



Ακαδημία Εμπορικού Ναυτικού Ασπροπύργου
Σχολή Μηχανικών

Πτυχιακή εργασία
Παρουσίαση των βασικών κατανομών πιθανότητας



Σπουδαστής: **Νάκης Αντώνιος** (AM 7592)

Επιβλέπων Καθηγητής: **Στέφανος Ι. Καρναβάς**, Μαθηματικός (M.Ed.),
Επίκουρος Καθηγητής

Ακαδημαϊκό έτος: **2020 – 2021**

Ημερομηνία ανάθεσης: 10.06. 2019
 Ημερομηνία κατάθεσης:06.2021
 Ημερομηνία εξέτασης: / 0.. / 2021

Α/Α	Ονοματεπώνυμο	Χαρακτηρισμός	Υπογραφή
1	Στέφανος Ι. Καρναβάς Μαθηματικός (Μ.Εδ.) Επίκουρος Καθηγητής	Διακεκριμένη	
2	Αντωνάτος Χαράλαμπος Ναυπηγός Μηχανολόγος Μηχανικός Καθηγητής		
3	Αθανασία Χρηστίδου Ηλεκτρολόγος Μηχανικός Επίκουρη Καθηγήτρια		
Τελικός χαρακτηρισμός			

Περιεχόμενα

1. Δοκιμές Bernoulli και διωνυμική κατανομή.....	3
1.1 Κατανομή Bernoulli	3
1.2 Ρίψη ιδανικού ζαριού μία φορά	4
1.3 Διωνυμική κατανομή.....	4
1.4 Συνάρτηση πιθανότητας της διωνυμικής κατανομής.....	4
1.5 Μέση τιμή και διακύμανση της διωνυμικής κατανομής.....	6
2. Γεωμετρική και αρνητική διωνυμική κατανομή.....	7
2.1 Γεωμετρική κατανομή.....	7
2.2 Συνάρτηση πιθανότητας της γεωμετρικής κατανομής.....	7
2.3 Μέση τιμή και διακύμανση της γεωμετρικής κατανομής.....	8
3. Αρνητική διωνυμική κατανομή	9
3.1 Συνάρτηση πιθανότητας της αρνητικής διωνυμικής κατανομής	10
3.2 Μέση τιμή και διακύμανση της αρνητικής διωνυμικής κατανομής	11
4. Η Υπεργεωμετρική κατανομή	12
4.1 Ορισμός υπεργεωμετρικής κατανομής.....	13
4.2 Συνάρτηση πιθανότητας της υπεργεωμετρικής κατανομής	13
4.3 Μέση τιμή και διακύμανση της υπεργεωμετρικής κατανομής	15
5. Κατανομή Poisson	15
5.1 Όριο της διωνυμικής κατανομής $b(n,p)$ για μεγάλα n	16
5.2 Ορισμός κατανομής Poisson	16
5.3 Η κατανομή Poisson ως όριο της διωνυμικής κατανομής	18

1. Δοκιμές Bernoulli και διωνυμική κατανομή

Οι **δοκιμές Bernoulli** των οποίων η ονομασία οφείλεται στο σουηδό μαθηματικό **James Bernoulli** (1654-1705) οδηγούν στον ορισμό της απλούστερης ίσως διακριτής τυχαίας μεταβλητής. Με τον όρο δοκιμή Bernoulli αναφερόμαστε σε ένα πείραμα με δύο δυνατά αποτελέσματα. Το ένα από τα δύο αποτελέσματα θα ονομάζεται **επιτυχία** (συμβολικά: ε) ενώ το άλλο **αποτυχία** (συμβολικά: α). Το πείραμα της ρίψης ενός νομίσματος αποτελεί τη βασικότερη περίπτωση μίας δοκιμής Bernoulli. Αν μας ενδιαφέρει η ένδειξη "κεφαλή" θα βαπτίσουμε ως επιτυχία ε το αποτέλεσμα αυτό και θα ορίσουμε ως αποτυχία α την ένδειξη "γράμματα". Αντίστοιχα, το τυχαίο "πείραμα" της γέννησης ενός παιδιού αποτελεί επίσης δοκιμή Bernoulli με 2 δυνατά αποτελέσματα αγόρι ή κορίτσι.

Τέλος, η εξέταση κατά πόσο ένα αντικείμενο μίας γραμμής παραγωγής είναι ελαττωματικό ή όχι είναι δοκιμή Bernoulli. Αν μας ενδιαφέρει να διατυπώσουμε πιθανοθεωρητικά συμπεράσματα για τα ελαττωματικά αντικείμενα που παράγονται, φαίνεται λογικό να ορίσουμε ως επιτυχία ε την εμφάνιση ελαττωματικού αντικειμένου και αποτυχία την εμφάνιση μη ελαττωματικού. Σε πολλές περιπτώσεις, μπορεί να έχουμε πειράματα με περισσότερα από δύο δυνατά αποτελέσματα και να φτάνουμε σε δοκιμές Bernoulli με κατάλληλη ομαδοποίηση των αποτελεσμάτων τους. Έτσι, κατά τη ρίψη ενός ζαριού μία φορά, αν μας ενδιαφέρει το αποτέλεσμα 6, μπορούμε να δημιουργήσουμε δοκιμές Bernoulli αν ορίσουμε ως επιτυχία ε την εμφάνιση της ένδειξης 6 και ως αποτυχία την εμφάνιση οποιασδήποτε από τις άλλες 5 ενδείξεις 1,2,3,4,5. Αν μας ενδιαφέρει η εμφάνιση άρτιου αποτελέσματος θα ορίσουμε ως επιτυχία την εμφάνιση μιας εκ των ενδείξεων 2,4,6 και ως αποτυχία την εμφάνιση 1,3,5. Σύμφωνα με όσα αναφέρθηκαν προηγουμένως ο δειγματικός χώρος του πειράματος που παράγει μία δοκιμή Bernoulli μπορεί να παρασταθεί ως: $\Omega = \{\varepsilon, \alpha\}$. Αν συμβολίζουμε με p, q τις πιθανότητες εμφάνισης των στοιχειωδών ενδεχόμενων $\{\varepsilon\}, \{\alpha\}$ αντίστοιχα είναι φανερό ότι θα έχουμε:

$$p+q=1, \quad 0 \leq p \leq 1, \quad 0 \leq q \leq 1$$

1.1 Κατανομή Bernoulli

Έστω χ ο αριθμός επιτυχιών σε μία δοκιμή Bernoulli με πιθανότητα επιτυχίας p και (αποτυχίας $q = 1 - p$). Η κατανομή της δίτιμης (μηδέν- ένα) τυχαίας μεταβλητής X καλείται **κατανομή Bernoulli** με παράμετρο p . Το σύνολο τιμών της τυχαίας μεταβλητής X είναι το $R_x = \{0,1\}$, ενώ για τη συνάρτηση πιθανότητας f έχουμε:

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

Σημειώνουμε πως ο τύπος της f μπορεί να γραφεί εναλλακτικά ως εξής:

$$f(x) = p^x q^{1-x}, \quad x = 0,1$$

Για τη συνάρτηση κατανομής F έχουμε

$$F(t) = \begin{cases} 0 & t < 0 \\ q & 0 \leq t \leq 1 \\ 1 & t \geq 1 \end{cases}$$

Η μέση τιμή $\mu = E(X)$ της X είναι ίση με

$$\mu = \sum_{x=0}^1 x f(x) = 0 \cdot f(0) + 1 \cdot f(1) = f(1) = p$$

Ενώ η ροπή δεύτερης τάξης υπολογίζεται ως ακολούθως

$$\mu_2 = E(X^2) = \sum_{x=0}^1 X^2 \cdot f(x) = 0^2 \cdot f(0) + 1^2 \cdot f(1) = f(1) = p$$

Επομένως, η διακύμανση $\sigma^2 = V(X)$ θα δίνεται από τον τύπο
 $\sigma^2 = V(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1 - p) = pq$

