ΚΕΝΤΡΙΚΟ ΟΡΙΑΚΟ ΘΕΩΡΗΜΑ

[#1]

ΣΥΓΚΛΙΣΗ ΚΑΤΑ ΚΑΤΑΝΟΜΗ: Έστω $(X_n)_{n \in \mathbb{N}^*}$ ακολουθία Τ.Μ. με κοινό σύνολο τιμών S και έστω $F_n(x) = P(X_n \leq x)$ η συνάρτηση κατανομής της X_n, $n \in \mathbb{N}_0$. Λέμε ότι η ακολουθία (X_n) συγκλίνει κατά κατανομή στην Τ.Μ. X που ακολουθεί μια κατανομή D και έχει συνάρτηση κατανομής $F(x) = P(X \leq x)$ και σύνολο τιμών S και γράφουμε $X_n \xrightarrow{d} D$, αν και μόνο εάν

$$\boxed{\lim_{n \to \infty} F_n(x) = F(x), \quad \forall x \in S \text{ όπου η } F(\cdot) \text{ είναι συνεχής.}}$$

Κεντρικό Οριακό Θεώρημα: Αν οι διακριτές ή συνεχείς $X_1, X_2, \dots, X_n$ ακολουθούν την ίδια κατανομή με μέση τιμή (μ) και (σ^2), τότε:

(i): $\boxed{S_n = X_1 + X_2 + \dots + X_n \longrightarrow N(n\mu, n\sigma^2)}$

(ii): $\boxed{\overline{X}_n = \dfrac{X_1 + X_2 + \dots + X_n}{n} \longrightarrow N\left(\mu, \dfrac{\sigma^2}{n}\right)}$

(iii): $\boxed{\overline{S}_n = \dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \dfrac{S_n - n\mu}{\sigma\sqrt{n}} = \dfrac{1}{\sigma\sqrt{n}} \cdot \sum_{i=1}^{n}(X_i - \mu) \longrightarrow N(0,1)}$

(iv): $\boxed{\lim_{n \to \infty} P(\overline{S}_n \leq x) = \Phi(x), \quad \forall x \in \mathbb{R}.}$

[#2]

Από την διατύπωση (iii) του Κεντρικού Οριακού Θεωρήματος [Central Limit Theorem – CLT] έχουμε ότι:

$$\bar{S}_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{(X_1+X_2+\cdots+X_n)-n\mu}{\sigma}\cdot\frac{1}{\sqrt{n}} \rightarrow$$

$$\bar{S}_n = \frac{(X_1-\mu)+(X_2-\mu)+\cdots+(X_n-\mu)}{\sigma}\cdot\frac{1}{\sqrt{n}} \Rightarrow$$

$$\boxed{\bar{S}_n = \left[\frac{X_1-\mu}{\sigma}+\frac{X_2-\mu}{\sigma}+\cdots+\frac{X_n-\mu}{\sigma}\right]\cdot\frac{1}{\sqrt{n}}}$$

Θεωρώντας τις Τ.Μ.: $X_i' = \dfrac{X_i-\mu}{\sigma}$, $\forall i \in [n]$, είναι τα μπορούμε να δείξουμε ότι:

$$\begin{cases} (\alpha): & E[X_i'] = 0 \\ (\beta): & Var[X_i'] = 1 \end{cases}$$

Ισχύει ότι $\begin{cases} E[aX+\beta] = a\,E[X] + \beta \\ Var[aX+\beta] = a^2 Var[X] \end{cases}$

$$E\left[\frac{X_i-\mu}{\sigma}\right] = \frac{1}{\sigma}E[X_i-\mu] = \frac{1}{\sigma}(E[X_i]-\mu) = \frac{1}{\sigma}(\mu-\mu) = 0$$

$$Var\left[\frac{X_i-\mu}{\sigma}\right] = \frac{1}{\sigma^2}Var[X_i-\mu] = \frac{1}{\sigma^2}\cdot Var[X_i] = \frac{1}{\sigma^2}\cdot\sigma^2 = 1$$

Εναλλακτική Διατύπωση Κεντρικού Οριακού Θεωρήματος [#3]

Αν οι Τ.Μ. $X_1, X_2, \dots, X_n$ είναι ανεξάρτητες με πανομοιότυπη κατανομή έτσι ώστε $E[X_i] = 0$ και $Var[X_i] = 1$, τότε η ποσότητα $\bar{S}_n = \dfrac{X_1 + X_2 + \dots + X_n}{\sqrt{n}} \xrightarrow{d} N(0,1)$.

