Κατανομές διακριτών τυχαίων μεταβλητών

Καθηγητής Χρήστος Γ. Μασούρος

Υπεργεωμετρική κατανομή

Ας θεωρήσουμε μια κάλπη που περιέχει Λ λευκές και Μ μαύρες σφαίρες. Εξάγουμε τυχαία **άνευ επαναθέσεως** ν σφαίρες. Έστω Χ το πλήθος των λευκών σφαιρών στο δείγμα των ν σφαιρών. Τότε λέμε ότι το Χ έχει την **υπεργεωμετρική** κατανομή. Η συνάρτηση πιθανότητας της Χ δίνεται από τον τύπο:

Υπεργεωμετρική κατανομή

$$P\left[X=\kappa\right]=\frac{\binom{\Lambda}{\kappa}\binom{M}{\nu-\kappa}}{\binom{\Lambda+\mathrm{M}}{\nu}}$$

Παράδειγμα

Σε μια δέσμη 50 τεμαχίων ενός προϊόντος τα 5 είναι ελαττωματικά. Εκλέγουμε τυχαία 3 τεμάχια από τη δέσμη.

α) ποια η πιθανότητα να υπάρχει ακριβώς ένα ελαττωματικό τεμάχιο στο δείγμα;

β) ποια η πιθανότητα τουλάχιστον ένα τεμάχιο να είναι ελαττωματικό;

γ) ποια η πιθανότητα το πολύ ένα τεμάχιο να είναι ελαττωματικό;

α) Έχουμε:

P (ένα ελαττωματικό)=

$$= \frac{\binom{5}{1}\binom{45}{2}}{\binom{50}{3}} = \frac{99}{392}$$

β) *P* (τουλάχιστον ένα ελαττωματικό)=

= 1-*P* (κανένα ελαττωματικό)=

$$= 1 - \frac{\binom{5}{0}\binom{45}{3}}{\binom{50}{3}} = \frac{541}{1960}$$

γ) *P* (το πολύ ένα ελαττωματικό)=

=*P* (κανένα ελαττωματικό) + *P* (ένα ελαττωματικό)=

$$=\frac{\binom{5}{0}\binom{45}{3}}{\binom{50}{3}}+\frac{\binom{5}{1}\binom{45}{2}}{\binom{50}{3}}=\frac{1419}{1960}+\frac{99}{392}=\frac{1914}{1960}$$

Δοκιμές Bernoulli

Μια ακολουθία ανεξαρτήτων δοκιμών, σε κάθε μια από τις οποίες διακρίνουμε δύο δυνατά αποτελέσματα (ενδεχόμενα) *A* (**επιτυχία**) και *A*΄ (**αποτυχία**) καλείται ***ακολουθία ανεξαρτήτων δοκιμών Bernoulli.***

Συνηθίζουμε να συμβολίζουμε με *p*, *q* τις πιθανότητες εμφάνισης των ενδεχομένων Α και Α΄ αντίστοιχα. Είναι φανερό ότι ***p+q=1***

