Ma 3b Practical – Recitation 5

February 27, 2025

Recall Poisson distribution, Bayer's theorem, Normal distribution, and central limit theorem.

Exercise 1. (Poisson process) Suppose that an office receives telephone calls as a Poisson process with $\lambda = 0.5$ per min. What is the probability of exactly one call in first 5 mininutes?

Exercise 2. (Random variables) The Rayleigh distribution has CDF

$$F(x) = 1 - e^{-x^2/2}, \quad x > 0$$

Compute the following

1. f(X)
2. E(X), Var(X)
3. P(X>2)

Intuition Let F be the CDF and f be the PDF of a continuous r.v. X. It's important that $f(x)$ is not a probability; for example, we could have $f(3) > 1$, and we know $P(X = 3) = 0$. But thinking about the probability of X being very close to 3 gives us a way to interpret $f(3)$. Specifically, the probability of X being in a tiny interval of length ϵ, centered at 3 , will essentially be $f(3)\epsilon$. This is because

$$P(3 - \epsilon/2 < X < 3 + \epsilon/2) = \int_{3-\epsilon/2}^{3+\epsilon/2} f(x)dx \approx f(3)\epsilon,$$

This also implies that $P(X \leqslant 3) = P(X < 3)$.

Exercise 3. (normal distribution) We estimate that the distribution of a birthdate, in days from conception, has a normal distribution with mean 270 (in days) and a variance of 100 . What is the probability that a child was born after the 300-th day or before the 240-th day after conception?

Exercise 4. (central limit theorem) We try to measure the intensity of a signal, but we can only make approximate measurements. These measurements are i.i.d. random variables X_i with mean μ (this is the true intensity, which is unknown) and variance 10 . How many measurements must be taken to find the intensity with an error of at most 1 unit, and 95% certainty?

Exercise 5. (central limit theorem) Using the CLT and a continuity correction[1] , estimate the probability that a Poisson r.v. with parameter $\lambda = 100$ is larger than 120 .

Exercise 6. (central limit theorem) We roll 100 unbiased dice and we want to estimate the probability that the sum of the results is between 300 and 400 (both included).

[1] *Let's forget about continuity correction if not covered yet. It's some minor improvement on estimation for discrete i.i.d random variables.*

Solution. The number of calls in a 5 -min. interval follows a Poisson distribution with parameter $\omega = 5\lambda = 2.5$. Thus, the probability of no calls in a 5 -min. interval is $e^{-2.5} = .082$. The probability of exactly one call is $2.5e^{-2.5} = .205$.

Notes: Section 5.6 of Introduction to Probability-Joseph K. Blitzstein, Jessica Hwang.

Recall that the Poisson distribution $\mathrm{Pois}(\lambda)$ has pmf $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$ where $\lambda = 2$.

A Poisson process is a sequence of arrivals occurring at different points on a timeline, such that the number of arrivals in a particular interval of time has a Poisson distribution. A process of arrivals in continuous time is called a Poisson process with rate λ if the following two conditions hold.

1. The number of arrivals that occur in an interval of length t is a Pois(λt) random variable, i.e. pmf $P_t(X = k) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}$.

2. The numbers of arrivals that occur in disjoint intervals are independent of each other. For example, the numbers of arrivals in the intervals (0, 10), [10, 12), and [15, 1) are independent.

Exercise 7. (normal distribution) We estimate that the distribution of a birthdate, in days from conception, has a normal distribution with mean 270 (in days) and a variance of 100 . What is the probability that a child was born after the 300-th day or before the 240-th day after conception?

Solution. Let X be the birthdate, in days from conception, then

$$X \sim \mathcal{N}(270, 100)$$

We compute

$$\begin{aligned}
&= 1 - P(240 \leqslant X \leqslant 300) \\
&= 1 - P\left(\frac{240 - 270}{10} \leqslant \frac{X - 270}{10} \leqslant \frac{300 - 270}{10}\right) \\
&= 1 - P(-3 \leqslant Z \leqslant 3) \quad \text{where } Z \stackrel{\circ}{=} \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1) \\
&= 1 - [P(Z \leqslant 3) - P(Z \leqslant -3)] = 1 - P(Z \leqslant 3) + 1 - P(Z \leqslant 3) \\
&= 2 - 2F_Z(3) \\
&= 2 - 2(0.9987) \quad (\text{ see the } \mathcal{N}(0, 1) \text{ c.d.f. table }) \\
&= 0.0026.
\end{aligned}$$

Exercise 8. (central limit theorem) We try to measure the intensity of a signal, but we can only make approximate measurements. These measurements are i.i.d. random variables X_i with mean μ (this is the true intensity, which is unknown) and variance 10 . How many measurements must be taken to find the intensity with an error of at most 1 unit, and 95% certainty?

