ΒΑΣΙΚΕΣ ΣΥΝΕΧΕΙΣ ΚΑΤΑΝΟΜΕΣ (ΣΥΝΕΧΕΙΑ)

Χαράλαμπος Α. Χαραλαμπίδης

9 Δεκεμβρίου 2009

Η σημαντικότερη κατανομή πιθανότητας της Θεωρίας Πιθανοτήτων και της Στατιστικής, με μεγάλο πεδίο εφαρμογών, είναι η **κανονική κατανομή**.

Η κατανομή αυτή χρησιμοποιήθηκε αρχικά από τους De Moivre και Laplace ως προσέγγιση της διωνυμικής κατανομής για μεγάλο αριθμό δοκιμών ν.

Ο Gauss χρησιμοποίησε την κανονική κατανομή ως προσεγγιστική της κατανομής των τυχαίων σφαλμάτων διαφόρων μετρήσεων. Για το λόγο αυτό η κατανομή αυτή αναφέρεται συχνά ως **Γκαουσιανή** κατανομή.

Η ονομασία **κανονική** δόθηκε πιο πρόσφατα από τον Karl Pearson.

Ορισμός

Έστω X μια συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty, \tag{1}$$

όπου $-\infty < \mu < \infty, 0 < \sigma < \infty$ είναι παράμετροι. Η κατανομή της τ.μ. X συμβολίζεται $N(\mu, \sigma^2)$ και καλείται κανονική κατανομή με παραμέτρους μ και σ^2.

Σημειώνουμε ότι η συνάρτηση (1) είναι μη αρνητική και εκτελώντας διαδοχικά τους μετασχηματισμούς $z = (x - \mu)/\sigma$ και $u = z/\sqrt{2}$ και το ολοκλήρωμα του Euler,

$$\int_0^\infty e^{-u^2} du = \frac{\sqrt{\pi}}{2},$$

συμπεραίνουμε ότι

$$\begin{aligned}\int_{-\infty}^\infty f(x)dx &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-z^2/2} dz = \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-u^2} du = 1,\end{aligned}$$

όπως απαιτείται από τον ορισμό της συνάρτησης πυκνότητας.

Θεώρημα

Έστω ότι η τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή με συνάρτηση πυκνότητας (1). Τότε η μέση τιμή και η διασπορά αυτής δίνονται από τις

$$E(X) = \mu, \quad V(X) = \sigma^2.$$

Απόδειξη.

Η μέση τιμή, σύμφωνα με τον ορισμό, δίνεται από την

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Θέτοντας $z = (x-\mu)/\sigma$, οπότε $x = \mu + \sigma z$, συνάγουμε την

$$\begin{aligned} E(X) &= \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz + \sigma \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz \\ &= \mu + \sigma \frac{1}{\sqrt{2\pi}} \Big[-e^{-z^2/2} \Big]_{-\infty}^{\infty} = \mu \end{aligned}$$

Η διασπορά δίνεται από την

$$V(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Θέτοντας και πάλιν $z = (x-\mu)/\sigma$, οπότε $x = \mu + \sigma z$, συνάγουμε την

$$V(X) = \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz$$

και ολοκληρώνοντας κατά παράγοντες, παίρνουμε

$$V(X) = \sigma^2 \frac{1}{\sqrt{2\pi}} \Big[-ze^{-z^2/2} \Big]_{-\infty}^{\infty} + \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = \sigma^2.$$

Το διάγραμμα της συνάρτησης πυκνότητας της κανονικής κατανομής είναι κωδωνοειδές.
Η μέση τιμή μ προσδιορίζει τη θέση της καμπύλης ως προς τον άξονα των x και η τυπική απόκλιση σ το οξύ ή πεπλατισμένο του σχήματος.

Η ειδική περίπτωση $\mu = 0$, $\sigma = 1$ παρουσιάζει ιδιαίτερο ενδιαφέρον. Σχετικά, η τυποποιημένη τυχαία μεταβλητή $Z = (X - \mu)/\sigma$ έχει συνάρτηση πυκνότητας

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < x < \infty \tag{2}$$

και συνάρτηση κατανομής

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad -\infty < z < \infty. \tag{3}$$

Η κατανομή $N(0, 1)$ καλείται *τυποποιημένη κανονική κατανομή*.

Η πινακοποίηση της συνάρτησης κατανομής της τυποποιημένης κανονικής κατανομής (3) διευκολύνει τη χρήση της. Τον σχετικό πίνακα συνεπικουρεί και η ακόλουθη ιδιότητά της.

Θεώρημα

Η συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής ικανοποιεί τη σχέση

$$\Phi(-z) = 1 - \Phi(z), \quad -\infty < z < \infty. \tag{4}$$

Απόδειξη.

