ΤΡΙΤΗ 16 ΦΕΒΡΟΥΑΡΙΟΥ 2021 - ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

Επανάληψη:

θέλω να δείξω αυτό

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

Περιληπτικά $G(k,\lambda) \equiv \frac{\lambda}{2} \chi^2_{2k}$ ή $\chi^2_{2k} = \frac{2}{\lambda} G(k,\lambda)$. (είδαμε την συνεχή περίπτωση)

Αλλαγή Μεταβλητών $\xrightarrow[\text{περίπτωση}]{\text{συνεχή}}$ **Μέθοδος Μετασχηματισμού**

Θεωρώ τον μετασχηματισμό $y = h(x) = \frac{2}{\lambda} x$.

(Πρέπει να βρω ας) τιμές τ.μ. Y, $y > 0$

O h είναι 1-1, γιατί $h'(x) > 0$, άρα $h \nearrow$.

Αφού h είναι 1-1, υπάρχει η αντίστροφη h^{-1} και είναι:

$x = h^{-1}(y) = \frac{\lambda}{2} y$ (Πρέπει να δω αν η παράγωγος της αντίστροφης ως προς y είναι συνεχής και $\neq 0$)

$\frac{d}{dy} h^{-1}(y) = \frac{\lambda}{2} \neq 0$ και συνεχής για $y > 0$.

απόλυτη τιμή

Άρα $f_Y(y) = f_X(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right| = f_X\left(\frac{\lambda}{2} y\right) \left| \frac{\lambda}{2} \right| = \frac{\lambda}{2} f_X\left(\frac{\lambda}{2} y\right)$, $y > 0$

Αφού $X \sim G(k,\lambda)$ (ξέρω την πυκνότητα του. Από τυπολόγιο) $\rightarrow f_X(x) = \frac{1}{\lambda^k \Gamma(k)} x^{k-1} e^{-x/\lambda}$, $x>0$

Επομένως, $f_Y(y) = \frac{\lambda}{2} \frac{1}{\lambda^k \Gamma(k)} \left(\frac{\lambda}{2} y\right)^{k-1} e^{-\left(\frac{\lambda}{2} y\right)/\lambda}$, $y > 0$

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

Η χ^2_ν έχει σ.π.π. $f_{\chi^2_\nu}(x) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{\frac{\nu}{2}-1} e^{-x/2}$, $x > 0$

Παρατηρούμε ότι αν $\nu = 2k$, η $f_Y(y) \equiv f_{\chi^2_{2k}}(y)$

Άρα $Y \sim \chi^2_{2k}$ ✓

Μέθοδος της Α.σ.κ. (αλλαγή μεταβλητών).

Παράδειγμα. Έστω συνεχής τ.μ. X με γνησίως αύξουσα α.σ.κ. $F_X(x)$. ΝΔΟ

$Y = F_X(X) \sim U(0,1)$.

Απ. (τέλεια στην συνεχή περίπτωση).

Τιμές της Y: $y = F_X(x) \in [0,1]$, αφού $F_X(x) = P(X \le x) \in [0,1]$ → δεν [illegible] [illegible] ειδικά με πλευρές α.σ. για ανοικτό [illegible] για ότι μάθαμε στη συνεχή περίπτωση.

Για $y \in (0,1)$: $F_Y(y) \overset{ορ}{=} P(Y \le y) = P(F_X(X) \le y)$

Αφού η X είναι γνησίως αύξουσα 1-1 αντιστρέψιμη

$$= P\left(F_X^{-1}(F_X(X)) \le F_X^{-1}(y)\right) = P(X \le F_X^{-1}(y)) \overset{ορ}{=} F_X(F_X^{-1}(y)) = y,\ y \in (0,1)$$

Επομένως:
$$F_Y(y) = \begin{cases} 0, & y \le 0 \\ y, & y \in (0,1) \\ 1, & y \ge 1 \end{cases} \quad (*)$$

Αν $W \sim U(\alpha,\beta)$ γνωρίζουμε ότι
$$F_W(w) = \begin{cases} 0, & w \le \alpha \\ \dfrac{w-\alpha}{\beta-\alpha}, & w \in (\alpha,\beta) \\ 1, & w \ge \beta \end{cases} \quad (**)$$

Από (*), (**) προφανώς $Y = F_X(X) \sim U(0,1)$.

