4.4. **Κατανομές Πιθανοτήτων: Διακριτές Μεταβλητές**

Στο κεφάλαιο 1 περιγράψαμε την έννοια των όρων **αριθμητικά δεδομένα, τιμές μεταβλητής**, και **τιμές μιας τυχαίας μεταβλητής**, που αναφέρονται στις τιμές των ποσοτικών χαρακτηριστικών που εξετάζουμε. Επίσης, αναφερθήκαμε στη διάκριση των μεταβλητών σε ασυνεχείς (διακριτές) και συνεχείς μεταβλητές. Σε αυτή την παράγραφο θα ασχοληθούμε με τις κατανομές πιθανοτήτων.

Για μία διακριτή (ασυνεχή) μεταβλητή, **κατανομή πιθανοτήτων** (probability distribution) ονομάζεται η καταγραφή όλων των δυνατών τιμών της μεταβλητής με τις αντίστοιχες πιθανότητες εμφάνισής τους. Για παράδειγμα. η κατανομή πιθανοτήτων για το αριθμητικό αποτέλεσμα της ρίψης ενός ζαριού, είναι:

Αριθμητικό αποτέλεσμα τυχαίας ρίψης (Χ)	Πιθανότητα (P)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
Σύνολο	**1**

Ας δούμε ένα άλλο παράδειγμα. Η (εμπειρική) κατανομή πιθανοτήτων του αριθμού των παιδιών ανά οικογένεια, σύμφωνα με την τελευταία απογραφή της Ελληνικής Στατιστικής Αρχής (ΕΛΣΤΑΤ), είναι η εξής (Διάγραμμα 4.1: υποθετικά δεδομένα):

Αριθμός παιδιών ανά οικογένεια (Χ)	**Πιθανότητα (P)**
0	0,15
1	0,25
2	0,35
3	0,15
4	0,07
5	0,02
6	0,01
Σύνολο	**1,00**

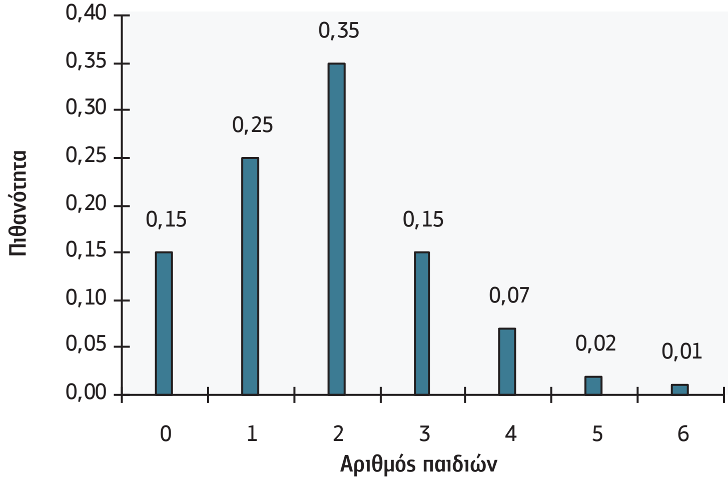
Και στα δύο παραδείγματα το άθροισμα των πιθανοτήτων είναι 1. Αυτό είναι λογικό, αφού οι τιμές της μεταβλητής Χ αναφέρονται σε όλα τα δυνατά αποτελέσματα, δηλαδή καλύπτουν ολόκληρο τον δειγματικό χώρο. Έτσι, μία κατανομή πιθανοτήτων δίνει όλες τις δυνατές τιμές της μεταβλητής Χ (X_i, $i = 1,\ldots, n$), και τις αντίστοιχες πιθανότητες $P(X_i)$. Η πιθανότητα η μεταβλητή Χ να πάρει την τιμή X_i συμβολίζεται με $P(X_i)$. Δηλαδή:

$$P(X = X_i) = P(X_i) \tag{4.10}$$

Όλοι οι κανόνες που περιγράψαμε πριν ισχύουν και στις κατανομές πιθανοτήτων. Για παράδειγμα, ποια η πιθανότητα μία οικογένεια να έχει τρία ή περισσότερα παιδιά; Με βάση τον κανόνα της άθροισης, και λαμβάνοντας υπόψη ότι σε μία κατανομή πιθανοτήτων τα ενδεχόμενα είναι αμοιβαίως αποκλειόμενα, έχουμε:

$$\begin{aligned} P(X \geq 3) &= P(3) + P(4) + P(5) + P(6) \\ &= 0{,}15 + 0{,}07 + 0{,}02 + 0{,}01 \\ &= 0{,}25 \;\; (25\% \;\text{ή}\; 1/4) \end{aligned}$$

Δηλαδή, μία στις τέσσερις οικογένειες έχει τουλάχιστον τρία παιδιά.



Διάγραμμα 4.1
Κατανομή Πιθανοτήτων Αριθμού Παιδιών ανά Οικογένεια

Τα βασικά χαρακτηριστικά των εμπειρικών κατανομών που περιγράψαμε στο κεφάλαιο 3 θα τα περιγράψουμε και για τις κατανομές πιθανοτήτων. Πιο συγκεκριμένα, για κάθε κατανομή πιθανοτήτων μας ενδιαφέρουν δύο παράμετροι. Ο μέσος όρος και η τυπική απόκλιση. Υπάρχει όμως μια διαφορά. Επειδή μία κατανομή πιθανοτήτων περιγράφει τη **θεωρητική** κατανομή μιας τυχαίας μεταβλητής, δηλαδή την πιθανότητα που έχει μια τιμή της X να εμφανιστεί, αντί του όρου μέσου αριθμητικού, χρησιμοποιούμε τον όρο **μέση αναμενόμενη τιμή** ή απλά **αναμενόμενη τιμή** (expected value), και συμβολίζεται με E(X) ή με **μ**, επειδή αντιστοιχεί στον πραγματικό μέσο του "πληθυσμού" των πιθανών τιμών της X. Υπολογίζεται δε από το γινόμενο κάθε πιθανής τιμής (X_i) επί την αντίστοιχη πιθανότητα $P(X_i)$, και στη συνέχεια αθροίζοντας όλα τα γινόμενα. Δηλαδή:

$$\mu = E(X) = X_1 \times P(X_1) + X_2 \times P(X_2) + \ldots + X_n \times P(X_n) = \Sigma\, X_i \times P(X_i) \qquad (4.11)$$

