Θεωρία Πιθανοτήτων

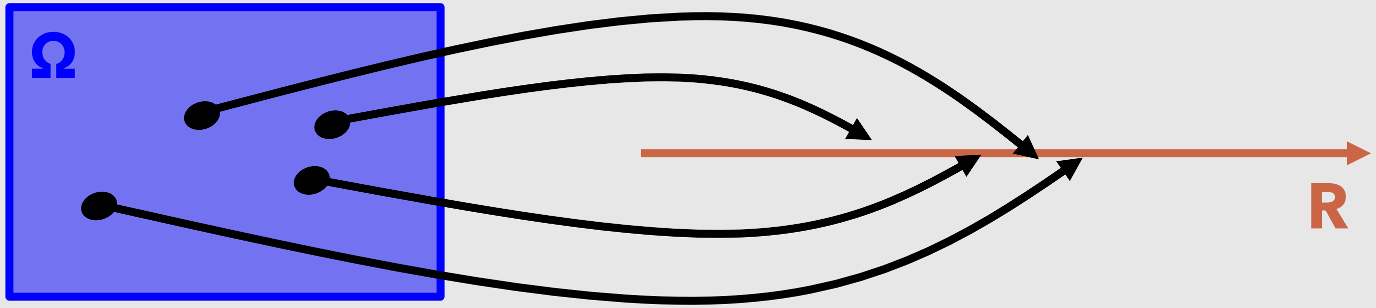
Ηλίας Σακέλλης
Γραφείο: IV-8 #9
Τηλέφωνο: 210 727 6727
Email: e_sakel@phys.uoa.gr

Βιβλιογραφία/Αναφορές

1. S. Ross. Βασικές αρχές θεωρίας πιθανοτήτων. Εκδόσεις Κλειδάριθμος.
2. Hoel, Port, Stone. Εισαγωγή στη θεωρία πιθανοτήτων. Πανεπιστημιακές Εκδόσεις Κρήτης.
3. ΠΑΠΟΥΛΗΣ, PILLAI - ΠΙΘΑΝΟΤΗΤΕΣ, ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ ΚΑΙ ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ.
4. Γ. Κοντογιάννης, Σ. Τουμπής (2015) - Στοιχεία πιθανοτήτων. Στα έγγραφα στην η-τάξη.
5. Χ. Δαμιανού, Ν. Παπαδάτος, Χ. Χαραλαμπίδης - Εισαγωγή στις Πιθανότητες και τη Στατιστική.
6. Χ. Χαραλαμπίδης. Θεωρία Πιθανοτήτων και Εφαρμογές. Εκδόσεις Συμμετρία.
7. Schaum's Outline of Theory and Problems of Probability and Statistic.
8. Ο. Χρυσαφινού, Α. Μπουρνέτας, Ε. Βαγγελάτου – Σημειώσεις Πιθανοτήτων – Στατιστικής.

Ορισμός : Τυχαία μεταβλητή

Τυχαία μεταβλητή καλείται μία πραγματική συνάρτηση η οποία αποδίδει έναν πραγματικό αριθμό σε κάθε ενδεχόμενο του δειγματικού χώρου ενός πειράματος.



Τις τυχαίες μεταβλητές τις συμβολίζουμε με κεφαλαία γράμματα ενώ τις αντίστοιχες τιμές τους με μικρά γράμματα.

Ορισμός : Διακριτές & Συνεχείς τυχαίες μεταβλητές

Μία τυχαία μεταβλητή μπορεί να είναι διακριτή αν το σύνολο των τιμών που λαμβάνει (πεδίο ορισμού) είναι πεπερασμένο ή απείρως αριθμήσιμο. Σε διαφορετική περίπτωση όπου παίρνει άπειρες τιμές λέγεται συνεχώς τυχαία μεταβλητή.

Παρατήρηση

Μία συνάρτηση τυχαίας μεταβλητής είναι και αυτή μία τυχαία μεταβλητή, για παράδειγμα X^n, e^{X}.