Κατανομή του max και min.

→ π.χ. ελάχιστος/μέγιστος χρόνος ζωής ασθενών.

έχουν την ίδια κατανομή

Παράδειγμα. Έστω τ.μ. $X_1, \dots, X_n$ ανεξάρτητες και ισόνομες. Έστω $X_{(1)} = \min\{X_1, \dots, X_n\}$ και $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$. Να βρεθεί η κατανομή της $X_{(n)}$ και $X_{(1)}$.

Απ. (Α.σ.κ.)

$$F_{X_{(n)}}(x) \overset{ορ}{=} P(X_{(n)} \le x) = P(\max\{X_1, X_2, \dots, X_n\} \le x) = P(X_i \le x,\ \forall i = 1, \dots, n) =$$

$x \in \mathbb{R}$

$$= P(X_1 \le x, X_2 \le x, \dots, X_n \le x) \overset{X_i \text{ ανεξ.}}{=} P(X_1 \le x) P(X_2 \le x) \cdots P(X_n \le x)$$

$$= F_{X_1}(x) F_{X_2}(x) \cdots F_{X_n}(x) = \prod_{i=1}^{n} F_{X_i}(x) \underset{F_{X_i}(x) = F_X(x)}{\overset{X_i \text{ ισόνομες}}{=}} \underbrace{[F_X(x)]^n}_{\text{η κοινή σ.κ.}}$$

Τελικά: $F_{X_{(n)}}(x) = [F_X(x)]^n,\ \forall x \in \mathbb{R}$.

Άρα $f_{X(n)}(x) = \frac{d}{dx} F_{X(n)}(x) = \frac{d}{dx}[F_X(x)]^n = n[F_X(x)]^{n-1} \frac{d}{dx} F_X(x)$

$$\Rightarrow \boxed{f_{X(n)}(x) = n\underbrace{[F_X(x)]^{n-1}}_{\text{η κοινή α.σ.κ.}} \underbrace{f_X(x)}_{\text{η κοινή πυκνότητα}}}$$

Για την $X_{(1)}$:

$F_{X(1)}(x) \overset{\text{ορ}}{=} P(X_{(1)} \leq x) = P(\min\{X_1, X_2, ..., X_n\} \leq x) = 1 - P(\min\{X_1, ..., X_n\} \geq x) =$

$= 1 - P(X_i \geq x \ \forall i) = 1 - P(X_1 \geq x, X_2 \geq x, ..., X_n \geq x)$

$\overset{X_i \text{ ανεξ.}}{=} 1 - P(X_1 \geq x) \cdots P(X_n \geq x) = 1 - [1 - P(X_1 \leq x)] \cdots [1 - P(X_n \leq x)]$

$= 1 - [1 - F_{X_1}(x)] \cdots [1 - F_{X_n}(x)]$

$\Rightarrow F_{X(1)}(x) = 1 - \prod_{i=1}^{n} [1 - F_{X_i}(x)] \underset{F_{X_i} = F_X}{\overset{X_i \text{ ισόνομες}}{=}} 1 - \prod_{i=1}^{n} [1 - F_X(x)]$

$\Rightarrow F_{X(1)}(x) = 1 - [1 - \underbrace{F_X(x)}_{\text{κοινή αθροισ.}}]^n, \ x \in \mathbb{R}$

Άρα $f_{X(1)}(x) = \frac{d}{dx} F_{X(1)}(x) = \frac{d}{dx}\{1 - [1 - F_X(x)]^n\} =$

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

$$\Rightarrow \boxed{f_{X(1)}(x) = n[1 - F_X(x)]^{n-1} f_X(x)}$$

Κατανομή αθροίσματος ανεξάρτητων τυχαίων μεταβλητών

Πρόβλημα έστω τ.μ. $X_1, X_2, \dots, X_n$ ανεξάρτητες.

