Hypothesis Testing

MLE/MAP are "point estimation" ← not always appropriate/meaningful in settings where priors are not sensible to assign.

Formulation: Consider $X \in \{0,1\}$ is an inference RV of interest.

It induces a continuum of tradeoffs based on observation.

Goal / $\max PCD = P(\hat{X}=1 \mid X=1)$ Probability of Correct Detection

s.t. $PFA = P(\hat{X}=1 \mid X=0) \le \beta$ Probability of False Alarm

Neyman Pearson (N-P)

1/ Observe Y

2/ 2 hypotheses – $H_0 : Y \sim f(y|0)$ Null

$H_1 : Y \sim f(y|1)$ Alternate

Decision rule – $r : \mathbb{R} \to \{0,1\}$

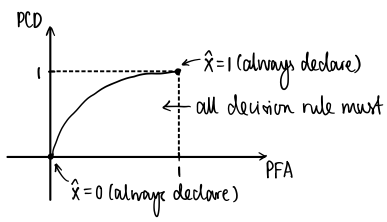
Error Types

	H_0	H_1
$r(y)=0$	✓	False Neg. (Type II)
$r(y)=1$	PFA False Pos (Type I)	PCD ✓ Power of test

significance level

$$\boxed{\begin{array}{l} \text{Goal / } \min \text{ Type II error} \quad \leftrightarrow \quad \max (1 - \text{Type II error}) = \text{Power of the Test} \\ \qquad \text{s.t. Type I error} \le \beta \end{array}}$$

ROC Curve



Likelihood Ratio

Def $\boxed{L(y) = \frac{f(y|1)}{f(y|0)}}$

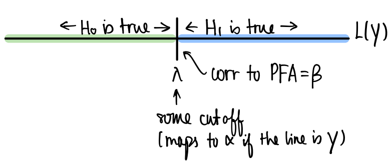
★ N-P Theorem: OPT Decision Rule

$$\hat{x} = r^*(y) = \begin{cases} 1 & \text{if } L(y) > \lambda \\ 0 & \text{if } L(y) < \lambda \\ 1 \text{ w.p. } \gamma & \text{if } L(y) = \lambda \end{cases}$$

Note: for monotonically decreasing $L(y)$, we flip the inequalities

where $\lambda > 0$ and $\gamma \in [0,1]$ chosen to ensure that PFA is equal to β

Implication:



$PFA = P[\hat{X}=1 \mid X=0] = P[Y > \alpha \mid X=0] \cdot 1 + P[Y = \alpha \mid X=0] \cdot \gamma = \beta$

Ex: Bias of a Coin

H_0: fair coin ($P(H) = 0.5$)

H_1: biased coin ($P(H) = 0.6$)

$X \in \{0,1\}$ where 0 = fair, 1 = biased

Observe Y_i for $i = 1, 2, \dots, n$

Goal / PFA $\leq 5\%$

Q/ What should the opt. decision rule be to decide if coin is fair or biased s.t. PFA $\leq 5\%$?

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n \mid X_0) = (0.5)^n \quad \text{fair}$$

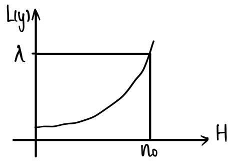
$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n \mid X_1) = (0.6)^H (0.4)^{n-H} \quad \text{biased}$$

$\uparrow$ H = heads observed

$$L(y_1, \dots, y_m) = \frac{P(Y_1, \dots, Y_m \mid X_1)}{P(Y_1, \dots, Y_m \mid X_0)} = \frac{(0.6)^H (0.4)^{n-H}}{(0.5)^n} = \left(\frac{0.4}{0.5}\right)^n \left(\frac{0.6}{0.4}\right)^H$$

H is also known as a "sufficient statistic" for the decision rule

Note: N-P thm $\Rightarrow$ $\hat{X} = \begin{cases} 1: H \geq n_0 \\ 0: H < n_0, \text{ where } P(H \geq n_0 \mid X = 0) = 0.05 \end{cases}$



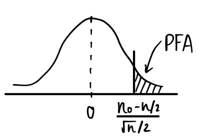
! A threshold on $L(y)$ corresponds to a threshold on H since $L(y)$ is monotic in H.

