(Jan. 25, 2021)

Hypothesis Testing

Y = random observation $\mathcal{G}$ = set of observation events

$y \in \Gamma$ values it takes

Decide among a finite # M hypotheses about stat. behavior of Y.

$M = 2 \implies$

$H_0 : Y \sim P_0$ null hypothesis   pmf $p_0(y)$ or pdf $f_0(y)$

vs

$H_1 : Y \sim P_1$ alternate hypothesis $p_1(y)$, $f_1(y)$

Decision Rule a.k.a. Hypothesis Test δ

A d.r. is equiv. to a partition

$\Gamma_0 \in \mathcal{G}$ $\qquad$ $\Gamma_1 = \Gamma_0^c$

Γ_0 → choose H_0 when $y \in \Gamma_0$ "acceptance region"

Γ_1 → choose H_1 when $y \in \Gamma_1$ "rejection region or critical region"

Since $\Gamma = \Gamma_0 \cup \Gamma_1$ $\qquad$ $\Gamma_0 \cap \Gamma_1 = \phi$ → partition

defines a function

$$\delta(y) = \begin{cases} 0 & y \in \Gamma_0 \\ 1 & y \in \Gamma_1 \end{cases}$$

[Goal: Choose δ in some optimal way.]

FORM C
APPROVED FOR USE IN
PURDUE UNIVERSITY

Bayesian Hypothesis Testing (M=2)

Costs $[c_{ij}]$ = costs incurred by choosing H_i when truth is H_j

$$\begin{bmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{bmatrix}$$

Unif costs
$c_{00} = c_{11} = 0$
$c_{01} = c_{10} = 1$

Conditional Risks ≜ expected cost incurred by a dr δ given H_j is true

$$R_j(\delta) = E_j\{C(\delta(Y), j)\} \qquad j = 0, 1$$

$$= c_{1,j} P_j(\delta(Y)=1) + c_{0,j} P_j(\delta(Y)=0)$$

$$= c_{1,j} P_j(\Gamma_1) + c_{0,j} P_j(\Gamma_0)$$

Spec. Case: unif costs

$R_0(\delta) = P_0(\Gamma_1)$ — false alarm prob.

$R_1(\delta) = P_1(\Gamma_0)$. — "miss prob"

<u>One Solution</u> Bayesian ... trying to find a single average criterion to minimize.

Prior Probs : $\pi_0 = P(H_0 \text{ is true})$ $\quad \pi_0 + \pi_1 = 1.$

$\pi_1 = P(H_1 \text{ " " })$

Bayes Risk

Average over the prior distribution ...

$J = 0, 1$ r.v.

$$r(\delta) = E\{E\{C(\delta(Y), J) \mid J\}\}$$

$$= \pi_0 R_0(\delta) + \pi_1 R_1(\delta)$$

A **Bayes rule** $\triangleq$ a d.r. δ_B which minimizes the Bayes risk.

Derivation of Bayes Rule for M=2

$$r(\delta) = \pi_0 R_0(\delta) + \pi_1 R_1(\delta)$$

$$= \pi_0 \left[c_{10} P_0(\Gamma_1) + c_{00} \underbrace{P_0(\Gamma_0)}_{1 - P_0(\Gamma_1)} \right] + \pi_1 \left[c_{11} P_1(\Gamma_1) + c_{01} \underbrace{P_1(\Gamma_0)}_{1 - P_1(\Gamma_1)} \right]$$

$$= \pi_0 c_{00} + \pi_1 c_{01} + \pi_0 (c_{10} - c_{00}) P_0(\Gamma_1) + \pi_1 (c_{11} - c_{01}) P_1(\Gamma_1)$$

Assume that P_1 and P_0 have densities (discrete or continuous) st

$$P_j(\Gamma_1) = \int_{\Gamma_1} f_j(y)\, dy \qquad \left[\text{or } P_j(\Gamma_1) = \sum_{y \in \Gamma_1} p_j(y) \text{ if disc.} \right]$$

Then

$$r(\delta) = \pi_0 c_{00} + \pi_1 c_{01} + \pi_0 (c_{10} - c_{00}) \int_{\Gamma_1} f_0(y)\, dy + \pi_1 (c_{11} - c_{01}) \int_{\Gamma_1} f_1(y)\, dy .$$

Since the integrals are over the same set may write as one integral

$$r(\delta) = \pi_0 c_{00} + \pi_1 c_{01}$$

$$+ \int_{\Gamma_1} \left[\pi_0 (c_{10} - c_{00}) f_0(y) + \pi_1 (c_{11} - c_{01}) f_1(y) \right] dy$$

Desire to minimize $r(\delta)$ over Γ_1 (i.e., over the choice of set over which we integrate), which is not really a very familiar optimization scenario.

Aside

Since the only thing dependent on Γ_1 is the integral part one might as well think of how to minimize a set-dependent quantity of the form

$$Q(\Gamma_1) = \int_{\Gamma_1} g(y)\, dy$$

Aside

where $g(y)$ is some (fixed) function of points $y \in \Gamma$ and $\Gamma_1 \subset \Gamma$. Note that $g(y)$ can be positive, or negative, or zero as y varies over Γ. Note also that the entire collection of numbers

$$Q(\Gamma_1) \qquad \Gamma_1 \subset \Gamma; \; \Gamma_1 \in \mathcal{G}$$

is bounded because of the fashion in which $g(y)$ was constructed from pdfs/pmfs.

