Βασικές διακριτές κατανομές

Στο προηγούμενο κεφάλαιο *(Παραδείγματα 5.3.5&5.3.15)* δείξαμε ότι αν X μια ***ομοιόμορφη διακριτή τυχαία μεταβλητή*** με σύνολο τιμών $R_X = \{x_1, x_2, \ldots, x_\nu\}$ και συνάρτηση πιθανότητας f με

$$f(x) = P(X = x) = \begin{cases} \dfrac{1}{\nu}, & x = x_1, x_2, \ldots, x_\nu \\ 0, & x \neq x_1, x_2, \ldots, x_\nu. \end{cases} \tag{6.1}$$

τότε

$$\mu = E(X) = \frac{x_1 + x_2 + \ldots + x_\nu}{\nu} \tag{6.2}$$

και

$$\sigma^2 = Var(X) = \frac{x_1^2 + x_2^2 + \ldots + x_\nu^2}{\nu} - \left(\frac{x_1 + x_2 + \ldots + x_\nu}{\nu} \right)^2. \tag{6.3}$$

Μια τέτοια κατανομή/μοντέλο πιθανοτήτων που εκχωρεί στις τιμές $x_1, x_2, \ldots, x_\nu$ της X ίσες πιθανότητες, ονομάζεται όπως είδαμε ***ομοιόμορφη διακριτή κατανομή***.

Έχοντας στη διάθεσή μας γενικά αποτελέσματα όπως τα *(6.1), (6.2)* και *(6.3),* κάθε φορά που θα χρειασθεί να μελετήσουμε μια *ομοιόμορφη διακριτή* τυχαία μεταβλητή, ανεξάρτητα με το τι αυτή κατά περίπτωση εκφράζει, θα μπορούμε ασφαλώς να τα αξιοποιούμε χωρίς να χρειάζεται να τα «ξαναανακαλύπτουμε». Έτσι, αν για παράδειγμα χρειασθεί να υπολογίσουμε τη *διακύμανση* της τυχαίας μεταβλητής που εκφράζει *την ένδειξη που εμφανίζεται κατά τη ρίψη ενός αμερόληπτου ζαριού*, έχουμε στη διάθεσή μας και μπορούμε να χρησιμοποιήσουμε τον τύπο *(6.3).* Αντίστοιχα, αν χρειασθεί να υπολογίσουμε το *μέσο αριθμό δοκιμών που θα απαιτηθούν μέχρι να βρεθεί το ελαττωματικό προϊόν που υπάρχει μεταξύ 50 προϊόντων (Παράδειγμα 5.3.3)*, μπορούμε να χρησιμοποιήσουμε τον τύπο *(6.2).*

Στο κεφάλαιο αυτό θα γνωρίσουμε τέσσερα ακόμη μοντέλα πιθανοτήτων που έχουν μελετηθεί συστηματικά και τα σχετικά με αυτά γενικά αποτελέσματα χρησιμοποιούνται και καλύπτουν ευρύτατο φάσμα εφαρμογών. Πρόκειται για την ***κατανομή Bernoulli***, τη ***διωνυμική κατανομή***, την ***κατανομή Poisson*** και τη ***διαδικασία Poisson.*** Θα κάνουμε επίσης μια σύντομη αναφορά στην ***πολυωνυμική κατανομή***. Άλλα διακριτά μοντέλα πιθανοτήτων που επίσης καλύπτουν ευρύ φάσμα εφαρμογών, στα οποία όμως δε θα αναφερθούμε, είναι η ***γεωμετρική κατανομή***, η ***αρνητική διωνυμική κατανομή*** και η ***υπεργεωμετρική κατανομή***.

6.1 Κατανομή Bernoulli και Διωνυμική κατανομή

Στα προηγούμενα, στο *Παράδειγμα 3.1.1α&δ,* γνωρίσαμε ένα πολύ απλό πείραμα τύχης, ίσως το απλούστερο, τη **δοκιμή Bernoulli (Bernoulli trial)**[1]. Πρόκειται για ένα πείραμα τύχης με μόνο δύο, αμοιβαίως αποκλειόμενα, δυνατά αποτελέσματα. Το ένα αποτέλεσμα έχει επικρατήσει να ονομάζεται ***επιτυχία (success)*** και το άλλο ***αποτυχία (failure)***.

Το πιο «δημοφιλές» στη βιβλιογραφία παράδειγμα *δοκιμής Bernoulli* είναι η ρίψη ενός νομίσματος μία φορά. Τα δυνατά αποτελέσματα είναι προφανώς μόνο δύο, «κεφαλή» και «γράμματα». Αν μας ενδιαφέρει η ένδειξη «κεφαλή» χαρακτηρίζουμε *επιτυχία* το αποτέλεσμα «κεφαλή» και *αποτυχία* το αποτέλεσμα «γράμματα» ενώ αν μας ενδιαφέρει η ένδειξη «γράμματα» χαρακτηρίζουμε *επιτυχία* το αποτέλεσμα «γράμματα» και *αποτυχία* το αποτέλεσμα «κεφαλή». Ας δούμε μερικά ακόμη παραδείγματα.

[1] *Φέρει το όνομα του Ελβετού μαθηματικού Jacob (James/Jacques) Bernoulli (1654-1705), της οικογένειας διακεκριμένων μαθηματικών.*

α) Η επιλογή ενός ζώου και η εξέτασή του για να διαπιστωθεί αν έχει προσβληθεί από μια συγκεκριμένη ασθένεια είναι *δοκιμή Bernoulli* γιατί τα δυνατά αποτελέσματα είναι μόνο δύο: το ζώο είτε έχει προσβληθεί *(επιτυχία)* είτε δεν έχει προσβληθεί *(αποτυχία)*.
β) Η λήψη δείγματος ελαιολάδου και ο έλεγχος για το αν είναι νοθευμένο είναι *δοκιμή Bernoulli* γιατί το ελαιόλαδο είτε είναι νοθευμένο *(επιτυχία)* είτε δεν είναι νοθευμένο *(αποτυχία)*.
γ) Η επιλογή και ο έλεγχος ενός φυτού για το αν η ξηρή μάζα του ξεπερνάει τα 150gr είναι *δοκιμή Bernoulli* γιατί το φυτό έχει ξηρή μάζα είτε μεγαλύτερη από 150gr *(επιτυχία)* είτε το πολύ 150gr *(αποτυχία)*.
δ) Φυτεύουμε ένα σπόρο πιπεριάς και ελέγχουμε αν βλάστησε. Πρόκειται για *δοκιμή Bernoulli* γιατί ο σπόρος είτε θα βλαστήσει *(επιτυχία)* είτε δε θα βλαστήσει *(αποτυχία)*.
ε) Η επιλογή και ο έλεγχος ενός φυτού για το αν έχει λιγότερα από 6 φύλλα είναι *δοκιμή Bernoulli* γιατί το φυτό έχει είτε λιγότερα από 6 *(επιτυχία)* είτε τουλάχιστον 6 φύλλα *(αποτυχία)*.
στ) Η επιλογή ενός μαθητή Α΄ Δημοτικού και ο έλεγχος της συγκέντρωσης της προσοχής του κατά τα πρώτα 20 λεπτά της πρώτης ώρας του προγράμματος μπορεί (με συγκεκριμένα κριτήρια) να χαρακτηρισθεί είτε ικανοποιητική είτε όχι ικανοποιητική. Είναι επομένως μια *δοκιμή Bernoulli*.
ζ) Ένα άτομο που επιλέγεται για να συμμετάσχει σε μια έρευνα γνώμης είναι είτε άνδρας είτε γυναίκα. Η επιλογή επομένως ενός ατόμου και η καταγραφή του φύλου του είναι μια *δοκιμή Bernoulli*. *Δοκιμή Bernoulli* είναι επίσης, η επιλογή ενός ατόμου και η υποβολή του σε ένα διαγνωστικό τεστ που έχει δύο δυνατά αποτελέσματα (είτε θετικό είτε αρνητικό) ή η επιλογή και ο έλεγχος ενός εξαρτήματος για να διαπιστωθεί αν είναι εντός ή εκτός προδιαγραφών.

Σημείωση 6.1.1: *Παρατηρείστε ότι οι δυνατές τιμές της ξηρής μάζας ενός φυτού στο παράδειγμα (γ) προφανώς δεν είναι μόνο δύο. Το ίδιο ισχύει και για τον αριθμό των φύλλων ενός φυτού στο παράδειγμα (ε). Όμως, με βάση το τι ενδιαφέρει στην αντίστοιχη έρευνα, ταξινομήσαμε τις δυνατές τιμές σε δύο κατηγορίες-αποτελέσματα και οδηγηθήκαμε έτσι σε δοκιμές Bernoulli. Δοκιμή Bernoulli έχουμε επίσης σε περιπτώσεις όπου, ενώ τα δυνατά εξαγόμενα είναι περισσότερα από δύο, μας ενδιαφέρει εάν συμβαίνει ή δε συμβαίνει, μόνο ένα συγκεκριμένο. Για παράδειγμα, ένα παιδί που επιλέγεται τυχαία από τα παιδιά μιας οικογένειας μπορεί να έχει το γονότυπο ΑΑ ή όχι*[2]*. Επίσης, από μία τράπουλα επιλέγουμε τυχαία ένα παιγνιόχαρτο το οποίο μπορεί να είναι ή να μην είναι «άσσος σπαθί».*

■

Σε μια *δοκιμή Bernoulli* ο αριθμός των επιτυχιών προφανώς είναι ή μια ή καμία (ή 1 ή 0). Αν η πιθανότητα *επιτυχίας* είναι p και η πιθανότητα αποτυχίας είναι q, προφανώς $p+q=1$ ή $q=1-p$.

Έστω X η διακριτή τυχαία μεταβλητή που εκφράζει τον αριθμό των επιτυχιών σε μια *δοκιμή Bernoulli* με πιθανότητα επιτυχίας p. Η κατανομή της X ονομάζεται ***κατανομή Bernoulli (Bernoulli distribution)*** με παράμετρο p και συμβολίζεται με $b(p)$. Επίσης γράφουμε $X \sim b(p)$.

