*Από την Πέμπτη 9/3/2017
μάθημα 12:00 - 14:40

Στατιστική Συμπερασματολογία

Σύγγραμμα: • Ηλιόπουλος $\longrightarrow$ 2/3 της ύλης περιέχει (εκτίμηση σε σημείο / -||- σε διάστημα)

Από Εύδοξο: • Παπαϊωάννου - Φερεντίνος "Μαθηματική Στατιστική" $\longrightarrow$ όλη την ύλη (εκτίμηση σε σημείο / -||- σε διάστημα / έλεγχος υποθέσεων)

και τα δύο:
+ Κουρούκλης κλπ (από Κάλλιπο σε pdf) $\longrightarrow$ 2/3 της ύλης
"Θέματα Παραμετρικής Στατιστικής Συμπερασματολογίας"

Αλλαγή μεταβλητών

Πρόβλημα: Έστω τυχαία μεταβλητή (τ.μ) X με γνωστή κατανομή
Ζητείται: Κατανομή της τ.μ. [redacted] $Y = g(X)$

Μέθοδος Μετασχηματισμού

Πρόταση: Έστω συνεχής τ.μ. X με σ.π.π. $f_X(x)$ και $x \in I \subseteq \mathbb{R}$.
Θεωρούμε το μετασχηματισμό $y = g(x)$ και υποθέτουμε:

Προϋποθέσεις:
(i) Ο g είναι 1-1 του I στο $g(I) = \{y : y = g(x),\ x \in I\}$
(ii) Η $\frac{d}{dy} g^{-1}(y)$ είναι συνεχής και $\neq 0$, $\forall y \in g(I)$

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

Παραδείγματα

Παραδ. 1: Έστω συνεχής τ.μ. X_{tr} με σ.π.π. $f_X(x) = \theta x^{-\theta-1}$, $x > 1$, $\theta > 0$
Να βρεθεί η κατανομή της τ.μ. $Y = \ln x$.

Θεωρώ το μετασχηματισμό $y = g(x) = \ln x$.
I: σύνολο τιμών της X: $I = (1, \infty)$ (αφού $x > 1$)
$g(I)$: σύνολο τιμών της Y: $g(I) = (0, +\infty)$

(I) Η g είναι 1-1 προφανώς (επειδή $g \uparrow$, αλλιώς για $x_1, x_2 \in I$ με $g(x_1) = g(x_2) \Rightarrow$
$\Rightarrow x_1 = x_2$)

$y = g(x) = \ln x \Rightarrow x = g^{-1}(y) = e^y$

(II) $\frac{d}{dy} g^{-1}(y) = \frac{d}{dy} e^{y} = e^{y}$ συνεχής και $e^{y} \neq 0$, $y \in (0, \infty)$

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\partial g^{-1}(y)}{\partial y} \right| = f_X(e^{y}) \left| e^{y} \right| = \theta (e^{y})^{-\theta-1} e^{y}$$

$$= \theta e^{-\theta y}, \ y > 0$$

Άρα $Y \sim Εκθ(\theta)$

Παραδ. 2: Αν $X \sim G(k, \lambda)$, τότε $Y = \frac{2}{\lambda} X \sim \chi^2_{2k}$

Συνοπτικά: $G(k, \lambda) \equiv \frac{\lambda}{2} \chi^2_{2k}$)

Θεωρώ το μετασχηματισμό $y = g(x) = \frac{2}{\lambda} x$

Προφανώς $I = (0, \infty) \xrightarrow{y = g(x) = \frac{2}{\lambda}x} g(I) = (0, \infty)$

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

(I) Η g είναι 1-1

(II) $\frac{d}{dy} g^{-1}(y) = \frac{\lambda}{2}$ συνεχής $\neq 0$

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

$$= f_X\left(\frac{\lambda}{2}, y\right) \left| \frac{\lambda}{2} \right|$$

Αφού $X \sim G(k, \lambda)$ $f_X(x) = \frac{1}{\lambda^{k} \Gamma(k)} x^{k-1} e^{-x/\lambda} =$

$$= \frac{1}{2^{2k/2} \Gamma(\frac{2k}{2})} y^{\frac{2k}{2} - 1} e^{-y/2}, \ y > 0 \Rightarrow$$

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

Μέθοδος α.σ.κ.

Παράδειγμα: Έστω τ.μ. X με ασκ $F_X(x)$
Να βρεθεί η κατανομή της τ.μ. $Y = F_X(x)$

Έστω F_Y η ασκ της τ.μ. Y που ζητάω να βρω.

