Φροντιστήριο 5

Οριακό Θεώρημα Poisson (Poisson Limit Theorem)
Το οριακό θεώρημα Poisson μας πληροφορεί για την χειροπιαστική σχέση των κατανομών της Διωνυμικής ($Bin(n,q)$) και της $Poisson(\lambda)$.

Έστω ακολουθία από $Bin(n,q)$, τέτοια ώστε $n \to \infty$, $0<q \to 0$ και $n \cdot q \to \lambda$. Με άλλα λόγια, ο αριθμός των ανεξάρτητων διωνυμικών πειραμάτων $n \to \infty$ και η πιθανότητα επιτυχίας $q \to 0$ με τέτοιο τρόπο ώστε το γινόμενο $n \cdot q$ να παραμένει σταθερό, δηλαδή $n \cdot q \to \lambda$.

Καταρχάς, επειδή $supp(Bin(n,q)) = \{0,1,2,\ldots,n\}$ είναι διαισθητικά εύλογο ότι το στήριγμα της κατανομής-όριο, αν υπάρχει, θα είναι το σύνολο των φυσικών αριθμών ($\mathbb{N}$), γιατί:

(1) κάθε φυσικός αριθμός θα περιλαμβάνεται στο στήριγμα της $Bin(n,q)$ για αρκετά μεγάλο n
(2) $supp(Bin(n,q)) \subset supp(Bin(n+1,q))$ για $\forall\, n>0$
(3) στα εν λόγω στηρίγματα περιλαμβάνονται μόνο φυσικοί αριθμοί.

Επομένως, αν η κατανομή-όριο υπάρχει, θα είναι διακριτή κατανομή με στήριγμα το σύνολο των φυσικών αριθμών ($\mathbb{N}$).

Τι πιθανότητες θα αποδίδει αυτή η κατανομή - όριο (αν υπάρχει) στα στοιχεία του δείγματός της??

Έστω $i \in \mathbb{N}$ (supp) Έχουμε ότι η $Bin(n,q)$ θα αποδίδει σε αυτό πιθανότητα:

$$\mathbb{P}(\{i\}) = \binom{n}{i} q^i (1-q)^{n-i} = \frac{n!}{i!\,(n-i)!}\, q^i (1-q)^{n-i} \overset{(*)}{=}$$

$$= \frac{(n-i)!\, \prod_{j=1}^{i} (n-i+j)}{i!\,(n-i)!}\, q^i (1-q)^{n-i} =$$

(*) Σε αυτή την ισότητα, χρησιμοποιούμε ότι $n! = (n-i)!\, \prod_{j=1}^{i} (n-i+j)$

$$= \frac{\prod_{j=1}^{i} (n-i+j)}{i!} \cdot q^i (1-q)^{n-i} =$$

$$= \frac{\prod_{j=1}^{i} (n-i+j)}{i!} \cdot q^i \left(1 - \frac{n \cdot q}{n}\right)^{n-i} =$$

$$= \frac{\prod_{j=1}^{i} (n-i+j)}{i!}\; q^i \cdot \frac{\left(1 - \frac{nq}{n}\right)^n}{\left(1 - \frac{nq}{n}\right)^i} =$$

$$= \frac{\prod_{j=1}^{i} (n-i+j)}{i!} \cdot q^i \cdot \frac{\left(1 - \frac{nq}{n}\right)^n}{\left(\frac{n-nq}{n}\right)^i} =$$

$$= \frac{\prod_{j=1}^{i} (n-i+j)}{i!} \cdot \left(\frac{q}{\frac{n-nq}{n}}\right)^i \cdot \left(1 - \frac{nq}{n}\right)^n =$$

$$= \frac{\prod_{j=1}^{i} (n-i+j)}{i!} \cdot \left(\frac{nq}{n(1-q)}\right)^i \left(1 - \frac{nq}{n}\right)^n =$$

$$= \frac{\prod_{j=1}^{i}(n-i+j)}{i!} \cdot \left(\frac{n \cdot q}{1-q}\right)^i \cdot \frac{1}{n^i} \cdot \left(1-\frac{nq}{n}\right)^n =$$

$$= \underbrace{\frac{\prod_{j=1}^{i}(n-i+j)}{i! \cdot n^i}}_{\alpha} \cdot \underbrace{\left(\frac{nq}{1-q}\right)^i}_{\beta} \underbrace{\left(1-\frac{nq}{n}\right)^n}_{\gamma} \quad (1)$$

Επομένως, έχουμε καταλήξει στη σχέση (1) που αποτελείται από 3 όρους, τον (α), (β) και (γ).

