Θεωρητικές Κατανομές

Στις προηγούμενες διαλέξεις είδαμε κάποια παραδείγματα κατανομών. Κάποιες από αυτές είναι γνωστές κατανομές που είναι χρήσιμες για να περιγράψουν συγκεκριμένα τυχαία πειράματα. Σκοπός μας είναι να βρούμε την άγνωστη κατανομή μιας τυχαίας μεταβλητής (π.χ. τη συνάρτηση πιθανότητας). Αν γνωρίζουμε ότι κάποιο πείραμα μπορεί να περιγραφεί -είτε ακριβώς είτε προσεγγιστικά- από κάποια γνωστή κατανομή τότε μας αρκεί να βρούμε τις άγνωστες παραμέτρους αυτής της κατανομής.

Τυποποίηση τυχαίας μεταβλητής

Πολλές φορές όπως θα δούμε είναι χρήσιμο να τυποποιούμε μια τυχαία μεταβλητή X με τον μέσο και την διακύμανση της

$$Z := \frac{X - E(X)}{\sqrt{Var(X)}}$$

οπότε για την καινούρια τυχαία μεταβλητή Z θα ισχύει ότι

$$E(Z) = \frac{E\left[X - E(X)\right]}{\sqrt{Var(X)}} = 0$$

και

$$Var(Z) = \frac{Var\left[X - E(X)\right]}{Var(X)} = \frac{Var(X)}{Var(X)} = 1$$

Άθροισμα τυχαίων μεταβλητών

Αν έχω τις τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$ τότε

$$\begin{aligned} E\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} E(X_i) \\ Var\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j) \end{aligned}$$

Στην περίπτωση που είναι και ανεξάρτητες τότε $Cov(X_i, X_j) = 0$ για $i \neq j$ οπότε

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Αν επιπλέον έχουν όλες οι τυχαίες μεταβλητές την ίδια κατανομή (οπότε αποτελούν ένα τυχαίο δείγμα από αυτή την κατανομή), οπότε έστω $E(X_i) = \mu \quad \forall i$ και $Var(X_i) = \sigma^2 \quad \forall i$, τότε

$$\begin{aligned} E\left(\sum_{i=1}^{n} X_i\right) &= n\mu \\ Var\left(\sum_{i=1}^{n} X_i\right) &= n\sigma^2 \end{aligned}$$

Διακριτές Κατανομές

Bernoulli

Η Bernoulli(p) έχει συνάρτηση (μάζας) πιθανότητας

$$f(x) = \begin{cases} p^x(1-p)^{1-x} & \text{για } x = 0, 1 \\ 0 & \text{αλλιώς} \end{cases}$$

Αν $X \sim$ Bernoulli(p) τότε:

- $E(X) = \sum_x x f(x) = 0 \cdot p^0 (1-p)^{1-0} + 1 \cdot p^1 (1-p)^{1-1} = p$
- $Var(X) = E(X^2) - [E(X)]^2 = \sum_x x^2 f(x) - [E(X)]^2 = 0^2 \cdot p^0 (1-p)^{1-0} + 1^2 \cdot p^1 (1-p)^{1-1} - p^2 = p - p^2 = p(1-p)$

Παράδειγμα κατανομής Bernoulli είναι η ρίψη κέρματος μία φορά όπου έχουμε $p = \frac{1}{2}$.

Διωνυμική

Η Διωνυμική $B(n,p)$ έχει συνάρτηση (μάζας) πιθανότητας

$$f(x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} & \text{για } x = 0, 1, 2, \ldots, n \\ 0 & \text{αλλιώς} \end{cases}$$

όπου $\begin{pmatrix} n \\ x \end{pmatrix} = \frac{n!}{x!(n-x)!}$ ο αριθμός των συνδυασμών n στοιχείων σε x θέσεις. Ισχύει ότι αν $X_1, X_2, \ldots, X_n$ ανεξάρτητες τυχαίες μεταβλητές όπου καθεμία κατανέμεται με την Bernoulli(p) τότε $\sum_{i=1}^n X_i \sim B(n,p)$. Αν $X \sim B(n,p)$ τότε:

- $E(X) = \sum_x x f(x) = E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np$
- $Var(X) = Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + \underbrace{2\sum_{i<j} Cov(X_i, X_j)}_{0} = np(1-p)$. Καθώς είναι ανεξάρτητες έχουμε ότι $Cov(X_i, X_j)$ για $i \neq j$.

Παράδειγμα Διωνυμικής κατανομής είναι η ρίψη του κέρματος n φορές όπου μας ενδιαφέρει πόσες φορές έρχονται γράμματα (ή κορώνα).

Συνεχείς Κατανομές

Κανονική Κατανομή

Η κανονική κατανομή (normal distribution) είναι η πιο σπουδαία από τις συνεχείς θεωρητικές κατανομές όπως θα δούμε. Η κανονική κατανομή με μέσο μ και διακύμανση σ^2 συμβολίζεται με $N(\mu, \sigma^2)$ και έχει την ακόλουθη συνάρτηση (πυκνότητας) πιθανότητας:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \quad \text{για } x \in \mathbb{R}.$$

άρα και η αθροιστική συνάρτηση κατανομής μπορεί να υπολογιστεί ως $F(x) = \int_{-\infty}^x f(u)du$.

