6. Βασικές Διακριτές Κατανομές

6.1 Η Διωνυμική Κατανομή

Η ***Διωνυμική κατανομή*** συνδέεται με ένα πολύ απλό πείραμα τύχης, ίσως το απλούστερο, τη ***δοκιμή Bernoulli***[1]. Όπως ήδη έχουμε αναφέρει *(δες Παράδειγμα 3.1 (α) & (δ))*, πρόκειται για ένα πείραμα τύχης με μόνο δύο, αμοιβαίως αποκλειόμενα, δυνατά αποτελέσματα. Το ένα αποτέλεσμα έχει επικρατήσει να ονομάζεται ***επιτυχία*** και το άλλο ***αποτυχία***.

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: Μας ενδιαφέρει να διατυπώσουμε πιθανοθεωρητικά συμπεράσματα
α) για τον αριθμό των ζώων μιας κτηνοτροφικής μονάδας που έχουν προσβληθεί από μια συγκεκριμένη ασθένεια. Η εξέταση ενός ζώου για το αν έχει προσβληθεί ή όχι από την ασθένεια είναι δοκιμή *Bernoulli* γιατί τα δυνατά αποτελέσματα είναι μόνο δύο: το ζώο είτε έχει προσβληθεί *(επιτυχία)* είτε δεν έχει προσβληθεί *(αποτυχία)*
β) για τον αριθμό αλλοιωμένων προϊόντων που είναι αποθηκευμένα στις εγκαταστάσεις μιας βιομηχανικής μονάδας επεξεργασίας αγροτικών προϊόντων. Ο έλεγχος ενός προϊόντος για το αν είναι αλλοιωμένο ή όχι είναι δοκιμή *Bernoulli* γιατί το προϊόν είτε έχει αλλοιωθεί *(επιτυχία)* είτε δεν έχει αλλοιωθεί *(αποτυχία)*
γ) για τον αριθμό των φυτών μιας καλλιέργειας που η ξηρή φυτική μάζα τους ξεπερνάει τα 150gr. Ο έλεγχος ενός φυτού για το αν η ξηρή μάζα του ξεπερνάει ή όχι τα 150gr είναι δοκιμή *Bernoulli* γιατί το φυτό έχει ξηρή μάζα είτε μεγαλύτερη από 150gr *(επιτυχία)* είτε το πολύ 150gr *(αποτυχία)*
δ) για τον αριθμό των φυτών μιας καλλιέργειας που έχουν λιγότερα από 6 φύλλα. 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)*.

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 (δ). 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 (όπως και στα παραδείγματα (γ) και (δ)), μας ενδιαφέρει εάν συμβαίνει ή όχι, 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».*

Σε μια *δοκιμή Bernoulli* είναι φανερό ότι ο αριθμός των επιτυχιών είναι ή 1 ή 0. Αν η πιθανότητα *επιτυχίας* είναι p προφανώς η πιθανότητα αποτυχίας είναι $1-p$ την οποία συμβολίζουμε με q. Ας συμβολίσουμε επίσης τον αριθμό των επιτυχιών σε μια *δοκιμή Bernoulli* με X. Η κατανομή της τυχαίας μεταβλητής X ονομάζεται ***κατανομή Bernoulli*** με παράμετρο p, συμβολίζεται με $b(p)$ και προφανώς έχει συνάρτηση πιθανότητας,

[1] *Φέρει το όνομα του Ελβετού μαθηματικού James Bernoulli (1654-1705), της οικογένειας διακεκριμένων μαθηματικών.*
[2] *Ο γονότυπος μπορεί να είναι ΑΑ, Αα ή αα.*

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

ή

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

Η *μέση τιμή* μ της X είναι

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

και η *διακύμανσή* της

$$\sigma^{2}=E(X^{2})-\left[E(X)\right]^{2}=0^{2}\cdot f(0)+1^{2}\cdot f(1)-p^{2}=p-p^{2}=p\cdot(1-p)=p\cdot q\,.$$

Παρότι, όπως αναφέραμε αλλά και διαπιστώσαμε, η *δοκιμή Bernoulli* είναι ένα πολύ απλό πείραμα, εντούτοις (ή μήπως γι' αυτό) βρίσκεται στον πυρήνα πολλών πραγματικών 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 αριθμό επαναλήψεων μιας δοκιμής Bernoulli που απαιτούνται μέχρι να συμβεί η πρώτη επιτυχία» ή «τον αριθμό επαναλήψεων μιας δοκιμής Bernoulli που απαιτούνται 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 το ερώτημα: *«πόσες επιτυχίες συμβαίνουν σε ν επαναλήψεις μιας δοκιμής Bernoulli»*. Ας δούμε ένα τέτοιο πρόβλημα.

Πρόβλημα 6.1.1: *Ο γεωπόνος ενός φυτώριου ισχυρίζεται ότι ποσοστό 90% ενός είδους φυτών που παράγει δίνει περισσότερους από 5 καρπούς/φυτό. Ένας αγρότης που είχε προμηθευθεί από το συγκεκριμένο φυτώριο μεγάλο αριθμό φυτών αυτού του είδους, θέλησε να ελέγξει τον ισχυρισμό του γεωπόνου. Για το σκοπό αυτό, κατά τη συγκομιδή, επέλεξε τυχαία 52 φυτά και μέτρησε τους καρπούς κάθε φυτού. Από τα 52 φυτά που εξέτασε, περισσότερους από 5 καρπούς είχαν μόνο τα 38, γεγονός που δημιούργησε αμφιβολίες στον αγρότη για τον ισχυρισμό του γεωπόνου. Είναι, άραγε, δικαιολογημένες οι αμφιβολίες του αγρότη;*

Απάντηση: Είναι προφανές ότι η τυχαία επιλογή και εξέταση ενός φυτού, για το αν έχει ή όχι περισσότερους από 5 καρπούς, είναι μια *δοκιμή Bernoulli:* το φυτό είτε έχει περισσότερους από 5 καρπούς είτε δεν έχει περισσότερους από 5 καρπούς. Η επιλογή επομένως και η εξέταση 52 φυτών είναι μια σειρά-ακολουθία 52 *δοκιμών Bernoulli*. Επειδή ο αγρότης ενδιαφέρεται για τα φυτά που έχουν περισσότερους από πέντε καρπούς είναι λογικό να ονομάσουμε *επιτυχία* το αποτέλεσμα *«το φυτό έχει περισσότερους από πέντε καρπούς»* και *αποτυχία* το αποτέλεσμα *«το φυτό δεν έχει περισσότερους από πέντε καρπούς»*.

