Παρασκευή 27/11/20

Η προσέγγιση κατά Bayes δίνει τη δυνατότητα να μεταφέρουμε στην εκ των προτέρων κατανομή την προσωπική μας άποψη για την σημαντικότητα των πιθανών τιμών της άγνωστης παραμέτρου.
Θα μπορούσαμε για κάθε κανόνα απόφασης θα για κάθε κανόνα απόφασης να σταθμίζουμε τη συνάρτηση κινδύνου του βάσει της εκ των προτέρων προσωπικής μας άποψης.

Στη στατιστική θεωρία κατά Bayes η αξιολόγηση ενός κανόνα αποφάσεων γίνεται βάση του κινδύνου Bayes

Συνάρτηση Κινδύνου Bayes: (Κίνδυνος του Bayes)
Αν d ένας κανόνας αποφάσεων τότε η ποσότητα
σε συνεχή χώρο

$$r_p(d) = E\{R(\theta, d)\} = \int\int_{\theta \cdot \underline{X}} L(\theta, d(x_1, \dots x_n)) f(x_1 \dots x_n|\theta) p(\theta) dx_1 \dots dx_n d\theta$$

ολοκλήρωμα n+1 διαστάσεων: Κάθε μια από τις παρατηρήσεις κάτι που λέει, επομένως πρέπει να της δίνω όλες υπόψην και

- Ο $r_p(d)$ δεν εξαρτάται από την άγνωστη παράμετρο θ ούτε από τα δεδομένα. Ο δείκτης p. έχει προστεθεί στο συμβολισμό για να τονίσει τον ρόλο της εκ των προτέρων κατανομής.

- Το τελικό αποτέλεσμα δεν είναι τ.μ.

- Σε κάθε κανόνα απόφασης d αντιστοιχεί μοναδική τιμή κινδύνου Bayes. $r_p(d)$

Αν $d_1 < d_2$ δυο κανόνες με $r_p(d_1) < r_p(d_2)$ τότε ο d_1 πρέπει να προτιμηθεί

Ορισμός: Κανόνας Bayes

Ένας κανόνας αποφάσεων d^* που ελαχιστοποιεί τον κίνδυνο Bayes λέγεται κανόνας Bayes (για την συγκεκριμένη συνάρτηση απώλειας και εκ των προτέρων κατανομή)

Δηλαδή $\boxed{r_p(d^*) = \min_d r_p(d)}$

Μετά την παρατήρηση του δείγματος ένας κανόνας απόφασης οδηγεί σε μια συγκεκριμένη απόφαση, βασισμένη στο παρατηρημένο δείγμα $X_1, \dots X_n$.

Bayes: Το δείγμα είναι συγκεκριμένο (μη τυχαίο) και η παράμετρος είναι τ.μ.

- Στην στατιστική κατά Bayes, οι αποφάσεις αξιολογούνται βάσει της εκ των υστέρων αναμενόμενης απώλειας (αφού έχει παρατηρηθεί το δείγμα).

Ορισμός: Εκ των υστέρων αναμενόμενη απώλεια απόφασης.

$$r_p^*(d|X_1,\dots X_n) = E_p^*\{L(\theta,d)|X_1,\dots X_n\} =$$

$$\boxed{= \int_{\Theta} L(\theta,d)\, p(\theta|X_1,\dots X_n)\, d\theta}$$

Για γνωστά δεδομένα η $r_p^*(d|X_1,\dots X_n)$ είναι συνάρτηση μόνο του $d \in D$.
Το ζητούμενο είναι η ελαχιστοποίηση της ποσότητας αυτής ως προς d.

Ορισμός Απόφαση Bayes → δεν είναι τ.μ.
Η απόφαση $d_p(X_1,\dots X_n)$ που ελαχιστοποιεί την εκ των υστέρων αναμενόμενη απώλεια $r_p^*(d|X_1,\dots X_n)$ λέγεται απόφαση Bayes. Για τη συγκεκριμένη συνάρτηση απώλειας και εκ των προτέρων κατανομή.

Όπως γνωρίζουμε από τη Θ. Αποφάσεων, όταν η απόφαση που καλούμαστε να πάρουμε είναι ~~αυτό~~ η εκτίμηση της $g(\theta)$, $\theta \in \Theta$, $g: \Theta \to g(\Theta)$ τότε ο χώρος αποφάσεων είναι ο $D = g(\Theta)$ και οι κανόνες αποφάσεων είναι οι εκτιμητές της $g(\theta)$.

