1º Μάθημα: Παρασκευή 06/03/20

Αλλαγή μεταβλητών:

Πρόβλημα: Έστω τ.μ X με γνωστή κατανομή. Ποιά είναι η κατανομή της $Y=g(X)$?

Μέθοδος του μετασχηματισμού:

Πρόταση: Έστω συνεχής τ.μ X με σ.π.π f_X και σύνολο τιμών της X το $I\subseteq\mathbb{R}$. Θεωρούμε τον μετασχηματισμό $g: I\to g(I)=\{y: y=g(x), x\in I\}$ για το οποίο:

(i) Η g είναι "1-1" απο το I στο $g(I)$.

(ii) Η $\frac{d}{dy}(y^{-1})\neq 0$ και συνεχής

Τότε $f_Y(y)=f_X(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right|$

Παράδειγμα: Αν $X\sim G(k,\lambda)$, $k,\lambda>0$ τοτε $Y=\frac{2}{\lambda}X\sim X^2_{2k}$

$\left(\frac{2}{\lambda}G(k,\lambda)\equiv X^2_{2k}, \quad G(k,\lambda)\equiv\frac{\lambda}{2}X^2_{2k}\right)$

Θεωρούμε τον μετασχηματισμό $Y=g(X)=\frac{2}{\lambda}X$, $X>0$

Αρα $y>0$ $\quad I=(0,+\infty)$ $\quad g(I)=(0,+\infty)$.

(i) Η g είναι "1-1"? ΝΑΙ γιατί: $g'(x)=\frac{2}{\lambda}>0\Rightarrow g\nearrow\Rightarrow g: 1-1$

$g(x_1)=g(x_2)\Rightarrow x_1=x_2\Rightarrow g$ '1-1'

Αφού η g είναι "1-1", $\exists g^{-1}$ και είναι:

$\cancel{g(y)=x \Rightarrow \frac{2}{\lambda}y=x}$ $\quad g(x)=y \Rightarrow \frac{2}{\lambda}x=y \Rightarrow x=\frac{\lambda}{2}y$

$\Rightarrow g^{-1}(x)=\frac{\lambda}{2}x$.

(ii) $\frac{dg^{-1}(x)}{dx}=\frac{\lambda}{2}\neq 0$ και συνεχής

Άρα $f_Y(y)=f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| =$

$$\left.\begin{aligned} &= f_X\left(\frac{\lambda}{2}y\right)\left|\frac{\lambda}{2}\right| = \frac{\lambda}{2}f_X\left(\frac{\lambda}{2}y\right) \\ &X\sim G(k,\lambda) \Rightarrow f_X(x)=\frac{1}{\lambda^k\Gamma(k)}x^{k-1}e^{-\frac{x}{\lambda}} \end{aligned}\right\} \Rightarrow$$

$$\left.\begin{aligned} &f_Y(y)=\frac{1}{2^k\Gamma(k)}y^{k-1}e^{-\frac{y}{2}}, \quad y>0 \\ &f_{\chi^2_{2k}}(y)=\frac{1}{2^{\frac{2k}{2}}\Gamma\left(\frac{2k}{2}\right)}y^{\frac{2k}{2}-1}e^{-\frac{y}{2}} \end{aligned}\right\} \Rightarrow f_Y(y)=f_{\chi^2_{2k}}(y)$$

$\Rightarrow Y\sim \chi^2_{2k}$.

Μέθοδος Α.Σ.Κ:

Π.χ.) Έστω συνεχής τ.μ. X με γνήσια αύξουσα α.σ.κ $F_X(x)$
Να βρεθεί η κατανομή της $Y = F_X(x)$

Αρκεί να βρώ την α.σ.κ Y.

Είναι $F_Y(y) \overset{\text{ορ.}}{=} P(Y \leq y) = P(F_X(x) \leq y) =$

$= P(F_X^{-1}(F_X(x)) \leq F_X^{-1}(y)) = P(X \leq F_X^{-1}(y)) \overset{\text{ορ.}}{=} F_X(F_X^{-1}(y)) = y.$

$f_Y(y) = \dfrac{d}{dy} F_Y(y) = 1, \quad y \in (0,1)$ γιατί $y = F_X(x) \Rightarrow$

$\Rightarrow Y \sim U(0,1).$

Κατανομή max και min τ.μ.:

Έστω τ.μ. $X_1, X_2, \dots, X_n$ ανεξάρτητες και ισόνομες (έχουν ακριβώς την ίδια κατανομή, έστω F_X).

Πρόβλημα: Ποια είναι η κατανομή της τ.μ. $X_{(n)} = \max X_i = \max\{X_1, X_2, \dots, X_n\}$?
Ποια είναι η κατανομή της τ.μ. $X_{(1)} = \min X_i = \min\{X_1, X_2, \dots, X_n\}$?