Το ερώτημα που, λογικά, προκύπτει είναι: *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 φυτών που παράγουν περισσότερους από 5 καρπούς είναι 90%, πόσο πιθανό είναι από τα 52 τυχαία επιλεγμένα φυτά να βρεθούν 38 τα οποία να έχουν περισσότερους από 5 καρπούς;* 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 φυτών που εξετάζονται είναι πολύ μικρός σε σχέση με τον αριθμό των φυτών της καλλιέργειας και επομένως, παρότι, η δειγματοληψία γίνεται χωρίς επανάθεση η πιθανότητα αυτή πρακτικά δεν αλλάζει. Είναι επίσης λογικό να θεωρήσουμε ότι καθοιονδήποτε τρόπο δεν επηρεάζεται/εξαρτάται το αποτέλεσμα οποιασδήποτε δοκιμής από το αποτέλεσμα προηγούμενων δοκιμών.

Το γενικότερο επομένως ερώτημα που τίθεται είναι:
Αν X 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: $P(X = x), \;\; x = 0,1,2,...,\nu$.

Στα προηγούμενα (θυμηθείτε την *Πρόταση 4.5 (δ))*, 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 φυτών (από τα 52 που επελέγησαν τυχαία) καθένα από τα οποία έχει περισσότερους από 5 καρπούς, τότε:

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

Δηλαδή, με την υπόθεση ότι αληθεύει ο ισχυρισμός του γεωπόνου, η πιθανότητα να συμβεί αυτό που συνέβη στον έλεγχο που έκανε ο αγρότης είναι πολύ μικρή (σχεδόν μηδενική). Άρα οι αμφιβολίες του αγρότη έχουν βάση[3]. *(Φυσικά, ο αγρότης δεν υπολόγισε την πιθανότητα $P(X = 38)$, απλώς υπολόγισε το 90% του 52 που είναι 46.8 και περίμενε να βρει περίπου 47 φυτά με περισσότερους από 5 καρπούς).*

Ας δώσουμε τώρα τον ορισμό της ***Διωνυμικής Κατανομής***[4]:

<u>Ορισμός:</u> *Αν X 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 X ονομάζεται* ***<u>διωνυμική κατανομή</u>*** *με παραμέτρους ν και p και συμβολίζεται με $B(\nu, \; p)$.*

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 τυχαία μεταβλητή X ορίσθηκε ως <u>άθροισμα</u> ν ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν την κατανομή *Bernoulli*. Δηλαδή, αν $X_i, \;\; i = 1,2,...,\nu$, είναι ν ανεξάρτητες τυχαίες μεταβλητές με

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

τότε

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

---

[3] *Περισσότερα επ' αυτού στη Στατιστική Συμπερασματολογία.*

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

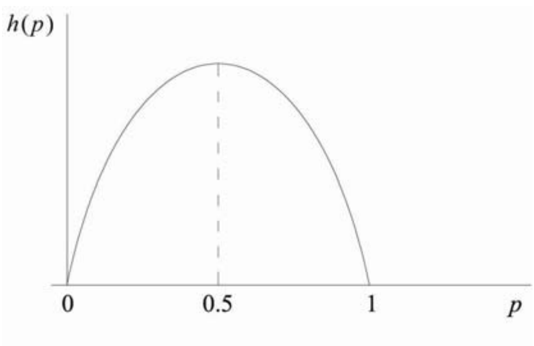
Παρατηρείστε, επίσης, ότι οι πιθανότητες των τιμών *0, 1, 2, ...,* ν, της *X* είναι οι όροι του διωνυμικού αναπτύγματος $[p+(1-p)]^{\nu}$ *(τύπος του Νεύτωνα*[5]*, δες και σελίδα 41 στην Ενότητα «Πώς απαριθμούμε;»).* Έτσι εξηγείται και το όνομα της *διωνυμικής κατανομής.*

Αποδεικνύεται ότι η *μέση τιμή* και η *διακύμανση* της *διωνυμικής* τυχαίας μεταβλητής δίνονται αντίστοιχα από τους τύπους:

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

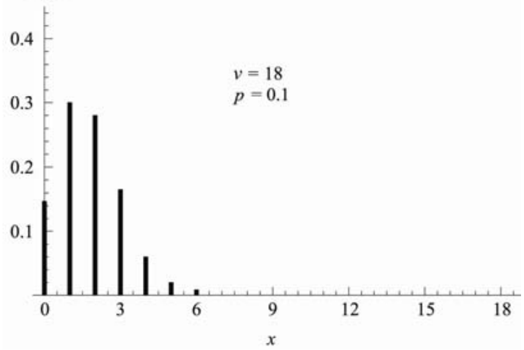
Έτσι, στο *Πρόβλημα 6.1.1*, η μέση τιμή και η *διακύμανση* του αριθμού *X* των φυτών που έχουν περισσότερους από 5 καρπούς είναι $\mu = 52 \cdot 0.9 = 46.8$ φυτά και $\sigma^2 = \nu p(1-p) = 52 \cdot 0.9 \cdot 0.1 = 4.68$ (ή *τυπική απόκλιση* $\sigma = \sqrt{4.68} = 2.16$ φυτά) αντίστοιχα.

