ΒΑΣΙΚΕΣ ΔΙΑΚΡΙΤΕΣ ΚΑΤΑΝΟΜΕΣ (Συνέχεια)

Χαράλαμπος Α. Χαραλαμπίδης

23 Νοεμβρίου 2009

Ορισμός

Έστω X ο αριθμός των δοκιμών μέχρι την πρώτη επιτυχία σε μια ακολουθία ανεξαρτήτων δοκιμών Bernoulli με πιθανότητα επιτυχίας p, σταθερή σε όλες τις δοκιμές.
Η κατανομή της τυχαίας μεταβλητής X καλείται γεωμετρική με παράμετρο p.

Θεώρημα

Η συνάρτηση πιθανότητας της γεωμετρικής κατανομής με παράμετρο p δίνεται από την

$$f(x) = P(X = x) = pq^{x-1}, \quad x = 1, 2, \dots \tag{1}$$

και η συνάρτηση κατανομής από την

$$F(x) = \begin{cases} 0, & -\infty < x < 1 \\ 1 - q^{[x]}, & 1 \leq x < \infty, \end{cases} \tag{2}$$

όπου $[x]$ παριστάνει το ακέραιο μέρος του x.

Απόδειξη.

Το ενδεχόμενο $\{X = x\}$, η πρώτη επιτυχία να πραγματοποιηθεί στη x-οστή δοκιμή, περιλαμβάνει ένα μόνο δειγματικό σημείο (στοιχειώδες ενδεχόμενο) και συγκεκριμένα το

$$\{(\alpha, \alpha, \ldots, \alpha, \varepsilon)\},$$

όπου στις $x - 1$ πρώτες θέσεις (δοκιμές) έχει αποτυχία και στη x-οστή θέση (δοκιμή) έχει επιτυχία.
Χρησιμοποιώντας ότι οι δοκιμές είναι ανεξάρτητες τούτο έχει πιθανότητα

$$q^{x-1}p.$$

Επομένως η συνάρτηση πιθανότητας της τ.μ. X είναι η

$$f(x) = P(X = x) = pq^{x-1}, \quad x = 1, 2, \ldots .$$

Σημειώνουμε ότι

$$f(x) > 0, \quad x = 1, 2, \ldots, \quad f(x) = 0, \quad x \notin \{1, 2, \ldots\}$$

και σύμφωνα με τον τύπο του αθροίσματος των απείρων όρων γεωμετρικής προόδου (σειράς),

$$\sum_{x=1}^{\infty} f(x) = p \sum_{x=1}^{\infty} q^{x-1} = p(1-q)^{-1} = 1,$$

όπως απαιτείται από τον ορισμό της συνάρτησης πιθανότητας.
Η συνάρτηση κατανομής δίνεται από την $F(x) = 0, -\infty < x < 1$, και την

$$F(x) = \sum_{\kappa \leq x} f(\kappa) = p \sum_{\kappa=0}^{[x]} q^{\kappa-1}, \quad 1 \leq x < \infty,$$

η οποία, χρησιμοποιώντας τον τύπο του αθροίσματος πεπερασμένου αριθμού όρων γεωμετρικής προόδου, συνεπάγεται την (2).

Θεώρημα

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

$$\mu = E(X) = \frac{1}{p}, \quad \sigma^2 = V(X) = \frac{q}{p^2}. \tag{3}$$

Απόδειξη.

Η μέση τιμή και η δεύτερης τάξης παραγοντική ροπή της γεωμετρικής κατανομής δίνονται από τις

$$\mu = E(X) = \sum_{x=1}^{\infty} xpq^{x-1} = p \sum_{x=1}^{\infty} xq^{x-1}$$

και

$$\mu_{(2)} = E[(X)_2] = E[X(X-1)] = pq \sum_{x=2}^{\infty} x(x-1)q^{x-2}.$$

Παρατηρούμε ότι, παραγωγίζοντας διαδοχικά ως προς q τη γεωμετρική σειρά

$$\sum_{x=0}^{\infty} q^x = (1-q)^{-1},$$

□

συνάγουμε τις σχέσεις

$$\sum_{x=1}^{\infty} xq^{x-1} = (1-q)^{-2}, \quad \sum_{x=2}^{\infty} x(x-1)q^{x-2} = 2(1-q)^{-3}.$$

