$\frac{d}{dx} \sin^{-1}\sqrt{x}$ Let $\sin^{-1}\sqrt{x} = \theta \Rightarrow \sqrt{x} = \sin\theta$

$\Rightarrow x = \sin^2\theta$

$$= \frac{d\theta}{d\sin^2\theta} = \frac{1}{\frac{d\sin^2\theta}{d\theta}} = \frac{1}{2\sin\theta\cos\theta} = \frac{1}{2\sqrt{x}\sqrt{1-x}}$$

Recap:

Maximum likelihood estimator.

(1) Bayes estimator (2) Minimax estimator.

Frequentist:

$X_1, \dots, X_n \overset{iid}{\sim} f_\theta(x)$

Goal: provide point estimate of θ.

Frequentist

data are r.v.'s and parameter θ is fixed but unknown.

Bayesian

data are fixed, parameter θ is a random variable.

Bayesian: Goal is to estimate the unknown distribution of the parameter θ.

Algorithm: (1) Start with a prior distribution on θ. This represents our prior belief on θ.

(2) Use this prior distribution and the data to find the posterior distribution of θ.

Algorithm: Let $X_1, \dots, X_n \overset{iid}{\sim} f_\theta(x)$

Let the prior dist. of θ be given by $\pi(\theta)$.

The posterior dist. of $\theta | X_1, \dots, X_n$, denoted by $\pi(\theta | X_1, \dots, X_n)$ is defined as

$$\pi(\theta|x_1,\dots,x_n) = \frac{\left[\prod_{i=1}^{n} f_\theta(x_i)\right]\pi(\theta)}{\int\left[\prod_{i=1}^{n} f_\theta(x_i)\right]\pi(\theta)\,d\theta}$$

Example: $X_1,\dots,X_n \overset{iid}{\sim} \text{Ber}(p)$.

Goal: Posterior dist. of p.

Since $0<p<1$, a reasonable prior dist. on p is given by $p \sim \text{Beta}(\alpha,\beta)$.

$$\pi(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{\text{Beta}(\alpha,\beta)}, \quad 0<p<1$$

$$\pi(p|x_1,\dots,x_n) = \frac{\left[\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}\right]\pi(p)}{\int\left[\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}\right]\pi(p)\,dp}$$

$$= \frac{p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}\,p^{\alpha-1}(1-p)^{\beta-1}}{\int p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}\,p^{\alpha-1}(1-p)^{\beta-1}\,dp}$$

$$= \frac{p^{\alpha+\sum_{i=1}^{n} x_i-1}(1-p)^{\beta+n-\sum_{i=1}^{n} x_i-1}}{\int p^{\alpha+\sum_{i=1}^{n} x_i-1}(1-p)^{\beta+n-\sum_{i=1}^{n} x_i-1}\,dp}$$

$$\Rightarrow p|x_1,\dots,x_n \sim \text{Beta}\left(\alpha+\sum_{i=1}^{n} x_i,\ \beta+n-\sum_{i=1}^{n} x_i\right)$$

Example: $X_1, \ldots, X_n \overset{iid}{\sim} \text{Pois}(\lambda)$, $\lambda \sim \text{Gama}(a, b)$

$$\pi(\lambda) = \frac{\lambda^{a-1} e^{-\lambda b}\, b^a}{\Gamma(a)}, \quad \lambda > 0$$

$$\pi(\lambda \mid x_1, \ldots, x_n) = \frac{\left[\prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}\right] \lambda^{a-1} e^{-\lambda b} \frac{b^a}{\Gamma(a)}}{\int \left[\prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}\right] \lambda^{a-1} e^{-\lambda b} \frac{b^a}{\Gamma(a)}\, d\lambda}$$

$$= \frac{e^{-(n+b)\lambda}\, \lambda^{\sum_{i=1}^{n} x_i + a - 1}}{\int e^{-(n+b)\lambda}\, \lambda^{\sum_{i=1}^{n} x_i + a - 1}\, d\lambda}$$

$$\lambda \mid x_1, \ldots, x_n \sim \text{Gama}\left(\sum_{i=1}^{n} x_i + a,\; n+b\right)$$

Conjugate family: ~~prior~~

Let $\mathcal{F}$ denote the class of pdfs or pmfs $f(x|\theta)$. A class Π of prior distributions is a conjugate family for $\mathcal{F}$ if the posterior distribution is in the class Π for all $f \in \mathcal{F}$, all priors in Π and all $x \in \mathcal{X}$.