Η αναμενόμενη τιμή της κατανομής του αριθμού των παιδιών ανά οικογένεια είναι:

$$E(X) = 0\times(0{,}15)+1\times(0{,}25)+2\times(0{,}35)+3\times(0{,}15)+4\times(0{,}07)+5\times(0{,}02)+6\times(0{,}01)$$
$$= 1{,}84 \text{ παιδιά}$$

Με ανάλογο τρόπο εκτιμούμε τη διακύμανση. Η διακύμανση (variance) σε μία κατανομή πιθανοτήτων ορίζεται σαν την αναμενόμενη τιμή των τετραγωνικών αποκλίσεων των τιμών της X από το μέσο μ. Δηλαδή:

$$Var(X) = \sigma^2 = E(X - \mu)^2 = \Sigma\,(X_i - \mu)^2 \times P(X_i) \qquad (4.12)$$

Η διακύμανση συμβολίζεται με Var(X) ή $\boldsymbol{\sigma^2}$, για τους ίδιους λόγους που η αναμενόμενη τιμή συμβολίζεται με $\boldsymbol{\mu}$. Δηλαδή, επειδή αντιστοιχεί στην πραγματική διακύμανση του "πληθυσμού" όλων των πιθανών τιμών της X. Στο σημείο αυτό πρέπει να διευκρινίσουμε ότι με ελληνικούς χαρακτήρες συμβολίζουμε τις τιμές των παραμέτρων του πληθυσμού, ενώ με λατινικούς χαρακτήρες τις εκτιμήσεις τους από τις τιμές ενός δείγματος. Έτσι, με $\overline{X}$ συμβολίζουμε την εκτίμηση του μ, και με s^2 την εκτίμηση του σ^2.

Η διακύμανση της κατανομής του αριθμού των παιδιών είναι:

$$\sigma^2 = (0-1{,}84)^2 \times (0{,}15) + (1-1{,}84)^2 \times (0{,}25) + (2-1{,}84)^2 \times (0{,}35) + (3-1{,}84)^2 \times (0{,}15)$$
$$+ (4-1{,}84)^2 \times (0{,}07) + (5-1{,}84)^2 \times (0{,}02) + (6-1{,}84)^2 \times (0{,}01) = 1{,}594 \text{ (παιδιά)}^2$$

Και η τυπική απόκλιση προκύπτει από τη διακύμανση, ως εξής:

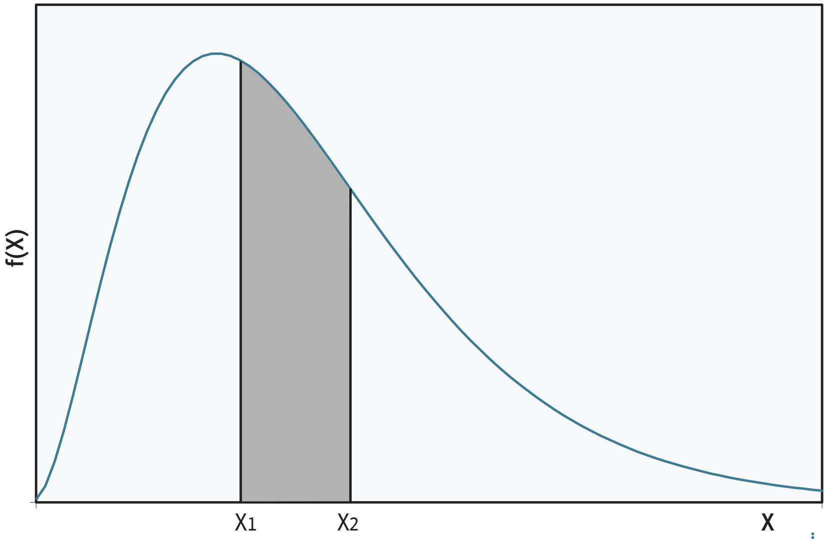
$$\sigma = [E(X - \mu)^2]^{1/2} = [\Sigma(X_i - \mu)^2 \times P(X_i)]^{1/2} = (1{,}594)^{1/2} = 1{,}263 \text{ παιδιά} \qquad (4.13)$$

4.5. **Κατανομές Πιθανοτήτων: Συνεχείς Μεταβλητές**

Εάν μια μεταβλητή είναι συνεχής, τότε δεν έχει νόημα να αναζητούμε την πιθανότητα να έχει μια συγκεκριμένη τιμή, αφού οι πιθανές τιμές είναι άπειρες. Αυτό που μπορούμε να ρωτήσουμε είναι: Ποια η πιθανότητα η μεταβλητή X να πάρει μια τιμή στο διάστημα X_1 έως X_2; Για παράδειγμα, είναι λογικό ένας ιατρός να θέλει να υπολογίσει την πιθανότητα ένα νεογέννητο βρέφος να έχει βάρος μεταξύ 3 και 3,5 κιλών (ή πάνω από τέσσερα κιλά κ.λπ.). Ενώ, στην ερώτηση "ποια είναι η πιθανότητα ένα βρέφος να γεννηθεί με βάρος 3,256 κιλά», η απάντηση εκτός από αφελής, είναι "μηδέν", λόγω της απειρίας των πιθανών τιμών του βάρους.