Δυωνυμική κατανομή

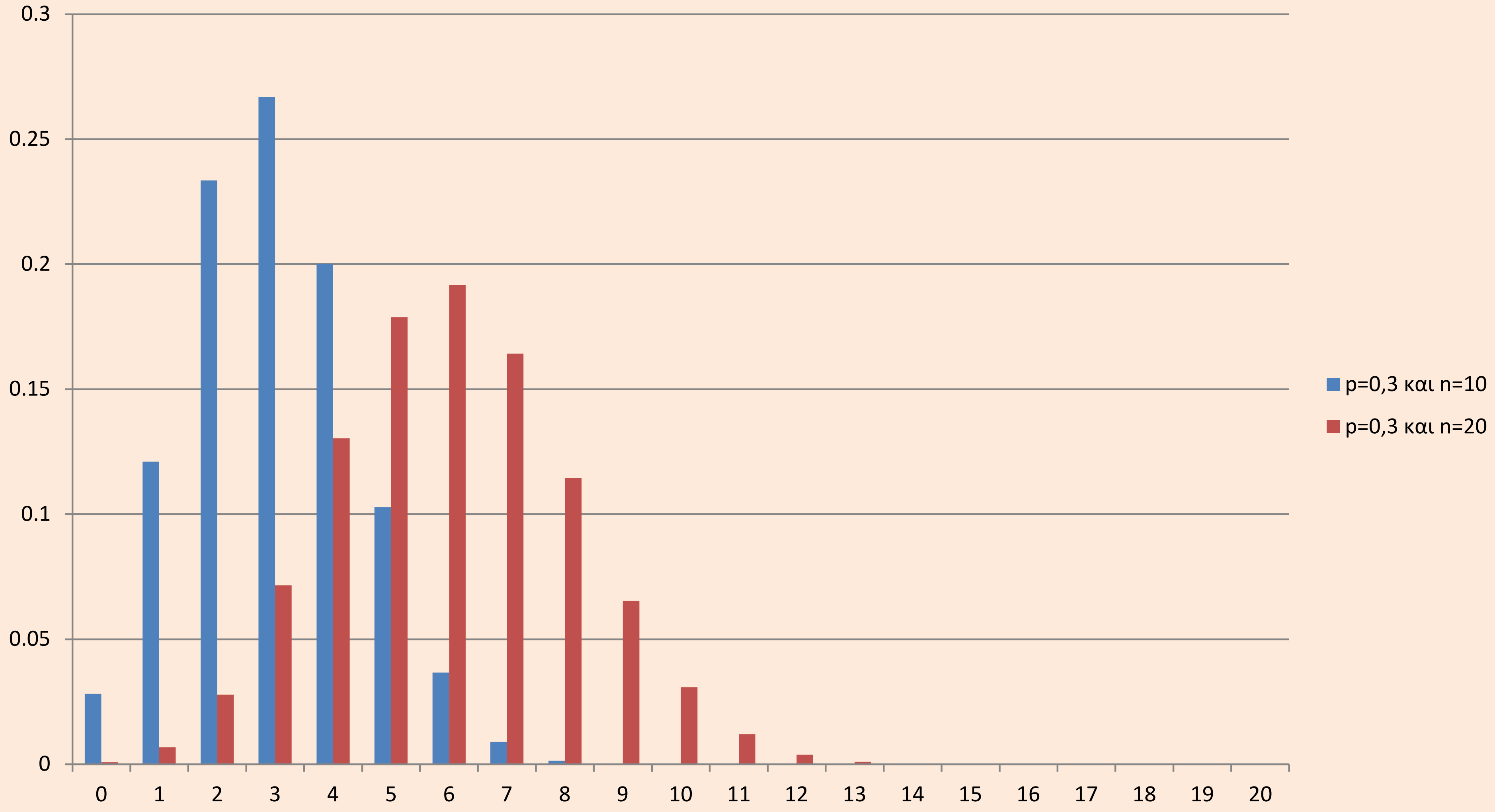
Ας θεωρήσουμε n δοκιμές Bernoulli, και έστω X ο αριθμός των επιτυχιών. Τότε λέμε ότι η X ακολουθεί την **δυωνυμική κατανομή**. Η *συνάρτηση πιθανότητας* της δυωνυμικής κατανομής:

P (k «επιτυχίες» στις n δοκιμές Bernoulli με πιθανότητα επιτυχίας p)

