OVERVIEW:
RANDOM SAMPLING MODELS AND INFERENCE

RANDOM SAMPLE:

- random variable x (univariate, for now)
- population distribution, density function $p(x|\Theta)$
- parameter Θ (maybe several scalar parameters)
- repeat sampling: $x_1, \ldots, x_n$, fixed sample size n
- independent draws: random sample $X = \{x_1, \ldots, x_n\}$
- joint density

$$p(X|\Theta) = \prod_{i=1}^{n} p(x_i|\Theta)$$

- defines sampling model: use to infer Θ
- statistical issues: sampling assumptions, model choice, ...

EXAMPLE 1: Diagnosis

- $n = 1$, $X = x_1 = x \in \{0, 1\}$, $\Theta = \theta \in \{0, 1\}$
- $p(X|\Theta) = \theta^x (1 - \theta)^{1-x}$

EXAMPLE 2: Bernoulli trials

- $x_i \in \{0, 1\}$ (failure/success), $\Theta = \theta$, $0 < \theta < 1$
- $p(X|\Theta) = \prod_{i=1}^{n} \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^y (1 - \theta)^{n-y}$

  where $y = \sum_{i=1}^{n} x_i$ = number of successes

- Compare Binimial model: $p(y|\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$
  - different sampling model: same likelihood function for θ
  - example of SUFFICIENT STATISTIC
  - principle of sufficiency

EXAMPLE 3: Binomial 2-Sample Model (VA hospitals example)

- $n = 2$, $y_i \sim Bin(n_i, \theta_i)$ for $i = 1, 2$
- $\Theta = \{\theta_1, \theta_2\}$
- $p(X|\Theta) = p(y_1, y_2|\theta_1, \theta_2) = p(y_1|\theta_1)p(y_2|\theta_2)$

Multiparameter model, Conditional independence structure

BAYESIAN INFERENCE on θ

- Prior density $p(\Theta)$
- Posterior density $p(\Theta|X)$
  - Bayes's theorem: Posterior $\propto$ Prior $\times$ Likelihood
  - Proportional form
  - (prior) predictive density $p(X) = \int p(X|\Theta)p(\Theta)d\Theta$
- Summarise posterior: mean, mode(s), posterior intervals, simulations, ...

Ranges of priors, sensitivity analyses

LIKELIHOOD AND MAXIMUM LIKELIHOOD ESTIMATES

$$p(\Theta|X) = kp(\Theta)p(X|\Theta)$$

"Flat" prior $p(\Theta) \propto$ constant ("uninformative" about Θ) implies

$$p(\Theta|X) = cp(X|\Theta)$$

Posterior PDF = Normalised likelihood function

Posterior mode = Maximum likelihood estimate of Θ
(may be multimodal)

Common POINT ESTIMATE of Θ

Also interesting when prior is *relatively* diffuse, spread compared to likelihood

LIKELIHOOD AND SUFFICIENCY

Suppose likelihood factorises as $p(\Theta|X) = a(X)l(y(X), \Theta)$

where

- $y(X)$ is a function of the data (= a STATISTIC)
- $a(\cdot)$ does not involve Θ

THEN

- $y(X)$ is a SUFFICIENT STATISTIC for Θ, or
- $y(X)$ is SUFFICIENT for Θ

RESULT:

For any prior, posterior depends on X only through the value of y

USE OF SUFFICIENCY

Data reduction: Only need to save/record/store value of y

e.g., $n = 1000$ Bernoulli trials (patients at a VA hospital)

For $\theta =$ population proportion “successes”, only need to record $y = y(X) = \sum_{i=1}^{n} x_i$

Also, alternative sampling model for y generates same likelihood:

as a function of Θ,

$$p(y|\Theta) \propto l(y, \Theta)$$

EXAMPLES: POISSON MODELS

Poisson distribution for count $x = 0, 1, 2, \ldots$; with $(x|\theta) \sim Pois(\theta)$

Sampling model

$$p(x|\theta) = \frac{\theta^x \exp(-\theta)}{x!}, \qquad x = 0, 1, 2, \ldots$$

- Mean $E(x|\theta) = \theta$, ("rate" of Poisson model), also variance
- Basic model for "purely random" outcomes – Numbers of:
  - patients presenting at ER per day
  - radioactive particles emitted from a source per second
  - gamma ray recordings on astronomical detector per second
  - leukemia cases in North Carolina per year (nb., out of 7million$\approx \infty$)
- Last example: Poisson approximation to Binomial

$$Bin(n, p) \approx Pois(\theta) \text{ with } \theta = np \text{ for small } p, \text{ large } n$$

Random sample: n seconds of gamma ray recording

$X = \{x_1, \ldots, x_n\}$

Likelihood function: $\prod_{i=1}^{n} p(x_i|\theta) = a(X)l(y, \theta)$

- $a(X) = 1/\prod_{i=1}^{n} x_i!$
- $y = \sum_{i=1}^{n} x_i$
- $l(y, \theta) = \theta^y \exp(-n\theta)$

y =total number recorded in n seconds; sufficient for θ

MLE: $\hat{\theta} = y/n = \bar{x}$ (match sample mean with "model" mean)

Conjugate priors: $\theta \sim Gamma(a, b)$

Posterior: $(\theta|X) \sim Gamma(a + y, b + n)$

Explore and review: Homework exercises

“Reference posterior”

- Prior is “worth” b observations with total a
- Limiting case of “uninformative” prior: $a, b \to 0$ – reference posterior

$$(\theta|X) \sim Gamma(y, n)$$

- Posterior mean $E(\theta|X) = y/n = \bar{x} = \text{MLE}$

- *Theory:* by transformation, when $\theta \sim Gamma(a, b)$ then $\theta = \phi/b$ where $\phi \sim Gamma(a, 1)$
- S-Plus: `t<-seq(,,); plot(t,dgamma(t*b,a)*b);`

```
rt<-rgamma(1000,a)/b;
qgamma( c(0.025,0.975), a)/b;
```
etc