BINOMIAL MODEL

Review:

- Independent Bernoulli trials $x_i, (i = 1, \ldots, n)$
- "Success" probability $\theta : p(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$
- $y =$ number of successes $= \sum_{i=1}^{n} x_i$
- $y \sim Bin(n, \theta)$

$$p(y|n,\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

  on $y = 0, 1, \ldots, n$

- Usually (not always) drop conditioning on n in notation
- $E(y|\theta) = n\theta, V(y|\theta) = n\theta(1-\theta)$
- S-Plus: `dbinom, pbinom, qbinom, rbinom`

VA Hospital Monitors

- Each hospital, one year: n patients, y "successes"
- e.g., Hospital 21/1992: $y = 306, n = 651$
  1993: $y = 300, n = 705$
  Hospital 34/1992: $y = 9, n = 25$
  1993: $y = 14, n = 34$
- Issues: changes in "success rates" year-to-year?
  Comparisons across hospitals?
- Assumptions for binomial model?
- n.b., Full VA data set: years 1988-1996, other hospital variables

Assumptions $\rightarrow$ sampling model $y \sim Bin(n, \theta)$

Better notation: $(y|\theta) \sim Bin(n, \theta)$

INFERENCE for θ : a probability to be estimated based on observed proportion $t = y/n$ (and n of course)

Common point estimate: proportion $t = y/n$

Sampling distribution:

- $E(t|\theta) = \theta$
- $V(t|\theta) = \theta(1 - \theta)/n$

  more precise for large n and small/high θ

Likelihood and Log Likelihood Function

- $p(y|\theta)$ for FIXED y and as θ varies over parameter space
- Most likely value of θ for this data?
  Maximum likelihood estimate of $\theta : \hat{\theta} = t = y/n$
- Sample proportion "estimates" population probability
- Posterior mode
- Calculations easier with log likelihood function

FUNCTIONS OF PARAMETERS:

- e.g., $o \equiv o(\theta) = \theta/(1-\theta)$ or $\theta \equiv \theta(o) = o/(1+o)$
- likelihood is same under transformation: $p(y|o) = p(y|\theta(o))$

More on likelihood

- Likelihood ratios: compare two values of θ
- Likelihood defined up to multiplicative (positive) constant
- Standardized (or relative) likelihood: relative to value at MLE

$$r(\theta) = \frac{p(y|\theta)}{p(y|\hat{\theta})}$$

- Same "answers" (from likelihood viewpoint) from
  - binomial data (y successes out of n)
  - observed Bernoulli data (list of successes/failures in order)

LIKELIHOOD INTERVALS

Interval of θ values such that $r(\theta) > p$

e.g., $p = 0.1, 0.2$ – θ values for that are not “too unlikely”

```
theta<-seq(0,1,length=5000)     # range for θ
y<-306; n<-651                  # data
length(theta)
lik<-dbinom(y,n,theta)          # likelihood
rlik<-lik/max(lik)              # relative likelihood
plot(theta,rlik,type='l')
i<-rlik>0.1                     # selects interval
range(theta[i])                 # find interval endpoints
mean(range(theta[i])); y/n      # MLE = rough midpoint
```

Rule-of-thumb estimates of intervals for large samples

For n in the hundreds or more, t not close to 0 or 1

Roughly speaking, intervals are of the form $t \pm a$ with $a \approx 0.03 - 0.04$ when $p = 0.1 - 0.2$

Rule-of-thumb uncertainties: $t\pm$ 3 or 4%

Theory: a better "rule-of-thumb" – Asymptotic approximation of likelihood and distribution theory

$$t \pm a \qquad\qquad a = \sqrt{(-2\log(p))\frac{t(1-t)}{n}}$$

A FIRST BAYESIAN CALCULATION

Likelihood looks like a density function:

Integrate over (0,1) to *normalize* and produce a density

$$c\,\theta^y(1-\theta)^{n-y} \qquad \text{for} \quad 0 < \theta < 1$$

with

$$c = \frac{1}{\int_0^1 \theta^y(1-\theta)^{n-y} d\theta}$$

- Integrated likelihood function: Shape unchanged, specific constant
- Same density obtained from Bayes' theorem with a uniform prior density ...

BAYES' THEOREM

Bayesian statistics describes and measures uncertainty via probability

Parameter θ : Initial (marginal or prior) uncertainty described by uniform density

$$p(\theta) = 1, \qquad 0 < \theta < 1$$

- "flat" density, each point "equally weighted"
- "uninformative" about true value

Bayes's theorem (continued):

Conditional on observed outcome y (and n)

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

over $0 < \theta < 1$

OR

$$p(\theta|y) \propto p(\theta)p(y|\theta)$$

subject to normalization to unit integral

Bayes' theorem in proportional form

Special Case: Uniform prior $p(\theta) = 1$ implies normalized likelihood

INTERPRETATIONS of distributions for PARAMETER θ

- Measures of *belief* about values of θ
- Prior uniform: any value equally likely (in example)
- Posterior: Data weighted prior - mapping from

  Prior$\rightarrow$ Posterior
- θ is *not* random – probability distribution for θ describes *uncertainty* about its actual value
- Random variables *vs.* Uncertain quantities