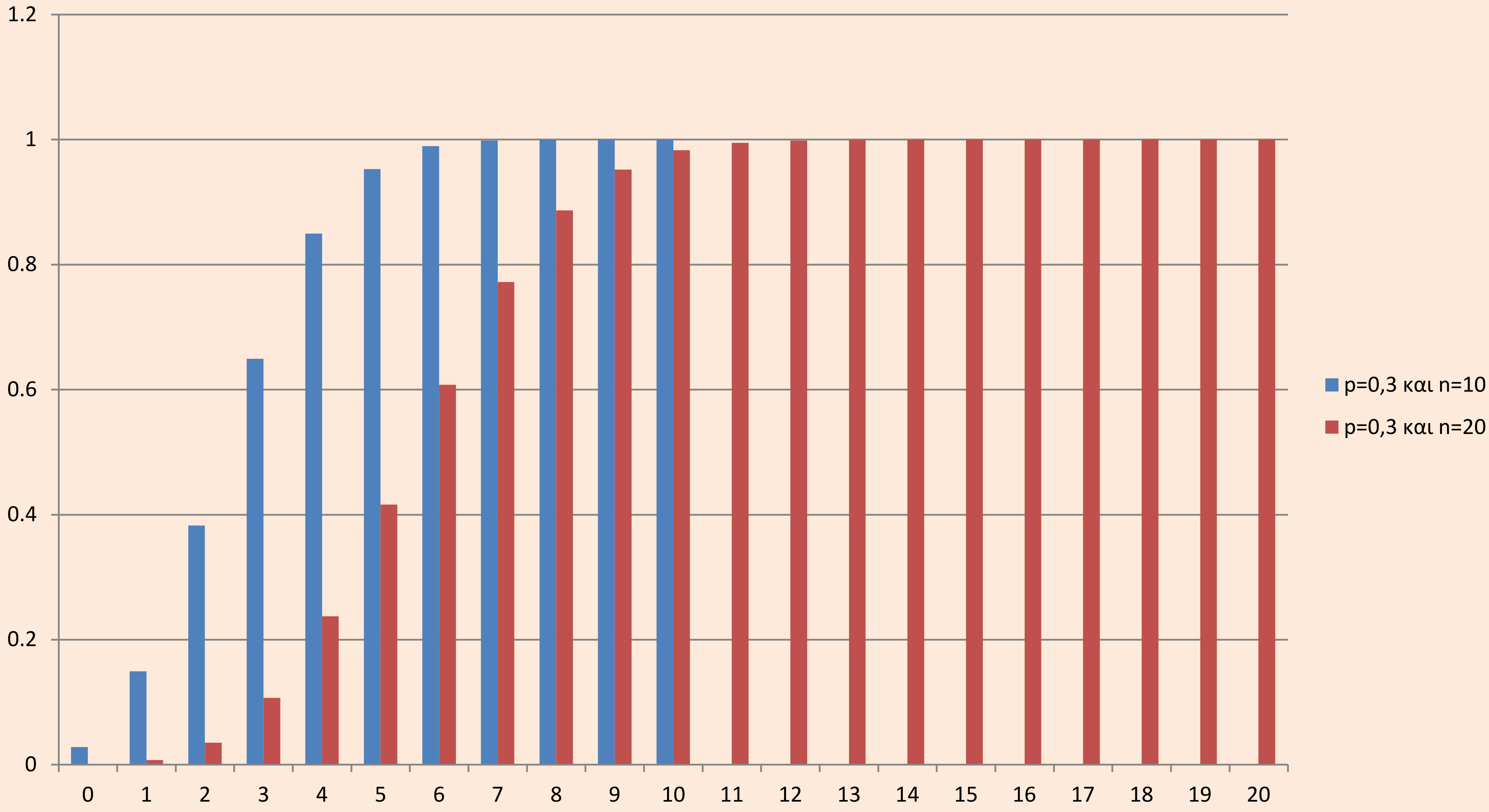
δίνεται από τον τύπο:

$$\mathrm{b}(\mathrm{k},\mathrm{n},\mathrm{p}) = \mathrm{P}(\mathrm{X}=k) = \binom{\mathrm{n}}{\mathrm{k}} \mathrm{p}^{\mathrm{k}}(1-\mathrm{p})^{\mathrm{n}-\mathrm{k}}\ ,\quad \mathrm{k}=0,1,\dots,\mathrm{n}$$

Γραφικές παραστάσεις των συναρτήσεων πιθανότητας για τις b(k, 10, 0,3) και b(k, 20, 0,3)



Γραφικές παραστάσεις των συναρτήσεων αθροιστικής πιθανότητας για τις b(k, 10, 0,3) και b(k, 20, 0,3)



Παραδείγματα

Ρίπτουμε ένα νόμισμα 10 φορές. Ποια η πιθανότητα να εμφανισθούν γράμματα 6 φορές;

$$b\left(6,10,\frac{1}{2}\right)=\binom{10}{6}\left(\frac{1}{2}\right)^{6}\left(\frac{1}{2}\right)^{10-6}=0,2051$$

Μια βιομηχανία παράγει εξαρτήματα από τα οποία το 10% είναι ελαττωματικά. Επιλέγουμε τυχαία 20 εξαρτήματα από την γραμμή παραγωγής.

α) Ποια η πιθανότητα να υπάρχουν 2 ελαττωματικά στο δείγμα που επιλέξαμε στην τύχη

$$b\left(2,20,\frac{10}{100}\right)=\binom{20}{2}\left(\frac{1}{10}\right)^{2}\left(\frac{9}{10}\right)^{20-2}=0,28518$$

β) Ποια η πιθανότητα να υπάρχουν το πολύ 2 ελαττωματικά στο δείγμα που επιλέξαμε στην τύχη;

$$F(2) = b\left(0, 20, \frac{10}{100}\right) + b\left(1, 20, \frac{10}{100}\right) + b\left(2, 20, \frac{10}{100}\right) =$$

$$= \binom{20}{0}\left(\frac{1}{10}\right)^{0}\left(\frac{9}{10}\right)^{20} + \binom{20}{1}\left(\frac{1}{10}\right)^{1}\left(\frac{9}{10}\right)^{19} + \binom{20}{2}\left(\frac{1}{10}\right)^{2}\left(\frac{9}{10}\right)^{18} =$$

$$= 0{,}12158 + 0{,}27017 + 0{,}28518 = 0{,}67639$$

γ) Ποια η πιθανότητα να υπάρχουν τουλάχιστον 2 ελαττωματικά στο δείγμα που επιλέξαμε στην τύχη;

$$1 - b\left(0, 20, \frac{10}{100}\right) - b\left(1, 20, \frac{10}{100}\right) = 1 - 0{,}12158 - 0{,}27017 = 0{,}60825$$

