Poisson process. Central limit theorem

Probability theory

Anton Afanasev

Higher School of Economics

DSBA 211

November 12, 2022

Seminar Overview

1. Quiz
2. Poisson process
   - Poisson distribution
   - Exponential distribution
3. Combinations of independent variables
   - Convolution
   - Marginalization
4. Limit theorems
   - De Moivre–Laplace theorem
   - Poisson limit theorem
   - Central limit theorem

Quiz

Random variables X and Y have a joint p.d.f.:

$$f(x, y) = (c + 0.1xy) \cdot I_{\{0 \le x, y \le 1\}}.$$

1. Are X and Y independent?
2. Find $\mathsf{P}(Y < X/2)$.

Poisson process

- It's a process with large number of rare occurrences.
- Properties:
  1. events in disjoint time intervals are independent;
  2. probability of more than 1 occurrence at the same time $\to 0$;
  3. the average rate of events occurring is assumed to be constant.

Example

- Radioactive decay.
- Telephone calls arrivals.
- Insurance payouts.
- Big earthquakes and meteorites occurrences.
- Queues dynamics.

Poisson distribution

- Poisson distribution of $X \iff X \sim \text{Poisson}(\lambda)$.
  X shows the number of occurrences in Poisson process with an average number of events λ.
- P.m.f.:
$$\mathsf{P}(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}, \quad k \in \mathbb{N}_0.$$
- Mean:
$$\mathsf{E}(X) = \lambda.$$
- Variance:
$$\mathsf{V}(X) = \lambda.$$

Poisson distribution

- To avoid confusion when dealing with different time units:

$$\mathsf{P}(k \text{ events in time period}) = \frac{e^{\left(-\frac{\text{events}}{\text{time unit}} \cdot \text{time period}\right)} \cdot \left(\frac{\text{events}}{\text{time unit}} \cdot \text{time period}\right)^k}{k!}.$$

- We can adapt p.m.f. using the notion of average rate r:

$$\frac{\text{events}}{\text{time unit}} \triangleq r, \qquad \text{time period} \triangleq t,$$

$$\mathsf{P}(X = k) = \frac{e^{-rt} \cdot (rt)^k}{k!}, \quad k \in \mathbb{N}_0.$$

Problem 1

Cars independently pass a point on a busy road at an average rate of 150 per hour.

1. Assuming a Poisson distribution, find the probability that none passes in a given minute.
2. What is the expected number passing in two minutes?
3. Find the probability that the expected number actually passes in a given two-minute period.

Other motor vehicles (vans, motorcycles etc.) pass the same point independently at the rate of 75 per hour. Assume a Poisson distribution for these vehicles too.

4. What is the probability of one car and one other motor vehicle in a two-minute period?

Problem 2

Telephone calls enter the hotel switchboard according to a Poisson process on the average of two every 3 minutes.

1. Calculate the probability, that there will be exactly one call in 3 minutes.
2. Calculate the probability, that there will be no more than one call in 3 minutes.

Let Y be the waiting time for the second call.

3. Find the probability $\mathsf{P}(Y > 3)$.

Exponential distribution

- Exponential distribution of $X \iff X \sim \text{Exp}(\lambda)$.
  X shows the time between events in a Poisson process with average number of events λ.
- P.d.f.:
$$f_X(x) = \lambda e^{-\lambda x} \cdot I_{\{x \geq 0\}}.$$
- C.d.f.:
$$F_X(x) = \left(1 - e^{-\lambda x}\right) \cdot I_{\{x \geq 0\}}.$$
- Mean:
$$\mathsf{E}(X) = \frac{1}{\lambda}.$$
- Variance:
$$\mathsf{V}(X) = \frac{1}{\lambda^2}.$$

Problem 3

1. Let $X_1, X_2, \dots X_n$ be i.i.d. exponential random variables with parameter λ. Prove that $\min\{X_1, X_2, \dots X_n\}$ has exponential distribution.
2. Prove *memoryless property* of an exponential random variable X:

$$\mathsf{P}(X > s + t \mid X > s) = \mathsf{P}(X > t) \qquad \forall t, s > 0.$$

3. When you enter the bank, you find that all three tellers are busy serving other customers, and there are no other customers in queue. Assume that the service times for you and for each of the customers being served are independent identically distributed exponential random variables. What is the probability that you will be the last to leave among the four customers?

