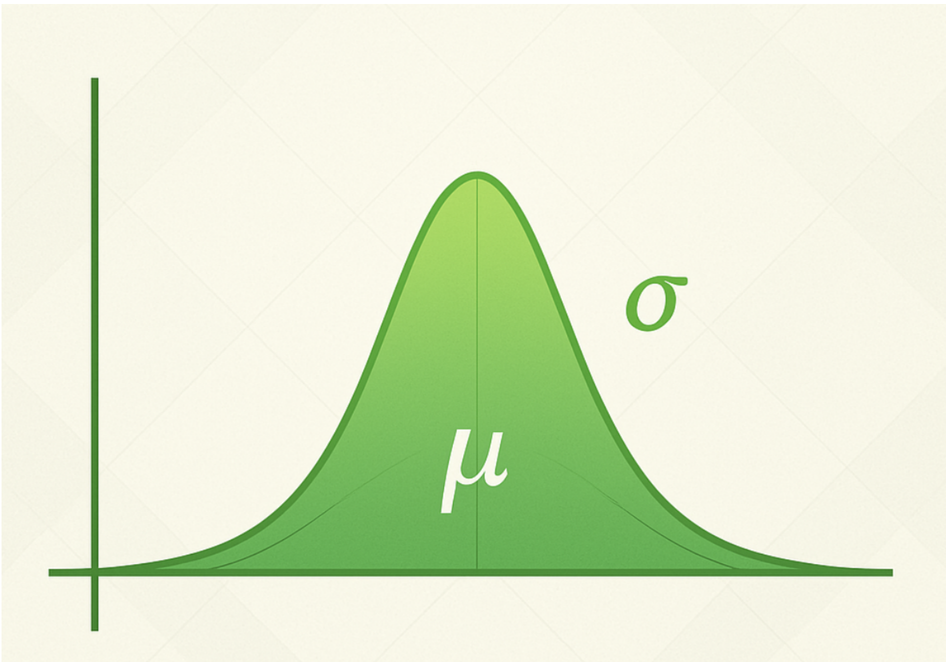
Higher School of Economics

Probability Theory and Statistics: Problem Sets for *Module 2*

Topics:

Seminars 9-10:	*Continuous Random Variables. Expectation of Continuous Distributions.*
Seminars 11-12:	*Continuos Random Vectors.*
Seminars 13-14:	*CLT and LLN.*
Seminars 15-16:	*Point Estimators and their Properties.*
Seminars 17-18:	*MLE and Method of Moments..*
Seminars 19-20:	*Least Square Estimators. Efficiency of Estimators.*
Seminars 21-22:	*CI for the Mean and Variance.*
Seminar 23:	*Review before Midterm.*



DSBA

Fall 2025

Continuous Random Variables. Expectation of Continuous Distributions.

During the seminars, we will work on problems across three different categories of difficulty, with Category 1 being the easiest and Category 3 the most challenging.

CATEGORY 1:

Poisson distribution: $X \sim \text{Poisson}(\lambda)$, where X shows the number of occurrences in Poisson process with an average number of events λ.

$$P(X = k) = \frac{e^{-k}\lambda^k}{k!}, \quad k \in \mathbb{N}_0, \quad E(X) = \lambda, \quad Var(X) = \lambda$$

Or if r is an average rate of occurrences, and t is a time period, then:

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

1. Cars independently pass a point on a busy road at an average rate of 150 per hour.
   (a) Assuming a Poisson distribution, find the probability that none passes in a given minute.
   (b) What is the expected number passing in two minutes?
   (c) 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.

   (d) What is the probability of one car and one other motor vehicle in a two-minute period?

$X \sim \text{Exp}(\lambda)$, where X shows the time between events in a Poisson process with average number of events λ.

$$f_X(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}, F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - \lambda e^{-\lambda x}, & x \geq 0 \end{cases}, \quad (X) = \frac{1}{\lambda}, (X) = \frac{1}{\lambda^2}$$

2. Telephone calls enter the hotel switchboard according to a Poisson process on the average of two every 3 minutes. Let Y be the waiting time for the second call.
   (a) Calculate the probability, that there will be exactly one call in 3 minutes.
   (b) Calculate the probability, that there will be no more than one call in 3 minutes.
   (c) Find the probability $P(Y > 3)$.

3. Prove the memoryless property of an exponential random variable $X \sim \text{Exp}(\lambda)$:

$$(X > s + t | X > s) = (X > t) \quad \forall t, s > 0.$$

Sum of normal distributions

All sums (linear combinations) of independent normally distributed random variables are also normally distributed.