Ορισμός : Συνάρτηση πιθανότητας

Για μια τυχαία μεταβλητή **X** με σύνολο τιμών x_1, x_2, …, $x_ν$ ορίζεται η συνάρτηση $p_{\mathbf{X}}(x)$: $\mathbb{R} \rightarrow [0, 1]$ πιθανότητας της **X** (ή συνάρτηση μάζας πιθανότητας)

$$p_{\mathbf{X}}(x) = \begin{cases} P(\mathbf{X} = x), & x \in \{x_1, x_2, \ldots, xν\} \\ 0, & \text{σε άλλη περίπτωση} \end{cases}$$

Ισχύει ότι: $\sum_{i=1}^{ν} p_{\mathbf{X}}(xi) = 1$

Ορισμός : Αθροιστική συνάρτηση κατανομής

Για μια τυχαία μεταβλητή **X** με σύνολο τιμών x_1, x_2, ..., x_ν ορίζεται η πραγματική συνάρτηση κατανομής ή αθροιστική συνάρτηση κατανομής
$F_{\mathbf{X}}: \mathbb{R} \rightarrow [0, 1]$

$$F_{\mathbf{X}}(x) = P(\mathbf{X} \leq x) = \sum_{x_i \leq x} [p_{\mathbf{X}}(xi)] \quad \forall x \in \mathbb{R}$$

Παρατήρηση

Η συνάρτηση πιθανότητας εκφράζει την πιθανότητα ή τυχαία μεταβλητή **X** να πάρει την τιμή x ενώ η συνάρτηση κατανομής εκφράζει την πιθανότητα η τυχαία μεταβλητή **X** να πάρει τιμή μικρότερη ή ίση της x.

Παράδειγμα

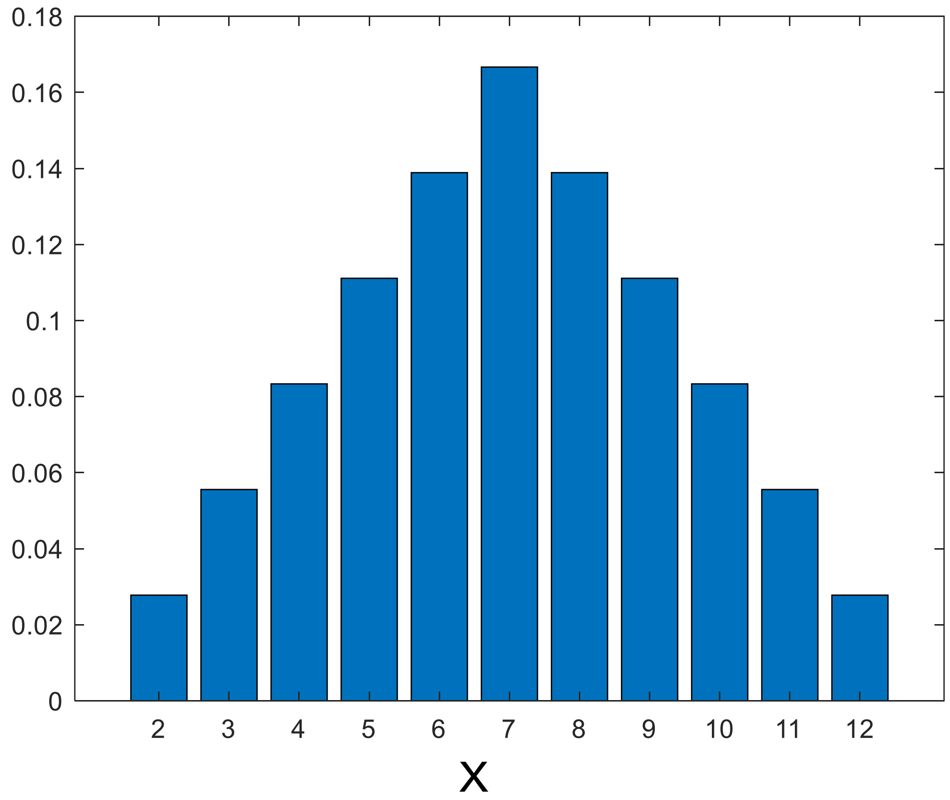
Ρίχνουμε 2 ζάρια.