THE BETA DISTRIBUTION

$$p(\theta) = \text{constant} \times \theta^{a-1}(1-\theta)^{b-1}, \text{ on } 0 < \theta < 1$$

- for specific values of the constants $a, b > 0$
- $\theta \sim Beta(a, b)$ or $Be(a, b)$ or $\beta(a, b)$
- Graphs: symmetric cases $a = b$, skewed cases, ..
- $E(\theta) = a/(a + b)$
- More concentrated, or "precise", for larger $a + b$
- $\theta \sim Beta(Mm, M(1 - m))$ where $m = a/M, M = a + b$
- $V(\theta) = m(1 - m)/(M + 1)$
- Modality:
  - unique mode at $(a - 1)/(a + b - 2)$ if $a, b > 1$
  - mode at 0 if $a < 1$, and/or one at 1 if $b < 1$

UNIFORM PRIOR ANALYSIS OF BINOMIAL DATA

Posterior

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

- Form is $Beta(y+1, n-y+1)$
  general form of posterior is $Beta(y+a, n-y+b)$
- Posterior mean $(y+1)/(n+1)$ and mode (usually) y/n
- More concentrated around mean/mode as n increases ("large sample")
- Easy to summarize – e.g., S-Plus
  - `dbeta, pbeta, qbeta, rbeta`
  - quantiles (percentiles, percentage points), e.g.,
    `qbeta(c(0.025,0.5,0.975),y+1,n-y+1)`

Posterior Intervals as summary "estimates" of θ

- *Names:*
  Credible intervals, (Bayesian) Confidence, Posterior intervals
- Central intervals (equal tails): e.g., 95% interval
  `qbeta(c(0.025,0.975),a,b)`
- Alternatives: Highest posterior density (HPD) intervals
- One-sided intervals, etc.

Shape and Spread of Beta Posteriors under Binomial Sampling:

Effects of binomial sample size n

- Plot densities for fixed $\hat{\theta} = y/n$ and $n = 10, 50, 100, 500, ..$
- More precise for larger n – Asymptotic Theory: concentrates around mode $\hat{\theta}$
- Closer to symmetric around mode as n increases

Variation in shape with "location" $\hat{\theta} = y/n$

- Plot densities for fixed n and $\hat{\theta} = 0.1, 0.25, 0.5, ...$
- More spread, uncertainty in central region
- Closer to symmetric in central region

SIMULATION FROM DISTRIBUTIONS (GCSR section 1.8)

Repeat random draws from distribution "Represent distribution"

- Random Sample: repeatedly sample without replacement
- CDF $F(\theta)$, inverse (=quantile function) $Q(\cdot)$ for $0 < u < 1$

  randomly sample a value of θ via $\theta = Q(u)$

  where $u \sim U(0,1)$
- Repeat: $\theta_1, \ldots, \theta_K$ for some large k
- Large random sample: HISTOGRAM approximates $p(\theta)$
- Cumulative ordered values approximate $F(\theta)$
- Sample moments approximate true moments, e.g., mean, ..

EXAMPLE: Beta distribution: `rbeta, hist, summary, ..`

COMPARISONS: A 2-SAMPLE PROBLEM

Comparisons of distributions – common & basic statistical problem

EXAMPLE: VA Hospital 21:
has true probability of success/failure changed between 1992 and 93?

DATA: $Y = \{y_1, n_1;\; y_2, n_2\}$

- In 1992, $y_1 = 306, n_1 = 651$
- In 1993, $y_2 = 300, n_2 = 705$

A BASIC MODEL:

- Independent binomial outcomes in each year: probabilities θ_1 and θ_2
- Independent continuous uniform priors $\rightarrow$ independent posteriors:

$$Beta(\theta_1|307, 346) \quad \text{and} \quad Beta(\theta_2|301, 406)$$

Graph posteriors: Densities and/or histograms of large samples

"Overlap?"

New parameter $\delta = \theta_2 - \theta_1$ measures *difference:* Inference on δ

We need $p(\delta|Y)$

- Immediately:
  $E(\delta|Y) = E(\theta_2|Y) - E(\theta_1|Y) = 0.426 - 0.470 = -0.044.$
- Is this significantly different from 0? Is it *really* negative?
- Immediately: $V(\delta|Y) = V(\theta_2|Y) + V(\theta_1|Y) = 0.0275^2$, S.D.$= 0.0275$

Can compute $p(\delta|Y)$ by transformation – but messy. Use Simulation . . .

Posterior simulation:

Large sample of k values for θ_1, similar for θ_2 and then compute δ

- `k<-5000;`
  `d<-rbeta(k,y2+1,n2-y2+1)-rbeta(k,y1+1,n1-y1+1)`
- `hist(d,nclass=30,prob=T); sum(d>0)/k; summary(d);`
  `mean(d)`

i.e., About a 95% probability that $\delta < 0$

the difference (of δ from 0) is "statistically significant at the 5% level"

For the VA, $\delta < 0$ represents an *improvement* in quality of care, so the data indicates a (very) likely improvement between 92 and 93

A 90% central posterior interval for δ is

```
quantile(d,prob=c(0.05,0.95))
```