Έτσι στην Θ. Α. Bayes έχουμε
Ορισμός: Εκτιμητής Bayes:
Ο κανόνας Bayes για το πρόβλημα της εκτίμησης της $g(\theta)$ λέγεται εκτιμητής Bayes για την συγκεκριμένη συνάρτηση απώλειας και την εκ των προτέρων κατανομή.

▷ Το πρόβλημα ~~Bayes~~ εύρεσης εκτιμητή Bayes ανάγεται στην ελαχιστοποίηση της συνάρτησης $r_p^*(d|X_1,\dots X_n)$ ως ~~προ~~ προς D.

Θεώρημα: Έστω $X_1, \dots X_n$ τ.δ από πληθυσμό με κατανομή $f(x|\theta)$ και $p(\theta)$ η εκ των προτέρων κατανομή του θ. ~~[illegible]~~
Αν $L(\theta,d) = L(g(\theta), d(X_1, \dots X_n))$ είναι η συνάρτηση απώλειας για την εκτίμηση της $g(\theta)$, τότε ο εκτιμητής Bayes της $g(\theta)$ είναι το στατιστικό d^* (ως συνάρτηση του d) που ελαχιστοποιεί την εκ των υστέρων αναμενόμενη απώλεια της απόφασης d.

Απόδειξη:
Το d^* είναι εκείνη η τιμή του d που ελαχιστοποιεί το $r_p(d)$ δηλ.

$$r_p(d^*) = \min_d r_p(d) = \min_d \iint_{\theta\, X} L(\theta, d(x_1, \dots x_n)) f(x_1, \dots x_n|\theta) p(\theta) dx_1 \dots dx_n d\theta$$

$$= \min_d \iint_{\theta\, X} L(\theta, d(x_1, \dots x_n)) f(x_1, \dots x_n|\theta) p(\theta|x_1, \dots x_n) m(x_1, \dots x_n) d(x_1 \dots dx_n d\theta$$

$$= \min_d \int_X \left\{ \int_\theta L(\theta, d(x_1, \dots x_n)) f(x_1, \dots x_n|\theta) p(\theta|x_1, \dots x_n) m(x_1, \dots x_n) d\theta \right\} dx_1 \dots dx_n$$

Η παραπάνω ελαχιστοποίηση ισοδυναμεί με ελαχιστοποίηση της ποσότητα $\int_\theta L(\theta, d(x_1, \dots x_n)) f(x_1, \dots x_n|\theta) p(\theta|x_1, \dots x_n) m(x_1, \dots x_n)\, d\theta$

Η οποία είναι η εκ των υστέρων αναμενόμενη απώλεια της απόφασης d.

Επομένως για τον εκτιμητή Bayes ~~[illegible]~~ δεν ελαχιστοποιούμε τον κίνδυνο Bayes αλλά ισοδύναμα την εκ των υστέρων αναμενόμενη ~~[illegible]~~ απώλεια της απόφασης d.

- αν ως συνάρτηση απώλειας έχουμε το ΜΤΣ τότε εκτιμητής Bayes: η αναμενόμενη τιμή της εκ των υστέρων κατανομής.

- αν έχουμε το απόλυτο σφάλμα τότε έχουμε εκτιμητής Bayes: διάμεσος εκ των υστέρων κατανομής

Διάμεσος:
κατανομής της τ.μ. Y λέγεται κάθε πραγματικός αριθμός μ για τον οποίο ισχύει $P(Y \leq \mu) \geq 1/2$. και $P(Y \geq \mu) \geq 1/2$.
Γενικά η διάμεσος δεν είναι μοναδική
Σε συμμετρική κατανομή Διάμεσος = επικρατούσα τιμή = μέση τιμή

Πρόταση: Εκ των υστέρων αναμενόμενη τιμή της $g(\theta)$.
Αν η συνάρτηση απώλειας είναι το τετραγωνικό σφάλμα τότε ο εκτιμητής Bayes της $g(\theta)$ είναι ο:

$$d^* = E\{g(\theta)|X_1, \dots X_n\} = \int_\Theta g(\theta)\, p(\theta|X_1, \dots X_n)\, d\theta.$$

Απόδειξη:
Για ευκολία έστω $\underline{X} = (X_1, \dots X_n)$ Απο το Θεώρημα, ο d^* προκύπτει απο ελαχιστοποίηση ως προς d της

$$\int_\Theta L(g(\theta), d(X))\, p(\theta|X)\, d\theta. \quad \text{όπου } L(\theta, d) = L(g(\theta), d(X)) = \{g(\theta) - d(X)\}^2$$

