

Mathematical Statistics Recitation 13

TA: Sonia Reilly

May 1, 2026

Final Exam Review

Probability Review (Lectures 1-6)

- expectation, variance, covariance
- conditional + marginal probability
- convergence in distribution/probability, MGFs
- $\bar{X}_n \xrightarrow{p} \mu$ and $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$

Sampling and Confidence Intervals (Lec 7-8)

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \rightarrow N(0, 1), \text{ or } \sim t_{n-1} \text{ if } X_i \text{ are normal}$$

where S_n^2 is an unbiased estimate of variance:

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

$1 - \alpha$ confidence interval (CI) for μ : (use $t_{n-1, \alpha/2}$ instead if needed)

$$\bar{X}_n - z_{\alpha/2} \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\alpha/2} \frac{S_n}{\sqrt{n}}$$

Parameter Estimation (Lec 9-12)

1. method of moments (MOM)

$$\text{solve } E_{\theta}[X^k] = \frac{1}{n} \sum_{i=1}^n X_i^k \text{ for } k \text{ up to } \dim(\theta)$$

2. maximum likelihood estimation (MLE)

$$\hat{\theta}_n = \arg \max_{\theta} \log L(\theta) = \arg \max_{\theta} \sum_i \log p_{\theta}(X_i)$$

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$$

Fisher information:

$$I(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2}{\partial \theta^2} \log p_\theta(X) \right]$$

Cramér–Rao: any unbiased estimator has variance $\geq \frac{1}{nI(\theta)}$ (MLE asymptotically lowest)

3. Bayesian perspective

θ viewed as random variable (as opposed to fixed but unknown)

Bayes law: posterior \propto prior \times likelihood

$$\pi(\theta|X_1, \dots, X_n) \propto \pi(\theta) f_\theta(X_1, \dots, X_n)$$

estimators: posterior mode, mean, etc.

Hypothesis Testing Basics (Lec 13-15)

- tradeoff between $\alpha = P(\text{reject } H_0|H_0)$ and $\beta = P(\text{not reject } H_0|H_1)$
- Neyman–Pearson lemma: LR test makes the optimal tradeoff in simple vs. simple case
- p -value = prob. of same or more extreme data given H_0
= smallest level at which we would reject
- Generalized LR test:

$$\text{LR} = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)}, \quad \text{reject } H_0 \text{ if } \text{LR} < c$$

$$-2 \log \text{LR} \xrightarrow{d} \chi_d^2 \quad \text{under } H_0, \quad d = \dim(\Theta) - \dim(\Theta_0)$$

Nonparametric Methods (Lec 16)

1. Testing if X_i 's are from CDF $F(x)$

Empirical CDF = $F_n(x) = \frac{1}{n}$ (# of X_i less than x).

Glivenko-Cantelli Theorem: $\sup_x |F_n(x) - F(x)| \xrightarrow{P} 0$.

Kolmogorov distribution:

$$\sqrt{n} \sup_x |F_n(x) - F(x)| \xrightarrow{d} K, \quad P(K > K(\alpha)) = \alpha$$

Kolmogorov-Smirnov Test: reject when

$$\sup_x |F_n(x) - F(x)| > \frac{K(\alpha)}{\sqrt{n}}$$

2. Bootstrap Method

For estimator $\hat{\theta}$, want approx $\text{Var}(\hat{\theta})$ given only data, no info about distribution.

- Generate multiple samples with replacement from original data
- Compute estimates $\hat{\theta}_j$ from each sample
- Compute sample variance of estimates

Two Sample Testing (Lec 17-18)

1. Normal data

$$H_0 : \mu_X = \mu_Y, \quad H_1 : \mu_X \neq \mu_Y$$

σ unknown: Two-sample t -test rejects for large

$$\sqrt{-2 \log LR} \approx T = \frac{\bar{X}_n - \bar{Y}_m}{\text{approx } \sqrt{\text{Var}(\bar{X}_n - \bar{Y}_m)}} \sim t_{m+n-2} \quad \text{under } H_0$$

σ known, $m = n$: plug in $\text{Var}(\bar{X}_n - \bar{Y}_n) = \frac{2\sigma^2}{n}$, and then $T \sim N(0, 1)$

2. Nonparametric: Mann-Whitney Test

Sort $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ lowest to highest, $T_Y = \sum_j \text{rank}(Y_j)$.