Επομένως

$$\mu = E(X) = p\sum_{x=1}^{\infty} xq^{x-1} = \frac{p}{(1-q)^2} = \frac{1}{p},$$

και

$$\mu_{(2)} = E[(X)_2] = pq\sum_{x=2}^{\infty} x(x-1)q^{x-2} = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2},$$

οπότε

$$\sigma^2 = V(X) = E[(X)_2] + E(X) - [E(X)]^2 = \frac{2q}{p} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

Θεώρημα

Έστω ότι η τυχαία μεταβλητή X ακολουθεί τη γεωμετρική κατανομή με συνάρτηση πιθανότητας την (1). Τότε

$$P(X > \kappa + r | X > \kappa) = P(X > r), \quad \kappa, r = 0, 1, \ldots . \tag{4}$$

Απόδειξη.

Η δεσμευμένη πιθανότητα του ενδεχομένου $\{\omega : X(\omega) > \kappa + r\}$ δεδομένου του ενδεχομένου $\{\omega : X(\omega) > \kappa\}$, χρησιμοποιώντας την (2) και το ότι $\{\omega : X(\omega) > \kappa + r\} \subseteq \{\omega : X(\omega) > \kappa\}$, είναι ίση με

$$P(X > \kappa + r | X > \kappa) = \frac{P(X > \kappa + r, X > \kappa)}{P(X > \kappa)} = \frac{P(X > \kappa + r)}{P(X > \kappa)}$$
$$= \frac{1 - F(\kappa + r)}{1 - F(\kappa)} = \frac{q^{\kappa + r}}{q^{\kappa}} = q^r$$

και επειδή

$$P(X > r) = 1 - F(r) = q^r$$

συμπεραίνουμε την (4). $\square$

Σημειώνουμε ότι η ιδιότητα αυτή σημαίνει **έλλειψη μνήμης** της γεωμετρικής κατανομής με την ακόλουθη έννοια:
Η πιθανότητα να απαιτηθούν επιπρόσθετα περισσότερες από r δοκιμές μέχρι την πρώτη επιτυχία δεδομένου ότι δεν έχει πραγματοποιηθεί επιτυχία στις κ πρώτες δοκιμές είναι η ίδια με την (μη δεσμευμένη) πιθανότητα να απαιτηθούν περισσότερες από r δοκιμές μέχρι την πρώτη επιτυχία.
Επομένως η πληροφορία μη επίτευξης του στόχου (επιτυχία) ξεχνιέται και η προσπάθεια συνεχίζεται όπως όταν πρωταρχίζει.

Παρατήρηση

Η συνάρτηση πιθανότητας του αριθμού Y των αποτυχιών μέχρι την πρώτη επιτυχία υπολογίζεται με τη χρησιμοποίηση της σχέσης $Y = X - 1$ και της (1) ως εξής:

$$f_Y(y) = P(Y = y) = P(X = y + 1) = pq^y, \quad y = 0, 1, \ldots . \tag{5}$$

Η κατανομή της τ.μ. Y καλείται επίσης γεωμετρική με παράμετρο p. Η μέση τιμή και η διασπορά αυτής προκύπτουν από τις (3):

$$E(Y) = E(X) - 1 = \frac{q}{p}, \quad V(Y) = V(X) = \frac{q}{p^2}. \tag{6}$$

Παράδειγμα

Το κόστος εκτέλεσης για πρώτη φορά ενός συγκεκριμένου πειράματος είναι 100 Ευρώ.
Αν το πείραμα αποτύχει, για ορισμένες μεταβολές που πρέπει να γίνουν πριν από την επόμενη εκτέλεσή του απαιτείται ένα πρόσθετο ποσόν 20 Ευρώ.
Υποθέτουμε ότι οι δοκιμές είναι ανεξάρτητες με πιθανότητα επιτυχίας $p = 4/5$ *και ότι συνεχίζονται μέχρι την πρώτη επιτυχία.*
Να υπολογισθούν (α) η πιθανότητα να απαιτηθούν 4 το πολύ δοκιμές μέχρι την πρώτη επιτυχία και (β) το αναμενόμενο κόστος μέχρι την πρώτη επιτυχία.

