The paper-and-pencil problems.

Actually, each of these problems can be solved in a fully digital environment, such as ipynb. However, you can submit handwritten solutions as well.

Strontium (15 points) Let T be an exponentially distributed random variable with parameter λ.
We call half-life the time h such that $P(T > h) = \frac{1}{2}$.
Strontium 90 is a dangerous radioactive isotope of strontium, which is released after a nuclear explosion. An atom of strontium 90 remains radioactive for a random time which follows an exponential distribution, after which it decays. Its half-life is approximatively 28 years.

- Express h in terms of λ. Find the value of λ for strontium 90.
- Find the probability that a given atom of strontium 90 has not decayed after 50 years.
- Calculate the number of years to wait so that 99% of the strontium 90 produced during a nuclear reaction disappears.

Portfolio (15 points). An investor puts \$2,000 into a deposit account with a fixed rate of return of 10% per year. A second sum of \$1,000 is invested in a fund with an expected rate of return of 16% and a standard deviation of 8% per year.

- Find the expected value of the total amount of money this investor will have after a year.
- Find the standard deviation of the total amount after a year.
- If the fund return is distributed normally, what is the probability that the total wealth will be less than \$3000?

Plane (15 points). The Aeroflot flight to Saint Petersburg on an Airbus has 380 seats. Suppose that any customer who booked this flight will cancel with probability $1 - p = 0.1$, independently of other customers. Suppose that the company sells a fixed number N of tickets, with $N \geq 380$. Let X be the number of customers present the day of the flight.

- What is the distribution of X?
- Using the central limit theorem, approximate the probability $P(X \geq n)$: plug the mean and variance from the original distribution into normal.
- What is the largest number of bookings the company can accept if it does not want to refuse customers with probability more than 0.05.
- Suppose the company chooses N equal the value calculated in the previous question. What is the probability that at least 35 seats are left unoccupied?

Worms (15 points). According to the U.S. Department of Agriculture, ten to twenty earthworms per cubic foot is a sign of healthy soil. The soil of a garden is checked by digging 8 holes, each of one-cubic-foot, and counting the earthworms, and the following counts are found: 5, 25, 15, 10, 7, 12, 16, 20.

1. Use the unbiased estimators to estimate the true mean and the true variance of worm density per cubic foot in this garden.
2. What are the mean and variance *of your estimate of the mean*? And how are they estimated?
3. Make a 95% confidence interval for the true mean (the interval between 2.5% and 97.5% quantiles of the estimated mean).

Maximum likelihood with cities (20 points). Pareto distribution describes various real-life variables, such as population of a city, number of citations of a paper, number of subscribers in a social network, etc. Its distribution (in the standardized form) can be defined by CDF $F_X(x) = P(X \leq x) = \begin{cases} 1 - (x/a)^{-b}, & if\ x \geq a \\ 0, & if\ x < a \end{cases}$,
where $a, b > 0$.

1) Derive a maximum likelihood estimate for the parameters a and b from an i.i.d sample $(x_1, \dots, x_n)$ of size n. That is:
   - Derive the PDF by differentiating the CDF with respect to x;
   - Write a formula for the log-likelihood function, using the PDF;
   - Find the parameters that maximize the log-likelihood:
     - Prove that whatever is b, log-likelihood is maximized when $a = \min(x_1, \dots, x_n)$
     - Find a value of b at which the derivative of LL with respect to b is 0.
2) Assuming that Russian cities have size distributed according to Pareto law, estimate the parameters. Use the data from the file `russian_cities_g_9k.csv`.
3) In large samples, log-likelihood estimator $\hat{b}$ of a parameter b has approximately normal distribution with mean b and variance $-(LL''_{bb})^{-1}$, where LL''_{bb} is the second derivative of log-likelihood w.r.t. b. Use this knowledge to construct a 99% confidence interval for b.

The coding problems (20 points)

Fill the gaps in the second part of the following notebook:

https://drive.google.com/file/d/1OzX879mnOJcdCyNW87JB8om9-QZ8hW0I/view?usp=sharing