For small m, n , can compute distribution exactly. For large m, n :

$$\frac{T_Y - \mathbb{E}[T_Y]}{\sqrt{\text{Var}(T_Y)}} \xrightarrow{d} N(0, 1) \quad \text{under } H_0$$

3. Paired Samples: (X_i, Y_i) independent, X_i not independent from Y_i

If data is normal: t -test. For $D_i = X_i - Y_i$,

$$\frac{\sqrt{n} \bar{D}_n}{\sqrt{\frac{1}{n-1} \sum (D_i - \bar{D}_n)^2}} \sim t_{n-1}$$

If nonparametric: Signed Rank Test

Order $|D_i|$ from lowest to highest, $W_+ = \sum_{D_i > 0} \text{rank}(|D_i|)$

$$\frac{W_+ - \mathbb{E}[W_+]}{\sqrt{\text{Var}(W_+)}} \xrightarrow{d} N(0, 1) \quad \text{under } H_0$$

Testing Categorical Data (Lec 19)

1. Fisher's Exact Test

2 rows, 2 columns. H_0 : distribution in col 1 = distribution in col 2.

Treat column, row sums as fixed. N_{11} determines distribution (hypergeometric).

Can calculate pmf exactly and compute 0.05 level rejection region.

Don't need $n \rightarrow \infty$. Impractical for large # samples.

2. Chi-Squared Test of Homogeneity

J distributions/columns, I categories/rows

Column sums fixed. Q: Are the distributions the same?

$$T = \sum_{i=1}^I \sum_{j=1}^J \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - n_i \cdot n_{.j} / n_{..})^2}{n_i \cdot n_{.j} / n_{..}} \rightarrow \chi_{(I-1)(J-1)}^2$$

3. Chi-Squared Test of Independence

Q: is the column variable independent of the row variable? Column assignment is random. Same test statistic, but for different question and setup.

Linear Regression (Lec 20-22)

1. Simple Linear Regression: Standard statistical model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad \mathbb{E}[e_i] = 0, \quad \text{Var}(e_i) = \sigma^2$$

Task: Given fixed x_i 's and the corresponding y_i 's, find best fit β_0, β_1 (intercept + slope).

Equations: easier to memorize the linear algebra ones and reduce to this case.

Can transform many nonlinear equations to linear form to do regression.

Plot residuals $\hat{e}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ to see if model is good. Should be independent of x_i .

Correlation ρ = slope of normalized x_i 's and y_i 's

2. Linear Algebra Formulation

$$y = X\beta + \epsilon$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots \\ 1 & x_{21} & x_{22} & \dots \\ \vdots & \vdots & \vdots & \dots \\ 1 & x_{n1} & x_{n2} & \dots \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

Each row of X is a data point, and each column is a feature.

Least squares:

$$\min_{\beta \in \mathbb{R}^p} \|y - X\beta\|_2^2.$$

Taking derivative in every β_i and setting to 0,

$$-2X^T(y - X\beta) = 0 \implies \underbrace{X^T X \hat{\beta} = X^T y}_{\text{"normal equations"}} \implies \hat{\beta} = (X^T X)^{-1} (X^T y).$$

Assumes $X^T X$ is invertible, which assumes $n \geq p$ (more data than features).

3. Statistics of Linear Regression

Random vectors:

- Covariance of z : $\Sigma = \Sigma_{zz} = \mathbb{E}[(z - \mathbb{E}[z])(z - \mathbb{E}[z])^T]$.
- $\Sigma_{ij} = \text{Cov}(z_i, z_j)$, $\Sigma_{ii} = \text{Var}(z_i)$, Σ is symmetric.
- Linearity of expectation: $\mathbb{E}[Az] = A\mathbb{E}[z]$.
- Covariance of linear transformation: $\Sigma_{Az, Az} = A\Sigma_{zz}A^T$.

Expectation and covariance of $\hat{\beta}$:

$$\mathbb{E}[\hat{\beta}] = \beta, \quad \Sigma_{\hat{\beta}\hat{\beta}} = \sigma^2 (X^T X)^{-1}$$

Estimated standard error:

$$\hat{s}_{\beta_i} = s \sqrt{(X^T X)^{-1}_{ii}}, \quad \text{where } s^2 = \frac{\|y - X\hat{\beta}\|_2^2}{n - p}$$

For CI's and hypothesis tests, use

$$\frac{\hat{\beta}_i - \beta_i}{\hat{s}_{\beta_i}} \xrightarrow{n \rightarrow \infty} N(0, 1) \quad \text{or if } e \text{ normal, } \frac{\hat{\beta}_i - \beta_i}{\hat{s}_{\beta_i}} \sim t_{n-p}$$

If x_i random, $\hat{\beta}$ is still unbiased but $\Sigma_{\hat{\beta}\hat{\beta}} = \sigma^2 \mathbb{E}[(X^T X)^{-1}]$.