Προφανώς η $X \sim b(p)$ έχει συνάρτηση πιθανότητας

$$f(x)=P(X=x)=\begin{cases} p, & x=1 \\ 1-p, & x=0 \end{cases}$$

ή

[2] *Ο γονότυπος μπορεί να είναι ΑΑ, Αα ή αα.*

$$f(x) = P(X = x) = p^x (1-p)^{1-x}, \ x = 0, 1.$$

Η *μέση τιμή* μ της X είναι

$$\mu = E(X) = 0 \cdot f(0) + 1 \cdot f(1) = p$$

και η *διακύμανσή* της

$$\sigma^2 = E(X^2) - [E(X)]^2 = 0^2 \cdot f(0) + 1^2 \cdot f(1) - p^2 = p - p^2 = p(1-p) = pq.$$

Παρότι η *δοκιμή Bernoulli* είναι ένα πολύ απλό πείραμα, εντούτοις (ή μήπως γι' αυτό;) βρίσκεται στον πυρήνα πολλών προβλημάτων με ευρύτατο φάσμα εφαρμογών που παρουσιάζουν μάλιστα ιδιαίτερο ενδιαφέρον και σε αρκετές περιπτώσεις ιδιαίτερες δυσκολίες στην αντιμετώπισή τους. Πιο συγκεκριμένα, πολλά πραγματικά προβλήματα αναλύονται σε μια σειρά–ακολουθία *δοκιμών Bernoulli* και τα τελικά ερωτήματα που απορρέουν από αυτά σχετίζονται με τον αριθμό των επιτυχιών που συμβαίνουν. Για παράδειγμα, το τελικό ερώτημα μετά την ανάλυση ενός τέτοιου προβλήματος μπορεί να αφορά *τον αριθμό των επιτυχιών που συμβαίνουν σε ν επαναλήψεις μιας δοκιμής Bernoulli ή τον απαιτούμενο αριθμό επαναλήψεων μιας δοκιμής Bernoulli μέχρι να συμβεί η πρώτη επιτυχία ή τον απαιτούμενο αριθμό επαναλήψεων μιας δοκιμής Bernoulli μέχρι να συμβούν r επιτυχίες ή το μεγαλύτερο μήκος ροής συνεχόμενων επιτυχιών σε ν επαναλήψεις μιας δοκιμής Bernoulli.*

Στη συνέχεια θα ασχοληθούμε με προβλήματα που αναλύονται σε μια ακολουθία *ν* ανεξάρτητων επαναλήψεων μιας *δοκιμής Bernoulli* και που τα πιθανοθεωρητικά συμπεράσματα που μας ενδιαφέρουν συνδέονται με τον αριθμό *επιτυχιών που συμβαίνουν σε ν επαναλήψεις μιας δοκιμής Bernoulli.* Ας δούμε ένα τέτοιο παράδειγμα.

Παράδειγμα 6.1.1: *Ο γεωπόνος ενός φυτώριου ισχυρίζεται ότι οι βολβοί τουλίπας που παράγονται στο φυτώριο βλαστάνουν σε ποσοστό 90%. Ένας αγρότης που είχε προμηθευθεί από το συγκεκριμένο φυτώριο μεγάλο αριθμό βολβών τουλίπας θέλησε να ελέγξει τον ισχυρισμό του γεωπόνου. Συγκεκριμένα, επέλεξε τυχαία 52 από τους βολβούς που φύτεψε και διαπίστωσε ότι βλάστησαν μόνο οι 38, γεγονός που του δημιούργησε αμφιβολίες για το αν πράγματι ευσταθεί ο ισχυρισμός του γεωπόνου. Είναι άραγε, δικαιολογημένες οι αμφιβολίες του αγρότη;*

<u>*Απάντηση:*</u> Είναι προφανές ότι η τυχαία επιλογή και ο έλεγχος ενός βολβού για το αν βλάστησε ή όχι, είναι μια *δοκιμή Bernoulli*. Η επιλογή επομένως και ο έλεγχος 52 βολβών είναι μια σειρά-ακολουθία 52 *δοκιμών Bernoulli*. Επειδή ο αγρότης ενδιαφέρεται για τους βολβούς που βλάστησαν, είναι λογικό να ονομάσουμε *επιτυχία* το αποτέλεσμα *«ο βολβός βλάστησε»* και *αποτυχία* το αποτέλεσμα *«ο βολβός δε βλάστησε».*

Το ερώτημα που λογικά προκύπτει είναι το εξής: *αν αποδεχθούμε τον ισχυρισμό του γεωπόνου, πόσο λογικό-πιθανό είναι το αποτέλεσμα του ελέγχου που έκανε ο αγρότης;* Δηλαδή, *με δεδομένο ότι το ποσοστό των βολβών που βλαστάνουν είναι 90%, πόσο πιθανό είναι από τους 52 τυχαία επιλεγμένους βολβούς να βρεθούν να έχουν βλαστήσει οι 38;* Ή αλλιώς, *ποια είναι η πιθανότητα, σε 52 επαναλήψεις μιας δοκιμής Bernoulli να συμβούν 38 επιτυχίες, με δεδομένο ότι σε κάθε επανάληψη η πιθανότητα επιτυχίας p είναι ίση με 0.9*.

Πρέπει να σημειώσουμε ότι σωστά θεωρήσαμε ότι σε κάθε επανάληψη της *δοκιμής Bernoulli* η πιθανότητα επιτυχίας παραμένει σταθερή και ίση με 0.9 γιατί ο αριθμός των βολβών που εξετάζονται είναι πολύ μικρός σε σχέση με τον αριθμό των βολβών που φυτεύτηκαν και επομένως, παρότι η δειγματοληψία γίνεται χωρίς επανάθεση, η πιθανότητα αυτή πρακτικά δεν αλλάζει. Είναι επίσης λογικό να θεωρήσουμε ότι δεν επηρεάζεται/εξαρτάται (καθοιονδήποτε τρόπο) το αποτέλεσμα οποιασδήποτε δοκιμής από το αποτέλεσμα προηγούμενων δοκιμών.

Το γενικότερο ερώτημα που τίθεται είναι το εξής:
Αν Χ ο αριθμός των επιτυχιών σε μια ακολουθία ν ανεξάρτητων δοκιμών Bernoulli με σταθερή πιθανότητα επιτυχίας p σε όλες τις δοκιμές, πώς μπορούμε να υπολογίσουμε τις πιθανότητες

$$P(X = x), \quad x = 0,1,2,...,\nu .$$

Στα προηγούμενα (ξαναδείτε την *Πρόταση 4.5.2δ*), δείξαμε ότι οι πιθανότητες αυτές δίνονται από τον τύπο

$$P(X = x) = \binom{\nu}{x} p^x (1-p)^{\nu-x}, \quad x = 0,1,2,...,\nu .$$

Έτσι, αν συμβολίσουμε με X τον αριθμό των βολβών που βλαστάνουν (από τους 52), τότε

$$P(X = 38) = \binom{52}{38} 0.9^{38} (1-0.9)^{14} \cong 0.0003 .$$

Δηλαδή, με την υπόθεση ότι αληθεύει ο ισχυρισμός του γεωπόνου, η πιθανότητα να βλαστήσουν μόνο 38 από τους 52 βολβούς είναι πολύ μικρή (περίπου μηδενική). Οι αμφιβολίες επομένως του αγρότη φαίνεται ότι έχουν βάση[3]. *(Ασφαλώς ο αγρότης δεν υπολόγισε την πιθανότητα $P(X = 38)$, απλώς υπολόγισε το 90% του 52 που είναι 46.8 και έτσι περίμενε από τους 52 βολβούς να έχουν βλαστήσει περίπου οι 47).*

■

Ας δώσουμε τώρα τον ορισμό της ***διωνυμικής κατανομής***[4]:

Ορισμός 6.1.1 (Διωνυμική κατανομή): *Αν Χ ο αριθμός των επιτυχιών σε μια ακολουθία ν ανεξάρτητων δοκιμών Bernoulli με σταθερή πιθανότητα επιτυχίας p σε όλες τις δοκιμές, τότε η κατανομή της τυχαίας μεταβλητής Χ ονομάζεται* ***διωνυμική κατανομή (binomial distribution)*** *με παραμέτρους ν και p και συμβολίζεται με $B(\nu, p)$. Επίσης, γράφουμε*

$$X \sim B(\nu, p) .$$

■

Όπως ήδη αναφέραμε, η συνάρτηση πιθανότητας της *διωνυμικής κατανομής* με παραμέτρους ν και p δίνεται από τον τύπο

$$f(x) = P(X = x) = \binom{\nu}{x} p^x (1-p)^{\nu-x}, \quad x = 0,1,2,...,\nu .$$

Παρατηρείστε, ότι η f έχει πράγματι τις ιδιότητες μιας συνάρτησης πιθανότητας.

Προφανώς είναι μη αρνητική συνάρτηση και επίσης

$$\sum_{x=0}^{\nu} f(x) = \sum_{x=0}^{\nu} \binom{\nu}{x} p^x (1-p)^{\nu-x} = [p + (1-p)]^{\nu} = 1^{\nu} = 1$$

αφού οι πιθανότητες των τιμών *0, 1, 2,...,ν*, της X είναι οι όροι του *διωνυμικού αναπτύγματος* $[p + (1-p)]^{\nu}$ (θυμηθείτε τον *τύπο του Νεύτωνα*[5]). Έτσι εξηγείται και το όνομα της *διωνυμικής κατανομής*.

---

[3] *Περισσότερα επ' αυτού στη στατιστική συμπερασματολογία.*

[4] *Τη διωνυμική κατανομή πρώτοι μελέτησαν ο De Moivre (1667-1754) και ο Jacob Bernoulli (1654-1705).*

[5] *Για οποιουσδήποτε πραγματικούς αριθμούς α, β ισχύει ότι* $(\alpha+\beta)^{\nu} = \sum_{x=0}^{\nu} \binom{\nu}{x} \alpha^x \beta^{\nu-x}, \ \nu = 1,2,...$

Πρόταση 6.1.1: *Η μέση τιμή και η διακύμανση της διωνυμικής κατανομής με παραμέτρους ν και p, είναι αντίστοιχα*

$$\mu = \nu p \text{ και } \sigma^2 = \nu p(1-p) = \nu pq .$$

<u>*Απόδειξη:*</u> Η *διωνυμική κατανομή* ορίσθηκε ως άθροισμα ν ανεξάρτητων τυχαίων μεταβλητών που η κάθε μια ακολουθεί μια κατανομή *Bernoulli* με την ίδια παράμετρο. Πράγματι, αν X_i, $i = 1,2,...,\nu$, ν ανεξάρτητες τυχαίες μεταβλητές με

$$X_i \sim b(p), \ \ i = 1,2,...,\nu$$

τότε

$$X = X_1 + X_2 + \ldots + X_\nu \sim B(\nu, p) .$$

Επειδή όπως δείξαμε, $E(X_i) = p$ ($i = 1,2,...,\nu$), από την *Πρόταση 5.3.5* έχουμε

$$\mu = E(X) = E(X_1 + X_2 + \ldots + X_\nu) = E(X_1) + E(X_2) + \ldots + E(X_\nu) =$$

$$= \underbrace{p + p + \ldots + p}_{\nu-όροι} = \nu p .$$

Επίσης, όπως δείξαμε $Var(X_i) = pq$ ($i = 1,2,...,\nu$) και επειδή οι τ.μ. $X_1, X_2, \ldots, X_\nu$ είναι **ανεξάρτητες**, από την *Πρόταση 5.5.1* έχουμε

$$\sigma^2 = Var(X) = Var(X_1 + X_2 + \ldots + X_\nu) = Var(X_1) + Var(X_2) + \ldots + Var(X_\nu) =$$

$$= \underbrace{pq + pq + \ldots + pq}_{\nu-όροι} = \nu pq .$$

Ας δούμε και μια διαφορετική απόδειξη (ως ένα παράδειγμα για το πώς μπορούμε να χρησιμοποιήσουμε την παραγοντική ροπή 2ης τάξης, $E[X(X-1)]$, για να υπολογίσουμε την $E(X^2)$).