Suppose $X_1, \ldots, X_n$ are independent normally distributed random variables, with $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n$, and $a_1, \ldots, a_n$ and b are constants. Then:

$$S = \sum_{i=1}^{n} a_i X_i + b \sim N(\mu, \sigma^2), \quad \text{where}$$

$$\mu = \sum_{i=1}^{n} a_i \mu_i + b \text{ and } \sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum \sum_{i<j} a_i a_j \mathsf{cov}(X_i, X_j)$$

4. Suppose that in the population of Russian people aged 16 or over:

   The heights of men (in cm) follow a normal distribution with mean 174.9 and standard deviation 7.39.

   The heights of women (in cm) follow a normal distribution with mean 161.3 and standard deviation 6.85.

   We select one man and one woman at random and independently of each other. What is the probability that the man is at most 10 cm taller than the woman?

χ^2 distribution

Let $Z_1, Z_2, \ldots, Z_k$ be independent $N(0,1)$ random variables. If:

$$Q = Z_1^2 + Z_2^2 + \ldots + Z_k^2 = \sum_{i=1}^{k} Z_i^2$$

the distribution of Q is the χ^2 distribution with k degrees of freedom. This is denoted by $Q \sim \chi^2(k)$ or $Q \sim \chi_k^2$. The χ_k^2 distribution is a continuous distribution, which can take values of $x \geq 0$. Its mean and variance are:

$$E(Q) = k$$

$$Var(Q) = 2k$$

If $Q_1, \ldots, Q_m$ are independent and distributed as $\chi_{k_1}^2, \ldots \chi_{k_m}^2$ respectively, then:

$$Q_1 + \ldots + Q_m \sim \chi_{k_1 + \ldots + k_m}^2$$

5. Let $X_1, \ldots, X_8$ be i.i.d. from $N(0, 16)$, $Y_1, \ldots, Y_5$ be i.i.d. from $N(1, 9)$; $Z_1, \ldots, Z_4$ be i.i.d. from $N(0, 1)$, and all X_i, Y_j, Z_k are independent.

   (a) Find $P\left(\sum_{i=1}^{4} Z_i^2 < 5\right)$

   (b) Find k such that $P\left(\sum_{i=1}^{4} Z_i^2 < k\right) = 0.4$

   (c) Find $P\left(\sum_{i=1}^{8} X_i^2 < 20\right)$

   (d) Find $P\left(\sum_{i=1}^{5} (Y_i - 1)^2 > 32\right)$

t-distribution

Let $Z \sim N(0,1)$ and $Q \sim \chi_k^2$ be independent random variables. Then the the distribution of random variable

$$T = \frac{Z}{\sqrt{Q/k}}$$

is t-distribution with k degrees of freedom (Student's t distribution)

$$E(T) = 0, k > 1$$

$$Var(T) = \frac{k}{k-2}, k > 2$$

When k increases t-distribution goes to normal distribution.

6. Let $X_1, \ldots, X_8$ be i.i.d. from $N(0,16)$, $Y_1, \ldots, Y_5$ be i.i.d. from $N(1,9)$, and all X_i, Y_j are independent.

   (a) Find $P\left(X_1 < \sqrt{\sum_{i=1}^{5}(Y_i-1)^2}\right)$

   (b) Find $P\left(X_1 + 2X_2 < \sqrt{\sum_{i=1}^{4}(Y_i-1)^2}\right)$

   (c) Find $P\left(Y_1 - 1 < \sqrt{\sum_{i=1}^{8} X_i^2}\right)$

F-distribution

Let $Q \sim \chi_k^2$ and $R \sim \chi_p^2$ be independent random variables. Then

$$F = \frac{Q/k}{R/p} \sim F_{k,p}$$

is F-distributed with degrees of freedom k and p.
By composition, inverse of F:

$$\frac{1}{F} \sim F_{p,k}$$

If $T \sim t_k$, then $T^2 \sim F_{1,k}$.

$$E(F) = \frac{k}{k-2}, k > 2$$

$$Var(F) = \frac{2k^2(p+k-2)}{p(k-2)^2(k-4)}, k > 4$$

7. Let $X_1, \ldots, X_8$ be i.i.d. from $N(0,16)$, $Y_1, \ldots, Y_5$ be i.i.d. from $N(1,9)$, and all X_i, Y_j are independent.

   (a) Find $P\left(\sum_{i=1}^{8} X_i^2 < \sum_{i=1}^{5}(Y_i-1)^2\right)$

   (b) Find $P\left(\sum_{i=1}^{4} X_i^2 < 7\sum_{i=1}^{3}(Y_i-1)^2\right)$

   (c) Find $P\left(\sum_{i=1}^{3} X_i^2 < \sum_{i=4}^{8} X_i^2\right)$

   (d) Find $P\left(\sum_{i=1}^{7} X_i^2 < \sum_{i=3}^{8} X_i^2\right)$

8. Suppose that X_1 and X_2 are independent $N(0,4)$ random variables. Compute $P(X_1^2 < 36.84 - X_2^2)$.

9. Suppose that X_1, X_2 and X_3 are independent $N(0,1)$ random variables, while Y (independently) follows a χ_5^2 distribution. Compute $P(X_1^2 + X_2^2 < 7.236Y - X_3^2)$.

