Συνεχείς τυχαίες μεταβλητές
Περίληψη

Αντώνης Οικονόμου
aeconom@math.uoa.gr

24 Απριλίου 2010

Μια τυχαία μεταβλητή X λέγεται (απόλυτα) **συνεχής** αν αντιστοιχεί διαισθητικά σε κάποιο συνεχές αριθμητικό χαρακτηριστικό ενός πειράματος τύχης. Σε μαθηματικό επίπεδο η X είναι συνεχής αν υπάρχει μια μη-αρνητική συνάρτηση f, που λέγεται **συνάρτηση πυκνότητας πιθανότητας**, τέτοια ώστε για κάθε σύνολο B να ισχύει

$$P(X \in B) = \int_B f(x)dx.$$

Ειδικότερα για τη συνάρτηση κατανομής της X έχουμε ότι δίνεται τότε από τη σχέση

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du.$$

Αντίστροφα, αν δίνεται η συνάρτηση κατανομής $F(x)$ μιας συνεχούς τυχαίας μεταβλητής X, τότε μπορούμε να βρούμε τη συνάρτηση πυκνότητας πιθανότητας $f(x)$ από τη σχέση

$$f(x) = \frac{d}{dx}F(x).$$

Η μέση τιμή της X ορίζεται τότε από τη σχέση

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

ενώ η διασπορά ορίζεται, όπως και για τις διακριτές τυχαίες μεταβλητές, από τη σχέση

$$Var[X] = E[(X - E[X])^2].$$

Η μέση τιμή $E[g(X)]$ μιας συνάρτησης τυχαίας μεταβλητής μπορεί να βρεθεί άμεσα από τον τύπο

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

Οι χρήσιμοι τύποι για τον υπολογισμό της μέσης τιμής και της διασποράς μιας τυχαίας μεταβλητής που ισχύουν για διακριτές τυχαίες μεταβλητές συνεχίζουν να ισχύουν. Συγκεκριμένα, αν X είναι μια τυχαία μεταβλητή, τότε ισχύουν οι σχέσεις:

$$\begin{aligned} E[aX+b] &= aE[X]+b, \\ Var[aX+b] &= a^2Var[X], \\ Var[X] &= E[X^2]-E[X]^2. \end{aligned}$$

Για μη-αρνητικές τυχαίες μεταβλητές είναι χρήσιμος ένας εναλλακτικός τύπος υπολογισμού της μέσης τιμής, μέσω της συνάρτησης κατανομής. Συγκεκριμένα έχουμε ότι

$$X \geq 0 \quad \Rightarrow \quad E[X] = \int_0^\infty (1 - F(x))dx.$$

Ένα σημαντικό πείραμα τύχης είναι η τυχαία επιλογή σημείου από ένα διάστημα $[a,b]$. Αν X είναι το αποτέλεσμα του πειράματος, τότε λέμε ότι η X ακολουθεί τη **συνεχή ομοιόμορφη κατανομή** στο $[a,b]$ (Uniform($[a,b]$)). Η συνάρτηση πυκνότητας πιθανότητας της X δίνεται τότε ως

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{αν } a \leq x \leq b \\ 0 & \text{διαφορετικά.} \end{cases}$$

Η μέση τιμή και η διασπορά της είναι

$$E[X] = \frac{a+b}{2}, \quad Var[X] = \frac{(b-a)^2}{12}.$$

Μια άλλη σημαντική συνεχής κατανομή παρουσιάζεται στη μελέτη πολλών χαρακτηριστικών πληθυσμών. Πρόκειται για την **κανονική κατανομή** με παραμέτρους μ και σ^2 ($N(\mu, \sigma^2)$) που έχει συνάρτηση πυκνότητας πιθανότητας

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

Η μέση τιμή και η διασπορά της είναι

$$E[X] = \mu, \quad Var[X] = \sigma^2.$$

Μια σημαντική ιδιότητα της κανονικής κατανομής είναι ότι κάθε γραμμική συνάρτησή της ακολουθεί επίσης κανονική κατανομή. Συγκεκριμένα έχουμε ότι

$$X \text{ ακολουθεί την } N(\mu, \sigma^2), \quad a \in \mathbb{R} \setminus \{0\}, \quad b \in \mathbb{R} \Rightarrow aX + b \text{ ακολουθεί την } N(a\mu + b, a^2\sigma^2).$$