$\Omega = \{(1, 1), (1, 2), \ldots, (6, 6)\}$

Έστω η τυχαία μεταβλητή Χ: το άθροισμα των ενδείξεων των 2 ζαριών.

$$p_{\mathbf{X}}(x) = \begin{cases} 1/36, & \text{αν } x = 2 \text{ ή } 12 \\ 2/36, & \text{αν } x = 3 \text{ ή } 11 \\ 3/36, & \text{αν } x = 4 \text{ ή } 10 \\ 4/36, & \text{αν } x = 5 \text{ ή } 9 \\ 5/36, & \text{αν } x = 6 \text{ ή } 8 \\ 6/36, & \text{αν } x = 7 \\ 0, & \forall \text{ άλλο } x \end{cases}$$

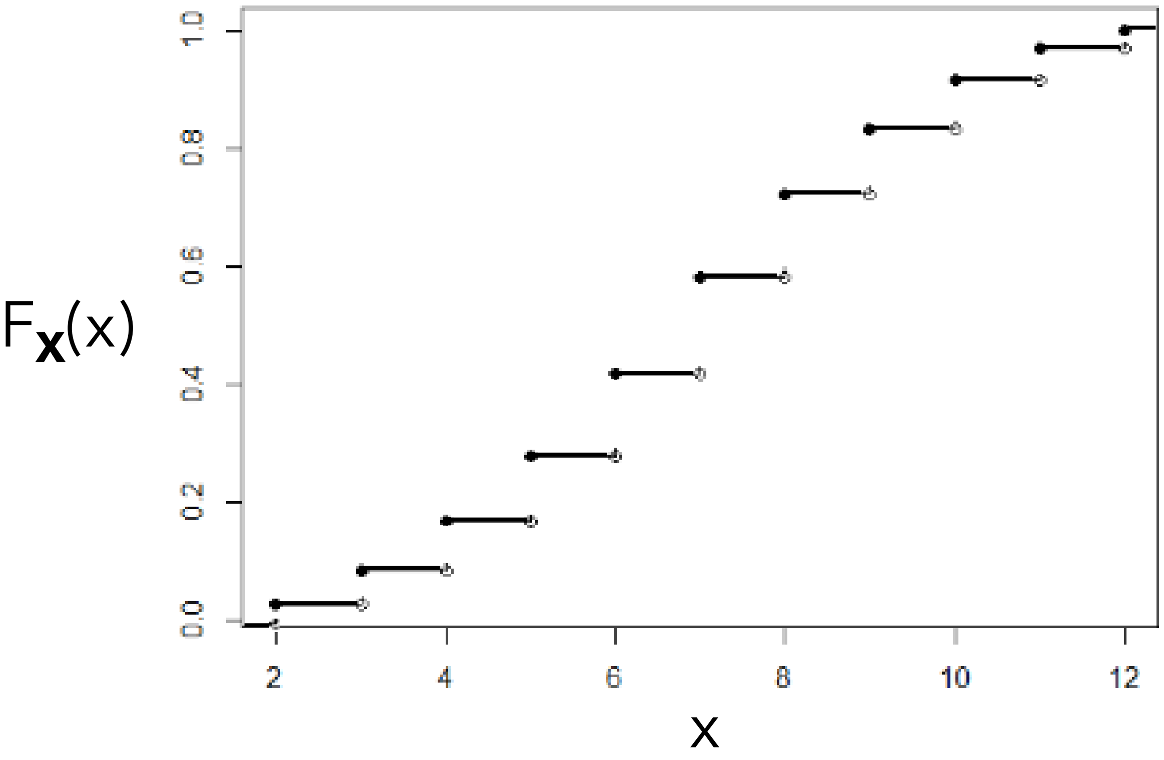
$p_{\mathbf{X}}(x) = P(x)$



Παράδειγμα (συνέχεια)

$$
p_{\mathbf{X}}(x) = \begin{cases} 1/36 \text{ , αν } x = 2 \text{ ή } 12 \\ 2/36 \text{ , αν } x = 3 \text{ ή } 11 \\ 3/36 \text{ , αν } x = 4 \text{ ή } 10 \\ 4/36 \text{ , αν } x = 5 \text{ ή } 9 \\ 5/36 \text{ , αν } x = 6 \text{ ή } 8 \\ 6/36 \text{ , αν } x = 7 \\ 0 \text{ , } \forall \text{ άλλο } x \end{cases}
$$