Από τον τύπο ορισμού της *μέσης τιμής* έχουμε

$$\mu = E(X) = \sum_{x=0}^{\nu} x f(x) = \sum_{x=0}^{\nu} x \binom{\nu}{x} p^x (1-p)^{\nu-x} = \sum_{x=1}^{\nu} x \binom{\nu}{x} p^x (1-p)^{\nu-x}$$

και επειδή

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

παίρνουμε

$$\mu = E(X) = \sum_{x=1}^{\nu} \nu \binom{\nu-1}{x-1} p^x (1-p)^{\nu-x} = \nu \sum_{x=1}^{\nu} \binom{\nu-1}{x-1} p p^{x-1} (1-p)^{\nu-x+1-1} =$$

$$= \nu p \sum_{x=1}^{\nu} \binom{\nu-1}{x-1} p^{x-1} (1-p)^{\nu-1-(x-1)} = \nu p \sum_{y=0}^{\nu-1} \binom{\nu-1}{y} p^y (1-p)^{\nu-1-y} = \nu p[p + (1-p)]^{\nu-1} = \nu p .$$

Για να υπολογίσουμε τη διακύμανση $\sigma^2 = Var(X)$ θα χρησιμοποιήσουμε τον τύπο

$$\sigma^2 = E(X^2) - \mu^2$$

και τη ροπή 2ης τάξης $E(X^2)$ θα την υπολογίσουμε μέσω της παραγοντικής ροπής 2ης τάξης από τον τύπο

$$E(X^2) = E[X(X-1)] + E(X) .$$

Για την παραγοντική ροπή 2ης τάξης $E[X(X-1)]$ έχουμε

$$E[X(X-1)] = \sum_{x=0}^{\nu} x(x-1) f(x) = \sum_{x=0}^{\nu} x(x-1) \binom{\nu}{x} p^x (1-p)^{\nu-x} = \sum_{x=2}^{\nu} x(x-1) \binom{\nu}{x} p^x (1-p)^{\nu-x}$$

και επειδή

$$x(x-1)\binom{\nu}{x} = x(x-1)\cdot\frac{\nu!}{x!(\nu-x)!} = \nu(\nu-1)\cdot\frac{(\nu-2)!}{(x-2)!(\nu-x)!} = \nu(\nu-1)\binom{\nu-2}{x-2}$$

παίρνουμε

$$E[X(X-1)] = \sum_{x=2}^{\nu}\nu(\nu-1)\binom{\nu-2}{x-2}p^x(1-p)^{\nu-x} =$$

$$= \nu(\nu-1)\sum_{x=2}^{\nu}\binom{\nu-2}{x-2}p^2 p^{x-2}(1-p)^{\nu-x+2-2} = \nu(\nu-1)p^2\sum_{x=2}^{\nu}\binom{\nu-2}{x-2}p^{x-2}(1-p)^{\nu-2-(x-2)} =$$

$$\nu(\nu-1)p^2\sum_{y=2}^{\nu-2}\binom{\nu-2}{y}p^y(1-p)^{\nu-2-y} = \nu(\nu-1)p^2$$

άρα

$$E(X^2) = E[X(X-1)] + E(X) = \nu(\nu-1)p^2 + \nu p\,.$$

Η διακύμανση επομένως είναι

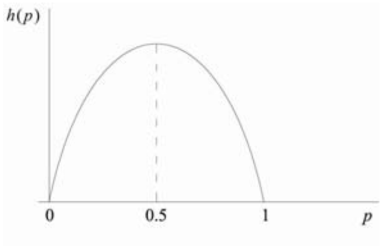
$$\sigma^2 = E(X^2) - \mu^2 = \nu(\nu-1)p^2 + \nu p - (\nu p)^2 = \nu^2 p^2 - \nu p^2 + \nu p - \nu^2 p^2 = \nu p(1-p)\,.$$

■

Έτσι, στο *Παράδειγμα 6.1.1* η *μέση τιμή*, η *διακύμανση* και η *τυπική απόκλιση* της τυχαίας μεταβλητής $X \sim B(52,\, 0.9)$ είναι αντίστοιχα

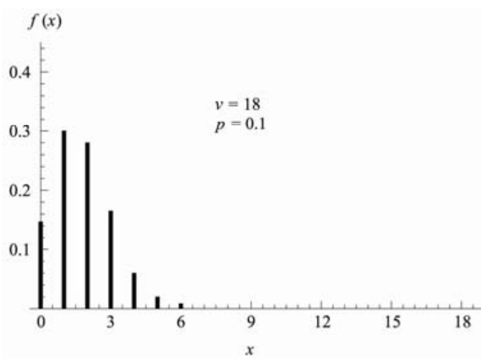
$$\mu = 52\cdot 0.9 = 46.8\,,\ \sigma^2 = \nu p(1-p) = 52\cdot 0.9\cdot 0.1 = 4.68 \text{ και } \sigma = \sqrt{4.68} = 2.16\,.$$

Σχόλιο 6.1.1 (για τη διακύμανση της διωνυμικής κατανομής): *Αν τη διακύμανση* $\sigma^2 = \nu p(1-p)$, *της διωνυμικής τυχαίας μεταβλητής* $X \sim B(\nu, p)$ *τη δούμε ως συνάρτηση, έστω h, του p με* $h(p) = \nu p(1-p) = -\nu p^2 + \nu p$, *παρατηρούμε ότι στο διάστημα* $[0,1/2]$, *είναι αύξουσα συνάρτηση του p ενώ στο διάστημα* $[1/2,1]$ *είναι φθίνουσα συνάρτηση του p και παρουσιάζει μέγιστο για* $p = 1/2$ *(Σχήμα 6.1.1).*

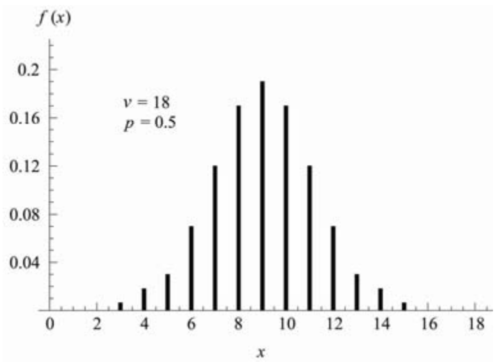


Σχήμα 6.1.1
Η διακύμανση της διωνυμικής κατανομής
ως συνάρτηση της παραμέτρου p

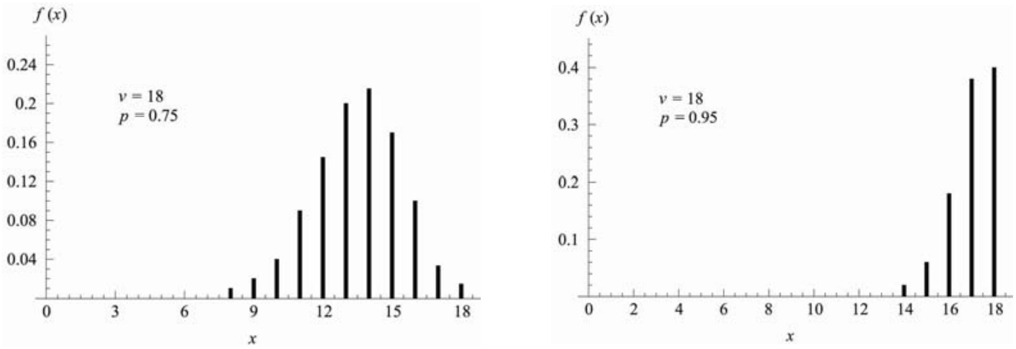
Δηλαδή, η διακύμανση ελαττώνεται όσο το p πλησιάζει το 0 ή το 1 και μεγιστοποιείται για $p = 1/2$. *Αυτό, είναι άραγε λογικό; Πριν απαντήσετε δείτε τα Σχήματα 6.1.2 όπου φαίνεται η συνάρτηση πιθανότητας της διωνυμικής κατανομής για* $\nu = 18$ *και* $p = 0.1,\ 0.5,\ 0.75, 0.95$.



(α)



(β)



(γ) *(δ)*

Σχήμα 6.1.2
Η συνάρτηση πιθανότητας της διωνυμικής κατανομής
για $\nu = 18$ *και* $p = 0.1,\ 0.5,\ 0.75, 0.95$

■

Ας δούμε ένα ακόμη παράδειγμα υπολογισμού πιθανοτήτων μιας διωνυμικής τυχαίας μεταβλητής.