Τιμές της τ.μ. X : $y = F_X(x) \overset{\text{ορ.}}{=} P(X \leq x)^{(*)}$

Άρα $y \in (0,1)$ είναι πιθανότητα το y λόγω $(*)$, άρα $0 < y < 1$.

$$
\begin{aligned}
F_Y(y) &\overset{\text{ορ.}}{=} P(Y \leq y) \\
&= P(F_X(x) \leq y) \\
&= P(F_X^{-1}(F_X(x)) \leq F_X^{-1}(y)) \\
&= P(X \leq F_X^{-1}(y)) \\
&\overset{\text{ορ.}}{=} F_X(F_X^{-1}(y)) \\
&= y, \quad y \in (0,1)
\end{aligned}
$$

Θυμόμαστε αν
$W \sim U(a, b)$ τότε

$$
f_W(w) = \begin{cases} 1/(b-a) , & a < w < b \\ 0 , & \text{αλλού} \end{cases}
$$

$$
F_W(w) = \begin{cases} 0 , & w \leq a \\ \dfrac{w-a}{b-a} , & a < w < b \\ 1 , & w \geq b \end{cases}
$$

Συγκρίνοντας την $f_Y(y)$ με την F_W, $W \sim U(a,b) \Rightarrow \boxed{Y = F_X(x) \sim U(0,1)}$

Κατανομή min και max τ.μ.

Έστω ανεξάρτητες και ισόνομες τ.μ. $X_1, \dots, X_n$

Ισόνομες $\equiv$ οι $X_1, \dots, X_n$ έχουν την ίδια κατανομή, διέπονται δηλαδή από τον ίδιο νόμο.

Έστω $X_{(1)} = \min\{X_1, \dots, X_n\}$

$X_{(n)} = \max\{X_1, \dots, X_n\}$

Νδο

① $f_{X_{(n)}}(x) = n\left[F_X(x)\right]^{n-1} f_X(x)$

② $f_{X_{(1)}}(x) = n\left[1 - F_X(x)\right]^{n-1} f_X(x)$

πολύ σημαντικοί

όπου f_X και F_X η κοινή σ.π.π. και α.σ.κ. των $X_1, \dots, X_n$.

Απόδειξη:

Απόδειξη ①

Εφαρμόζω μέθοδο α.σ.κ.

$$F_X(x) \overset{\text{ορ}}{=} P(X_{(n)} \le x) = P(\max\{X_1, \dots, X_n\} \le x) =$$

$$= P(X_i \le x, \forall i = 1, \dots, n) = P(X_1 \le x, X_2 \le x, \dots, X_n \le x) =$$

$$\underset{X_1, X_2, \dots, X_n}{\overset{\text{ανεξαρτησία:}}{=}} P(X_1 \le x) \dots P(X_n \le x)$$

$$\Rightarrow F_{X_{(n)}}(x) = \prod_{i=1}^{n} P(X_i \le x) \underset{\text{α.σ.κ}}{\overset{\text{ορ}}{=}} \prod_{i=1}^{n} F_{X_i}(x) = \prod_{i=1}^{n} F_X(x) = \left[F_X(x)\right]^n$$

$$f_{X_{(n)}}(x) = \frac{d}{dx} F_{X_{(n)}}(x)$$

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

Απόδειξη ②

$$F_{X_{(1)}}(x) \overset{op}{=} P(X_{(1)} \leq x)$$

$$= P(\min\{X_1, \dots, X_n\} \leq X)$$

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

$$= 1 - P(X_1 \geq x, \dots, X_n \geq x)$$

αν ήταν ≥ τότε θα λέγαμε ότι κάθε $X_i, i=1,\dots,n$ είναι $X_i \geq X$ οπότε κατασκευάζουμε τη σχέση "≥"

$$\overset{\text{ανεξαρτησία}}{=} 1 - \prod_{i=1}^{n} P(X_i \geq x) \Rightarrow F_{X_{(1)}}(x) = 1 - \prod_{i=1}^{n} (1 - P(X_i \leq x))$$

$$= 1 - \prod_{i=1}^{n} (1 - F_{X_i}(x))$$

$$\xrightarrow[X_1, \dots, X_n]{\text{ισόνομα}} 1 - \prod_{i=1}^{n} (1 - F_X(x)) = 1 - (1 - F_X(x))^n$$

$$f_{X_{(1)}}(x) = \frac{\partial}{\partial x} F_{X_{(1)}}(x) = \frac{\partial}{\partial x} \{1 - (1 - F_X(x))^n\}$$

$$= -n (1 - F_X(x))^{n-1} \frac{\partial}{\partial x} (1 - F_X(x))$$

$$= n (1 - F_X(x))^{n-1} \frac{\partial}{\partial x} (1 - F_X(x))$$

$$= n (1 - F_X(x))^{n-1} \frac{\partial}{\partial x} F_X(x)$$

$$= n (1 - F_X(x))^{n-1} f_X(x)$$

Κατανομή αθροίσματος (ανεξαρτήτων και ισόνομων) τυχαίων μεταβλητών

Έστω τ.μ. $X_1, \dots, X_n$ — ανεξάρτητες, ισόνομες

Έστω $T = \sum_{i=1}^{n} X_i$

Ζητείται κατανομή της T.
Εφαρμόζω μέθοδο ροπογεννήτριας

$$M_T(t) \overset{op}{=} E(e^{tT}) = E\left(e^{t \sum_{i=1}^{n} X_i}\right)$$

$$t \in \mathbb{R} \quad = E\left(e^{tX_1} \cdot e^{tX_2} \cdots e^{tX_n}\right)$$

$$M_T(t) \xrightarrow[X_1, \dots, X_n]{\text{ανεξαρτ.}} \prod_{i=1}^{n} E(e^{tX_i})$$

$$\Rightarrow m_T(t) \overset{op}{=} \prod_{i=1}^{n} m_{x_i}(t)$$

$$\xrightarrow[m_{x_1}(t)=m_x(t)]{\text{ισόνομα}} \boxed{m_T(t) = [m_x(t)]^n}$$

Αν από $[m_x(t)]^n$ προκύψει αναγνωρίσιμος τύπος ροπογεννήτριας, τότε από Θεώρημα μοναδικότητας ροπογεννητριών η κατανομή της T ταυτίζεται με την κατανομή που έχει ροπή $[m_x(t)]^n$