δ) Ποια η πιθανότητα να υπάρχουν ελαττωματικά σε μια συσκευασία που επιλέγουμε στην τύχη;

$$1-b\left(0,20,\frac{10}{100}\right)=1-\binom{20}{0}\left(\frac{1}{10}\right)^{0}\left(\frac{9}{10}\right)^{20}=1-1\cdot 1\cdot\left(\frac{9}{10}\right)^{20}=$$

$$=1-0,12158=0,8784$$

Μέση τιμή και διακύμανση της διωνυμικής κατανομής

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

- $\mu = E(X) = \nu p$
- $\sigma^2 = V(X) = \nu p q$

Γεωμετρική κατανομή

Ας θεωρήσουμε μια ακολουθία δοκιμών Bernoulli, η οποία τερματίζεται όταν εμφανισθεί επιτυχία για πρώτη φορά. Έστω X ο αριθμός των απαιτουμένων δοκιμών. Τότε λέμε ότι η X ακολουθεί την **γεωμετρική κατανομή**. Το X λαμβάνει κάθε θετική ακέραια τιμή, και ισχύει ο τύπος:

$$p_\upsilon = P[X=\upsilon] = (1-p)^{\upsilon-1}p = q^{\upsilon-1}p$$

δεδομένου ότι οι πρώτες υ-1 δοκιμές πρέπει να δώσουν **αποτυχία** η δε ν-οστη **επιτυχία**.

Παράδειγμα

Η πιθανότητα να βρεθεί ελεύθερο ταξί σε κεντρικό δρόμο μιας πόλης σε ώρα αιχμής είναι 10%. Ποια η πιθανότητα να επιβιβασθούμε στο 7ο ταξί που θα περάσει από μπροστά μας;

Απάντηση: Καθώς ένα διερχόμενο ταξί είναι ελεύθερο ή κατειλημμένο, έχουμε μια ακολουθία δοκιμών Bernoulli, η οποία τερματίζεται όταν εμφανισθεί για πρώτη φορά ελεύθερο ταξί, στο οποίο και επιβιβαζόμαστε. Έχουμε συνεπώς μια **γεωμετρική κατανομή** με $\upsilon=7$, $p=0{,}1$ και $q=0{,}9$:

$$P[X=\upsilon] = q^{\upsilon-1}p \Rightarrow$$

$$P[X=7] = 0{,}9^{7-1}0{,}1 = 0{,}053$$

Γεωμετρική κατανομή

Για την συνάρτηση κατανομής της Γεωμετρικής κατανομής έχουμε:

$$F(k)=p\sum_{n=1}^{k}q^{n-1}=p\left(1+q+...+q^{k-1}\right)=p\frac{1-q^k}{1-q}=p\frac{1-q^k}{p}=1-q^k$$

Επομένως:

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

Παράδειγμα

Ένα ζάρι ρίχνεται συνεχώς μέχρι να εμφανισθεί άσσος. Ποια η πιθανότητα να συμβεί αυτό

- α) στην $10^{η}$ ρίψη
- β) πριν από την $10^{η}$ ρίψη
- γ) μετά την $10^{η}$ ρίψη

α) $p_{10} = P[X=10] = (1-\frac{1}{6})^{10-1}\frac{1}{6}$ = **3,2%**

β) $P[X\leq 9] = F(9) = 1-(\frac{5}{6})^{9} \approx$ **80.6%**

γ) $P[X>10] = 1-P[X \leq 10] = 1-F(10) =$

$$= 1 - [1 - (\frac{5}{6})^{10}] = (\frac{5}{6})^{10} \approx \textbf{16,2\%}$$

Κατανομή Pascal (αρνητική διωνυμική)

Θεωρούμε μια ακολουθία (ανεξαρτήτων) 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** ή **κατανομή του Pascal**, με παραμέτρους r και p και συμβολίζεται $Nb(r,b)$. Η συνάρτηση πιθανότητας αυτής της κατανομής δίνεται από τον τύπο:

$$f(x) = P[X = x] = \binom{x-1}{r-1} p^r q^{x-r}$$

Παράδειγμα

Πολεμικό πλοίο βάλει εναντίον εχθρικού σκάφους με πυροβόλο, οι βολές του οποίου έχουν πιθανότητα να επιτύχουν τον στόχο ίση προς ¾. Αν απαιτούνται 4 επιτυχείς βολές για να βυθισθεί το αντίπαλο πλοίο, ποια η πιθανότητα να χρειασθούν τουλάχιστον 6 βολές;