(α) Ο αριθμός X των δοκιμών μέχρι την πρώτη επιτυχία ακολουθεί τη γεωμετρική κατανομή με συνάρτηση πιθανότητας

$$f(x) = P(X = x) = \frac{4}{5}\left(\frac{1}{5}\right)^{x-1}, \quad x = 1, 2, \ldots$$

και συνάρτηση κατανομής

$$F(x) = \begin{cases} 0, & -\infty < x < 1 \\ 1 - \left(\frac{1}{5}\right)^{[x]}, & 1 \leq x < \infty. \end{cases}$$

Επομένως

$$P(X \leq 4) = F(4) = 1 - \left(\frac{1}{5}\right)^{4} = 0{,}9984.$$

(β) Αν K είναι το κόστος μέχρι την πρώτη επιτυχία τότε

$$K = 100X + 20(X-1) = 120X - 20$$

και

$$E(K) = 120E(X) - 20.$$

Η μέση τιμή της τυχαίας μεταβλητής X είναι ίση με

$$E(X) = \frac{1}{p} = \frac{5}{4}$$

και συνεπώς

$$E(K) = 130.$$

Ορισμός

Έστω X ο αριθμός των δοκιμών μέχρι την r-οστή επιτυχία σε μια ακολουθία ανεξαρτήτων δοκιμών Bernoulli με πιθανότητα επιτυχίας p, σταθερή σε όλες τις δοκιμές.
Η κατανομή της τυχαίας μεταβλητής X καλείται κατανομή Pascal με παραμέτρους r και p.

Θεώρημα

Η συνάρτηση πιθανότητας της κατανομής Pascal με παραμέτρους r και p δίνεται από την

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r q^{x-r}, \quad x = r, r+1, \ldots . \tag{7}$$

Απόδειξη.

Το ενδεχόμενο $\{X = x\}$ περιλαμβάνει τα δειγματικά σημεία (στοιχειώδη ενδεχόμενα) $(\omega_1, \omega_2, \ldots, \omega_{x-1}, \varepsilon)$, με $r-1$ επιτυχίες στις $x-1$ πρώτες δοκιμές και επιτυχία στη x-οστή δοκιμή, τα οποία είναι πλήθους $\binom{x-1}{r-1}$, όσα και ο αριθμός των επιλογών των $r-1$ θέσεων για τις επιτυχίες από τις $x-1$ δυνατές θέσεις.

Επιπλέον κάθε τέτοιο δειγματικό σημείο έχει πιθανότητα $p^r q^{x-r}$. □

Επομένως

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r q^{x-r}, \quad x = r, r+1, \ldots, .$$

Σημειώνουμε ότι

$$f(x) > 0, \quad x = r, r+1, \ldots, \quad f(x) = 0, \quad x \notin \{r, r+1, \ldots\}$$

και χρησιμοποιώντας το αρνητικό διωνυμικό ανάπτυγμα,

$$\sum_{x=0}^{\infty} \binom{r+x-1}{x} t^x = (1-t)^{-r}, \quad -1 < t < 1, \tag{8}$$

συνάγουμε τη σχέση

$$\sum_{x=r}^{\infty} f(x) = \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r q^{x-r} = p^r \sum_{y=0}^{\infty} \binom{r+y-1}{y} q^y = p^r (1-q)^{-r} = 1,$$

όπως απαιτείται από τον ορισμό της συνάρτησης πιθανότητας.

Θεώρημα

Έστω ότι η τυχαία μεταβλητή X ακολουθεί την κατανομή Pascal με συνάρτηση πιθανότητας την (7).
Τότε η μέση τιμή και η διασπορά αυτής δίνονται από τις

$$\mu = E(X) = \frac{r}{p}, \quad \sigma^2 = V(X) = \frac{rq}{p^2}. \tag{9}$$

Απόδειξη.