Convolution of probability distributions

- X and Y – independent random variables with known probability distributions.
- What is probability distribution for $X + Y$?
  - Discrete case:

$$\mathsf{P}(X+Y=z)=\sum_{y\in Y}\mathsf{P}(Y=y,X+Y=z)=\sum_{y\in Y}\mathsf{P}(Y=y,X=z-y)=$$

$$_{\text{(independence)}}=\sum_{y\in Y}\mathsf{P}(Y=y)\cdot\mathsf{P}(X=z-y).$$

  - Continuous case:

$$f_{X+Y}(z)=\int_{\mathbb{R}} f_Y(y)\,f_X(z-y)\,dy.$$

Problem 4

X_1 and X_2 are independent exponential random variables with rate λ. Let $Y = X_1 + X_2$. What is the p.d.f. of Y?

Use this result to compute probability from Problem 2.3.

Problem 5

Let X and Y be independent Poisson random variables with means λ and μ respectively. Find p.m.f. of $Z = X + Y$.

Preservation of distribution for sum of variables

- Normal distribution
  - $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.
  - $\text{Corr}(X, Y) = \rho$.
  - $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y)$.
- Poisson distribution
  - $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$.
  - X and Y are independent.
  - $X + Y \sim \text{Poisson}(\lambda + \mu)$.
- Binomial distribution
  - $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$.
  - X and Y are independent with identical probability of success p.
  - $X + Y \sim \text{Bin}(n + m, p)$.

Problem 6

Let X and Y be independent random variables with uniform distribution on $[0, 2]$ each. Find p.d.f. and c.d.f. of $Z = X + Y$.

Marginalization

- X and Y – independent random variables with known probability distributions.
- How to find $\mathsf{P}(g(X,Y) \le 0)$, where g is arbitrary?
  - Discrete case:

$$\mathsf{P}(g(X,Y) \le 0) = \sum_{y \in Y} \mathsf{P}(g(X,Y) \le 0, Y = y) = \sum_{y \in Y} \mathsf{P}(g(X,y) \le 0, Y = y) =$$

$$\underset{\text{(independence)}}{=} \sum_{y \in Y} \mathsf{P}(g(X,y) \le 0) \cdot \mathsf{P}(Y = y).$$

  - Continuous case:

$$\mathsf{P}(g(X,Y) \le 0) = \int_{\mathbb{R}} \mathsf{P}(g(X,y) \le 0) \cdot f_Y(y)\, dy.$$

Problem 7

Let $X_1, X_2, \dots X_n$ be independent random variables with X_i having an $\mathrm{Exp}(\lambda_i)$ distribution. Find the probability that X_i is the minimum.

Problem 8

Solve Problem 3.3 using marginalization technique.

Problem 9

A fair coin is flipped 1000 times. Let X be the number of heads, turned up in the experiment.

1. Find $\mathsf{P}(X \geq 1)$.
2. Find $\mathsf{P}(X \geq 10)$.
3. Find $\mathsf{P}(467 < X \leq 513)$.

De Moivre–Laplace theorem

- Let $X_n \sim \text{Bin}(n, p)$, and $q = 1 - p$.

Theorem (De Moivre–Laplace)

$$\frac{X_n - np}{\sqrt{npq}} \xrightarrow[n\to\infty]{d} Z \sim \mathcal{N}(0, 1)$$

- Approximation can be used when $np \gg 1$. Practical boundary is $np > 5$ and $nq > 5$. Cases with $p = 0$ or 1 can't be used.