$X = 0$: $H \sim Bin(n, 1/2)$

$\mathbb{E}[H] = \frac{n}{2}$, $Var(H) = \frac{n}{4}$

Use CLT $\Rightarrow$ $P(H \geq n_0 \mid X = 0) = P\left(\underbrace{\frac{H - \mathbb{E}[H]}{\sigma_H}}_{\to N(0,1) \text{ as } n \to \infty} > \frac{n_0 - \mathbb{E}[H]}{\sigma_H}\right)$

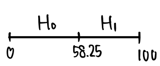
$\hookrightarrow$ $PFA = P\left(N(0,1) > \frac{n_0 - n/2}{\sqrt{n}/2}\right) = 1 - \Phi\left(\frac{n_0 - n/2}{\sqrt{n}/2}\right)$



$$\frac{n_0 - n/2}{\sqrt{n}/2} = \Phi^{-1}(0.95) = 1.65$$

$n_0 = 0.825\sqrt{n} + \frac{n}{2}$ $\leftarrow$ decision rule is independent of bias! but PCD will obviously differ (see Uniformly Most Powerful (UMP) rule)

e.g. $n = 100$, $n_0 = 58.25$ w/ 5% PFA



Outline of why randomized rule is needed:

No obs Y (you have to meet PFA $\leq \beta$)

2 choices:

a) $r = 0$ always $\rightarrow$ $P(\hat{X} = 1 \mid X = 0) = 0 \leq \beta$ but PCD = 0

b) r is a RV $\rightarrow$ $r = \begin{cases} 1 & \text{w.p. } \beta \\ 0 & \text{else} \end{cases}$ $\rightarrow$ $P(\hat{X} = 1 \mid X = 0) = \beta$, $PCD = P(\hat{X} = 1 \mid X = 1) = \beta$

Proof of N-P

Consider a binary HT idea of proof is to show that any other decision rule than N-P having the same "PFA spec" will not result in a better "PCD spec".

i.e. Let $\tilde{x}$ be an alt decision rule, then

If $P(\tilde{x}=1|X=0)\leq\beta$, then $P(\tilde{x}=1|X=1)\leq P(\hat{x}=1|X=1)$

Lemma: $(\hat{x}-\tilde{x})(L(y)-\lambda)\geq 0$

If $L(y)>\lambda$, $\hat{x}=1 \Rightarrow \hat{x}-\tilde{x}\geq 0$

If $L(y)<\lambda$, $\hat{x}=0 \Rightarrow \hat{x}-\tilde{x}\leq 0$

If $L(y)=\lambda$, $0\geq 0$

① $\hat{x}L(y)-\tilde{x}L(y)\geq\lambda\hat{x}-\lambda\tilde{x}$ condition on $X=0$ & take $\mathbb{E}$

$$\mathbb{E}[\hat{x}L(y)|X=0]-\mathbb{E}[\tilde{x}L(y)|X=0]\geq\lambda\{\underbrace{\mathbb{E}[\hat{x}|X=0]}_{PFA=P(\hat{x}=1|X=0)=\beta}-\underbrace{\mathbb{E}[\tilde{x}|X=0]}_{P(\tilde{x}=1|X=0)\leq\beta}\}\geq 0$$

$r(y)$

$$\boxed{\mathbb{E}[\hat{x}L(y)|X=0]\geq\mathbb{E}[\tilde{x}L(y)|X=0]}$$

$$LHS=\int r(y)L(y)\cdot f_{Y|X}(y|0)\,dy$$

$$=\int r(y)\frac{f(y|1)}{\cancel{f(y|0)}}\cancel{f(y|0)}\,dy$$

$$=\mathbb{E}[\hat{x}|X=1]$$

$$=P(\hat{x}=1|X=1)$$

$RHS=P(\tilde{x}=1|X=1)$ same as above

$\Rightarrow$ PCD(N-P) $\geq$ PCD(any other rule) □