Η συνάρτηση

$$\Phi(-z) = \int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt,$$

θέτοντας $t = -u$, μετσχηματίζεται στην

$$\Phi(-z) = -\int_{\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = \int_{z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

Επομένως

$$\begin{aligned} \Phi(z) + \Phi(-z) &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt + \int_{z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 1 \end{aligned}$$

$\square$

Ο υπολογισμός πιθανοτήτων που αφορούν μία κανονική τυχαία μεταβλητή X με μέση τιμή $E(X) = \mu$ και διασπορά $V(X) = \sigma^2$ ανάγεται στον υπολογισμό πιθανοτήτων που αφορούν την τυποποιημένη τυχαία μεταβλητή $Z = (X - \mu)/\sigma$ και γίνεται με τη χρήση των πινάκων της $\Phi(z)$, σύμφωνα με το ακόλουθο πόρισμα.

Πόρισμα

Αν η τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή $N(\mu, \sigma^2)$, τότε

$$P(a \leq X \leq \beta) = \Phi\left(\frac{\beta - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right), \quad a \leq \beta \tag{5}$$

και ειδικότερα

$$P(X \leq \beta) = \Phi\left(\frac{\beta - \mu}{\sigma}\right), \quad P(X \geq a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right). \tag{6}$$

Απόδειξη.

Χρησιμοποιώντας το ότι η τυποποιημένη τ.μ. $Z = (X - \mu)/\sigma$ ακολουθεί την $N(0, 1)$ με συνάρτηση κατανομής τη $\Phi(z)$, παίρνουμε τη ζητούμενη έκφραση

$$P(a \leq X \leq \beta) = P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{\beta-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{\beta-\mu}{\sigma}\right)$$
$$= \Phi\left(\frac{\beta-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Επίσης

$$P(X \leq \beta) = P\left(\frac{X-\mu}{\sigma} \leq \frac{\beta-\mu}{\sigma}\right) = P\left(Z \leq \frac{\beta-\mu}{\sigma}\right) = \Phi\left(\frac{\beta-\mu}{\sigma}\right),$$

και

$$P(X \geq a) = 1 - P(X \leq a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Παράδειγμα

Έστω ότι η τυχαία μεταβλητή X, η οποία εκφράζει ποσοτικά κάποιο χαρακτηριστικό (ύψος, βάρος) των ατόμων ενός πληθυσμού, ακολουθεί την κανονική κατανομή $N(\mu, \sigma^2)$.
Να υπολογισθεί η πιθανότητα η X να απέχει από τον μέσο μ το πολύ κ τυπικές αποκλίσεις σ, για $\kappa = 1, 2, 3$.

Εισάγοντας την τυποποιημένη τ.μ. $Z = (X - \mu)/\sigma$, συνάγουμε για τη ζητούμενη πιθανότητα την έκφραση

$$\begin{aligned} P(|X - \mu| \leq \kappa\sigma) &= P(|Z| \leq \kappa) = P(-\kappa \leq Z \leq \kappa) \\ &= \Phi(\kappa) - \Phi(-\kappa) = 2\Phi(\kappa) - 1. \end{aligned}$$

Από τον πίνακα της τυποποιημένης κανονικής κατανομής έχουμε $\Phi(1) = 0{,}8413$, $\Phi(2) = 0{,}9772$ και $\Phi(3) = 0{,}9987$ και έτσι

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(|X - \mu| \leq \sigma) = 2\Phi(1) - 1 = 0{,}6826 \cong 68\%,$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(|X - \mu| \leq 2\sigma) = 2\Phi(2) - 1 = 0{,}9544 \cong 95\%,$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = P(|X - \mu| \leq 3\sigma) = 2\Phi(3) - 1 = 0{,}9974 \cong 99{,}7\%,$$

Επομένως, *το 68% περίπου των τιμών ενός κανονικού πληθυσμού βρίσκονται σε απόσταση το πολύ μιας τυπικής απόκλισης, το 95% περίπου σε απόσταση δύο τυπικών αποκλίσεων και το 99,7% περίπου σε απόσταση τριών αποκλίσεων από τον μέσο μ του πληθυσμού.*
Τα συμπεράσματα αυτά είναι χρήσιμα στη στατιστική συμπερασματολογία.

Παράδειγμα

Ας υποθέσουμε ότι η διάρκεια κύησης X μιας γυναίκας ακολουθεί την κανονική κατανομή με μέση τιμή $\mu = 270$ ημέρες και τυπική απόκλιση $\sigma = 30$ ημέρες. Να υπολογισθεί η πιθανότητα η διάρκεια της κύησης να είναι μικρότερη από επτά μήνες.