1.2 Ρίψη ιδανικού ζαριού μία φορά

Κατά τη ρίψη ενός ζαριού μία φορά ορίζουμε ως επιτυχία την εμφάνιση ένδειξης μεγαλύτερης του 4. Να δοθεί η συνάρτηση πιθανότητας, η συνάρτηση κατανομής, η μέση τιμή και η διακύμανση της αντίστοιχης κατανομής Bernoulli αν X είναι ο αριθμός των φόρων που εμφανίστηκε ένδειξη μεγαλύτερη του 4 (στη μία ρίψη) έχουμε

$$R_x = \{0,1\}$$

Και

$$p = P(X=1) = P(\text{εμφανίζεται 5 ή 6}) = \frac{2}{6} = \frac{1}{3}$$

Επομένως,

$$f(x) = \left(\frac{1}{3}\right)^x \left(1 - \frac{1}{3}\right)^{1-x} = \frac{2^{1-x}}{3}, x = 0,1$$

$$F(t) = \begin{cases} 0 & , t < 0 \\ 2/3 & , 0 \leq t \leq 1 \\ 1 & , t \geq 1 \end{cases}$$

$$E(X) = \left(\frac{1}{3}\right) \quad V(X) = \frac{1}{3} \left(1 - \frac{1}{3}\right) = \frac{2}{9}$$

Παρότι η ιδέα της δοκιμής Bernoulli είναι εξαιρετικά απλή αν θεωρήσουμε πολλές επαναλήψεις του ίδιου πειράματος μπορούν να προκύψουν ενδιαφέροντα και σε αρκετές περιπτώσεις δύσκολα να απαντηθούν ερωτήματα. Για τη διεξαγωγή ενός πειράματος Bernoulli απαραίτητη προϋπόθεση είναι η ύπαρξη αυστηρά παρόμοιων συνθηκών σε κάθε επανάληψη ώστε να υπάρξει ουδέτερο αποτέλεσμα. Στη συνέχεια θα θεωρήσουμε ότι γίνονται n ανεξάρτητες επαναλήψεις μιας δοκιμής Bernoulli και θα μελετήσουμε το πλήθος X των επιτυχιών που εμφανίζονται σε αυτές.

1.3 Διωνυμική κατανομή

Έστω X ο αριθμός των επιτυχιών σε μία ακολουθία n (ανεξάρτητων) δοκιμών Bernoulli με σταθερή πιθανότητα επιτυχίας p και αποτυχίας $q = 1 - p$ σε όλες τις δοκιμές. Η κατανομή της τυχαίας μεταβλητής X καλείται **διωνυμική κατανομή** με παραμέτρους n και p και συμβολίζεται με $b(n,p)$. Από τον ορισμό είναι φανερό ότι το σύνολο τιμών της τυχαίας μεταβλητής X είναι το

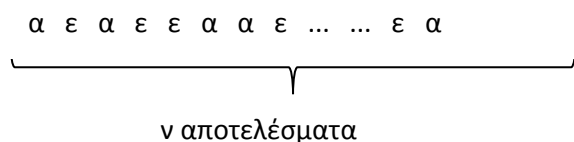
$$R_x = \{0, 1, \dots, n\}$$

1.4 Συνάρτηση πιθανότητας της διωνυμικής κατανομής

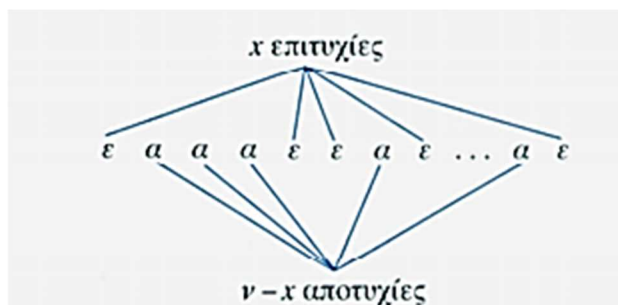
Η συνάρτηση πιθανότητας της διωνυμικής κατανομής με παραμέτρους n και p δίνεται από τον τύπο

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

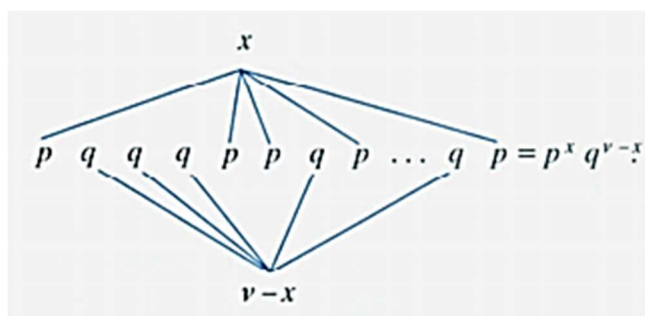
Τα στοιχεία του δειγματικού χώρου Ω του πειράματος (εκτέλεση μιας ολόκληρης σειράς δοκιμών) είναι ν-άδες από επιτυχίες (ε) και αποτυχίες (α) δηλαδή έχουν τη μορφή



Έστω τώρα έχει ένας συγκεκριμένος αριθμός του συνόλου τιμών Rx δηλαδή $0 \leq x \leq \nu$. Το ενδεχόμενο $\{X=x\}$ αποτελείται από εκείνα τα στοιχεία του Ω τα οποία περιέχουν x αποτελέσματα ίσα με ε και $\nu-x$ για αποτελέσματα ίσα με α δηλαδή είναι της μορφής



Λόγω της ανεξαρτησίας των δοκιμών, η πιθανότητα εμφάνισης ενός τέτοιου αποτελέσματος (με τη συγκεκριμένη σειρά διαδοχής επιτυχιών-αποτυχιών) είναι ίση με



Όμως υπάρχουν $\binom{\nu}{x}$ τέτοιες ακολουθίες οι οποίες περιέχουν x αποτελέσματα ίσα με ε και $\nu-x$ αποτελέσματα ίσα με α (αρκεί να διαλέξουμε από τις ν διαθέσιμες θέσεις, τις x όπου θα τοποθετηθούν οι επιτυχίες ε και να συμπληρώσουμε τις υπόλοιπες $\nu-x$ θέσεις με αποτυχίες α). Επομένως

