203hotline-bootstrap

B-MAT-400

Previously: random variables and distributions

- Random variable: values that depends of the outcome of an experiment, defined as a function $X: \Omega \to E$
  - Example: sum of 2 dice rolls, $E = \{2,3,4,5,6,7,8,9,10,11,12\}$

- Probability distribution: function that maps every value of a random variable to its probability

$$p(x) = P(X = x), x \in E$$

Bernoulli trials

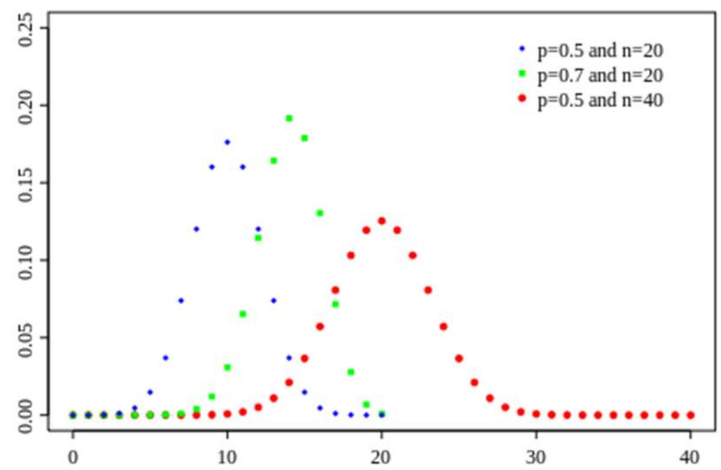
- Random experiment with exactly two possible outcomes:
  - “success” (event A)
  - “failure” (event A^C)
- If p is the probability of success:
  - $P(A) = p$
  - $P(A^C) = 1 - p$
- Example: rolling a 6 (event A)
  - $P(A) = \frac{1}{6}$
  - $P(A^C) = \frac{5}{6}$

Binomial distribution

- The random variable X is the number of successes in n independent Bernoulli trials of probability p
- X follows a binomial distribution: $X \sim B(n, p)$

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

- C_n^k is the binomial coefficient: $C_n^k = \frac{n!}{k!(n-k)!}$



Binomial distribution – Example

- An urn with 2 white balls and 3 black ones
- Experiment: drawing a ball and putting it back
- Random variable: number of white balls picked after n draws
- Distribution: $P(X = k) = \mathrm{C}_{\mathrm{n}}^{\mathrm{k}} \left(\frac{2}{5}\right)^{k} \left(\frac{3}{5}\right)^{n-k}$
- Picking 4 white balls in 5 draws:

$$P(X = 4) = C_5^4 \left(\frac{2}{5}\right)^4 \left(\frac{3}{5}\right) \approx 0.0768 = 7.68\%$$

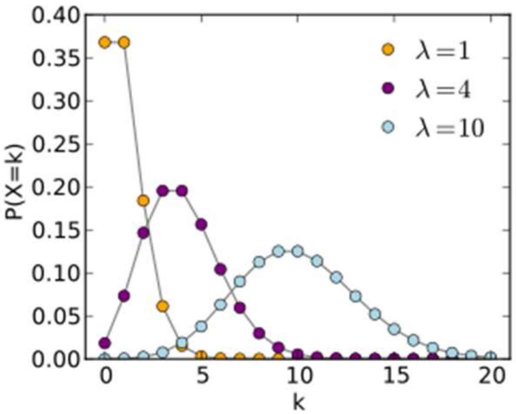
Poisson distribution

- Probability of a given number of events occurring in a fixed interval of time (or space) when knowing the average amount of events per interval
- Used to count rare events
- Examples:
  - Number of accidents on a highway each day
  - Number of phone calls in a call center
  - Number of defects on a cable length
  - Number of bacteria on a microscope slide
  - …

Poisson distribution

- Hypotheses:
  - Events are independent
  - Only one event can occur at the same time
  - Events occurs at a constant rate
- X is the number of events observed in an interval
- λ is the average number of events in an interval
- X follows a Poisson distribution $P(\lambda)$ if

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



Poisson distribution – Example

- In average, 1.9 customers enter a store per minute
- Random variable: number of customer entering the store in a minute
- Distribution: $P(X=k)=\frac{e^{-1.9}1.9^k}{k!}$
- Probability that 5 customers enter the store in a minute:

$$P(X=5)=\frac{e^{-1.9}1.9^5}{5!}\approx 0.0309=3.09\%$$

Approximating a binomial distribution

- Computing a binomial distribution when n is high is expensive
- A binomial distribution $B(n, p)$ can be approximated by a Poisson distribution $P(\lambda)$ if:
  - p is low (< 0.1)
  - n is large (> 50)
  - $np \leq 10$
- The parameter of the Poisson distribution is $\lambda = np$

203hotline

- Call center with 25 phones, 3500 calls per day (8 hours)
- Probability of an overload?
- Goals:
  - Compute efficiently binomial coefficients
  - Compute the binomial distribution for the number of simultaneous phone calls, given the average duration of a call
  - Compute the approximation by the Poisson distribution

203hotline

- Inputs: [n k | d]
  - n, k: values for the computation of C_n^k
  - d: average duration of a phone call (in seconds)
- Output (if n and k passed as parameters):
  - Value of C_n^k
- Output (if d passed as parameter):
  - For both Binomial and Poisson distributions:
    - Probability of getting n simultaneous phone calls for n from 0 to 50
    - Probability of getting an overload
    - Computation time

Exercise: Computing binomial coefficients

$$C_n^k = \frac{n!}{k!\,(n-k)!}$$

- What is the relationship between C_n^k and C_{n-1}^{k-1}?
- What is the relationship between C_n^k and C_n^{n-k}?
- Create a function that computes C_n^k without using factorial
- Beware, when n increases, C_n^k becomes quite large…

Exercise: p in the binomial distribution

- 3500 calls per day (8 hours) in average
- Average duration of a call = d seconds
- The random variable of the distribution is the number of **simultaneous** calls
- What is the event we want to observe?
- What is its probability p ?
- Create a function that computes p given the average call duration d