Η μέση τιμή της τ.μ. X δίνεται από την

$$\mu = E(X) = \sum_{x=r}^{\infty} x\binom{x-1}{r-1}p^r q^{x-r},$$

οπότε, χρησιμοποιώντας τη σχέση

$$x\binom{x-1}{r-1} = x\frac{(x-1)!}{(r-1)!(x-r)!} = r\frac{x!}{r!(x-r)!} = r\binom{x}{x-r}$$

και την (8), συνάγουμε την έκφραση

$$E(X) = rp^r \sum_{x=r}^{\infty}\binom{x}{x-r}q^{x-r} = rp^r \sum_{y=0}^{\infty}\binom{r+y}{y}q^y = rp^r(1-q)^{-r-1} = \frac{r}{p}.$$

□

Η δεύτερης τάξης ανοδική παραγοντική ροπή της τ.μ. X δίνεται από την

$$\mu_{[2]} = E[X(X+1)] = \sum_{x=r}^{\infty} x(x+1)\binom{x-1}{r-1} p^r q^{x-r},$$

οπότε, χρησιμοποιώντας τη σχέση

$$\begin{aligned} x(x+1)\binom{x-1}{r-1} &= x(x+1)\frac{(x-1)!}{(r-1)!(x-r)!} \\ &= r(r+1)\frac{(x+1)!}{(r+1)!(x-r)!} = r(r+1)\binom{x+1}{x-r} \end{aligned}$$

και την (8), συνάγουμε την έκφραση

$$\begin{aligned} E[X(X+1)] &= r(r+1)p^r \sum_{x=r}^{\infty} \binom{x+1}{x-r} q^{x-r} = r(r+1)p^r \sum_{y=0}^{\infty} \binom{r+y+1}{y} q^y \\ &= r(r+1)p^r(1-q)^{-r-2} = r(r+1)p^{-2}. \end{aligned}$$

Επομένως, η διασπορά της τ.μ. X είναι

$$\sigma^2 = V(X) = E[X(X+1)] - E(X) - [E(X)]^2 = \frac{r(r+1)}{p^2} - \frac{r}{p} - \frac{r^2}{p^2} = \frac{rq}{p^2}.$$

Παρατήρηση

Αρνητική διωνυμική κατανομή. *Ας θεωρήσουμε τον αριθμό Y των αποτυχιών μέχρι τη r-οστή επιτυχία σε μια ακολουθία ανεξαρτήτων δοκιμών Bernoulli με πιθανότητα επιτυχίας p. Η συνάρτηση πιθανότητας της τυχαίας αυτής μεταβλητής δύναται να υπολογισθεί είτε απευθείας είτε με τη χρησιμοποίηση της σχέσης $Y = X - r$ και της συνάρτησης πιθανότητας (7) της X. Έχουμε*

$$f_Y(y) = P(Y = y) = P(X = r + y) = \binom{r+y-1}{y} p^r q^y, \quad y = 0, 1, \ldots. \quad (10)$$

Η κατανομή της τ.μ. Y καλείται αρνητική διωνυμική με παραμέτρους r και p. Η μέση τιμή και η διασπορά αυτής δύνανται να προκύψουν από τις (9) ως εξής:

$$E(Y) = E(X) - r = \frac{r}{p} - r = \frac{rq}{p}, \quad V(Y) = V(X) = \frac{rq}{p^2}. \quad (11)$$

Παράδειγμα

Μια γυναίκα εξακολουθεί να τεκνοποιεί μέχρι να αποκτήσει δύο αγόρια. Έστω ότι η πιθανότητα γέννησης αγοριού είναι $p = 0{,}49$.
Να υπολογισθούν
(α) η πιθανότητα όπως η γυναίκα αυτή αποκτήσει το πολύ 4 παιδιά μέχρι να πετύχει το σκοπό της και
(β) ο αναμενόμενος αριθμός παιδιών μέχρι τη γέννηση του δεύτερου αγοριού.

(α) Έστω X ο αριθμός των παιδιών μέχρι και τη γέννηση του δεύτερου αγοριού. Τότε η τ.μ. X έχει την κατανομή Pascal με παραμέτρους $r = 2$, $p = 0{,}49$ και έτσι

$$\begin{aligned} P(X \leq 4) &= \sum_{\kappa=2}^{4} (\kappa - 1)(0{,}49)^2 (0{,}51)^{\kappa-2} \\ &= (0{,}49)^2 \{1 + 2(0{,}51) + 3(0{,}51)^2\} = 0{,}67. \end{aligned}$$

