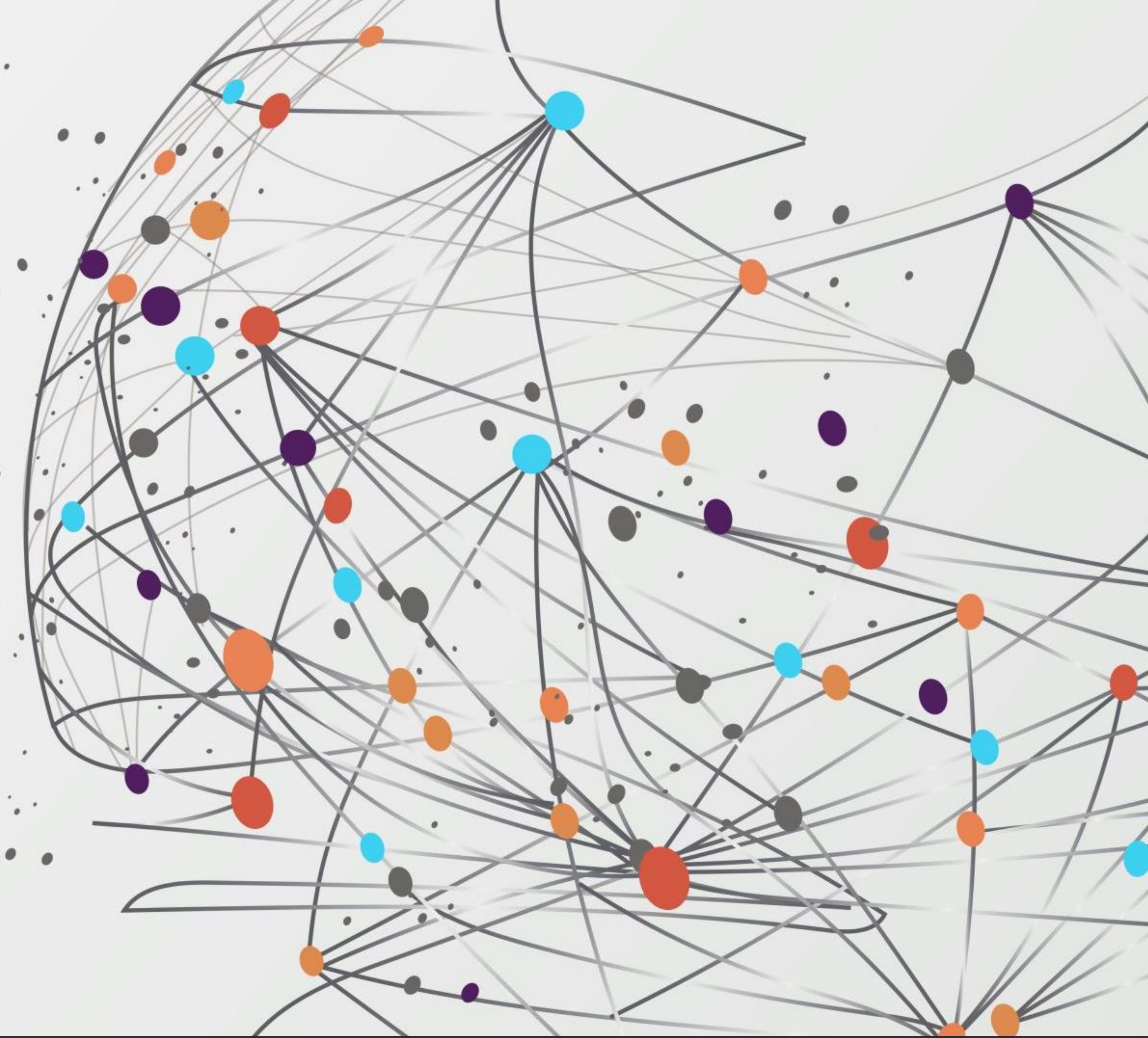
Θεωρία Πιθανοτήτων



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

Βιβλιογραφία

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 : Αρνητική διωνυμική κατανομή (κατανομή Pascal)

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

$$P(\mathbf{X}=N) = \binom{N-1}{r-1} \cdot p^{r} \cdot q^{N-r} \quad \text{όπου } N = r, r+1, \ldots$$

Η αρνητική διωνυμική κατανομή συμβολίζεται με τους εξής τρόπους: **X** ~ nbin(r, p) ή **X** ~ NB(r, p). Εκφράζει δηλαδή την πιθανότητα να χρειαστούν N δοκιμές μέχρις ότου πραγματοποιηθούν r «επιτυχίες».

