Διακριτές κατανομές

Πιθανότητες και Στατιστική

Τμήμα Πληροφορικής, Πανεπιστήμιο Πειραιώς

2023-2024

Κατανομή Bernoulli

Ορισμός

Μια διακριτή ΤΜ X λέμε ότι ακολουθεί την *κατανομή Bernoulli με παράμετρο* p και γράφουμε $X \sim \mathrm{Bern}(p)$, όπου $p \in (0,1)$, ανν έχει σύνολο τιμών το $S_X = \{0,1\}$ και PMF f_X, με

$$f_X(1) = P(X=1) = p, \text{ και } f_X(0) = P(X=0) = 1-p.$$

Άμεσα προκύπτουν η CDF (συνάρτηση κατανομής), η μέση τιμή και η διακύμανση:

$$F_X(x) = P(X \le x) = \begin{cases} 0, & x < 0, \\ 1-p, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

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

και

$$\sigma^2 = V(X) = E((X-p)^2) = (1-p)^2 p + (0-p)^2(1-p) = p(1-p).$$

Κατανομή Bernoulli

- Η κατανομή Bernoulli εμφανίζεται σε τυχαία φαινόμενα, όπου υπάρχουν δύο δυνατά αποτελέσματα (επιτυχία - αποτυχία).
- Κλασικό παράδειγμα ΤΜ που ακολουθεί την κατανομή Bernoulli είναι η έκβαση (επιτυχία: 1, αποτυχία: 0) της ρίψης ενός νομίσματος.
- Η κατανομή Bernoulli είναι η βάση για να ορίσουμε πολλές κατανομές.
- Σημειώνεται ότι η δείκτρια συνάρτηση I_A ενός ενδεχομένου A είναι μια ΤΜ Bernoulli με παράμετρο $p = P(A)$ (την πιθανότητα πραγματοποίησης του A).

Διωνυμική κατανομή

Εισαγωγικό παράδειγμα

Έστω ότι ρίχνουμε ένα νόμισμα N φορές για το οποίο γνωρίζουμε ότι $P(K) = p$ και $P(\Gamma) = 1 - p$ σε κάθε μια από τις ρίψεις. Να υπολογισθεί η PMF της TM

$$X = \text{αριθμός των ρίψεων που φέραμε Κορώνα.}$$

Λύση

Η X έχει σύνολο τιμών $S = \{0, 1, 2, \ldots, N\}$. Θεωρούμε τις TM X_i, $i \in [N]$, με

$$X_i = \begin{cases} 1, & \text{αν έρθει κορώνα στην } i\text{-οστή ρίψη,} \\ 0, & \text{αν έρθει γράμματα στην } i\text{-οστή ρίψη.} \end{cases}$$

Οι TM $X_1, X_2, \ldots, X_N$ είναι ανεξάρτητες και ακολουθούν την κατανομή Bernoulli με παράμετρο p και $X = X_1 + X_2 + \cdots + X_N$.

Διωνυμική κατανομή

Λύση (συνέχεια)

Κωδικοποιούμε την Κορώνα με 1 και τα Γράμματα με 0, οπότε, αν x_i είναι το αποτέλεσμα της i-οστής ρίψης, τότε η ακολουθία

$$x = x_1 x_2 \cdots x_N$$

είναι μια δυαδική λέξη με N γράμματα και το ενδεχόμενο $\{X = k\}$, όπου $k \in S$, πραγματοποιείται όταν η x περιέχει ακριβώς k μονάδες (και $N - k$ μηδενικά). Υπάρχουν $\binom{N}{k}$ τέτοιες δυαδικές λέξεις x και κάθε μία εμφανίζεται με πιθανότητα $p^k(1-p)^{N-k}$. Επομένως, η PMF της TM X είναι η

$$f_X(k) = P(X = k) = \binom{N}{k} p^k (1-p)^{N-k}, \;\; k = 0, 1, 2, \ldots, N.$$

Διωνυμική κατανομή

Ορισμός

Μια διακριτή ΤΜ X λέμε ότι ακολουθεί την *διωνυμική κατανομή με παραμέτρους* N *και* p, για κάποιο $N \in \mathbb{N}^*$ και $p \in (0,1)$, ανν έχει σύνολο τιμών $S_X = \{0, 1, 2, \ldots, N\}$ και PMF

$$f_X(k) = P(X = k) = \binom{N}{k} p^k (1-p)^{N-k}.$$

(Συμβολισμός: $X \sim \mathrm{Binom}(N, p)$, ή $X \sim \text{Διων}(N, p)$.)

Πρακτικά, η ΤΜ $X \sim \mathrm{Binom}(N, p)$ αντιπροσωπεύει το πλήθος επιτυχιών μετά από N ανεξάρτητες επαναλήψεις, όταν η πιθανότητα επιτυχίας είναι p.
Η συνάρτηση κατανομής της X δίνεται από τον τύπο

$$F_X(x) = P(X \le x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{N}{k} p^k (1-p)^{N-k}.$$

Διωνυμική κατανομή

Η μέση τιμή και η διακύμανση της X δίνονται από τους ακόλουθους τύπους:

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

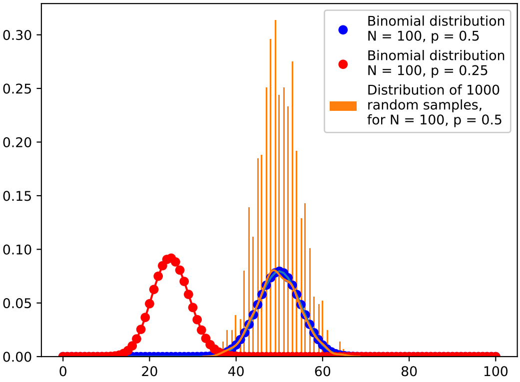
Πράγματι, επειδή $X = X_1 + X_2 + \cdots + X_N$ όπου $X_1, X_2, \ldots, X_N \sim \text{Bern}(p)$ και οι $X_1, X_2, \ldots, X_N$ είναι ανεξάρτητες, έπεται ότι

$$E(X) = E(X_1 + X_2 + \cdots + X_N) = E(X_1) + E(X_2) + \cdots + E(X_N) = Np.$$

και

$$V(X) = V(X_1 + X_2 + \cdots + X_N) = V(X_1) + V(X_2) + \cdots + V(X_N) = Np(1-p).$$

Διωνυμική κατανομή



Το παραπάνω σχήμα απεικονίζει γραφικά 2 διωνυμικές κατανομές (μπλε - κόκκινο) με διαφορετικές παραμέτρους και την κατανομή ενός τυχαίου δείγματος (πορτοκαλί) που παράχθηκε με βάση την (μπλε) διωνυμική κατανομή. Το σχήμα παράγεται από τον ακόλουθο κώδικα:

Διωνυμική κατανομή

```
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
import seaborn as sns

trials = 100 #N
P, P2 = 0.5, 0.25 #success probability
S = 1000 #sample size = #experiments
rs = np.random.binomial(n=trials, p=P, size=S) #draw a random sample of size S using
      numpy.random
#rs = stats.binom.rvs(n=trials, p=P, size=S) #draw a random sample of size S using scipy
      .stats
vals = np.arange(trials+1) #list of values 0,1,2,..., trials
f1 = stats.binom.pmf(vals, trials, P)
f2 = stats.binom.pmf(vals, trials, P2)

label1 = "Binomial distribution\n" + "N = " + str(trials) + ", p = " + str(P)
plt.plot(vals,f1, 'bo',label = label1)
plt.plot(vals,f1)
label2 = "Binomial distribution\n" + "N = " + str(trials) + ", p = " + str(P2)
plt.plot(vals,f2, 'ro', label = label2)
plt.plot(vals,f2, color='red')
label3 = "Distribution of " + str(S) + "\nrandom samples,\n" + "for N = " + str(trials)
     + ", p = " + str(P)
sns.distplot(rs, hist=True, kde=True, bins = trials+1, label = label3)
plt.legend()
plt.show()
```

Παράδειγμα

Για ένα αεροπλάνο 50 θέσεων έχουν γίνει 55 κρατήσεις (overbooking). Η πιθανότητα ο καθένας από αυτούς να εμφανισθεί στο αεροδρόμιο είναι 90% (ανεξάρτητα από τους υπολοίπους).

i) Ποιος είναι ο αναμενόμενος αριθμός επιβατών που θα εμφανισθούν για check-in;

ii) Ποια είναι η πιθανότητα κάποιοι επιβάτες να μείνουν εκτός πτήσης;

Λύση

Έστω X ο αριθμός των επιβατών που θα εμφανισθούν και έστω

$$X_i = \begin{cases} 1, & \text{αν εμφανισθεί ο } i\text{-οστός επιβάτης,} \\ 0, & \text{αλλιώς} \end{cases} \qquad i \in [55].$$

Τότε, $X = X_1 + X_2 + \cdots + X_{55}$, όπου $X_1, X_2, \ldots, X_{55} \sim \text{Bern}(0.9)$. Άρα, $X \sim \mathrm{Binom}(55, 0.9)$.

i) Κατά μέσο όρο, αναμένονται $E(X) = 55 \cdot 0.9 = 49.5$ επιβάτες.

ii) Η πιθανότητα κάποιοι επιβάτες να μείνουν εκτός πτήσης ισούται με

$$\begin{aligned} P(X > 50) &= P(X=51) + P(X=52) + P(X=53) + P(X=54) + P(X=55) \\ &= \sum_{k=51}^{55} \binom{55}{k} 0.9^k 0.1^{55-k} \\ &= 0.3451. \end{aligned}$$

Διωνυμική κατανομή

Παράδειγμα

Σε ένα τετρακινητήριο αεροπλάνο, το ενδεχόμενο ο i-οστός κινητήρας να πάθει βλάβη εν ώρα πτήσης είναι ανεξάρτητο από την πιθανότητα βλάβης των υπολοίπων. Μια πτήση είναι ασφαλής αν λειτουργούν τουλάχιστον οι δύο κινητήρες. Να βρεθεί η πιθανότητα μια πτήση να γίνει με ασφάλεια.

Λύση

Αν X το πλήθος των κινητήρων που έπαθαν βλάβη, τότε

$$\begin{aligned} F_X(2) = P(X \leq 2) &= \sum_{k=0}^{2} \binom{4}{k} p^k (1-p)^{4-k} \\ &= (1-p)^4 + 4p(1-p)^3 + 6p^2(1-p)^2 \end{aligned}$$

όπου p η πιθανότητα βλάβης. Για παράδειγμα, αν $p = 0.1$, τότε $F_X(2) = 0.9963$.

Διωνυμική κατανομή

Παράδειγμα

Ένας αδιάβαστος φοιτητής, ο οποίος απαντά στην τύχη, συμμετέχει σε μια εξέταση με 20 ερωτήσεις πολλαπλής επιλογής με 4 απαντήσεις η κάθεμια, και μόνο μια σωστή από αυτές.

i) Να βρεθεί η μέση τιμή των σωστών απαντήσεων.

ii) Να βρεθεί η πιθανότητα να δώσει ακριβώς 10 σωστές απαντήσεις.

Λύση.

Το πλήθος σωστών απαντήσεων X ακολουθεί διωνυμική κατανομή με παραμέτρους $N = 20$ και $p = 1/4$. Επομένως,

$$E(X) = Np = 5$$

και

$$P(X = 10) = \binom{N}{10} p^{10}(1-p)^{N-10} \simeq 0.00992. \qquad \square$$

Γεωμετρική κατανομή

Ορισμός

Μια ΤΜ X λέμε ότι ακολουθεί την γεωμετρική κατανομή με παράμετρο p, όπου $p \in (0,1)$, ανν έχει σύνολο τιμών $S_X = \{1,2,3,\ldots\} = \mathbb{N}^*$ και PMF

$$f_X(k) = P(X = k) = (1-p)^{k-1}p, \quad k \in \mathbb{N}^*.$$

(Συμβολισμός $X \sim \Gamma\varepsilon\omega\mu(p)$ ή $X \sim \mathrm{Geom}(p)$.)

Η X αντιπροσωπεύει τον αριθμό επαναλήψεων ενός πειράματος μέχρι την πρώτη επιτυχία, όταν η πιθανότητα επιτυχίας είναι ίση με p.
Η συνάρτηση κατανομής της X είναι η

$$F_X(x) = P(X \le x) = \sum_{i=1}^{\lfloor x \rfloor} (1-p)^{i-1}p = p \sum_{i=0}^{\lfloor x \rfloor - 1} (1-p)^i$$

$$= p\frac{1-(1-p)^{\lfloor x \rfloor}}{1-(1-p)} = 1-(1-p)^{\lfloor x \rfloor}$$

Γεωμετρική κατανομή

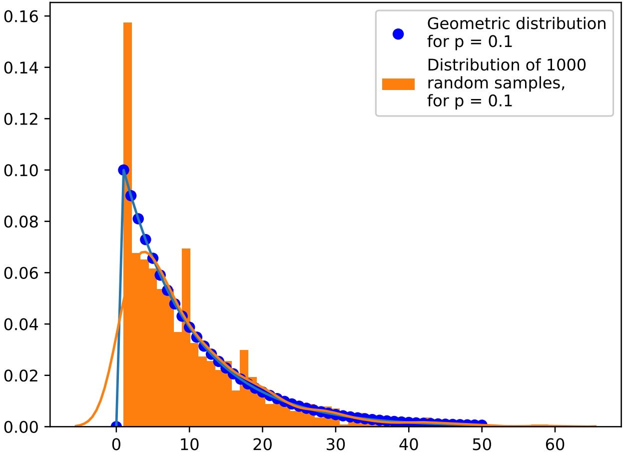
Η μέση τιμή και η διακύμανση δίνονται από τους τύπους

$$E(X) = \frac{1}{p}, \qquad V(X) = \frac{1-p}{p^2}$$

Η απόδειξη των παραπάνω τύπων βασίζεται στη γνωστή δυναμοσειρά

$$\frac{1}{1-p} = \sum_{k=0}^{\infty} p^k, \qquad |p| < 1$$

και τις 2 πρώτες παραγώγους αυτής και δίδεται στη λυμένη άσκηση 3.2.