10. Suppose $X_i \sim N(0,3)$, for $i = 1,2,3,4$. Assume all these random variables are independent. Derive the value of k in each of the following:

    a) $P(X_1 + 3X_2 > 4) = k$
    b) $P(X_1^2 + X_2^2 + X_3^2 + X_4^2 < k) = 0.99$
    c) $P(X_1 < k\sqrt{X_2^2 + X_3^2}) = 0.05$
    d) $P(X_1^2 + X_2^2 > 39(X_3^2 + X_4^2)) = k$

CATEGORY 2:

11. Find the expected value and the variance of a random variable such that

    (a) $\xi \sim U[a,b]$.
    (b) $\xi \sim \exp(\lambda)$.
    (c) $\xi \sim \mathcal{N}(a,\sigma)$

12. The PDF of η is given by

$$p_\eta(x) = \begin{cases} A\sin 2x & , x \in [0, \frac{\pi}{2}]; \\ 0 & , \text{otherwise.} \end{cases}$$

    Find (a) Value of A. (b) CDF of η, (c) $\mathsf{E}\eta$, $\mathsf{Var}\eta$. (d) $\mathsf{P}(-\frac{\pi}{6} \le \eta \le \frac{\pi}{6})$.

13. The coordinates of two points on the line are independent and have an uniform distribution on $[0,1]$. Find the expected value and the variance of the distance between these two points.

CATEGORY 3:

14. Find the expected value and the variance of a random variable such that

    (a) $\xi \sim \Gamma(\alpha,\beta)$
    (b) ξ has Beta distribution with parameters (α,β).

15. Let $\xi_1,\ldots,\xi_n$ be independent random variables with standard normal distribution. Prove that the random variable $\eta = \xi_1^2 + \ldots + \xi_n^2$ has distribution $\Gamma(1/2, n/2)$.

16. Let $\xi_1,\xi_2,\ldots,\xi_n$ be independent random variables with uniform distribution in $[0,1]$. Find $\mathsf{E}\xi_{(k)}$ and $\mathsf{Var}\xi_{(k)}$.

17. Let $\xi \sim \mathcal{N}(0,\sigma)$. Find $\mathsf{E}|\xi|$.

Continuos Random Vectors.

During the seminars, we will work on problems across three different categories of difficulty, with Category 1 being the easiest and Category 3 the most challenging.

CATEGORY 1:

1. Let X and Y be two jointly continuous random variables with joint PDF:

$$f_{XY}(x,y) = \begin{cases} x + cy^2, & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

   (a) Find the constant c.

   (b) Find $\mathrm{P}(0 \le X \le 0.5, 0 \le Y \le 0.5)$.

2. The probability distribution function (PDF) of a random vector (ξ, η) is given by

$$p_{\xi,\eta}(x,y) = \begin{cases} Axy & ,(x,y) \in D; \\ 0 & ,(x,y) \notin D. \end{cases}$$

   where $D : \{x \ge 0, y \ge 0, x + y \le 2\}$. Find

   a) The value of A.

   b) $\mathsf{P}((\xi,\eta) \in G)$, where $G : \{0 \le x \le 1, 0 \le y \le 1\}$.

   c) $p_\xi(x)$, $p_\eta(y)$. Are ξ and η independent random variables?

   d) $\mathsf{E}\xi$, $\mathsf{E}\eta$.

   e) $\mathsf{Var}\xi$, $\mathsf{Var}\eta$.

   f) $\mathsf{Cov}(\xi,\eta)$ and $\mathsf{Corr}(\xi,\eta)$

3. Let X and Y be two jointly continuous random variables with joint PDF:

$$f_{XY}(x,y) = \begin{cases} cxy^2, & 0 \le y \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

   (a) Find th constant c.

   (b) Find R_{XY} and show it in the xy plane.

   (c) Find marginal PDFs, $f_X(x)$ and $f_Y(y)$.

   (d) Find $\mathrm{P}\left(Y \le \frac{X}{2}\right)$.

   (e) Find $\mathrm{P}\left(Y \le \frac{X}{4} \middle| Y \le \frac{X}{2}\right)$.

4. The joint probability density function of the random variables X and Y is given by

$$p_{X,Y}(x,y) = \begin{cases} \frac{3}{2}(x^2+y^2) & ,0 \le x \le 1,\ 0 \le y \le 1; \\ 0 & ,\text{otherwise}. \end{cases}$$

   Find $\mathsf{cov}(X,Y)$ and $\mathsf{corr}(X,Y)$.

5. The joint probability density function of the random variables X and Y is given by

$$p_{X,Y}(x,y) = \begin{cases} 4xy & ,0 \le x \le 1,\ 0 \le y \le 1; \\ 0 & ,\text{otherwise}. \end{cases}$$

   Find $\mathsf{E}X^2Y$ and $\mathsf{cov}(X,Y)$.