Παράδειγμα 6.1.2: *Μια βιομηχανία κατασκευάζει μεταλλικά ελάσματα για να αντέχουν σε συγκεκριμένη καταπόνηση. Σύμφωνα με τις προδιαγραφές παραγωγής, κάθε τέτοιο έλασμα αντέχει στη συγκεκριμένη καταπόνηση με πιθανότητα 0.8. Επιλέγουμε τυχαία 9 τέτοια ελάσματα και τα υποβάλλουμε στη συγκεκριμένη καταπόνηση. Ποια είναι η πιθανότητα να αντέξουν α) το πολύ 2 ελάσματα β) περισσότερα από 7 ελάσματα γ) τουλάχιστον 2 ελάσματα και δ) λιγότερα από 6 και τουλάχιστον 4 ελάσματα.*

Απάντηση: Έστω X ο αριθμός των ελασμάτων που θα αντέξουν την καταπόνηση (από τα 9 που θα ελεγχθούν). Προφανώς[6], $X \sim B(9,\ 0.8)$ και επομένως για τις ζητούμενες πιθανότητες έχουμε

α) $P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) =$

$$= \binom{9}{0} \cdot 0.8^0 0.2^9 + \binom{9}{1} \cdot 0.8^1 0.2^8 + \binom{9}{2} \cdot 0.8^2 0.2^7 = 0.0003\,.$$

β) $P(X > 7) = P(X = 8) + P(X = 9) = \binom{9}{8} \cdot 0.8^8 0.2^1 + \binom{9}{9} \cdot 0.8^9 0.2^0 = 0.436207\,.$

γ) $P(X \ge 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = \ldots = 1 - 0.000019 = 0.999981\,.$

δ) $P(4 \le X < 6) = P(X = 4) + P(X = 5) = \binom{9}{4} \cdot 0.8^4 0.2^5 + \binom{9}{5} \cdot 0.8^5 0.2^4 = 0.0825754\,.$

■

Σημείωση 6.1.2: *Είναι προφανές, ότι όταν το* ν *είναι μεγάλο και το* x *όχι πολύ κοντά στο 0 ή το* ν*, οι διωνυμικοί συντελεστές* $\binom{\nu}{x}$ *που εμφανίζονται στον τύπο*

$$P(X = x) = \binom{\nu}{x}\ p^x (1-p)^{\nu - x}, \quad x = 0,1,2,\ldots,\nu$$

παίρνουν μεγάλες τιμές με συνέπεια να γίνεται προβληματικός ο υπολογισμός των πιθανοτήτων. Ένας τρόπος αντιμετώπισης του προβλήματος είναι ο υπολογισμός των πιθανοτήτων μέσω αναδρομικού τύπου. Πράγματι, εύκολα αποδεικνύεται[7] *ο ακόλουθος αναδρομικός τύπος.*

[6] *Τι υποθέσεις πρέπει να κάνουμε;*
[7] *Η απόδειξη του αναδρομικού τύπου είναι πολύ απλή:*

Για τις πιθανότητες, $P(X=x),\ x=1,\ldots,\nu$ της τυχαίας μεταβλητής $X \sim B(\nu,p)$ ισχύει

$$P(X=x)=\frac{\nu-x+1}{x}\cdot\frac{p}{1-p}P(X=x-1)$$

με αρχική συνθήκη $P(X=0)=(1-p)^{\nu}$.

Έτσι, αν υπολογίσουμε την πιθανότητα $P(X=0)=(1-p)^{\nu}$, μπορούμε μέσω αυτού του αναδρομικού τύπου να υπολογίσουμε όλες τις πιθανότητες $P(X=1)$, $P(X=2),\ldots,P(X=\nu)$ χωρίς να χρειασθεί να υπολογίσουμε τους αντίστοιχους διωνυμικούς συντελεστές. Ας δούμε ένα παράδειγμα.

Έστω $X \sim B(5,\, 0.3)$. Η πιθανότητα $P(X=3)$ μέσω του αναδρομικού τύπου υπολογίζεται ως εξής:

$$P(X=0)=0.7^5=0.168$$

$$P(X=1)=\frac{5-1+1}{1}\frac{0.3}{1-0.3}0.168=0.360$$

$$P(X=2)=\frac{5-2+1}{2}\frac{0.3}{1-0.3}0.360=0.309$$

$$P(X=3)=\frac{5-3+1}{3}\frac{0.3}{1-0.3}0.309=0.132.$$

Είναι βέβαια φανερό ότι αυτή η μέθοδος υπολογισμού των πιθανοτήτων έχει το μειονέκτημα ότι για να υπολογίσουμε μια συγκεκριμένη πιθανότητα $P(X=x)$ που μας ενδιαφέρει, πρέπει να προηγηθεί ο υπολογισμός των πιθανοτήτων

$$P(X=0),\ P(X=1),\ldots,P(X=x-1).$$

Έτσι, όσο αυξάνει το x, αυξάνει (...δυστυχώς) και ο αριθμός των πιθανοτήτων που πρέπει να υπολογισθούν.

Στη συνέχεια, όταν θα μιλήσουμε για την κατανομή Poisson και την κανονική κατανομή, θα δούμε και άλλο τρόπο αντιμετώπισης των δυσκολιών υπολογισμού (στις περιπτώσεις που υπάρχουν) των πιθανοτήτων της διωνυμικής τυχαίας μεταβλητής[8].

Σημείωση 6.1.3 (η πιο πιθανή τιμή μιας διωνυμικής τ.μ.): *Αν $X \sim B(\nu,p)$, άραγε από τις τιμές 0, 1, 2,....,ν που μπορεί να πάρει η X, ποια είναι η πιο πιθανή. Παρατηρείστε στα Σχήματα 6.1.2 ότι όσο οι τιμές της X αυξάνουν από το 0 στο ν, οι αντίστοιχες πιθανότητες $P(X=x),\ x=0,1,\ldots,\nu$, μέχρι ένα σημείο (ας το συμβολίσουμε με x_0) αυξάνουν και στη συνέχεια φθίνουν. Πράγματι έτσι είναι. Ας το αποδείξουμε και ας προσδιορίσουμε αυτή την τιμή x_0 της X.*

Από τον αναδρομικό τύπο που αποδείξαμε προηγουμένως προκύπτει ότι

$$\frac{P(X=x)}{P(X=x-1)}=\frac{\nu-x+1}{x}\cdot\frac{p}{1-p}.$$

Δηλαδή

$$P(X=x)>P(X=x-1)$$

αν και μόνο αν

$$\frac{P(X=x)}{P(X=x-1)}=\frac{\dfrac{\nu!}{(\nu-x)!x!}p^x(1-p)^{\nu-x}}{\dfrac{\nu!}{(\nu-x+1)!(x-1)!}p^{x-1}(1-p)^{\nu-x+1}}=\frac{\nu-x+1}{x}\cdot\frac{p}{1-p}.$$

[8] *Οι δυσκολίες υπολογισμών, δεν λύθηκαν με τη χρήση Η/Υ. Και στους υπολογιστές τίθενται σχετικά ζητήματα όπως, απαιτούμενου υπολογιστικού χρόνου, μνήμης κτλ.*

$$(\nu - x + 1)p > x(1-p) \text{ ή } x < (\nu+1)p.$$

Αυτό σημαίνει ότι:

α) *Όταν το γινόμενο* $(\nu+1)p$ *δεν είναι ακέραιος αριθμός, η πιο πιθανή τιμή* x_0 *της X είναι το ακέραιο μέρος του, δηλαδή*

$$x_0 = [(\nu+1)p]$$

διότι οι πιθανότητες $P(X = x)$ *αυξάνουν γνησίως για* $x \le [(\nu+1)p]$ *και φθίνουν γνησίως για* $x \ge [(\nu+1)p]+1$.

β) *Όταν το γινόμενο* $(\nu+1)p$ *είναι ακέραιος αριθμός, οι πιθανότητες* $P(X = x)$ *αυξάνουν γνησίως για* $x \le (\nu+1)p-1$ *και φθίνουν γνησίως για* $x \ge (\nu+1)p+1$. *Και επειδή* $P[X = (\nu+1)p] = P[X = (\nu+1)p-1]$, *στην περίπτωση αυτή, οι τιμές της X με τη μεγαλύτερη πιθανότητα είναι δύο*

$$x_0 = (\nu+1)p \text{ και } x_0' = (\nu+1)p-1.$$

Συνοψίζοντας, όταν το γινόμενο $(\nu+1)p$ *δεν είναι ακέραιος αριθμός, η πιο πιθανή τιμή* x_0 *της* $X \sim B(\nu, p)$ *είναι το ακέραιο μέρος του, δηλαδή,* $x_0 = [(\nu+1)p]$ *ενώ όταν το γινόμενο* $(\nu+1)p$ *είναι ακέραιος αριθμός, οι πιο πιθανές τιμές είναι δύο, η* $x_0 = (\nu+1)p$ *και η* $x_0' = (\nu+1)p-1$.

Έτσι, στο Παράδειγμα 6.1.1, επειδή το γινόμενο $(\nu+1)p = 53 \cdot 0.9 = 47.7$ *δεν είναι ακέραιος, η πιο πιθανή τιμή του αριθμού X των βολβών (από τους 52) που βλαστάνουν είναι η* $x_0 = [47.7] = 47$. *Δηλαδή ο αγρότης σωστά ανέμενε αυτή την τιμή ως την πιο πιθανή!*

Στο Παράδειγμα 6.1.2, επειδή το γινόμενο $(\nu+1)p = 10 \cdot 0.8 = 8$ *είναι ακέραιος, οι τιμές με τη μεγαλύτερη πιθανότητα είναι δύο: η* $x_0 = (\nu+1)p = 8$ *και η* $x_0' = (\nu+1)p - 1 = 7$.

Δείτε επίσης την τιμή με τη μεγαλύτερη πιθανότητα σε κάθε μια από τις τέσσερις διωνυμικές κατανομές που φαίνονται στα Σχήματα 6.1.2. Επιβεβαιώστε/επαληθεύστε ότι πράγματι τα προηγούμενα συμπεράσματα ισχύουν.

■

Δείτε την Άσκηση 6.1.1 και θυμηθείτε τη σχέση *Πιθανοτήτων* και *Στατιστικής* που συζητήσαμε στο *1ο Κεφάλαιο*. Ξανασκεφθείτε επίσης, το *Παράδειγμα 6.1.1*.

Άσκηση 6.1.1: *Ρίχνουμε ένα νόμισμα 10 φορές. Αν η ένδειξη «γράμματα» έρθει από 3 έως το πολύ 7 φορές, δεχόμαστε ότι το νόμισμα είναι αμερόληπτο, ενώ σε κάθε άλλη περίπτωση θεωρούμε το νόμισμα μεροληπτικό. α) Να υπολογίσετε την πιθανότητα να θεωρήσουμε ότι το νόμισμα είναι μεροληπτικό, ενώ στην πραγματικότητα είναι αμερόληπτο* ***(πιθανότητα σφάλματος τύπου Ι)****. β) Να υπολογισθεί η πιθανότητα να θεωρήσουμε ότι το νόμισμα είναι αμερόληπτο όταν είναι μεροληπτικό με πιθανότητα εμφάνισης της ένδειξης «γράμματα» ίση με 0.7* ***(πιθανότητα σφάλματος τύπου ΙΙ)****.*

Απάντηση: α) $1 - \sum_{x=3}^{7} \binom{10}{x} \left(\frac{1}{2}\right)^{10} \cong 0.11$. β) $\sum_{x=3}^{7} \binom{10}{x} \cdot 0.7^x \cdot 0.3^{10-x} \cong 0.615$.