όμως $$\frac{\partial}{\partial d}\int_\Theta L(g(\theta), d(X))\, p(\theta|X)\, d\theta = 0 \Rightarrow$$

$$\Rightarrow \int_\Theta \frac{\partial}{\partial d}\{g(\theta) - d(X)\}^2 p(\theta|X)\, d\theta = 0 \Rightarrow \cancel{\int_\Theta g(\theta) p(\theta|X)} = \int$$

$$\Rightarrow -2\int_\Theta \{g(\theta) - d(X)\}\, p(\theta|X)\, d\theta = 0 \Rightarrow$$

$$\Rightarrow \int_{\Theta} g(\theta) p(\theta|x) d\theta = \int_{\Theta} d(x) p(\theta|x) d\theta \Rightarrow$$

$$\Rightarrow d(x) \overbrace{\int_{\Theta} p(\theta|x) d\theta}^{=1} = E\{g(\theta)|x\} \Rightarrow d^* = E\{g(\theta)|x\}$$

Ακόμα $\frac{\partial^2}{\partial d^2} \int_{\Theta} \{g(\theta) - d(x)\}^2 p(\theta|x) d\theta = 2\int_{\Theta} p(\theta|x) d\theta$

$= 2 > 0$ άρα d^* ελάχιστο

Πρόταση: Εκ των υστέρων διάμεσος της $g(\theta)$.
Αν η συνάρτηση απώλειας είναι το απόλυτο σφάλμα $L(g(\theta), d(X_1,\dots X_n)) = |g(\theta) - d(X_1,\dots X_n)|$. τότε ένας εκτιμητής Bayes d^* της $g(\theta)$ είναι μια διάμεσος της $g(\theta)$ ως προς την εκ των υστέρων κατανομή.

Απόδειξη:
Έστω M μια διάμεσος της εκ των υστέρων κατανομής της $g(\theta)$ και d μια εκτίμηση της $g(\theta)$.
Χ.β.τ.γ υποθέτουμε ότι $d<M$ τότε

$$L(\theta,d) - L(\theta,M) = |g(\theta) - d| - |g(\theta) - M| = \begin{cases} d-M, & g(\theta) \le d \\ 2g(\theta) - d - M, & d < g(\theta) \le M \\ M-d, & g(\theta) \ge M \end{cases}$$

και έχουμε $L(\theta,d) - L(\theta,M) \ge (d-M) I_{(-\infty,M]}(g(\theta)) + (M-d) I_{(M,+\infty)}(g(\theta))$
όπου η ισότητα ισχύει στην περίπτωση (i) και (iii)
ενώ για την περίπτωση 2 ισχύει
$g(\theta) > d \Rightarrow 2g(\theta) - d > d \Rightarrow 2g(\theta) - d - M > d - M$.

Η διαφορά των εκ των υστέρων αναμενόμενων απωλειών ισούται με:

$r_p(d|X) - r_p(M|X) \geq (d-M) E_p\{ I_{(-\infty, M)}(g(\theta)|X)\} +$
$+ (M-d) E_p\{I_{(M,+\infty)}(g(\theta)|X)\}$
$= (d-M) P_p(g(\theta) < M|X) + (M-d) P_p(g(\theta) \geq M|X)$
$= (M-d)\{ P_p(g(\theta) \geq M|X) - P_p(g(\theta) < M|X)\}$
και $r_p(d|X_1, \ldots X_n) = r_p(M|X)$
αφού $p(g(\theta) \geq M|X) \geq 1/2 \geq P(g(\theta) \geq M|X)$.

Άρα $d^* = M$

Παράδειγμα 1:
Είναι δεδομένο ότι αν έχουμε μια μόνο παρατήρηση X από την κατανομή $N(\theta, 1)$ στην κλάση των κανόνων αποφάσεων $d_c(X) = c \cdot X$ και με συνάρτηση απώλειας το τετραγωνικό σφάλμα, η συνάρτηση κινδύνου είναι $R(\theta, d_c) = c^2 + (1-c)^2\theta^2$. Σε αυτή τη περίπτωση ο d_1 είναι R-καλύτερος όταν $c \geq 1$ ενώ για $0 \leq c \leq 1$ οι κανόνες είναι μη συγκρίσιμοι.
Ποιος είναι ο κανόνας Bayes αν θεωρήσουμε την εκ των προτέρων κατανομή $\theta \sim N(0, \tau^2)$;