Επέκταση $\longrightarrow$ Γραμμικών συνδυασμών

Έστω $x_1, \dots, x_n \leftarrow$ ανεξάρτητα ισόνομα

Έστω $T = \sum_{i=1}^{n} a_i x_i$

$$m_T(t) \overset{op}{=} E(e^{tT}) = E\left(e^{t\sum_{i=1}^{n} a_i x_i}\right) = E\left(e^{ta_1x_1} \cdots e^{ta_nx_n}\right)$$

$$\overset{\text{ανεξαρτ.}}{=} \prod_{i=1}^{n} E\left(e^{ta_ix_i}\right)$$

$$\overset{op.}{=} \prod_{i=1}^{n} m_{x_i}(ta_i) \overset{\text{ισονομ.}}{=} \prod_{i=1}^{n} m_x(ta_i)$$

Παράδ: Έστω $x_1, \dots, x_n$ ανεξάρτητα και $x_i \sim Poison(\theta)$ $\forall i=1,\dots,n$.
Να βρεθεί η κατανομή αθροίσματος $T = \sum_{i=1}^{n} x_i$

Όταν έχω αθροίσματα σκέφτομαι ροπογεννήτρια

$$m_T(t) = E(e^{tT}) = E(e^{t\sum x_i}) = E(e^{tx_1} \cdots e^{tx_n})$$

$$\overset{\text{ανεξ.}}{=} \prod_{i=1}^{n} E(e^{tx_i}) = \prod_{i=1}^{n} m_{x_i}(t) = \prod_{i=1}^{n} e^{\theta(e^t-1)}$$

$$\left.\begin{array}{l} \Rightarrow m_T(t) = e^{n\theta(e^t-1)} \\ \text{Α } w \sim Poisson(\lambda) \quad m_w(\lambda) = e^{\lambda(e^t-1)} \end{array}\right\} \quad T \sim Poisson(n\theta)$$

Παράδειγμα ②: Έστω $X_1, \dots, X_n$ ανεξ. και ισόνομες $N(\mu, \sigma^2)$

Να βρεθεί η κατανομή της $T = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

$$m_T(t) \overset{op}{=} E(e^{tT}) = E(e^{t\frac{1}{n}\sum_{i=1}^{n} X_i})$$

$$= E(e^{\frac{t}{n}X_1} e^{\frac{t}{n}X_2} \cdots e^{\frac{t}{n}X_n})$$

$$\underset{\text{ανεξαρτ.}}{=} \prod_{i=1}^{n} E(e^{\frac{t}{n}X_i}) \underset{\text{ροπογ.}}{\overset{op}{=}} \prod_{i=1}^{n} m_{X_i}\left(\frac{t}{n}\right)$$

Αν $X_1, \dots, X_n$ ανεξάρτητες τότε και $g_1(X_1), \dots, g_n(X_n)$ ανεξάρτητες. τότε $E(g_1(X_1)\cdots g_n(X_n)) = \prod_{i=1}^{n} E(g(X_i))$

$$= \prod_{i=1}^{n} \exp\left\{\mu \frac{t}{n} + \frac{1}{2}\sigma^2\left(\frac{t}{n}\right)^2\right\}$$

$$= \prod_{i=1}^{n} e^{\frac{\mu}{n}t + \frac{1}{2}\frac{\sigma^2}{n^2}t^2}$$

$$= e^{n\left(\frac{\mu}{n}t + \frac{1}{2}\frac{\sigma^2}{n}t^2\right)}$$

$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Αυτή είναι η ροπογεννήτρια της $N(\mu, \frac{\sigma^2}{n})$

Άρα $T \equiv \bar{X} \sim N(\mu, \frac{\sigma^2}{\mu})$

Ασκήσεις

Έστω $X_1, \dots, X_n$ ανεξάρτητες και ισόνομες με

ⓐ $X_i \sim \text{Εκθ}(\theta)$, $\theta > 0$. Τότε $T \equiv \sum X_i \sim G(n, \frac{1}{\theta})$

ⓑ $X_i \sim \text{Εκθ}\left(\frac{1}{\theta}\right)$. Τότε $T = \sum X_i \sim G(n, \theta)$

ⓒ Αν $X_i \sim \text{Bernoulli}(p)$ $(\text{Bernoulli}(p) \equiv B(1, p))$

Τότε $T = \sum_{i=1}^{n} X_i \sim B(n, p)$