Επομένως, στις συνεχείς μεταβλητές δεν υπάρχουν απλές πιθανότητες, αλλά αυτό που γνωρίζουμε (ή προσπαθούμε να προσεγγίσουμε) είναι η μαθηματική συνάρτηση της κατανομής f(X) (Διάγραμμα 4.2). Η συνάρτηση μιας συνεχούς μεταβλητής ονομάζεται **συνάρτηση πυκνότητας πιθανότητας** (probability density function). Δηλαδή, είναι η συνάρτηση που εκφράζει την **πυκνότητα** (συχνότητα) εμφάνισης των παρατηρήσεων στα διάφορα διαστήματα τιμών.

Όσα αναφέραμε πριν σε ότι αφορά τις συναρτήσεις πιθανοτήτων για διακριτές μεταβλητές ισχύουν και για τις συναρτήσεις πιθανοτήτων για συνεχείς μεταβλητές, με τη διαφορά ότι, αντί του αθροιστή Σ, θα χρησιμοποιήσουμε την ολοκλήρωση. Έτσι, μια συνάρτηση f(X) για να αποτελεί συνάρτηση πυκνότητας πιθανότητας πρέπει να ικανοποιεί τη σχέση:



Διάγραμμα 4.2
Κατανομή Συνάρτησης Πιθανότητας f(X)

$$\int_{-\infty}^{+\infty} f(X)dx = 1 \tag{4.14}$$

Τα όρια -∞ και +∞ έχουν συμβολικό χαρακτήρα και αναφέρονται σε όλες τις δυνατές τιμές της συνεχούς μεταβλητής X. Είναι λογικό ότι, για συγκεκριμένο χαρακτηριστικό, η μεταβλητή X παίρνει τις αντίστοιχες τιμές στο πεδίο ορισμού της (π.χ. η μεταβλητή X = βάρος νεογέννητου βρέφους παίρνει τιμές στο διάστημα 1.500 - 4.500 γραμμάρια, η μεταβλητή X = βάρος συσκευασίας καφέ 50 γραμμαρίων στο διάστημα 46 - 54 γραμμάρια, κ.λπ.).

Η πιθανότητα η μεταβλητή X να πάρει μια τιμή στο διάστημα X_1 έως X_2 ισούται με:

$$P(X_1 < X < X_2) = \int_{X_1}^{X_2} f(X)dx \tag{4.15}$$

Η σχέση (4.15) ισχύει ανεξάρτητα από το εάν το διάστημα τιμών (X_1, X_2) είναι “κλειστό” ή “ανοικτό”, δηλαδή περιλαμβάνει ή όχι τις τιμές X_1 και X_2 με την αυστηρή μαθηματική έννοια.

Η αναμενόμενη τιμή της κατανομής της συνεχούς μεταβλητής X, με συνάρτηση πυκνότητας πιθανότητας f(X), ισούται με:

$$\mu = E(X) = \int_{-\infty}^{+\infty} X \times f(X)dx \tag{4.16}$$

και η διακύμανση της κατανομής πιθανότητας της X είναι:

$$\mathrm{Var}(X) = \sigma^2 = E(X - \mu)^2 = \int_{-\infty}^{+\infty} (X-\mu)^2 \times f(X)dx \qquad (4.17)$$

Στο επόμενο κεφάλαιο θα περιγράψουμε την πιο γνωστή από τις θεωρητικές κατανομές συνεχών μεταβλητών, την κανονική κατανομή. Τώρα, θα περιγράψουμε δύο θεωρητικές κατανομές ασυνεχών (διακριτών) μεταβλητών. Τη διωνυμική κατανομή και την κατανομή Poisson.

4.6. Η Διωνυμική Κατανομή

Η **διωνυμική κατανομή** (binomial distribution) είναι μια ασυνεχής κατανομή με πολλές εφαρμογές στην καθημερινή ζωή. Η εφαρμογή της βασίζεται στις εξής τέσσερις προϋποθέσεις:

α) Οι παρατηρήσεις ταξινομούνται σε δύο αμοιβαίως αποκλειόμενες κατηγορίες, που συνήθως ονομάζονται **επιτυχία** (ή εμφάνιση) και **αποτυχία** (ή μη εμφάνιση). Για παράδειγμα, ένας ψηφοφόρος να ψηφίζει το κόμμα Α ή να μην το ψηφίζει, ένας καταναλωτής αγοράζει το προϊόν Α ή δεν το αγοράζει, μία γυναίκα εργάζεται ή δεν εργάζεται κ.λπ.).

β) Η πιθανότητα μια παρατήρηση να ανήκει στην κατηγορία **επιτυχία** ισούται με p, και είναι σταθερή για όλες τις παρατηρήσεις. Έτσι, η πιθανότητα μια παρατήρηση να ανήκει στη συμπληρωματική κατηγορία **αποτυχία** ισούται με 1-p, και είναι σταθερή για όλες τις παρατηρήσεις.