Now we will learn how to create a good estimator using this Bayesian knowledge.

Let $\delta(x)$ be an estimator of θ. The loss in estimating θ by $\delta(x)$ is represented by a function ~~given~~ known as the loss function. We denote the loss fn. by $L(\theta, \delta) / L(\theta, \delta(x))$

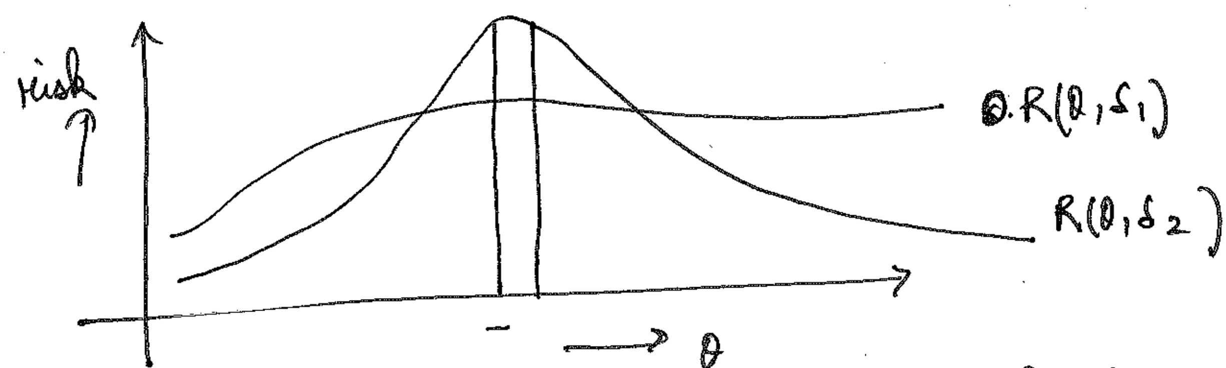
$R(\theta, \delta) = E_{X|\theta}[L(\theta, \delta(X))]$ is known as the risk function of δ.

$L(\theta, \delta(X)) = (\theta - \delta(X))^2 \rightarrow$ squared error loss function

$R(\theta, \delta) = \underset{X|\theta}{E}[L(\theta, \delta(X))] = \underset{X|\theta}{E}[(\theta - \delta(X))^2] \rightarrow$ Mean squared error.

Given two estimators δ_1 and δ_2, we say δ_1 is better than δ_2 if

$R(\theta, \delta_1) \leq R(\theta, \delta_2)$ $\forall \theta$ and the inequality is strict at least for one θ.



We want to create a summary of the entire risk fn.

1. Average risk $= E_\theta[R(\theta, \delta(X))] = E_\theta E_{X|\theta}[L(\theta, \delta(X))]$

2. Supremum risk $= \sup_\theta R(\theta, \delta(X))$

If $\delta(X)$ is an estimator of θ that minimizes the average risk over all estimators, then $\delta(X)$ is called the Bayes estimator of θ.

Average risk $= E_\theta E_{X|\theta}[L(\theta, \delta(X))]$

$= E_{X,\theta}[L(\theta, \delta(X))]$

$= E_X E_{\theta|X}[L(\theta, \delta(X))]$

If $\exists$ an estimator $\delta(X)$ that minimizes

$E_{\theta|X}[L(\theta, \delta(X))] \quad \forall X$

then that is going to be the minimizer of the average risk.

Finding the Bayes estimator is equivalent to finding an estimator $\delta(X)$ that minimizes

$E_{\theta|X}[L(\theta, \delta(X))]$

If $L(\theta, \delta(X)) = (\theta - \delta(X))^2$

$\underset{\delta(X)}{\arg\min}\ E_{\theta|X}[(\theta - \delta(X))^2]$

$E_{\theta|X}[(\theta - \delta(X))^2] = E_{\theta|X}[(\theta - E_{\theta|X}(\theta) + E_{\theta|X}(\theta) - \delta(X))^2]$

$= E_{\theta|X}[(\theta - E_{\theta|X}(\theta))^2] + \underline{2 E_{\theta|X}[(\theta - E_{\theta|X}(\theta))(E_{\theta|X}(\theta) - \delta(X))]}$

