High Dimensional Statistics

Spring 2025

Sharif University of Technology

Dr. Amir Najafi

Homework 1 Classical Statistics and Concentration Bounds Due: 1403/12/24

Problem 1: MLE in Linear Regression *[12 points]*

Given the values $(x_i)_{i=1}^n$ and the linear regression problem with data points $\{(x_i, y_i)\}_{i=1}^n$, the relationship is defined as:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where $\epsilon_i \sim N(0, \sigma^2)$, and β_0, β_1 are non-negative parameters.

(a) Prove that the maximum likelihood estimates (MLEs) of β_0 and β_1 are equivalent to the values that minimize the mean squared error.

(b) Prove that the obtained estimators are unbiased and follow the distributions:

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2 \sum_i x_i^2}{n \sum_i (x_i - \bar{x})^2}\right), \quad \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}\right)$$

(c) Investigate whether the MLE estimator belongs to the family of linear estimators defined by:

$$\tilde{\beta}_1 = \frac{\sum \gamma_i y_i}{\sum \gamma_i x_i} \quad \text{such that} \quad \sum_i \gamma_i = 1.$$

If it does, find the relation between γ_i and the inputs.

(d) Prove that every estimator in the above family is unbiased.

(e) Prove that for any choice of γ_i in the above family, the following inequality holds.

$$Var(\hat{\beta}_1) \leq Var(\tilde{\beta}_1)$$

.

Problem 2: Two-Sample Mean Comparison Test *[10 points]*

An experiment is planned to compare the mean of a control group to the mean of an independent sample of a group given a treatment. Suppose that there are to be 25 samples in each group. Suppose that the observations are approximately normally distributed and that the standard deviation of a single measurement in either group is $\sigma = 5$.

(a) What will the standard error of $\overline{Y} - \overline{X}$ be?

(b) With a significance level $\alpha = 0.05$, what is the rejection region of the test of the null hypothesis $H_0 : \mu_Y = \mu_X$ versus the alternative $H_A : \mu_Y > \mu_X$?

(c) What is the power of the test if $\mu_Y = \mu_X + 1$?

(d) Suppose that the p-value of the test turns out to be 0.07. Would the test reject at significance level $\alpha = 0.10$?

(e) What is the rejection region if the alternative is $H_A : \mu_Y \neq \mu_X$? What is the power if $\mu_Y = \mu_X + 1$?

Problem 3: Cramér-Rao Lower Bound *[12 points]*

Assume that $T = t(X)$ is an estimator with expectation $\psi(\theta)$ (based on the observations X), i.e., that

$$\mathbb{E}(T) = \psi(\theta).$$

Prove that, for all θ,

$$\mathrm{var}(t(X)) \geq \frac{[\psi'(\theta)]^2}{I(\theta)}.$$

where $I(\theta)$ is the **Fisher information**, defined as:

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta} \log f(X;\theta)\right)^2\right],$$

where $f(X;\theta)$ is the probability density (or mass) function of X given the parameter θ.

Problem 4: Wishart Matrices *[10 points]*

Let $W \sim \mathcal{W}_p(n, \Sigma)$ be a random matrix distributed according to the Wishart distribution with p dimensions, n degrees of freedom, and scale matrix Σ. Specifically, the random matrix W is formed by sampling $x_1, \ldots, x_n$ i.i.d. samples from a p-variate normal distribution with zero mean and covariance matrix Σ, and then constructing W as follows:

$$W = \sum_{i=1}^{n} x_i x_i^T.$$

(a) Compute $\mathbb{E}[W]$.

(b) Calculate the covariance between each pair of elements of W.

(c) The moment generating function (MGF) of a random matrix is generally defined as $M_X(T) = \mathbb{E}\left[\exp\left(\mathrm{tr}(TX)\right)\right]$, where T is a $p \times p$ matrix. Prove that, given that $I - 2\Sigma T$ is positive semi-definite, the MGF of the Wishart distribution is:

$$M_W(T) = \det(I - 2\Sigma T)^{-n/2}.$$

Problem 5: Mills Ratio *[12 points]*

Let $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ be the density function of a standard normal $Z \sim \mathcal{N}(0,1)$ variate.

(a) Show that $\varphi'(z) + z\varphi(z) = 0$.

(b) Use part (a) to show that

$$\varphi(z)\left(\frac{1}{z} - \frac{1}{z^3}\right) \leq P[Z \geq z] \leq \varphi(z)\left(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5}\right)$$

for all $z > 0$.