γ) Η προηγούμενη προϋπόθεση (β) σημαίνει ότι οι παρατηρήσεις είτε προέρχονται από άπειρους πληθυσμούς χωρίς επανατοποθέτηση είτε από πεπερασμένους πληθυσμούς με επανατοποθέτηση. Επειδή η πιθανότητα επιτυχίας ή αποτυχίας πρέπει να είναι σταθερή, σημαίνει ότι εάν οι παρατηρήσεις προέρχονται από πληθυσμό με πεπερασμένο αριθμό μελών (π.χ. 50 άτομα γενικής συνέλευσης μετόχων, 40 σπουδαστές ενός τμήματος ΜΒΑ, κ.λπ.), μετά από κάθε δειγματοληψία, ο “κλήρος” επανατοποθετείται στην “κληρωτίδα”, έτσι ώστε να μην αλλοιώνεται το περιεχόμενο και ως εκ τούτου η πιθανότητα επιτυχίας. Για παράδειγμα, θέλετε να επιλέξετε μια πενταμελή επιτροπή από τους 50 μετόχους της Γενικής Συνέλευσης, για να εξετάσουν το θέμα της αύξησης του μετοχικού κεφαλαίου. Εάν είναι γνωστό ότι 20 μέτοχοι (δηλαδή το 40%) είναι υπέρ της αύξησης και 30 (60%) είναι κατά, τότε η πιθανότητα ο πρώτος που θα επιλεγεί τυχαία να είναι υπέρ είναι 20/50 = 40%. Εάν αυτός **δεν** συμμετέχει στην επόμενη τυχαία επιλογή, τότε η πιθανότητα για το δεύτερο που θα επιλεγεί τυχαία να είναι υπέρ είναι 19/49 = 38,8%. Με αυτόν τον τρόπο παραβιάζεται η προϋπόθεση της σταθερότητας της πιθανότητας, και γι’ αυτό ο “κλήρος” πρέπει να επανατοποθετηθεί. Όμως, στους μεγάλους (άπειρους) πληθυσμούς αυτό το πρόβλημα δεν υπάρχει, διότι η μεταβολή της πιθανότητας είναι μηδενική. Για παράδειγμα, τι επίδραση μπορεί να υπάρχει στην πιθανότητα να επιλεγεί τυχαία μια γυναίκα που να εργάζεται, όταν η πιθανότητα στην πρώτη επιλογή είναι, π.χ., (1.200.000/3.800.000), και στη δεύτερη (1.199.999/3.799.999);

δ) Τέλος, το αποτέλεσμα (επιτυχία ή αποτυχία) κάθε παρατήρησης πρέπει να είναι ανεξάρτητο από τα αποτελέσματα των άλλων παρατηρήσεων (κάτι που ισχύει εάν ικανοποιείται η προϋπόθεση της σταθερότητας των πιθανοτήτων).

Με βάση τα παραπάνω, η ασυνεχής μεταβλητή, που ακολουθεί τη διωνυμική κατανομή, είναι ο αριθμός των επιτυχιών σε δείγμα n παρατηρήσεων (επαναλήψεις του πειράματος). Για παράδειγμα, ποια η πιθανότητα σε δείγμα 9 μετόχων οι 5 να είναι υπέρ της αύξησης του μετοχικού κεφαλαίου, εάν γνωρίζουμε ότι το 40% των μετόχων είναι υπέρ της αύξησης; Ποια η πιθανότητα σε παραγγελία 50 μονάδων ενός συγκεκριμένου ανταλλακτικού, 2 να είναι ελαττωματικά εάν γνωρίζουμε ότι το ποσοστό παραγωγής ελαττωματικών μονάδων είναι 3%; κ.ο.κ.

Τα παραδείγματα είναι άπειρα και τα συναντάμε καθημερινά σε όλες τις δραστηριότητες. Έτσι, η διωνυμική κατανομή μας δίνει την πιθανότητα P(X) να εμφανιστεί το χαρακτηριστικό που εξετάζουμε (ελαττωμματικό προϊόν, μέτοχος υπέρ της αύξησης, ψηφοφόρος του κόμματος Α, κλπ.) X φορές σε δείγμα n παρατηρήσεων. Η μαθηματική σχέση που δίνει τις τιμές των P(X) προκύπτει ως εξής:

Εάν η επιτυχία εμφανιστεί X φορές σε δείγμα μεγέθους n (δηλαλή n επαναλήψεις του τυχαίου πειράματος), τότε το ερώτημα είναι με ποια σειρά θα εμφανιστούν οι επιτυχίες. Εάν εμφανιστεί τις πρώτες X φορές, με βάση τον κανόνα του πολλαπλασιασμού, η πιθανότητα είναι:

$$P(X) = p^X \times (1\text{-}p)^{n\text{-}X}$$

Μπορεί, όμως, να εμφανιστεί την 1η φορά και τις τελευταίες (X-1), τις πρώτες 2 και τις τελευταίες (X-2) κ.ο.κ. Έτσι, χρησιμοποιώντας τους κανόνες της συνδυαστικής ανάλυσης, έχουμε τους συνδυασμούς (combinations) της σειράς εμφάνισης των X επιτυχιών σε n επαναλήψεις (παρατηρήσεις). Δηλαδή τους συνδυασμούς των n ανά X που συμβολίζεται με C^n_X και ισούται με:

$$C^n_X = \frac{n!}{(n\text{-}X)! \times X!} = \frac{n \times (n\text{-}1) \times (n\text{-}2) \times \ldots \times 1}{(n\text{-}X) \times (n\text{-}X\text{-}1) \times \ldots \times 1 \times X \times (X\text{-}1) \times \ldots \times 1} \qquad (4.18)$$

Επομένως, το ενδεχόμενο της εμφάνισης της επιτυχίας X φορές αποτελείται από το σύνολο των C^n_X ενδεχομένων, και η πιθανότητα εμφάνισης της επιτυχίας X φορές σε n επαναλήψεις (ανεξάρτητα από τη σειρά εμφάνισης) ισούται με:

$$P(X) = \frac{n!}{(n\text{-}X)! \times X!} \times p^X \times (1\text{-}p)^{n\text{-}X} \qquad (4.19)$$

που αποτελεί και τον τελικό τύπο προσδιορισμού των πιθανοτήτων της διωνυμικής κατανομής. Ας δούμε ένα παράδειγμα. Ποια είναι η πιθανότητα σε μια παρέα τριών ατόμων ηλικίας 20-30 ετών οι δύο να είναι παντρεμένοι, όταν είναι γνωστό ότι το 40%

των ατόμων αυτής της ηλικιακής ομάδας είναι παντρεμένοι; Οι τιμές των δεδομένων του προβλήματος είναι, n=3, X=2, και p=0,4. Έτσι:

$$P(2) = \frac{3!}{(3\text{-}2)! \times 2!} \times 0{,}4^2 \times (1\text{-}0{,}4)^{3-2}$$

$$= \frac{3 \times 2 \times 1}{1 \times 2 \times 1} \times 0{,}4^2 \times 0{,}6^1 = 0{,}288 \text{ (ή 28,8\%)}$$