■

6.2. Πολυωνυμική κατανομή

Στην ***πολυωνυμική δοκιμή*** έχουμε ήδη αναφερθεί όταν μιλήσαμε για τα ανεξάρτητα πειράματα. Αποτελεί μια ευθεία και εύλογη γενίκευση της ***δοκιμής Bernoulli***. Είναι ένα πείραμα με $k \ge 2$ αμοιβαίως αποκλειόμενα δυνατά αποτελέσματα (και όχι κατ' ανάγκη μόνο δύο όπως συμβαίνει σε μια *δοκιμή Bernoulli*). Αντίστοιχα, η ***πολυωνυμική κατανομή*** αποτελεί γενίκευση της ***διωνυμικής κατανομής*** και ορίζεται ως εξής.

Έστω ένα πείραμα τύχης που αποτελείται από ν ανεξάρτητες *πολυωνυμικές δοκιμές*, όπου σε κάθε δοκιμή τα δυνατά αποτελέσματα είναι k, έστω $r_1, r_2, \ldots, r_k$, και η πιθανότητα να εμφανισθεί το αποτέλεσμα r_i, $i = 1, 2, \ldots, k$ σε μια (οποιαδήποτε) δοκιμή είναι p_i. Αν X_i ($i = 1, 2, \ldots, k$) τυχαία μεταβλητή που εκφράζει πόσες φορές εμφανίσθηκε το αποτέλεσμα r_i στις ν ανεξάρτητες επαναλήψεις της *πολυωνυμικής δοκιμής*, τότε η τυχαία μεταβλητή $(X_1, X_2, \ldots, X_k)$ λέμε ότι ακολουθεί την ***πολυωνυμική κατανομή (polynomial distribution)*** με παραμέτρους $\nu, p_1, p_2, \ldots, p_k$ και συμβολίζεται με $M(\nu, p_1, p_2, \ldots, p_k)$. Επίσης, γράφουμε

$$(X_1, X_2, \ldots, X_k) \sim M(\nu, p_1, p_2, \ldots, p_k).$$

Αποδεικνύεται *(δες Πρόταση 4.5.3)* ότι

$$P(X_1 = \nu_1, X_2 = \nu_2, \ldots, X_k = \nu_k) = \frac{\nu!}{\nu_1! \cdot \nu_2! \cdot \ldots \cdot \nu_k!} p_1^{\nu_1} p_2^{\nu_2} \ldots p_k^{\nu_k}$$

με $\sum_{i=1}^{k} \nu_i = \nu$ και $\sum_{i=1}^{k} p_i = 1$.

Δηλαδή, η πιθανότητα, στις ν ανεξάρτητες *πολυωνυμικές δοκιμές* να εμφανισθούν ν_1 αποτελέσματα r_1, ν_2 αποτελέσματα r_2, … και ν_k αποτελέσματα r_k είναι ίση με

$$\frac{\nu!}{\nu_1! \cdot \nu_2! \cdot \ldots \cdot \nu_k!} p_1^{\nu_1} p_2^{\nu_2} \ldots p_k^{\nu_k}.$$

Επίσης, αποδεικνύεται ότι για κάθε $i = 1, 2, \ldots, k$ είναι $E(X_i) = \nu p_i$, δηλαδή, για κάθε ενδεχόμενο (αποτέλεσμα) r_i, η **αναμενόμενη συχνότητα** εμφάνισής του στις ν δοκιμές είναι $E_i = \nu p_i$. Επίσης, για κάθε $i = 1, 2, \ldots, k$ είναι $V(X_i) = \nu p_i (1 - p_i)$. *(Αν σκεφθούμε τις ν δοκιμές ως δοκιμές Bernoulli με «επιτυχία» το ενδεχόμενο (αποτέλεσμα) r_i και «αποτυχία» όλα τα υπόλοιπα, τότε προφανώς $X_i \sim B(\nu, p_i)$ και επομένως $E(X_i) = \nu p_i$ και $Var(X_i) = \nu p_i (1 - p_i)$)*

Ας δούμε δύο παραδείγματα.

Παράδειγμα 6.2.1: *Σύμφωνα με ένα μοντέλο κληρονομικότητας, οι τρεις τύποι απογόνων Α, Β και Γ που προκύπτουν από μια ορισμένη διασταύρωση πειραματόζωων, βρίσκονται σε αναλογία 9:3:4, αντίστοιχα. Αν στο πλαίσιο ενός πειράματος προέκυψαν από μια τέτοια διασταύρωση 64 απόγονοι, πόσοι αναμένεται να είναι τύπου Α, πόσοι τύπου Β και πόσοι τύπου Γ;*

Απάντηση: Η ταξινόμηση ενός απογόνου σε (ακριβώς) μια από τρεις κατηγορίες (τύπος Α, τύπος Β, τύπος Γ) είναι μια *πολυωνυμική δοκιμή* με $k = 3$ δυνατά αποτελέσματα. Πρόκειται επομένως για ένα πείραμα που αποτελείται από $\nu = 64$ *πολυωνυμικές δοκιμές* η κάθε μια από τις οποίες έχει $k = 3$ δυνατά αποτελέσματα: τύπος Α, τύπος Β, τύπος Γ. Η πιθανότητα πραγματοποίησης καθενός από τα τρία δυνατά αποτελέσματα είναι αντίστοιχα, $p_1 = 9/16$, $p_2 = 3/16$ και $p_3 = 4/16$. Έστω X_i, $i = 1, 2, 3$ τυχαίες μεταβλητές που αντίστοιχα εκφράζουν πόσες φορές στις 64 επαναλήψεις εμφανίσθηκε απόγονος τύπου Α, τύπου Β και τύπου Γ.

Οι αναμενόμενες τιμές των X_i, $i = 1,2,3$ στις $\nu = 64$ επαναλήψεις της *πολυωνυμικής δοκιμής* είναι αντίστοιχα

$$E(X_1) = 64(9/16) = 36\,,\ E(X_2) = 64(3/16) = 12 \text{ και } E(X_3) = 64(4/16) = 16\,.$$

Δηλαδή, σύμφωνα με το μοντέλο κληρονομικότητας, από τους 64 απογόνους οι 36 αναμένεται να είναι τύπου Α, οι 12 τύπου Β και οι 16 τύπου Γ.

Στο Β΄ Μέρος θα δούμε πώς αξιοποιούμε τις αναμενόμενες τιμές $E(X_i) = \nu p_i$ ($i = 1,2,\ldots,k$) , για να κάνουμε X^2 *στατιστικούς ελέγχους*.

■

Παράδειγμα 6.2.2: *Από τους άνδρες ηλικίας 18-24 ετών που κατοικούν σε συγκεκριμένη περιοχή, ποσοστό 87% έχει συστολική πίεση μικρότερη από 140mmHg, ποσοστό 12% έχει συστολική πίεση μεγαλύτερη ή ίση από 140 αλλά μικρότερη από 160mmHg και ποσοστό 1% έχει συστολική πίεση μεγαλύτερη ή ίση από 160mmHg. Επιλέγουμε τυχαία 10 άνδρες ηλικίας 18-24 ετών από τη συγκεκριμένη περιοχή και μετράμε τη συστολική τους πίεση. Ποια είναι η πιθανότητα έξι από αυτούς να έχουν συστολική πίεση μικρότερη από 140mmHg, τρεις να έχουν συστολική πίεση μεγαλύτερη ή ίση από 140 αλλά μικρότερη από 160mmHg και ένας μεγαλύτερη ή ίση από 160mmHg;*

Απάντηση: Έστω *Y* η τυχαία μεταβλητή που εκφράζει τη συστολική πίεση των ανδρών ηλικίας 18-24 ετών που κατοικούν στη συγκεκριμένη περιοχή. Οι τιμές της *Y* ταξινομούνται/κατατάσσονται σε τρεις κατηγορίες/κλάσεις, $[0,140)$, $[140,160)$ και $[160,+\infty)$. Επομένως, παρότι η *Y* δεν είναι μια ποιοτική μεταβλητή, η ταξινόμηση μιας (οποιασδήποτε) τιμής της σε μία από τις τρεις αυτές κλάσεις, αποτελεί μια *πολυωνυμική δοκιμή* με $k = 3$ δυνατά αποτελέσματα: μια οποιαδήποτε τιμή της *Y* είτε θα ανήκει στην κλάση $[0,140)$, είτε στην $[140,160)$, είτε στην $[160,+\infty)$.

Έτσι, το πρόβλημα που μελετάμε είναι ένα πρόβλημα $\nu = 10$ *πολυωνυμικών δοκιμών* η κάθε μια από τις οποίες έχει $k = 3$ δυνατά αποτελέσματα: η συστολική πίεση ενός τυχαία επιλεγμένου άνδρα είτε θα ανήκει στην κλάση $[0,140)$ είτε στην κλάση $[140,160)$ είτε στην κλάση $[160,+\infty)$.

Έστω X_i, $i = 1,2,3$ τυχαία μεταβλητή που εκφράζει πόσες φορές στις 10 επαναλήψεις εμφανίσθηκε τιμή της *Y* που ανήκει, αντίστοιχα, στην κλάση $[0,140)$, $[140,160)$, $[160,+\infty)$. Προφανώς ζητάμε την πιθανότητα $P(X_1 = 6,\ X_2 = 3,\ X_3 = 1)$.

