Αναμενόμενη Τιμή & Ροπές Κατανομών

Όπως στην περιγραφική στατιστική περιγράφαμε ένα σύνολο δεδομένων, μπορούμε με παρόμοιο τρόπο να περιγράψουμε την κατανομή πιθανότητας μιας τυχαίας μεταβλητής. Σημειώστε ότι η τυχαία μεταβλητή (μαζί με την κατανομή πιθανότητάς της) παίζει το ρόλο του πληθυσμού στη στατιστική, και η περιγραφή του πληθυσμού τελικά συνίσταται στην περιγραφή της κατανομής πιθανότητας μιας τυχαίας μεταβλητής. Όπως είδαμε το μέτρο πιθανότητας μιας τυχαίας μεταβλητής μπορεί να αναπαρασταθεί πλήρως χρησιμοποιώντας συναρτήσεις κατανομών (αθροιστική και συνάρτηση πιθανότητας). Λειτουργώντας αφαιρετικά (όπως είχαμε κάνει και στην περιγραφική στατιστική με τα αριθμητικά μέτρα) αυτές τις συναρτήσεις κατανομών μπορούμε να τις περιγράψουμε με κάποιες παραμέτρους, που περιγράφουν για παράδειγμα τη θέση και τη διασπορά μιας κατανομής.

Για το σκοπό αυτό χρειαζόμαστε την έννοια της αναμενόμενης τιμής μιας τυχαίας μεταβλητής η οποία αντικαθιστά την έννοια του μέσου όρου που είδαμε στην περιγραφική στατιστική. Η αναμενόμενη τιμή είναι ένας σταθμισμένος μέσος όρος των τιμών που μπορεί να πάρει η τυχαία μεταβλητή όπου οι σταθμίσεις είναι οι πιθανότητες των τιμών που αυτή παίρνει. Οπότε δουλεύει όπως και στην περιγραφική στατιστική αλλά τώρα αντί για σχετικές συχνότητες έχουμε πιθανότητες.

Αναμενόμενη Τιμή

Η αναμενόμενη τιμή ή ο μέσος (μ) μιας διακριτής τυχαίας μεταβλητής X ορίζεται ως:

$$E(X) = \sum_x x f(x)$$

όπου f είναι η συνάρτηση μάζας πιθανότητας της X, και η άθροιση γίνεται ως προς όλες τις τιμές που παίρνει η X με θετική πιθανότητα (δηλ. $f(x) > 0$). Σημειώστε ότι η αναμενόμενη τιμή μιας τυχαίας μεταβλητής δίνει μια αριθμητική τιμή, οπότε δεν είναι τυχαία μεταβλητή.

Αν η τυχαία μεταβλητή είναι συνεχής τότε η αναμενόμενη τιμή της θα ορίζεται ως:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Παράδειγμα (διακριτή) Αν η X έχει συνάρτηση (μάζας) πιθανότητας

$$f(x) = \begin{cases} \frac{1}{2} & \text{για } x = 0, 1 \\ 0 & \text{αλλιώς} \end{cases}$$

τότε η αναμενόμενη τιμή της είναι $E(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$.

Παράδειγμα (συνεχής) Αν η X έχει συνάρτηση (πυκνότητας) πιθανότητας

$$f(x) = \begin{cases} 1 & \text{για } x \in [0, 1] \\ 0 & \text{αλλιώς} \end{cases}$$

τότε η αναμενόμενη τιμή της είναι $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$.

Αναμενόμενη τιμή μετασχηματισμού τυχαίας μεταβλητής

Αν έχουμε μια τυχαία μεταβλητή X τότε για οποιαδήποτε συνάρτηση $g : \mathbb{R} \to \mathbb{R}$ η $Y = g(X)$ είναι επίσης τυχαία μεταβλητή, η οποία έχει προκύψει από μετασχηματισμό της X. Οπότε η καινούρια τυχαία μεταβλητή Y θα έχει συνάρτηση κατανομής η οποία προφανώς θα συνδέεται με την συνάρτηση κατανομής της X. Στην περίπτωση π.χ. που η X είναι διακριτή, η συνάρτηση (μάζας) πιθανότητας της $g(X)$ θα είναι η συνάρτηση $f_Y(y) = P(g(X) = y) = \sum_{x:g(x)=y} f(x)$.