Στη διωνυμική κατανομή η μεταβλητή X (αριθμός εμφάνισης των επιτυχιών) παίρνει τιμές από 0 έως και n, και το άθροισμα όλων των P(X) ισούται με 1. Για το παραπάνω παράδειγμα έχουμε:

$$P(0)+P(1)+P(2)+P(3) = \frac{3!}{(3\text{-}0)! \times 0!} \times 0{,}4^0 \times (1\text{-}0{,}4)^{3-0}$$

$$+ \frac{3!}{(3\text{-}1)! \times 1!} \times 0{,}4^1 \times (1\text{-}0{,}4)^{3-1}$$

$$+ \frac{3!}{(3\text{-}2)! \times 2!} \times 0{,}4^2 \times (1\text{-}0{,}4)^{3-2}$$

$$+ \frac{3!}{(3\text{-}3)! \times 3!} \times 0{,}4^3 \times (1\text{-}0{,}4)^{3-3} =$$

$$= 0{,}216 + 0{,}432 + 0{,}288 + 0{,}064$$

$$= 1{,}000$$

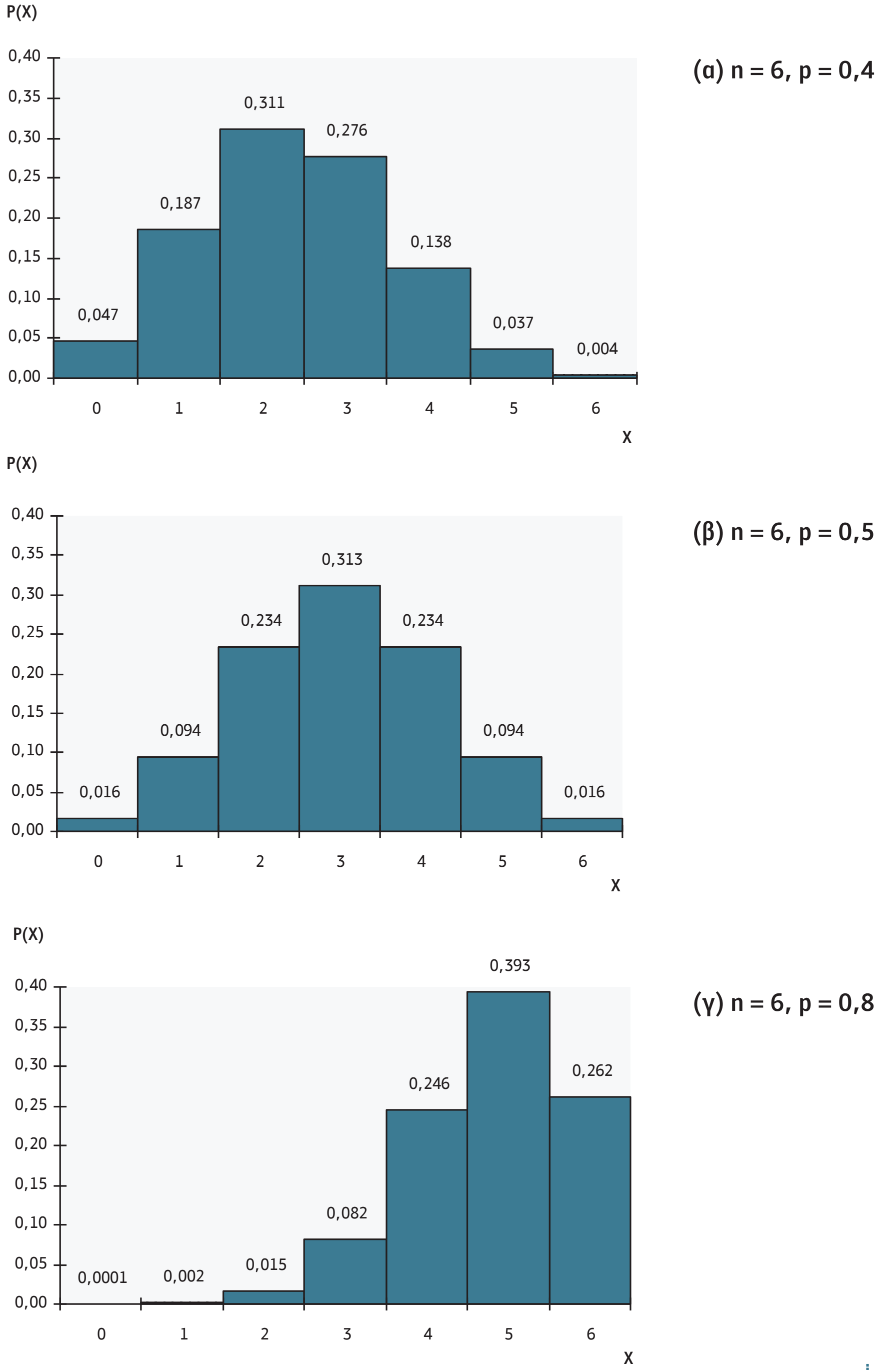
Ο μέσος της διωνυμικής κατανομής, προκύπτει σύμφωνα με τον τύπο (4.11), και ισούται με:

$$\mu = E(X) = n \times p \qquad (4.20)$$

Για το παραπάνω παράδειγμα έχουμε:

$$\mu = E(X) = n \times p = 3 \times 0{,}4 = 1{,}2$$

που σημαίνει ότι εάν συνεχίσουμε να επιλέγουμε στην τύχη τρία άτομα ηλικίας 20 – 30 ετών, κατά μέσο όρο, 1,2 άτομα (από τα τρία) θα είναι παντρεμένοι.



