Παρασκευή 25/10/20 Μάθημα 3ο

Χώρος αποφάσεων:

Συμβολίζεται με D και είναι το σύνολο όλων των δυνατών ενεργειών/πράξεων d που μπορεί να λάβει κάποιος

- στην εκτιμητική ο χώρος D είναι ίδιος με τον παραμετρικό Θ
- στον έλεγχο υποθέσεων είναι $D = \{$ απόρριψη H_0 ή μη απορ. $H_0\}$

Συνάρτηση απόφασης

Είναι μια μετρήσιμη συνάρτηση d με Π.Ο το $\mathbb{R}^n$ και Π.Τ. τον D.

Η τιμή $d(X)$ στο σημείο $X = (X_1, .. X_n)$ ονομάζεται απόφαση

Συνάρτηση Απώλειας:

Είναι η συνάρτηση $L(\theta, D)$ που εκφράζει τη ζημιά που θα έχουμε αν πάρουμε την απόφαση $d(X)$ ενώ η πραγματική τιμή που θέλουμε να εκτιμήσουμε είναι θ.

$L(\theta, D) \geq 0 \ \forall \theta$

$L(\theta, \theta) = 0$

Παραδείγματα συναρτήσεων απώλειας.

1) Απόλυτο σφάλμα $L(\theta, d) = |d - \theta|$

2) Τετραγωνικό σφάλμα $L(\theta, d) = (d - \theta)^2$

3) $L(\theta, d) = |d - \theta|^k, \ k > 0$

4) Linex $L(\theta, d) = e^{k(d-\theta)} - k(d - \theta) - 1$

5) Stein $L(\theta, d) = \frac{d}{\theta} - 1 - \log\left(\frac{d}{\theta}\right)$

Linex.

$L(\theta,\theta) = e^{k(\theta-\theta)} - k(\theta-\theta) - 1 = 0$ άρα $L(\theta,\theta)=0$

ν.δ.ο $L(\theta,d) \geq 0$

Θέτω $k(\theta-d) = x \overset{\text{ν.δ.ο}}{\Rightarrow} e^x - x - 1 \geq 0$

$f(x) = e^x - x - 1$

$f'(x) = e^x - 1$, $f'(x) = 0 \Rightarrow x = 0$

Στο $[0,+\infty)$ $f(x) \geq 0$

Άρα Linex συνάρτηση απώλειας

Stein

$L(\theta,\theta) = \frac{\theta}{\theta} - 1 - \log\left|\frac{\theta}{\theta}\right| = 0 \Rightarrow L(\theta,\theta) = 0$

ν.δ.ο $L(\theta,d) \geq 0$ $\forall \theta$

Θέτω $\frac{d}{\theta} = x > 0$ Άρα έχω

$f(x) = x - 1 - \log x$,

$f'(x) = 1 - \frac{1}{x}$ $f'(x) = 0 \Rightarrow x = 1$ Άρα στο $[1,+\infty)$ $f \uparrow$

$\forall x > 0$ $f(x) \geq 0$

Συνάρτηση Κινδύνου:

Ορίζεται από την σχέση $R(\theta,d) = E\{L(\theta,d(X))\}$.

Πρόταση:

Όταν η συνάρτηση απώλειας είναι το τετραγωνικό σφάλμα $L(\theta,d) = (d-\theta)^2$ τότε

$R(\theta,d) = V(d) + (Ed-\theta)^2$.

Απόδειξη:

$R(\theta,d) = E(L(\theta,d(x))) = E(d-\theta)^2$.

$EX^2 = Var X + (EX)^2$.

εφαρμόζωντας για $X = d - \theta$ οπότε

$R(\theta,d) = V(d) + (Ed-\theta)^2$.

• η ποσότητα d αντιστοιχεί στην απόκλιση
• $Ed-\theta$ ποσό μεροληψίας
$Ed<\theta$ υποεκτίμηση, $Ed>\theta$ υπερεκτίμηση
$Ed=\theta$ αμερόληπτη εκτίμηση

Παράδειγμα:
Έστω τ.δ $X_1,\dots X_n \sim N(\theta,\sigma^2)$ όπου $\theta\in\mathbb{R}$ και σ^2 γνωστή διακύμανση. Να υπολογιστεί το Μ.Τ.Σ του $\bar{X}$ και η συνάρτηση κινδύνου με συνάρτηση απώλειας το απόλυτο σφάλμα του $\bar{X}$.