Αρνητική διωνυμική κατανομή: Μέση τιμή, Διασπορά, Τυπική απόκλιση

1. Μέση τιμή

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

2. Διασπορά

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

3. Τυπική απόκλιση

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

Ορισμός : Υπεργεωμετρική κατανομή

Μια διακριτή τυχαία μεταβλητή **X** ακολουθεί αρνητική υπεργεωμετρική κατανομή με παραμέτρους M_1, M-M_1(=M_2) και n αν έχει συνάρτηση πιθανότητας

$$P(\mathbf{X}=\kappa) = \frac{\binom{M_1}{\kappa}\cdot\binom{M-M_1}{n-\kappa}}{\binom{M}{n}} \quad \text{όπου } \kappa = 0, 1, \ldots, n$$

Η υπεργεωμετρική κατανομή συμβολίζεται με τους εξής τρόπους: **X** ~ HGeom(M_1, M_2, n) ή **X** ~ Υπερ(M, M_1, n) ή **X** ~ HG(M_1, M_2, n). Εκφράζει δηλαδή την πιθανότητα από έναν πληθυσμό M μελών με M_1 δυνατές «επιτυχίες», αν επιλέξουμε ένα δείγμα n μελών, χωρίς επανατοποθέτηση, να παρατηρηθούν κ «επιτυχίες».

Υπεργεωμετρική κατανομή: Μέση τιμή, Διασπορά, Τυπική απόκλιση

1. Μέση τιμή

$$\mu = E(\mathbf{X}) = n \cdot p \text{ , με } p = \frac{M_1}{M}$$

2. Διασπορά

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

3. Τυπική απόκλιση

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

Παρατηρήσεις

1. Σε αντίθεση με τις υπόλοιπες κατανομές οι δοκιμές στην υπεργεωμετρική κατανομή δεν είναι ανεξάρτητες και η πιθανότητα «επιτυχίας» σε κάθε δοκιμή είναι διαφορετική.
2. Αν ο πληθυσμός Ν είναι πολύ μεγάλος και το δείγμα n μικρό τότε μπορεί να προσεγγισθεί η υπεργεωμετρική κατανομή από τη διωνυμική, αφού η πιθανότητα p δεν αλλάζει σημαντικά μεταξύ των δοκιμών. Εξ' ου και ο διορθωτικός παράγοντας λόγω του πεπερασμένου πληθυσμού $\frac{M-n}{M-1}$ τείνει στη μονάδα αν το M>>n, δηλαδή ο πληθυσμός γίνεται τόσο μεγάλος που η μη επανατοποθέτηση γίνεται ασήμαντη.

Ορισμός : Κατανομή Poisson

Μια διακριτή τυχαία μεταβλητή **X** η οποία ακολουθεί την διωνυμική κατανομή **X** ~ B(N, p) με τις παραμέτρους $N \to \infty$ και $p \to 0$ έτσι ώστε $\mu = N \cdot p \to \lambda$ με $\lambda > 0$, τότε στο όριο αυτό ακολουθεί την κατανομή Poisson με παράμετρο λ και η συνάρτηση πιθανότητάς της δίνεται από την σχέση

$$p_{\mathbf{X}}(x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!} \quad \text{όπου } x = 0, 1, \ldots$$

Η κατανομή Poisson συμβολίζεται με τους εξής τρόπους: **X** ~ P(λ) ή **X** ~ Pois(λ) ή **X** ~ Po(λ). Εκφράζει δηλαδή την πιθανότητα x «επιτυχιών» σε έναν πολύ μεγάλο αριθμό δοκιμών N όπου η πιθανότητα «επιτυχίας» ανά δοκιμή είναι πολύ μικρή.