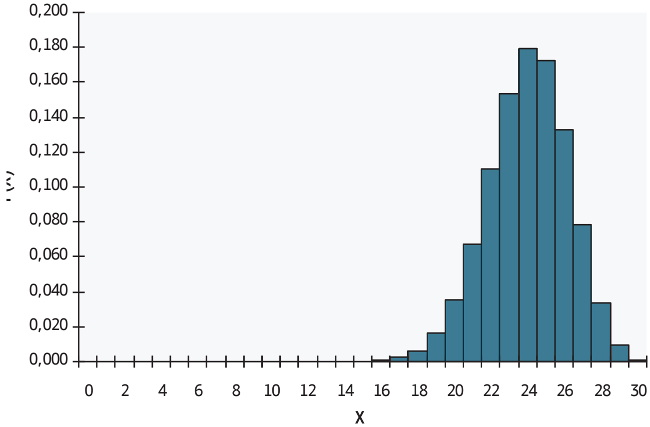
Διάγραμμα 4.3
Τρεις Διωνυμικές Κατανομές

Με ανάλογο τρόπο και σύμφωνα με τον τύπο (4.12), μπορούμε να αποδείξουμε ότι η διακύμανση της διωνυμικής κατανομής ισούται με:

$$Var(X) = \sigma^2 = E(X - \mu)^2 = n \times p \times (1\text{-}p) \quad (4.21)$$

και η τυπική απόκλιση με

$$\sigma = \sqrt{n \times p \times (1\text{-}p)} \quad (4.22)$$



Διάγραμμα 4.4
Διωνυμική Κατανομή για n = 30 και p = 0,8

Τέλος, όσον αφορά το σχήμα της διωνυμικής κατανομής, για μικρές τιμές του n, εξαρτάται κυρίως από την τιμή της πιθανότητας p. Σύμφωνα με το Διάγραμμα 4.3, όσο η τιμή της p πλησιάζει το 0,5 τόσο πιο συμμετρική γίνεται η διωνυμική κατανομή και αντίστροφα. Όμως, όσο αυξάνεται ο αριθμός των επαναλήψεων του πειράματος n (μέγεθος δείγματος) τόσο η διωνυμική κατανομή τείνει προς τη συμμετρία ανεξάρτητα από την τιμή της πιθανότητας p, όπως φαίνεται και στο Διάγραμμα 4.4.

4.7. **Κατανομή Poisson**

Η κατανομή Poisson είναι επίσης μια διακριτή κατανομή πιθανοτήτων με πολλές πρακτικές εφαρμογές. Αποτελεί προέκταση της διωνυμικής κατανομής σε περιπτώσεις που το μέγεθος του δείγματος (παρατηρήσεις ή επαναλήψεις του πειράματος) είναι πολύ μεγάλο και, σε πρακτικούς όρους, θα λέγαμε άπειρο. Συνήθως, χρησιμοποιείται για να περιγράψει καταστάσεις «ουρών» (queues), δηλαδή, άφιξη πελατών σε τράπεζες, ή πολυκαταστήματα, αριθμός κλήσεων σε τηλεφωνικά κέντρα, αριθμός διερχόμενων οχημάτων από συγκεκριμένη διασταύρωση κλπ. Έτσι, ισχύουν και εδώ οι προϋποθέσεις εφαρμογής της διωνυμικής κατανομής, αλλά με τη βασική διαφορά ότι η πιθανότητα επιτυχίας (εμφάνισης) σε μια παρατήρηση είναι πολύ μικρή, και ο αριθμός των παρατηρήσεων άπειρος.

Ας δούμε ένα παράδειγμα. Μία τράπεζα αναδιοργανώνει το κεντρικό της κατάστημα με αντικείμενο την καλύτερη εξυπηρέτηση των πελατών. Η αύξηση του κύκλου εργασιών έχει οδηγήσει σε μεγάλη προσέλευση πελατών και αύξηση του χρόνου αναμονής τους. Το πρόβλημα που αντιμετωπίζει η Τράπεζα είναι διπλό. Από τη μια την ενδιαφέρει το είδος του πελάτη, δηλαδή εταιρία (corporate banking) ή ιδιώτης (retail banking), και από την άλλη η συχνότητα άφιξης των πελατών κυρίως τις ώρες αιχμής. Το ποσοστό των πελατών που αφορούν εταιρίες είναι 30%, και από την ηλεκτρονική συσκευή ασφαλείας που παρακολουθεί τις αφίξεις πελατών, έχει εκτιμηθεί ότι τις ώρες αιχμής εισέρχονται στο κατάστημα, κατά μέσο όρο, 5 πελάτες ανά λεπτό. Η τράπεζα προγραμματίζει τη δημιουργία διαφορετικών τμημάτων εξυπηρέτησης των πελατών ανάλογα με την κατηγορία που ανήκουν (εταιρίες ή ιδιώτες) και, παράλληλα, την αύξηση του προσωπικού και των θέσεων εργασίας για την εξυπηρέτηση μεγαλύτερου αριθμού πελατών.

Προσέξτε τώρα το είδος των στατιστικών προβλημάτων που αντιμετωπίζουν οι υπεύθυνοι της Τράπεζας. Ποια η πιθανότητα στο χρονικό διάστημα **11:00 με 11:01** να φθάσουν στην τράπεζα 10 πελάτες από τους οποίους οι μισοί να είναι ιδιώτες και οι άλλοι μισοί από εταιρίες; Η δεύτερη ερώτηση αφορά τη διωνυμική κατανομή. Γνωρίζουμε την πιθανότητα ένας πελάτης να προέρχεται από εταιρία (0,30), η άφιξη κάθε πελάτη είναι ανεξάρτητη από τις αφίξεις των άλλων πελατών, και το μέγεθος του δείγματος είναι σχετικά μικρό ($n = 10$).

