θ (which we will consider), a Bayesian probability interval for θ, a specific action decision (Yes/No), etc.
- In order to find the best $\hat{a}(y)$, we need something called a loss function $L(\theta, a)$
  - → Describes the "loss" or "cost" associated with choosing action a, as a function of θ
  - For example: Wish to obtain a point estimate a of θ, then a frequently used loss function is the squared error loss $L(\theta, a) = (a-\theta)^2$
- We then have the risk function as a function of the parameter, which is the expected loss

$$R(a,\theta) = E\,L(\theta,a) = \int L(\theta, a(y)) \cdot p(y|\theta)\, dy$$ ← a random function because θ is random

sep 8-13:23

Taking the expectation of $R(a,\theta)$ with respect to the prior distribution of θ $(p_\theta(\theta))$, we get the Bayes risk:

$$BR(a, p_\theta) = E[R(a,\theta)] = \int R(a,\theta)\cdot p_\theta(\theta)\,d\theta$$

The Bayes solution (or Bayes estimate if a is a point estimate) is then the $\hat{a}$, among all possible decisions $\hat{a}$, that minimizes the risk, i.e.

$$\hat{a} = \underset{a}{\operatorname{argmin}}\; BR(a, p_\theta)$$

Given data y, we have that

$$\hat{a}(y) = \underset{a}{\operatorname{argmin}}\; E[L(\theta,a)\mid y] = \underset{a}{\operatorname{argmin}} \int_\theta L(\theta,a)\,p(\theta\mid y)\,d\theta$$

(Expectation of $L(\theta,a)$ taken w.r.t. the posterior distribution of θ given y)

sep 8-13:33

Proof

$$BR(a,p_\theta) = \int_\Theta R(a,\theta)\,p_\theta(\theta)\,d\theta = \int_\Theta \int_Y L(\theta,a(y))\,p(y\mid\theta)\,dy\;p_\theta(\theta)\,d\theta$$

$$= \int_\Theta \int_Y L(\theta, a(y))\,\underbrace{p(y\mid\theta)\,p_\theta(\theta)}_{p(\theta\mid y)\cdot p(y)}\,dy\,d\theta$$

$$= \int_Y \underbrace{\int_\Theta L(\theta, a(y))\cdot p(\theta\mid y)\,d\theta}_{E[L(\theta,a(y))\mid y]}\;p(y)\,dy = \int_Y E[L(\theta,a(y))\mid y]\,p(y)\,dy$$

This is minimized by, for a given y, minimizing $E[L(\theta,a(y))\mid y]$

sep 8-13:41

Ex: For finding the Bayes estimate $\hat{\theta}$ of θ with loss function $L(\theta, a) = (a-\theta)^2$, we must find the value of $\hat{\theta}$ that minimizes

$$E[L(\theta,\hat{\theta}) \mid y] = E[(\hat{\theta}-\theta)^2 \mid y] = \int_\Theta (\hat{\theta}-\theta)^2 \, p(\theta \mid y)\, d\theta$$

Differentiating

$$\frac{d}{d\hat{\theta}} \int_\Theta (\hat{\theta}-\theta)^2 p(\theta \mid y)\, d\theta = 2\int_\Theta (\hat{\theta}-\theta)\, p(\theta \mid y)\, d\theta$$

$$= 0 \quad \text{for} \quad \hat{\theta} \underbrace{\int_\Theta p(\theta \mid y)\, d\theta}_{=1} = \int_\Theta \theta\, p(\theta \mid y)\, d\theta$$

$$\hat{\theta} = E[\theta \mid y]$$

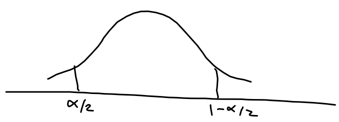
For $L(\theta, a) = |a-\theta|$, the posterior median minimizes the posterior expected loss.

sep 8-13:48

Posterior quantiles and intervals

- Posterior uncertainty is often reported as a $100\cdot(1-\alpha)\%$ central posterior interval, which is equivalent to reporting the $\alpha/2$ and $1-\alpha/2$ quantiles
- An alternative is the "$100\cdot(1-\alpha)\%$ highest posterior density region", the set of values that (1) contains $100(1-\alpha)\%$ of the posterior probability AND (2) the density within the region is never lower than that outside

Symmetric, unimodal:



Other examples, see for example Fig. 2.2

sep 8-13:56