$$MSE(\bar{X}) = Var(\bar{X}) + \{E(\bar{X}-\theta)\}^2 = \frac{\sigma^2}{n}.$$

$$E|\bar{X}-\theta| = \int |\bar{x}-\theta| f_{\bar{X}}(x)\, d\bar{x}$$

• $L(\theta,d) = L(\theta,\bar{X}) = |\bar{X}-\theta|$
Θέτω $\bar{x}=y$ τότε $d\bar{x}=y$.

$$R(\theta,\bar{x}) = EL(\theta,\bar{X}) = E|\theta,\bar{X}| = \int_{-\infty}^{+\infty} |y-\theta|(2n^{-1}\pi\sigma^2)^{-1/2} \cdot e^{-\frac{(y-\theta)^2}{2n^{-1}\sigma^2}} dy$$

Θέτουμε $y-\theta=z \Rightarrow dy=dz$

$$\int_{-\infty}^{+\infty} |z|(2n^{-1}\pi\sigma^2)^{-1/2} e^{-\frac{z^2}{2n^{-1}\sigma^2}} dz \overset{f(z)=f(-z)}{=} 2\int^{+\infty} \frac{z}{(2n^{-1}\pi\sigma^2)^{1/2}} e^{-\frac{z^2}{2n^{-1}\sigma^2}} dz \quad \dots$$

Καταλήγουμε στο $R(\theta,d) = R(\theta,\bar{X}) = E|\bar{X}-\theta| = \frac{\sqrt{2}\sigma}{\sqrt{\pi n}}$.

Παράδειγμα 2:

Έστω τ.δ. $X_1, \ldots X_n \sim N(\mu, \sigma^2=\theta)$ σ^2 άγνωστη, μεσοσυνάρτηση απώλειας το τετραγωνικό σφάλμα.
Ποια από τις $d_1 = \frac{1}{n}\sum(X_i-\mu)^2$ και $d_2 = k d_1$, $k>0$ είναι καλύτερη και πότε;

Θα συγκρίνουμε τα $R(\theta, d_1)$ και $R(\theta, d_2)$

$R(\theta, d_1) = V(d_1) + (Ed_1 - \theta)^2$

Ξέρουμε ότι για $X_1, \ldots X_n \sim N(\mu, \sigma^2)$ ισχύει ότι $Z = (X_i-\mu)\sigma^{-1} \sim N(0,1)$ και $Z^2 \sim X_1^2$

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim X_n^2 = \Gamma\left(\frac{n}{2}, 2\right)$$

Άρα $E\left(\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right) = n$ $\quad V\left(\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right) = 2n.$

$$V\left[\sum_{i=1}^{n}(X_i-\mu)^2\right] = 2n\sigma^4.$$

$$R(\theta, d_1) = \frac{1}{n^2}2n\theta^2 + \frac{1}{n}n\theta - \theta^2 = \frac{2\theta^2}{n}.$$

$$R(\theta, d_2) = V\left[\frac{k}{n}\sum(X_i-\mu)^2\right] + \left[\frac{k}{n}E\left\{\sum(X_i-\mu)^2\right\} - \theta\right]^2$$

$$= \frac{k^2}{n^2}2n\theta^2 + \left[\frac{k}{n}n\theta - \theta\right]^2 = \frac{2\theta^2 k^2}{n} + (k-1)^2\theta^2.$$

$$R(\theta, d_2) - R(\theta, d_1) = \left[\frac{2\theta^2 k^2}{n} + (k-1)^2 - \frac{2}{n}\right]\theta^2.$$

έστω: $f(k,n) = \frac{2\theta^2 k^3}{n} + (k-1)^2 - \frac{2}{n} =$

$= \frac{1}{n}\left[(2+n)k^2 - 2nk + (n-2)\right]$

η οποία είναι 2ου βαθμού ως προς k και έχει δύο ρίζες 1 και $\frac{n-2}{n-1} < 1$

Άρα για $k \in \left(\frac{n-2}{n-1}, 1\right)$ $f(k,n) < 0$ οπότε

$R(\theta, d_2) < R(\theta, d_1)$ καλύτερος ο d_2

Αντίθετα για $k \in \left[0, \frac{n-2}{n-1}\right) \cup (1, +\infty)$ $f(k,n) > 0$

καλύτερος ο d_1