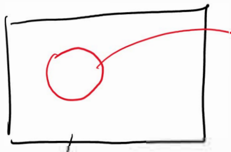
Wednesday, October 13, 2021 3:44 PM

Bernoulli



→ region Z

N_Z = #points in Z (control)

n_Z = #case points in Z

→ Study area

N : total #points (control)

n : total #case points

Likelihoods: (probability of observing the data)

$$L(p,q) = p^{n_Z} \cdot (1-p)^{N_Z - n_Z} \cdot q^{n-n_Z} \cdot (1-q)^{[(N-N_Z)-(n-n_Z)]}$$

H_0 assumes $p = q$

H_1 assumes $p > q$ (just keep p and q as seperate terms)

Maximum likelihood estimation:

- p and q are unknown. We find best estimates of p and q by maximizing the likelihood
- Achieved by setting derivatives w.r.t. p and q to 0
- To make it easier for derivatives, we take the log of $L(p,q)$ (note: $\log xy = \log x + \log y$)
  → This doesn't change solution because if $x_1 > x_2$ then $\log x_1 > \log x_2$

$$\log L(p,q) = n_Z \cdot \log p + (N_Z - n_Z)\log(1-p) + (n-n_Z)\log q + [(N-N_Z)-(n-n_Z)]\cdot \log(1-q)$$

(note: $\frac{d \log x}{dx} = \frac{1}{x}$)

H_1: Take derivative w.r.t. p and set to 0: (ignore last two terms as their values have NO relationship to p; i.e., p doesn't affect their values)

↓ $p \neq q$ (keep seperate)

$$n_Z \cdot \frac{1}{p} + \underbrace{(N_Z - n_Z) \cdot \frac{1}{1-p} \cdot (-1)}_{\text{this is the chain rule from your calculus class}} = 0$$

$$\boxed{p = \frac{n_Z}{N_Z}}, \quad \text{similarly,} \quad \boxed{q = \frac{n - n_Z}{(N - N_Z)}}$$

→ practice yourself

H_0: Take derivative w.r.t. p and set to 0

$p = q$ (just replace q with p in $L(p,q)$ before taking derivatives)

$$\boxed{p = q = \frac{n}{N}}$$

→ practice yourself

Plug these values back to $L(p,q)$

$$\text{Likelihood ratio } LR = \frac{L_1(p,q)}{L_0(p,q)}$$

$L_1(p,q)$ → use p,q solutions from H_1

$L_0(p,q)$ → use p,q ($p=q$) solution from H_0