<u>Σχόλιο 6.1.1:</u> *Αν δούμε τη διακύμανση,* $\sigma^2 = \nu p(1-p)$*, της διωνυμικής τυχαίας μεταβλητής ως συνάρτηση του p, έστω* $h(p) = \nu p(1-p) = -\nu p^2 + \nu p$*, παρατηρούμε ότι αν* $0 \le p \le \frac{1}{2}$*, η h είναι αύξουσα συνάρτηση του p ενώ αν* $\frac{1}{2} \le p \le 1$ *είναι φθίνουσα συνάρτηση του p και παρουσιάζει μέγιστο για* $p = \frac{1}{2}$ *(Σχήμα 6.1.1).*

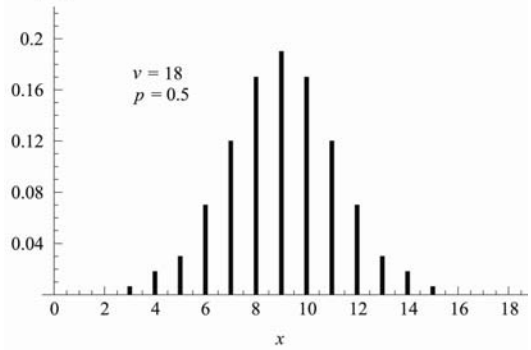


Σχήμα 6.1.1

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



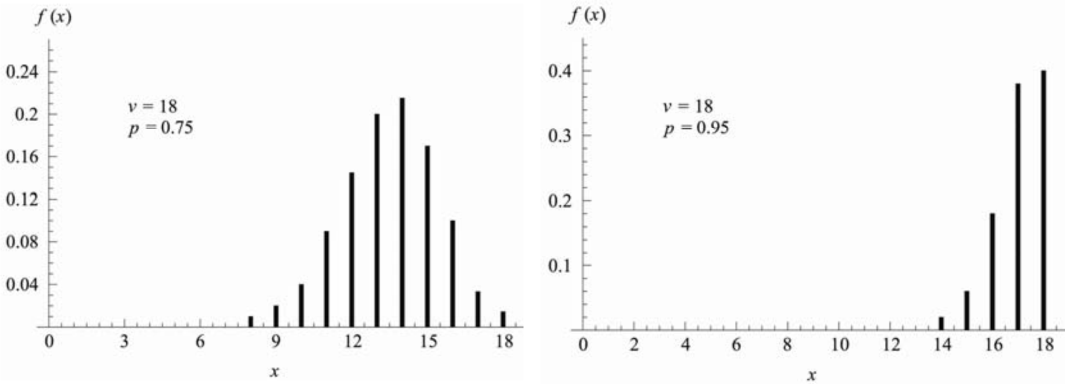
(α)



(β)

---

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



(γ) *(δ)*

Σχήμα 6.1.2

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

<u>Πρόβλημα 6.1.2:</u> *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.*

<u>Απάντηση:</u> 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) $P(X > 7) = P(X=8) + P(X=9) = \binom{9}{8}\cdot 0.8^{8}0.2^{2} + \binom{9}{8}\cdot 0.8^{8}0.2^{2} = 0.436207$

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

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

<u>Παρατηρήσεις 6.1.1:</u>

1. *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, αποδεικνύεται[7] ότι:

---

[6] *Τι υποθέσεις πρέπει να κάνουμε;*

[7] *Η απόδειξη του αναδρομικού τύπου είναι πολύ απλή:*

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

Για τις πιθανότητες, $P(X=x),\ \ x=0,1,\ldots,\nu$ *της τυχαίας μεταβλητής* $X\sim B(\nu,p)$
ισχύει, $P(X=x)=\dfrac{\nu-x+1}{x}\cdot\dfrac{p}{1-p}P(X=x-1)$
με αρχική συνθήκη: $P(X=0)=(1-p)^{\nu}$.

Έτσι, αν υπολογίσουμε μόνο την πιθανότητα $P(X=0)=(1-p)^{\nu}$, *μπορούμε μέσω αυτού του αναδρομικού τύπου να υπολογίσουμε όλες τις πιθανότητες* $P(X=1)$, $P(X=2),...,P(X=\nu)$ *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 όμως φανερό ότι αυτή η μέθοδος υπολογισμού των πιθανοτήτων έχει το εξής μειονέκτημα. Για να υπολογίσουμε μια συγκεκριμένη πιθανότητα $P(X=x)$ *που μας ενδιαφέρει, πρέπει να προηγηθεί ο υπολογισμός των πιθανοτήτων* $P(X=0)$, $P(X=1),...,P(X=x-1)$. *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].

2. *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(α)-(δ) 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* 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, $P(X=x)>P(X=x-1)$ <u>*αν και μόνο αν*</u> $(\nu-x+1)p>x(1-p)$ *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*

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

$P[X=(\nu+1)p]=P[X=(\nu+1)p-1]$, στην περίπτωση αυτή, οι τιμές της Χ με τη μεγαλύτερη πιθανότητα είναι δύο: η $x_0=(\nu+1)p$ και η $x_0'=(\nu+1)p-1$.

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 Πρόβλημα 6.1.1, επειδή το γινόμενο $(\nu+1)p=53\cdot 0.9=47.7$ δεν είναι ακέραιος, η πιο πιθανή τιμή του αριθμού Χ των φυτών που έχουν περισσότερους από 5 καρπούς είναι η $x_0=[47.7]=47$[9].

Στο Πρόβλημα 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(α)-(δ). Επιβεβαιώστε/επαληθεύστε ότι πράγματι τα προηγούμενα συμπεράσματα ισχύουν.

Δείτε την παρακάτω άσκηση και θυμηθείτε τη σχέση *Θεωρίας Πιθανοτήτων* και *Στατιστικής* που συζητήσαμε στο εισαγωγικό μάθημα. Ξανασκεφθείτε, επίσης, το *Πρόβλημα 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 υπολογισθεί η πιθανότητα **(σφάλμα τύπου Ι)** να θεωρήσουμε ότι το νόμισμα είναι μεροληπτικό, ενώ στην πραγματικότητα είναι αμερόληπτο. β) Να υπολογισθεί η πιθανότητα **(σφάλμα τύπου ΙΙ)** να θεωρήσουμε ότι το νόμισμα είναι αμερόληπτο όταν είναι μεροληπτικό με πιθανότητα εμφάνισης της ένδειξης «γράμματα» ίση με 0.7.*

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. Η πολυωνυμική κατανομή (polynomial distribution).

Στα προηγούμενα, όταν μιλήσαμε για τα ανεξάρτητα πειράματα, γνωρίσαμε την ***πολυωνυμική δοκιμή*** η οποία αποτελεί μια ευθεία, προφανή και εύλογη γενίκευση της ***δοκιμής Bernoulli***. Πράγματι, η *πολυωνυμική δοκιμή* είναι ένα πείραμα με $k\geq 2$ 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*

[9] *Δηλαδή ο αγρότης σωστά ανέμενε αυτή την τιμή ως την πιο πιθανή!!!*

δοκιμής, τότε η πολυδιάστατη τυχαία μεταβλητή $(X_1, X_2, \ldots, X_k)$ λέμε ότι ακολουθεί την ***πολυωνυμική κατανομή*** με παραμέτρους $\nu, p_1, p_2, \ldots$ και 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 4.6)* 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 $V(X_i) = \nu p_i (1 - p_i)$)

Ας δούμε δύο παραδείγματα.