έστω $T=\sum_{i=1}^{n} X_i$, ή γενικότερα $T=\sum_{i=1}^{n} a_i X_i$ → κάποιος γραμμικός συνδυασμός των X_i
ή γενικότερα κάποιος συνδυασμός των X_i

Ποια η κατανομή της T;

Απ

Μέθοδος της ροπογεννήτριας (βολεύει αυτή η μέθοδος)

$$m_T(t) \overset{\text{ορ}}{=} E(e^{tT}) = E\left(e^{t\sum_{i=1}^{n} X_i}\right) = E\left(e^{tX_1+tX_2+\dots+tX_n}\right) = E\left(e^{tX_1}e^{tX_2}\cdots e^{tX_n}\right) \underset{X_i \text{ ανεξ.}}{=} E(e^{tX_1})E(e^{tX_2})\cdots E(e^{tX_n}) = \prod_{i=1}^{n} E(e^{tX_i}) \overset{\text{ορ}}{=} \prod_{i=1}^{n} m_{X_i}(t)$$

Αν επιπλέον X_i ισόνομες με κοινή ροπογεννήτρια $m_X(t)$, τότε:

$$m_T(t)=[m_X(t)]^n.$$

Ανάλογα, αν $T=\sum_{i=1}^{n} a_i X_i$ $\quad m_T(t)=\dots=\prod_{i=1}^{n} m_{X_i}(a_i t)$

και αν οι $X_1,\dots,X_n$ ισόνομες : $m_T(t)=\dots=\prod_{i=1}^{n} m_X(a_i t)$

Παράδειγμα έστω $X_1, X_2, \dots, X_n$ ανεξάρτητες και $X_i \sim P(\theta)$ $\forall i=1,\dots,n$, $\theta>0$. Να βρεθεί η κατανομή $T=\sum_{i=1}^{n} X_i$

έχουν όλες την ίδια κατανομή

Απ

$$m_T(t) = E(e^{tT}) = \dots = \prod_{i=1}^{n} m_{X_i}(t) = [m_X(t)]^n$$

Αλλά, αν $X \sim P(\theta)$, $m_X(t)=e^{\theta(e^t-1)}$, $t\in\mathbb{R}$

$$\Rightarrow m_T(t)=\left[e^{\theta(e^t-1)}\right]^n \Rightarrow \boxed{m_T(t)=e^{n\theta(e^t-1)},\ t\in\mathbb{R}}$$

επειδή η m_T ταυτίζεται με την ροπογεννήτρια της Poisson με παράμετρο $n\theta$, από το θεώρημα μοναδικότητας των ροπογεννητριών $T=\sum_{i=1}^{n} X_i \sim P(n\theta)$

Ασκήσεις.

(1) Έστω ανεξάρτητες $X_1, X_2, \ldots, X_n \sim Bernoulli(p)$, $p \in (0,1)$ (Δηλ. $P_{X_i}(x_i) = p^{x_i}(1-p)^{1-x_i}$, $i = 1, \ldots, n$, $x_i = 0, 1$). Τότε $T = \sum_{i=1}^{n} X_i \sim B(n, p)$

(2) Έστω ανεξάρτητες $X_1, X_2, \ldots, X_n$ με $X_i \sim Εκθ(1/\theta)$, $\theta > 0$. Νδο $T = \sum_{i=1}^{n} X_i \sim G(n, \theta)$

με ροπογεννήτρια.

(3) Έστω ανεξάρτητες $X_1, \ldots, X_n$ με $X_i \sim N(\mu, \sigma^2)$. Νδο $T = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$