Definition

Sequence of r.v.-s $\{X_1, \dots, X_n : \Omega \to \mathbb{R}\}$ converges in distribution to X:

$$X_n \xrightarrow[n\to\infty]{d} X,$$

$$\text{if } \forall x : F_X(x) \text{ is continuous} : \qquad \lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

Problem 9

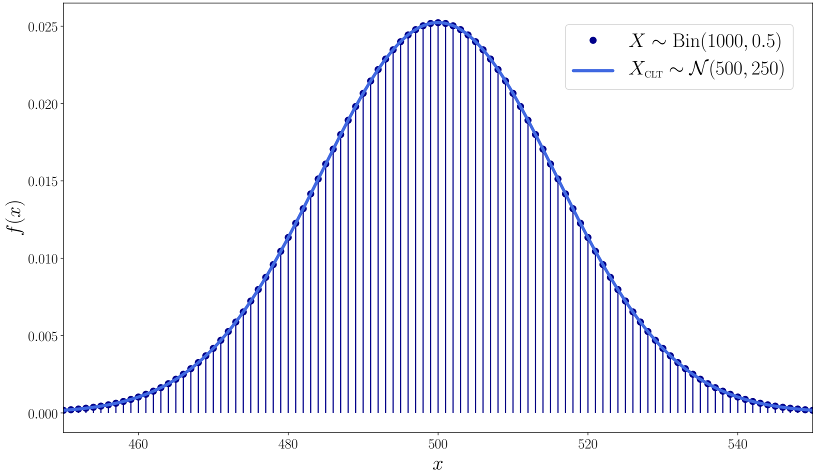

Figure: Approximation of $X \sim \text{Bin}\left(1000, \frac{1}{2}\right)$ with $X_{\text{CLT}} \sim \mathcal{N}(500, 250)$.

Problem 9

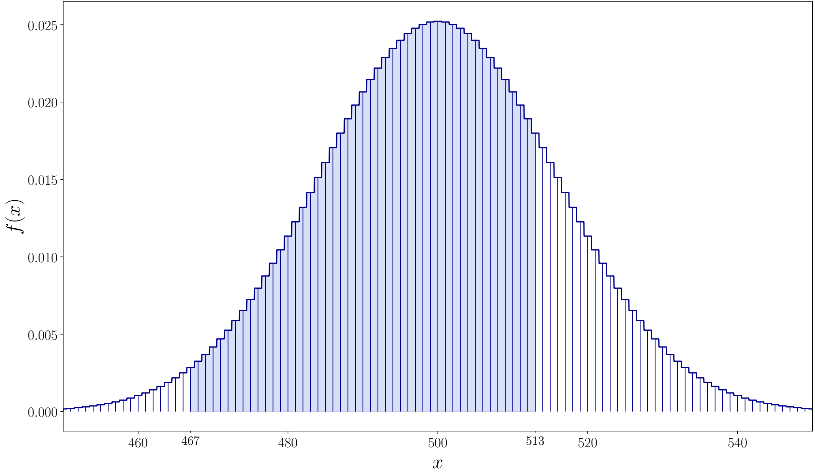

Figure: Pseudo-binomial p.d.f. to be approximated with $\mathsf{P}(467 < X_{\text{CLT}} \leq 513)$.

