Bayesian approach

- Posterior risk of assigning an observation x to the jth class: a_j

posterior $\propto f(x|\theta)\,g(\theta) = \text{joint}$

$$\mathrm{PR}(j|x) = \sum_{i=1}^{m} l_{ij}\, h(\theta_i|x) = \sum_{i\neq j} h(\theta_i|x) = \frac{\sum_{i\neq j}\pi_i f(x|\theta_i)}{\sum_{k=1}^{m}\pi_k f(x|\theta_k)}$$

fixed; a_j; same

same derivation as for (*) in LNp.30

$$= P_{\Theta|X}(\Theta\neq\theta_j|x) = 1 - P_{\Theta|X}(\Theta=\theta_j|x) = 1-h(\theta_j|x)$$

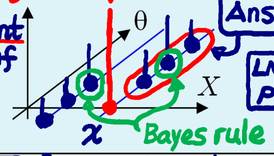
- $\mathrm{PR}(j|x)$ is minimized for that value of j such that $P_X(x)$: marginal

Sum=1 (?) Ans. No

Bayes rule $\leftarrow P_{\Theta|X}(\Theta=\theta_j|x)$ is maximized, i.e., maximize joint pmf

(reasonable?) mode of $\theta|X$

6/3 for the class that has maximum posterior probability.



LN, CHI~6 p.21

cf. Q: Why not use mean or median of $\theta|X$?

Example 10.11 (waiting times between emissions, 2nd Ed., TBp. 582)

- Data: $\mathbf{X}=(X_1,X_2,\dots,X_n)=n$ waiting times between emissions of alpha particles from a radioactive substance.

- On the basis of $\mathbf{X}$, a decision is to be made as to whether to classify the particles as coming from substance I or substance II.
- Statistical Modeling: $\Omega=\{\theta_1,\theta_2\}$: 2 classes
  - $X_1,X_2,\dots,X_n$ are i.i.d. $\sim E(\theta_1)$ if particles came from substance I

    exponential; assumed known
  - $X_1,X_2,\dots,X_n$ are i.i.d. $\sim E(\theta_2)$ if particles came from substance II
  - prior: $\pi_1=\pi_2=1/2$

    Q: What if we treat it as a testing problem? e.g. H_0: $E(\theta_1)$ vs. H_A: $E(\theta_2)$ cf. more protection

risk = misclassification probability

- loss function: 0-1 loss

Bayesian approach $\xleftrightarrow{cf.}$ Q: What is minimax rule under Frequentist approach? (exercise)

- posterior probabilities:

$$h(\theta_i|\mathbf{x}) = \frac{\tfrac{1}{2}\,f(\mathbf{x}|\theta_i)}{\tfrac{1}{2}f(\mathbf{x}|\theta_1)+\tfrac{1}{2}f(\mathbf{x}|\theta_2)} = \frac{\theta_i^{\,n}\exp\left(-\theta_i\sum_j x_j\right)}{\theta_1^{\,n}\exp\left(-\theta_1\sum_j x_j\right)+\theta_2^{\,n}\exp\left(-\theta_2\sum_j x_j\right)}$$

- Notice that

$<$ decision: $d(x)=\theta_2$; $\times\frac{1}{2}=$ joint ; $f(\underline{x}|\theta_1) \lesseqgtr f(\underline{x}|\theta_2)$

decision: $d(x)=\theta_1$

$$h(\theta_1|\mathbf{x}) > h(\theta_2|\mathbf{x}) \;\Leftrightarrow\; \theta_1^{\,n}\exp\left(-\theta_1\sum_j x_j\right) > \theta_2^{\,n}\exp\left(-\theta_2\sum_j x_j\right)$$

$$\Leftrightarrow\; n\log\theta_1-\theta_1\sum_j x_j > n\log\theta_2-\theta_2\sum_j x_j$$

taking log

$$\Leftrightarrow\; (\theta_2-\theta_1)\sum_j x_j > n\log(\theta_2/\theta_1) \quad \text{reasonable?}$$

$$\Leftrightarrow\; \sum_j x_j \;\; > \text{ or } < \;\; n[\log(\theta_2/\theta_1)]/(\theta_2-\theta_1)$$

$Z\sim E(\theta)$, $E(Z)=1/\theta$; $\theta_2-\theta_1>0$; $\theta_2-\theta_1<0$; known

Why known? Why not determined by a significance level α? Note. The protection on H_0 is already reflected in π_1. (exercise)

- The Bayesian classification rule is thus based on the size of $\sum_{j=1}^{n} x_j$. $\rightarrow$ same statistic to determie "more extreme" in MP test.