Παρ' όλα αυτά, δεν χρειάζεται να βρούμε την συνάρτηση πιθανότητας της Y για να υπολογίσουμε την αναμενόμενη τιμή της, αρκεί να ξέρουμε την συνάρτηση πιθανότητας της X! Μπορούμε να υπολογίσουμε την αναμενόμενη τιμή της τυχαίας μεταβλητής $g(X)$ ως εξής:

Διακριτή περίπτωση: $E(Y) = E[g(X)] = \sum_x g(x)f(x)$

Συνεχής περίπτωση: $E(Y) = E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$

Για παράδειγμα η αναμενόμενη τιμή της X^2 είναι:

Διακριτή περίπτωση: $E(Y) = E(X^2) = \sum_x x^2 f(x)$

Συνεχής περίπτωση: $E(Y) = E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$

Αναμενόμενη τιμή συνάρτησης 2 τυχαίων μεταβλητών

Ας πούμε ότι έχουμε δύο τυχαίες μεταβλητές X, Y με από κοινού συνάρτηση πιθανότητας $f_{XY}(x, y)$. Τότε μπορώ να υπολογίσω την αναμενόμενη τιμή οποιασδήποτε συνάρτησης που εξαρτάται και από τις 2 τυχαίες μεταβλητές. Έστω $g : \mathbb{R}^2 \to \mathbb{R}$ μια τέτοια συνάρτηση. Μπορώ να υπολογίσω την αναμενόμενη τιμή της $Z = g(X, Y)$ ως εξής:

Διακριτή περίπτωση: $E(Z) = E[g(X, Y)] = \sum_y \sum_x g(x, y) f_{XY}(x, y)$

Συνεχής περίπτωση: $E(Z) = E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)dxdy$

Για παράδειγμα η αναμενόμενη τιμή της $Z = XY$ είναι:

Διακριτή περίπτωση: $E(Z) = E(XY) = \sum_y \sum_x xy f_{XY}(x, y)$

Συνεχής περίπτωση: $E(Z) = E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x, y)dxdy$

Κανόνες υπολογισμού αναμενόμενων τιμών

Έστω τυχαίες μεταβλητές X, Y

1. $E(a) = a$ για οποιαδήποτε σταθερά a.
2. $E(aX) = aE(X)$
3. $E(X + Y) = E(X) + E(Y)$
   (a) $E(a + bX) = a + bE(X)$, όπου a, b σταθερές.
   (b) $E(g(X) + h(Y)) = E[g(X)] + E[h(Y)]$

Ροπές τυχαίων μεταβλητών

- Η ροπή k τάξης (γύρω από το μηδέν), όπου $k = 1, 2, 3, \ldots$, ορίζεται ως
  - Διακριτή: $E(X^k) = \sum_x x^k f(x)$
  - Συνεχής: $E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx$

Για $k = 1$ έχουμε τον μέσο: $\mu = E(X)$

- Η κεντρική ροπή k τάξης (γύρω από τον μέσο), όπου $k = 1, 2, 3, \ldots$, ορίζεται ως
  - Διακριτή: $E\left[(X - \mu)^k\right] = \sum_x (x - \mu)^k f(x)$
  - Συνεχής: $E\left[(X - \mu)^k\right] = \int_{-\infty}^{\infty} (x - \mu)^k f(x)dx$

Για $k = 2$ έχουμε την διακύμανση: $\sigma^2 = Var(X) = E\left[(X - \mu)^2\right]$

Για $k = 3$ έχουμε $E\left[(X - \mu)^3\right]$ και για $k = 4$ έχουμε $E\left[(X - \mu)^4\right]$. Μπορούμε να κατασκευάσουμε τον συντελεστή συμμετρίας και τον συντελεστή κύρτωσης ως $\frac{E\left[(X-\mu)^3\right]}{\sigma^3}$ και $\frac{E\left[(X-\mu)^4\right]}{\sigma^4}$ αντίστοιχα.

