Statistical Techniques. Week 2

Objectives (for today)

- Data and sampling distributions
- Statistic
- MLE
- Markov, Chebyshev inequalities
- LLM
- CLT

Theory (recap)

Data and sampling distributions

Data distribution – distribution of the original dataset

Sampling distribution – distribution of a statistic calculated on many samples drawn from the original dataset

Sampling distribution != sample distribution (do not mix!!!)

Statistic

Statistic – any function of the random variables constituting a random sample (Walpole R. E. et al. Probability and statistics for engineers and scientists)

Examples: mean, median, standard deviation.

Sampling distribution of mean

Each $X_i, i = 1, \dots, n$ has a normal distribution with mean μ and variance σ^2. Then

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

has a a normal distribution with mean $\mu_x = \mu$ and variance $\sigma_x^2 = \frac{\sigma^2}{n}$.

MLE

Maximum likelihood estimation (MLE) is a method that can be used to estimate the parameters of a given distribution.

Algorithm:

1. Write the PDF;
2. Write the likelihood function (the product of the PDF for the observed values);
3. Write the log likelihood function;
4. Calculate the derivative of the log likelihood function with respect to the parameters;
5. Set the derivative equal to zero and solve for the parameters.

Markov and Chebyshev inequalities

$$P(X \geq a) \leq \frac{E(X)}{a}$$

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

1. What are $E(X)$, μ, σ?
2. What are these formulas?

LLN (weak)

Given a collection of iid samples from a random variable with finite mean, the sample mean converges in probability to the expected value. Than is, for any $\epsilon > 0$:

$$\lim_{n \to \infty} P\left(\left|\overline{X_n} - \mu\right| < \epsilon\right) = 1$$

CLT

If is $\bar{X}$ the mean of a random sample of size n taken from a population with mean μ and finite variance σ^2, then the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

as $\mathrm{n} \to \infty$, is the standard normal distribution $\mathrm{n}(z; 0, 1)$.

Break, 5 min

Pen and paper part

Sampling distribution: tasks

1. If the standard deviation of the mean for the sampling distribution of random samples of size 36 from a large or infinite population is 2, how large must the sample size become if the standard deviation is to be reduced to 1.2.

   Answer: 100

2. The heights of 1000 students are approximately normally distributed with a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. Suppose 200 random samples of size 25 are drawn from this population and the means recorded to the nearest tenth of a centimeter. Determine the mean and standard deviation of the sampling distribution.

   Answer: 174.5 and 1.38

Statistic: tasks

The lengths of time, in minutes, that 10 patients waited in a doctor's office before receiving treatment were recorded as follows: 5, 11, 9, 5, 10, 15, 6, 10, 5, and 10. Treating the data as a random sample, find

1. the mean;
2. the median;
3. the mode.

Answer: 8.6
Answer: 9.5
Answer: 5, 10

Which statistic describes the data better?

MLE: tasks

Estimate the parameter λ of a Poisson distribution. The probability function of Poisson distribution is:

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

Answer: $\lambda = \frac{1}{n}\sum_{j=1}^{n} x_j$

Markov, Chebyshev inequalities: tasks

1. Suppose that the average grade on the upcoming math exam is 70%. Give an upper bound on the proportion of students who score at least 90%.

Answer: 7/9

2. If the distribution of Y is $b(n; 0.25)$, give a lower bound for $P(\left|\frac{Y}{n} - 0.25\right| < 0.05)$ when $n = 100$.

Answer: 0.25

3. Suppose a fair coin is flipped 100 times. Find a bound on the probability that the number of times the coin lands on heads is at least 60 or at most 40.

Answer: 0.25

CLT: tasks

If a certain machine makes electrical resistors having a mean resistance of 40 ohms and a standard deviation of 2 ohms, what is the probability that a random sample of 36 of these resistors will have a combined resistance of more than 1458 ohms?

Answer: 0.0668

Coding part

Practice with Python

Click here

The end