Fangyuan's Collection of Exercises in Probability Theory and Statistics

Fangyuan Lin

August 23, 2025

Contents

Linear Algebra

1.1 Fundamentals

1.1.1 Eigenvalue and Eigenvector

The eigenvalues of A^2

If A has eigenvalues λ_i. Then the eigenvalues of A^2 are λ_i^2.

Proof. Well, my first intuition is to think about the diagonalization of A and the result becomes clear.
A rigorous proof is also not hard:

1. $$A\vec{v} = \lambda\vec{v} \implies A^2\vec{v} = A\lambda\vec{v} = \lambda^2\vec{v}.$$

2. The algebraic multiplicity of the eigenvalues λ_i^2 of A^2 is the same as the eigenvalues λ_i of A:

$$\det\left(A^2 - \lambda^2 I\right) = \det(A + \lambda I)\det(A - \lambda I).$$

This means that

□

1.1.2 Orthogonality

Orthogonal Matrices have the following properties:

- $U^{-1} = U^T$
- Rows and columns are orthogonal unit vectors.
- Preserves the inner product of vectors: $\langle x, y\rangle = \langle Ux, Uy\rangle$.
- Isometric: length/distance preserving.
- Rigit rotation, reflection, rotoreflection.

1.1.3 Singular Value Decomposition

The Principal Component Analysis PCA is a byproduct of SVD. Let X denote the original data matrix. Let X^* be the (column-)centered matrix. (In machine learning, each row represents a data point and each column represents a feature). Let $\hat{X}$ denote the centered, normalized matrix.

- ${X^*}^T X^*$ is the covariance matrix.
- $\hat{X}^T \hat{X}$ is the correlation matrix.
- $X^T X$ is the cross-product.

Consider an $n \times d$ matrix A

Singular Value Decomposition

Let A be an $n \times d$ matrix with singular vectors $v_1, \dots, v_r$ and corresponding singular values $\sigma_1, \dots, \sigma_r$. Then $u_i = \frac{1}{\sigma_i} A v_i$ are the left singular vectors and A can be decomposed into a sum of rank one matrices as

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

- Compute $A^T A$. This is a symmetric and positive semi-definite matrix, so it has orthonormal eigenvectors v_i and non-negative eigenvalues λ_i by the Spectral theorem.
- The singular values are $\sigma_i = \sqrt{\lambda_i}$
- $V = [v_i]$ is the matrix of right singular values.
- Define $u_i = \frac{1}{\sigma_i} A v_i$

Intuition behind SVD

Consider the optimization problem:

$$\max_{\|v\|=1} \|Av\| = \max_{\|v\|=1} v^T A^T A v$$

This is a Rayleigh quotient of $A^T A$, so its maximum value corresponds to its largest eigenvalue!

- If we define $u_i = \frac{1}{\sigma_i} A v_i$, then we get $A v_i = \sigma_i u_i$ and u_i are orthonomal vectors.

Let's verify the algorithm is correct:

$$u_i^T A v_i = (\frac{1}{\sigma_i} A v_i)^T A v_j = \frac{\sigma_i^2}{\sigma_i} = \sigma_i$$

- We can think of A as three steps: rotation V^T, then horizontal/vertical scaling (Σ), lastly rotation U.

- Use

```
from scipy import linalg
U, s, Vh = linalg.svd(X)
```

  to compute the SVD of matrix X.

- Any matrix can be quickly decomposed into SVD form.
- One important and obvious application of SVD is data compression.

Apply SVD to a two by two matrix

Let $M = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, rotation by ninety degrees clockwise. Note that this matrix has complex eigenvalue of $-i$. Let's compute the SVD of it.

$$SVD(M) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

- V^T is reflection about the y-axis / rotation by 180 degress about y-axis.
- Σ is the trivial scaling.
- U is rotation by 180 degress about the origin. (reflection about both the x and y-axes).
- The rank 1 approximation would be $\begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}$

1.1.4 Determinant

The Determinant is the Product of Eigenvalues

Proof. Let A be a matrix with eigenvalues λ_i. The key idea of the proof uses the characteristic polynomial.

1. Consider the characteristic polynomial

$$p(\lambda) = |\lambda I - A| = c_0 + c_1\lambda + \cdots + \lambda^n$$

   Note that the characteristic polynomial is monic.