$+ E_{\theta|X}[(E_{\theta|X}(\theta) - \delta(X))^2]$

$= E_{\theta|X}[(\theta - E_{\theta|X}(\theta))^2] + 2\,(\cancel{E_{\theta|X}(\theta) - \delta(X)})[E_{\theta|X}(\theta) - E_{\theta|X}(\theta)]$

$+ E_{\theta|X}[(E_{\theta|X}(\theta) - \delta(X))^2]$

$\geq E_{\theta|X}[(\theta - E_{\theta|X}(\theta))^2]$

Thus $\delta(X) = E_{\theta|X}(\theta)$ is the Bayes estimator in this case.

$$\underset{\delta(x)}{\operatorname{argmin}} E_{\theta|x}\left((\theta-\delta(x))^2\right)$$

$$\frac{d}{d\delta} E_{\theta|x}\left((\theta-\delta)^2\right) = E_{\theta|x}\left(\frac{d}{d\delta}(\theta-\delta)^2\right) = 2E_{\theta|x}\left((\theta-\delta)\right) = 0$$

$$\Rightarrow \delta = E_{\theta|x}(\theta)$$

$L(\theta, \delta(x)) = \omega(\theta)(\theta-\delta(x))^2 \longrightarrow$ weighted loss fn.

$$\underset{\delta(x)}{\operatorname{argmin}} E_{\theta|x}\left(\omega(\theta)(\theta-\delta(x))^2\right)$$

$$\frac{d}{d\delta} E_{\theta|x}\left(\omega(\theta)(\delta-\theta)^2\right) = 0$$

$$\Rightarrow 2E_{\theta|x}\left(\omega(\theta)(\delta-\theta)\right) = 0 \Rightarrow \delta\, E_{\theta|x}(\omega(\theta)) = E_{\theta|x}(\theta\,\omega(\theta))$$

$$\Rightarrow \delta(x) = \frac{E_{\theta|x}(\theta\,\omega(\theta))}{E_{\theta|x}(\omega(\theta))}$$

Example: $X_1, \ldots, X_n \overset{iid}{\sim} \text{Ber}(p)$, $p \sim \text{Beta}(\alpha, \beta)$.

$$\pi(p|x_1, \ldots, x_n) = \frac{p^{\sum_{i=1}^n x_i + \alpha - 1}(1-p)^{n-\sum_{i=1}^n x_i + \beta - 1}}{\text{Beta}(\alpha + \sum_{i=1}^n x_i,\ n - \sum_{i=1}^n x_i + \beta)}$$

$$E[p|x_1, \ldots, x_n] = \int p\,\pi(p|x_1 \ldots x_n)\,dp$$

$$= \int \frac{p^{\sum_{i=1}^n x_i + \alpha}(1-p)^{n-\sum_{i=1}^n x_i + \beta - 1}\,dp}{\text{Beta}(\alpha + \sum_{i=1}^n x_i,\ n - \sum_{i=1}^n x_i + \beta)}$$

$$= \frac{\text{Beta}(\alpha + \sum_{i=1}^n x_i + 1,\ n - \sum_{i=1}^n x_i + \beta)}{\text{Beta}(\alpha + \sum_{i=1}^n x_i,\ n - \sum_{i=1}^n x_i + \beta)}$$

back ①

$$= \frac{\dfrac{\Gamma\left(\alpha+\sum_{i=1}^{n}x_i+1\right)\Gamma\left(n-\cancel{\sum_{i=1}^{n}x_i+\beta}\right)}{\Gamma(\alpha+1+n+\beta)}}{\dfrac{\Gamma\left(\alpha+\sum_{i=1}^{n}x_i\right)\Gamma\left(n-\cancel{\sum_{i=1}^{n}x_i+\beta}\right)}{\Gamma(\alpha+n+\beta)}}$$

$$= \frac{\sum_{i=1}^{n}x_i+\alpha}{\alpha+n+\beta} \quad \left(\text{Using the fact. } \Gamma(\alpha+1)=\alpha\,\Gamma(\alpha)\right)$$

$$= \underbrace{\bar{x}}_{\text{data mean}} \cdot \frac{n}{\alpha+\beta+n} + \frac{(\alpha+\beta)}{(\alpha+\beta+n)} \cdot \underbrace{\frac{\alpha}{(\alpha+\beta)}}_{\text{prior mean}}$$

posterior mean is the convex combination of data and prior.