<u>Παράδειγμα 6.2.1:</u> *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;*

<u>Απάντηση:</u> 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, $i = 1, 2, 3$ τυχαία μεταβλητή που εκφράζει πόσες φορές στις 64 επαναλήψεις εμφανίσθηκε απόγονος τύπου Α, τύπου Β και τύπου Γ αντίστοιχα.

Οι αναμενόμενες τιμές των τυχαίων μεταβλητών X_i, $i = 1, 2, 3$ στις $\nu = 64$ επαναλήψεις της *πολυωνυμικής δοκιμής* είναι, αντίστοιχα, $E(X_1) = 64(9/16) = 36$, $E(X_2) = 64(3/16) = 12$ 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, $E(X_i) = \nu p_i$ ($i = 1, 2, \ldots, k$), στους X^2 *στατιστικούς ελέγχους*.

<u>Παράδειγμα 6.2.2:</u> *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 τουλάχιστον 140 αλλά μικρότερη από 160mmHg και ποσοστό 1% έχει τουλάχιστον 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 τουλάχιστον 140 αλλά μικρότερη από 160mmHg και ένας τουλάχιστον 160mmHg;*

<u>Απάντηση:</u> 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 (ακριβώς) μία από τρεις κατηγορίες/κλάσεις, $[0,140)$, $[140,160)$ και $[160,+\infty)$. Επομένως, παρότι η Y δεν είναι μια ποιοτική μεταβλητή, η ταξινόμηση μιας (οποιαδήποτε) τιμής της σε μία από τις τρεις αυτές κλάσεις, αποτελεί μια *πολυωνυμική δοκιμή* με $k=3$ δυνατά αποτελέσματα: μια τιμή της Y (οποιαδήποτε) 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

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

6.3 Η Κατανομή Poisson

Σε μια κτηνοτροφική περιοχή υπάρχουν 300000 αιγοπρόβατα. Κάθε χρόνο όλα τα αιγοπρόβατα εμβολιάζονται για προστασία από κάποια ασθένεια. Σύμφωνα με την άδεια χρήσης 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[10] 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 (κυρίως για πολύ μεγάλο ν*)*, 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…

$$P(X=15)=\binom{300000}{15}\cdot(2\cdot 10^{-5})^{15}(1-2\cdot 10^{-5})^{299985}=\frac{300000!}{15!\cdot 299985!}\cdot(0.00002)^{15}\cdot(0.99998)^{299985}=$$
$$=\frac{299986\cdot 299987\cdot\ldots\cdot 300000}{1\cdot 2\cdot 3\cdot\ldots\cdot 15}\cdot(0.00002)^{15}\cdot(0.99998)^{299985}=\ldots.=0.0008912.$$

Οι δυσκολίες αυτές, αναγνωρίσθηκαν από πολύ νωρίς. Ο Γάλλος μαθηματικός *Simeon Denis Poisson (1781-1840)* αναζήτησε τρόπο αντιμετώπισης του προβλήματος και το 1837, σε βιβλίο του με εφαρμογές της θεωρίας πιθανοτήτων σε θέματα δικαστικών υποθέσεων, δημοσίευσε το παρακάτω εντυπωσιακό για την εποχή του οριακό θεώρημα.

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, να ισχύει $\nu p\to\lambda$, τότε

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

[10] *Με την υπόθεση ότι κάθε ζώο έχει την ίδια πιθανότητα να πεθάνει από τον εμβολιασμό σε ένα έτος.*

Πολύ αργότερα[11] διαπιστώθηκε ότι η συνάρτηση με τύπο $f(x) = e^{-\lambda}\dfrac{\lambda^x}{x!}, \quad x = 0,1,2,...$, που εισάγεται με το *οριακό θεώρημα του Poisson,* έχει όλες τις ιδιότητες μιας συνάρτησης πιθανότητας. Έτσι ορίσθηκε η ***κατανομή Poisson***.

<u>Ορισμός:</u> *Έστω X μια διακριτή τυχαία μεταβλητή με συνάρτηση πιθανότητας*

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

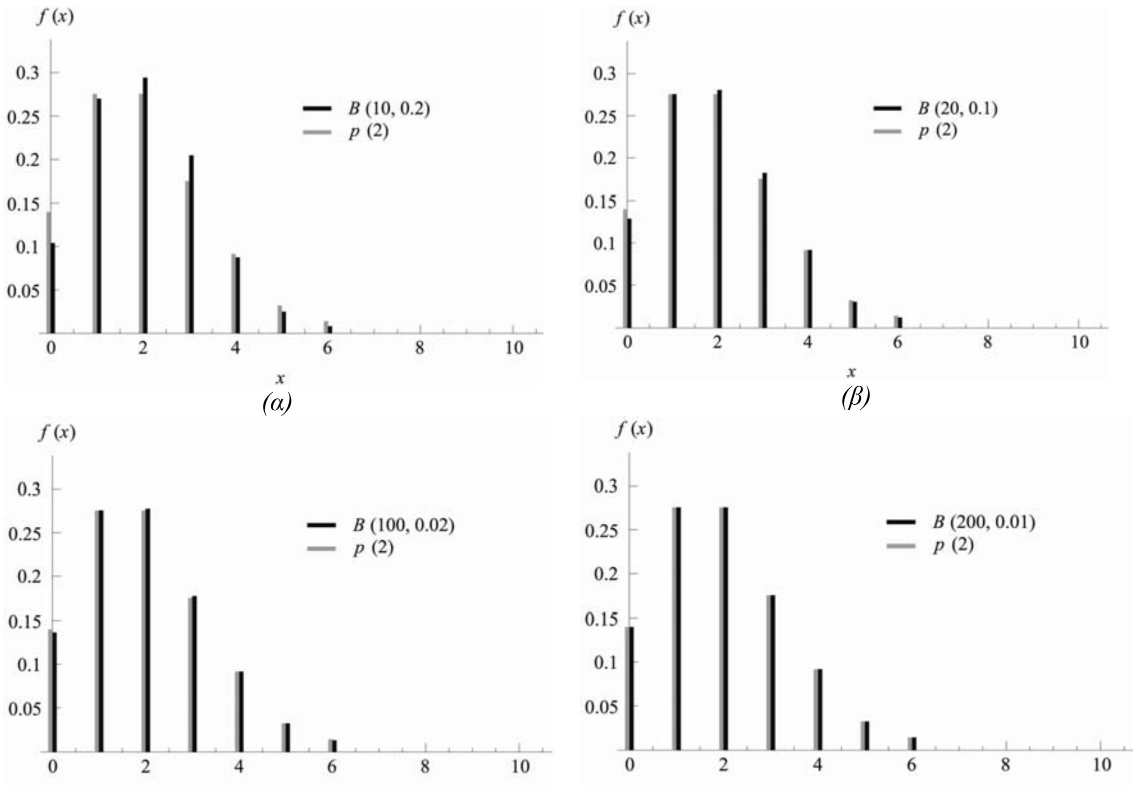
όπου $\lambda > 0$. Η κατανομή της τυχαίας μεταβλητής X ονομάζεται ***κατανομή Poisson*** *με παράμετρο λ και συμβολίζεται με $P(\lambda)$.*