Problem 6: Sharp Upper Bounds on Binomial Tails *[14 points]*

Let $\{X_i\}_{i=1}^n$ be an i.i.d. sequence of Bernoulli variables with parameter $\alpha \in (0, 1/2]$, and consider the binomial random variable $Z_n = \sum_{i=1}^n X_i$. The goal of this exercise is to prove, for any $\delta \in (0, \alpha)$, a sharp upper bound on the tail probability $P[Z_n \leq \delta n]$.

(a) Show that $P[Z_n \leq \delta n] \leq e^{-nD(\delta \| \alpha)}$, where the quantity

$$D(\delta \| \alpha) := \delta \log \frac{\delta}{\alpha} + (1-\delta) \log \frac{1-\delta}{1-\alpha}$$

is the Kullback–Leibler divergence between the Bernoulli distributions with parameters δ and α, respectively.

(b) Show that the bound from part (a) is strictly better than the Hoeffding bound for all $\delta \in (0, \alpha)$.

Problem 7: Upper Bounds for Sub-Gaussian Maxima *[14 points]*

Let $\{X_i\}_{i=1}^n$ be a sequence of zero-mean random variables, each sub-Gaussian with parameter σ.

(a) Prove that

$$\mathbb{E}\left[\max_{i=1,\ldots,n} X_i\right] \leq \sqrt{2\sigma^2 \log n} \quad \text{for all } n \geq 1.$$

(b) Prove that the random variable $Z = \max_{i=1,\ldots,n} |X_i|$ satisfies

$$\mathbb{E}[Z] \leq \sqrt{2\sigma^2 \log(2n)} \leq 2\sqrt{\sigma^2 \log n},$$

valid for all $n \geq 2$.

Problem 8: Concentration Around Medians and Means *[12 points]*

Given a scalar random variable X, suppose that there are positive constants c_1, c_2 such that

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq c_1 e^{-c_2 t^2} \quad \text{for all } t \geq 0.$$

(a) Prove that $\text{var}(X) \leq \frac{c_1}{c_2}$.

(b) A median m_X is any number such that $\mathbb{P}(X \geq m_X) \geq 1/2$ and $\mathbb{P}(X \leq m_X) \geq 1/2$. Show by example that the median need not be unique.
(c) Show that whenever the mean concentration bound holds, then for any median m_X, we have

$$\mathbb{P}(|X - m_X| \geq t) \leq c_3 e^{-c_4 t^2} \quad \text{for all } t \geq 0,$$

where $c_3 := 4c_1$ and $c_4 := \frac{c_2}{8}$.
(d) Conversely, show that whenever the median concentration bound holds, then mean concentration holds with $c_1 = 2c_3$ and $c_2 = \frac{c_4}{4}$.

Problem 9: Tail Bounds under Moment Conditions *[12 points]*

Suppose that $\{X_i\}_{i=1}^n$ are zero-mean and independent random variables such that, for some fixed integer $m \geq 1$, they satisfy the moment bound $\|X_i\|_{2m} := (\mathbb{E}[X_i^{2m}])^{\frac{1}{2m}} \leq C_m$. Show that

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \geq \delta\right] \leq B_m \left(\frac{1}{\sqrt{n}\delta}\right)^{2m} \quad \text{for all } \delta > 0,$$

where B_m is a universal constant depending only on C_m and m.
Hint: You may find the following form of Rosenthal's inequality to be useful. Under the stated conditions, there is a universal constant R_m such that

$$\mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^{2m}\right] \leq R_m \left(\sum_{i=1}^n \mathbb{E}[X_i^{2m}] + \left(\sum_{i=1}^n \mathbb{E}[X_i^2]\right)^m\right).$$

Problem 10: Sub-Gaussian Norm Bound *[12 points]*

Let $X \in \mathbb{R}^d$ be a sub-Gaussian random vector with variance proxy σ^2. Then

$$\mathbb{E}\left[\max_{\theta \in \mathcal{B}_2} \theta^\top X\right] = \mathbb{E}\left[\max_{\theta \in \mathcal{B}_2} |\theta^\top X|\right] \leq 4\sigma\sqrt{d}.$$

Moreover, for any $\delta > 0$, with probability $1 - \delta$, it holds

$$\max_{\theta \in \mathcal{B}_2} \theta^\top X = \max_{\theta \in \mathcal{B}_2} |\theta^\top X| \leq 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)}.$$