Παρατηρούμε τα εξής, σχετικά με τα όρια του κάθε όρου:

(α) Καθώς το $n \to \infty$ τότε $\prod_{j=1}^{i}(n-i+j) \simeq n^i$ άρα ο πρώτος όρος (α) έχει το ακόλουθο όριο:

$$\lim_{n \to \infty} \frac{\prod_{j=1}^{i}(n-i+j)}{i!\, n^i} = \frac{1}{i!}$$

(β) Επειδή $nq \to \lambda$, $q \to 0$ τότε $\lim_{n \to \infty} \left(\frac{nq}{1-q}\right)^i = \lambda^i$ που είναι το όριο του δεύτερου όρου (β).

(γ) Από τον ορισμό του εκθετικού και επειδή $n \to \infty$, $n \cdot q \to \lambda$, $q \to 0$, το όριο του τρίτου όρου (γ) θα είναι:

$$\lim_{n \to \infty} \left(1-\frac{nq}{n}\right)^n = e^{-\lambda}$$

Επομένως, τα παραπάνω συνεπάγονται από τη σχέση 1 ότι:

$$\lim_{n\to\infty} \mathbb{P}(\{i\}) = \lim_{n\to\infty}\left[\frac{\prod_{j=1}^{i}(n-i+j)}{i!\,n^i}\cdot\left(\frac{nq}{1-q}\right)^i\left(1-\frac{nq}{n}\right)^n\right] \Rightarrow$$

αντικαθιστώντας τα όρια του κάθε όρου

$$\Rightarrow \lim_{n\to\infty}\mathbb{P}(\{i\}) = e^{-\lambda}\cdot\frac{\lambda^i}{i!}$$

Καταλήγουμε, επομένως, στο συμπέρασμα ότι η κατανομή-όριο θα έχει ως στήριγμα (supp) το σύνολο των φυσικών αριθμών $\mathbb{N}$, ενώ σε κάθε στοιχείο του στηρίγματος $i\in\mathbb{N}$, θα αποδίδει πιθανότητα $\mathbb{P}(\{i\}) = e^{-\lambda}\cdot\frac{\lambda^i}{i!}$.

Άρα, έχουμε ότι:

$$\lim_{n\to\infty,\, q\to 0,\, nq\to\lambda} Bin(n,q) = Poiss(\lambda)$$

Το παραπάνω ονομάζεται οριακό θεώρημα Poisson (**Poisson Limit Theorem**) και μας πληροφορεί, ότι οι κατανομές Poisson προκύπτουν ως κατάλληλα όρια διωνυμικών κατανομών.

Παράδειγμα

Έστω εξέταση πανεπιστημιακού μαθήματος, αποτελούμενη από 30 ερωτήσεις πολλαπλής επιλογής. Κάθε ερώτηση είναι ανεξάρτητη από οποιαδήποτε άλλη και εμπεριέχει 4 επιλογές, εκ των οποίων μόνο η μία είναι σωστή. Επομένως, σε κάθε ερώτηση, η πιθανότητα ο εξεταζόμενος να απαντήσει σωστά επιλέγοντας τυχαία την απάντηση από τις δεδομένες επιλογές είναι $\frac{1}{4}$.

Αρχικά, θέλουμε να βρούμε την κατανομή πιθανότητας επί του $\mathbb{R}$ που περιγράφει την κάθε ερώτηση πολλαπλής επιλογής ξεχωριστά και στη συνέχεια, θα βρούμε την κατανομή πιθανότητας επί του $\mathbb{R}$ για το πιο σύνθετο πείραμα, εκείνο της επίδοσης του εξεταζόμενου εφόσον αυτός απαντήσει σε όλες τις ερωτήσεις του διαγωνίσματος, δηλαδή και στις 30 ερωτήσεις πολλαπλής επιλογής, κάνοντας τυχαία επιλογή.