Κανόνες υπολογισμού διακύμανσης

Έστω τυχαίες μεταβλητές X, Y

1. $Var(a) = 0$ για οποιαδήποτε σταθερά a
2. $Var(aX) = a^2 Var(X)$
3. $Var(X) = E(X^2) - [E(X)]^2$
4. $Var(X + a) = Var(X)$ για οποιαδήποτε σταθερά a
5. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
6. Γενικότερα $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$

όπου $Cov(X, Y) = E\left[(X - E(X))(Y - E(Y))\right]$.

Συνδιακύμανση και συντελεστής συσχέτισης Η συνδιακύμανση δύο τυχαίων μεταβλητών X και Y ορίζεται ως

$$\begin{aligned} Cov(X,Y) &= E\left\{[X - E(X)]\,[Y - E(Y)]\right\} \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Αν

- $Cov(X, Y) > 0$ τότε λέμε ότι οι X, Y είναι θετικά συσχετισμένες
- $Cov(X, Y) < 0$ τότε λέμε ότι οι X, Y είναι αρνητικά συσχετισμένες
- $Cov(X, Y) = 0$ τότε λέμε ότι οι X, Y είναι ασυσχέτιστες

Προσέξτε ότι $Cov(X, X) = Var(X)$.

Ο συντελεστής συσχέτισης των μεταβλητών X, Y ορίζεται ως

$$\rho_{XY} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

και παίρνει τιμές μεταξύ 0 και 1, δηλαδή $-1 \le \rho_{XY} \le 1$.

Σημειώστε ότι δύο μεταβλητές που είναι ασυσχέτιστες δεν είναι απαραίτητα και ανεξάρτητες. Η συνδιακύμανση είναι ένα μέτρο γραμμικής εξάρτησης. Αλλά αν είναι ανεξάρτητες τότε είναι και ασυσχέτιστες καθώς τότε $f_{XY}(x, y) = f_X(x)f_Y(y)$ οπότε $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy = \int_{-\infty}^{\infty} xf_X(x)\int_{-\infty}^{\infty} yf_Y(y)dydx = \int_{-\infty}^{\infty} xf_X(x)E(Y)dx = E(Y)\int_{-\infty}^{\infty} xf_X(x)dx = E(Y)E(X)$ άρα $Cov(X, Y) = 0$.

Δεσμευμένη αναμενόμενη τιμή & Δεσμευμένη διακύμανση Αν έχουμε 2 τυχαίες μεταβλητές X, Y τότε μπορούμε να ορίσουμε τη δεσμευμένη αναμενόμενη τιμή της Y δεδομένης κάποιας τιμής για την X ως εξής

- Διακριτή περίπτωση

$$E(Y|X \quad = \quad x) = \sum_y y f_{Y|X}(y|x)$$

- Συνεχής περίπτωση

$$E(Y|X \quad = \quad x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy$$

Αντίστοιχα και για την X δεδομένης κάποιας τιμής για τη Y.

Η δεσμευμένη διακύμανση της Y δεδομένης κάποιας τιμής για την X ορίζεται ως εξής:

$$\begin{aligned} V(Y|X) &= E\left\{[Y - E(Y|X=x)]^2 | X = x\right\} \\ &= E(Y^2|X=x) - [E(Y|X=x)]^2 \end{aligned}$$

δηλαδή στον τύπο της διακύμανσης της Y αντικαθιστούμε όλες τις αναμενόμενες τιμές με δεσμευμένες αναμενόμενες τιμές.