CATEGORY 2:

6. Let ξ_1, ξ_2 be independent random variables, uniformly distributed on $[0, 2]$. Find $\mathsf{P}(1 \leq \xi_2\xi_2 \leq 2)$.

7. The joint probability density function of the random variables X and Y is given by

$$p_{X,Y}(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}.$$

Find $\mathsf{E}\sqrt{X^2 + Y^2}$.

CATEGORY 3:

8. The CDF of a random vector (ξ, η) is given by

$$F_{\xi,\eta}(x, y) = \begin{cases} 0 & , x \leq 0 \text{ or } y \leq 0, ; \\ 1 - e^{-x^2} - e^{-2y} + e^{-x^2-2y} & , x \geq 0, y \geq 0. \end{cases}$$

Find

a) The joint PDF $p_{\xi,\eta}(x, y)$.

b) The probabilities $\mathsf{P}(-2 \leq \xi < 2, 1 \leq \eta < 3)$, $\mathsf{P}(\xi \geq 0, \eta \geq 1)$, $\mathsf{P}(-2 \leq \xi < 2, 1 \leq \eta < 3)$, $\mathsf{P}(\xi < 1, \eta \geq 2)$

c) CDFs F_ξ and F_η.

9. Let the random vector (X, Y) have the following density

$$p_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)}$$

Find the covariance matrix of (X, Y) and the distribution of X.

CLT and LLN.

During the seminars, we will work on problems across three different categories of difficulty, with Category 1 being the easiest and Category 3 the most challenging.

CATEGORY 1:

Weak Law of Large Numbers (WLLN). Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with finite mean μ. Then

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$$

as $n \to \infty$.

Central Limit Theorem (CLT). Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with finite mean μ and finite variance $\sigma^2 > 0$. Then

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

as $n \to \infty$.

Poisson Limit Theorem. Let $X_n \sim \text{Binomial}(n, p_n)$ where $p_n \to 0$ and $np_n \to \lambda > 0$ as $n \to \infty$. Then

$$\lim_{n\to\infty} \Pr\big(X_n = k\big) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

i.e., X_n converges in distribution to a Poisson random variable with parameter λ.

1. 1800 dice are thrown. Find an approximate value for the probability that the total number of occurrences of 2 and 6 is not less than 620.
2. When typing, the stenographer makes a mistake in the character with probability of 0.0005. Find an approximate value for the probability that, when typing 10.000 characters, the stenographer will make a mistake no more than three times.
3. A bank teller serves customers standing in the queue one by one. Suppose that the service time ξ_i for customer i has mean $\mathsf{E}\xi_i = 2$ (minutes) and $\mathsf{Var}(\xi_i) = 1$. We assume that service times for different bank customers are independent. Let η be the total time the bank teller spends serving 50 costumers. Find $\mathsf{P}(90 < \eta < 110)$.
4. The probability that a light bulb lasts more than 1000 hours is $\frac{1}{3}$. Estimate the probability that out of 1800 light bulbs, the lifetime of at least 580 bulbs exceeds 1000 hours.

CATEGORY 2:

5. A die is thrown n times. Let the random variable S_n correspond to the sum of the points obtained. Find n such that $\mathsf{P}(\left|\frac{S_n}{n} - 3.5\right| \geq 0.1) \geq 0.05$
   a) Using the Chebyshev's inequality.
   b) Using the Central Limit Theorem.
6. How many times should we throw a die if we want that the sum of points obtained was at least 4500 with probability $p \geq 0,975$? (use the central limit theorem).

CATEGORY 3:

7. Prove that
$$\lim_{n\to\infty} \frac{1}{e^n} \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{2},$$
using the Central Limit Theorem.

8. Let $\{\xi_n\}$ be a sequence of independent identically distributed random variables with distribution $\mathcal{U}(0,1)$. Set $M_n = \max\{\xi_1, \ldots, \xi_n\}$. Prove that
$$n(1 - M_n) \xrightarrow{\mathsf{d}} \mathsf{Exp}(1)$$

9. Let $\{\xi_n,\, n \in \mathcal{N}\}$ be independent random variables with finite variance. Prove that for any $x \in \mathbb{R}$ the following limit exists:
$$\lim_{n\to\infty} P(\xi_1 + \ldots + \xi_n \leq x)$$
and that it equals 0, 1, or 1/2. Specify the conditions under which each of these values occurs.

Point Estimators and their Properties.

During the seminars, we will work on problems across three different categories of difficulty, with Category 1 being the easiest and Category 3 the most challenging.