Εισάγοντας την τυποποιημένη τ.μ. $Z = (X - \mu)/\sigma$, όπου $\mu = 270$ και $\sigma = 30$, συνάγουμε για τη ζητούμενη πιθανότητα την έκφραση

$$P(X < 210) = P\left(\frac{X-270}{30} < \frac{210-270}{30}\right) = P(Z < -2) = \Phi(-2) = 1 - \Phi(2).$$

Από τον πίνακα της τυποποιημένης κανονικής κατανομής έχουμε $\Phi(2) = 0{,}9773$ και έτσι

$$P(X < 210) = \Phi(-2) = 1 - \Phi(2) = 1 - 0{,}9773 = 0{,}0227 \cong 2\%.$$

Παράδειγμα

Έστω ότι ο χρόνος εμφάνισης X ενός φωτογραφικού φιλμ ακολουθεί κανονική κατανομή με μέση τιμή $\mu = 30$ min και τυπική απόκλιση $\sigma = 1{,}2$ min.
Να υπολογισθούν (α) η πιθανότητα ο χρόνος εμφάνισης ενός φιλμ να μη υπερβεί τα 28 min και
(β) η πιθανότητα σε δύο τουλάχιστο από 10 φιλμ ο χρόνος εμφάνισης να μη υπερβεί τα 28 min.

(α) Η πιθανότητα ο χρόνος εμφάνισης να μην υπερβεί τα 28 min δίνεται από την

$$P(X \leq 28) = P\left(\frac{X-30}{1{,}2} \leq \frac{28-30}{1{,}2}\right)$$
$$= P(Z \leq -1{,}67) = 1 - \Phi(1{,}67) = 0{,}0475 \cong 5\%.$$

(β) Η πιθανότητα σε δύο τουλάχιστο από 10 φιλμ ο χρόνος εμφάνισης να μη υπερβεί τα 28 min βρίσκεται αν θεωρήσουμε τον αριθμό Y των φιλμ (από τα 10) με χρόνο εμφάνισης το πολύ 28 min.
Τότε η τ.μ. Y ακολουθεί τη διωνυμική κατανομή, με $\nu = 10$ δοκιμές και πιθανότητα επιτυχίας $p = P(X \leq 28) = 0{,}05$, και η ζητούμενη πιθανότητα είναι

$$\begin{aligned} P(Y \geq 2) &= 1 - P(Y < 2) = 1 - P(Y = 0) - P(Y = 1) \\ &= 1 - \binom{10}{0}(0{,}05)^0(0{,}95)^{10} - \binom{10}{1}(0{,}05)^1(0{,}95)^9 = 0{,}086. \end{aligned}$$

4.1. Έστω ότι η τυχαία μεταβλητή X έχει την ομοιόμορφη κατανομή στο διάστημα $[\alpha, \beta]$. Αν $E(X) = 1$ και $V(X) = 3$,
(α) να υπολογισθούν οι σταθερές α και β,
(β) να προσδιορισθεί η πυκνότητα της τυχαίας μεταβλητής $Y = |X|$ και
(γ) να βρεθούν οι $E(Y)$ και $V(Y)$.

(α)

$$E(X) = \frac{\alpha + \beta}{2} = 1, \quad V(X) = \frac{(\beta - \alpha)^2}{12} = 3$$

$$\alpha + \beta = 2, \quad (\beta - \alpha)^2 = 36, \quad \beta - \alpha > 0,$$

$$\alpha = -2, \quad \beta = 4$$

Επομένως

$$f_X(x) = \frac{1}{6}, \quad -2 \leq x \leq 4$$

(β)

$$f_Y(y) = \begin{cases} f_X(y) + f_X(-y), & 0 \leq y \leq 2 \\ f_X(y), & 2 \leq x \leq 4. \end{cases}$$

Συνεπώς

$$f_Y(y) = \begin{cases} 1/3, & 0 \leq y \leq 2 \\ 1/6, & 2 \leq x \leq 4. \end{cases}$$

(γ)

$$E(Y) = \frac{1}{3}\int_0^2 y dy + \frac{1}{6}\int_2^4 y dy = \left[\frac{y^2}{6}\right]_0^2 + \left[\frac{y^2}{12}\right]_2^4 = \frac{8+16-4}{12} = \frac{5}{3}$$

$$E(Y^2) = E(X^2) = \frac{a^2 + a\beta + \beta^2}{3} = 4, \ \ V(Y) = 4 - \left(\frac{5}{3}\right)^2 = \frac{36-25}{9} = \frac{11}{9}.$$