$$f(x) = P(X=x) = \binom{\nu}{x} p^x q^{\nu-x}$$

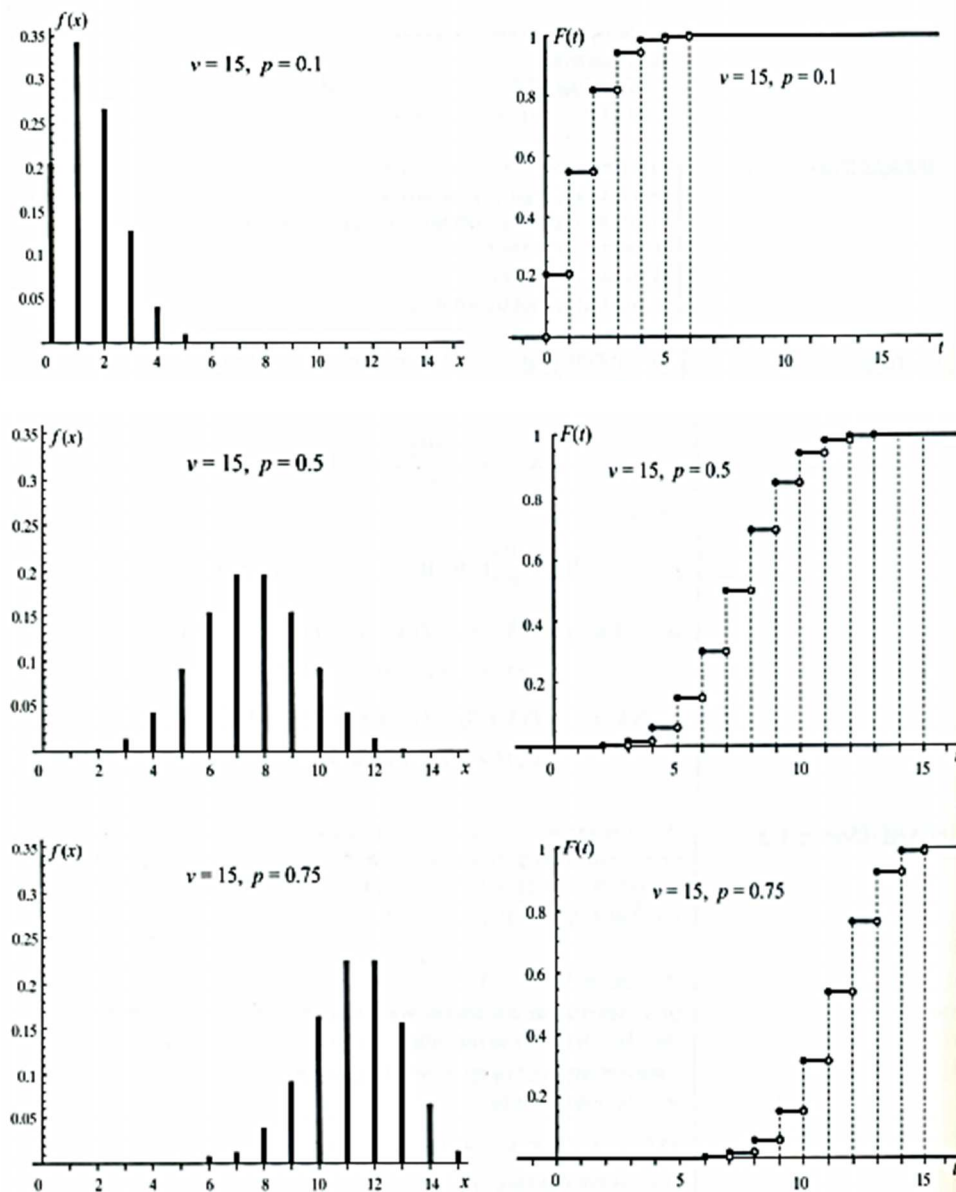
Η διωνυμική κατανομή $b(\nu, p)$ στην ειδική περίπτωση $\nu=1$ ανάγεται στην κατανομή Bernoulli όπως φαίνεται αμέσως και από το αποτέλεσμα της παραπάνω πρότασης αν θέσουμε $\nu=1$. Αν ορίσουμε τις επόμενες ν τυχαίες μεταβλητές

$$X_i = \begin{cases} 0 \\ 1 \end{cases}$$

$X=1$ αν στη δοκιμή i εμφανίστηκε ε και $X=0$ αν στη δοκιμή i εμφανίστηκε α (οι οποίες ακολουθούν την κατανομή Bernoulli με παράμετρο p) τότε το άθροισμα

$$X_1 + X_2 + \dots + X_\nu$$

Θα δίνει τον αριθμό επιτυχιών στις n δοκιμές, δηλαδή θα ακολουθεί τη διωνυμική κατανομή. Στο παρακάτω σχήμα δίνεται η συνάρτηση πιθανότητας και η συνάρτηση κατανομής της $b(n,p)$ για $n=15$ και $p=0.1, 0.5, 0.75$



1.5 Μέση τιμή και διακύμανση της διωνυμικής κατανομής

Η μέση τιμή και η διακύμανση της διωνυμικής κατανομής με παραμέτρους n και p δίνεται από τους τύπους

$$\mu = E(X) = np, \quad \sigma^2 = V(X) = npq$$

Ο πρώτος τύπος έχει ήδη αποδειχθεί. Για το δεύτερο, αρκεί να παρατηρήσουμε ότι $V(X) = E(X^2) - \mu^2 = E[X(X-1) + X] - \mu^2 = E[X(X-1)] + E(X) - \mu^2 = \mu_2 + \mu - \mu^2$. Και να αντικαταστήσουμε τα

$$\mu_{(2)} = n(n-1)p^2, \quad \mu = np$$

2. Γεωμετρική και αρνητική διωνυμική κατανομή

Οι δύο κατανομές που έχουν σχέση με το πλήθος των δοκιμών Bernoulli που πρέπει να εκτελεστούν **μέχρι να παρατηρηθεί ορισμένος αριθμός επιτυχιών**. Για το λόγο αυτό αναφέρονται συνήθως ως **κατανομές του χρόνου αναμονής** (στην ερμηνεία αυτή το ρόλο της χρονικής μονάδας τον παίζει η μία δοκιμή).

2.1 Γεωμετρική κατανομή

Θεωρούμε μία ακολουθία ανεξάρτητων δοκιμών Bernoulli με πιθανότητα επιτυχίας p (και αποτυχίας $q=1-p$) σταθερή για όλες τις δοκιμές. Έστω x ο αριθμός των δοκιμών μέχρι την εμφάνιση της πρώτης επιτυχίας. Η κατανομή της τυχαίας μεταβλητής X καλείται **γεωμετρική κατανομή** με παράμετρο p και συμβολίζεται με $G(p)$. Το ενδεχόμενο $\{X=x\}$, $1 \leq x$ σημαίνει ότι στις δοκιμές $1, 2, \dots, x-1$ εμφανίστηκε επιτυχία α και στην επόμενη δοκιμή εμφανίστηκε επιτυχία ε . Επομένως, το ενδεχόμενο αυτό φέρει ένα και μοναδικό αποτέλεσμα (στοιχείο του δειγματικού χώρου του πειράματος), το

$$\underbrace{\alpha \alpha \dots \alpha}_{x-1 \text{ φορές}} \varepsilon$$

Αφού οι δοκιμές είναι ανεξάρτητες, η πιθανότητα εμφάνισης του αποτελέσματος αυτού θα είναι:

$$\underbrace{q \ q \ \dots \ q}_{x-1 \text{ φορές}} \ , \quad p = q^{x-1} p$$

2.2 Συνάρτηση πιθανότητας της γεωμετρικής κατανομής

Η συνάρτηση πιθανότητας της γεωμετρικής κατανομής $G(x)$ θα ισούται με

$$f(x) = P(X=x) = q^{x-1} p \ , \ x = 1, 2, \dots$$

Ο τυπικός έλεγχος ότι η συνάρτηση f είναι πράγματι συνάρτηση πιθανότητας μιας διακριτής τυχαίας μεταβλητής είναι άμεσος αφού $f(x) > 0$ ή $f(x) = 0$ για όλα τα $x \in \mathbb{R}_x = \{1, 2, \dots\}$, ενώ ισχύει

$$\sum_{x \in \mathbb{R}_x} f(x) = \sum_{x=1}^{\infty} q^{x-1} p = p \sum_{x=1}^{\infty} q^{x-1} = p \frac{1}{1-q} = p \frac{1}{p} = 1$$

Για τη συνάρτηση κατανομής της $G(p)$ μπορούμε να γράψουμε

$$F(t) = \begin{cases} 0, & t < 1 \\ \sum_{x=1}^{[t]} f(x), & t \geq 1 \end{cases}$$

