STAT 530

Bayesian hypothesis testing

or

How good are the drugs on the pharmacy shelf?

April 7, 2010

Consider testing/approving a steady stream of candidate pharmaceutical drugs...

$\theta_i = 0$ i-th drug innefective
$\theta_i > 0$ i-th drug effective

Data: $\bar{X}_i \sim N(\theta_i, \sigma^2/n)$

Test statistic: $Z_i = \bar{X}_i/(\sigma/\sqrt{n})$.
$Z_i \leq 1.65$ i-th drug not approved
$Z_i > 1.65$ i-th drug approved

What proportion of drugs on the market are ineffective?

Say in the population of tested drugs,
$\theta \sim pN(0,0) + (1-p)N^+(0,\tau^2)$

$$\begin{aligned} Pr(\text{ineffective}|\text{approved}) &= Pr(\theta_i = 0 | Z_i > 1.65) \\ &= \end{aligned}$$

$Pr(\text{ineffective}|\text{approved})$

$p = 0.5$, $\sigma = 2$

	$n=20$	$n=200$	$n=\infty$
$\tau = 0.5$			
$\tau = 0.2$			

Alternate approval strategy

Be Bayesian, and use a prior. Say $\theta \sim p^* N(0,0) + (1-p^*)N^+(0,\tau^{*2})$.

Approve i-th drug $\leftrightarrow$ $Pr(\theta_i > 0|\bar{x}_i) > 0.95$

Compute $Pr(\theta_i = 0|\bar{X}_i = \bar{x}_i)$

```
prnull <- function(xbar, sgbar, p, tau) {

  ### model xbar|sgbar ~ N(theta, sgbar^2)
  ### prior theta ~ sim p*N(0,0)+ (1-p)*N+(0,tau^2)
  ### return Pr(theta=0|X=x)

  num <- p*dnorm(xbar,0,sgbar)

  tmp.sd <- sqrt(1/(1/sgbar^2 + 1/tau^2))
  tmp.mn <- (xbar/sgbar^2) / (1/sgbar^2 + 1/tau^2)
  den <- (1-p)*2*dnorm(xbar, 0, sqrt(sgbar^2+tau^2))*
         (1-pnorm(0,tmp.mn,tmp.sd))

  bf <- num/den
  pstprb <- bf/(1+bf)  ### post prob of null
}
```

How does this work under 'right prior' conditions?

```
set.seed(13);  NREP <- 50000

th <- rep(0,NREP) + rbinom(NREP,size=1,prob=.5)*
                    abs(rnorm(NREP,0,0.2))

xbar  <- rnorm(NREP, th, sqrt(4/200))

pstprb <- rep(NA, NREP)
for (i in 1:NREP) {
  pstprb[i] <- prnull(xbar[i], sqrt(4/200), 0.5, 0.2)
}
```

Operating characteristics

```
> table(pstprb<.05, th>0)

            FALSE  TRUE
      FALSE 24831 22799
      TRUE     45  2325
```

Compared to frequentist approach

```
> table( (xbar/sqrt(4/200))>1.65,  th>0)

              FALSE  TRUE
      FALSE 23630 16963
      TRUE   1246  8161
```

Generally, P-value << $P(H_0|\text{Data})$