Geometric distribution
for p = 0.1
Distribution of 1000
random samples,
for p = 0.1
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
10
20
30
40
50
60

Γεωμετρική κατανομή

Παράδειγμα

Ρίχνουμε ένα νόμισμα συνεχώς. Το νόμισμα έχει πιθανότητα p να φέρει Κορώνα και πιθανότητα $1-p$ να φέρει Γράμματα. Έστω X η ρίψη στην οποία θα εμφανισθεί για πρώτη φορά Κορώνα. Να βρεθεί η πιθανότητα $P(X=k)$.

Λύση

Αν θεωρήσουμε τη δείκτρια ΤΜ X_i για το ενδεχόμενο εμφάνισης κορώνας στην i-οστή ρίψη, $i \in \mathbb{N}^*$, τότε, λόγω ανεξαρτησίας, ισχύει ότι

$$
\begin{aligned}
P(X=k) &= P(X_1=0, X_2=0, \ldots, X_{k-1}=0, X_k=1) \\
&= P(X_1=0)P(X_2=0)\cdots P(X_{k-1}=0)P(X_k=1) \\
&\underbrace{(1-p)(1-p)\cdots(1-p)}_{k-1 \text{ φορές}}\, p = (1-p)^{k-1} p.
\end{aligned}
$$

Άρα, η ΤΜ X ακολουθεί την γεωμετρική κατανομή με παράμετρο p.

Παρατήρηση: Η γεωμετρική κατανομή ικανοποιεί την ιδιότητα

$$P(X \geq m + n | X > n) = P(X \geq m),$$

όπου το ``$\geq$'' μπορεί να αντικατασταθεί και από ``$=$'' ή ``$>$''.
Η ιδιότητα αυτή καλείται **έλλειψη μνήμης**.
Η απόδειξη ζητείται στην άλυτη άσκηση 3.7.
Πρακτικά, σύμφωνα με την ιδιότητα αυτή, αν έχουμε ήδη ρίξει ένα αμερόληπτο κέρμα n φορές και έφερε Γράμματα σε όλες, η πιθανότητα να φέρει Γράμματα στην επόμενη ρίψη είναι $1/2$, δηλαδή το παρελθόν δεν έχει σημασία.

Παράδειγμα

Κάποιος επιβάτης έχει υπολογίσει ότι η πιθανότητα να συναντήσει ελεγκτές εισιτηρίων σε μια διαδρομή με λεωφορείο είναι $p = 1/100$.

i) Να βρεθεί η πιθανότητα να κάνει 60 διαδρομές χωρίς να συναντήσει καθόλου ελεγκτές εισιτηρίων.

ii) Να βρεθεί ο μέσος αριθμός ελέγχων που θα γίνουν στον επιβάτη μέσα σε 60 διαδρομές.

iii) Αν το πρόστιμο για μια διαδρομή χωρίς εισιτήριο είναι 84 ευρώ, ποιο είναι το αναμενόμενο πρόστιμο για 60 διαδρομές χωρίς εισιτήριο;

Λύση

i) Έστω X ο αριθμός της διαδρομής στην οποία ο επιβάτης θα συναντήσει για πρώτη φορά ελεγκτές. Προφανώς, $X \sim \mathrm{Geom}(p)$, οπότε η πιθανότητα να μην συναντήσει ελεγκτές μέσα σε 60 διαδρομές ισούται με $P(X > 60) = 1 - F_X(60) = (1-p)^{60} = (99/100)^{60} = 0.552$.

Λύση (συνέχεια)

ii) Έστω Y ο αριθμός των ελέγχων που θα έχει ο επιβάτης μέσα σε 60 διαδρομές και έστω

$$Y_i = \begin{cases} 1, & \text{αν γίνει έλεγχος στην } i\text{-οστή διαδρομή} \\ 0, & \text{αλλιώς.} \end{cases}, i = 1, 2, \ldots, 60.$$

Ισχύει ότι $Y_i \sim \text{Bern}(p)$, για κάθε $i \in [60]$. Επειδή $Y = Y_1 + Y_2 + \cdots + Y_{60}$ και οι Y_i, $i \in [60]$, είναι ανεξάρτητες, έπεται ότι η $Y \sim \mathrm{Binom}(60, \mathrm{p})$. Επομένως,

$$E(Y) = 60 \cdot \frac{1}{100} = 0.6,$$

δηλαδή ο επιβάτης θα συναντήσει ελεγκτή κατά μέσο 6 φορές ανά 600 διαδρομές.

iii) Το αναμενόμενο πρόστιμο για 60 διαδρομές ισούται με $E(Y) \cdot 84 = 0.6 \cdot 84 = 50.4$ ευρώ.

Αρνητική διωνυμική κατανομή

Ορισμός

Μια διακριτή ΤΜ X λέμε ότι ακολουθεί την αρνητική διωνυμική κατανομή με παραμέτρους r, p και γράφουμε $X \sim NB(r, p)$, ανν έχει σύνολο τιμών $S_X = \{r, r+1, r+2, \dots\}$ και PMF

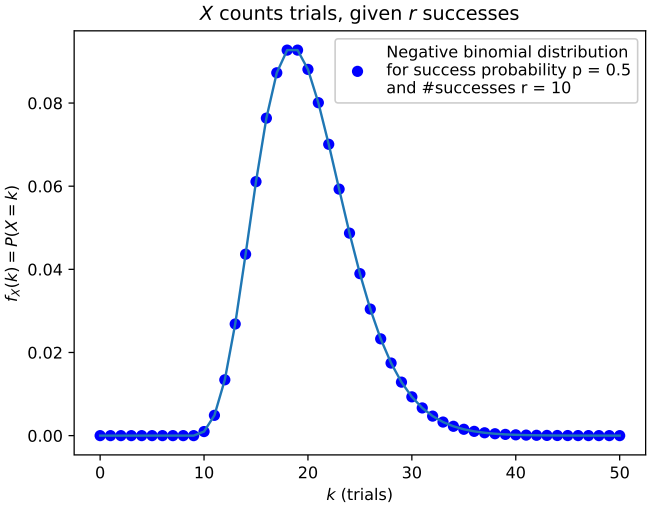
$$f_X(k) = P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \qquad 0 < p < 1, r \in \mathbb{N}^*.$$

Η X αντιπροσωπεύει τον αριθμό επαναλήψεων ενός πειράματος μέχρι την r-οστή επιτυχία, όταν η πιθανότητα επιτυχίας είναι ίση με p.
Η μέση τιμή και η διακύμανση είναι αντίστοιχα ίσες με

$$E(X) = \frac{r}{p} \qquad \text{και} \qquad V(X) = \frac{r(1-p)}{p^2}.$$

Η απόδειξη των παραπάνω τύπων δίνεται στη λυμένη άσκηση 3.3.