(β) Ο αναμενόμενος αριθμός παιδιών μέχρι τη γέννηση του δεύτερου αγοριού, σύμφωνα με την πρώτη από τις (9), είναι

$$\mu = E(X) = \frac{2}{0{,}49} = 4{,}08.$$

Ας θεωρήσουμε έναν πεπερασμένο πληθυσμό του οποίου τα στοιχεία, σύμφωνα με κάποιο χαρακτηριστικό, κατατάσσονται σε δύο κατηγορίες. Έστω ότι ένα δείγμα συγκεκριμένου μεγέθους εκλέγεται από τον πληθυσμό αυτό, χωρίς επανάθεση.
Ο αριθμός των στοιχείων της μιας ή της άλλης κατηγορίας που περιλαμβάνονται στο δείγμα αποτελεί αντικείμενο πιθανοθεωρητικής μελέτης. Σχετικά θέτουμε τον ακόλουθο ορισμό.

Ορισμός

Έστω ότι από μια κάλπη που περιέχει r άσπρα και s μαύρα σφαιρίδια εξάγονται διαδοχικά το ένα μετά το άλλο, χωρίς επανάθεση, υ σφαιρίδια.
Στο τυχαίο (στοχαστικό) αυτό πείραμα έστω X ο αριθμός των άσπρων σφαιριδίων τα οποία εξάγονται. Η κατανομή της τ.μ. X καλείται υπεργεωμετρική με παραμέτρους r, s και υ.

Θεώρημα

Η συνάρτηση πιθανότητας της υπεργεωμετρικής κατανομής με παράμετρους r, s και υ δίνεται από την

$$f(x) = P(X = x) = \binom{r}{x}\binom{s}{\upsilon - x}\Big/\binom{r+s}{\upsilon}, \quad x = 0, 1, \ldots, \upsilon. \tag{12}$$

Απόδειξη.

Ο δειγματικός χώρος Ω περιλαμβάνει $N(\Omega) = \binom{r+s}{\upsilon}$ δειγματικά σημεία, όσα και ο αριθμός των υ-άδων σφαιριδίων που δύνανται να εξαχθούν. Τα δειγματικά αυτά σημεία είναι ισοπίθανα.
Το ενδεχόμενο $\{X = x\}$ περιλαμβάνει $\binom{r}{x}\binom{s}{\upsilon-x}$ υ-άδες σφαιριδίων με x άσπρα από τα r και $\upsilon - x$ μαύρα από τα s. □

Σημειώνουμε ότι

$$f(x) \geq 0, \quad x = 0, 1, \ldots, \upsilon, \quad f(x) = 0, \quad x \notin \{0, 1, \ldots, \upsilon\}$$

και σύμφωνα με τον τύπο του Cauchy,

$$\sum_{x=0}^{\upsilon} \binom{r}{x} \binom{s}{\upsilon - x} = \binom{r+s}{\upsilon}, \tag{13}$$

ισχύει

$$\sum_{x=0}^{\upsilon} f(x) = \sum_{x=0}^{\upsilon} \binom{r}{x} \binom{s}{\upsilon - x} \Big/ \binom{r+s}{\upsilon} = 1,$$

όπως απαιτείται από τον ορισμό της συνάρτησης πιθανότητας.

Θεώρημα

Έστω ότι η τυχαία μεταβλητή X ακολουθεί την υπεργεωμετρική κατανομή με συνάρτηση πιθανότητας την (9). Τότε η μέση τιμή και η διασπορά αυτής δίνονται από τις

$$\mu = E(X) = \frac{\upsilon r}{r+s}, \quad \sigma^2 = V(X) = \frac{\upsilon r}{r+s} \cdot \frac{s}{r+s} \cdot \frac{r+s-\upsilon}{r+s-1}. \tag{14}$$

Θεώρημα

Έστω ότι η τυχαία μεταβλητή X έχει την υπεργεωμετρική συνάρτηση πιθανότητας (12) με $N = r + s$.
Αν $N, r, s \longrightarrow \infty$ έτσι ώστε $\lim_{N\to\infty}(r/N) = p$, τότε

$$\lim_{N\to\infty} \binom{r}{x}\binom{s}{\upsilon - x}\bigg/\binom{N}{\upsilon} = \binom{\upsilon}{x}p^x(1-p)^{\upsilon-x}, \quad x = 0, 1, \ldots, \upsilon. \tag{15}$$