Κατανομή του $X_{(n)}$ (n-οστό διατεταγμένο στατιστικό)

$F_{X_{(n)}}(x) \overset{\text{ορ.}}{=} P(X_{(n)} \leq x) = P(\max\{X_1, \dots, X_n\} \leq x) =$

$= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \overset{X_i \text{ ανεξ.}}{=} \prod_{i=1}^{n} P(X_i \leq x) = \prod_{i=1}^{n} F_{X_i}(x) \overset{X_i \text{ ισονομ.}}{=}$

$= \prod_{i=1}^{n} F_X(x).$

$$\boxed{F_{X_{(n)}}(x) = [F_X(x)]^n}$$ Αθροιστική συνάρτηση κατανομής μεγίστου.

$$f_{X_{(n)}}(x) = \frac{dF_{X_{(n)}}(x)}{dx} = \frac{d}{dx}[F_X(x)]^n = n(F_X(x))^{n-1} \cdot f_X(x)$$

$$\boxed{f_{X_{(n)}}(x) = n\, f_X(x)\,[F_X(x)]^{n-1}}$$ Πυκνότητα του μεγίστου.

<u>Κατανομή $X_{(1)}$:</u>

$$F_{X_{(1)}}(x) \overset{\text{ορ.}}{=} P(X_{(1)} \le x) = P(\min\{X_1, X_2, \dots X_n\} \le x) =$$

$$= 1 - P(\min\{X_1, X_2, \dots, X_n\} \ge x) = 1 - P(X_1 \ge x, X_2 \ge x, \dots X_n \ge x) =$$

$$\overset{X_i \text{ ανεξ.}}{=} 1 - \prod_{i=1}^{n} P(X_i \ge x) = 1 - \prod_{i=1}^{n}[1 - P(X_i \le x)] =$$

$$= 1 - \prod_{i=1}^{n}[1 - F_{X_i}(x)] \overset{X_i \text{ ισόνομ.}}{=} 1 - \prod_{i=1}^{n}(1 - F_X(x))$$

$$\Rightarrow \boxed{F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n}$$ Αθροιστική συνάρτηση ελαχίστου

$$f_{X_{(1)}}(x) = \frac{dF_{X_{(1)}}(x)}{dx} = \frac{d}{dx}\{1 - (1 - F_X(x))^n\} =$$

$$= -n(1 - F_X(x))^{n-1} \frac{d}{dx}(1 - F_X(x)) = n(1 - F_X(x))^{n-1} f_X(x)$$

$$\Rightarrow \boxed{f_{X_{(1)}}(x) = n\, f_X(x)\,(1 - F_X(x))^{n-1}}$$ Πυκνότητα του ελαχίστου.

Κατανομή Αθροίσματος Ανεξάρτητων τ.μ.:

Έστω ανεξάρτητες τ.μ $X_1, X_2, \dots, X_n$. Έστω ότι είναι και ισόνομες οπότε έχουν την ίδια κατανομή (p_X, F_X, f_X η m_X).

Ποια η κατανομή του $T = \sum_{i=1}^{n} X_i$ (η $T = \sum_{i=1}^{n} a_i X_i$)?

Μέθοδος ροπογεννήτριας (κορυφαία στην διερεύνηση)

$$m_T(t) \overset{op}{=} E(e^{tT}) = E(e^{t\sum_{i=1}^{n} X_i}) = E(e^{t(X_1+X_2+\dots+X_n)}) = E(e^{tX_1} \cdot e^{tX_2} \cdot \dots \cdot e^{tX_n})$$

$$\overset{\text{ανεξ.}}{=} E(e^{tX_1}) \cdot E(e^{tX_2}) \cdot \dots \cdot E(e^{tX_n}) = \prod_{i=1}^{n} E(e^{tX_i}) = \prod_{i=1}^{n} m_{X_i}(t) \overset{\text{ισον.}}{=} \prod_{i=1}^{n} m_X(t)$$

$$\Rightarrow \boxed{m_T(t) = (m_X(t))^n}$$

Ακόμα και αν δεν είναι ροπογεννήτρια γνωστής κατανομής, μπορούμε να βρούμε την κατανομή καθώς υπάρχει τύπος που συνδέει την ροπογεννήτρια με την πυκνότητα.

Εαν είχαμε γραμμικό συνδυασμό τ.μ.
$$m_T(t) = E(e^{tT}) = E(e^{t\sum a_i X_i}) = E(e^{ta_1X_1} \cdot \dots \cdot e^{ta_nX_n}) =$$

$$= E(e^{ta_1X_1}) \cdot \dots \cdot E(e^{ta_nX_n}) = \prod_{i=1}^{n} m_{X_i}(ta_i) \overset{\text{ισον.}}{=} \prod_{i=1}^{n} m_X(ta_i)$$

<u>Παράδειγμα:</u> Έστω ανεξάρτητες τ.μ. $X_1, X_2, \dots, X_n$ και ισόνομες με κατανομή $N(\mu, \sigma^2)$. Ποια η κατανομή του $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$?