CATEGORY 1:

Unbiased Estimator. An estimator $\hat{\theta}$ of a parameter θ is said to be **unbiased** if

$$\mathbb{E}[\hat{\theta}] = \theta.$$

Consistent Estimator. An estimator $\hat{\theta}_n$ of a parameter θ is said to be **consistent** if

$$\hat{\theta}_n \xrightarrow{P} \theta \quad \text{as } n \to \infty,$$

i.e., $\hat{\theta}_n$ converges in probability to θ as the sample size n increases.
Mean Squared Error (MSE). The **mean squared error** of an estimator $\hat{\theta}$ of a parameter θ is defined as

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}\big[(\hat{\theta} - \theta)^2\big].$$

It can be decomposed as:

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \Big(\mathrm{Bias}(\hat{\theta})\Big)^2.$$

Efficient Estimator. An estimator $\hat{\theta}$ is called **efficient** if it is unbiased and achieves the lowest possible variance among all unbiased estimators of θ. For regular problems, this variance lower bound is given by the Cramér-Rao lower bound.
Sample Mean. Given a sample $X_1, X_2, \ldots, X_n$, the **sample mean** $\bar{X}$ is defined as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Sample Variance. Given a sample $X_1, X_2, \ldots, X_n$, the **sample variance** S^2 is defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\big(X_i - \bar{X}\big)^2.$$

Sample Proportion. Given a sample of n Bernoulli trials with success indicator variables $X_1, X_2, \ldots, X_n$, the **sample proportion** $\hat{p}$ of successes is defined as

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

1. Random variable X assumes values 0 and 1, each with probability 1/2. You have 9 independent observations of X. Consider the following estimators of the population mean μ:

   - $\hat{\mu}_1 = 0.45$
   - $\hat{\mu}_2 = X_1$
   - $\hat{\mu}_3 = \overline{X}$
   - $\hat{\mu}_4 = \frac{2}{3}X_1 + \frac{2}{3}X_2 - \frac{1}{3}X_3$

   Are these estimator biased? Which one is the most efficient?

2. Let $X_1, \ldots, X_n$ be a sample from $\mathrm{U}(0, \theta)$. Verify if the following estimators for the parameter θ are unbiased and consistent:

   (a) $2\overline{X}$

3. Let $X_1, \ldots, X_n$ be a sample from some distribution with parameter σ^2. Assume additionally that $\mathsf{Var} X_1 = \sigma^2$. Define the uncorrected sample variance as the statistic $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Prove that:

   (a) $s^2 = \overline{X^2} - \overline{X}^2$;

   (b) Is s^2 an unbiased estimator for σ^2?

CATEGORY 2:

4. Let $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ and $Y_1, \ldots, Y_m \sim \mathcal{N}(2\mu, 2\sigma^2)$ and consider the estimator $\hat{\mu} = a\bar{X} + b\bar{Y}$. For which a and b the estimator $\hat{\mu}$ is unbiased and most efficient?

5. Show that
$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \Big(\mathrm{Bias}(\hat{\theta})\Big)^2.$$

6. Let $X_1, \ldots, X_n$ be a sample from $\mathrm{U}(0, \theta)$. Verify if the following estimators for the parameter θ are unbiased and consistent:

   (a) $\overline{X} + X_{(n)}/2$

   (b) $(n+1)X_{(1)}$

   (c) $X_{(1)} + X_{(n)}$

   (d) $\frac{n+1}{n}X_{(n)}$

7. Let $X_1, \ldots, X_n$ be a sample from some distribution with parameter σ^2. Assume additionally that $\mathsf{Var} X_1 = \sigma^2$. Define the uncorrected sample variance as the statistic $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Prove that:

   (a) s^2 is a consistent estimator for σ^2;

CATEGORY 3:

8. Show that, if $\mathrm{MSE}(\hat{\theta}) \to 0$ as $n \to \infty$, then the Estimator $\hat{\theta}$ is consistent.

MLE and Method of Moments.

During the seminars, we will work on problems across three different categories of difficulty, with Category 1 being the easiest and Category 3 the most challenging.

CATEGORY 1:

Maximum Likelihood Estimator (MLE)

Definition. Let $X_1, \dots, X_n$ be a random sample from a probability distribution with density (or mass) function $p_\theta(x)$ depending on a parameter θ. The *likelihood function* is defined as

$$L(\theta) = \prod_{i=1}^{n} p_\theta(X_i).$$

The *log-likelihood function* is

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log p_\theta(X_i).$$

A *maximum likelihood estimator* (MLE) of θ is any value $\hat{\theta}$ that maximizes $L(\theta)$ (or equivalently, $\ell(\theta)$), i.e.,

$$\hat{\theta} = \arg\max_{\theta} L(\theta).$$

Properties of the MLE

Under suitable regularity conditions, the MLE $\hat{\theta}$ has the following properties:

- **Consistency:** $\hat{\theta} \to \theta$ in probability as $n \to \infty$.
- **Asymptotic efficiency:** The MLE achieves the Cramér-Rao lower bound asymptotically.
- **Invariance:** If $g(\cdot)$ is a differentiable function, then $g(\hat{\theta})$ is the MLE of $g(\theta)$.