Ας εξετάσουμε τώρα το πρώτο ερώτημα. Μας ενδιαφέρει η πιθανότητα να φθάσουν στην Τράπεζα 10 πελάτες σε μία συγκεκριμένη χρονική στιγμή (11:00 με 11:01). Προφανώς η πιθανότητα είναι πολύ μικρή, σχεδόν μηδαμινή, διότι δεν ρωτάμε γενικά ποια η πιθανότητα να φθάσουν 10 πελάτες στη διάρκεια ενός λεπτού, αλλά στη διάρκεια ενός συγκεκριμένου λεπτού μιας συγκεκριμένης ημέρας. Ταυτόχρονα τα χρονικά διαστήματα του ενός λεπτού που αποτελείται η εργάσιμη ημέρα των 8 ωρών είναι μεγάλος ($8 \times 60 = 480$). Και εάν η ανάλυση καλύψει μεγάλο χρονικό διάστημα (π.χ. έξι μηνών), τότε οι παρατηρήσεις θα ανέλθουν σε $120 \times 8 \times 60 = 57.600$, με την υπόθεση ότι ένας μήνας έχει 20 εργάσιμες ημέρες. Έτσι, σύμφωνα με το συγκεκριμένο παράδειγμα, βλέπουμε ότι, ενώ οι συνθήκες είναι ανάλογες με εκείνες που εφαρμόζουμε τη διωνυμική κατανομή (σταθερή πιθανότητα άφιξης ενός πελάτη, ανεξαρτησία μεταξύ των αφίξεων κ.λπ.), υπάρχουν δύο ουσιώδεις διαφορές:

α) η πιθανότητα άφιξης ενός πελάτη (εμφάνιση του χαρακτηριστικού) σε ένα συγκεκριμένο χρονικό διάστημα (μια παρατήρηση) είναι πάρα πολύ μικρή, σχεδόν μηδενική, και

β) ο αριθμός των χρονικών διαστημάτων που εξετάζουμε (μέγεθος δείγματος ή αριθμός επανάληψης του πειράματος) είναι πολύ μεγάλος (σχεδόν άπειρος).

Είναι προφανές ότι οι υπεύθυνοι της Τράπεζας δεν ενδιαφέρονται για το συγκεκριμένο χρονικό διάστημα του ενός λεπτού (11:00 με 11:01 της συγκεκριμένης ημέρας), αλλά γενικά για την κατανομή του αριθμού των πελατών που φθάνουν στην τράπεζα στη διάρκεια ενός λεπτού. Η συχνότητα άφιξης των πελατών, σε συνδυασμό με τον μέσο χρόνο που απαιτείται για την εξυπηρέτηση ενός πελάτη, θα οδηγήσει τους υπεύθυνους να αποφασίσουν πόσα νέα σημεία εξυπηρέτησης χρειάζονται.

Συμπερασματικά, η κατανομή Poisson χρησιμοποιείται αντί της διωνυμικής κατανομής στις περιπτώσεις που το p είναι πολύ μικρό (μικρότερο από 0,01) και το n πολύ μεγάλο (π.χ. μεγαλύτερο από 100), έτσι ώστε το γινόμενο $n \times p$ να είναι ένας σχετικά μικρός αριθμός (μικρότερος του 10). Οι τιμές της διακριτής μεταβλητής που ακολουθεί την κατανομή Poisson παίρνει τις τιμές 1, 2, 3, …, n, και επειδή το n είναι πολύ μεγάλο (θεωρητικά άπειρο), και η X παίρνει (θεωρητικά) τιμές μέχρι το άπειρο.

Τώρα είμαστε σε θέση να περιγράψουμε τα ερωτήματα στα οποία καλείται να απαντήσει η κατανομή Poisson. Ποια η πιθανότητα να φθάσουν στην τράπεζα κατά τη διάρκεια ενός λεπτού 0, 1, 2, 3, …, ∞ πελάτες; Ποια η πιθανότητα ένα τηλεφωνικό κέντρο να δεχτεί κατά τη διάρκεια 30 δευτερολέπτων 0, 1, 2, 3, …, ∞ τηλεφωνήματα; Ποια η πιθανότητα να περάσουν από συγκεκριμένη διασταύρωση στη διάρκεια 90 δευτερολέπτων (όσο διαρκεί το κόκκινο φανάρι) 0, 1, 2, …, ∞ αυτοκίνητα; Μην παρασύρεστε από το άνω όριο των δυνατών τιμών της X που είναι το άπειρο. Μόνο θεωρητική σημασία έχει. Όπως θα δούμε στη συνέχεια, οι πιθανότητες P(X) φθίνουν πολύ γρήγορα και τείνουν προς το μηδέν για σχετικά μικρές τιμές της X (μικρότερες του 50).

Για να εφαρμόσουμε την κατανομή Poisson το μόνο που χρειαζόμαστε είναι η τιμή του μέσου όρου που συμβολίζεται με λ (το αντίστοιχο του $n \times p$ της διωνυμικής κατανομής). Η πιθανότητα να έχουμε X επιτυχίες (αφίξεις) δίνεται από τη σχέση:

$$P(X) = \frac{e^{-\lambda} \times \lambda^{\chi}}{X!} \tag{4.23}$$

όπου

P(X) = πιθανότητα X επιτυχιών
λ = μέσος αναμενόμενος αριθμός επιτυχιών
e = η βάση των νεπέρειων λογαρίθμων (≈ 2,71828)
X = αριθμός επιτυχιών ανά παρατήρηση

Για παράδειγμα, ένα υποκατάστημα μιας Τράπεζας μετά από παρατηρήσεις τεσσάρων εβδομάδων εκτίμησε ότι φθάνουν στην Τράπεζα, κατά μέσο όρο, 3 πελάτες το λεπτό. Ο υπεύθυνος του τμήματος εξυπηρέτησης πελατών ρωτά: ποια η πιθανότητα να φθάσουν μέσα σε ένα λεπτό 5 πελάτες, και ποια η πιθανότητα να φθάσουν πάνω από 5 πελάτες; Για την πρώτη ερώτηση έχουμε:

$$P(5) = \frac{e^{-\lambda} \times \lambda^{\chi}}{X!} = \frac{2{,}71828^{-3} \times 3^5}{5!}$$

$$= 0{,}1008 \ \ (\text{ή } 10{,}08\%)$$

Για τη δεύτερη ερώτηση έχουμε:

$$\begin{aligned} P(X > 5) &= 1 - P(X \leq 5) \\ &= 1 - P(0) - P(1) - P(2) - P(3) - P(4) - P(5) \\ &= 1 - 0{,}0498 - 0{,}1494 - 0{,}2240 - 0{,}2240 - 0{,}1680 - 0{,}1008 \\ &= 1 - 0{,}9161 \\ &= 0{,}0839 \ (8{,}39\%) \end{aligned}$$