• Application of Decision Theory: Hypothesis Testing

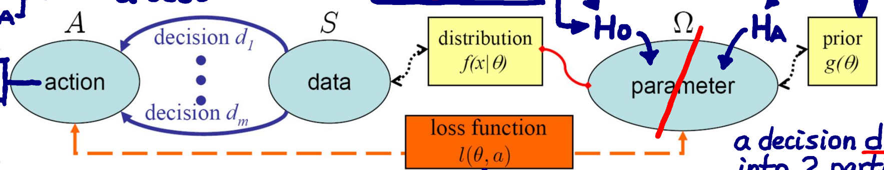
Hypothesis testing can be regarded as a classification problem with a constraint on the probability of miscalssification (i.e., the probability of a type I error). $\leftarrow$ significance level α

Risk function (θ, d) $\leftarrow$ LNp.30 ; H_0, H_A ; a test

a_1: not reject H_0

a_2: reject H_0

α: protects H_0 class ; 2 classes ; cf.



0-1 loss

Both type I and type II errors are misclassification errors

a decision d splits S into 2 partitions: RR and AR

2nd Edition, Ch15, p.33

- Recall (Frequentist approach for simple vs. simple hypotheses): (EX10.11, LNp.31) $X_1,\cdots,X_n \overset{iid}{\sim} E(\theta)$
  - $H_0: \mathbf{X} \sim f_0(\mathbf{x})$ vs. $H_A: \mathbf{X} \sim f_1(\mathbf{x})$ (e.g., $H_0: \theta=\theta_1$ vs. $H_A: \theta=\theta_2$ or $H_0: \theta=\theta_2$ vs. $H_A: \theta=\theta_1$) protect
  - Likelihood ratio (or MP) test accepts H_0 if

f_0: $f(\underset{\sim}{x}|\theta_1)$, f_1: $f(\underset{\sim}{x}|\theta_2)$

In both Frequentist & Bayesian — determine which data $\underset{\sim}{x}$ is more extreme

$$\frac{f_0(\mathbf{x})}{f_1(\mathbf{x})} > c \quad \Leftrightarrow \quad \frac{1}{c}\left(\frac{f_0(\mathbf{x})}{f_1(\mathbf{x})}\right) > 1$$

usually <1

and rejects H_0 otherwise.

use significance level α

  - The constant c is chosen to control the probability of type I error. (under H_0)

- Bayesian approach: critical value ($c\downarrow$ when $\alpha\downarrow$)

probability of misclassification: H_0 (true class) $\rightarrow$ H_A (classified class)

for classification

  - Assign a prior probability to H_0 and H_A:
$$P(H_0)=\pi, \quad \text{and} \quad P(H_A)=1-\pi.$$

risk function at H_0 under 0-1 loss (LNp.30)

Q: how about $\beta=P_{H_A}$(type II error)?

  - Posterior probabilities are:
$$P(H_0|\mathbf{x}) = \frac{P(H_0,\mathbf{x})}{P(\mathbf{x})} = \frac{P(\mathbf{x}|H_0)\,P(H_0)}{P(\mathbf{x})} \quad \text{and} \quad P(H_A|\mathbf{x}) = \frac{P(\mathbf{x}|H_A)P(H_A)}{P(\mathbf{x})}$$

  - The ratio of posterior probabilities:

How can Bayesian approach protect (if we want) H_0? Ans. assign a larger π in prior.

For $\underset{\sim}{x}$ satisfy this, Bayes rule $d(\underset{\sim}{x})=H_0$

$$\frac{P(H_0|\mathbf{x})}{P(H_A|\mathbf{x})} = \frac{P(H_0)}{P(H_A)}\,\frac{P(\mathbf{x}|H_0)}{P(\mathbf{x}|H_A)} = \frac{\pi}{1-\pi}\,\boxed{\frac{f_0(\underset{\sim}{x})}{f_1(\underset{\sim}{x})}}$$

  - So,
$$\frac{P(H_0|\mathbf{x})}{P(H_A|\mathbf{x})} > 1 \iff \boxed{\frac{P(\mathbf{x}|H_0)}{P(\mathbf{x}|H_A)}} > \frac{P(H_A)}{P(H_0)} = \frac{1-\pi}{\pi} \equiv c'$$

<1: reject H_0; accept H_0

Question 7.9 (LN, CH9, p.19) cf.

cf.

2nd Edition, Ch15, p.34

Theorem 10.9 (Neyman-Pearson Lemma from Bayesian viewpoint, 2nd Ed., TBp.583)

LN, CH9, p.17

Let
- d^* be a test that accepts H_0 if $\dfrac{f_0(\mathbf{x})}{f_1(\mathbf{x})} > c$, — likelihood ratio $>c$
  a decision function for classification; RR(rejection region)=AR^c ← AR(acceptance region)
- α^* be the significance level of d^*, and ← under Frequentist approach; determine c
- d be another test that has significance level $\alpha \le \alpha^*$. — level-α^* test

Then, the power of d is less than or equal to the power of d^*. — most powerful test

$1-$risk function at H_A $=P$(correct class.$|H_A$) $=P(\ell=0|H_A)$ $=P(RR|H_A)$ $=1-\beta=$ power