Method of Moments Estimator (MME)

Definition. Let $X_1, \dots, X_n$ be a random sample from a distribution with parameter θ. The theoretical moments of order k are defined as

$$\mu'_k(\theta) = \mathbb{E}\big[X^k\big].$$

The *method of moments* equates the first r sample moments to the corresponding theoretical moments to solve for the parameters. That is, for parameters $\theta = (\theta_1, \dots, \theta_r)$, solve

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k = \mu'_k(\theta), \quad k = 1, \dots, r.$$

The solution $\hat{\theta}$ is called the *method of moments estimator* (MME).

Properties of the Method of Moments Estimators

Under suitable conditions, the MME has the following properties:

- **Consistency:** MME is consistent under mild conditions, especially if the moments uniquely determine the distribution.
- **Simplicity:** Often simpler to compute than MLE, since it avoids maximizing complex likelihood functions.
- **Efficiency:** MMEs are generally *less efficient* than MLEs, meaning they may have larger variance asymptotically.

1. Find the maximum likelihood estimators (MLEs) for the parameters of the following distributions:

   (a) $\mathrm{Geom}(p)$

(b) $\mathcal{N}(a, \sigma^2)$ in three cases: when only one parameter is unknown and when both parameters are unknown.

(c) the parameter λ of the distribution Gamma(α, λ), assuming α is known.

2. A random sample $\{X_1, X_2, \ldots, X_n\}$ is drawn from the following probability distribution:

$$p(x;\, \lambda) = \frac{\lambda^{2x} e^{-\lambda^2}}{x!}, \quad \text{for } x = 0, 1, 2, \ldots$$

and $p(x;\, \lambda) = 0$ for all other values of x, with $\lambda > 0$.

(i) Derive the maximum likelihood estimator of λ.

(ii) State the maximum likelihood estimator of $\theta = \lambda^3$.

3. A random sample $\{X_1, X_2, \ldots, X_n\}$ is drawn from the following probability distribution:

$$f(x;\, \theta) = \begin{cases} \dfrac{2x}{1-\theta^2} & \text{for } \theta \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Derive the maximum likelihood estimator of θ and explain why this is a maximum.

(ii) If $n = 4$, such that $x_1 = 0.70$, $x_2 = 0.92$, $x_3 = 0.21$, and $x_4 = 0.34$, determine the maximum likelihood estimate of θ.

(iii) As $n \to \infty$, would you expect the maximum likelihood estimate of θ to get closer to θ? Briefly explain your answer.

4. Find the method of moments estimators for the parameters of the following distributions:

(a) Pois(λ)

(b) Geom(p)

(c) Beta(α, β)

(d) $\Gamma(\alpha, \beta)$

(e) Bin(m, p)

(f) $\mathcal{N}(a, \sigma^2)$

CATEGORY 2:

5. Find the maximum likelihood estimators (MLEs) for the parameters of the following distributions:

(a) U$(0, a)$

(b) Bin(n, p).

6. Find the method of moments estimators for the parameters of the following distributions:

(a) U(a, b)

CATEGORY 3:

7. Find the maximum likelihood estimate for the location parameter in the Cauchy distribution model,

$$p_\theta(x) = \frac{1}{\pi\big(1 + (x-\theta)^2\big)},$$

if the sample consists of (a) one observation, (b) two observations (i.e. $n = 1, 2$).

Least Square Estimators. Efficiency of Estimators.

During the seminars, we will work on problems across three different categories of difficulty, with Category 1 being the easiest and Category 3 the most challenging.

CATEGORY 1:

Least Squares Estimator (LSE)
Definition. Consider the linear regression model:

$$Y = X\beta + \varepsilon,$$

where:

- Y is an $n \times 1$ vector of observations.
- X is an $n \times p$ matrix of known predictors (design matrix).
- β is a $p \times 1$ vector of unknown parameters.
- ε is an $n \times 1$ vector of random errors, typically assumed to satisfy:

$$\mathbb{E}[\varepsilon] = 0 \quad \text{and} \quad \mathrm{Var}(\varepsilon) = \sigma^2 I_n.$$

The **least squares estimator** (LSE) of β minimizes the sum of squared residuals:

$$S(\beta) = (Y - X\beta)^\top (Y - X\beta).$$

The solution is given by:

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y,$$

provided that $X^\top X$ is invertible.

Properties of the Least Squares Estimator:

(a) **Unbiasedness:**

$$\mathbb{E}[\hat{\beta}] = \beta.$$

(b) **Variance-Covariance Matrix:**

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 (X^\top X)^{-1}.$$

(c) **Efficiency (Gauss-Markov Theorem):** Among all linear unbiased estimators, $\hat{\beta}$ has the minimum variance.