Λύση: Βάση του ορισμού ο κίνδυνος Bayes δίνεται από τον τύπο $r_p(d) = E_p\{R(\theta, d_c)\} = E_p(c^2 + (1-c)^2\theta^2)$
$= c^2 + (1-c)^2 E_p(\theta^2) = c^2 + (1+c)^2\{Var(\theta) + [E(\theta)]^2\} =$
$= c^2 + (1+c)^2(\tau^2 + 0) = c^2 + (1+c)^2\tau^2$.
Για την δοθείσα εκ των προτέρων κατανομή $p(\theta)$ ο κανόνας Bayes, θα είναι ο $d^* = d_{c_0}$, όπου $c = c_0$ η τιμή που ελαχιστοποιεί τον κίνδυνο Bayes.

$$\frac{\partial r_p(d_c)}{\partial c} = 0 \Rightarrow 2c - 2(1-c)\tau^2 = 0 \Rightarrow$$

$$\Rightarrow c_0 = \frac{\tau^2}{1+\tau^2}$$

Επιβεβαιώνουμε ότι στο c_0 ο κίνδυνος Bayes παρουσιάζει ελάχιστο καθώς $\frac{\partial^2 r_p(d_c)}{dc^2} = 2+2\tau^2 > 0$

Έτσι για την εν λόγω ~~σε~~ εκ των προτέρων κατανομή ο κανόνας Bayes είναι ο $d^* = \frac{\tau^2}{1+\tau^2} X$.

και ο σχετικός κίνδυνος Bayes είναι ίσος με

$$r_p(d^*) = c_0^2 + (1-c_0)^2\tau^2 = \left(\frac{\tau^2}{1+\tau^2}\right)^2 + \left(\frac{\tau^2}{1+\tau^2}\right)^2 \tau^2 =$$

$$= \frac{\tau^4}{(1+\tau^2)^2} + \frac{\tau^6}{(1+\tau^2)^2} = \frac{\tau^4(1+\tau^2)}{(1+\tau^2)^2} = \frac{\tau^4}{1+\tau^2}$$

Παράδειγμα 2:

Αν η τιμή μιας διωνυμικής τ.μ X, δηλ $X \sim B(n,\theta)$ όπου παράμετρος ενδιαφέροντος είναι η πιθανότητα επιτυχίας θ και έστω Βήτα η εκ των προτέρων κατανομή δηλ $\theta \sim B(a,b)$. Ο εκτιμητής Bayes για συνάρτηση απώλειας το ΤΣ είναι η μέση τιμή της εκ των υστέρων κατανομής και γράφεται ως κυρτός συνδυασμός του συνήθους εκτιμητή του θ, δηλ $\hat{\theta} = X/n$. και της αρχικής (εκ των προτέρων) εκτίμησης του $\frac{a}{a+b}$.

Ποιος θα ήταν ο εκτιμητής Bayes της $g(\theta) = \theta(1-\theta)$

Για τη συνάρτηση απώλειας το τετραγωνικό σφάλμα από το ~~[illegible]~~ γνωστή πρόταση ο εκτιμητής Bayes θα είναι η εκ των υστέρων μέση τιμή της $g(\theta)$

δηλ. $E_p\{g(\theta)|x\} = \int_\theta \theta(1-\theta)\, p(\theta|x)\, d\theta$

$$p(\theta|x) = \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(b+n-x)} \theta^{a+x-1}(1-\theta)^{b+n-x-1}, \quad 0<\theta<1$$

αφού $\theta|x \sim Beta(a+x, b+n-x)$

οπότε $E_p\{g(\theta)|X\} = \int_0^1 \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(b+n-x)} \theta^{a+x}(1-\theta)^{b+n-x} d\theta$.

$= \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(b+n-x)} \int_0^1 \theta^{(a+x+1)-1}(1-\theta)^{(b+n-x+1)-1} d\theta$

$= \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(b+n-x)} \cdot \frac{\Gamma(a+x+1)\Gamma(b+n-x+1)}{\Gamma(a+b+n+2)} = \frac{(a+x)(b+n-x)}{(a+b+n)(a+b+n+1)}$

Παράδειγμα 3:

Έστω $X_1, \ldots X_n$ τ.δ από $N(\theta,1)$. Να βρεθεί ο εκτιμητής Bayes του $\theta \sim N(0,1)$ και η συνάρτηση απώλειας είναι $L(\theta,d) = (\theta-d)^2$, δηλ το τ.σ.