Από τα παραπάνω αποτελέσματα προκύπτει ότι, με μονάδα μέτρησης του χρόνου το λεπτό, 10% του χρόνου λειτουργίας του υποκαταστήματος θα φθάνουν 5 πελάτες και, περίπου, 8,4% του χρόνου λειτουργίας θα φθάνουν πάνω από 5 πελάτες. Το Διάγραμμα 4.5 δείχνει τα σχήματα δύο κατανομών Poisson με μέσο (λ) 3 και 8, αντίστοιχα. Από το διάγραμμα προκύπτει ότι όσο αυξάνει η τιμή του λ τόσο η κατανομή γίνεται συμμετρική. Προσέξτε, επίσης, πόσο γρήγορα φθίνουν και τείνουν στο μηδέν οι τιμές των πιθανοτήτων P(X). Η διακύμανση της κατανομής Poisson ισούται με λ. Επομένως έχουμε:

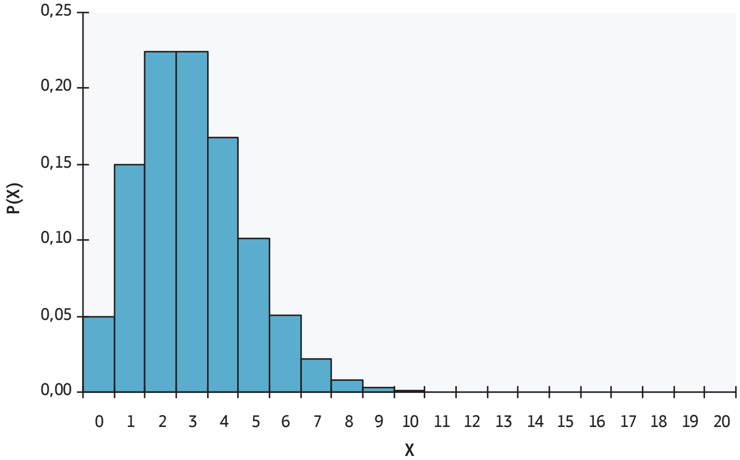
$$Var(X) = \sigma^2 = E(X - \mu)^2 = \lambda \tag{4.24}$$

και

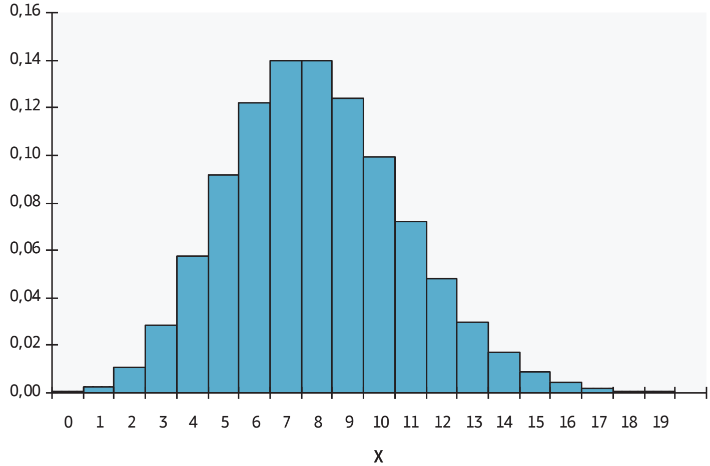
$$\sigma = \sqrt{\lambda} \tag{4.25}$$

Όπως ήδη αναφέραμε η διωνυμική κατανομή τείνει προς την κατανομή Poisson όταν το p είναι μικρό και το n είναι μεγάλο. Επειδή η διωνυμική προσεγγίζει πολύ γρήγορα την Poisson, μπορούμε να χρησιμοποιήσουμε την Poisson ως προσέγγιση των πιθανοτήτων της διωνυμικής για σχετικά μικρές τιμές του n.Ένας πρακτικός κανόνας είναι να χρησιμοποιούμε την κατανομή Poisson αντί της διωνυμικής για $n \geq 20$ και $p \leq 0{,}05$.

Τέλος, θα πρέπει να τονίσουμε ότι τόσο η διωνυμική όσο και η Poisson τείνουν γρήγορα προς τη συμμετρία όσο το μέγεθος του δείγματος n και του λ αυξάνουν (Διαγράμματα 4.4 και 4.5). Αυτή η τάση προς τη συμμετρία σημαίνει ότι και οι δύο κατανομές τείνουν προς την **κανονική κατανομή**, που θα παρουσιάσουμε στο επόμενο κεφάλαιο. Η κανονική κατανομή είναι ευκολότερη στη χρήση της και δίνει με πολύ καλή προσέγγιση τις τιμές των πιθανοτήτων τόσο της διωνυμικής όσο και της Poisson. Στο Κεφάλαιο 15 δίνονται οι τιμές των πιθανοτήτων της διωνυμικής κατανομής για επιλεγμένες τιμές των p και n. Επίσης, δίνονται οι τιμές των πιθανοτήτων της κατανομής Poisson για επιλεγμένες τιμές του μέσου λ και περιγράφεται ο τρόπος υπολογισμού των πιθανοτήτων των δύο κατανομών με τη χρήση του προγράμματος Excel.



(α) λ = 3



(β) λ = 8

Διάγραμμα 4.5
Δύο Κατανομές Poisson