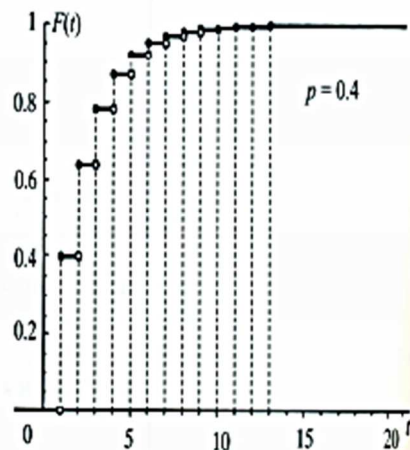
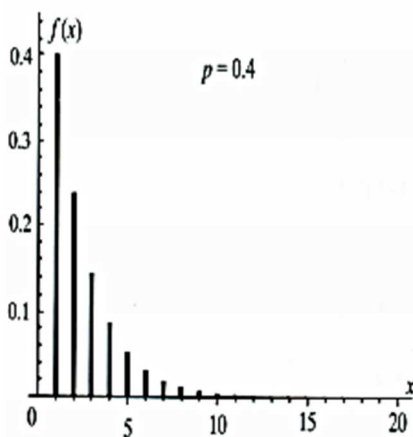
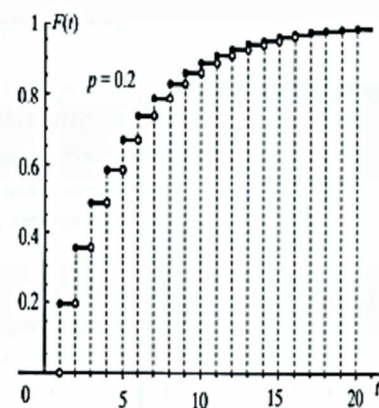
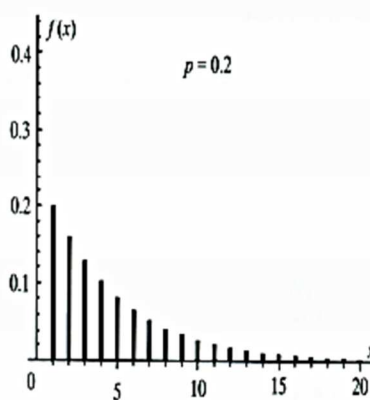
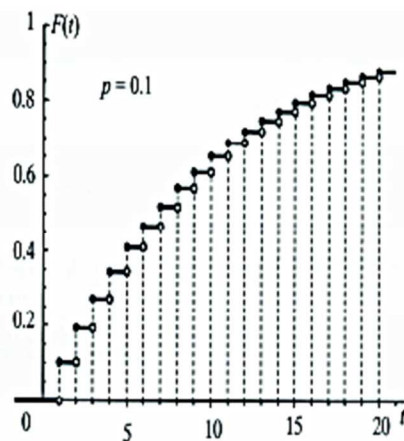
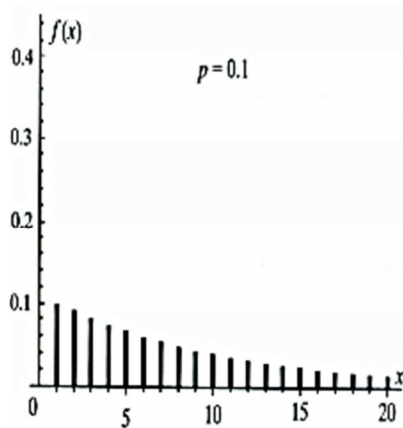
Και παρατηρώντας ότι, για $[t]=k$ έχουμε

$$\sum_{x=1}^k f(x) = p \sum_{x=1}^k q^{x-1} = p(1 + q + \dots + q^{k-1}) = p \frac{1-q^k}{1-q} = p \frac{1-q^k}{p} = 1 - q^k$$

Έτσι, καταλήγουμε στον εξής τύπο

$$F(t) = \begin{cases} 0, & t < 1 \\ 1 - q^{[t]}, & t \geq 1 \end{cases}$$

Στα παρακάτω σχήματα δίνεται η συνάρτηση πιθανότητας και η συνάρτηση γεωμετρικής κατανομής της $G(p)$ για $p=0.1$, $p=0.2$ και $p=0.4$ αντίστοιχα



2.3 Μέση τιμή και διακύμανση της γεωμετρικής κατανομής

Αν η τυχαία μεταβλητή p ακολουθεί τη γεωμετρική κατανομή με παράμετρο τότε

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

Παρατηρούμε αρχικά ότι η μέση τιμή $E(X)$ και η παραγοντική ροπή δεύτερης τάξης

$$\mu_2 = E[(X)_2] = E[X(X-1)]$$

της X είναι ίσες με

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

και

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

Αντίστοιχα. Ο υπολογισμός των δύο αθροισμάτων που εμφανίστηκαν στο δεξιό μέλος των παραπάνω τύπων μπορεί να γίνει αν παραγωγίσουμε δύο φορές τον γνωστό τύπο

$$\sum_{x=0}^{\infty} t^x = 1 + t + t^2 + t^3 + \dots + t^{\nu} + \dots = \frac{1}{1-t}, |t| \leq 1$$

οπότε θα πάρουμε

$$\sum_{x=1}^{\infty} xt^{x-1} = 1 + 2t + 3t^2 + \dots + \nu t^{\nu-1} + \dots = \left(\frac{1}{1-t} \right)' = \frac{1}{(1-t)^2}$$

και

$$\begin{aligned} \sum_{x=2}^{\infty} x(x-1)t^{x-2} &= 2 + 6t + \dots + \nu(\nu-1)t^{\nu-2} + \dots = \left(\frac{1}{1-t} \right)' \\ &= \left(\frac{1}{(1-t)^2} \right)' = \frac{2}{(1-t)^3} \end{aligned}$$

Θέτοντας στις τελευταίες $t=q$, και αντικαθιστώντας στις εκφράσεις που έχουμε για το μ και μ_2 έχουμε

$$\mu = p \frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

$$\mu_2 = pq \frac{2}{(1-q)^3} = pq \frac{2}{p^3} = \frac{2q}{p^2}$$

Επομένως $\mu = E(X) = 1/p$, ενώ για την εύρεση της διακύμανσης $V(X)$ αρκεί να κάνουμε για μία ακόμη φορά χρήση του τύπου (βλέπε την προαναφερθέντα απόδειξη)

$$V(X) = \mu_2 + \mu - \mu^2$$

Οπότε θα έχουμε

$$V(X) = \frac{2q}{p^2} + \frac{1}{p} - \left(\frac{1}{p} \right)^2 = \frac{2(1-p) + p - 1}{p^2} = \frac{1-p}{p^2} = \frac{p}{p^2}$$

3. Αρνητική διωνυμική κατανομή

Μια άμεση γενίκευση της γεωμετρικής κατανομής προκύπτει αν εξετάσουμε το χρόνο αναμονής μέχρι την r επιτυχία, όπου r είναι ένας θετικός ακέραιος αριθμός. Πιο συγκεκριμένα, έχουμε τον επόμενο ορισμό. Θεωρούμε μια ακολουθία (ανεξάρτητων) δοκιμών Bernoulli με πιθανότητα επιτυχίας p (και αποτυχίας $q=1-p$) σταθερή για όλες τις δοκιμές. Έστω X ο αριθμός των δοκιμών μέχρι την εμφάνιση της r επιτυχίας. Η κατανομή της τυχαίας μεταβλητής X καλείται **αρνητική διωνυμική κατανομή** με παραμέτρους r και p συμβολίζεται με $Nb(r,p)$.

Το σύνολο τιμών της τυχαίας μεταβλητής X είναι $R_x = \{r, r+1, r+2, \dots\}$

Για την εύρεση της συνάρτησης πιθανότητας της X αρκεί να παρατηρήσουμε ότι το ενδεχόμενο $\{X=x\}$ είναι ισοδύναμο με την τομή των ανεξάρτητων ενδεχομένων

A: στη δοκιμή x έχουμε επιτυχία,

B: στις δοκιμές 1, 2,...,k-1 εμφανίστηκαν ακριβώς r-1 επιτυχίες. Επομένως $P(X=x)=P(AB)=P(A)P(B)$ Και αφού $P(A)=p$
 $P(B)=\binom{x-1}{r-1}p^{r-1}q^{(x-1)-(r-1)}=\binom{x-1}{r-1}p^{r-1}q^{x-r}$

3.1 Συνάρτηση πιθανότητας της αρνητικής διωνυμικής κατανομής

Η συνάρτηση πιθανότητας της αρνητικής διωνυμικής κατανομής $Nb(r,p)$ δίνεται από τον τύπο

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

Τόσο από τους άνωθεν ορισμούς όσο και με σύγκριση των αποτελεσμάτων των προαναφερθέντων προτάσεων προκύπτει άμεσα ότι η γεωμετρική κατανομή συμπίπτει με την ειδική περίπτωση $r=1$ της αρνητικής διωνυμικής $Nb(r,p)$. Η ονομασία “αρνητική διωνυμική κατανομή” οφείλεται στο γεγονός ότι, η συνάρτηση πιθανότητας $f(x)$ εμφανίζεται στο διωνυμικό ανάπτυγμα της αρνητικής δύναμης $(1-q)^{-r}$. Πράγματι, από την προηγούμενη πρόταση έχουμε

$$(1-q)^{-r} = \sum_{k=0}^{\infty} \binom{r+k-1}{k} q^k = \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} q^k$$