★ Έχουμε αποδείξει ότι:

$$\boxed{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2} + tx} = e^{\frac{t^2}{2}} = M_X(t)} \quad (1)$$

$$\boxed{\text{για } X \sim N(0,1)}$$

★ Θα χρειαστούμε επιπλέον την τιμή του ορίου:

$$\boxed{\lim_{n \to \infty} \left(1 + \frac{a}{n}\right)^n} \xrightarrow[n \to +\infty \Rightarrow t \to 0^+]{\frac{a}{n} = t,\ n = \frac{a}{t}} \lim_{t \to 0^+} (1+t)^{\frac{a}{t}} =$$

$$= \lim_{t \to 0^+} e^{\ln (1+t)^{\frac{a}{t}}} = \lim_{t \to 0^+} e^{\frac{a}{t} \cdot \ln(1+t)} \overset{(*)}{=} \boxed{e^a} \quad (2)$$

(*) $\displaystyle \lim_{t \to 0^+} \frac{a}{t} \cdot \ln(1+t) = a \lim_{t \to 0^+} \frac{\ln(t+1)'}{t'} =$

$$= a \cdot \lim_{t \to 0^+} \frac{(t+1)'}{t+1} \cdot \frac{1}{1} = a \cdot \lim_{t \to 0^+} \frac{1}{t+1} = a$$

★ Γνωρίζουμε επιπλέον ότι:

$$Var\left[\frac{x_i}{\sqrt{n}}\right] = \left(\frac{1}{\sqrt{n}}\right)^2 Var[x_i] = \frac{1}{n} \quad (3)$$

$$Var\left[\left(\frac{x_i}{\sqrt{n}}\right)\right] = E\left[\left(\frac{x_i}{\sqrt{n}}\right)^2\right] - E^2\left[\left(\frac{x_i}{\sqrt{n}}\right)\right] \quad (4)$$

$$\boxed{Var\left[\frac{x_i}{\sqrt{n}}\right] = \frac{1}{n} = E\left[\left(\frac{x_i}{\sqrt{n}}\right)^2\right], \forall i \in [n] \quad (5)}$$

Θα κάνουμε χρήση του μετασχηματισμού $\boxed{y_i = \frac{x_i}{\sqrt{n}}, \forall i \in [n]}$ για τον οποίο έχουμε δείξει ότι:

$$\boxed{\begin{cases} E[y_i] = 0, \forall i \in [n] \\ E[y_i^2] = \frac{1}{n}, \forall i \in [n] \end{cases} \quad (6)}$$

Στην συνέχεια θα πρέπει να υπολογισθεί η ροπογεννήτρια συνάρτηση της Τ.Μ. y_i:

$$\boxed{M_{y_i}(t) = 1 + \frac{E[y_i]}{1!} \cdot t + \frac{E[y_i^2]}{2!} \cdot t^2 + \frac{E[y_i^3]}{3!} t^3 + \dots \quad (7)}$$

Από τις εξισώσεις (6) και (7) έχουμε ότι:

$$\boxed{M_{y_i}(t) = 1 + 0 + \frac{\left(\frac{1}{n}\right)}{2!} \cdot t^2 + O(t^3) \quad (8)}$$

→ higher order terms!