$$
F_{\mathbf{X}}(x) = \begin{cases} 0 \text{ , αν } x < 2 \\ 1/36 \text{ , αν } 2 \leq x < 3 \\ 3/36 \text{ , αν } 3 \leq x < 4 \\ 6/36 \text{ , αν } 4 \leq x < 5 \\ 10/36 \text{ , αν } 5 \leq x < 6 \\ . \\ . \\ . \\ 33/36 \text{ , αν } 10 \leq x < 11 \\ 35/36 \text{ , αν } 11 \leq x < 12 \\ 1 \text{ , αν } x \geq 12 \end{cases}
$$



Ιδιότητες Αθροιστικής συνάρτησης κατανομής
1. Έχει σύνολο τιμών [0, 1].
2. Είναι αύξουσα συνάρτηση.
3. $\lim_{x \to +\infty} F_X() = 1, \lim_{x \to -\infty} F_X() = 0.$
4. Είναι συνεχής από δεξιά.

Ιδιότητες Αθροιστικής συνάρτησης κατανομής

Για μία τυχαία μεταβλητή **X** με αθροιστική συνάρτηση κατανομής $F_{\mathbf{X}}$ με α<β

όπου α, $\beta \in \mathbb{R}$ και $\lim_{x \to x_0^-} F_{\mathbf{X}}(x) = F_{\mathbf{X}}(x_0^-)$ και $\lim_{x \to x_0^+} F_{\mathbf{X}}(x) = F_{\mathbf{X}}(x_0)$ ισχύουν:

1. $P(\mathbf{X} = x_0) = F_{\mathbf{X}}(x_0) - F_{\mathbf{X}}(x_0^-)$
2. $P(\mathbf{X} < x_0) = F_{\mathbf{X}}(x_0^-)$
3. $P(\mathbf{X} \leq x_0) = F_{\mathbf{X}}(x_0)$
4. $P(\alpha < \mathbf{X} < \beta) = F_{\mathbf{X}}(\beta^-) - F_{\mathbf{X}}(\alpha)$
5. $P(\alpha \leq \mathbf{X} < \beta) = F_{\mathbf{X}}(\beta^-) - F_{\mathbf{X}}(\alpha^-)$
6. $P(\alpha \leq \mathbf{X} \leq \beta) = F_{\mathbf{X}}(\beta) - F_{\mathbf{X}}(\alpha^-)$
7. $P(\mathbf{X} > \alpha) = 1 - F_{\mathbf{X}}(\alpha)$
8. $P(\mathbf{X} \geq \alpha) = 1 - F_{\mathbf{X}}(\alpha^-)$

Ορισμός : Μέση τιμή
(ή μαθηματική ελπίδα ή αναμενόμενη τιμή ή προσδοκώμενη τιμή)

Η μέση τιμή διακριτής μεταβλητής **X** με σύνολο τιμών $S = \{x_1, x_2, \ldots, x_\nu\}$ και συνάρτηση πιθανότητας $P(x)$ ορίζεται ως

$$\mu = E[\mathbf{X}] = \sum_{x_i\ S} x_i \cdot P(x_i)$$

Για οποιαδήποτε συνάρτηση $f(x)$ η μέση τιμή της ορίζεται από

$$E[f(x_i)] = \sum_{x_i \in S} f(x_i) \cdot P(x_i)$$

Ισχύουν $E(c) = c$ και $E[c \cdot f(x)] = c \cdot E[f(x)]$

Παρατήρηση

Η μέση τιμή μπορεί να ερμηνευτεί ως ο μέσος όρος ενός πεπερασμένου πληθυσμού ή ως το «κέντρο βάρους» μιας κατανομής ή ως το αναμενόμενο κέρδος ανά παιχνίδι σε ένα παιχνίδι με πολλές επαναλήψεις.