Αρνητική διωνυμική κατανομή



Αρνητική διωνυμική κατανομή

Παρατήρηση: Έστω Y το πλήθος των αποτυχιών μέχρι την r-οστή επιτυχία. Προφανώς, είναι $S_Y = \mathbb{N}$ και $X = Y + r$. Η PMF της Y είναι η

$$f_Y(y) = \binom{y+r-1}{r-1} p^r (1-p)^y \overset{k=y+r}{=} \binom{k-1}{r-1} p^r (1-p)^{k-r} = f_X(k),$$

άρα

$$f_X(k) = f_Y(k-r).$$

Ορισμένες φορές στη βιβλιογραφία, όπως και στη βιβλιοθήκη scipy.stats της Python, θεωρείται η Y (και όχι η X) ως η ΤΜ που ακολουθεί την $\mathrm{NB}(r, p)$, οπότε το τελευταίο σχήμα παράγεται καλώντας την PMF της βιβλιοθήκης ως εξής:

```
stats.nbinom.pmf(k-r, r, p)
```

Ορισμός

Η διακριτή ΤΜ X λέμε ότι ακολουθεί τη υπεργεωμετρική κατανομή με παραμέτρους M, n, N και γράφουμε $Y \sim \mathrm{HGeom(M, n, N)}$, ανν έχει σύνολο τιμών $S_X = [\max\{0, N-b\}, \min\{N, a\}]$, $a, b, N \in \mathbb{N}$ και PMF

$$f_X(k) = P(X = k) = \frac{\binom{a}{k}\binom{b}{N-k}}{\binom{a+b}{N}} = \frac{\binom{n}{k}\binom{M-n}{N-k}}{\binom{M}{N}}, \quad k \in S_X, n = a, M = a+b.$$

Η X αντιπροσωπεύει τον αριθμό άσπρων (καλών) αντικειμένων που επιλέγονται μετά από N επιλογές **χωρίς επανατοποθέτηση** μεταξύ $a = n$ άσπρων και $b = M - n$ μαύρων (κακών) αντικειμένων. Η μέση τιμή και η διακύμανση είναι αντίστοιχα ίσες με

$$E(X) = \frac{Na}{a+b} \qquad \text{και} \qquad V(X) = \frac{Nab(a+b-N)}{(a+b)^2(a+b-1)}.$$

Παρατήρηση: Αθροίζοντας για όλες τις δυνατές τιμές του k, προκύπτει ότι

$$\sum_k \binom{a}{k}\binom{b}{N-k} = \binom{a+b}{N} \sum_k P(X=k) = \binom{a+b}{N}.$$

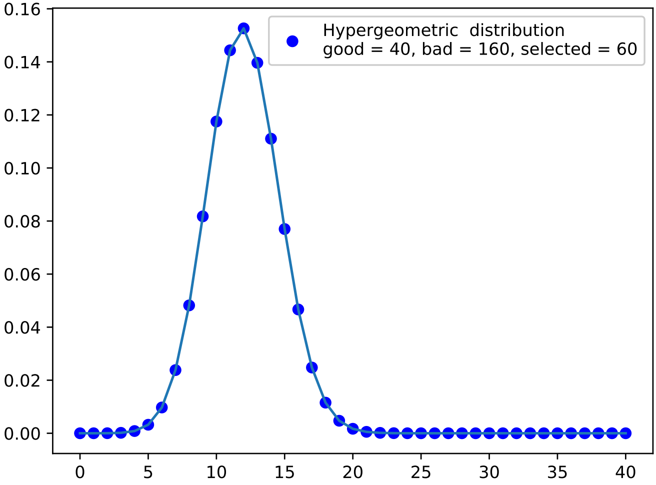
Η παραπάνω σχέση είναι γνωστή ως ταυτότητα Vandermonde ή τύπος του Cauchy και με τη βοήθειά της αποδεικνύονται οι παραπάνω εκφράσεις για τη μέση τιμή και τη διακύμανση.

Προσέγγιση υπεργεωμετρικής κατανομής από τη διωνυμική κατανομή: Αν $a/M = n/M \to p$ καθώς $M = a+b \to +\infty$, τότε

$$\mathrm{HGeom(M,n,N)} \to \mathrm{Binom}(N,p), \qquad \text{καθώς } M \to \infty.$$

Η απόδειξη αφήνεται ως άσκηση. Πρακτικά, η υπεργεωμετρική κατανομή αντιστοιχεί στο παραπάνω πείραμα χωρίς επανατοποθέτηση, ενώ η διωνυμική αντιστοιχεί στο ίδιο πείραμα αλλά με επανατοποθέτηση. Όταν ο συνολικός πληθυσμός M είναι πολύ μεγάλος, η επανατοποθέτηση επηρεάζει ελάχιστα την πιθανότητα επιλογής του κάθε ατόμου.

Υπεργεωμετρική κατανομή



Κατανομή Poisson

Ορισμός

Μια ΤΜ X λέμε ότι ακολουθεί την κατανομή Poisson με παράμετρο $\lambda > 0$, αν $S_X = \mathbb{N}$ και έχει PMF

$$f_X(k) = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

(Συμβολισμός: $X \sim \text{Poisson}(\lambda)$.)

Η μέση τιμή και η διακύμανση της X είναι αντίστοιχα ίσες με

$$E(X) = \lambda \qquad \text{και} \qquad V(X) = \lambda.$$

Πράγματι, είναι

$$E(X) = \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$
$$= e^{-\lambda}\lambda e^{\lambda} = \lambda$$

και

$$V(X) = E(X^2) - (E(X))^2 = -\lambda^2 + \sum_{k=0}^{\infty} k^2 P(X=k)$$
$$= -\lambda^2 + e^{-\lambda}\sum_{k=0}^{\infty} k^2\frac{\lambda^k}{k!} = -\lambda^2 + \lambda e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^{k-1}}{(k-1)!}$$
$$= -\lambda^2 + \lambda e^{-\lambda}\sum_{k=0}^{\infty}(k+1)\frac{\lambda^k}{k!} = -\lambda^2 + \lambda e^{-\lambda}\left(\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} + \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\right)$$
$$= -\lambda^2 + \lambda e^{-\lambda}\left(\lambda e^{\lambda} + e^{\lambda}\right) = \lambda.$$

Η κατανομή Poisson ονομάζεται και νόμος των μικρών αριθμών. Χρησιμοποιείται ως μοντέλο για τυχαία φαινόμενα όπως

- ο αριθμός των τροχαίων που συμβαίνουν σε μια περιοχή σε κάποιο χρονικό διάστημα
- ο αριθμός των πελατών που φτάνουν σε ένα κατάστημα σε κάποιο διάστημα χρόνου
- ο αριθμός των πλοίων που φτάνουν σε ένα λιμάνι σε κάποιο διάστημα χρόνου
- ο αριθμός των τυπογραφικών λαθών που εμφανίζονται σε κάποια σελίδα ενός βιβλίου
- ο αριθμός των ψαριών (poisson στα Γαλλικά) που πιάνονται στα δίκτυα μας σε κάποιο χρονικό διάστημα.