Και αν στο τελευταίο άθροισμα θέσουμε $r+k=x$ προκύπτει

$$(1-q)^{-r} = \sum_{x=r}^{\infty} \binom{x-1}{r-1} q^{x-r}$$

Ή ακόμη πολλαπλασιάζοντας επί p^x

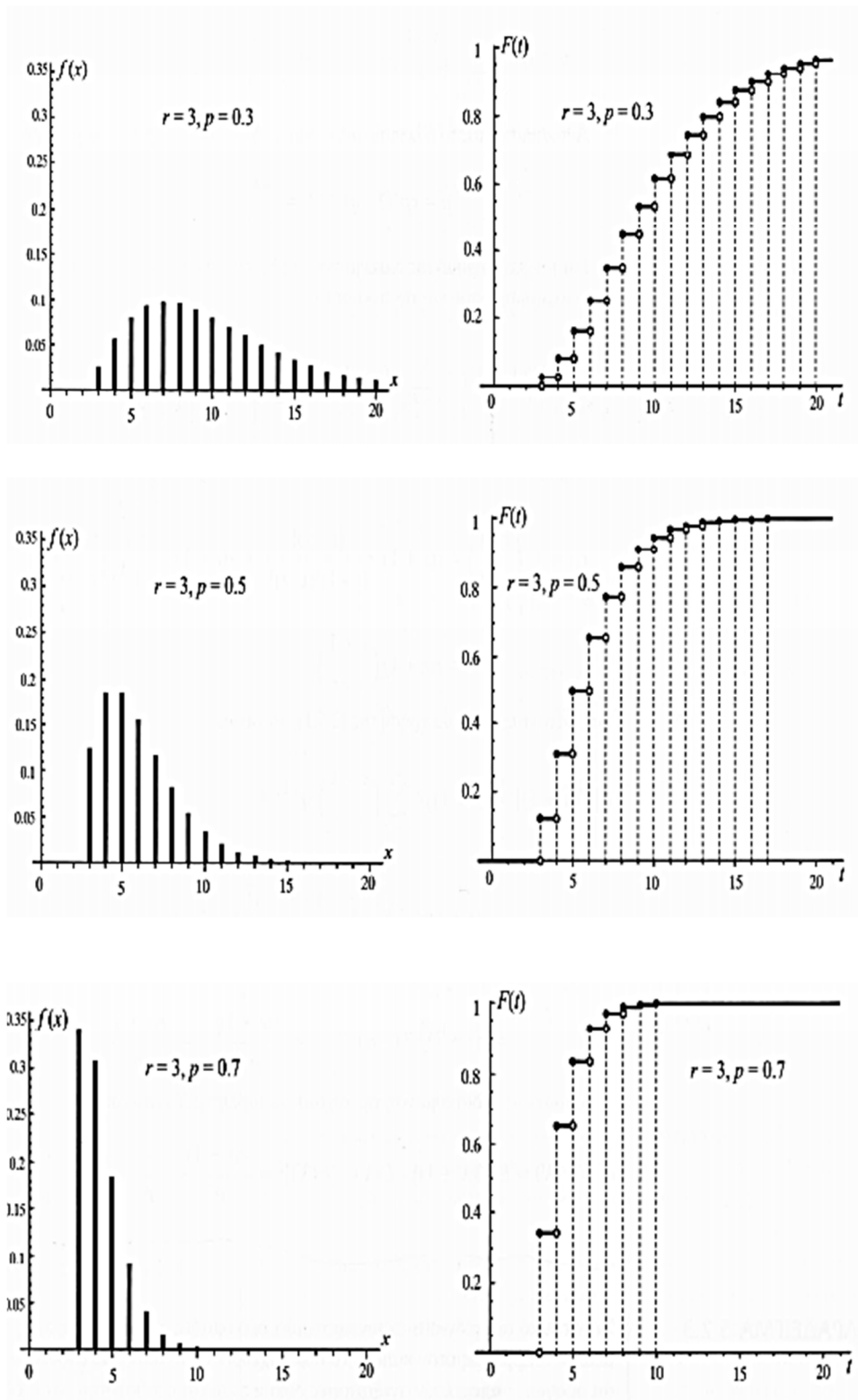
$$p(1-q)^{-r} = \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r q^{x-r} = \sum_{x=r}^{\infty} f(x)$$

Η τελευταία ισότητα δείχνει μάλιστα ότι η f είναι πράγματι μια συνάρτηση πιθανότητας μιας τυχαίας μεταβλητής με σύνολο τιμών $R_x = \{r, r+1, r+2, \dots\}$, αφού ισχύει $f(x) \geq 0$ για όλα τα $x \in R_x$ και επιπλέον

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

Η συνάρτηση κατανομής της αρνητικής διωνυμικής κατανομής με παραμέτρους r και p δίνεται από τον τύπο

$$F(t) = \begin{cases} 0, & t < r \\ \sum_{x=r}^{\lfloor t \rfloor} f(x) = \sum_{x=r}^{\lfloor t \rfloor} \binom{x-1}{r-1} p^r q^{x-r}, & t \geq r \end{cases}$$



3.2 Μέση τιμή και διακύμανση της αρνητικής διωνυμικής κατανομής

Αν η τυχαία μεταβλητή X ακολουθεί την αρνητική διωνυμική και τις παραμέτρους r και p , τότε

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

Και χρησιμοποιώντας τη σχέση

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

Βρίσκουμε

$$\mu = r p^r \sum_{x=r}^{\infty} \binom{x}{r} q^{x-r}$$

Η ακόμη, θέτοντας $x+1=y$

$$\mu = r p^r \sum_{y=r+1}^{\infty} \binom{y-1}{r} q^{y-r-1} = r p^r \sum_{y=r+1}^{\infty} \binom{y-1}{(r+1)-1} q^{y-(r+1)}$$

Αντικαθιστώντας το άθροισμα με βάση την παραπάνω σχέση, έχουμε

$$\mu = r p^r (1-q)^{-(r+1)} = \frac{r p^r}{p^{r+1}} = \frac{r}{p}$$

Για τον υπολογισμό της διακύμανσης $V(X)$, διευκολύνει να ξεκινήσουμε με την αναζήτηση τύπου για την ποσότητα

$$E[X(X+1)] = \sum_{x \in R_x} x(x+1)f(x) = \sum_{x=1}^{\infty} x(x+1) \binom{x-1}{r-1} p^r q^{x-r}$$

Όμως

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

Και προκύπτει

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

Επομένως, η διακύμανση της τυχαίας μεταβλητής X είναι ίση με

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

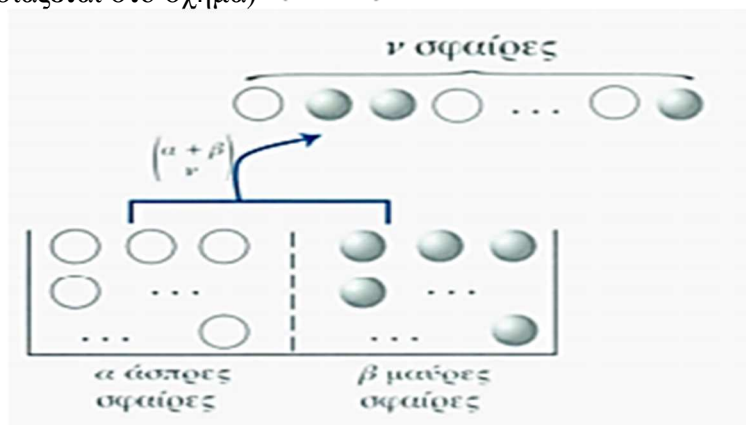
4. Η Υπεργεωμετρική κατανομή

Έστω ότι τα άτομα πεπερασμένου πληθυσμού ταξινομούνται σε δύο κατηγορίες ανάλογα με τις τιμές κάποιου χαρακτηριστικού που μας ενδιαφέρει π.χ. άνδρες-γυναίκες (χαρακτηριστικό που νέας ενδιαφέρει: φύλλο), καπνιστές- μη καπνιστές, καλό – ελαττωματικό αντικείμενο μιας γραμμής παραγωγής, ψηφίο 0-1 κατά την εκπομπή δυαδικών σημάτων, υπερτασικό– μη υπερτασικό άτομο (χαρακτηριστικό: τιμή νέας πίεσης), υπέρβαρο-μη υπέρβαρο άτομο (χαρακτηριστικό: βάρος) κτλ. Ένα δείγμα μεγέθους n επιλέγεται τυχαία από τον πληθυσμό αυτό, χωρίς επανατοποθέτηση. Όταν θέλουμε να μελετήσουμε το πλήθος των στοιχείων του δείγματος (ατόμων) τα οποία ανήκουν σε καθεμία από νέας δύο κατηγορίες η κατανομή που εμφανίζεται είναι η λεγόμενη **υπεργεωμετρική κατανομή**. Είναι πολύ εύκολο να διαπιστωθεί ότι, όλα τα προαναφερθέντα παραδείγματα, μπορούν να

ενταχθούν στο επόμενο μοντέλο κάλπης το οποίο χρησιμοποιείται για την εισαγωγή νέας κατανομής.