Since $E[\bar{x}]=p \Rightarrow$ the above estimator can't be an unbiased estimator of p.

Thm: Under squared error loss, No unbiased estimator $\delta(X)$ of θ can be a Bayes estimator unless $E_\theta E_{X|\theta}\left((\theta-\delta(X))^2\right)=0$

Pf: Let $\delta(X)$ be an unbiased estimator of θ which is also a Bayes estimator. Thus,

$E_{X|\theta}[\delta(X)]=\theta \longrightarrow$ as $\delta(X)$ is unbiased

and

$E_{\theta|X}(\theta)=\delta(X) \longrightarrow$ as $\delta(X)$ is Bayes estimator.

back ②

$$E_\theta E_{X|\theta}\left[\delta(x)\theta\right] = E_\theta\left[\theta E_{X|\theta}[\delta(x)]\right] = E_\theta\left[\theta^2\right] \quad \cdots (1)$$

Also, $$E_X E_{\theta|X}\left[\delta(x)\theta\right] = E_X\left[\delta(x) E_{\theta|X}(\theta)\right] = E_X\left[\delta(x)^2\right] \quad \cdots (2)$$

As, $$E_\theta E_{X|\theta}\left[\delta(x)\theta\right] = E_X E_{\theta|X}\left[\delta(x)\theta\right]$$

$\Rightarrow$ by (1) & (2)

$$E_\theta E_{X|\theta}\left[\delta(X)^2\right] = E_\theta E_{X|\theta}\left[\delta(x)\theta\right] = E_\theta\left[\theta^2\right] \quad \cdots (3)$$

Hence,

$$E_{\theta,x}\left[(\delta(x)-\theta)^2\right] = E_{\theta,x}\left[\theta^2 - 2\delta(x)\theta + \delta(x)^2\right]$$

$$= E_\theta E_{X|\theta}\left[\delta(x)^2\right] + E_\theta\left[\theta^2\right] - 2E_\theta E_{X|\theta}\left[\delta(x)\theta\right]$$

$$= 0 \quad (\text{by } (3))$$

$\Rightarrow$ ~~the above equation needs to be satisfied for $\delta(x)$ as both Bayes~~ unbiased estimator, $\delta(x)$ to be a Bayes and ~~minimax~~ the above equation needs to be satisfied.

Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 is known, $\bar{X}$ is the UMVUE and

$$E_\mu E_{X|\mu}\left((\bar{x}-\mu)^2\right) = E_\mu\left(\frac{\sigma^2}{n}\right) = \frac{\sigma^2}{n} \neq 0.$$

$\Rightarrow \bar{x}$ can't be a Bayes estimator.

Minimax estimator minimizes $\sup_{\theta} E_{X|\theta}[L(\theta, \delta(X))]$.

Minimax estimator is difficult to find in general. But there are certain cases in which Bayes estimator becomes a minimax estimator.

Least favorable distribution

A prior dist. $\pi(\theta)$ on θ is known to be a least favorable prior if

$$E_{\theta} E_{X|\theta}[L(\theta, \delta_{\pi}(X))] \geq E_{\theta} E_{X|\theta}[L(\theta, \delta_{\pi'}(X))]$$

for all prior dist. π' on θ. Here δ_{π} and $\delta_{\pi'}$ are Bayes estimators w.r.t. priors π and π' respectively.

Recap:

Least favorable distribution.

A prior dist. $\pi(\theta)$ of θ is a least favorable prior if $E_\theta E_{X|\theta}\left[L(\theta, \delta_\pi(X))\right] \geq E_\theta E_{X|\theta}\left[L(\theta, \delta_{\pi'}(X))\right]$ for all prior dist. π' on θ. δ_π and $\delta_{\pi'}$ are the Bayes estimators corresponding to priors π and π'.

Result: If $\pi(\theta)$ is a prior dist. for which

$\int E_{X|\theta}\left[L(\theta, \delta_\pi)\right]\pi(\theta)\,d\theta = \sup_\theta E_{X|\theta}\left[L(\theta, \delta_\pi)\right]$, where δ_π is the Bayes estimator under the prior $\pi(\theta)$, then

a) δ_π is minimax

b) π is least favorable.