Προσέγγιση της διωνυμικής κατανομής από την κατανομή Poisson: Αν μια ΤΜ X ακολουθεί την Binom(n, p), τότε για μεγάλες τιμές του n και μικρές τιμές του p, η X προσεγγίζεται από μια ΤΜ η οποία ακολουθεί την κατανομή Poisson(np), δηλαδή

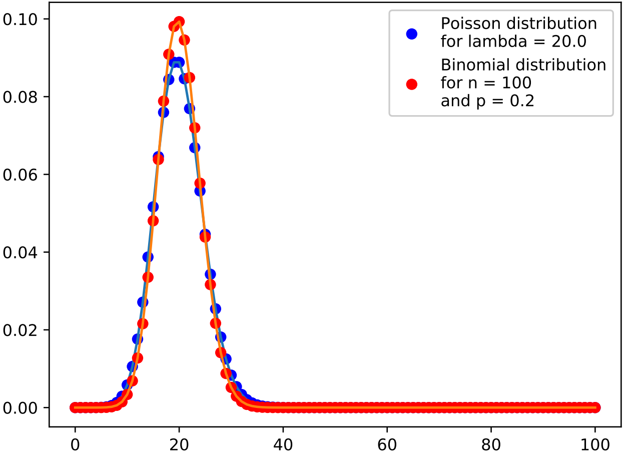
$$\text{Αν } \lim_{n\to\infty} np = \lambda \in \mathbb{R}, \text{ τότε } \lim_{n\to\infty} \text{Binom}(n, p) = \text{Poisson}(\lambda).$$

(Η απόδειξη δίνεται στην λυμένη άσκηση 3.5.)
Στις εφαρμογές, χρησιμοποιείται η προσέγγιση αυτή για την απλοποίηση των υπολογισμών, δηλαδή αν $X \sim \mathrm{Binom}(n, p)$, τότε

$$P(X = k) = f_X(k) \approx f(k),$$

όπου f η PMF μιας ΤΜ Poisson με $\lambda = np$. Η προσέγγιση αυτή είναι πολύ καλή, όταν το n είναι μεγάλο και το p μικρό.

Κατανομή Poisson



Παράδειγμα

Ένας server δέχεται κατά μέσο όρο 5 αιτήσεις από clients ανά δευτερόλεπτο.

i) Ποια είναι η πιθανότητα να δεχθεί 5 αιτήσεις στο επόμενο δευτερόλεπτο;

ii) Ποια είναι η πιθανότητα να δεχθεί 1000 αιτήσεις στο επόμενο λεπτό;

Λύση

i) Αν X το πλήθος των αιτήσεων σε ένα δευτερόλεπτο, τότε $X \sim \mathrm{Poisson}(\lambda)$, όπου $\lambda = 5$. Επομένως, η ζητούμενη πιθανότητα είναι

$$P(X = 5) = e^{-5} \frac{5^5}{5!} \simeq 0.17546737.$$

Λύση (συνέχεια)

ii) Αν X_i το πλήθος των αιτήσεων στο i-οστό δευτερόλεπτο του λεπτού, όπου $i \in [t]$ και $t = 60$, τότε $X_i \sim \mathrm{Poisson}(\lambda)$, με $\lambda = 5$.
Οι ΤΜ X_i είναι ανεξάρτητες, οπότε το πλήθος αιτήσεων στο λεπτό είναι

$$X = \sum_{i=1}^{t} X_i \sim \mathrm{Poisson}(\lambda t).$$

(Η απόδειξη ζητείται στην άλυτη άσκηση 3.8.)
Κατόπιν τούτου, είναι

$$P(X = 1000) = e^{-300} \frac{300^{1000}}{1000!}$$

Ο αριθμός αυτός προσεγγίζεται πιο εύκολα με τη βοήθεια της κανονικής κατανομής που θα δούμε στα επόμενα.

Παράδειγμα

Η πιθανότητα ένας άνθρωπος να έχει αλλεργία σε ένα εμβόλιο είναι $p = 1/1000$. Κάνουμε ένεση σε $N = 2000$ άτομα. Να βρεθεί η πιθανότητα να παρουσιάσουν αλλεργία πάνω από 2 άτομα.

Λύση

Έστω X ο αριθμός των ατόμων που θα πάθουν αλλεργία. Προφανώς, $X \sim \mathrm{Binom}(N, p)$. Για την ζητούμενη πιθανότητα $P(X > 2)$, ισχύει ότι

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) = 1 - P(X = 2) - P(X = 1) - P(X = 0) \\ &= 1 - \sum_{i=0}^{2} \binom{N}{i} p^i (1-p)^{N-i} = 0.3233. \end{aligned}$$

Λύση (συνέχεια)

Μπορούμε να προσεγγίσουμε την ΤΜ X από μια ΤΜ $Y \sim \text{Poisson}(\lambda)$, όπου $\lambda = 2000 \cdot \frac{1}{1000} = 2$, δηλαδή

$$\begin{aligned} P(X > 2) \simeq P(Y > 2) &= 1 - P(Y \leq 2) \\ &= 1 - P(Y = 2) - P(Y = 1) - P(Y = 0) \\ &= 1 - e^{-2}\frac{2^2}{2!} - e^{-2}\frac{2^1}{1!} - e^{-2}\frac{2^0}{0!} \\ &= 1 - 5e^{-2} = 0.3233. \end{aligned}$$

Παρατήρηση: Η προσέγγιση της ΤΜ X από την Y είναι πολύ καλή. Παρ' όλα αυτά, οι πιθανότητες που υπολογίσαμε δεν ταυτίζονται απόλυτα. Συγκεκριμένα, οι αντίστοιχες πιθανότητες με περισσότερη ακρίβεια είναι 0.3233235612 και 0.3233235838.

Παράδειγμα

Έστω ότι για κάποιο λαχείο εκδίδονται $2 \cdot 10^6$ δελτία, από τα οποία 100 έχουν σημαντικό κέρδος.
Πόσα δελτία πρέπει να αγοράσουμε ώστε με πιθανότητα τουλάχιστον 95% να εξασφαλίσουμε κάποιο σημαντικό κέρδος;

Λύση

Η πιθανότητα ένα δέλτιο να κερδίσει είναι $p = \dfrac{10^2}{2 \cdot 10^6} = \dfrac{1}{2 \cdot 10^4}$.
Έστω ότι αγοράζουμε N δελτία και έστω X ο αριθμός των δελτίων που κερδίζουν ανάμεσα στα N δελτία που αγοράσαμε.
Η X ακολουθεί την διωνυμική κατανομή με παραμέτρους N και p.
Μπορούμε να την προσεγγίσουμε χρησιμοποιώντας κατανομή Poisson με παράμετρο $\lambda = Np = \dfrac{N}{2 \cdot 10^4}$.

Λύση (συνέχεια)

Θέλουμε

$$
\begin{aligned}
P(X \geq 1) \geq 0.95 \Leftrightarrow P(X < 1) \leq 0.05 \Leftrightarrow P(X = 0) \leq 0.05 \\
&\Leftrightarrow e^{-\lambda}\frac{\lambda^0}{0!} \leq 0.05 \\
&\Leftrightarrow e^{\lambda} \geq 20 \\
&\Leftrightarrow \lambda = \frac{N}{2 \cdot 10^4} \geq \ln 20 \\
&\Leftrightarrow N \geq 2 \cdot 10^4 \ln 20 = 59914.6.
\end{aligned}
$$

Άρα, πρέπει να αγοράσουμε τουλάχιστον $N = 59615$ δελτία.