Δηλαδή, η *κατανομή Poisson* ορίσθηκε ως οριακή κατανομή της *διωνυμικής, έτσι:*

*Η **<u>διωνυμική κατανομή προσεγγίζεται από την κατανομή Poisson</u>** αν για ν μεγάλο (θεωρητικά $\nu \to \infty$), η πιθανότητα επιτυχίας p συγκλίνει στο 0 ($p \to 0$) έτσι ώστε η μέση τιμή της κατανομής 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 $\nu \geq 20$ και $p \leq \dfrac{10}{\nu}$ 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(α)-(δ)* που ακολουθούν 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.



Σχήμα 6.3.1

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

Ως παράδειγμα εφαρμογής των παραπάνω ας δούμε πάλι το εισαγωγικό πρόβλημα. Επειδή $\nu = 300000 \geq 20$ 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:

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

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

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

Αποδεικνύεται ότι, αν $X \sim P(\lambda)$ τότε,

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

Στην πράξη, η παράμετρος λ, συνήθως δεν υπολογίζεται από τα ν και p αλλά εκτιμάται εμπειρικά (από στατιστικά στοιχεία).

Ερώτηση: Είναι άραγε αναμενόμενο (λογικό) η μέση τιμή της κατανομής *Poisson* να είναι ίση με τη διακύμανση της[12];

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 αναφέρεται και ως ***κατανομή των σπάνιων ενδεχομένων (distribution of rare events[13])***.

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 την $B(\nu, p)$. Επειδή ο αριθμός των αυτοκινήτων ν που διέρχονται από το συγκεκριμένο σημείο τη συγκεκριμένη χρονική περίοδο είναι μεγάλος και η πιθανότητα ατυχήματος (επιτυχίας!) p είναι πολύ μικρή[14], 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 τη συγκεκριμένη χρονική περίοδο και όπως αναφέραμε, στην πράξη, συνήθως εκτιμάται εμπειρικά (από στατιστικά στοιχεία).

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

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

[14] *Φοβάμαι ότι θλιβερή εξαίρεση αποτελεί η Ελλάδα (με τους οδηγούς της και τους δρόμους της!!).*

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[15] η κατανομή του αριθμού των τυπογραφικών λαθών/σελίδα προσεγγίζεται ικανοποιητικά από την κατανομή *Poisson* με $\lambda = \nu p$. Η παράμετρος λ εκφράζει τον μέσο αριθμό τυπογραφικών λαθών/σελίδα και όπως αναφέραμε και στο προηγούμενο παράδειγμα, στην πράξη, συνήθως εκτιμάται εμπειρικά (από στατιστικά στοιχεία).
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 ατόμων ενός πληθυσμού που ζουν περισσότερα από 100 χρόνια.
6. Ο αριθμός X των παιδιών ενός πληθυσμού που θα γίνουν ψηλότερα από 1.95μέτρα.
7. Ο αριθμός X των βακτηριδίων σε 1cm^2 μιας πλάκας Petri.
8. Ο αριθμός X των πελατών ενός super market σε μια ημέρα, που θα αγοράσουν σοκολατάκια για σκύλους.
9. Ο αριθμός X των ελαττωματικών προϊόντων που παράγονται από μια συγκεκριμένη γραμμή παραγωγής σε ένα ορισμένο χρονικό διάστημα.
10. 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.
11. Ο αριθμός X των φυσαλίδων σε υαλοπίνακα συγκεκριμένης επιφάνειας.
12. Ο αριθμός X των θανάτων σε μια πόλη από μια σπάνια ασθένεια σε ένα μήνα.
13. 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.).
14. Ο αριθμός X των επιβατών μιας αεροπορικής πτήσης που ενώ έχουν κάνει κράτηση θέσης δεν εμφανίζονται την ώρα αναχώρησης.
15. Ο αριθμός X των α-σωματίων που εκπέμπονται από ραδιενεργό υλικό σε συγκεκριμένο χρονικό διάστημα.
16. Ο αριθμός X των σεισμών μεγέθους περισσότερο των 5 βαθμών της κλίμακας *Richter* που πλήττουν μια σεισμογόνο περιοχή σε ένα έτος.
17. Ο αριθμός X των ελαττωματικών σημείων που υπάρχουν σε συγκεκριμένο μήκος καλωδίου.
18. Ο αριθμός X των βακτηριδίων σε διάλυμα συγκεκριμένου όγκου.
19. Ο αριθμός X των αστεριών σε μια γαλαξιακή περιοχή συγκεκριμένου όγκου.
20. Ο αριθμός X των προβλημάτων (ρωγμές και λακκούβες) στο οδόστρωμα ενός εθνικού δρόμου ανά Km.
21. Ως τελευταίο παράδειγμα αναφέρουμε εφαρμογές στη χωροδιάταξη (spatial pattern) φυτών, ζώων κτλ που είναι τυχαία διασκορπισμένα σε μια μεγάλη έκταση 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, όταν ικανοποιούνται τρεις συνθήκες[16], τότε και οι περιπτώσεις αυτές είναι *«διωνυμικές καταστάσεις»* με ν πολύ

[15] *Εκτός αν η δακτυλογράφος έχει πρόβλημα.*
[16] *που είναι αρκετά απλές και φυσιολογικές-λογικές*

μεγάλο και p πολύ μικρό. Οι συνθήκες αυτές για την περίπτωση χρονικού διαστήματος είναι οι εξής[17]:

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[18].