Αφού κάθε $X_i \sim N(\mu, \sigma^2)$ η ροπογεννήτρια $m_{X_i}(t) = m_X(t) =$
$= e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.

$m_{\bar X}(t) \overset{op.}{=} E(e^{t\bar X}) = E(e^{\frac{t}{n}x_1 + \dots + \frac{t}{n}x_n}) = E(e^{\frac{t}{n}x_1} \cdots e^{\frac{t}{n}x_n}) =$

$= \prod_{i=1}^{n} E(e^{\frac{t}{n}x_i}) = \prod_{i=1}^{n} m_{x_i}(\frac{t}{n}) = \prod_{i=1}^{n} m_x(\frac{t}{n}) =$

$= [m_x(\frac{t}{n})]^n = [e^{\frac{\mu t}{n} + \frac{1}{2}\sigma^2 (\frac{t}{n})^2}]^n =$

$= e^{\mu t + \frac{1}{2}\frac{\sigma^2}{n} t^2} \Rightarrow \bar X \sim N(\mu, \frac{\sigma^2}{n})$

Παραδείγματα / Ασκήσεις

Αν $X_1, \dots, X_n \sim Poisson(\theta)$. $T = \sum X_i \sim P(n\theta)$

Αν $X_1, \dots, X_n \sim Εκθ(1/\theta)$ $T = \sum X_i \sim G(n,\theta)$

Αν $X_1, \dots, X_n \sim Εκθ(\theta)$ $T = \sum X_i \sim G(n, 1/\theta)$

Αν $X_1 \dots X_n \sim Bernoulli(\theta) \equiv$ Διωνυμική με $n=1$. $T = \sum X_i \sim B(1,\theta)$

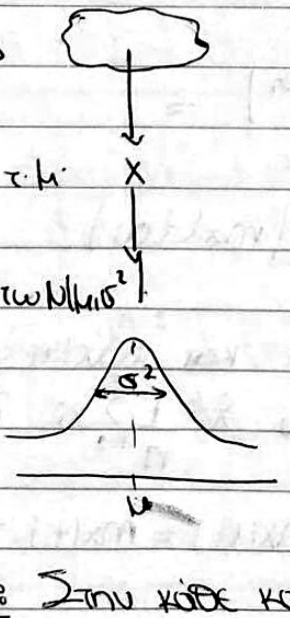
Στατιστική Συμπερασματολογία:

Πληθυσμός

Μια ή πολλές τ.μ. X

κατανομή έστω $N(\mu,\sigma^2)$

μ, σ^2 άγνωστα



Πρόβλημα: Στην κάθε κατανομή υπάρχουν άγνωστες παράμετροι τις οποίες πρέπει να βρω τρόπο να προσεγγίσω (εκτιμήσω).

Στατιστική Συμπερασματολογία: Μεθοδολογίες στατιστικές για την προσέγγιση (εκτίμηση) των αγνώστων παραμέτρων της κατανομής

1) Εκτίμηση σε σημείο
2) Εκτίμηση σε διάστημα
3) Στατιστικό Τεστ.

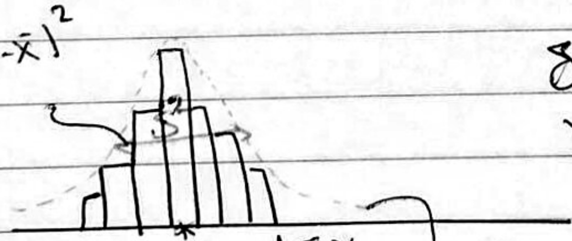
Πως μπορεί να προσεγγιστεί το πρόβλημα αυτό
Τι κάνει ο στατιστικός?

→ Διαλέγει ένα δείγμα: n- πειραματικές μονάδες όσο πιο αντιπροσωπευτικές γίνεται του πληθυσμού

→ Μετρώντας το χαρακτηριστικό X από τις n μονάδες που επέλεξα δημιουργώ το τυχαίο δείγμα: $X_1, X_2, \ldots X_n$
(X_i παριστάνει τη μέτρηση της X στην i-πειραματική μονάδα).

Τα $X_1, X_2, \ldots, X_n$ αξιοποιούνται για το ιστόγραμμα.

$\sigma^2 \approx S^2 = \frac{1}{n-1} \Sigma (X_i - \bar{x})^2$



δείχνει την κατανομή των δεδομένων $X_1, \ldots, X_n$.

$\mu \approx \bar{x} = \frac{1}{n} \Sigma X_i$

→ Παρατηρώ αν υπάρχει έστω συμμετρία ή γενικά αν μπορώ να προσεγγίσω την κατανομή μέσω γραφικής παράστασης.

Στατιστικό ή Στατιστική συνάρτηση:
Κάθε συνάρτηση του τυχαίου δείγματος (τ.δ) $X_1, \ldots, X_n$.