Pf:- a) For any estimator $\delta(X)$,

$$\sup_\theta E_{X|\theta}\left[L(\theta, \delta)\right] \geq E_\theta E_{X|\theta}\left[L(\theta, \delta)\right]$$
$$\geq E_\theta E_{X|\theta}\left[L(\theta, \delta_\pi)\right] = \sup_\theta E_{X|\theta}\left[L(\theta, \delta_\pi)\right]$$

thus δ_π minimizes supremum risk over all estimators. Hence δ_π is minimax estimator.

b) Note that

$$E_\theta E_{X|\theta}\left[L(\theta, \delta_\pi)\right] = \sup_\theta E_{X|\theta}\left[L(\theta, \delta_\pi)\right]$$
$$\geq \int E_{X|\theta}\left[L(\theta, \delta_\pi)\right]\pi'(\theta)\,d\theta \quad \text{(For any other prior dist. } \pi'\text{)}$$
$$\geq \int E_{X|\theta}\left[L(\theta, \delta_{\pi'})\right]\pi'(\theta)\,d\theta$$
$$= E_\theta E_{X|\theta}\left[L(\theta, \delta_{\pi'})\right]$$

⇒ ~~π' is the~~ π is the least favourable dist.

Example: $X_1, \dots, X_n \overset{iid}{\sim} \text{Ber}(p)$, $p \sim \text{Beta}(\alpha, \beta)$.

Bayes estimator of p under squared error loss

$$\delta_\pi(\underset{\sim}{X}) = \frac{\alpha + \sum_{i=1}^{n} X_i}{n+\alpha+\beta}.$$

$$E_{X|p}\left((\delta_\pi(\underset{\sim}{X}) - p)^2\right) = \frac{1}{(\alpha+\beta+n)^2}\left[\alpha^2 + \{\underbrace{n - 2\alpha(\alpha+\beta)}\}p + \{\underbrace{(\alpha+\beta)^2 - n}\}p^2\right]$$

We want to make this risk fn. constant fn. of p

thus we set,

$$n = 2\alpha(\alpha+\beta) \quad \text{and} \quad (\alpha+\beta)^2 = n$$

$$\Rightarrow \alpha = \frac{\sqrt{n}}{2}, \quad \beta = \frac{\sqrt{n}}{2}$$

Thus, under the prior dist. $p \sim \text{Beta}\left(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right)$, Bayes estimator produces constant risk.

Now use the result to argue that this Bayes estimator

$$\frac{\frac{\sqrt{n}}{2} + \sum_{i=1}^{n} X_i}{n+\sqrt{n}}$$

is a minimax estimator under the prior $\text{Beta}\left(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right)$.

Of course, by the same result $\text{Beta}\left(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right)$ is the least favourable dist.

3. Testing of Hypothesis

Statistical hypothesis testing is all about

1. Beginning with a tentative idea about the unknown parameter.
2. Want to test validity of this tentative idea based on sample information.

(3) Existing tentative idea : H_0 (null hypothesis), new idea : H_1 (alternative hypothesis).

(4) We begin by assuming that the null hypothesis is true. Only when there is an overwhelming evidence contradicting the null do we reject it in favour of alternative.

	H_0 is true	H_0 is false
Do not reject H_0	Correct	Type 2 error.
reject H_0	Type 1 error	correct.

Our goal is to minimize

P(Type 1 error) = P(reject H_0 | H_0 is true)

P (Type 2 error) = P(not rejecting H_0 | H_0 is false).

They can't be minimized together. Thus we fix P(type 1 error) and minimize P(Type 2 error)

Minimizing P(Type 2 error)

$\Longleftrightarrow$ Maximizing 1 − P(Type 2 error)

= P(rejecting H_0 | H_0 is false)

P(type 1 error) is called the level of the test and 1 − P(type 2 error) is called the power of the test.

Parametric tests: Let $X_1, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$

We test $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$, where Θ_0 and Θ_1 are disjoint sets. If Θ_0 is a singleton set, then the null hypothesis is called a simple null hypothesis. If Θ_0 has multiple elements, then the null hypothesis is called a composite null hypothesis.