Επομένως, η μαθηματική μορφοποίηση του τυχαίου πειράματος που αφορά την επίδοση του εξεταζόμενου σε μία μεμονωμένη ερώτηση πολλαπλής επιλογής, θα έχει ως χώρο πιθανότητας $(\Omega, 2^{\Omega}, \mathbb{P})$, εκείνον τον χώρο που θα αποτελείται από:

- το σύνολο αναφοράς Ω (δηλαδή, το σύνολο όλων των δυνατών αποτελεσμάτων του πειράματος τύχης), άρα
  $\Omega = \{$ορθή τυχαία επιλογή, λάθος τυχαία επιλογή$\}$

- το δυναμοσύνολο του Ω, 2^{Ω}, το οποίο είναι:

$2^{\Omega} = \{\ \phi,\ \{$ορθή τυχαία επιλογή$\}$, $\{$λάθος τυχαία επιλογή$\}$, $\Omega\}$

την κατανομή πιθανότητας $\mathbb{P}: 2^{\Omega} \to [0,1]$ που ορίζεται από:

$\mathbb{P}(\emptyset) = 0$

$\mathbb{P}(\{\text{ορθή τυχαία επιλογή}\}) = \frac{1}{4}$

$\mathbb{P}(\{\text{λάθος τυχαία επιλογή}\}) = \frac{3}{4}$

$\mathbb{P}(\Omega) = 1$

Χρησιμοποιώντας την τυχαία μεταβλητή $X: \Omega \to \mathbb{R}$ που ορίζεται ως:

$X(\text{ορθή τυχαία επιλογή}) = 1$

$X(\text{λάθος τυχαία επιλογή}) = 0$

Θα μεταφέρουμε την κατανομή πιθανότητας στο $\mathbb{R}$ και θα βρούμε από ποια "οικεία" διακριτή κατανομή πιθανότητας αναπαρίσταται το παραπάνω πείραμα τύχης στους πραγματικούς αριθμούς ($\mathbb{R}$).

Επομένως, σύμφωνα με τον ορισμό της τυχαίας μεταβλητής, βρίσκουμε την αντίστροφη εικόνα, μέσω της τυχαίας μεταβλητής X, για $\forall A \in \Sigma_{\mathbb{R}}$, άρα:

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} = \begin{cases} \emptyset & , \text{αν } 0,1 \notin A \\ \{\text{ορθή τυχαία επιλογή}\} & , \text{αν } 0 \notin A, 1 \in A \\ \{\text{λάθος τυχαία επιλογή}\} & , \text{αν } 0 \in A, 1 \notin A \\ \Omega & , \text{αν } 0,1 \in A \end{cases}$$

Συνεπώς, η κατανομή από μεταφορά $\mathbb{P}^*$ που ορίζεται ως: $\mathbb{P}^*(A) = \mathbb{P}(X^{-1}(A))$, για $\forall A \in \Sigma_{\mathbb{R}}$, θα είναι:

$$\mathbb{P}^*(A) = \begin{cases} \mathbb{P}(\emptyset) & , \text{αν } 0,1 \notin A \\ \mathbb{P}(\{\text{ορθή τυχαία επιλογή}\}) & , \text{αν } 0 \notin A, 1 \in A \\ \mathbb{P}(\{\text{λάθος τυχαία επιλογή}\}) & \text{αν } 0 \in A, 1 \notin A \\ \mathbb{P}(\Omega) & , \text{αν } 0,1 \in A \end{cases} = \begin{cases} 0 & , \text{αν } 0,1 \notin A \\ \frac{1}{4} & , \text{αν } 0 \notin A, 1 \in A \\ \frac{3}{4} & , \text{αν } 0 \in A, 1 \notin A \\ 1 & , \text{αν } 0,1 \in A \end{cases}$$

Από τα παραπάνω, παρατηρούμε ότι το στήριγμα (supp) της κατανομής από μεταφορά $\mathbb{P}^*$ είναι το σύνολο $\{0,1\}$.
Άρα, $supp\,\mathbb{P}^* = \{0,1\}$.

Επομένως, η παραπάνω κατανομή πιθανότητας αναπαρίσταται στο $\mathbb{R}$ από την διακριτή κατανομή Bernoulli. Συγκεκριμένα, αναπαρίσταται από την κατανομή $Ber\left(\frac{1}{4}\right)$, γιατί το στήριγμα (supp) της $\mathbb{P}^*$ είναι το διακριτό σύνολο $\{0,1\}$ που είναι, όπως γνωρίζουμε, το supp της Bernoulli και επίσης, η πιθανότητα επιτυχίας (q) (δηλαδή, η πιθανότητα ορθής τυχαίας επιλογής) είναι $\frac{1}{4}$. Άρα, $q = \frac{1}{4}$ και $1-q = \frac{3}{4}$.

Επομένως, το πείραμα τύχης που περιγράψαμε αναπαρίσταται στο $\mathbb{R}$ από την $Ber\left(\frac{1}{4}\right)$.

