Πανεπιστήμιο Αιγαίου

Πιθανότητες και Στατιστική

Ελισάβετ Κωνσταντίνου

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Κανονική κατανομή

- Αν $X \sim N(\mu, \sigma^2)$, τότε $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$
- Αν $X_1 \sim N(\mu_1, \sigma_1^2)$ και $X_2 \sim N(\mu_2, \sigma_2^2)$, τότε
  $Y = X_1 + X_2 \sim N(\mu, \sigma^2)$

  όπου $\mu = \mu_1 + \mu_2$

  και $\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2Cov(X, Y) = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$

  Γενικά αν $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, ..., n$, **ανεξάρτητες**, τότε

  $Y = a_1X_1 + ... + a_nX_n + b \sim N(\mu, \sigma^2)$

  όπου $\mu = a_1\mu_1 + ... + a_n\mu_n + b$

  και $\sigma^2 = a_1^2\sigma_1^2 + ... + a_n^2\sigma_n^2$
- Αν $X_i \sim N(\mu, \sigma^2)$, $i = 1, ..., n$, **ανεξάρτητες**, τότε

  $Y = (X_1 + ... + X_n) / n \sim N(\mu, \sigma^2 / n)$

Διωνυμική κατανομή

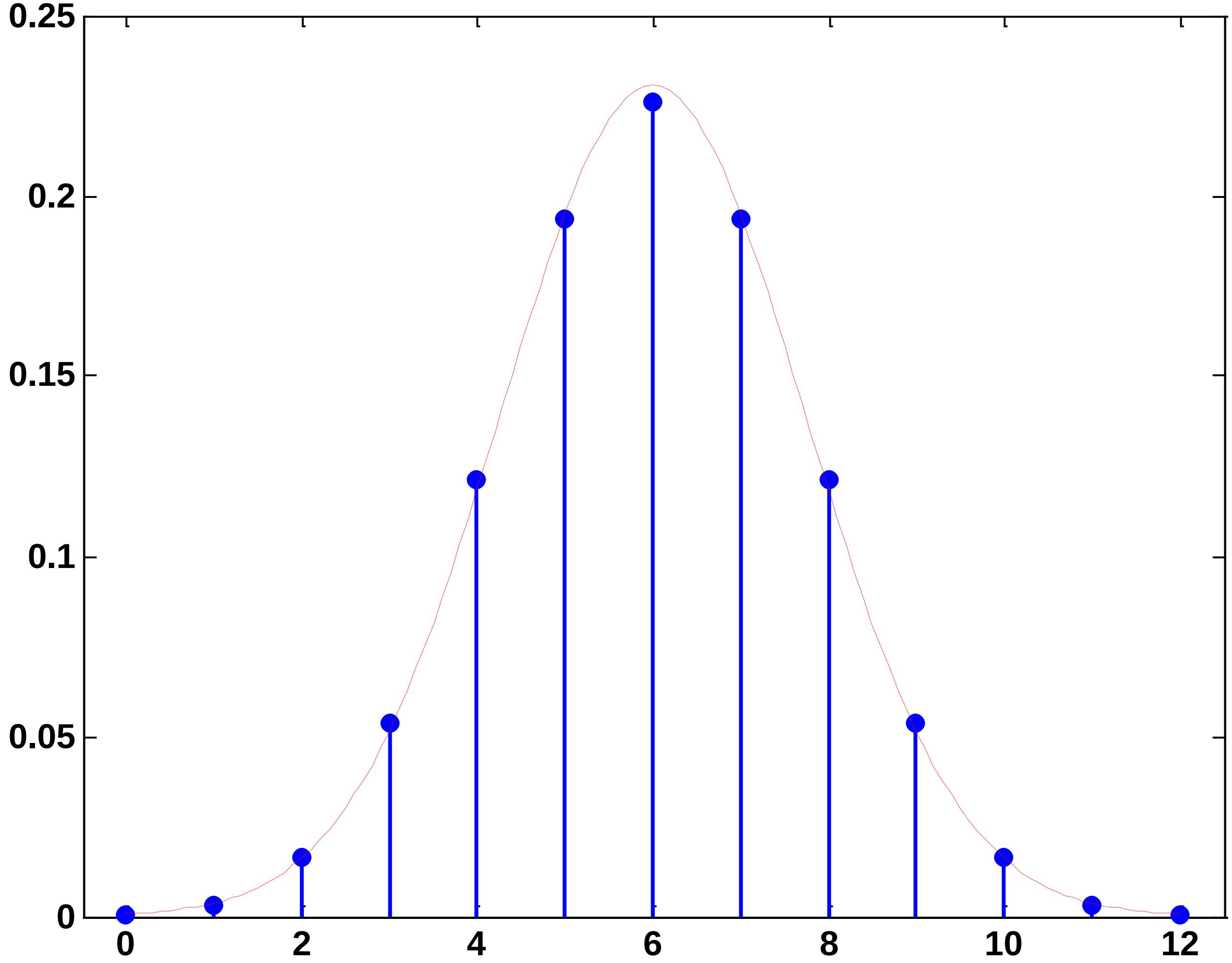
- Αν ένα πείραμα αποτελείται από n ανεξάρτητα πειράματα Bernoulli, το καθένα με πιθανότητα επιτυχίας p, τότε ο συνολικός αριθμός επιτυχιών είναι διωνυμική τ.μ. με παραμέτρους n και p:

  $X \sim B(n, p)$

  $$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- $E\{X\} = np$, $Var\{X\} = np(1-p)$
- Για $p = 0.5$, η σ.π. είναι συμμετρική γύρω από τη μέση τιμή $n / 2$.

Προσέγγιση της δυωνυμικής από την κανονική



Προσέγγιση της δυωνυμικής από την κανονική κατανομή

- Οταν το n είναι αρκετά μεγάλο και το p δεν είναι πολύ κοντά στα όρια 0 ή 1, η διωνυμική B(n, p) προσεγγίζεται από μια κανονική με την ίδια μέση τιμή και διασπορά:

  $B(n, p) \rightarrow N(\mu, \sigma)$

  όπου $\mu = np$ και $\sigma^2 = np(1–p)$.

  Γενικά η προσέγγιση θεωρείται ικανοποιητική για $np \geq 5$ και $n(1–p) \geq 5$.

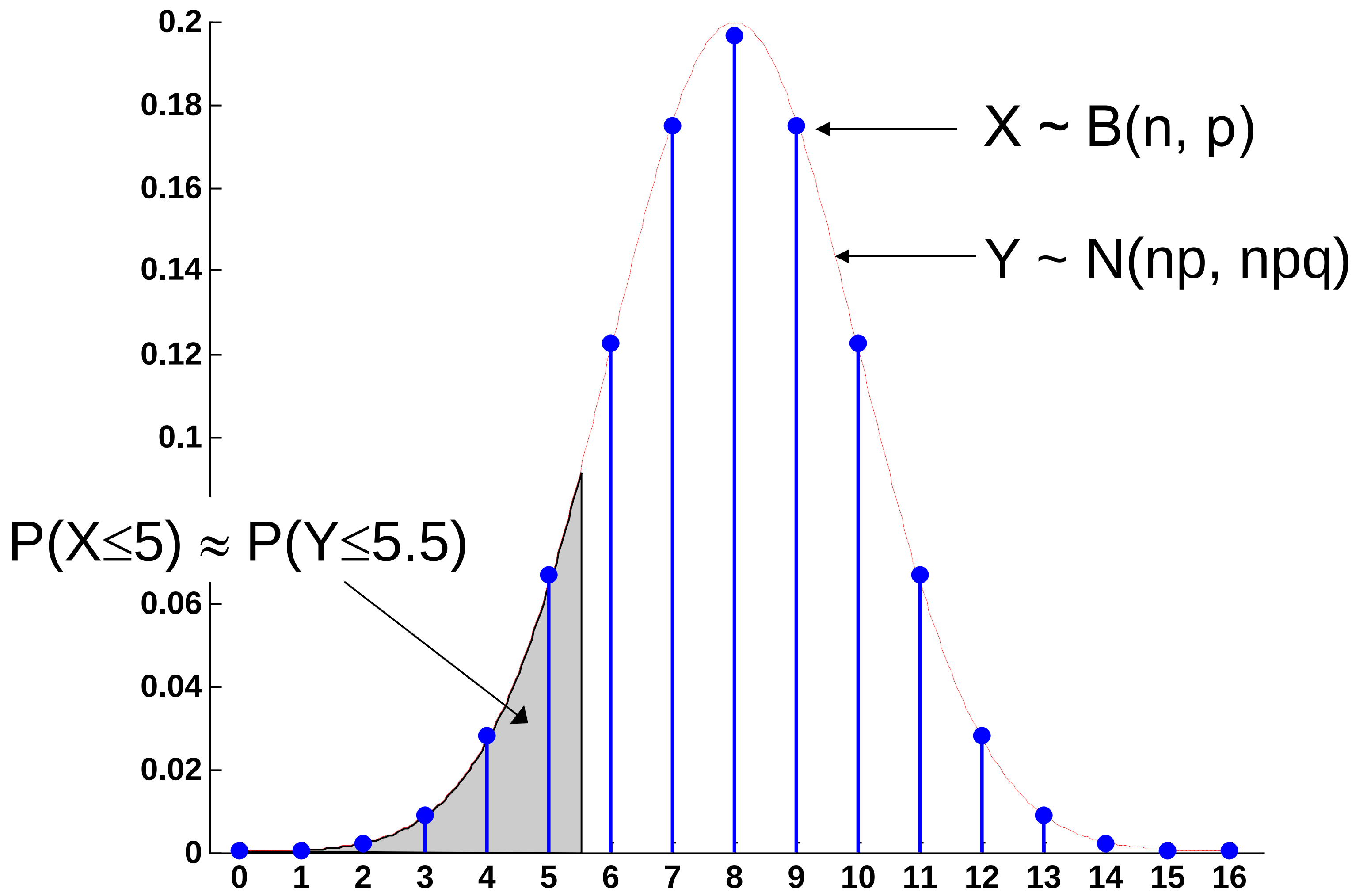
Παράδειγμα

- Εχουμε n = 16 συσκευές με πιθανότητα βλάβης της καθεμιάς στο επόμενο χρόνο p = 0.5. Ο αριθμός των βλαβών θα είναι X ~ B(16, 0.5) με E{X} = 8, Var{X} = 4.
- Η πιθανότητα να έχουμε μέχρι και 5 βλάβες είναι:

$$P(X \le 5) = \sum_{k=0}^{5} \binom{16}{k} \times 0.5^{k} \times 0.5^{16-k} = 0.1051$$

- Αν προσεγγίσουμε την X με μία Y ~ N(μ = 8, σ = 2):

$$P(Y \le 5.5) = \Phi\left(\frac{5.5-8}{2}\right) = 0.1056$$



X ~ B(n, p)
Y ~ N(np, npq)
P(X≤5) ≈ P(Y≤5.5)
0.2
0.18
0.16
0.14
0.12
0.1
0.06
0.04
0.02
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Κεντρικό οριακό θεώρημα

- Αν οι $X_1, \ldots, X_n$ είναι **ανεξάρτητες** τ.μ. με την ίδια κατανομή **(οποιαδήποτε)**, με μέση τιμή μ και διασπορά σ^2, τότε για αρκετά μεγάλο n:
  – το άθροισμά τους προσεγγίζει την $N(n\mu, n\sigma^2)$
  – ο μέσος όρος τους προσεγγίζει την $N(\mu, \sigma^2 / n)$
- Παράδειγμα: ο χρόνος X μεταξύ βλαβών στο δίκτυο είναι εκθετική τ.μ. με $\lambda = 1/15$.