Λυμένες ασκήσεις

Άσκηση

Από τα 10^7 άτομα ενός πληθυσμού, οι 10^5 έχουν οπτική ίνα στο σπίτι. Από αυτόν τον πληθυσμό επιλέγουμε τυχαία (με επανατοποθέτηση) 150 άτομα για να συμμετάσχουν σε μια έρευνα αγοράς.
Τότε το πλήθος Y των ατόμων με οπτική ίνα ανάμεσα στους 150 επιλεχθέντες ακολουθεί την διωνυμική κατανομή με παραμέτρους $N = 150$ και $p = 10^5/10^7 = 0.01$.

Λύση

Θεωρούμε ένα τυχαία επιλεγμένο άτομο και θέτουμε A το ενδεχόμενο να έχει οπτική ίνα και B το ενδεχόμενο να επιλεχθεί στους $N = 150$ της έρευνας. Είναι

$$p := P(A|B) = P(A \cap B)/P(B) = P(A)P(B)/P(B) = P(A) = 1/100.$$

Επομένως, κάθε άτομο που συμμετέχει στην έρευνα έχει οπτική ίνα με πιθανότητα p και άρα $Y \sim \mathrm{Binom}(N, p)$.

Άσκηση

Αν $X_1 \sim \mathrm{Geom}(p_1)$ και $X_2 \sim \mathrm{Geom}(p_2)$ ανεξάρτητες, να ευρεθεί η PMF της $Y = \min\{X_1, X_2\}$.

Λύση

Για $k \in \mathbb{N}^*$, είναι

$$\begin{aligned} P(Y = k) &= P(X_1 = k)P(X_2 \geq k) + P(X_1 > k)P(X_2 = k) \\ &= (1 - p_1)^{k-1} p_1 (1 - p_2)^{k-1} + (1 - p_2)^{k-1} p_2 (1 - p_1)^k \\ &= ((1 - p_1)(1 - p_2))^{k-1} (p_1 + p_2(1 - p_1)) \\ &= ((1 - p_1)(1 - p_2))^{k-1} (1 - (1 - p_1)(1 - p_2)) \\ &= (1 - q)^{k-1} q, \end{aligned}$$

όπου $q = 1 - (1 - p_1)(1 - p_2)$. Επομένως, $Y \sim \mathrm{Geom}(q)$.

Άσκηση

Να αποδειχθεί ότι αν $X_1 \sim \mathrm{Binom}(N_1, p)$ και $X_2 \sim \mathrm{Binom}(N_2, p)$ ανεξάρτητες, τότε $X_1 + X_2 \sim \mathrm{Binom}(N_1 + N_2, p)$.

Λύση

Το ενδεχόμενο $\{X_1 + X_2 = n\}$ διαμερίζεται στα $\{X_1 = k, X_2 = n - k\}$, με $k \in \{0, 1, 2, \ldots, n\}$, οπότε (και λόγω ανεξαρτησίας των X_1, X_2)

$$P(X_1+X_2=n) = \sum_{k=0}^{n} P(X_1=k, X_2=n-k) = \sum_{k=0}^{n} P(X_1=k)P(X_2=n-k)$$

$$= \sum_{k=0}^{n} \binom{N_1}{k} p^k (1-p)^{N_1-k} \binom{N_2}{n-k} p^{n-k} (1-p)^{N_2-n+k}$$

$$= p^n (1-p)^{N_1+N_2-n} \sum_{k=0}^{n} \binom{N_1}{k} \binom{N_2}{n-k} \stackrel{1}{=} p^n (1-p)^{N_1+N_2-n} \binom{N_1+N_2}{n}$$

[1] Από την ταυτότητα Vandermonde.

Άσκηση

Να αποδειχθεί ότι αν $X_1 \sim \mathrm{Poisson}(\lambda_1)$ και $X_2 \sim \mathrm{Poisson}(\lambda_2)$ ανεξάρτητες, τότε $X_1 + X_2 \sim \mathrm{Poisson}(\lambda_1 + \lambda_2)$.

Λύση

Το ενδεχόμενο $\{X_1 + X_2 = n\}$ διαμερίζεται στα $\{X_1 = k, X_2 = n - k\}$, με $k \in \{0, 1, 2, \ldots, n\}$, οπότε (και λόγω ανεξαρτησίας των X_1, X_2)

$$
\begin{aligned}
P(X_1 + X_2 = n) &= \sum_{k=0}^{n} P(X_1 = k, X_2 = n - k) = \sum_{k=0}^{n} P(X_1 = k)P(X_2 = n - k) \\
&= \sum_{k=0}^{n} e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k} = e^{-(\lambda_1+\lambda_2)} \frac{1}{n!} (\lambda_1 + \lambda_2)^n
\end{aligned}
$$

Άσκηση

Να αποδειχθεί ότι αν $X \sim \mathrm{Geom}(p)$ και $m, n \in \mathbb{N}^*$, τότε

$$P(X \geq m + n | X > n) = P(X \geq m).$$

Λύση

Είναι

$$\begin{aligned} P(X \geq m + n | X > n) &= \frac{P(X \geq m + n, X > n)}{P(X > n)} = \frac{P(X \geq m + n)}{P(X > n)} \\ &= \frac{(1-p)^{m+n-1}}{(1-p)^n} = (1-p)^{m-1} = P(X \geq m). \end{aligned}$$

Άσκηση

Έστω ΤΜ $X \sim \mathrm{Binom}(n, p)$, $p \in (0, 1)$. Να αποδειχθεί ότι

i) $f_X(k+1) = \dfrac{p}{1-p}\dfrac{n-k}{k+1} f_X(k)$, για $k \in \{0, 1, 2, \dots, n-1\}$.

ii) Η $f_X(k)$ είναι αύξουσα για $k \leq \lfloor (n+1)p \rfloor$ και φθίνουσα για $k \geq \lfloor (n+1)p \rfloor$.

Λύση

Για $k \in \{0, 1, 2, \dots, n-1\}$, είναι

$$\frac{f_X(k+1)}{f_X(k)} = \frac{\binom{n}{k+1} p^{k+1}(1-p)^{n-k-1}}{\binom{n}{k} p^k (1-p)^{n-k}}$$

$$= \frac{p}{1-p}\frac{n!}{(k+1)!(n-k-1)!}\frac{k!(n-k)!}{n!} = \frac{p}{1-p}\frac{n-k}{k+1}.$$

Λύση (συνέχεια)

Επομένως,

$$\frac{f_X(k+1)}{f_X(k)} \geq 1 \Leftrightarrow \frac{p}{1-p}\frac{n-k}{k+1} \geq 1 \Leftrightarrow pn - pk \geq k + 1 - pk - p$$
$$\Leftrightarrow k+1 \leq (n+1)p \Leftrightarrow k+1 \leq \lfloor (n+1)p \rfloor$$

και ομοίως,

$$\frac{f_X(k+1)}{f_X(k)} < 1 \Leftrightarrow k+1 \geq \lfloor (n+1)p \rfloor + 1.$$

Άρα, για $m = \lfloor (n+1)p \rfloor$, είναι

$$f_X(0) < \cdots < f_X(m-1) \leq f_X(m) > f_X(m+1) > \cdots > f_X(n),$$

με την ισότητα να ισχύει ανν $(n+1)p \in \mathbb{N}^*$.

Η τιμή m ονομάζεται επικρατούσα τιμή (mode) της κατανομής.