Ορισμός : Διακύμανση ή Διασπορά

Η διακύμανση ή διασπορά μίας τυχαίας μεταβλητής **X** ορίζεται από τη σχέση

$$\mathrm{Var}(\mathbf{X}) = \sigma^2 = E((\mathbf{X} - \mu)^2) = \sum_{x_i} (x_i - \mu)^2 \cdot p_{\mathbf{X}}(x_i)$$

Η διακύμανση εκφράζει το βαθμό διασποράς των τιμών γύρω από τη μέση τιμή.

Η ποσότητα $\sqrt{\mathrm{Var}(\mathbf{X})} = \sigma$ ονομάζεται τυπική απόκλιση και έχει μονάδες ίδιες με αυτές του πρωτογενούς μεγέθους.

Ιδιότητες διακύμανσης

1. Αν $\mathbf{X} = c$ τότε $Var(\mathbf{X}) = 0$.
2. $Var(\mathbf{X}) \geq 0$.
3. $Var(\mathbf{X}) = E(\mathbf{X}^2) - E^2(\mathbf{X}) = E(\mathbf{X}^2) - \mu^2$
4. $Var(\alpha\mathbf{X}+\beta) = \alpha^2 \cdot Var(\mathbf{X})$
5. $Var(f(\mathbf{X})) = E(f^2(\mathbf{X})) - E^2(f(\mathbf{X}))$

Ορισμός : Ομοιόμορφη κατανομή

Μια τυχαία μεταβλητή **X** ακολουθεί την διακριτή ομοιόμορφη κατανομή αν μπορεί να πάρει κάθε τιμή από το σύνολο τιμών της με την ίδια πιθανότητα. Δηλαδή:

Η συνάρτηση πιθανότητας είναι $p_{\mathbf{X}}(x) = \begin{cases} \frac{1}{\nu}, & x \in \{x_1, x_2, \dots, x\nu\} \\ 0, & \text{σε άλλη περίπτωση} \end{cases}$

ενώ η αθροιστική συνάρτηση κατανομής είναι $F_{\mathbf{X}}(x) = \frac{x}{\nu}$

Η μέση τιμή της ομοιόμορφης κατανομής είναι $\mu = \frac{\nu+1}{2}$ ενώ η διασπορά είναι $\mathrm{Var}(\mathbf{X}) = \frac{\nu^2-1}{12}$

Ορισμός : Δοκιμή Bernoulli

Συχνά έχουμε να κάνουμε με ερωτήματα/πειράματα τα οποία επιδέχονται δυο μόνο απαντήσεις/αποτελέσματα, όχι απαραιτήτως ισοπίθανων, αλλά με κάποια πιθανότητα p να έχουμε «επιτυχία» και $q=1-p$ να έχουμε «αποτυχία». Αυτό το πείραμα τύχης καλείται δοκιμή Bernoulli.

Δοκιμή Bernoulli: Μέση τιμή, Διασπορά, Τυπική απόκλιση

1. Μέση τιμή

$$\mu = E(\mathbf{X}) = p$$

2. Διασπορά

$$\sigma^2 = Var(\mathbf{X}) = p \cdot (1 - p)$$

3. Τυπική απόκλιση

$$\sigma = \sqrt{Var(\mathbf{X})} = \sqrt{p \cdot (1 - p)}$$

Ορισμός : Διωνυμική κατανομή

Μια διακριτή τυχαία μεταβλητή **X** ακολουθεί διωνυμική κατανομή με παραμέτρους Ν και p αν έχει συνάρτηση πιθανότητας

$$p_{\mathbf{X}}(x) = \begin{cases} \binom{N}{x} \cdot p^{x} \cdot q^{N-x}, & x = 0,\ 1, \dots, \nu \\ 0, & \text{σε άλλη περίπτωση} \end{cases} \quad \text{όπου } q = 1 - p$$

Η διωνυμική κατανομή συμβολίζεται με τους εξής τρόπους: **X** ~ Διων(N, p) ή **X** ~ Bin(N, p) ή **X** ~ B(N, p) ή **X** ~ b(x, N, p). Εκφράζει δηλαδή την πιθανότητα x «επιτυχιών» σε N δοκιμές.