Αν p_i, $i = 1,2,3$, η πιθανότητα μια τυχαία τιμή της *Y* να ανήκει, αντίστοιχα, στην κλάση $[0,140)$, $[140,160)$, $[160,+\infty)$, αν δηλαδή, $p_1 = P(Y < 140)$, $p_2 = P(140 \le Y < 160)$ και $p_3 = P(Y \ge 160)$ τότε

$$P(X_1 = 6,\ X_2 = 3,\ X_3 = 1) = \frac{10!}{6!\cdot 3!\cdot 1!} p_1^6 p_2^3 p_3^1$$

και επειδή δίνεται ότι

$p_1 = P(Y < 140) = 0.87$, $p_2 = P(140 \le Y < 160) = 0.12$ και $p_3 = P(Y \ge 160) = 0.01$

έχουμε

$$P(X_1 = 6,\ X_2 = 3,\ X_3 = 1) = \frac{10!}{6!\cdot 3!\cdot 1!} 0.87^6 0.12^3 0.01^1 = 0.0063\,.$$

■

Σε μια κτηνοτροφική περιοχή υπάρχουν 300000 αιγοπρόβατα. Κάθε χρόνο όλα τα αιγοπρόβατα εμβολιάζονται για προστασία από μια συγκεκριμένη ασθένεια. Σύμφωνα με την άδεια χορήγησης του εμβολίου, υπάρχει πολύ μικρή πιθανότητα, ίση με $2 \cdot 10^{-5}$*, το εμβόλιο να προκαλέσει μια πολύ σοβαρή παρενέργεια που οδηγεί στο θάνατο του ζώου που εμβολιάζεται. Μας ενδιαφέρει να βρούμε την πιθανότητα, στη συγκεκριμένη περιοχή, να πεθάνουν σε ένα έτος x ζώα από την παρενέργεια που προκαλεί ο εμβολιασμός.*

Είναι προφανές[9] ότι ο αριθμός X των ζώων που πεθαίνουν από τον εμβολιασμό στη συγκεκριμένη περιοχή σε ένα έτος είναι *διωνυμική* τυχαία μεταβλητή με $X \sim B(300000,\, 2 \cdot 10^{-5})$. Ενδιαφερόμαστε για τις πιθανότητες $P(X = x)$, όπου $x = 0,1,...,300000$.

Έτσι, για την πιθανότητα *σε ένα έτος να μην υπάρξουν θάνατοι ζώων από τον εμβολιασμό* έχουμε

$$P(X = 0) = \binom{300000}{0} \cdot (2 \cdot 10^{-5})^0 (1 - 2 \cdot 10^{-5})^{300000} = 1 \cdot 1 \cdot (0.99998)^{300000} = 0.0024786$$

και για την πιθανότητα *σε ένα έτος να υπάρξουν ακριβώς 2 θάνατοι από τον εμβολιασμό* έχουμε

$$P(X = 2) = \binom{300000}{2} \cdot (2 \cdot 10^{-5})^2 (1 - 2 \cdot 10^{-5})^{299998} = \frac{300000!}{2! \cdot 299998!} \cdot (0.00002)^2 \cdot (0.99998)^{299998} =$$

$$= \frac{299999 \cdot 300000}{2} \cdot (0.00002)^2 \cdot (0.99998)^{299998} = 0.0446165.$$

Ήδη από αυτά τα δύο αριθμητικά παραδείγματα φαίνεται ότι για τιμές των παραμέτρων της *διωνυμικής* κατανομής όπως αυτές του προβλήματός μας ο τύπος της συνάρτησης πιθανότητάς της δεν είναι πολύ πρακτικός για τον υπολογισμό πιθανοτήτων (ιδιαίτερα χωρίς υπολογιστική μηχανή). Μάλιστα, για τιμές x που δεν είναι κοντά στο 0 ή το ν το πρόβλημα γίνεται μεγάλο. Η πιθανότητα, για παράδειγμα, *σε ένα έτος να υπάρξουν ακριβώς 15 θάνατοι από τον εμβολιασμό* είναι…

$$P(X = 15) = \binom{300000}{15} \cdot (2 \cdot 10^{-5})^{15} (1 - 2 \cdot 10^{-5})^{299985} =$$

$$= \frac{300000!}{15! \cdot 299985!} \cdot (0.00002)^{15} \cdot (0.99998)^{299985} =$$

$$= \frac{299986 \cdot 299987 \cdot ... \cdot 300000}{1 \cdot 2 \cdot 3 \cdot ... \cdot 15} \cdot (0.00002)^{15} \cdot (0.99998)^{299985} = ... = 0.0008912.$$

Οι δυσκολίες αυτές αναγνωρίσθηκαν από πολύ νωρίς. Ο Γάλλος μαθηματικός *Simeon Denis Poisson (1781-1840)* αναζήτησε τρόπο αντιμετώπισης του προβλήματος και το 1837 δημοσίευσε το παρακάτω εντυπωσιακό για την εποχή του οριακό θεώρημα.

Θεώρημα 6.3.1 (οριακό θεώρημα Poisson): *Έστω ότι η τυχαία μεταβλητή X ακολουθεί τη διωνυμική κατανομή* $B(\nu,\, p)$ *με συνάρτηση πιθανότητας*

$$P(X = x) = \binom{\nu}{x} p^x (1 - p)^{\nu - x}, \qquad x = 0,1,2,...,\nu.$$

Αν για $\nu \to +\infty$ *το* $p \to 0$ *έτσι ώστε η μέση τιμή της* X *να συγκλίνει προς μια θετική σταθερά* λ*, δηλαδή, έτσι ώστε* $\nu p \to \lambda$*, τότε*

[9] *Με την υπόθεση ότι κάθε ζώο έχει την ίδια πιθανότητα να πεθάνει από τον εμβολιασμό.*

$$\lim_{\nu\to+\infty}\binom{\nu}{x} p^x(1-p)^{\nu-x}=e^{-\lambda}\frac{\lambda^x}{x!},\quad x=0,1,2,...$$

■

Πολύ αργότερα[10] παρατηρήθηκε ότι η συνάρτηση με τύπο

$$f(x)=e^{-\lambda}\frac{\lambda^x}{x!},\quad x=0,1,2,...$$

που εισάγεται με το *οριακό θεώρημα του Poisson,* έχει τις ιδιότητες μιας συνάρτησης πιθανότητας. Πράγματι, προφανώς είναι μη αρνητική συνάρτηση και επίσης

$$\sum_{x=0}^{\infty} f(x)=\sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!}=1$$

αφού είναι γνωστό ότι

$$\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}=e^{\lambda}$$

και επομένως

$$\sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!}=e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}=e^{-\lambda}e^{\lambda}=1.$$

Ορίσθηκε έτσι, η ***κατανομή Poisson***.

Ορισμός 6.3.1 (Η κατανομή Poisson): *Έστω Χ μια διακριτή τυχαία μεταβλητή με συνάρτηση πιθανότητας*

$$f(x)=P(X=x)=e^{-\lambda}\frac{\lambda^x}{x!},\quad x=0,1,2,...$$

όπου $\lambda>0$*. Η κατανομή της τυχαίας μεταβλητής Χ ονομάζεται* ***κατανομή Poisson*** *με παράμετρο* λ *και συμβολίζεται με* $P(\lambda)$.

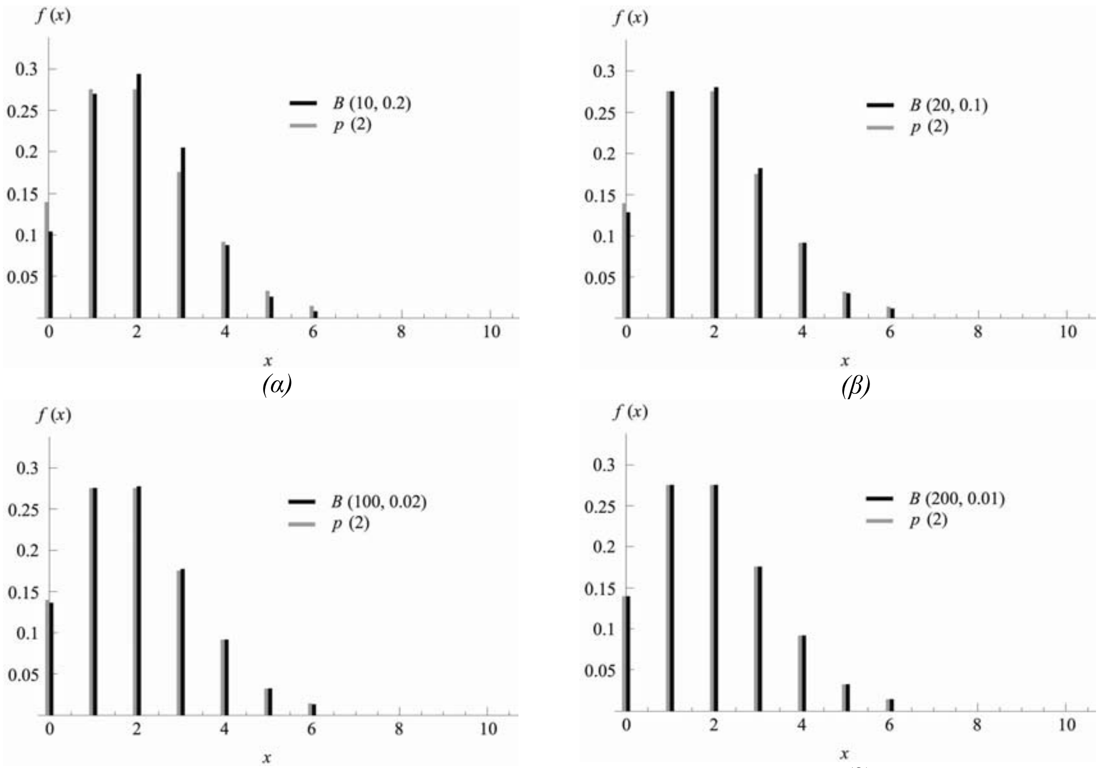
■

Δηλαδή, η *κατανομή Poisson* ορίσθηκε (γεννήθηκε) ως οριακή κατανομή της *διωνυμικής. Έτσι, η* ***διωνυμική κατανομή προσεγγίζεται από την κατανομή Poisson*** *αν για* ν *μεγάλο (θεωρητικά* $\nu\to\infty$*), η πιθανότητα επιτυχίας p συγκλίνει στο 0 (*$p\to 0$*) έτσι ώστε η μέση τιμή της να συγκλίνει σε μια θετική σταθερά* λ *(*$\nu p\to\lambda$*).*

Πρακτικά όμως, ***πόσο μεγάλο πρέπει να είναι το ν*** *και* ***πόσο μικρό το p*** για να είναι ικανοποιητική η προσέγγιση; Έχει παρατηρηθεί ότι *αν* $\nu\ge 20$ *και* $p\le\frac{10}{\nu}$ *ώστε η μέση τιμή* νp *να παίρνει μέτριες τιμές (στην πράξη, μικρότερες του 10), η ακρίβεια της προσέγγισης είναι ικανοποιητική.*

Στα *Σχήματα 6.3.1* φαίνεται γραφικά η σύγκλιση της *διωνυμικής* κατανομής $B(\nu,\ p)$ στην κατανομή *Poisson* με $\lambda=\nu p=2$ όταν $\nu=10,20,100,200$ και $p=0.2,\ 0.1,\ 0.02,\ 0.01$ αντίστοιχα. Παρατηρείστε ότι για $\nu=200$ και $p=0.01$ η προσέγγιση είναι τέλεια.