Λυμένες ασκήσεις

Άσκηση

Ένα πείραμα έχει r δυνατά αποτελέσματα, τα $1, 2, \ldots, r$, με πιθανότητες $p_1, p_2, \ldots, p_r$ αντίστοιχα, όπου $\sum_{i=1}^{r} p_i = 1$. Αν το πείραμα επαναλαμβάνεται n φορές και X_i είναι το πλήθος των επαναλήψεων με αποτέλεσμα $i \in [r]$,

i) Ποια είναι η κατανομή της X_1;

ii) Είναι οι X_1 και X_2 ανεξάρτητες;

iii) Ποια είναι η κατανομή της $X_1 + X_2$;

iv) Για $k < r$, ποια είναι η κατανομή της $X = \sum_{i=1}^{k} X_i$;

Λύση

i) Επειδή το κάθε πείραμα έχει πιθανότητα επιτυχίας p_1, ως προς το αποτέλεσμα 1, έπεται ότι $X_1 \sim \mathrm{Binom}(n, p_1)$.
ii) Οι X_1, X_2 δεν είναι ανεξάρτητες, αφού για παράδειγμα είναι $P(X_1 = 1 | X_2 = n) = 0 \neq P(X_1 = 1)$.

Λύση (συνέχεια)

iii) Έστω $Y = X_1 + X_2$ και έστω Y_i η δείκτρια ΤΜ του ενδεχομένου η i-οστή επανάληψη να έχει αποτέλεσμα 1 ή 2. Προφανώς, οι Y_i, $i \in [n]$, είναι ανεξάρτητες, ακολουθούν την $\mathrm{Bernoulli}(p_1 + p2)$ και $Y = Y_1 + \cdots + Y_n$, οπότε $Y \sim \mathrm{Binom}(n, p_1 + p_2)$.

iv) Ομοίως με το (*iii*), θέτοντας $Y = \sum_{i=1}^{k} X_i$, για $k < r$, και Y_i την δείκτρια ΤΜ του ενδεχομένου η i-οστή επανάληψη να έχει αποτέλεσμα από 1 έως k, προκύπτει ότι

$$Y \sim \mathrm{Binom}(n, p_1 + p_2 + \cdots + p_k).$$

Άσκηση

Το πλήθος των φορών που ένας άνθρωπος αρρωσταίνει το χρόνο με κρύωμα ακολουθεί κατανομή Poisson με παράμετρο $\lambda = 3$.
Ένα νέο φάρμακο μειώνει αυτή τη συχνότητα σε $\lambda = 2$ για το 75% του πληθυσμού, ενώ δεν έχει αποτέλεσμα στο υπόλοιπο 25%.
Αν ένας άνθρωπος που πήρε το φάρμακο αρρώστησε 0 φορές κατά τη διάρκεια του έτους, ποια η πιθανότητα να τον ωφέλησε το φάρμακο;

Λύση

Έστω X το πλήθος φορών που αρρώστησε το άτομο σε ένα χρόνο και έστω A το ενδεχόμενο να τον ωφέλησε το φάρμακο. Δίνονται

$$P(X=0|A)=e^{-2},\quad P(X=0|\overline{A})=e^{-3},\quad p(A)=3/4.$$

Εφαρμόζοντας τον τύπο του Bayes, προκύπτει ότι

$$\begin{aligned} P(A|X=0) &= \frac{P(X=0|A)P(A)}{P(X=0|A)P(A)+P(X=0|\overline{A})P(\overline{A})} \\ &= \frac{3e^{-2}/4}{3e^{-2}/4+e^{-3}/4} \\ &= \frac{3}{3+1/e} \approx 0.89. \end{aligned}$$

Άσκηση

Ένα ηλεκτρονικό κατάστημα παραλαμβάνει ένα φορτίο με 100 ηλεκτρονικές συσκευές. Ελέγχει 10 τυχαία επιλεγμένες από αυτές και αν βρει πάνω από μία ελαττωματική, τότε επιστρέφει το φορτίο στον προμηθευτή. Αν το φορτίο περιέχει 20 ελαττωματικές συσκευές, ποια είναι η πιθανότητα να μην επιστραφεί;

Λύση

Αν X είναι το πλήθος των μη ελαττωματικών από τις 10 που ελέγχονται, τότε, κατά τα γνωστά, $X \sim \mathrm{HGeom}(M, n, N)$, με $M = 100, n = 80, N = 10$.
Η ζητούμενη πιθανότητα είναι η $P(X \geq 9) = 1 - F_X(8) \approx 0.363$.

Το τελικό αποτέλεσμα μπορεί να υπολογισθεί με οποιαδήποτε από τις παρακάτω εντολές:

```
1-stats.hypergeom.cdf(8, 100, 80, 10)
stats.hypergeom.sf(8, 100, 80, 10)
```

Παρατήρηση: Αν στο προηγούμενο αποτέλεσμα χρησιμοποιηθεί η προσέγγιση

$$\mathrm{HGeom}(M, n, N) \to \mathrm{Binom}(N, n/M),$$

τότε βρίσκουμε

$$P(X \geq 9) \approx 0.3758096384,$$

αντί για το ακριβές αποτέλεσμα

$$P(X \geq 9) = 0.3630494342076581.$$

Άσκηση

Ο Γιώργος και ο Δημήτρης είναι συγκάτοικοι και κάθε φορά που έχουν να πλύνουν πιάτα ρίχνουν ένα νόμισμα. Όποιος φέρει πρώτος κορώνα κερδίζει και ο άλλος πλένει τα πιάτα. Ο Γιώργος παρατήρησε ότι όταν ρίχνει πρώτος κερδίζει συχνότερα, ενώ όταν ξεκινάει δεύτερος καταλήγει συχνότερα στο νεροχύτη. Υπολογίστε τις πιθανότητες και στις δυο περιπτώσεις. Είναι ο ισχυρισμός του Γιώργου σωστός;

Λύση

Έστω X ο αριθμός ρίψεων μέχρι να έρθει κορώνα για πρώτη φορά. Η X ακολουθεί την γεωμετρική κατανομή με παράμετρο $p = 1/2$, οπότε

$$P(X = k) = (1 - p)^{k-1} p,$$

για κάθε $k \in \mathbb{N}^*$.

Λύση (συνέχεια)

Όταν ο Γιώργος ξεκινάει πρώτος, τότε κερδίζει όταν $X = 1, 3, 5, 7, \ldots$, με πιθανότητα

$$\sum_{k=0}^{\infty} P(X = 2k+1) = \sum_{k=0}^{\infty} (1-p)^{2k} p = p \sum_{k=0}^{\infty} ((1-p)^2)^k = \frac{p}{1-(1-p)^2}$$
$$= \frac{p}{p(2-p)} = \frac{1}{2-p},$$

ενώ, όταν ξεκινάει δεύτερος, τότε κερδίζει όταν $X = 2, 4, 6, 8, \ldots$, με πιθανότητα

$$\sum_{k=1} P(X = 2k) = \sum_{k=1} (1-p)^{2k-1} p = \frac{p}{1-p} \sum_{k=1}^{\infty} ((1-p)^2)^k$$
$$= \frac{p}{1-p} \frac{(1-p)^2}{1-(1-p)^2} = \frac{1-p}{2-p}.$$

Λύση (συνέχεια)

Επομένως, η πιθανότητα να κερδίσει είναι μεγαλύτερη στην πρώτη περίπτωση, όταν ξεκινά πρώτος.