Problem 9

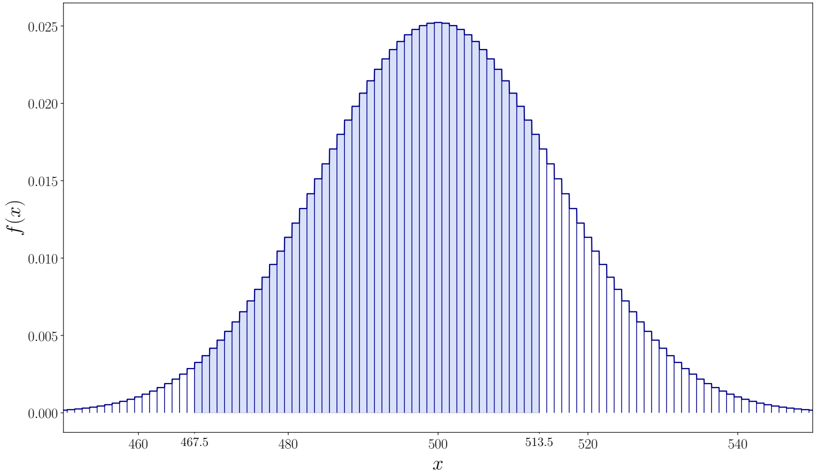

Figure: Continuity corrected area is given by $\mathsf{P}(467.5 < X_{\text{CLT}} \leq 513.5)$.

Poisson limit theorem

- Let $X_n \sim \text{Bin}(n, p)$, and $q = 1 - p$.

Theorem (Poisson)

$$X_n \xrightarrow[\substack{n\to\infty \\ p\to 0 \\ np=\lambda}]{d} Y \sim \text{Poisson}(\lambda)$$

- Approximation can be used when np^2 is small. For example $p = 1/100, n = 100$.
- Let $Y \sim \text{Poisson}(\lambda)$, then:

$$\frac{Y - \lambda}{\sqrt{\lambda}} \xrightarrow[\lambda\to\infty]{d} Z \sim \mathcal{N}(0, 1)$$

- If $\lambda > 10$, approximation is good with continuity correction.

Problem 10

An emergency towing service receives on average seventy calls per day for assistance. For any given day, what is the probability that fewer than fifty calls will be received?

Central limit theorem

- Let $X_1, \ldots, X_n$ be i.i.d. (collectively independent and identically distributed) random variables with $\mathsf{V}(X_i) = \sigma^2 > 0$, $\mathsf{E}(X_i) = \mu$, and $\mathsf{E}\left(X_i^2\right) < \infty$.

$$S_n = X_1 + \ldots + X_n,$$

$$Z_n = \frac{S_n - \mathsf{E}(S_n)}{\sigma(S_n)} = \frac{S_n - n\mu}{\sqrt{n}\sigma}.$$

Theorem (CLT)

$$Z_n \xrightarrow[n\to\infty]{d} Z \sim \mathcal{N}(0,1)$$

- Binomial case (De Moivre–Laplace):

$$X_i \sim \text{Bernoulli}(p),$$
$$S_n \sim \text{Bin}(n, p),$$

$$Z_n = \frac{S_n - np}{\sqrt{npq}} \quad \xrightarrow[n\to\infty]{d} \quad Z \sim \mathcal{N}(0, 1).$$

- Poisson case:

$$X_i \sim \text{Poisson}(1),$$
$$S_n \sim \text{Poisson}(n),$$

$$Z_n = \frac{S_n - n \cdot 1}{\sqrt{n} \cdot 1} \quad \xrightarrow[n\to\infty]{d} \quad Z \sim \mathcal{N}(0, 1).$$

Problem 11

Suppose that the step length is uniformly distributed in the interval $[0.9, 1.1]$ meters. A person walks 100 steps. Find the probability that the walked distance will be more than 105 meters.

Problem 12

Your gain in one round of a game with zero mean is 1 (with probability 0.5), 0 (with probability 0.4), and -5 (with probability 0.1). Estimate the probability that your total gain after 100 independent rounds of the game will be greater than 20.

Problem 13

Let $X_1, \ldots, X_{48}$ be a random sample of size 48 from the distribution with p.d.f.:

$$f(x) = \frac{1}{x^2} \cdot I_{\{x>1\}}.$$

Approximate probability, that at most 10 of these random variables have values greater than 4.

Look at the time!