Μπορούμε, τώρα, να εξετάσουμε το πιο σύνθετο πείραμα της επίδοσης του εξεταζόμενου εφόσον αυτός απαντήσει σε όλες τις ερωτήσεις πολλαπλής επιλογής του διαγωνίσματος κάνοντας τυχαία επιλογή. Εξαιτίας της ανεξαρτησίας (που έχουμε υποθέσει) μεταξύ των 30 ερωτήσεων πολλαπλής επιλογής του διαγωνίσματος (άρα, $n=30$), η κατανομή πιθανότητας στο $\mathbb{R}$ που θα περιγράφει το πείραμα θα είναι η διωνυμική (Binomial) με παραμέτρους $n=30$ και $q=\frac{1}{4}$ (η πιθανότητα ορθής τυχαίας επιλογής σε κάθε ερώτηση του διαγωνίσματος)
άρα έχουμε την κατανομή $Bin\left(30, \frac{1}{4}\right)$.

Αφού γνωρίζουμε την κατανομή που περιγράφει το πείραμά μας, μπορούμε να υπολογίσουμε πιθανότητες χρησιμοποιώντας τη συνάρτηση μάζας πιθανότητας της κατανομής, η οποία στην γενική της μορφή είναι:

$$\mathbb{P}(\{i\}) = \binom{n}{i} q^i (1-q)^{n-i}$$

Στη συγκεκριμένη περίπτωση, η $Bin\left(30, \frac{1}{4}\right)$ περιγράφεται από:

$$supp\left(Bin\left(30, \tfrac{1}{4}\right)\right) = \{0, 1, 2, \dots, 30\}$$

$$\mathbb{P}(\{i\}) = \binom{30}{i} \cdot \left(\frac{1}{4}\right)^{i} \cdot \left(1 - \frac{1}{4}\right)^{30-i}, \quad i \in supp$$

Συνεπώς, μπορούμε, τώρα, να υπολογίσουμε την πιθανότητα ο εξεταζόμενος να λάβει π.χ τον βαθμό 5 - εφόσον όλες οι ερωτήσεις είναι ισόβαθμες και απαντάει σε όλες τυχαία. Για να πάρει 5 ο εξεταζόμενος πρέπει να απαντήσει σωστά στις 15 από τις 30 ερωτήσεις του διαγωνίσματος. Άρα, για $i = 15 \in supp$, έχουμε

$$\mathbb{P}(\{15\}) = \binom{30}{15} \cdot \left(\frac{1}{4}\right)^{15} \cdot \left(1 - \frac{1}{4}\right)^{30-15} = \binom{30}{15} \cdot \left(\frac{1}{4}\right)^{15} \cdot \left(\frac{3}{4}\right)^{15} =$$

$$= \frac{30!}{15!\,(30-15)!} \cdot \left(\frac{1}{4}\right)^{15} \left(\frac{3}{4}\right)^{15} = \frac{30!}{15!\,15!} \left(\frac{1}{4}\right)^{15} \left(\frac{3}{4}\right)^{15} \approx 0.0019$$

Στο παραπάνω παράδειγμα, αν επιτρέποταν στο πλήθος των ερωτήσεων (n) του διαγωνίσματος να αποκλίνει στο άπειρο (δηλαδή, $n \to \infty$) και στο πλήθος των πιθανών επιλογών σε κάθε ερώτηση (έστω, m: το πλήθος των επιλογών σε κάθε ερώτηση) να αποκλίνει, επίσης, στο άπειρο με τον ίδιο ρυθμό, που συνεπάγεται ότι η πιθανότητα ορθής τυχαίας επιλογής (q) σε κάθε ερώτηση θα τείνει στο 0 (αφού $q = \frac{1}{m}$ και $m \to \infty$ τότε $q \to 0$), τότε η κατανομή πιθανότητας που θα περιέγραφε την επίδοση του εξεταζόμενου, βάσει αποκλειστικά τυχαίων επιλογών, σε αυτό το ασυμπτωτικό διαγώνισμα θα ήταν η Poisson(1), σύμφωνα με το Οριακό Θεώρημα της Poisson. Γιατί $\lambda = 1$?? Επειδή, όπως γνωρίζουμε από θεώρημα: $n \cdot q \to \lambda$, όμως $q = \frac{1}{m}$ και εφόσον υποθέτουμε ότι $n \to \infty$ και $m \to \infty$ με τον ίδιο ρυθμό άρα $n \cdot q = n \cdot \frac{1}{m} \longrightarrow 1$, αν $n \to \infty$, $m \to \infty$