- Παίρνει τιμές σε όλους τους πραγματικούς αριθμούς.
- Επίσης είναι συμμετρική γύρω από το μ. Οπότε $f(\mu - x) = f(\mu + x) \quad \forall x \in \mathbb{R}$. Επίσης $F(\mu - x) = 1 - F(\mu + x) \quad \forall x \in \mathbb{R}$.

Γενικά όταν λέμε π.χ. $X \sim N(\mu, \sigma^2)$ θα εννοούμε ότι η τυχαία μεταβλητή X κατανέμεται με την κατανομή $N(\mu, \sigma^2)$, άρα θα έχει την παραπάνω συνάρτηση πυκνότητας πιθανότητας.

- Για την X ισχύει ότι $E(X) = \mu$ και $Var(X) = \sigma^2$ (δεν το αποδεικνύουμε).

Προφανώς για διαφορετικά μ, σ^2 έχουμε διαφορετική κανονική κατανομή οπότε η παραπάνω συνάρτηση πιθανότητας παριστάνει μια οικογένεια κατανομών. Επομένως αν θέλουμε να υπολογίσουμε πιθανότητες ολοκληρώνοντας την παραπάνω συνάρτηση θα πρέπει για κάθε ζευγάρι τιμών για τα μ, σ^2 να υπολογίζουμε διαφορετικό ολοκλήρωμα.

Για $\mu = 0$ και $\sigma^2 = 1$ έχουμε την τυποποιημένη κανονική κατανομή, $N(0,1)$. Επίσης σημειώστε ότι ο γραμμικός μετασχηματισμός μίας τυχαίας μεταβλητής που ακολουθεί την κανονική κατανομή, ακολουθεί και αυτός την κανονική κατανομή. Αυτό έχει ως συνέπεια αν $X \sim N(\mu, \sigma^2)$ τότε τυποποιώντας παίρνω πάλι κανονική κατανομή!

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1).$$

Αυτό με διευκολύνει στο να υπολογίζω πιθανότητες που αφορούν την $N(\mu, \sigma^2)$ αν ξέρω την αθροιστική συνάρτηση της $N(0,1)$. Δηλαδή

$$\begin{aligned} P\left(x_1 < X < x_2\right) &= P\left(\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{x_2-\mu}{\sigma}\right) \\ &= P\left(\frac{x_1-\mu}{\sigma} < Z < \frac{x_2-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right) \end{aligned}$$

όπου με $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$ συμβολίζουμε την αθροιστική συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής. Επίσης, καθώς είναι συμμετρική γύρω από το μηδέν θα έχω $\Phi(x) = 1 - \Phi(-x)$.

Αντίστροφα, αν έχω $Z \sim N(0,1)$ τότε $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$. Σημειώστε ότι αυτό δεν ισχύει γενικά σε όλες τις κατανομές!

Τέλος, αν έχουμε X_1, X_2 ανεξάρτητες τυχαίες μεταβλητές με $X_1 \sim N(\mu_1, \sigma_1^2)$ και $X_2 \sim N(\mu_2, \sigma_2^2)$ τότε

$$Y = X_1 + X_2 \sim N\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

Κατανομή χ^2

Αν $Z \sim N(0,1)$ τότε $Z^2 \sim \chi^2(1)$. Επίσης, αν έχουμε n ανεξάρτητες τυχαίες μεταβλητές $Z_1, Z_2, \ldots, Z_n$ όπου καθεμία κατανέμεται με την $N(0,1)$ τότε $\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$. Η παράμετρος n λέγεται "βαθμοί ελευθερίας" της κατανομής.

Είναι εύκολο να δούμε ότι αν $X \sim \chi^2(n)$ τότε ο μέσος της X θα είναι $E(X) = E\left(\sum_{i=1}^{n} Z_i^2\right) = \sum_{i=1}^{n} E(Z_i^2) = \sum_{i=1}^{n} 1 = n$.

Επίσης θα έχουμε $Var(X) = Var\left(\sum_{i=1}^{n} Z_i^2\right) = \sum_{i=1}^{n} Var(Z_i^2) + 2\sum_{i<j} Cov(Z_i^2, Z_j^2) \overset{\text{ανεξ.}}{=} \sum_{i=1}^{n} Var(Z_i^2) = \sum_{i=1}^{n} 2 = 2n$, καθώς για την τυποποιημένη κανονική κατανομή έχουμε $Var(Z_i^2) = 2$ (χωρίς απόδειξη).

Κατανομή Student t

Για $Z \sim N(0,1)$ και $U \sim \chi^2(v)$ και Z, U ανεξάρτητες έχουμε ότι $\frac{Z}{\sqrt{U/v}} \sim t_v$ (κατανομή t με v βαθμούς ελευθερίας). Επίσης, η κατανομή είναι συμμετρική γύρω από τον μέσο και ο μέσος είναι το 0.