Ειδικά για $p = \dfrac{1}{2}$, οι αντίστοιχες πιθανότητες είναι

$$\frac{1}{2-\frac{1}{2}} = \frac{2}{3} \qquad \text{και} \qquad \frac{1-\frac{1}{2}}{2-\frac{1}{2}} = \frac{1}{3},$$

δηλαδή αυτός που αρχίζει να ρίχνει το νόμισμα πρώτος κερδίζει 2 στις 3 φορές.

Άσκηση

Οι ημερήσιες συχνότητες επιθέσεων σε ένα site τον πρώτο χρόνο λειτουργίας του δίδονται στον επόμενο πίνακα

Αριθμός επιθέσεων	0	1	2	3	4
Συχνότητα (ημέρες)	223	110	27	4	1

Αν υποθέσουμε ότι ο αριθμός X των επιθέσεων που γίνονται στο site μέσα σε μια ημέρα ακολουθεί την κατανομή Poisson, να βρεθεί η παράμετρος λ η οποία προσεγγίζει τις παραπάνω παρατηρήσεις. Στην συνέχεια, να βρεθεί με βάση τον τύπο της κατανομής Poisson ο αναμενόμενος αριθμός ημερών στις οποίες θα έχουμε 0, 1, 2, 3, 4 επιθέσεις αντίστοιχα μέσα στον επόμενο χρόνο.
(Υπόδειξη: Να υπολογισθεί ο μέσος αριθμός επιθέσεων σε μια μέρα με βάση τα στοιχεία του πίνακα.)

Λύση

Ο μέσος αριθμός επιθέσεων μέσα σε μια μέρα είναι

$$\mu = \frac{0 \cdot 223 + 1 \cdot 110 + 2 \cdot 27 + 3 \cdot 4 + 4 \cdot 1}{365} = \frac{36}{73} = 0.493.$$

Επομένως, $X \sim \mathrm{Poisson}(\lambda)$. Άρα με $\lambda = \mu = 36/73 = 0.493 \approx 0.5$.

$P(X = 0) = e^{-0.5}\frac{(0.5)^0}{0!} = 0.606,$ $\quad P(X = 1) = e^{-0.5}\frac{(0.5)^1}{1!} = 0.303.$

$P(X = 2) = e^{-0.5}\frac{(0.5)^2}{2!} = 0.075,$ $\quad P(X = 3) = e^{-0.5}\frac{(0.5)^3}{3!} = 0.012.$

$P(X = 4) = e^{-0.5}\frac{(0.5)^3}{4!} = 0.0015.$

Επομένως, μέσα σε ένα χρόνο (365 ημέρες), ο μέσος αριθμός ημερών με 0, 1, 2, 3 και 4 επιθέσεις θα είναι αντίστοιχα

$365 \cdot P(X = 0) = 221.38 \approx 221$, $365 \cdot P(X = 1) = 110.69 = 111$,

$365 \cdot P(X = 2) = 27.67 \approx 28$, $365 \cdot P(X = 3) = 4.61 \approx 5$ και

$365 \cdot P(X = 4) = 0.57 \approx 1$.

Άσκηση

Είναι γνωστό ότι οι βίδες που παράγει μια συγκεκριμένη εταιρεία είναι ελαττωματικές με πιθανότητα 0.01, ανεξάρτητα η μια από την άλλη. Η εταιρεία πουλάει τις βίδες σε συσκευασίες των 10 και προσφέρει εγγύηση επιστροφής χρημάτων αν υπάρχει πάνω από 1 ελαττωματική βίδα σε μια συσκευασία. Το ποσοστό των πουλημένων συσκευασιών πρέπει να αντικαταστήσει η εταιρεία;

Άσκηση

Στο λεγόμενο παιχνίδι του τροχού της τύχης ρίχνονται 3 αμερόληπτα ζάρια (ισοδύναμα περιστρέφεται ένας τρόχος όπου σε κάθε φέτα εμφανίζονται όλες οι πιθανές διατεταγμένες τριάδες αριθμών από το 1 έως το 6). Κάθε παίχτης στοιχηματίζει a χρηματικές μονάδες σε ένα αριθμό από το 1 έως το 6. Αν κανένα από τα ζάρια δεν φέρει αυτόν τον αριθμό τότε χάνει το ποσό που στοιχημάτισε, αλλιώς κερδίζει το ποσό ka όπου k είναι ο αριθμός των φορών που εμφανίστηκε στα 3 ζάρια ο αριθμός που στοιχημάτισε.
Να εξετασθεί αν το παιχνίδι είναι δίκαιο για τον παίχτη.

Άσκηση

Υποθέστε ότι το πλήθος των τυπογραφικών λαθών σε μια σελίδα ενός μαθηματικού βιβλίου έχει κατανομή Poisson με παράμετρο $\lambda = 2$.
α) Υπολογίστε την πιθανότητα να υπάρχει τουλάχιστον ένα λάθος σε μια συγκεκριμένη σελίδα
β) Υπολογίστε την πιθανότητα να υπάρχουν τουλάχιστον 3 λάθη σε 2 σελίδες.

Άσκηση

Ένας έμπορος ηλεκτρικών εξαρτημάτων αγοράζει εξαρτήματα σε συσκευασίες των 10. Η πολιτική του είναι να επιθεωρεί τυχαία 3 εξαρτήματα από κάθε συσκευασία και να δέχεται να την αγοράσει αν και μόνο αν και τα 3 δεν είναι ελαττωματικά.
Αν το 30 τοις εκατό των συσκευασιών έχει ακριβώς 4 ελαττωματικά εξαρτήματα και 70 τοις εκατό έχει μόνο 1, τι πόσοστο συσκευασιών αναμένεται να απορρίψει ο έμπορος;

Άσκηση

Έστω ότι ο αριθμός των γκολ που πετυχαίνει η Εθνική ομάδα ποδοσφαίρου στα εκτός έδρας παιχνίδια ακολουθεί την κατανομή Poisson. Επίσης, είναι γνωστό ότι στα εκτός έδρας παιχνίδια η Εθνική έχει την ίδια πιθανότητα να πετύχει ένα ή δύο γκολ. Να βρεθεί η πιθανότητα στο επόμενο εκτός έδρας παιχνίδι η Εθνική να πετύχει τέσσερα γκολ.

Λύση.

Έστω X ο αριθμός των γκόλ που πετυχαίνει η εθνική. Γνωρίζουμε ότι η X ακολουθεί την κατανομή Poisson με παράμετρο λ και έχει PMF $f_X(k) = P(X = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$. Επίσης,

$$P(X = 1) = P(X = 2) \Leftrightarrow e^{-\lambda}\frac{\lambda^1}{1!} = e^{-\lambda}\frac{\lambda^2}{2!} \Leftrightarrow \lambda = 2.$$

Επομένως,

$$P(X = 4) = e^{-2}\frac{2^4}{4!} = \frac{2}{3e^2} = 0.09. \qquad \square$$