Solution. We want to choose n so that $P\left(\left|\bar{X}_n - \mu\right| \leqslant 1\right) = 0.95$.

Equivalently, $P\left(\left|\frac{\bar{X}_n-\mu}{\sqrt{10/n}}\right| \leqslant \frac{1}{\sqrt{10/n}}\right) = 0.95$.

According to the CLT, this is a normal distribution

$$P\left(\left|\frac{\bar{X}_n - \mu}{\sqrt{10/n}}\right| \leqslant a\right) = 0.95. \iff a = 1.96$$

So $\frac{1}{\sqrt{10/n}} = 1.96$ and $n = 10(1.96)^2 = 38.416$.

Solution. Let $X_i \overset{\text{i.i.d.}}{\sim} \text{Pois}(1)$, then we know that $X \overset{\circ}{=} \sum_{i=1}^{100} X_i \sim \text{Pois}(100)$. Note that $E[X_i] = 1$ and $\text{Var}(X_i) = 1$. Therefore, by the CLT (also applying a continuity correction), we have

$$\begin{aligned}
P(X > 120) &= P\left(\frac{X - 100\cdot 1}{\sqrt{100\cdot 1}} > \frac{120 - 100\cdot 1}{\sqrt{100\cdot 1}}\right) \\
&\approx 1 - F_Z\left(\frac{120 + 1/2 - 100}{10}\right) \qquad \text{where } Z \sim \mathcal{N}(0,1) \\
&= 1 - F_Z(2.05) \\
&= 1 - 0.9798 \\
&= 0.0202
\end{aligned}$$

Solution. Let $X_i, i \in \{1, 2, \dots, 100\}$, be i.i.d. r.v.s such that

$$P(X_i = k) = 1/6 \quad \text{for all } k \in \{1, 2, \dots, 6\}.$$

We have

$$\begin{aligned}
E[X_i] &= \sum_{k=1}^{6} k\frac{1}{6} = \frac{6\cdot 7}{2}\frac{1}{6} = 7/2 = 3.5 \\
E\left[X_i^2\right] &= \sum_{k=1}^{6} k^2\frac{1}{6} = \frac{6\cdot 7\cdot 13}{6}\frac{1}{6} = 91/6 \\
\text{Var}(X_i) &= E\left[X_i^2\right] - (E[X_i])^2 = 91/6 - (7/2)^2 = \frac{182 - 147}{12} = 2.91\overline{6}
\end{aligned}$$

Let $X \overset{\circ}{=} \sum_{i=1}^{100} X_i$, and note that

$$E[X] = 100 \cdot 3.5 = 350, \quad \text{and} \quad \text{Var}(X) = 100 \cdot 2.91\overline{6} = 291.\overline{6}$$

By the CLT (also applying a continuity correction), we have

$$
\begin{aligned}
P(X \in [300, 400]) &= P(|X - 350| \leqslant 50) \\
&= P\left(\left|\frac{X - 350}{\sqrt{291.\overline{6}}}\right| \leqslant \frac{50}{\sqrt{291.\overline{6}}}\right) \\
&= 2F_Z\left(\frac{50 + 1/2}{\sqrt{291.\overline{6}}}\right) - 1 \\
&= 2F_Z(2.96) - 1 = 2 \cdot 0.9985 - 1 = 0.997
\end{aligned}
$$

1 Remark on continuity correction(optional)

Since the cumulative distribution function of a binomial r.v. jumps at discrete values and the one for the normal is continuous, the approximation can be improved if we apply what is called a continuity correction. When we look at the probability that S_n is inside some interval of the form $(c, d), [c, d), (c, d], [c, d]$,

- we add $+1/2$ to the LEFT endpoint if it is strict (it doesn't include the endpoint);
- we add $-1/2$ to the LEFT endpoint if it is not strict (it includes the endpoint);
- we add $-1/2$ to the RIGHT endpoint if it is strict (it doesn't include the endpoint);
- we add $+1/2$ to the RIGHT endpoint if it is not strict (it includes the endpoint);

2 Theorem (Central limit theorem (CLT))

2Let $X_1, X_2, X_3, \dots$ be a sequence of i.i.d. r.v.s with expectation μ and (finite) variance σ^2, and let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, we have the following convergence in law :

$$\lim_{n\to\infty} P\left(\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \leqslant a\right) = F_Z(a) \stackrel{\circ}{=} \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz, \quad \forall a \in \mathbb{R},$$

where $Z \sim \mathcal{N}(0, 1)$. We can also write:

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \leqslant a\right) = F_Z(a), \quad \forall a \in \mathbb{R}.$$