[#5]

Θέτοντας $\bar{S}_n = \sum_{i=1}^{n} Y_i$, θα έχουμε ότι:

$$\boxed{M_{\bar{S}_n}(t) = \prod_{i=1}^{n} M_{Y_i}(t) \quad (a)}$$

(καθώς οι Y_i, $i \in [n]$ είναι i.i.d)

Από την εξίσωση (a) έχουμε ότι:

$$\boxed{M_{\bar{S}_n}(t) = \left(1 + \frac{1}{2n} t^2 + O(t^3)\right)^n \quad (10)}$$

Λαμβάνοντας υπόψιν ότι η συνεισφορά των όρων υψηλότερης τάξης είναι αμελητέα, έχουμε ότι:

$$\boxed{M_{\bar{S}_n}(t) = \left(1 + \frac{1}{2n} t^2\right)^n \quad (11)}$$

Θέλουμε να βρούμε όμως, το όριο της ροπογεννήτριας συνάρτησης καθώς $n \to +\infty$:

$$\boxed{\lim_{n \to +\infty} M_{\bar{S}_n}(t) = \lim_{n \to +\infty} \left(1 + \frac{\frac{1}{2}t^2 = a}{n}\right)^n \overset{(2)}{=\!=} e^{\frac{1}{2}t^2} \quad (12)}$$

Η συνάρτηση $e^{\frac{1}{2}t^2}$ γνωρίζουμε ότι αντιστοιχεί στην ροπογεννήτρια συνάρτηση μιας Τ.Μ. $Z \sim N(0,1)$.

★ Επομένως, η ποσότητα $\bar{S}_n \xrightarrow{d} N(0,1)$ (ο.έ.δ).

Problem: Let $X_1, X_2, \ldots, X_n$ be independent random samples from variables, where $X_i \sim \text{Gamma}(a, \beta)$, with unknown shape and rate parameters a and β. Derive an approximation $\hat{a}$ and $\hat{\beta}$ for the parameters of the Gamma distribution as the number of samples $n \to +\infty$, assuming that the true mean and variance $E[\bar{X}_n]$ and $Var[\bar{X}_n]$ are known.

Solution: We know that the PDF of each X_i, $i \in [n]$, is given by:

$$\boxed{f_{X_i}(x; a, \beta) = \frac{\beta^a \cdot x^{a-1} \cdot e^{-\beta x}}{\Gamma(a)}, \quad x > 0 \quad (1)}$$

We have already established that:

$$\begin{cases} E[X_i] = \dfrac{a}{\beta} & (2) \\ \quad \forall i \in [n] \\ Var[X_i] = \dfrac{a}{\beta^2} & (3) \end{cases}$$

Assuming that we aggregate a number of samples n, $X_1, X_2, \ldots, X_n$ whose true PDF is given by Eq. (1) (knowing that the samples are independent), we may use the CLT to derive that the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ follows the normal distribution as:

$$\boxed{\bar{X}_n \sim N\left(E[X_i], \frac{Var[X_i]}{n}\right) \quad (4)}$$

Substituting Eq. (2) and (3) into Eq. (4) yields:

$$\boxed{\bar{X}_n \sim N\left(\frac{a}{\beta}, \frac{a}{n\beta^2}\right) \quad (5)}$$

Thus, we may write that:

$$\boxed{\begin{cases} E[\bar{X}_n] = \dfrac{\hat{a}}{\hat{\beta}} & (5) \\ \\ Var[\bar{X}_n] = \dfrac{\hat{a}}{n\hat{\beta}^2} & (6) \end{cases}}$$

From Eqs. (5) and (6), we have that:

$$\begin{cases} \dfrac{\hat{a}}{\hat{\beta}} = E[\bar{X}_n] & (7) \\ \\ \dfrac{\hat{a}}{\hat{\beta}^2} = n \cdot Var[\bar{X}_n] & (8) \end{cases} \rightarrow \frac{\hat{a}}{\hat{\beta}} \cdot \frac{1}{\hat{\beta}} = n \cdot Var[\bar{X}_n] \xrightarrow{(7)}$$

$$\frac{E[\bar{X}_n]}{\hat{\beta}} = n \cdot Var[\bar{X}_n] \rightarrow$$

$$\boxed{\hat{\beta} = \frac{E[\bar{X}_n]}{n \cdot Var[\bar{X}_n]} \quad (8)}$$

From Eq. (7), we have that:

$$\hat{a} = \hat{\beta} \cdot E[\bar{X}_n] \xrightarrow{(8)} \boxed{\hat{a} = \frac{E^2[\bar{X}_n]}{n\,Var[\bar{X}_n]} \quad (9)}$$