Problems

Linear Regression

1. Suppose $X_0 = 0$ and X_1, \dots, X_n are given by the following model, known as a Gaussian autoregressive process:

$$X_i = \theta X_{i-1} + \varepsilon_i$$

with $\theta \in [0, 1)$ unknown, and the errors $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d. with $\sigma^2 > 0$ known.

- (a) Write down the log-likelihood function for the unknown parameter θ .
- (b) Find $\hat{\theta}_{\text{MLE}}$.
- (c) Find the least squares estimate $\tilde{\theta}$.

Solution:

- (a) At each step X_{i-1} is known and X_i is determined by the random ε_i , so

$$f_{X_i|X_{i-1}}(X_i|X_{i-1}) = f_{\varepsilon_i}(\varepsilon_i) \quad \text{and} \quad f_{X_1, \dots, X_n}(X_0, \dots, X_n) = f_{\varepsilon_1, \dots, \varepsilon_n}(\varepsilon_1, \dots, \varepsilon_n).$$

$$\varepsilon_i = X_i - \theta X_{i-1} \quad \text{has distribution } \mathcal{N}(0, \sigma^2),$$

or pdf

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\varepsilon_i}{\sigma}\right)^2\right).$$

So the likelihood is

$$\mathcal{L}_n(\theta) = \prod_{i=1}^n f(\varepsilon_i) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \theta X_{i-1})^2\right).$$

and the log-likelihood is

$$\ell_n(\theta) = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \theta X_{i-1})^2.$$

- (b)

$$\ell'_n(\theta) = \frac{1}{\sigma^2} \sum (X_i - \theta X_{i-1}) X_{i-1} = 0$$

$$\Rightarrow \sum X_i X_{i-1} - \theta \sum X_{i-1}^2 = 0 \quad \Rightarrow \quad \hat{\theta}_{\text{MLE}} = \frac{\sum X_i X_{i-1}}{\sum X_{i-1}^2}.$$

- (c) Set up as a linear regression problem:

$$\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} X_0 \\ \vdots \\ X_{n-1} \end{bmatrix} \theta + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \text{where } \theta \text{ is 1-dimensional.}$$

$$X_{\text{new}} = X_{\text{old}} \theta + \varepsilon$$

$$\tilde{\theta} = (X_{\text{old}}^\top X_{\text{old}})^{-1} (X_{\text{old}}^\top X_{\text{new}}) = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2}.$$

(normal equations: $\hat{\beta} = (X^\top X)^{-1} X^\top Y$)

MLE and linear regression give the same answer!

Hypothesis Testing

1. You collect data from one population and you want to test if the data follows a particular distribution. What test would you run in the following scenarios, or have we not covered an appropriate test?
 - (a) The data is categorical (or discrete).
 - i. You wish to test for the distribution (i.e. nonparametric test). **Pearson's Chi-squared Test**
 - ii. You wish to test for a parameter estimate. **Normality of MLE**
 - (b) The data is (continuous) numerical.
 - i. You wish to test for the distribution (i.e. nonparametric test). **Kolmogorov-Smirnov**
 - ii. You wish to test for a parameter estimate. **Normality of MLE**
2. You collect data from two populations and want to compare the distributions for both populations. What test would you run?
 - (a) The data is categorical and both populations can only take on two values.
 - i. Data is unpaired. **Chi-squared test of homogeneity, or Fisher's exact test**
 - ii. Data is paired **not covered**
 - (b) The data is continuous:
 - Data is unpaired
 - Populations follow normal distribution. **Two-sample t-test**
 - We want to compare the unknown distributions. **Mann-Whitney**
 - Data is paired
 - Populations follow normal distribution. **Paired t-test**
 - We want to compare the unknown distributions. **Signed Rank Test**
 - (c) The data is categorical and both populations can take on multiple values. **Chi-squared test of homogeneity**
 - (d) What test would you run if you wanted to compare the distributions of multiple populations? **If categorical, chi-squared. Else, not covered.**