4.1 Ορισμός υπεργεωμετρικής κατανομής

Από μία κάλπη η οποία περιέχει A άσπρες και B μαύρες σφαίρες εξάγουμε διαδοχικά τη μία μετά την άλλη, n σφαίρες χωρίς επανατοποθέτηση. Έστω X ο αριθμός των άσπρων σφαιρών που περιέχονται στο δείγμα. Η κατανομή της τυχαίας μεταβλητής X καλείται υπεργεωμετρική κατανομή με παραμέτρους α, β και n και θα συμβολίζεται με $h(n, \alpha, \beta)$. Τα αποτελέσματα του πειράματος (επιλογή n σφαιρών) είναι όλες οι δυνατές επιλογές n σφαιρών από τις $\alpha + \beta$ συνολικά σφαίρες που υπάρχουν στην κάλπη. Επομένως, το πλήθος των δυνατών αποτελεσμάτων είναι $\binom{\alpha + \beta}{n}$ (όπως παρουσιάζεται στο σχήμα)



Αφού η επιλογή των σφαιρών γίνεται με τυχαίο τρόπο, όλα αυτά τα αποτελέσματα είναι ισοπίθανα και το καθένα από αυτά έχει η πιθανότητα $1/\binom{\alpha + \beta}{n}$ να εμφανιστεί. Συνεπώς για τον υπολογισμό πιθανοτήτων μπορούμε να χρησιμοποιήσουμε τον κλασικό ορισμό.

4.2 Συνάρτηση πιθανότητας της υπεργεωμετρικής κατανομής

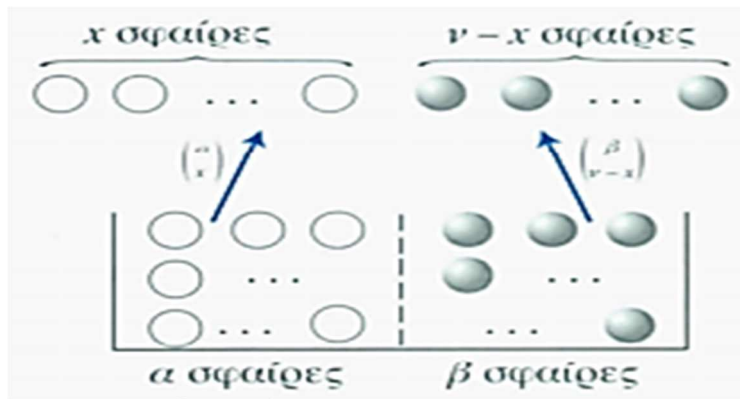
Η συνάρτηση πιθανότητας της υπεργεωμετρικής κατανομής με παραμέτρους α, β και n δίνεται από τον τύπο

$$f(x) = P(X = x) = \frac{\binom{\alpha}{x} \binom{\beta}{n-x}}{\binom{\alpha + \beta}{n}}, \quad x = \max(0, n-\beta), \dots, \min(n, \alpha)$$

Απόδειξη

Όπως φαίνεται και στο σχήμα τα ευνοϊκά αποτελέσματα για το ενδεχόμενο $\{X=x\}$ προκύπτουν επιλέγοντας x άσπρες σφαίρες από τις α που είναι διαθέσιμες και $n-x$ μαύρες από τις β που είναι διαθέσιμες. Το πλήθος των τρόπων επιλογής των άσπρων σφαιρών είναι και για κάθε τέτοια επιλογή, υπάρχουν $\binom{\beta}{n-x}$ δυνατές επιλογές για τις $n-x$ μαύρες σφαίρες που χρειάζονται για να συμπληρωθεί το δείγμα.

Σύμφωνα με την πολλαπλασιαστική αρχή ο αριθμός των τρόπων επιλογής των ευνοϊκών αποτελεσμάτων θα είναι $\binom{\alpha}{x} \cdot \binom{\beta}{n-x}$



Αφού υπάρχουν $\binom{\alpha + \beta}{\nu}$ διαφορετικοί τρόποι επιλογής των δυνατών αποτελεσμάτων του πειράματος η ζητούμενη πιθανότητα θα δίνεται από τον τύπο

$$f(x) = P(X = x) = \frac{\binom{\alpha}{x} \binom{\beta}{\nu - x}}{\binom{\alpha + \beta}{\nu}}$$

Το X θα μπορούσε θεωρητικά να πάρει οποιαδήποτε ακέραια μία αρνητική τιμή χωρίς όμως να μπορεί να ξεπεράσει το μέγεθος του δείγματος δηλαδή

$$X = 0, 1, 2, \dots, \nu$$

Όμως Αφού υπάρχουν μόνο άσπρες σφαίρες στην κάλπη θα πρέπει να ισχύει

$$x \leq \alpha$$

ενώ λόγω του περιορισμού ότι υπάρχουν μόνο β μαύρες σφαίρες θα πρέπει να έχουμε επιπλέον

$$0 \leq \nu - x \leq \beta$$

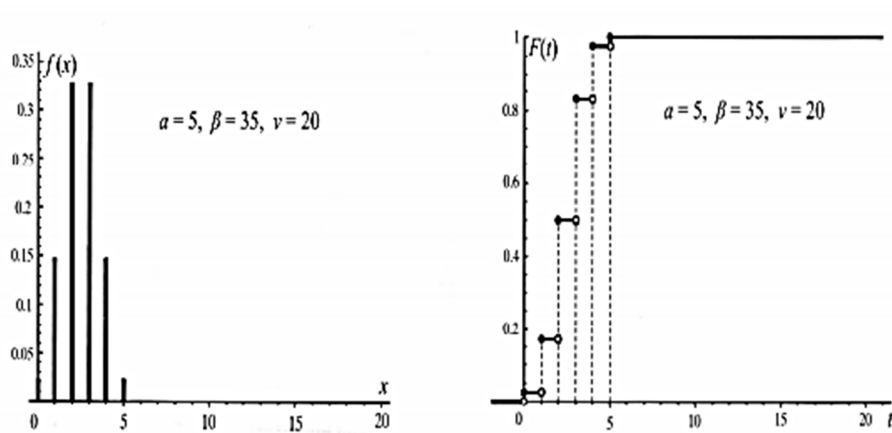
Συνδυάζοντας όλους τους ορισμούς καταλήγουμε στις ανισότητες