Από την ιδιότητα αυτή έπεται ότι δοθείσης μιας κανονικής τυχαίας μεταβλητής με μέση τιμή μ και διασπορά σ^2, η τυχαία μεταβλητή

$$Z = \frac{X - \mu}{\sigma}$$

είναι κανονική με μέση τιμή 0 και διασπορά 1. Μια τέτοια τυχαία μεταβλητή αναφέρεται ως **τυποποιημένη κανονική**. Όλοι οι υπολογισμοί που αφορούν τη X μπορούν να γίνουν χρησιμοποιώντας τη Z. Επειδή η συνάρτηση κατανομής της κανονικής κατανομής δεν δίνεται σε κλειστή μορφή αλλά μέσω ολοκληρώματος, χρησιμοποιούμε τους πίνακες τιμών που υπάρχουν για τη **συνάρτηση κατανομής $\Phi(z)$ της τυποποιημένης κανονικής κατανομής**. Έτσι έχουμε ότι

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

Η σημασία της κανονικής κατανομής έγγειται στο ότι προσεγγίζει ικανοποιητικά την κατανομή αθροισμάτων ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Το αποτέλεσμα αυτό που αναφέρεται ως **κεντρικό οριακό θεώρημα** αποτελεί ένα από τα σημαντικότερα θεωρήματα της θεωρίας πιθανοτήτων. Στην πιο απλή μορφή του λέει ότι για αρκετά μεγάλα n, η διωνυμική κατανομή $\text{Bin}(n,p)$, δηλαδή το πλήθος των επιτυχιών σε n ανεξάρτητες δοκιμές Bernoulli με πιθανότητα επιτυχίας p σε κάθε δοκιμή, μπορεί να προσεγγιστεί από την κανονική κατανομή με ίδιες μέση τιμή και διασπορά, δηλαδή από την κανονική κατανομή $N(np, np(1-p))$. Πιο συγκεκριμένα έχουμε **το θεώρημα των DeMoivre-Laplace**, σύμφωνα με το οποίο, αν S_n είναι ο αριθμός των επιτυχιών σε n ανεξάρτητες δοκιμές Bernoulli, με πιθανότητα επιτυχίας p ανά δοκιμή, τότε για κάθε πραγματικούς αριθμούς $a < b$ έχουμε

$$\lim_{n\to\infty} P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \Phi(b) - \Phi(a).$$

Μια άλλη σημαντική συνεχής κατανομή χρησιμοποιείται για τη μοντελοποίηση χρόνων ζωής εξαρτημάτων και χρόνων γενικότερα. Η **εκθετική κατανομή** με παράμετρο λ (Exp(λ)) έχει συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{αν } x \geq 0 \\ 0 & \text{διαφορετικά.} \end{cases}$$

Η μέση τιμή και η διασπορά της είναι

$$E[X] = \frac{1}{\lambda}, \quad Var[X] = \frac{1}{\lambda^2}.$$

Μια βασική ιδιότητα της εκθετικής κατανομής είναι η λεγόμενη **αμνήμονη ιδιότητα** η οποία λέει ότι για οποιαδήποτε $s, t > 0$, ισχύει

$$P(X > s + t | X > t) = P(X > s).$$

Διαισθητικά, αν X παριστάνει το χρόνο ζωής ενός εξαρτήματος, η ιδιότητα αυτή λέει ότι η δεσμευμένη πιθανότητα ο υπολειπόμενος χρόνος ζωής του μηχανήματος να είναι μεγαλύτερος από s, δεδομένου ότι έχει ήδη ζήσει t χρονικές μονάδες είναι η ίδια με την πιθανότητα ο χρόνος ζωής ενός καινούργιου μηχανήματος να είναι μεγαλύτερος από s. Δηλαδή η ηλικία δεν παίζει ρόλο για το πόσο θα ζήσει ακόμα το εξάρτημα. Η εκθετική κατανομή είναι η μοναδική κατανομή που έχει την αμνήμονη ιδιότητα που περιγράψαμε παραπάνω (για κάθε $s, t > 0$).

Μια άλλη κατανομή που χρησιμοποιείται για τη μοντελοποίηση χρόνων ζωής είναι η **κατανομή Γάμμα** με παραμέτρους α και λ (Gamma(α, λ)), η οποία έχει συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{αν } x \geq 0. \\ 0 & \text{διαφορετικά.} \end{cases}$$

Η ποσότητα $\Gamma(\alpha)$ είναι η γνωστή συνάρτηση Γάμμα υπολογισμένη στο σημείο α, η οποία δίνεται από το ολοκλήρωμα

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}.$$

Η μέση τιμή και η διασπορά της είναι

$$E[X] = \frac{\alpha}{\lambda}, \quad Var[X] = \frac{\alpha}{\lambda^2}.$$

Η κατανομή Γάμμα γενικεύει την εκθετική κατανομή, αφού η εκθετική κατανομή προκύπτει για $\alpha = 1$. Γενικότερα η κατανομή Γάμμα είναι γνωστή ως κατανομή Erlang όταν η παράμετρος α είναι θετικός ακέραιος.