Αν συμβολίσουμε με X τον αριθμό των βολών που θα απαιτηθούν μέχρι να βληθεί 4 φορές ο αντίπαλος στόχος, τότε η τυχαία μεταβλητή X ακολουθεί την κατανομή ***Nb*(4, ¾)**, και συνεπώς:

$$P[X=x]=\binom{x-1}{4-1}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^{x-4}, \quad x=4,5,\ldots$$

Οπότε

$$P[X\geq 6]=1-P[X=4]-P[X=5]=$$

$$=1-\binom{4-1}{4-1}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^{4-4}-\binom{5-1}{4-1}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^{5-4}\cong 0,37$$

Κατανομή Poisson

- Αναφερόμαστε σε μια τυχαία μεταβλητή X που μετρά τον αριθμό των εμφανίσεων ενός φαινομένου σε κάποιο συγκεκριμένο διάστημα (χρονικό διάστημα, ή διάστημα μήκους, κλπ) .
- Τα σημαντικά στοιχεία είναι:
- 1) η X παίρνει τιμές αρχίζοντας από 0 χωρίς να έχει θεωρητικά ένα άνω όριο,
- 2) πρέπει να αναφερόμαστε σε ένα συγκεκριμένο διάστημα και να είναι ξεκάθαρο πόσο είναι αυτό, ένας χρόνος, μία μέρα, μία ώρα, μία σελίδα. Πάντα πρέπει να γνωρίζουμε σε ποιο διάστημα ακριβώς αναφέρεται ο αριθμός των εμφανίσεων του φαινομένου που μετρά η X.

Κατανομή Poisson

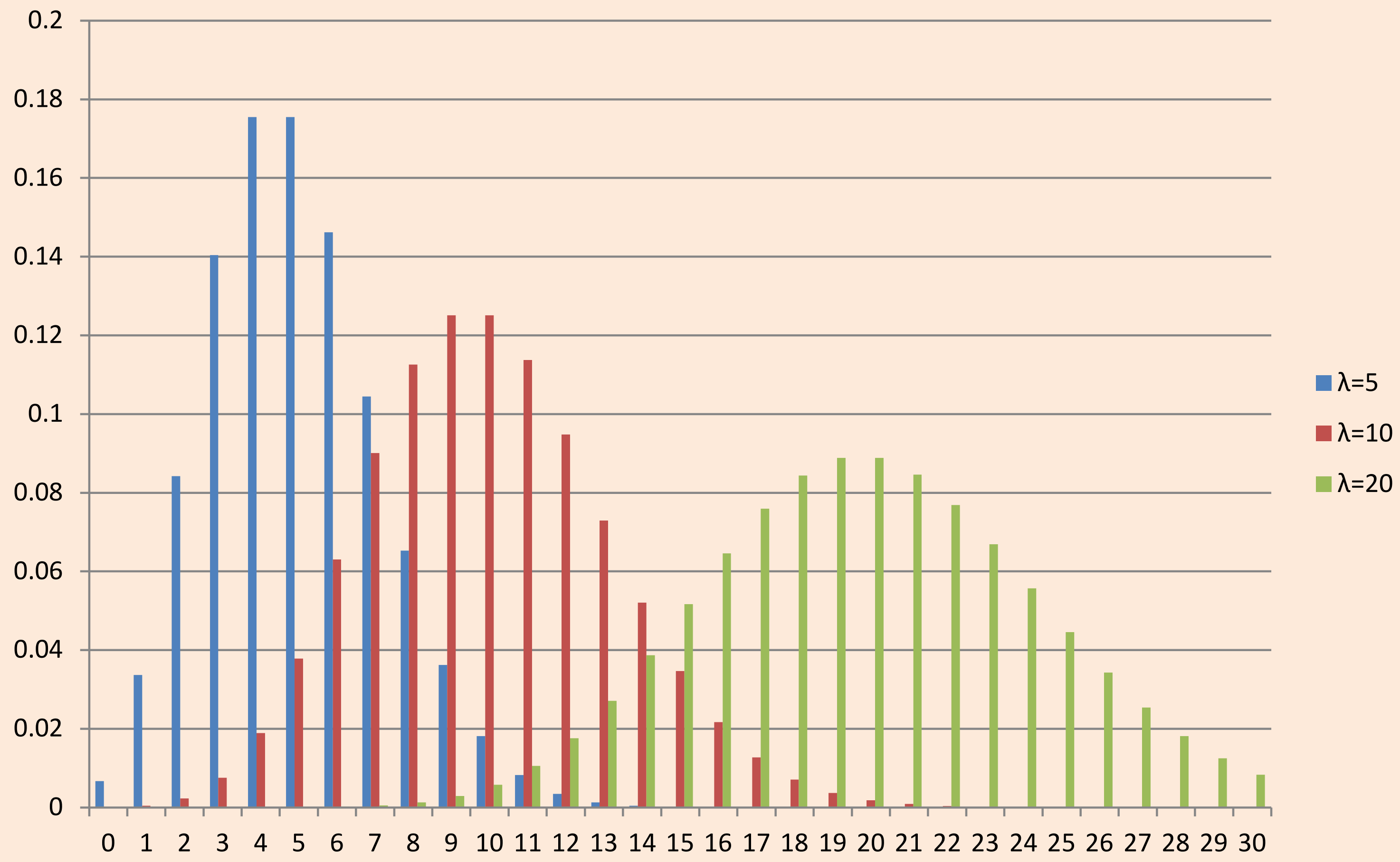
- Η ***συνάρτηση πιθανότητας*** της κατανομής Poisson είναι:

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

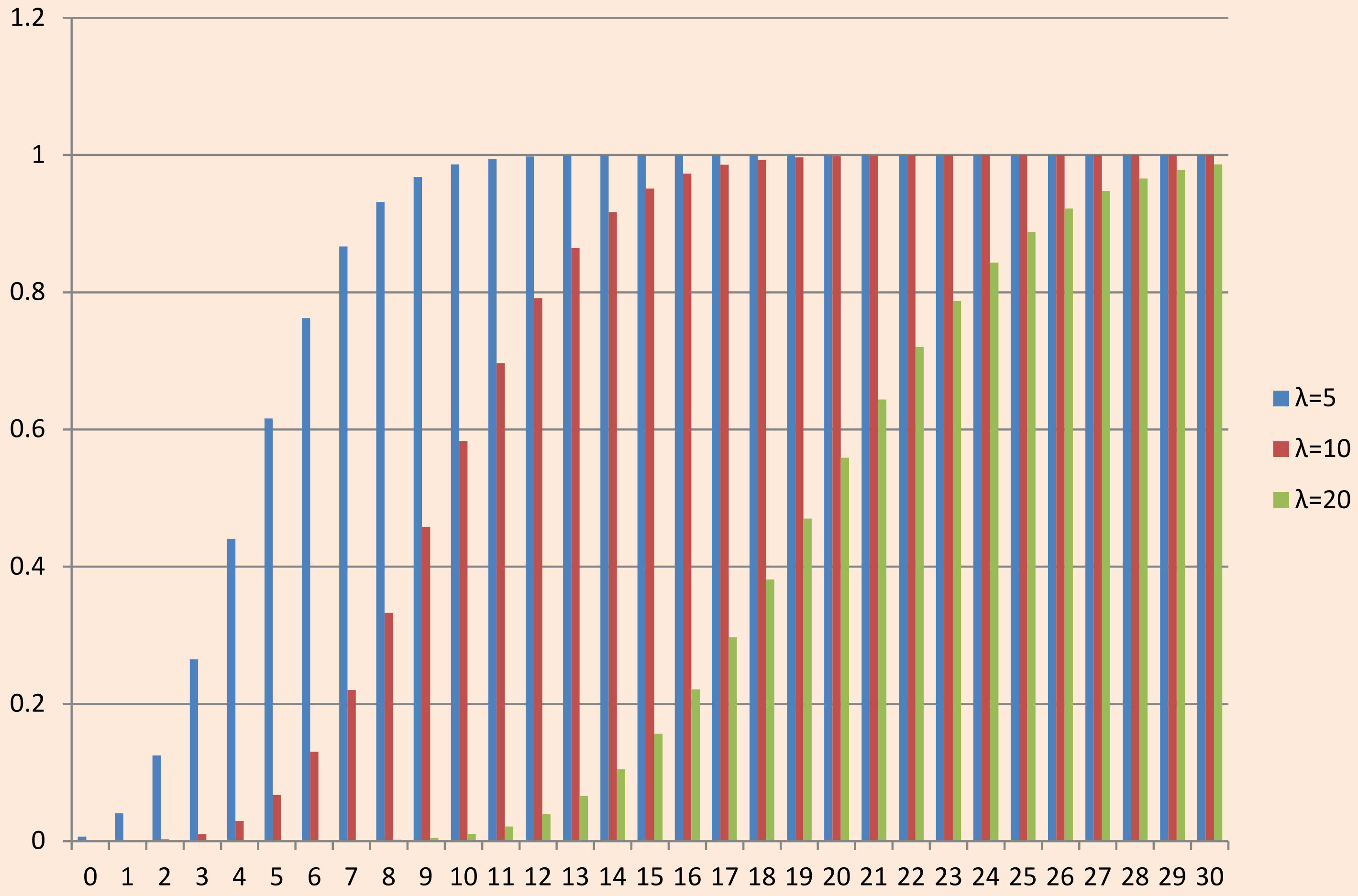
- Η ***αθροιστική συνάρτηση πιθανότητα*** (ή ***συνάρτηση κατανομής***) να έχουμε μέχρι και x εμφανίσεις ισούται με