(d) **Normality (if errors are normal):**

$$\hat{\beta} \sim \mathcal{N}\big(\beta, \sigma^2 (X^\top X)^{-1}\big).$$

1. Suppose that you are given independent observations y_1, y_2 and y_3 such that:

$$\begin{aligned} y_1 &= 3\alpha + \beta + \varepsilon_1, \\ y_2 &= \alpha - 2\beta + \varepsilon_2, \\ y_3 &= -\alpha + 2\beta + \varepsilon_3. \end{aligned}$$

The random variables ε_i, for $i = 1, 2, 3$, are normally distributed with a mean of 0 and a variance of 4.

(i) Find the least squares estimators of the parameters α and β, and verify that they are unbiased estimators.

(7 marks)

(ii) Calculate the variance of the estimator of α.

(3 marks)

2. Suppose that you are given observations y_1, y_2 and y_3 such that:

$$\begin{aligned} y_1 &= 2\alpha + \beta + \varepsilon_1, \\ y_2 &= -\alpha - \beta + \varepsilon_2, \\ y_3 &= -\alpha + \beta + \varepsilon_3. \end{aligned}$$

The random variables ε_i, for $i = 1, 2, 3$, are independent and normally distributed with mean 0 and variance σ^2, while α and β are unknown parameters. Find the least squares estimator of β, verify that it is an unbiased estimator of β, and find the variance of this estimator.

CATEGORY 2:

3. There are 2 pieces of cheese with weights a and b. Using the same scales, the first piece, the second piece, and then both pieces together were weighed. Find the least squares estimators for a and b, as well as an unbiased estimator of the variance of the measurement error.

4. Let $X_1, \ldots, X_n$ be a sample from a uniform distribution on the interval $[0, \theta]$. Compare the following estimators of the parameter θ in the mean square error sense, that is, which one is the most efficient?

$$2\overline{X}, \quad (n+1)\, X_{(1)}, \quad \frac{n+1}{n}\, X_{(n)}.$$

CATEGORY 3:

5. Let $X_i = \beta_1 + i\,\beta_2 + \varepsilon_i$, for $i = 0, 1, \ldots, n$, where β_1 and β_2 are unknown parameters, and $\varepsilon_0, \ldots, \varepsilon_n$ are independent random variables distributed as $\mathcal{N}(0, \sigma^2)$. Reduce the problem to a linear model and find the least squares estimators for β_1 and β_2, as well as an unbiased estimator for σ^2.

6. In the quadrilateral $ABCD$, independent measurements of the angles ABD, DBC, ABC, BCD, CDB, BDA, CDA, DAB (in degrees) gave the results

$$50.78, \quad 30.25, \quad 78.29, \quad 99.57, \quad 50.42, \quad 40.59, \quad 88.87, \quad 89.86,$$

respectively. Assuming that the measurement errors are normally distributed according to $\mathcal{N}(0, \sigma^2)$, find the optimal estimates of the angles $\beta_1 = ABD$, $\beta_2 = DBC$, $\beta_3 = CDB$, $\beta_4 = BDA$, and the unknown variance σ^2.

CI for the Mean and Variance.

During the seminars, we will work on problems across three different categories of difficulty, with Category 1 being the easiest and Category 3 the most challenging.

CATEGORY 1:

Confidence Intervals for Mean and Variance of a Normal Distribution

Model:
Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution:

$$X_i \sim \mathcal{N}(\mu, \sigma^2), \quad i = 1, \ldots, n.$$

Theorem 1 (Confidence Interval for the Mean when Variance is Known).
If σ^2 is known, then:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Hence, a $(1 - \alpha) \times 100\%$ confidence interval for μ is:

$$\overline{X} \ \pm \ z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

Theorem 2 (Confidence Interval for the Mean when Variance is Unknown).
If σ^2 is unknown, then:

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Therefore, a $(1 - \alpha) \times 100\%$ confidence interval for μ is:

$$\overline{X} \ \pm \ t_{\alpha/2,\, n-1} \cdot \frac{S}{\sqrt{n}},$$

where $t_{\alpha/2,\, n-1}$ is the upper $\alpha/2$ quantile of the t-distribution with $n - 1$ degrees of freedom.

Theorem 3 (Confidence Interval for the Variance).
Under normality:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Hence, a $(1 - \alpha) \times 100\%$ confidence interval for σ^2 is:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,\, n-1}}, \ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,\, n-1}} \right),$$

where $\chi^2_{p,\, n-1}$ is the p-quantile of the chi-squared distribution with $n - 1$ degrees of freedom.

Remark.
Confidence intervals for σ can be obtained by taking square roots of the limits of the interval for σ^2.

1. You are interested in the average emergency room (ER) wait time at your local hospital. You take a random sample of 50 patients who visit the ER over the past week. From this sample, the mean wait time was 30 minutes and the standard deviation was 20 minutes. Find a 95% confidence interval for the average ER wait time for the hospital.