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***. Για τη συνάρτηση πιθανότητας της X_t, όπως αναφέραμε προηγουμένως, 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).

Ας δούμε ένα παράδειγμα.
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 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

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

---

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

[18] *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 Σ1, Σ3 ότι είναι ανεξάρτητες με σταθερή πιθανότητα επιτυχίας $\frac{\lambda \cdot t}{\nu}$, όπου λ ο αναμενόμενος αριθμός εμφανίσεων στη μονάδα χρόνου, χώρου, κτλ.*

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:

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

<u>Παρατηρήσεις 6.3.1:</u>

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

Για τις πιθανότητες $P(X = x)$, $x = 1,2,...$, *της τυχαίας μεταβλητής* $X \sim P(\lambda)$ *ισχύει,* $P(X = x) = \frac{\lambda}{x}\cdot P(X = x-1)$ *με αρχική συνθήκη,:* $P(X = 0) = e^{-\lambda}$.

2. *Όπως και στη διωνυμική κατανομή, εύκολα αποδεικνύεται ότι η πιο πιθανή τιμή της* $X \sim P(\lambda)$ *είναι η* $x_0 = [\lambda]$ *όταν ο θετικός αριθμός λ <u>δεν είναι</u> ακέραιος, ενώ όταν <u>είναι</u> ακέραιος οι τιμές της X με τη μεγαλύτερη πιθανότητα είναι δύο: η* $x_0 = \lambda$ *και η* $x_0' = \lambda - 1$.

Προβλήματα και Ασκήσεις

1. Έχει διαπιστωθεί ότι το 30% των ζώων που κατά τον εμβολιασμό εκτίθενται σε μολυσμένη βελόνα από ηπατίτιδα Β, αναπτύσσει ηπατίτιδα Β. Επιλέγουμε τυχαία 5 ζώα από τον πληθυσμό των ζώων που έχουν εμβολιασθεί με μολυσμένη βελόνα. α) Ποια είναι πιθανότητα, από τα 5 ζώα να βρεθούν άρρωστα από ηπατίτιδα Β i) ακριβώς 2 β) το πολύ 1 iii) τουλάχιστον 3. β) Πόσα ζώα (από τα 5) αναμένεται να βρεθούν άρρωστα από ηπατίτιδα Β.

2. Σε έναν πληθυσμό της μύγας *Drosophila melanogaster* το 25% είναι μαύρες λόγω κάποιας μετάλλαξης, ενώ το υπόλοιπο 75% έχουν το φυσιολογικό τους γκρι χρώμα. Από 7 μύγες που παγιδεύσαμε τυχαία από αυτό τον πληθυσμό α) πόσες αναμένεται να είναι μαύρες β) ποια είναι η πιθανότητα να βρεθούν να είναι μαύρες i) όλες ii) καμία iii) τουλάχιστον μια iv) λιγότερες από 7 v) ακριβώς 3 vi) λιγότερες από 3 vii) τουλάχιστον 3 viii) το πολύ 3 γ) πόσες μύγες είναι πιο πιθανό να βρεθούν να είναι μαύρες.

3. Είναι γνωστό ότι από τους σπόρους συγκεκριμένου είδους βλαστάνει μόνο το 80%. Από μια συσκευασία 10 σπόρων αυτού του είδους α) πόσοι σπόροι αναμένεται να βλαστήσουν β) ποια είναι η πιθανότητα να βλαστήσει i) τουλάχιστον ένας ii) όλοι iii) τουλάχιστον οκτώ. γ) πόσοι σπόροι είναι πιο πιθανό να βλαστήσουν;

4. Μια αεροπορική εταιρεία έχει παρατηρήσει ότι το 5% των ατόμων που κάνουν κράτηση για να ταξιδέψουν δεν εμφανίζονται. Αν σε μια πτήση που γίνεται με ένα μικρό αεροσκάφος χωρητικότητας 50 ατόμων η εταιρεία κάνει κράτηση για 52 άτομα, ποια είναι η πιθανότητα να υπάρχει διαθέσιμο κάθισμα για καθένα άτομο που εμφανίζεται για να ταξιδέψει (από τα 52).

5. Το δίκτυο ομβρίων μιας αγροτικής περιοχής δεν μπορεί να ανταποκριθεί σε δυσμενείς καιρικές συνθήκες που εμφανίζονται στην περιοχή κατά μέσο όρο, μια φορά στα 50 χρόνια. Να υπολογισθεί η πιθανότητα, σε μια χρονική περίοδο 10 ετών να πλημμυρίσει η περιοχή α) τουλάχιστον μια χρονιά β) τουλάχιστον δύο χρονιές, γνωρίζοντας ότι είχε πλημμυρίσει τουλάχιστον μια χρονιά.

6. Ένα άτομο ισχυρίζεται ότι έχει την ικανότητα να προβλέπει το μέλλον. Για να ελεγχθεί ο ισχυρισμός του, ρίχνεται ένα νόμισμα 8 φορές και του ζητείται να προβλέψει τα αποτελέσματα των ρίψεων. Αν προβλέψει σωστά τα 6 αποτελέσματα, νομίζετε ότι έχει βάση ο ισχυρισμός του;

7. **Αξιοπιστία:** Ένα σύστημα αποτελείται από 10 εξαρτήματα που λειτουργούν ανεξάρτητα το ένα από το άλλο. Η αξιοπιστία κάθε εξαρτήματος είναι 0.8 και το σύστημα λειτουργεί μόνο αν τουλάχιστον 8 εξαρτήματά του λειτουργούν. α) Να βρεθεί η αξιοπιστία R του συστήματος β) Να βρεθεί η πιθανότητα να υποστούν βλάβη τουλάχιστον δύο εξαρτήματα δεδομένου ότι έχει υποστεί βλάβη τουλάχιστον ένα.