Κατανομή Poisson: Μέση τιμή, Διασπορά, Τυπική απόκλιση

1. Μέση τιμή

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

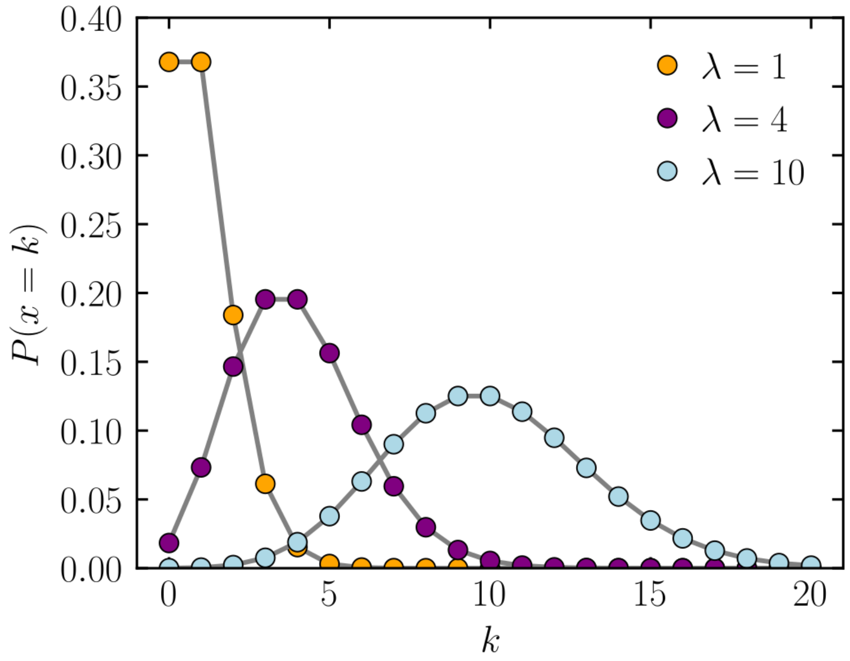
2. Διασπορά

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

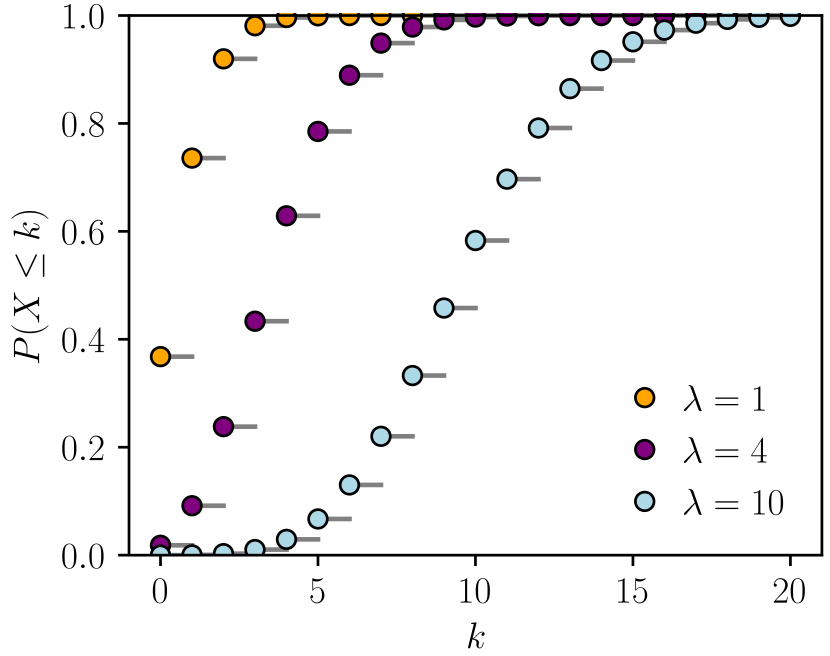
3. Τυπική απόκλιση

$$\sigma = \sqrt{Var(\mathbf{X})} = \sqrt{\lambda}$$