2. The yearly salary for mathematics assistant professors are normally distributed. A random sample of 8 math assistant professor's salaries are listed below in thousands of dollars. Estimate the population mean salary with a 99% confidence interval.

   66.0 75.8 70.9 73.9 63.4 68.5 73.3 65.9

3. The weight of the world's smallest mammal is the bumblebee bat (also known as Kitti's hog-nosed bat or *Craseonycteris thonglongyai*) and is approximately normally distributed with a mean of 1.9 grams. Such bats are roughly the size of a large bumblebee.

   A chiropterologist believes that the Kitti's hog-nosed bats in a new geographical region under study have a different average weight than 1.9 grams. A sample of 10 bats weighed in grams in the new region are shown below:

Weight (g)	1.9	2.24	2.13	2.00	1.54	1.96	1.79	2.18	1.81	2.30

   Use the **confidence interval** method to test the claim that the **mean weight** for all bumblebee bats is not 1.9 g using a **10% level of significance**.

4. A researcher is interested in estimating the average salary of teachers. She wants to be 95% confident that her estimate is correct. In a previous study, she found the population standard deviation was $1,175. How large a sample is needed to be accurate within $100?

5. A random sample of 20 nominally measured 2 mm diameter steel ball bearings is taken and the diameters are measured precisely. The measurements, in mm, are as follows:

2.02	1.94	2.09	1.95	1.98	2.00	2.03	2.04	2.08	2.07
1.99	1.96	1.99	1.95	1.99	1.99	2.03	2.05	2.01	2.03

   Assuming that the diameters are normally distributed with unknown mean μ and unknown variance σ^2:

   (a) Find a two-sided 95% confidence interval for the variance, σ^2.

   (b) Find a two-sided confidence interval for the standard deviation, σ.

6. The feeding habits of two species of net-casting spiders are studied. The species, the *deinopis* and *menneus*, coexist in eastern Australia. The following data were obtained on the size, in millimeters, of the prey of random samples of the two species:

   Size of Random Prey Samples of the Deinopis Spider in Millimeters

sample 1	**sample 2**	**sample 3**	**sample 4**	**sample 5**	**sample 6**	**sample 7**	**sample 8**	**sample 9**	**sample 10**
12.9	10.2	7.4	7.0	10.5	11.9	7.1	9.9	14.4	11.3

   Size of Random Prey Samples of the Menneus Spider in Millimeters

sample 1	**sample 2**	**sample 3**	**sample 4**	**sample 5**	**sample 6**	**sample 7**	**sample 8**	**sample 9**	**sample 10**
10.2	6.9	10.9	11.0	10.1	5.3	7.5	10.3	9.2	8.8

   What is the difference, if any, in the mean size of the prey (of the entire populations) of the two species?

7. Let's return to the example, in which the feeding habits of two species of net-casting spiders are studied. The species, the *deinopis* and *menneus*, coexist in eastern Australia. The following summary statistics were obtained on the size, in millimeters, of the prey of the two species:

Adult DEINOPIS	**Adult MENNEUS**
$n = 10$	$m = 10$
$\overline{x} = 10.26$ mm	$\overline{y} = 9.02$ mm
$s_X^2 = (2.51)^2$	$s_Y^2 = (1.90)^2$

   What is the difference in the mean sizes of the prey (of the entire populations) of the two species?

Confidence Intervals for the Difference of Means
Model:
Suppose we have two independent samples:

$$X_1, \dots, X_{n_1} \sim \mathcal{N}(\mu_1, \sigma_1^2),$$
$$Y_1, \dots, Y_{n_2} \sim \mathcal{N}(\mu_2, \sigma_2^2).$$

Theorem 4 (Difference of Means, Equal Variances).
If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, define the pooled variance:

$$S_p^2 = \frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}.$$

Then:

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}.$$

Hence, a $(1-\alpha) \times 100\%$ confidence interval for $\mu_1 - \mu_2$ is:

$$\left(\overline{X} - \overline{Y} \pm t_{\alpha/2,\, n_1+n_2-2} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right).$$

Theorem 5 (Difference of Means, Unequal Variances – Welch's Approximation).
If variances are not assumed equal, then:

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}} \sim t_\nu,$$

where the approximate degrees of freedom ν are given by:

$$\nu = \frac{\left(\dfrac{S_X^2}{n_1} + \dfrac{S_Y^2}{n_2}\right)^2}{\frac{\left(\frac{S_X^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_Y^2}{n_2}\right)^2}{n_2 - 1}}.$$

Thus, a $(1-\alpha) \times 100\%$ confidence interval for $\mu_1 - \mu_2$ is:

$$\left(\overline{X} - \overline{Y} \pm t_{\alpha/2,\, \nu} \cdot \sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}\right).$$