Now for simplicity imagine that Γ is a discrete set and that the integral is a sum. Then minimizing $Q(\Gamma_1)$ means trying to make it most negative (or zero) by the choice of which $y \in \Gamma$ should be put into Γ_1 ... i.e., we build up Γ_1 point by point

$$\Gamma_1 = \{ y \in \Gamma : g(y) \leq 0 \}$$

→ actually this definition is not unique because points y st $g(y) = 0$ are don't cares.

Also note:

① A legitimate possible choice for Γ_1 is $\Gamma_1 = \phi$ for which $Q(\phi) = 0$ ... can always get at least this small.

② "Optimal" choice of Γ_1 does not depend on assumption of a discrete Γ. Can argue almost the same way for pdfs which are continuous functions of y. Of course it also holds for Borel measurable functions of y.

Back to $r(\delta)$

From previous argument we should pick

$$\Gamma_1 = \left\{ y \in \Gamma : \pi_0 (c_{10} - c_{00}) f_0(y) + \pi_1 (c_{11} - c_{01}) f_1(y) \leq 0 \right\}$$

$$= \left\{ y \in \Gamma : \pi_1 (c_{11} - c_{01}) f_1(y) \leq \pi_0 (c_{00} - c_{10}) f_0(y) \right\}$$

Other Forms for Bayes Rule – M = 2

Suppose $c_{11} - c_{01} < 0$

$$\Gamma_1 = \left\{ y \in \Gamma : f_1(y) \geq \underbrace{\frac{\pi_0 (c_{00} - c_{10})}{\pi_1 (c_{11} - c_{01})}}_{\text{a threshold } \tau} f_0(y) \right\}$$

$$= \left\{ y \in \Gamma : \underbrace{\frac{f_1(y)}{f_0(y)}}_{\text{Likelihood ratio}} \geq \tau \right\}$$

Bayes Rule is a L.R.T.

Example: Location Testing with Gaussian Error ... i.i.d. Observations

$H_0: Y_k = \varepsilon_k + \mu_0$

vs. $\qquad k = 1, 2, \ldots, K$

$H_1: Y_k = \varepsilon_k + \mu_1$

$\mu_1 > \mu_0$

$\varepsilon_k \sim N(0, \sigma^2)$ and indep. statistically

(i.i.d. $\triangleq$ independent and identically distributed)

$$\Rightarrow \quad f_i(y_1 \ldots y_K) = \prod_{k=1}^{K} f_i(y_k)$$

$$= \frac{1}{(2\pi\sigma^2)^{K/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^{K} (y_k - \mu_i)^2 \right\} \qquad i = 0, 1$$

As before the Bayes test is a LRT

$$L(\underline{y}) = L(y_1 \ldots y_K) = \frac{f_1(y_1 \ldots y_K)}{f_0(y_1 \ldots y_K)} = \exp\left\{ \frac{-1}{2\sigma^2} \left[\sum_{k=1}^{K} (y_k - \mu_1)^2 - \sum_{\ell=1}^{K} (y_\ell - \mu_0)^2 \right] \right\}$$

$$L(\underline{y}) = \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{k=1}^{K}\left(y_k^2 - 2y_k\mu_1 + \mu_1^2 - y_k^2 + 2y_k\mu_0 - \mu_0^2\right)\right]\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{K}\left(2y_k(\mu_0-\mu_1) + \mu_1^2 - \mu_0^2\right)\right\}$$

$$= \exp\left\{\frac{\mu_1-\mu_0}{\sigma^2}\sum_{k=1}^{K} y_k - \frac{1}{2\sigma^2}K(\mu_1-\mu_0)(\mu_1+\mu_0)\right\}$$

Just as before the Bayes test is $L(y) \underset{<}{\geq} \tau$ $\begin{matrix}\text{say } H_1 \\ \text{say } H_0\end{matrix}$

Can simplify by taking log:

$$\frac{\mu_1-\mu_0}{\sigma^2}\sum_{k=1}^{K} y_k - \frac{K(\mu_1-\mu_0)(\mu_1+\mu_0)}{2\sigma^2} \underset{<}{\geq} \ln\tau \quad \begin{matrix}\text{say } H_1 \\ \text{say } H_0\end{matrix}$$

$$K\,\frac{\mu_1-\mu_0}{\sigma^2}\,\frac{1}{K}\sum_{k=1}^{K} y_k \;\underset{<}{\geq}\; \ln\tau + \frac{K(\mu_1-\mu_0)(\mu_1+\mu_0)}{2\sigma^2}$$

$$\frac{1}{K}\sum_{k=1}^{K} y_k \;\underset{<}{\geq}\; \frac{\sigma^2}{K(\mu_1-\mu_0)}\ln\tau + \frac{\mu_1+\mu_0}{2} \qquad \begin{array}{l}\text{say } H_1\\ \text{say } H_0\end{array}$$

Note: Test depends on observations only via the sample mean

$$\frac{1}{K}\sum_{k=1}^{K} y_k$$

... an example of concept of sufficient statistic