Συνάρτηση πιθανότητας

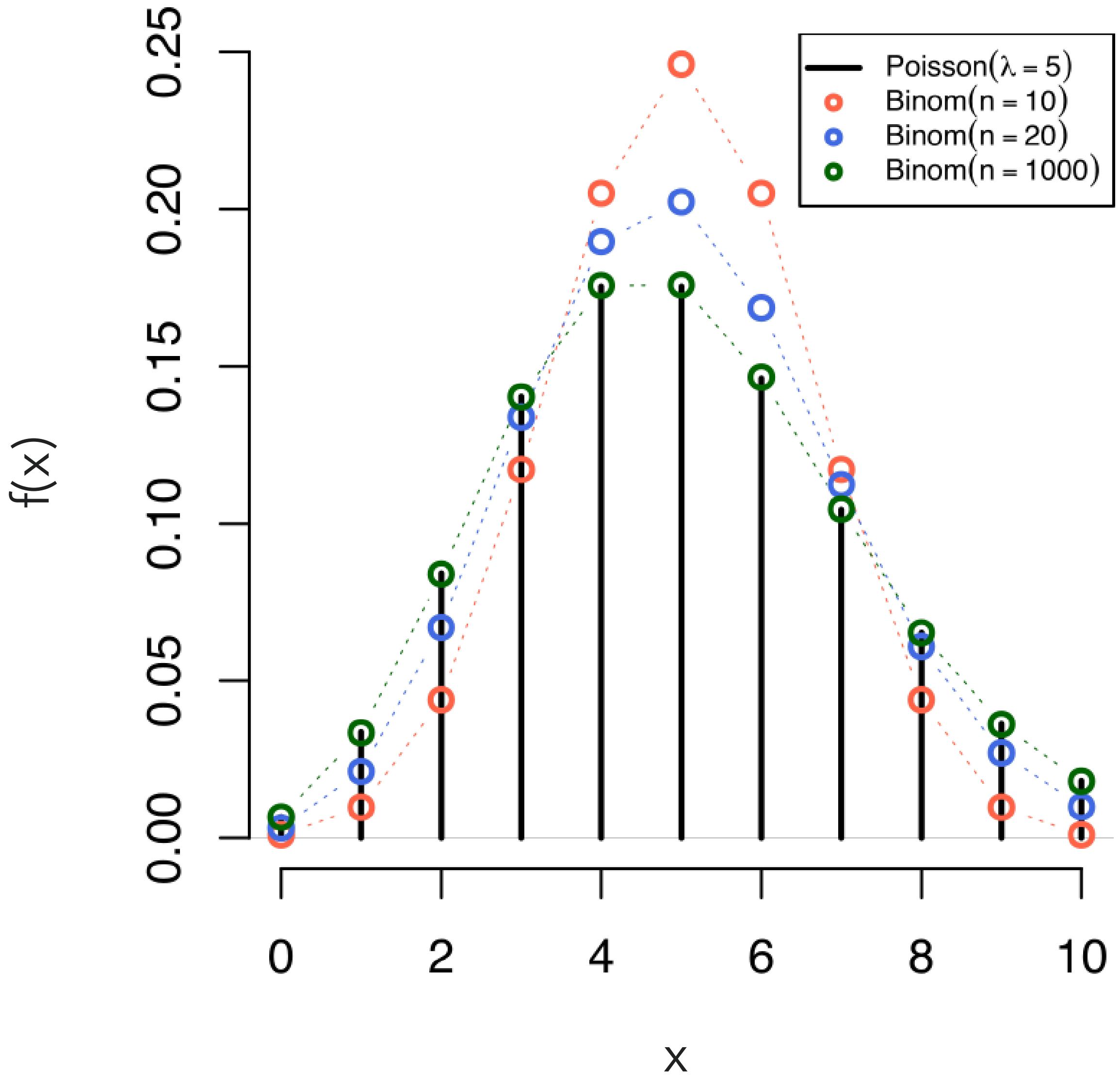


Αθροιστική συνάρτηση κατανομής



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

Προσέγγιση της Διωνυμικής κατανομής από την κατανομή Poisson



Καθώς το N γίνεται μεγαλύτερο, η κατανομή Poisson προσεγγίζει ολοένα και καλύτερα τη διωνυμική κατανομή.