[10] *Το 1889, από τον Ρωσο-Γερμανό μαθηματικό L.V. Bortriewicz.*



(γ) (δ)

Σχήμα 6.3.1
Σύγκλιση της διωνυμικής κατανομής $B(\nu,\ p)$
στην κατανομή Poisson $P(\nu p)$

Ως παράδειγμα εφαρμογής των παραπάνω ας δούμε πάλι το εισαγωγικό πρόβλημα. Επειδή

$$\nu = 300000 \geq 20 \text{ και } p = 0.00002 \leq \frac{10}{300000} = 0.00003$$

μπορούμε να υπολογίσουμε πολύ πιο εύκολα τις ζητούμενες πιθανότητες αν χρησιμοποιήσουμε την προσέγγιση της *διωνυμικής* κατανομής $B(300000,\ 0.00002)$ από την κατανομή *Poisson* με παράμετρο $\lambda = \nu p = 300000 \cdot 0.00002 = 6$. Έτσι έχουμε

$$P(X = x) \cong e^{-6}\frac{6^x}{x!}, \quad x = 0,1,2,...$$

Άρα

$$P(X = 0) \cong e^{-6}\frac{6^0}{0!} = 0.0024788$$

$$P(X = 2) \cong e^{-6}\frac{6^2}{2!} = 0.0446176$$

$$P(X = 15) \cong e^{-6}\frac{6^{15}}{15!} = 0.0008913.$$

■

Πρόταση 6.3.1: Η *μέση τιμή* και η *διακύμανση* της κατανομής *Poisson* με παράμερο λ, αντίστοιχα είναι

$$\mu = E(X) = \lambda \text{ και } \sigma^2 = Var(X) = \lambda.$$

<u>*Απόδειξη*</u>: Από τον τύπο ορισμού της *μέσης τιμής* έχουμε

$$\mu = E(X) = \sum_{x=0}^{\infty} xf(x) = \sum_{x=0}^{\infty} xe^{-\lambda}\frac{\lambda^x}{x!} = \lambda e^{-\lambda}\sum_{x=1}^{\infty} x\frac{\lambda^{x-1}}{x!} = \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

Για να υπολογίσουμε τη *διακύμανση* $\sigma^2 = Var(X)$ θα χρησιμοποιήσουμε τον τύπο

$$\sigma^2 = E(X^2) - \mu^2$$

όπου τη ροπή $2^{ης}$ τάξης $E(X^2)$ θα την υπολογίσουμε μέσω της παραγοντικής ροπής $2^{ης}$ τάξης από τον τύπο

$$E(X^2) = E[X(X-1)] + E(X)\,.$$

Για την παραγοντική ροπή $2^{ης}$ τάξης $E[X(X-1)]$ έχουμε

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)f(x) = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=2}^{\infty} x(x-1)\frac{\lambda^x}{x!} =$$

$$= \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} = \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2$$

άρα

$$E(X^2) = E[X(X-1)] + E(X) = \lambda^2 + \lambda\,.$$

Η διακύμανση επομένως της X είναι

$$\sigma^2 = E(X^2) - \mu^2 = \lambda^2 + \lambda - \lambda^2 = \lambda\,.$$

Ερώτηση: *Είναι άραγε αναμενόμενο (λογικό) η μέση τιμή της κατανομής Poisson να είναι ίση με τη διακύμανση της[11];*

■

Σημείωση 6.3.1: α) *Όπως και στη διωνυμική κατανομή, εύκολα αποδεικνύεται ο παρακάτω αναδρομικός τύπος υπολογισμού των πιθανοτήτων $P(X = x)$, $x = 0,1,2,...$ μιας Poisson τυχαίας μεταβλητής X.*

Για τις πιθανότητες $P(X = x)$, $x = 1,2,...$, της τυχαίας μεταβλητής $X \sim P(\lambda)$ ισχύει

$$P(X = x) = \frac{\lambda}{x}P(X = x-1)$$

με αρχική συνθήκη $P(X = 0) = e^{-\lambda}$.

β) Εύκολα επίσης αποδεικνύεται ότι η πιο πιθανή τιμή της $X \sim P(\lambda)$ είναι η $x_0 = [\lambda]$ όταν ο θετικός αριθμός λ δεν είναι ακέραιος, ενώ όταν είναι ακέραιος οι τιμές της X με τη μεγαλύτερη πιθανότητα είναι δύο: η $x_0 = \lambda$ και η $x_0' = \lambda - 1$.

■

Η κατανομή *Poisson* ως οριακή κατανομή της *διωνυμικής* κατανομής έχει, όπως και η *διωνυμική*, μεγάλο εύρος εφαρμογών σε διάφορες επιστημονικές περιοχές. Πιο συγκεκριμένα, χρησιμοποιείται για τη μοντελοποίηση *«διωνυμικών καταστάσεων»* όπου ενδιαφέρει **ο αριθμός εμφανίσεων σπάνιων ενδεχομένων σε μεγάλους πληθυσμούς** (δηλ. όταν σε κάθε επανάληψη, η πιθανότητα επιτυχίας p είναι πολύ μικρή και ο αριθμός επαναλήψεων ν πολύ μεγάλος). Γι' αυτό το λόγο, στη βιβλιογραφία συναντάται και ως ***κατανομή των σπάνιων ενδεχομένων (distribution of rare events[12])***.

Ας δούμε μερικά χαρακτηριστικά παραδείγματα τυχαίων μεταβλητών που ακολουθούν την κατανομή *Poisson*.

1. Ο αριθμός X των τροχαίων ατυχημάτων σε ένα τμήμα (με μεγάλη κυκλοφορία) του οδικού δικτύου μιας χώρας στη διάρκεια ενός Σαββατοκύριακου (ή μιας ημέρας, ή μιας εβδομάδας, ή ενός μήνα κτλ.). Με την υπόθεση ότι κάθε αυτοκίνητο που περνάει από το συγκεκριμένο σημείο έχει την ίδια πιθανότητα p να εμπλακεί σε τροχαίο ατύχημα, η X ακολουθεί μια $B(\nu, p)$. Επειδή ο αριθμός

---

[11] *Είναι. Θυμηθείτε τη μέση τιμή και τη διακύμανση της διωνυμικής και παρατηρείστε ότι όταν συγκλίνει στην Poisson το 1-p είναι περίπου 1.*

[12] *Αναφέρεται επίσης ως **νόμος των μικρών αριθμών.***

των αυτοκινήτων ν που διέρχονται από το συγκεκριμένο σημείο στη διάρκεια ενός Σαββατοκύριακου είναι μεγάλος και η πιθανότητα ατυχήματος (επιτυχίας!) p είναι πολύ μικρή[13], η κατανομή της X προσεγγίζεται ικανοποιητικά από την κατανομή *Poisson* με $\lambda = \nu p$. Η παράμετρος λ εκφράζει τον μέσο αριθμό ατυχημάτων στο συγκεκριμένο σημείο στη διάρκεια ενός Σαββατοκύριακου και στην πράξη συνήθως εκτιμάται εμπειρικά από στατιστικά στοιχεία.

2. Ο αριθμός X των τυπογραφικών λαθών σε μια δακτυλογραφημένη σελίδα (ή σε ένα σύνολο σελίδων). Όπως και στο προηγούμενο παράδειγμα πρόκειται για *διωνυμική* κατανομή $B(\nu,\ p)$. Επειδή το ν είναι μεγάλο (πολλά γράμματα στο κείμενο) και η πιθανότητα τυπογραφικού λάθους (επιτυχίας!) p είναι μικρή[14] η κατανομή του αριθμού των τυπογραφικών λαθών ανά σελίδα προσεγγίζεται ικανοποιητικά από την κατανομή *Poisson* με $\lambda = \nu p$. Η παράμετρος λ εκφράζει τον μέσο αριθμό τυπογραφικών λαθών ανά σελίδα και όπως αναφέραμε και στο προηγούμενο παράδειγμα, στην πράξη συνήθως εκτιμάται εμπειρικά.
3. Ο αριθμός X των κλήσεων στο help desk ενός μεγάλου Internet provider σε μια ημέρα (ή σε μια ώρα, ή σε μια εβδομάδα κτλ.).
4. Ο αριθμός X των βλαβών μιας μηχανής σε μια ημέρα (ή σε μια εβδομάδα κτλ.).
5. Ο αριθμός X των εργατικών ατυχημάτων που συμβαίνουν σε μια ημέρα σε συγκεκριμένη βιομηχανική ζώνη.
6. Ο αριθμός X των ατόμων ενός πληθυσμού που ζουν περισσότερα από 100 χρόνια.
7. Ο αριθμός X των παιδιών ενός πληθυσμού που θα γίνουν ψηλότερα από 1.95μέτρα.
8. Ο αριθμός X των βακτηριδίων σε 1cm^2 ενός τρυβλίου Petri.
9. Ο αριθμός X των πελατών ενός super market σε μια ημέρα, που θα αγοράσουν σοκολατάκια για σκύλους.
10. Ο αριθμός X των ελαττωματικών προϊόντων που παράγονται από μια συγκεκριμένη γραμμή παραγωγής σε ορισμένο χρονικό διάστημα.
11. Ο αριθμός X των λανθασμένων τηλεφωνικών κλήσεων (άλλος αριθμός πληκτρολογείται και άλλος καλείται) σε μια ημέρα. Επίσης, ο αριθμός X των τηλεφωνικών κλήσεων που φθάνουν σε ένα τηλεφωνικό κέντρο σε μια συγκεκριμένη χρονική περίοδο.
12. Ο αριθμός X των φυσαλίδων σε υαλοπίνακα συγκεκριμένης επιφάνειας.
13. Ο αριθμός X των θανάτων σε μια πόλη από μια σπάνια ασθένεια σε ένα μήνα.
14. Ο αριθμός X των πελατών που φθάνουν σε ένα κέντρο εξυπηρέτησης (τράπεζα, ταχυδρομικό γραφείο, κατάστημα κτλ.) σε μια ημέρα (ή σε μια ώρα, ή σε μια εβδομάδα, κτλ.).
15. Ο αριθμός X των επιβατών μιας αεροπορικής πτήσης που ενώ έχουν κάνει κράτηση θέσης δεν εμφανίζονται την ώρα αναχώρησης.
16. Ο αριθμός X των α-σωματίων που εκπέμπονται από ραδιενεργό υλικό σε συγκεκριμένο χρονικό διάστημα.
17. Ο αριθμός X των σεισμών μεγέθους μεγαλύτερου των 5.5 βαθμών της κλίμακας *Richter* που συμβαίνουν σε μια σεισμογόνο περιοχή σε ένα έτος.
18. Ο αριθμός X των ελαττωματικών σημείων που υπάρχουν σε συγκεκριμένο μήκος καλωδίου.
19. Ο αριθμός X των βακτηριδίων σε διάλυμα συγκεκριμένου όγκου.
20. Ο αριθμός X των αστεριών σε μια γαλαξιακή περιοχή συγκεκριμένου όγκου.
21. Ο αριθμός X των προβλημάτων (ρωγμές και λακκούβες) στο οδόστρωμα ενός εθνικού δρόμου ανά Km.
22. Ως τελευταίο παράδειγμα αναφέρουμε εφαρμογές στη χωροδιάταξη (spatial pattern) φυτών, ζώων κτλ που είναι τυχαία διασκορπισμένα σε μια μεγάλη έκταση

[13] *Φοβάμαι ότι θλιβερή εξαίρεση αποτελεί η Ελλάδα (με τους οδηγούς της και τους δρόμους της!!).*
[14] *Εκτός αν η δακτυλογράφος έχει πρόβλημα.*

ώστε κάθε δειγματοληπτική μονάδα (τετράγωνο «μικρού» εμβαδού) να έχει πολύ μικρή πιθανότητα να «φιλοξενήσει» ένα φυτό ή ζώο.