Proof. Let $1/c = \pi/(1-\pi)$. Then d^* accepts if
$$\frac{\pi}{1-\pi}\,\frac{f_0(\mathbf{x})}{f_1(\mathbf{x})} > 1$$

set $\pi=\dfrac{1}{1+c}$; $d^*(\underset{\sim}{x}) = \begin{cases} H_0, & \text{if } \underset{\sim}{x} \\ H_A, & \text{o.w.}\end{cases}$ — interpreted from Bayesian approach

and (d^* is thus a Bayes rule for a classification problem with prior probabilities π $(=1/(1+c))$ and $1-\pi$ for H_0 and H_A respectively, and with 0-1 loss.) So,

cf. • Freq.: min β, s.t. $\alpha \le \alpha^*$ • Bayes: min $\pi\alpha+(1-\pi)\beta$

$\ell=1$, if $\underset{\sim}{x}\in RR$

$$B(d^*) \le B(d). \leftarrow \text{Bayes risks}$$

risk function at H_A $=P$(misclassification$|H_A$) $=P(\ell=1|H_A)$ $=P(AR|H_A)$ $=P$(type II error$|H_A$)$=\beta$

But,
$$\begin{aligned} B(d^*)-B(d) &= \pi\left\{E_{\mathbf{X}}[l(H_0,d^*(\mathbf{X}))] - E_{\mathbf{X}}[l(H_0,d(\mathbf{X}))]\right\} \\ &\quad + (1-\pi)\left\{E_{\mathbf{X}}[l(H_A,d^*(\mathbf{X}))] - E_{\mathbf{X}}[l(H_A,d(\mathbf{X}))]\right\} \\ &= \pi(\alpha^*-\alpha) + (1-\pi)\left\{E_{\mathbf{X}}[l(H_A,d^*(\mathbf{X}))] - E_{\mathbf{X}}[l(H_A,d(\mathbf{X}))]\right\} \le 0 \end{aligned}$$

risk function at H_0 $=P$(misclassification$|H_0$) $=P(\ell=1|H_0)=P(RR|H_0)$ $=P$(type I error$|H_0$)

$\ell=1$, if $\underset{\sim}{x}\in AR$

≥ 0; RR; ≤ 0

Thus, $E_{\mathbf{X}}[l(H_A,d^*)] \le E_{\mathbf{X}}[l(H_A,d)]$, i.e., $P(d^* \text{ reject } H_0|H_A) \ge P(d \text{ reject } H_0|H_A)$.

$\boxed{1-}$ $\ge$ $\boxed{1-}$; power

$\rightarrow$ EX10.11 (LNp.31-32)

cf.

2nd Edition, Ch15, p.35

Example 10.12 (hypothesis testing of exponential distribution, 2nd Ed., TBp. 583)

parameter

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables following the exponential(θ) distribution. Consider $H_0: \theta = \theta_1$ vs. $H_A: \theta = \theta_2$

$\pi = 1/(1+c)$
if $\pi = 1/2 \Leftrightarrow c = 1$
$\Leftrightarrow \alpha = ?$ (exer)
if $\alpha = 0.05 \Leftrightarrow c = ?$ (exer)
$\Leftrightarrow \pi = ?$ (exer)

- Likelihood ratio test rejects H_0 for small values of

Bayes rule

$$c > \frac{f(\mathbf{x}|\theta_1)}{f(\mathbf{x}|\theta_2)} = \left(\frac{\theta_1}{\theta_2}\right)^n \exp\left[(\theta_2 - \theta_1)\sum_j x_j\right] \Leftrightarrow \sum_j x_j \ \substack{\le \\ \ge} \ \frac{\log(c)}{\theta_2 - \theta_1} + \frac{n\log(\theta_2/\theta_1)}{\theta_2 - \theta_1}, \ \substack{\text{if } \theta_2 > \theta_1 \\ \text{if } \theta_2 < \theta_1}$$

- The critical value c is calculated for any desired significance level by using the fact that $\sum_j X_j$ follows gamma(n, θ_1) distribution under H_0. $\boxed{\text{reasonable?}}$

test statistic

❖ **Reading**: textbook (2nd ed.), 15.2.3

$X_1, \cdots, X_n \overset{i.i.d.}{\sim} E(\theta_1) \Rightarrow \sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \theta_1) \leftarrow$ null dist.

Future Statistics Courses

統計學 (MATH2820)

理論
- 數統導論 (STAT3875, 大學)
- 數理統計 (STAT5110, 碩)

應用
- 統計資料分析 (STAT2622, 大學)
- 迴歸分析 (STAT, 大學)
- 線性模式 (STAT5410, 碩)
- 應用多變量分析 (STAT5191, 碩)
- 時間序列分析 (STAT5210, 碩)
- 類別資料分析 (STAT5230, 碩)

線性代數