Προϋποθέσεις διωνυμικής κατανομής

1. Καθορισμένος αριθμός δοκιμών N
2. Ανεξάρτητες μεταξύ τους δοκιμές
3. Κάθε δοκιμή έχει 2 μόνο αποτελέσματα «επιτυχία» ή «αποτυχία»
4. Η πιθανότητα «επιτυχίας» p («αποτυχίας» q) είναι η ίδια σε κάθε δοκιμή

Διωνυμική κατανομή: Μέση τιμή, Διασπορά, Τυπική απόκλιση

1. Μέση τιμή

$$\mu = E(\mathbf{X}) = N \cdot p$$

2. Διασπορά

$$\sigma^2 = Var(\mathbf{X}) = N \cdot p \cdot (1 - p)$$

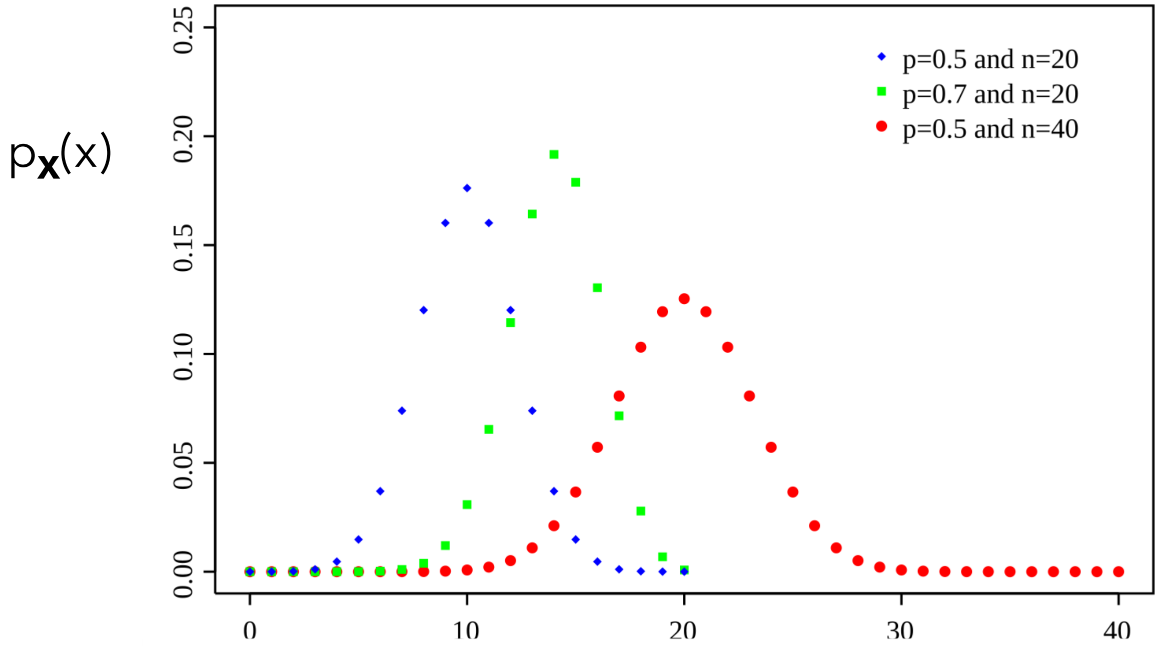
3. Τυπική απόκλιση

$$\sigma = \sqrt{Var(\mathbf{X})} = \sqrt{N \cdot p \cdot (1 - p)}$$

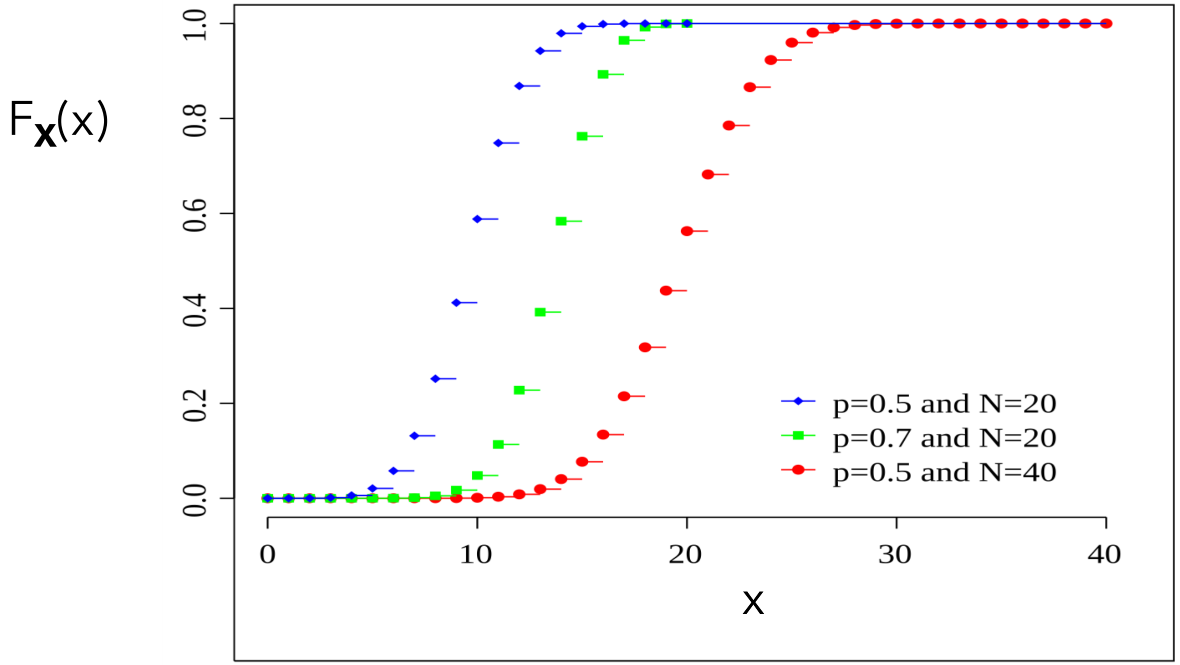
Παρατήρηση

Για $N = 1$ η διωνυμική κατανομή ανάγεται στην κατανομή Bernoulli η οποία έχει $\mu = p$ και $Var(\mathbf{X}) = p \cdot q$

Συνάρτηση πιθανότητας



Αθροιστική συνάρτηση κατανομής



https://en.wikipedia.org/wiki/Binomial_distribution

Ορισμός : Γεωμετρική κατανομή

Μια διακριτή τυχαία μεταβλητή **X** ακολουθεί γεωμετρική κατανομή με παράμετρο p αν έχει συνάρτηση πιθανότητας

$$p_{\mathbf{X}}(x) = \begin{cases} (1-p)^{x-1} \cdot p, & x = 1,\ 2, \dots, \nu \\ 0, & \text{σε άλλη περίπτωση} \end{cases} \quad \text{όπου } q = 1 - p$$

Η γεωμετρική κατανομή συμβολίζεται με τους εξής τρόπους: **X** ~ Geo(p) ή **X** ~ Γεωμ(p) ή **X** ~ g(x:p). Εκφράζει δηλαδή την πιθανότητα σε x δοκιμές να έχουμε «επιτυχία» για πρώτη φορά.