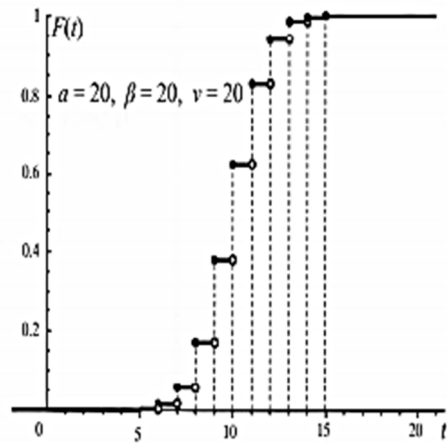
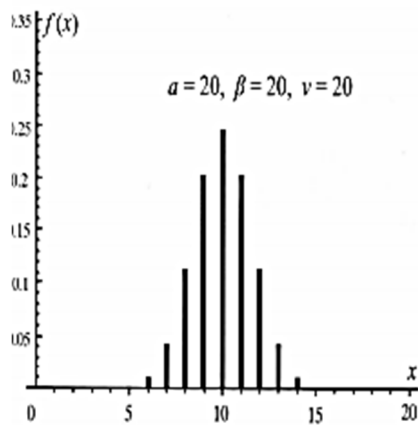
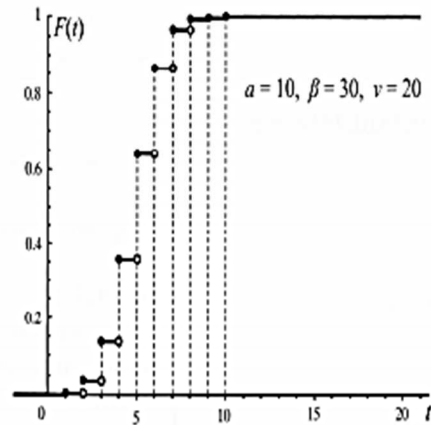
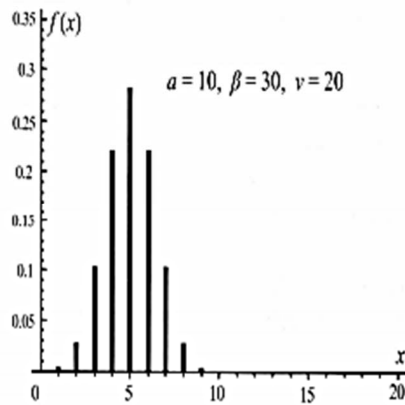
$$\nu \leq \alpha, \quad \nu \leq x$$

και

$$x \geq 0, \quad x \geq \nu - \beta$$

Επομένως, η μικρότερη τιμή που μπορεί να πάρει το x είναι το $\max(0, \nu - \beta)$ ενώ η μεγαλύτερη το $\min(\nu, \alpha)$.





4.3 Μέση τιμή και διακύμανση της υπεργεωμετρικής κατανομής

Αν η τυχαία μεταβλητή X ακολουθεί την υπεργεωμετρική κατανομή με παραμέτρους α, β και ν τότε

$$\mu = E(X) = \nu \frac{\alpha}{\alpha + \beta}, \quad \sigma^2 = V(X) = \nu \frac{\alpha}{\alpha + \beta} \frac{\beta}{\alpha + \beta} \left(1 - \frac{\nu - 1}{\alpha + \beta - 1} \right)$$

Ο πρώτος τύπος προκύπτει άμεσα από την έκφραση που βρήκαμε για την παραγοντική ροπή $r = 1$. Για τη διακύμανση, αρκεί να παρατηρήσουμε ότι

$$E[X(X-1)] = \mu_2 = \frac{(\alpha)_2 (\nu)_2}{(\alpha + \beta)_2} = \frac{\alpha(\alpha-1)\nu(\nu-1)}{(\alpha + \beta)(\alpha + \beta - 1)}$$

Και να χρησιμοποιήσουμε για άλλη μια φορά τον τύπο $V(X) = \mu_{(2)} + \mu - \mu^2$

5. Κατανομή Poisson

Ένα από τα σημαντικότερα προβλήματα τα οποία αντιμετωπίζουμε κατά τη χρήση των διωνυμίων συντελεστών $\binom{\nu}{x}, x=0, 1, \dots, \nu$ είναι οι μεγάλες τιμές που παίρνουν όταν το ν είναι μεγάλο και το x όχι πολύ κοντά στο 0 ή το ν . Δεδομένου ότι αυτοί οι συντελεστές εμφανίζονται στον τύπο της διωνυμικής κατανομής $b(\nu, p)$

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

Ο απευθείας υπολογισμός πιθανοτήτων μέσω αυτού του τύπου καθίσταται προβληματικός, τουλάχιστον για τις περιπτώσεις που αναφέρθηκαν παραπάνω. Ένας τρόπος για να ξεπεράσει κανείς το πρόβλημα, είναι να καταφύγει στην αναδρομική σχέση.

$$f(x) = \left(\alpha + \frac{\beta}{x}\right)f(x-1), \quad x=0,1,\dots,v$$

Όπου

$$\alpha = -\frac{p}{1-p}, \beta = -\frac{(\nu+1)p}{1-p}$$

Η τελευταία, σε συνδυασμό με την αρχική προσθήκη $f(0) = (1-p)^\nu$ δίνει ένα σχετικά εύκολο σχήμα υπολογισμού των πιθανοτήτων για όλα τα $x=0,1,\dots,v$. Ένα μειονέκτημα αυτής της προσέγγισης είναι ότι, ακόμη και αν μας ενδιαφέρει να υπολογίσουμε μόνο την πιθανότητα για κάποια μεγάλη τιμή του x , θα πρέπει να περάσουμε αναγκαστικά από όλες τις τιμές $f(0), f(1), \dots, f(x-1)$. Ο Simeon -Denis Poisson (1781-1840), αναγνωρίζοντας τη δυσκολία αυτή, πραγματοποίησε μια συστηματική μελέτη της διωνυμικής κατανομής με στόχο την ανακάλυψη κάποιας έμμεσης μεθόδου υπολογισμού των πιθανοτήτων στην περίπτωση που

- Το ν είναι πολύ μεγάλος αριθμός ($\nu \rightarrow \infty$)
- Η πιθανότητα επιτυχίας p είναι πολύ μικρή ($p \rightarrow 0$)
- Ο μέσος αριθμός επιτυχιών στις ν δοκιμές παίρνει μέτριες τιμές, δηλαδή το γινόμενο νp συγκλίνει προς κάποιο σταθερό αριθμό $\lambda > 0$.

Το αποτέλεσμα της προσπάθειας του ήταν το επόμενο εντυπωσιακό για την εποχή εκείνη **οριακό θεώρημα**, το οποίο δημοσιεύτηκε το 1837, σε σχετική μονογραφία που εξέδωσε.

5.1 Όριο της διωνυμικής κατανομής $b(\nu, p)$ για μεγάλα ν

Έστω ότι η τυχαία μεταβλητή X ακολουθεί τη διωνυμική κατανομή $b(\nu, p)$ με συνάρτηση πιθανότητας

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

Αν, για ($\nu \rightarrow \infty$), η πιθανότητα επιτυχίας p συγκλίνει στο 0 ($p \rightarrow 0$) έτσι ώστε το γινόμενο $\mu = E(X) = \nu p$ να συγκλίνει προς μία θετική σταθερά λ ($\nu p \rightarrow \lambda > 0$) τότε

$$\lim_{\nu \rightarrow \infty} f(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x=0,1,\dots$$

