Σημαντικά πράγματα να θυμόμαστε (1)

- Τυχαίες Μεταβλητές

διακριτές | συνεχείς

συνάρτηση πιθανότητας | συνάρτηση πυκνότητας πιθανότητας

$$E(X) = \sum_{x \in A} x \cdot \underbrace{P(X=x)}_{f(x)} \qquad E(X) = \int_{\mathbb{R}} x \cdot f_X(x)\, dx$$

A: αριθμήσιμο σύνολο $\qquad$ f_X: σ.π.π.

$$V(X) = E(X-\mu)^2 = E(X^2) - E^2(X)$$

$\mu \to E(X)$

\+ Ιδιότητες μέσης τιμής + διασποράς

π.χ. $E(aX+b) = aE(X)+b$, $V(aX+b) = a^2 V(X)$.

Όταν έχουμε X, Y τ.μ.

$$V(X+Y) = V(X) + V(Y) + 2\,Cov(X,Y)$$

$$V(X-Y) = V(X) + V(Y) - 2\,Cov(X,Y)$$

ανεξάρτητες $\Longrightarrow$ ασυσχέτιστες

$\not\Longleftarrow$ $\qquad\qquad\Updownarrow$

$$Cov(X,Y) = 0.$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)\,V(Y)}}$$

$\downarrow$

συντελεστής γραμμικής συσχέτισης

$$|\rho(X,Y)| \le 1$$

$\rho(X,Y) = -1 \rightarrow$ πλήρως αρνητική συσχ.

$\rho(X,Y) = 0 \Longleftrightarrow$ ασυσχέτιστες τ.μ.

$\rho(X,Y) = 1 \rightarrow$ πλήρως θετική συσχ.

$Y \overset{\rho=1}{=} aX + b$ ($a>0$), $\quad Y \overset{\rho=-1}{=} aX + b$ ($a<0$).

3

Όλο το πρώτο κομμάτι περί υπολογισμού πιθανοτήτων είναι χρήσιμο για να εφαρμόζεται σε συνδυασμό με τυχαίες μεταβλητές.

π.χ. συνδυασμούς και διατάξεις.
μας βοηθά στους υπολογισμούς πιθανοτήτων.

Τα είδαμε σε σχέση με τη διωνυμική κατανομή που εμφανίζεται ο διωνυμικός συντελεστής

Βασικές Ιδιότητες πιθανότητας — Θεώρημα Ολικής Πιθανότητας — Τύπος του Bayes (δεσμευμένη πιθανότητα)

4

Ανεξαρτησία Τυχαίων Μεταβλητών.
και συνέπειες

π.χ. $f(x,y) = f(x)\cdot f(y)$.

$$E(XY) = E(X)\,E(Y)$$

$$V(X+Y) = V(X) + V(Y)$$

Βασικά είδη κατανομών.

Διακριτές (Στήριγμα / $E(X)$, $V(X)$) Συνεχείς

$Be(p)$

Διακριτή Ομοιόμορφη.

$Geo(p)$ [2 μορφές].

$Bin(n, p)$

$P(\lambda)$ [Poisson].

$NegBin(n,p)$ (ή Pascal).

$\boxed{\text{+ σχέση με ακολουθία ανεξάρτητων δοκιμών } Be(p).}$

$Exp(\theta)$, (2 εκφράσεις) → ρυθμού/ κλίμακας

$\mathcal{U}(\alpha, \beta)$ ή $\mathcal{U}[\alpha, \beta]$

$G(\alpha, \theta)$ [2 μορφές]

$N(\mu, \sigma^2)$

Εφαρμογές στη Στατιστική

$\chi^2_n \equiv G\left(\frac{n}{2}, \frac{1}{2}\right)$ ή $G\left(\frac{n}{2}, 2\right)$

t_n (+ αναπαράστασης)

(1) → Ισχυρός Νόμος των Μεγάλων Αριθμών.

$$\overline{X}_n \xrightarrow{a.s.} \mu.$$

(2) → Ασθενής Νόμος των Μεγάλων Αριθμών.

$$\overline{X}_n \xrightarrow{P} \mu.$$

(3) → Κεντρικό Οριακό Θεώρημα.

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) \xrightarrow{d.} N(0,1)$$

Συγκλίσεις Ακολουθιών Τυχαίων Μεταβλητών.