8. **Αξιοπιστία:** Ένα σύστημα αποτελείται από ν_1 εξαρτήματα τύπου Ι και ν_2 εξαρτήματα τύπου ΙΙ που λειτουργούν ανεξάρτητα το ένα από το άλλο. Η αξιοπιστία κάθε εξαρτήματος τύπου Ι είναι R_1 ενώ η αξιοπιστία κάθε εξαρτήματος τύπου ΙΙ είναι R_2. Το σύστημα λειτουργεί μόνο αν λειτουργούν συγχρόνως τουλάχιστον 2 εξαρτήματα τύπου Ι και τουλάχιστον 2 εξαρτήματα τύπου ΙΙ. Να βρεθεί η αξιοπιστία R του συστήματος.

9. Οι σπόροι του μπιζελιού *Pisum sativum* είναι πράσινοι ή κίτρινοι. Μια συγκεκριμένη διασταύρωση δίνει σπόρους σε αναλογία 3 κίτρινοι:1 πράσινο. Αν από τους σπόρους που προέκυψαν από μια τέτοια διασταύρωση επιλέξουμε τυχαία τέσσερις, ποια είναι η πιθανότητα να βρούμε (α) τρεις κίτρινους και έναν πράσινο (β) και τους τέσσερις πράσινους (γ) και τους τέσσερις του ιδίου χρώματος.

10. Να βρεθεί η συνάρτηση πιθανότητας της τυχαίας μεταβλητής X στα ακόλουθα προβλήματα: α) Ο αριθμός X των ελαττωματικών προϊόντων που περιέχονται σε ένα δείγμα 20 προϊόντων που έχουν επιλεγεί χωρίς επανάθεση από ένα σωρό 100 προϊόντων στον οποίο υπάρχουν 16 ελαττωματικά προϊόντα. β) Ο αριθμός X των ελαττωματικών προϊόντων του ερωτήματος (α) στην περίπτωση που η επιλογή των 20 προϊόντων του δείγματος γίνει με επανάθεση.

11. Έχει παρατηρηθεί ότι η πιθανότητα να συμβεί σοβαρό ατύχημα με γεωργικό μηχάνημα σε μια μεγάλη αγροτική περιοχή είναι 0.0001. Αν κατά τη διάρκεια μιας εργάσιμης ημέρας στην περιοχή αυτή χρησιμοποιούνται 1000 γεωργικά μηχανήματα, ποια είναι η πιθανότητα να γίνουν τουλάχιστον δύο ατυχήματα; (Για την επιλογή της κατανομής που θα χρησιμοποιήσετε παρατηρείστε ότι ο αριθμός τον επαναλήψεων είναι πολύ μεγάλος και η πιθανότητα «επιτυχίας» πολύ μικρή).

12. Έχει παρατηρηθεί ότι στις αεροπορικές πτήσεις που εξυπηρετούνται από μεγάλα αεροπλάνα, από τα άτομα που έχουν κάνει κράτηση θέσης δεν εμφανίζονται κατά μέσο όρο 10 άτομα. α) Ποια είναι η πιθανότητα ένα άτομο που βρίσκεται στην τρίτη θέση της λίστας αναμονής για μια τέτοια πτήση να μπορέσει να ταξιδέψει β) Αν τρία άτομα που είχαν κάνει κράτηση σε μια συγκεκριμένη πτήση έχουν ήδη ακυρώσει την κράτηση τους, ποια είναι η πιθανότητα το έκτο άτομο της λίστας αναμονής να μην μπορέσει να ταξιδέψει.

13. Σε μια γραμμή παραγωγής τυποποιημένων τροφίμων, τα παραγόμενα τεμάχια (μικρά σε όγκο) συσκευάζονται σε μεγάλα κιβώτια και αποστέλλονται στους προμηθευτές. Τα ελαττωματικά προϊόντα σε κάθε κιβώτιο είναι κατά μέσο όρο 0.5. α) Ποια είναι η πιθανότητα ένα κιβώτιο να περιέχει 2 τουλάχιστον ελαττωματικά προϊόντα β) Ας υποθέσουμε ότι γίνεται έλεγχος όλων των τεμαχίων σε κάθε κιβώτιο και όταν εντοπίζεται κιβώτιο με τουλάχιστον δύο ελαττωματικά προϊόντα πραγματοποιείται ρύθμιση των μηχανημάτων κατασκευής. Ποια είναι η πιθανότητα μέχρι τον έλεγχο των 10 πρώτων κιβωτίων να χρειασθεί ρύθμιση των μηχανημάτων ακριβώς 5 φορές.

14. Από παρατηρήσεις πολλών ετών, έχει επαληθευθεί ότι ο αριθμός X_t των σεισμών μεγέθους μεγαλύτερου των 5 Richter που πλήττουν μια σεισμογενή περιοχή σε χρόνο t , περιγράφεται ικανοποιητικά από μια διαδικασία Poisson. Αν ο ρυθμός εμφάνισής τους είναι 4 ανά έτος, α) ποια είναι η πιθανότητα να υπάρξουν τουλάχιστον δύο σεισμοί μεγέθους μεγαλύτερου των 5 Richter σε ένα χρονικό διάστημα i. ενός έτους ii. τεσσάρων μηνών β) ποια είναι η πιθανότητα στα επόμενα 10 έτη να υπάρξουν ακριβώς 3 έτη, στα οποία να συμβούν τουλάχιστον 2 σεισμοί μεγέθους μεγαλύτερου των 5 Richter.

15. Έχει παρατηρηθεί ότι ο αριθμός X_t ενός σπάνιου είδους φυτών σε έκταση εμβαδού *t*, περιγράφεται ικανοποιητικά από μια στοχαστική διαδικασία *Poisson* με μέσο αριθμό φυτών ανά στρέμμα 3 φυτά. **α)** Να βρεθεί η πιθανότητα i. σε μια έκταση ενός στρέμματος να υπάρχουν τουλάχιστον δύο φυτά ii. σε μια έκταση μισού στρέμματος να υπάρχει το πολύ ένα φυτό iii. σε μια έκταση 2 στρεμμάτων να υπάρχουν τουλάχιστον τρία φυτά. **β)** Σε μια έκταση 3 στρεμμάτων ποιος είναι

ο πιθανότερος αριθμός φυτών; Επίσης, να βρεθεί η πιθανότητα εμφάνισης αυτού του αριθμού φυτών.