$$P(X \leq x) = \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^{k}}{k!}, \quad x = 0,1,2,\ldots$$

Γραφικές παραστάσεις των συναρτήσεων πιθανότητας Poisson με λ=5, 10 και 20



Γραφικές παραστάσεις των συναρτήσεων αθροιστικής πιθανότητας Poisson με λ=5, 10 και 20



Μέση τιμή και διακύμανση της κατανομής Poisson

- Η μέση τιμής μιας τυχαίας μεταβλητής X που ακολουθεί την κατανομή Poisson είναι

$$\mu = \mathrm{E}(\mathrm{X}) = \lambda$$

- Ενώ η διακύμανση είναι

$$\sigma^2 = \mathrm{Var}(\mathrm{X}) = \lambda$$

Υπολογισμός πιθανότητας Poisson

Ο αριθμός *X* των τυπογραφικών λαθών σε μια σελίδα ενός βιβλίου ακολουθεί την κατανομή Poisson με παράμετρο λ = 5. Να υπολογιστεί η πιθανότητα μια σελίδα να περιέχει 2 ακριβώς λάθη.

Κάνοντας χρήση της κατανομής Poisson για λ=5 θα έχουμε:

$$\mathrm{P}(\mathrm{X} = 2) = \frac{\mathrm{e}^{-5} 5^2}{2!} = 0{,}084$$

Παράδειγμα

Μια βιοτεχνία καθαρισμού ρούχων λειτουργεί καθημερινά 8 ώρες. Η βιοτεχνία δέχεται κατά μέσο όρο 24 παραγγελίες την ημέρα για καθαρισμό ενδυμάτων.

- **(ι)** Να υπολογισθεί η πιθανότητα να δεχτεί 4 παραγγελίες την επόμενη ώρα.

Οι παραγγελίες ακολουθούν την κατανομή Poisson. Σύμφωνα με τα δεδομένα ο μέσος αριθμός παραγγελιών είναι 24 παραγγελίες ανά ημέρα, δηλαδή **3** παραγγελίες την ώρα. Με βάση τα παραπάνω έχουμε:

$$P(4 \text{ παραγγελίες την ώρα}) = P(X=4 \mid \lambda = 3) = \frac{e^{-3} \cdot 3^4}{4!} = 0{,}168$$

Άρα η ζητούμενη πιθανότητα είναι 16,8%.

Παράδειγμα

(ιι) Να υπολογισθεί η πιθανότητα να δεχτεί το πολύ 2 παραγγελίες την επόμενη ώρα.

P (το πολύ 2 παραγγελίες την ώρα)=

$$=P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

$$P(X=0 \mid \lambda=3) = \frac{e^{-3} \cdot 3^0}{0!} = 0{,}0498$$

$$P(X=1 \mid \lambda=3) = \frac{e^{-3} \cdot 3^1}{1!} = 0{,}1494$$

$$P(X=2 \mid \lambda=3) = \frac{e^{-3} \cdot 3^2}{2!} = 0{,}2240$$

Επομένως P (το πολύ 2 παραγγελίες την ώρα) =

= 0,0498 + 0,1494 + 0,2240 = 0,4232.

Άρα η ζητούμενη πιθανότητα είναι **42,32%**.

Παράδειγμα

- **(ιιι)** Αν υποθέσουμε ότι το πλυντήριο καθαρισμού έπαθε κάποια βλάβη και χρειάστηκαν 2 ώρες να επισκευαστεί,

  (α) ποιος είναι ο αναμενόμενος αριθμός παραγγελιών στο διάστημα αυτό, και
  (β) ποια είναι η πιθανότητα η βιοτεχνία να έχει δεχθεί τουλάχιστον 3 παραγγελίες στο συγκεκριμένο δίωρο;

(iii-α) Ο αναμενόμενος αριθμός παραγγελιών στις 2 ώρες ισούται με το μέσο της κατανομής Poisson για το διάστημα αυτό. Δηλαδή, αναμένουμε λ= 3·2 = 6 παραγγελίες.