(1) → $X_n \xrightarrow{a.s.} X$

→ $X_n \xrightarrow{P} X$

→ $X_n \xrightarrow{d} X$

Παραμετρική Στατιστική

Μοντελοποίηση με άγνωστες παράμετρο

$\{f(x;\theta)\}$ → θ άγνωστο
θέλουμε να το εκτιμήσουμε.

$x_1 \ldots x_n$

↓

$X_1 \ldots X_n$ τυχαίο δείγμα → $\{f(x;\theta)\}$ έχει σ.π. ή σ.π.π.

(παραμετρική οικογένεια.

Εκτιμήτριες Παραμέτρων

Τι είναι εκτιμήτρια ?

Μεθόδους εύρεσης εκτιμητριών.

Εκτιμήτριες ροπών

$\boxed{\text{Εκτιμήτριες Μέγιστης Πιθανοφάνειας}}$

→ Ιδιότητες Εκτιμητριών:

- μεροληψία (κοιτάμε αμεροληψία)
- ΜΤΣ (Μέσο Τετραγωνικό Σφάλμα). (θέλουμε να το ελαχιστοποιήσουμε)

Συγκρίνουμε εκτιμήτριες με βάση μεροληψία + ΜΤΣ

- Συνέπεια (απαραίτητη)

→ Κατασκευή Διαστημάτων Εμπιστοσύνης για μία παράμετρο.

$(1-\alpha)$ - Δ.Ε. [συντελεστής].

Τυχαία διαστήματα τα οποία θέλουμε να περιλαμβάνουν την άγνωστη παράμετρο με δεδομένη πιθανότητα.

⇒ πιο σημαντικά αυτά που συνδέονται με την κανονική κατανομή για μέση τιμή & διασπορά (προσοχή σα άνω ποσοστημόρια).

→ Έλεγχος Υποθέσεων

Μια υπόθεση κινητοποιεί τη διεξαγωγή ενός ελέγχου υποθέσεων, στη βάση του οποίου θα αποφασίσουμε αν τα δεδομένα του προβλήματος δίνουν ισχυρές ενδείξεις για την απόρριψη της

H_0 : η υπόθεση "status quo" → αν δεν είχαμε δεδομένα. τη δεχόμαστε

vs H_1 : η υπόθεση που μας ενδιαφέρει και η ισχύ της επιφέρει κάποιες αλλαγές, κάποια μέτρα που πρέπει να πάρουμε.

ε.σ.σ. (επίπεδο στατιστικής σημαντικότητας) (12)

που ελέγχει την πιθανότητα σφάλματος τύπου I

, δηλ. λανθασμένη απόρριψη της H_0.

→ 2 είδη σφαλμάτων → τύπου I
→ τύπου II (λανθασμένη αποδοχή της H_0!)

Ισχύς του ελέγχου, συνάρτηση ισχύος.

→ Βασικότεροι έλεγχοι που σχετίζονται

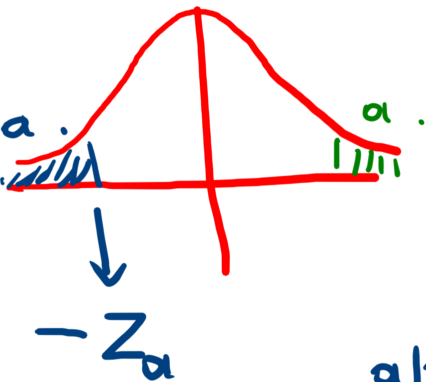
με τον έλεγχο μέσης τιμής της κανονικής κανονικής.

με γνωστή ή άγνωστη διασπορά.

Τα άνω ποσοστημόρια.

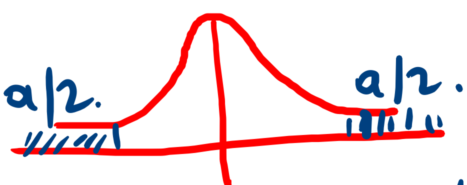
Η βασική ελεγχοσυνάρτηση η Z και η T αντίστοιχα.

↳ είδος της κρίσιμης περιοχής.



$H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$

vs $H_1: \mu > \mu_0$



$H_1: \mu \neq \mu_0$.

+ ισοδυναμία διεξαγωγής $|Z| > Z_{\frac{a}{2}}$. (όλα αυτά θα γίναν T και $t_{n-1,a}$ ή $t_{n-1,\frac{a}{2}}$ έχουμε άγνωστη διασπορά

του ελέγχου με p-value

↳ αξιολογεί τη σημαντικότητα των όποιος λάβουν