16. *Συνέχεια της άσκησης 5.11:* Παίρνουμε ένα τυχαίο δείγμα τριών τέτοιων εξαρτημάτων. Ποια είναι η πιθανότητα ακριβώς ένα από αυτά να λειτουργήσει από 200 έως 250 ώρες.

17. Επιλέγονται τυχαία 6 άτομα, υποβάλλονται στην εξέταση και το αποτέλεσμα και στις έξι περιπτώσεις είναι θετικό. Ποια είναι η πιθανότητα τουλάχιστον τρία από αυτά πράγματι να πάσχουν από την ασθένεια.

18. Σε μια πολύ μεγάλη μονάδα θερμοκηπίων έχει εγκατασταθεί σύστημα αυτόματου ποτίσματος. Έχει παρατηρηθεί ότι ο αριθμός X_t των ελαττωματικών μπεκ σε μήκος σωλήνα t, περιγράφεται ικανοποιητικά από μια στοχαστική διαδικασία *Poisson* με μέσο αριθμό *ελαττωματικών* μπεκ, 3 μπεκ ανά κομμάτι σωλήνα. α) Να βρεθούν οι πιθανότητες: i. σε ένα κομμάτι σωλήνα να υπάρχουν τουλάχιστον δύο ελαττωματικά μπεκ ii. σε μισό κομμάτι σωλήνα να υπάρχει το πολύ ένα ελαττωματικό μπεκ iii. σε μήκος σωλήνα όσο δύο κομμάτια να υπάρχουν τουλάχιστον τρία ελαττωματικά μπεκ iv. από τρία τυχαία επιλεγμένα κομμάτια σωλήνα, να υπάρχει το πολύ ένα κομμάτι με τουλάχιστον δύο ελαττωματικά μπεκ. β) Σε μήκος σωλήνα όσο τρία κομμάτια, ποιος είναι ο πιθανότερος αριθμός ελαττωματικών μπεκ.

19. Από τις αιτήσεις υποψηφίων για εγγραφή σε ένα πανεπιστήμιο, το 60% γίνονται δεκτές, το 35% δε γίνονται δεκτές και το 5% γίνονται δεκτές υπό κάποιες προϋποθέσεις. Από 500 αιτήσεις που έγιναν φέτος πόσες αναμένεται να γίνουν δεκτές και πόσες να γίνουν δεκτές υπό προϋποθέσεις.

20. Πανελλαδική έρευνα που έγινε πριν ένα έτος για τη διερεύνηση της στάσης των νέων ηλικίας 18-25 ετών απέναντι στην απαγόρευση του καπνίσματος σε δημόσιους χώρους, έδωσε τα ακόλουθα αποτελέσματα.

Στάση απέναντι στην απαγόρευση του καπνίσματος

Πολύ αρνητική	*Αρνητική*	*Αδιάφορη*	*Θετική*	*Πολύ θετική*
11.6%	*21.1%*	*11.2%*	*32.5%*	*23.6%*

Επιλέγουμε με βάση ένα σχέδιο τυχαίας δειγματοληψίας 6 νέους ηλικίας 18-25 ετών. Αν δεχθούμε ότι τα αποτελέσματα της (προ ενός έτους) έρευνας ισχύουν ακόμη, ποια είναι η πιθανότητα δύο (από τους 6) να είναι αρνητικοί, τρεις να τους αφήνει αδιάφορους μια τέτοια απαγόρευση και ένας να είναι πολύ θετικός;

21. Σε ένα κλασσικό πείραμα του Mendel με μπιζέλια, από μια συγκεκριμένη διασταύρωση (φυτά που έχουν στρογγυλούς κίτρινους σπόρους με φυτά που έχουν ρυτιδωμένους πράσινους σπόρους) παράγονται απόγονοι τεσσάρων ειδών. Στρογγυλοί κίτρινοι, στρογγυλοί πράσινοι, ρυτιδωμένοι κίτρινοι και ρυτιδωμένοι πράσινοι. Σύμφωνα με το μοντέλο κληρονομικότητας του Mendel τα τέσσερα είδη απογόνων βρίσκονται σε αναλογία 9:3:3:1 αντίστοιχα. Αν επιλέξουμε τυχαία τέσσερις απογόνους που προκύπτουν από μια τέτοια διασταύρωση, ποια είναι η πιθανότητα να εμφανισθεί ένας απόγονος από κάθε είδος;

Απαντήσεις

1. *α) i)0.309 ii) 0.528 iii) 0.163. β) 1.5*
2. *α) 1.75 β) i) 0.00006 ii) 0.1335 iii) 0.8665 iv) 0.99994 v) 0.1730 vi) 0.7565 vii) 0.2430 viii) 0.9295 γ) 1 ή 2*
3. *α) 8 β) i)* ≈ 1 *ii) 0.1074 iii) 0.6778 γ) 8*
4. *0.74*
5. *α) 0.183 β) 0.087*
6. *Ναι (η πιθανότητα* $P(X \geq 6) = 0.1445$ *είναι μικρή και επίσης E(X)=4)*
7. *α) 0.6778 β) 0.699*
8. $R = \prod_{i=1}^{2} \left\{ 1 - [(1-R_i)^{\nu_i} + \nu_i R_i (1-R_i)^{\nu_i - 1}] \right\}$
9. *α) 0.4219 β) 0.0039 γ) 0.3203*
10. *α)* $f(x) = P(X = x) = \dfrac{\binom{16}{x} \cdot \binom{84}{20-x}}{\binom{100}{20}}, \ x = 0,1,\ldots,16$

    β) $f(x) = P(X = x) = \binom{20}{x} \cdot 0.16^{x} \cdot 0.84^{20-x}, \ \ x = 0,1,\ldots,20$
11. *0.00467884*
12. *α) 0.997 β) 0.0645*
13. *α) 0.09 β) 0.0009*
14. *α) i.0.9084 ii. 0.3849 β) 0.00000487*
15. *α) i)* $1 - 4e^{-3}$ *ii)* $\dfrac{5}{2} e^{-3/2}$ *iii)* $1 - 25e^{-6}$ *β) 9 και 8, 0.1318*
16. *0.2454*
17. *0.2*
18. *α) i)* $1 - 4e^{-3}$ *ii)* $\dfrac{5}{2} e^{-3/2}$ *iii)* $1 - 25e^{-6}$ *iv) 0.104 β) 9 και 8*
19. *300, 25*
20. *0.000886*
21. *0.0297*