Γεωμετρική κατανομή: Μέση τιμή, Διασπορά, Τυπική απόκλιση

1. Μέση τιμή

$$\mu = E(\mathbf{X}) = \frac{1}{p}$$

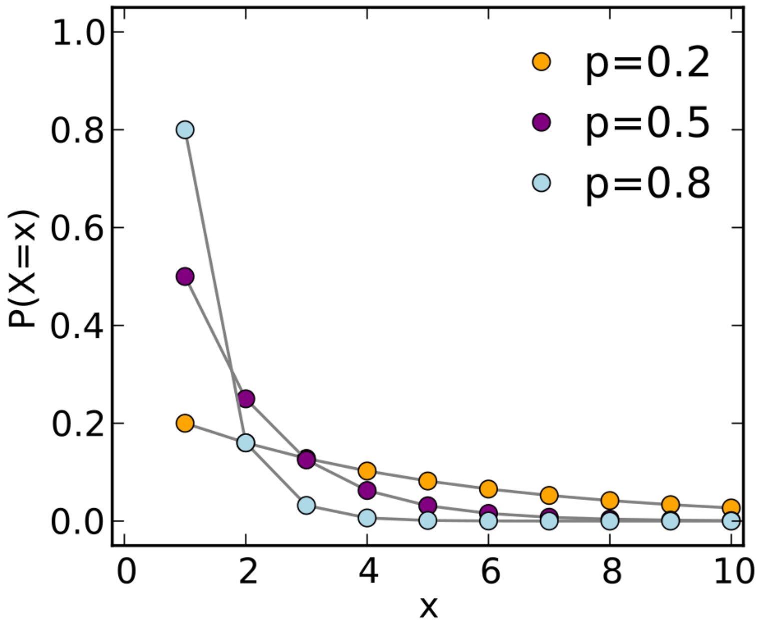
2. Διασπορά

$$\sigma^2 = Var(\mathbf{X}) = \frac{1-p}{p^2}$$

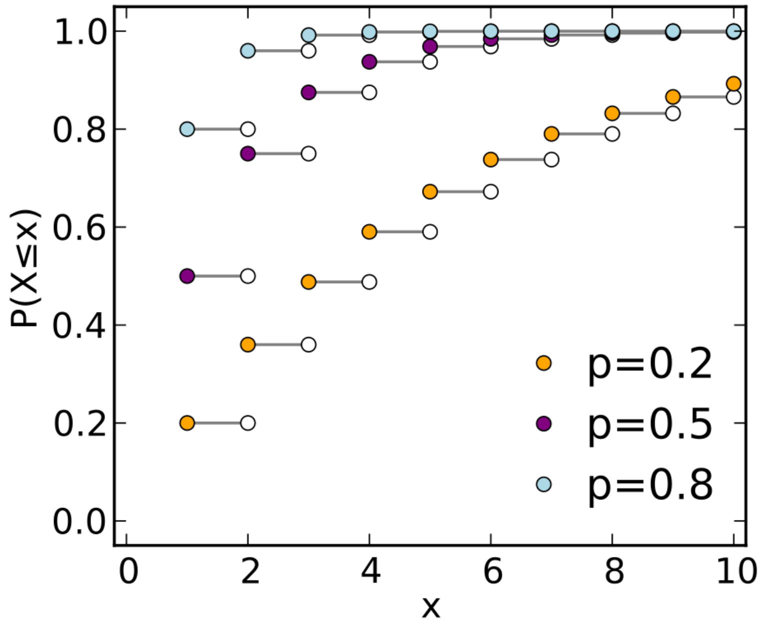
3. Τυπική απόκλιση

$$\sigma = \sqrt{Var(\mathbf{X})} = \sqrt{\frac{1-p}{p^2}}$$

Συνάρτηση πιθανότητας



Αθροιστική συνάρτηση κατανομής



https://en.wikipedia.org/wiki/Geometric_distribution

Ιδιότητες Γεωμετρικής κατανομής

1. Ουρά: $\forall\, x \geq 1 \quad P(\mathbf{X} > x) = (1\text{-}p)^x$
2. Έλλειψη μνήμης: $P(\mathbf{X} > x + y \mid \mathbf{X} > y) = P(\mathbf{X} > x)$

Δηλαδή η γνώση ότι δεν έχουμε «επιτυχία» σε y προηγούμενες προσπάθειες δεν επηρεάζει το πόσες προσπάθειες x θα απαιτηθούν κατόπιν μέχρι να υπάρξει «επιτυχία».

$$P(\mathbf{X} > x \mid \mathbf{X} > y) = P(\mathbf{X} > x - y)$$

$$P(\mathbf{X} \geq x \mid \mathbf{X} > y) = P(\mathbf{X} > x - y - 1)$$

$$P(\mathbf{X} \geq x \mid \mathbf{X} \geq y) = P(\mathbf{X} > x - y)$$

$$P(\mathbf{X} > x \mid \mathbf{X} \geq y) = P(\mathbf{X} > x - y + 1)$$

3. $P(\mathbf{X} \geq x) = P(\mathbf{X} > x - 1) = q^{x-1}$

   $P(\mathbf{X} \leq x) = 1 - P(\mathbf{X} > x) = 1 - q^x$