Rejection region: Let $R = \{\underline{x} \in \mathcal{X} \mid H_0 \text{ is rejected for } \underline{x}\}$

be known as the rejection region or critical region of a test.

Let $\phi(\underline{x})$ = Prob. of rejecting H_0 when $\underline{x}$ is observed. the power function of a test is given by

$\beta(\theta) = \underset{\underline{X}|\theta}{E}[\phi(\underline{X})]$. Power function is a function of the parameter θ.

Consider the situation $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$.

$\beta(\theta_0) = E_{\underline{X}|\theta_0}[\phi(\underline{X})]$ = level of the test.

$\beta(\theta_1) = E_{\underline{X}|\theta_1}[\phi(\underline{X})]$ = power of the test.

Under the composite null hypothesis, $H_0 : \theta \in \Theta_0$. the level of the test α implies

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha.$$

Given a certain level, we want to find the most powerful test.

X	0	1	2	3
f_0	1/8	1/8	1/4	1/2
f_1	1/2	1/4	1/8	1/8

$X \sim f_\theta \quad \theta = 0, 1$

How to find the best (most powerful level 1/8 test).

level 1/8 means rejection region R must have prob. 1/8 under H_0.

$R_1 = \{0\}$ test 1

$R_2 = \{1\}$ test 2.

power of test 1 = P(rejecting H_0 | H_0 is false)

$= P(X \in R_1 \mid H_0 \text{ is false}) = 1/2$

power of test 2 $= P(X \in R_2 \mid H_0 \text{ is false}) = 1/4$.

test 1 is more powerful than test 2 under the same level.

$\frac{f_1}{f_0}$ is higher at 0 than at 1.

Let us decide to reject H_0 with a rejection region that contains high values of $\frac{f_1}{f_0}$.

power $= P(X \in R \mid f_1 \text{ is true})$

Theorem: Neyman-Pearson lemma:

Consider testing $H_0: \theta = \theta_0$ vs. $H_1: \theta = \theta_1$, where pdf or pmf corresponding to θ_i is $f(x|\theta_i)$, $i = 0, 1$. Using a test with rejection region R that satisfies

$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } f(\underset{\sim}{x}|\theta_1) > k f(x|\theta_0) \Leftrightarrow \dfrac{f(x|\theta_1)}{f(x|\theta_0)} > k \\ 0 & \text{o.w.} \end{cases}$$

for some $k \geq 0$, and $\alpha = P_{\theta_0}(\underset{\sim}{X} \in R)$. Then

a) Any test that satisfies the above is the most powerful level α-test.

b) If there exists a test satisfying the above, then it is a most powerful level α-test of its level.

Note:
$$\frac{f(\underset{\sim}{x}|\theta_1)}{f(\underset{\sim}{x}|\theta_0)} = \frac{g(T(\underset{\sim}{x})|\theta_1)\,\cancel{h(\underset{\sim}{x})}}{g(T(\underset{\sim}{x})|\theta_0)\,\cancel{h(\underset{\sim}{x})}}$$

MP test can be written as

$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } g(T(\underset{\sim}{x})|\theta_1) > k\, g(T(\underset{\sim}{x})|\theta_0) \\ 0 & \text{o.w.} \end{cases}$$

Example: $X_1, X_2 \overset{iid}{\sim} \text{Ber}(\theta)$. Want to test $H_0: \theta = \frac{1}{2}$ vs. $H_1: \theta = \frac{3}{4}$.

$\sum_{i=1}^{2} X_i$ = sufficient stat.

$\sum_{i=1}^{2} X_i \sim \text{Bin}(2, \theta)$.

$$\frac{f(0 \mid \theta = 3/4)}{f(0 \mid \theta = 1/2)} = \frac{1}{4}, \quad \frac{f(1 \mid \theta = \frac{3}{4})}{f(1 \mid \theta = \frac{1}{2})} = \frac{3}{4}, \quad \frac{f(2 \mid \theta = \frac{3}{4})}{f(2 \mid \theta = \frac{1}{2})} = \frac{9}{4}.$$

if we choose $\frac{3}{4} < K < \frac{9}{4} \Rightarrow R = \{2\}$

MP level α test for $\alpha = P\left(\sum_{i=1}^{2} x_i = 2 \mid \theta = \frac{1}{2}\right) = \frac{1}{4}$.