Όπως φαίνεται από τα παραπάνω παραδείγματα, η κατανομή *Poisson* βρίσκει εφαρμογή και σε περιπτώσεις όπου σε ένα τυχαίο πείραμα μας ενδιαφέρει **πόσες φορές εμφανίζεται ένα ενδεχόμενο σε <u>*χρονικό διάστημα t*</u> ή σε <u>*μήκος t*</u> ή σε <u>*επιφάνεια t*</u> ή σε <u>*όγκο t*</u>**. Αυτό συμβαίνει διότι και οι περιπτώσεις αυτές, όταν ικανοποιούνται τρεις συγκεκριμένες συνθήκες[15], είναι *«διωνυμικές καταστάσεις»* με ν πολύ μεγάλο και p πολύ μικρό. Οι συνθήκες αυτές για την περίπτωση χρονικού διαστήματος είναι οι εξής[16]:

Σ1. *Η πιθανότητα να εμφανισθεί το ενδεχόμενο σε ένα μικρό χρονικό διάστημα μια φορά είναι ανάλογη του μήκους του.*
Σ2. *Η πιθανότητα να εμφανισθεί το ενδεχόμενο δύο ή περισσότερες φορές σε ένα μικρό χρονικό διάστημα είναι αμελητέα.*
Σ3. *Οι εμφανίσεις του ενδεχομένου σε δύο ξένα χρονικά διαστήματα είναι ανεξάρτητα ενδεχόμενα.*

Υποθέτουμε επίσης ότι οι συνθήκες του πειράματος παραμένουν αμετάβλητες (στο χρόνο, το χώρο, κτλ.). Η εξήγηση, διαισθητικά, γιατί υπό τις συνθήκες Σ1, Σ2 και Σ3, οι περιπτώσεις αυτές είναι «διωνυμικές καταστάσεις» με ν πολύ μεγάλο και p πολύ μικρό είναι σχετικά απλή[17].

Με βάση τα προηγούμενα, είναι φανερό ότι για κάθε $t \geq 0$ έχουμε μια τυχαία μεταβλητή X_t που εκφράζει πόσες φορές εμφανίσθηκε το ενδεχόμενο σε διάστημα *t*. Πρόκειται δηλαδή για μια οικογένεια τυχαίων μεταβλητών $\{X_t,\ t \geq 0\}$ η οποία ονομάζεται ***στοχαστική διαδικασία (ανέλιξη) Poisson (Poisson stochastic process)***. Για τη συνάρτηση πιθανότητας της X_t, όπως προηγουμένως αναφέραμε, αποδεικνύεται ότι:

Αν ένα ενδεχόμενο εμφανίζεται σε χρονικό διάστημα t (ή σε μήκος t ή σε επιφάνεια t ή σε όγκο t) έτσι ώστε να ικανοποιούνται οι συνθήκες Σ1, Σ2 και Σ3, τότε υπάρχει ένας θετικός αριθμός λ τέτοιος ώστε η κατανομή του αριθμού X_t των εμφανίσεων του ενδεχομένου σε χρονικό διάστημα t (ή σε μήκος t ή σε επιφάνεια t ή σε όγκο t), να δίνεται από τον τύπο $$P(X_t = x) = e^{-\lambda t}\frac{(\lambda t)^x}{x!}, \quad x = 0,1,2,\ldots$$

Δηλαδή, η X_t ακολουθεί την κατανομή *Poison* με *μέση τιμή* λt. Το λ εκφράζει τον μέσο αριθμό εμφανίσεων του ενδεχομένου στη μονάδα του χρόνου (ή μήκους ή επιφάνειας ή όγκου) ή αλλιώς τον ρυθμό εμφάνισης του ενδεχομένου. Στην πράξη, το λ εκτιμάται εμπειρικά από στατιστικά στοιχεία.

Παράδειγμα 6.3.1: *Στο help desk ενός μεγάλου Internet provider φθάνουν αιτήματα πελατών με ρυθμό 3 αιτήματα ανά λεπτό. Ποια είναι η πιθανότητα α) σε ένα λεπτό να*

---

[15] *που είναι αρκετά απλές και φυσιολογικές-λογικές*

[16] *Ανάλογα διατυπώνονται για μήκος, επιφάνεια ή όγκο.*

[17] *Αν χωρίσουμε το διάστημα [0, t) σε ν υποδιαστήματα ίδιου πλάτους $\frac{t}{\nu}$ (όπου ν πολύ μεγάλο ώστε $\frac{t}{\nu} \to 0$), η συνθήκη Σ2 εξασφαλίζει ότι πρόκειται για δοκιμές Bernoulli και οι Σ3, Σ1 ότι είναι ανεξάρτητες με σταθερή πιθανότητα επιτυχίας $\frac{\lambda \cdot t}{\nu}$, όπου λ ο αναμενόμενος αριθμός εμφανίσεων στη μονάδα χρόνου, χώρου, κτλ.*

φθάσουν το πολύ 2 αιτήματα β) σε μισό λεπτό να φθάσουν το πολύ 2 αιτήματα γ) σε 2 λεπτά να φθάσουν το πολύ 4 αιτήματα και δ) σε 3 διαφορετικά χρονικά διαστήματα του ενός λεπτού να βρεθούν τουλάχιστον δύο τέτοια διαστήματα σε καθένα από τα οποία να έχουν φθάσει το πολύ 2 αιτήματα.

Απάντηση: Ο αριθμός X_t των αιτημάτων που φθάνουν στο help desk σε διάστημα t λεπτών ακολουθεί *κατανομή Poisson* με

$$P(X_t = x) = e^{-3t}\frac{(3t)^x}{x!}, \quad x = 0,1,2, \ldots$$

Επομένως, έχουμε

α) $P(X_1 \le 2) = P(X_1 = 0) + P(X_1 = 1) + P(X_1 = 2) = \sum_{x=0}^{2} e^{-3}\frac{3^x}{x!} = 0.4232$.

β) $P(X_{1/2} \le 2) = P(X_{1/2} = 0) + P(X_{1/2} = 1) + P(X_{1/2} = 2) = \sum_{x=0}^{2} e^{-1.5}\frac{1.5^x}{x!} = 0.8088$.

γ) $P(X_2 \le 4) = \sum_{x=0}^{4} e^{-6}\frac{6^x}{x!} = 0.2851$.

δ) Τα τρία χρονικά διαστήματα του ενός λεπτού μπορούν να θεωρηθούν ως τρεις ανεξάρτητες *δοκιμές Bernoulli* στις οποίες επιτυχία σημαίνει: σε ένα λεπτό φθάνουν το πολύ δύο αιτήματα. Έτσι αν συμβολίσουμε με Y τον αριθμό των επιτυχιών στις 3 δοκιμές είναι προφανές ότι $Y \sim B(3,\ 0.4232)$ δηλαδή

$$P(Y = y) = \binom{3}{y}(0.4232)^y (0.5768)^{3-y}, \ y = 0,1,2,3$$

και επομένως η ζητούμενη πιθανότητα είναι

$$P(Y \ge 2) = \binom{3}{2}(0.4232)^2 (0.5768)^1 + \binom{3}{3}(0.4232)^3 (0.5768)^0 = 0.3857.$$

■

Σχόλιο 6.3.1: *Όπως ήδη αναφέραμε, στην πράξη η παράμετρος λ συνήθως δεν υπολογίζεται από τις παραμέτρους ν και p αλλά εκτιμάται εμπειρικά από στατιστικά στοιχεία. Για παράδειγμα, ας θυμηθούμε την τυχαία μεταβλητή Y του Παραδείγματος 5.3.1 που εκφράζει τον αριθμό των ατυχημάτων που συμβαίνουν σε μια ημέρα σε μια συγκεκριμένη βιομηχανική ζώνη. Είναι λογικό να θεωρήσουμε ότι πρόκειται για τ.μ. που εκφράζει τον αριθμό εμφανίσεων ενός σπάνιου ενδεχομένου (συμβαίνει ατύχημα στη βιομηχανική ζώνη) και επομένως περιμένουμε να περιγράφεται από μια κατανομή Poisson. Πράγματι, όπως παρατηρήσαμε (Σχόλιο 5.3.1) αλλά και όπως θα διαπιστώσουμε στη συνέχεια, κάνοντας κατάλληλο στατιστικό έλεγχο (Παράδειγμα 14.1.3) δε μπορούμε να απορρίψουμε την ιδέα ότι η Y περιγράφεται από την κατανομή Poisson με παράμετρο $\lambda = 1$. Αυτή η εκτίμηση της παραμέτρου λ προκύπτει από τα δεδομένα που φαίνονται στον πίνακα που ακολουθεί (πρόκειται για στοιχεία που πήραμε από την Επιθεώρηση Εργασίας και αναφέρονται σε 1500 εργάσιμες ημέρες).*

Αριθμός ατυχημάτων σε μια ημέρα y	0	1	2	3	4	5	
Αριθμός ημερών που συνέβησαν y ατυχήματα	549	555	273	93	24	6	1500

Πράγματι,

$$\lambda \cong \frac{0\cdot 549 + 1\cdot 555 + 2\cdot 273 + 3\cdot 93 + 4\cdot 24 + 5\cdot 6}{1500} \cong 1.$$

■