(iii-β) η πιθανότητα η βιοτεχνία να έχει δεχθεί τουλάχιστον 3 παραγγελίες στο συγκεκριμένο δίωρο είναι

P (τουλάχιστον 3 παραγγελίες το δίωρο) $= P(X \geq 3) = 1 - P(X=0) - P(X=1) - P(X=2)$

Όμως

$$P(X = 0 \mid \lambda = 6) = \frac{e^{-6} \cdot 6^0}{0!} = 0{,}0025$$

$$P(X = 1 \mid \lambda = 6) = \frac{e^{-6} \cdot 6^1}{1!} = 0{,}0149$$

$$P(X = 2 \mid \lambda = 6) = \frac{e^{-6} \cdot 6^2}{2!} = 0{,}0446$$

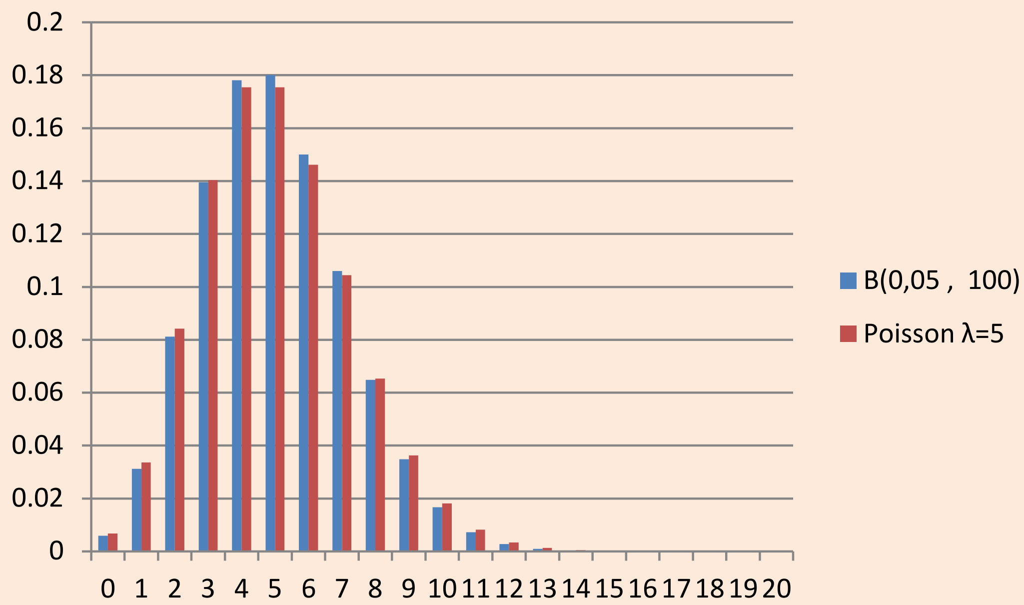
Επομένως P (τουλάχιστον 3 παραγγελίες το δίωρο)=

=1-0,0025-0,0149-0,0446=0,9380

Άρα η ζητούμενη πιθανότητα είναι **93,80%.**

Η προσέγγιση της Δυωνυμικής κατανομής από την κατανομή Poisson

- Αν η X ακολουθεί Δυωνυμική κατανομή με n μεγάλο και p κοντά στο μηδέν (με $np<10$), τότε η κατανομή της X προσεγγίζεται ικανοποιητικά από την κατανομή Poisson. Το λ της Poisson ως μέση τιμή της θα ισούται με $\lambda=np$.



Παράδειγμα

- Η διακριτή τυχαία μεταβλητή X ακολουθεί Δυωνυμική κατανομή με $n = 100$ και $p = 0{,}03$. Να βρεθεί η $P[X=2]$,
- α) με χρήση της Δυωνυμικής κατανομής, και
- β) με τη χρήση της προσέγγισης από την κατανομή Poisson.

Παράδειγμα

- α) $P(X=2)=\binom{100}{2}0{,}03^2\cdot 0{,}97^{98}=0{,}225$

- β) Προσεγγίζουμε τη Δυωνυμική κατανομή με την κατανομή Poisson υπολογίζοντας το λ = n · p = 100·0,03=3.

$$P(X=2)=\frac{e^{-3}3^2}{2!}=0{,}224$$