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

Παρατηρήσεις

1. Η κατανομή Poisson λέγεται και κατανομή σπάνιων ενδεχομένων (ή νόμος των μικρών αριθμών) λόγω του ότι εφαρμόζεται σε πειράματα με πολύ μεγάλο πλήθος επαναλήψεων N και με πολύ μικρή πιθανότητα «επιτυχίας» p ανά δοκιμή.
2. Πόσο μεγάλο είναι το N και πόσο μικρό είναι το p; Μία επαρκώς ικανοποιητική προσέγγιση είναι για $N>100$ και $p<0.05$ ή $N\cdot p<5$ (ή εναλλακτικά $N\cdot p<10$)

Προσέγγιση παραγοντικών μεγάλων αριθμών - Προσεγγίσεις Stirling

1. $\ln N! = N \cdot \ln N - N - 1$

$N >> 1 \quad \ln N! \simeq N \cdot \ln N - N$

Άλλες προσεγγίσεις του παραγοντικού:

2. $N! = N^N \cdot e^{-N} \cdot \sqrt{2\pi N}$

ή καλύτερη προσέγγιση

3. $N! = N^N \cdot e^{-N} \cdot \sqrt{2\pi N} \cdot \left(1 + \frac{1}{12 \cdot N}\right)$

Ορισμός: Διαδικασία Poisson

Θα θεωρούμε ως ανεξάρτητη δοκιμή την παρατήρηση ή όχι ενός ενδεχόμενου σε ένα στοιχειώδες χρονικό διάστημα με ίδια πιθανότητα πραγματοποίησης. Έστω ότι το ενδεχόμενο πραγματοποιείται με πιθανότητα ανάλογη της διαμέρισης του χρόνου $p \propto \frac{t}{N}$ σε Ν επιμέρους στοιχειώδη διαστήματα, ενώ υπάρχει κάποιος στατιστικά προσδιορισμένος συντελεστής αναλογίας r τέτοιος ώστε $p = r \cdot \frac{t}{N}$ εκφράζει δηλαδή τον ρυθμό πραγματοποίησης του ενδεχομένου. Αν ο χρόνος διαμεριστεί σε πάρα πολλά στοιχειώδη διαστήματα δηλαδή $N \to \infty$ τότε $p \to 0$, το $N \cdot p = r \cdot t$ και πληρούνται όλες οι προϋποθέσεις ώστε να περιγράφει το πείραμα ως μια κατανομή Poisson.

Αυτή η διαδικασία καταμέτρησης του πλήθους των φορών που πραγματοποιήθηκε το ενδεχόμενο σε διάστημα t εκφράζεται από μια τυχαία μεταβλητή **X** η οποία προσεγγίζεται από την κατανομή Poisson και λέγεται διαδικασία Poisson με μέσο ρυθμό εμφάνισης r (δηλαδή τον πλήθος των φορών που πραγματοποιήθηκε το ενδεχόμενο στη μονάδα του χρόνου) και συνάρτηση πιθανότητας

$$P(\mathbf{X}=x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!} \quad \text{όπου } \lambda = r \cdot t$$

Παραδείγματα διαδικασιών Poisson

Ως ανεξάρτητες δοκιμές μπορούν να θεωρηθούν εκτός από χρονικά διαστήματα, χωρικά διαστήματα ή στοιχειώδες επιφάνειες ή στοιχειώδεις όγκοι.

Ως διαδικασία Poisson μπορούν να ιδωθούν προβλήματα όπως:

- ο αριθμός πασχόντων από μια σπάνια πάθηση σε ορισμένο χρονικό διάστημα
- ο αριθμός των πελατών κλήσεων ή email που φθάνουν σε μια ώρα, μια ημέρα, μια εβδομάδα
- ο αριθμός των σωματιδίων α που εκπέμπονται σε ορισμένο χρονικό διάστημα από πυρήνες ασταθών ραδιενεργά στοιχείων
- ο αριθμός των ηλεκτρονίων που εκπέμπονται από μια θερμιονική πηγή
- ο αριθμός των μεταλλάξεων κατά μήκος μιας αλυσίδας DNA
- ο αριθμός των κρυσταλλικών ατελειών στον όγκο ενός τρισδιάστατου υλικού ή στην επιφάνεια του
- ο αριθμός των ουρανίων σωμάτων σε μια περιοχή του γαλαξία
- ο αριθμός των βακτηρίων σε μια επιφάνεια