  Ο χρόνος εμφάνισης της 10ης βλάβης: $Y \sim N(150, 2250)$

  Αρα με πιθανότητα 97% θα έχουμε τη 10η βλάβη σε 55 έως 245 μέρες από σήμερα.

Εκθετική κατανομή

- Το χρονικό διάστημα Τ μεταξύ αφίξεων του λεωφορείου στη στάση είναι εκθετική τ.μ. με Ε{Τ} = 15 λεπτά.

  $P(T \leq t) = 1 - e^{-\lambda t}, t > 0$
- Πιθανότητα να περιμένουμε περισσότερο από 15 λεπτά:

  $P(T > 15) = e^{-1} = 0.368$
- Αν είμαστε ήδη στη στάση 15 λεπτά, η πιθανότητα να περιμένουμε τουλάχιστον άλλα 15 είναι:

$$P(T > 30 \mid T > 15) = \frac{P(T > 30 \,\&\, T > 15)}{P(T > 15)} = \frac{P(T > 30)}{P(T > 15)} = 0.368$$

- Η εκθετική κατανομή «δεν έχει μνήμη».

Κατανομή Poisson

- Αν τα χρονικά διαστήματα μεταξύ διαδοχικών συμβάντων είναι ανεξάρτητες εκθετικές τ.μ. με την ίδια παράμετρο λ, τότε ο αριθμός των συμβάντων N σε μια περίοδο T είναι τ.μ. Poisson με παράμετρο λT:

$$p(N = k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!}$$

- και τα N_1, N_2 σε μη επικαλυπτόμενα διαστήματα T_1, T_2 είναι επίσης ανεξάρτητα.
- Ισχύει και το αντίστροφο.

Διδιάστατη κανονική κατανομή

- Αν $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ και $Cov(X_1, X_2) = 0$ (X_1, X_2 ανεξάρτητες), η από κοινού π.π. είναι:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ -\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(y-\mu_2)^2}{2\sigma_2^2} \right\}$$

- Οι υπό συνθήκη κατανομές $f_{X|Y}$ και $f_{Y|X}$ είναι επίσης κανονικές.
- Η παραπάνω σχέση γενικεύεται όταν οι X_1, X_2 δεν είναι ανεξάρτητες.

$X \sim N(0,1)$, $Y \sim N(0,1)$, $\rho = 0$