Αφού η συνάρτηση απώλειας είναι το τετραγωνικό σφάλμα έχουμε ότι η μέση τιμή της εκ των υστέρων κατανομής της θ θα είναι ο εκτιμητής Bayes.

$\theta | X_1 \ldots X_n \sim N\left(\frac{n\bar{X}}{n+1}, \frac{1}{n+1}\right)$

άρα $d^* = E(\theta | X_1 \ldots X_n) = \frac{n\bar{X}}{n+1}$.

Παράδειγμα 4:

Έστω $X_1, \ldots X_n$ τ.δ από $N(\theta, \sigma^2)$ με $\theta \in \Theta = \mathbb{R}$ και $\sigma^2 \geq 0$ γνωστή. Να βρεθεί ο εκτιμητής Bayes του θ αν $\theta \sim N(0, \tau^2)$ και η συνάρτηση απώλειας είναι το τετραγωνικό σφάλμα.

Η συνάρτηση πιθανοφάνειας του δείγματος ισούται με

$f(X_1 \ldots X_n | \theta) = \prod_{i=1}^n \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\{-1/2\sigma^2 (x_i-\theta)^2\}\right] =$

$= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\theta)^2\} = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\{-\frac{1}{2\sigma^2} \sum (x_i-\bar{x})^2 + n(\bar{x}-\theta)^2\}$

οπότε τελικά έχουμε

$$f(x_1 \dots x_n|\theta) \propto \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right\}$$

Η εκ των προτέρων κατανομή είναι:

$$p(\theta) = \frac{1}{\tau\sqrt{2\pi}} \exp\left\{-\frac{1}{2\tau^2}(\theta-\mu)^2\right\}$$

Για την εκ των υστέρων κατανομή ισχύει

$$p(\theta|x_1 \dots x_n) \propto f(x_1 \dots x_n|\theta)\, p(\theta) \propto \exp\left\{\frac{-n}{2\sigma^2}(\bar{x}-\theta)^2\right\} \exp\left\{\frac{-1}{2\tau^2}(\theta-\mu)^2\right\}$$

$$= \exp\left\{-\frac{n\bar{x}}{2\sigma^2} - \frac{n\theta}{2\sigma^2} + \frac{n\bar{x}\theta}{2\sigma^2} - \frac{\theta^2}{2\tau^2} - \frac{\mu^2}{2\tau^2} + \frac{2\theta\mu}{2\tau^2}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\theta^2 \cancel{\frac{1}{2}}\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu}{\tau^2}\right)\theta\right\} =$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{n\tau^2+\sigma^2}{\tau^2\sigma^2}\right)\left(\theta^2 - 2\theta\,\frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{n\tau^2+\sigma^2}{\tau^2\sigma^2}\right)\left[\theta^2 - 2\theta\,\frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2} + \left(\frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{n\tau^2+\sigma^2}{\tau^2\sigma^2}\right)\left[\theta^2 - \left(\frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}\right)^2\right]\right\}$$

Επειδή το μεταβλητό μέρος μιας πυκνότητας καθορίζει μονοσήμαντα την κατανομή αναγνωρίζουμε ότι το παραπάνω μεταβλητό μέρος αντιστοιχεί σε κανονική κατανομή με μέση τιμή $\frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}$ και διακύμανση $\frac{\tau^2\sigma^2}{n\tau^2+\sigma^2}$

άρα $\theta|x_1 \dots x_n \sim N\left(\frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2}, \frac{\tau^2\sigma^2}{n\tau^2+\sigma^2}\right)$

και έχουμε ως συνάρτηση απώλειας το τ.σ άρα ο εκτιμητής Bayes είναι ο

$$d_B(X_1 \dots X_n) = \frac{n\tau^2 \bar{X} + \sigma^2 \mu}{n\tau^2+\sigma^2} = \frac{n\tau^2}{n\tau^2+\sigma^2}\bar{X} + \frac{\sigma^2}{n\tau^2+\sigma^2}\mu .$$

δηλαδή ένας κυρτός συνδυασμός του εκτιμητή μέγιστης πιθανοφάνειας της παραμέτρου θ, $\hat{\theta} = \bar{X}$ και της μέσης τιμής μ της εκ των προτέρων κατανομής.