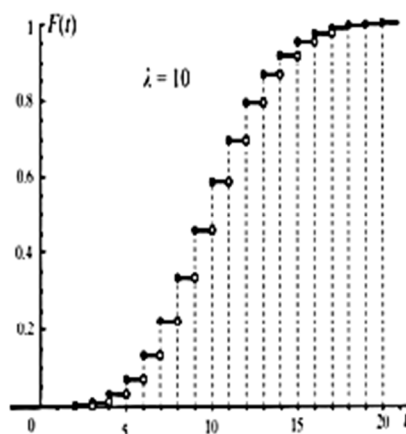
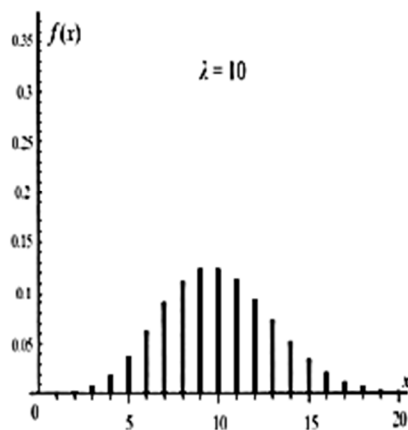
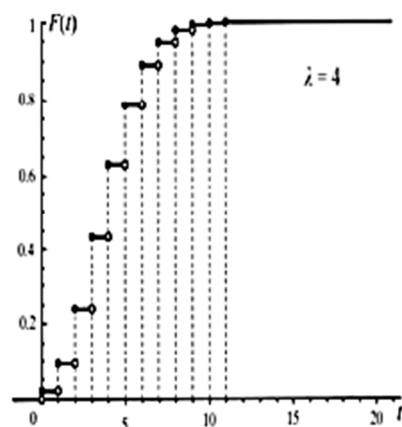
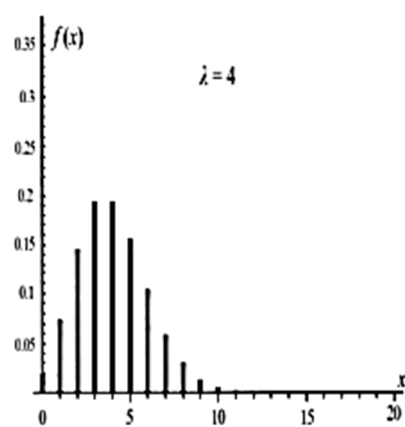
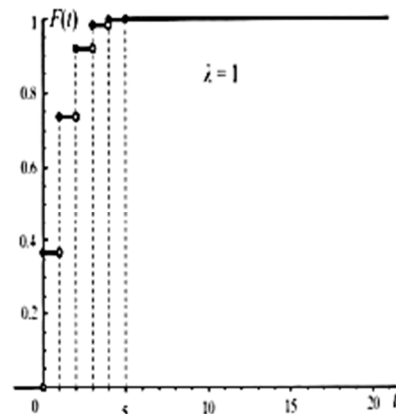
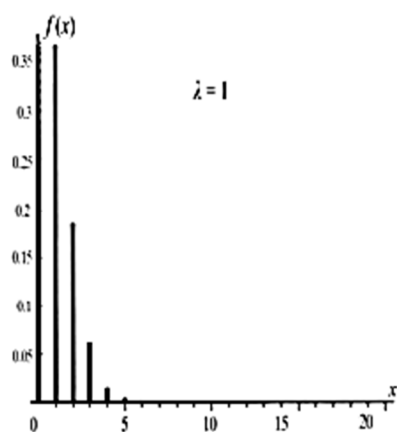
5.2 Ορισμός κατανομής Poisson

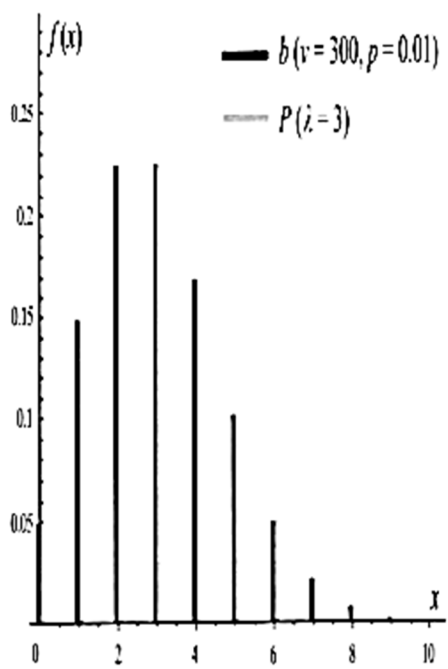
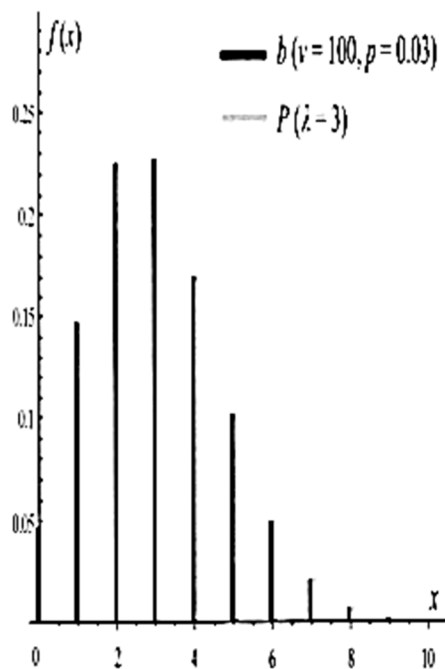
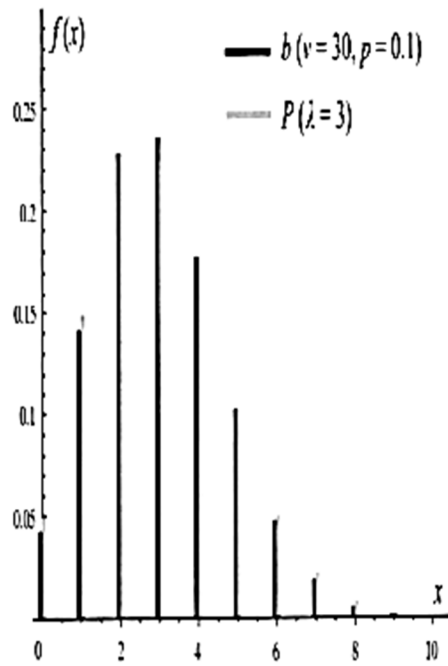
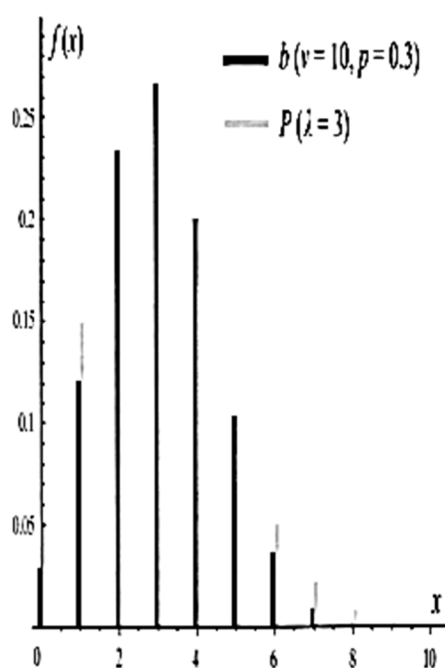
Έστω X μία διακριτή τυχαία μεταβλητή με σύνολο τιμών $R_X = \{0,1,2,\dots\}$ και συνάρτηση πιθανότητας

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

Όπου $\lambda > 0$. Η κατανομή της τυχαίας μεταβλητής X λέγεται κατανομή Poisson με παράμετρο λ και συμβολίζεται με $P(\lambda)$. Για τη συνάρτηση κατανομής $P(\lambda)$ μπορούμε να γράψουμε

$$F(t) = \begin{cases} 0 & t < 0 \\ \sum_{x=0}^{|t|} f(x) & t \geq 0 \end{cases} = \begin{cases} 0 & t < 0 \\ e^{-\lambda} \sum_{x=0}^{|t|} \frac{\lambda^x}{x!} & t \geq 0 \end{cases}$$





5.3 Η κατανομή Poisson ως όριο της διωνυμικής κατανομής

Έστω ότι η τυχαία μεταβλητή X ακολουθεί τη διωνυμική κατανομή με παραμέτρους n και p . Αν, για $n \rightarrow \infty$, η πιθανότητα επιτυχίας συγκλίνει στο 0, έτσι ώστε ο μέσος της κατανομής np να συγκλίνει σε μία θετική σταθερά λ , τότε η κατανομή της τυχαίας μεταβλητής X , μπορεί να προσεγγιστεί από την κατανομή Poisson με παράμετρο λ . Στο παρακάτω σχήμα φαίνεται γραφικά η σύγκλιση της διωνυμικής κατανομής $b(n,p)$ προς την κατανομή Poisson με παράμετρο $\lambda=np$ όταν $n \rightarrow \infty$, $p \rightarrow 0$. Στις εφαρμογές όπου θα κάνουμε χρήση του παραπάνω αποτελέσματος θα γράφουμε συνήθως την προσέγγιση στη μορφή

$$f(x) = \binom{\nu}{x} p^x q^{\nu-x} \cong e^{-\nu p} \frac{(\nu p)^x}{x!}, \quad x=0,1,2,\dots$$

Η ακόμη $\underline{b}(\nu, p) \approx \underline{P}(\nu, p)$