2. We can obtain c_0 by

$$p(0) = c_0 = |0 \cdot I - A| = (-1)^n \det A$$

3. Note that the eigenvalues λ_i are roots of the characteristic polynomial so

$$p(0) = \prod_i (0 - \lambda_i) = (-1)^n \lambda_i$$

4. Lastly,

$$c_0 = (-1)^n \prod_i \lambda_i = (-1)^n \det A$$

so

$$\det A = \prod_i \lambda_i$$

□

1.1.5 Trace

Trace is Equal to the Sum of Eigenvalues

Trace is Equal to the Sum of Eigenvalues

Let A be an $n \times n$ matrix with eigenvalues λ_i. Show that

$$\mathrm{Tr}(A) = \sum_i \lambda_i$$

Solution. 1. The proof is similar to that of "the determinant is product of eigenvalues," i.e. we work with the characteristic polynomial. TO BE FILLED IN

■

An Inequality relating Trace and Determinant

2018 Summer Practice Problem, # 18

Suppose Σ is a non-negative definite matrix of $n \times n$ real entries and real eigenvalues. Show that

$$\mathrm{Tr}\left(\Sigma^2\right) \geq n \cdot \det(\Sigma)^{2/n}.$$

Solution. 1. Let $\{\lambda_i\}$ be the eigenvalues of Σ. To make some progress, let's write the trace as

$$\mathrm{Tr}(\Sigma) = \sum_i \lambda_i^2$$

2. By the Arithmetic Mean - Geometric Mean inequality,

$$\frac{\sum_{i=1}^n \lambda_i^2}{n} \geq \sqrt[n]{\prod_{i=1}^n \lambda_i^2} \implies \mathrm{Tr}\left(\Sigma^2\right) \geq n \det(\Sigma)^{\frac{2}{n}}$$

■

1.2 Core Competency Exam Questions

1.2.1 Problem 1

1: 2020 September Exam, #8

For every $n \geq 1$, let A_n be an $n \times n$ symmetric matrix with non-negative entries. Let $R_n(i) := \sum_{j=1}^n A_n(i,j)$ denote the ith row/column sum of A_n. Assume that

$$\lim_{n\to\infty} \max_{1\leq i\leq n} |R_n(i) - 1| = 0.$$

Let $\lambda_n \geq 0$ denote an eigenvalue with the largest absolute value, and let $\vec{x} = (x_1, \dots, x_n)$ denote its corresponding eigenvector.

- Show that
$$\frac{1}{n}\sum_{i,j=1}^n A_n(i,j) \to 1$$
- Show that $\lambda_n |x_i| \leq \max_{1\leq j\leq n} |x_j| R_n(i)$.
- Using parts one and two, show that
$$\lambda_n \to 1.$$

Solution. For the first part, let's just write something down:

1.
$$\frac{1}{n}\sum_{i,j=1}^n A_n(i,j) = \frac{1}{n}\sum_{i=1}^n R_n(i)$$

2.
$$\left|\frac{1}{n}\sum_{i,j=1}^n A_n(i,j) - 1\right| = \left|\frac{1}{n}\sum_{i=1}^n R_n(i) - 1\right|$$
$$\leq \max_{1\leq i\leq n} |R_n(i) - 1| \to 0$$

For the second part,

1. By assumption,

$$A_n\vec{x} = \lambda_n\vec{x}, \quad \lambda_n x_i = \sum_{j=1}^n A_n(i,j)x_j$$
$$\lambda_n|x_i| \leq \sum_{j=1}^n A_n(i,j)|x_j| = R_n(i)\max_{1\leq j\leq n}|x_j|.$$

For the third part, we first use the Rayleigh quotient. For any nonzero vector $v \in \mathbb{R}^n$,

$$\begin{aligned}\lambda_n &= \max_{\|u\|_2=1} u^T A_n u \geq \max_{\|u\|_2=1} \sum_{i,j=1}^{n} A_n(i,j) u_i u_j \\ &\geq \sum_{i,j=1}^{n} A_n(i,j) \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n}} = \frac{1}{n} \sum_{i,j} A_n(i,j) \to 1.\end{aligned}$$

For the other direction, we use part two. Choose k such that $|x_k| = \max_j |x_j|$

$$\lambda_n \leq \frac{x_k}{x_k} R_n(k) \to 1$$

■

1.2.2 Problem 2

2: (Straightforward) 2021 May Exam, #7

Suppose that $A = (a_{ij})_{1 \leq i,j \leq 2}$ is a 2×2 symmetric matrix, with $a_{11} = a_{22} = \frac{3}{4}$ and $a_{12} = a_{21} = \frac{1}{4}$.

- Find the eigenvalues and eigenvectors of the matrix A.
- Compute $\lim_{n \to +\infty} a_{12}^{(n)}$ where $a_{i,j}^{(n)}$ denotes the ijth entry of the matrix A^n.

Solution. The first part is standard. Set up the characteristic polynomial and solve for its roots:

$$p(\lambda) = \det(A - \lambda I) = 0 \implies \lambda = \frac{1}{2}, 1$$

The eigenvector corresponding to $\lambda = 1$ is $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. The eigenvector corresponding to $\lambda = \frac{1}{2}$ is $\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$.

For the second part. We should use diagonalization; otherwise, matrix exponential would be hard to compute.

$$A = PDP^{-1}$$

where D is the diagonal matrix whose diagonal entries are the eigenvalues. P^{-1} is the matrix whose the columns are the corresponding eigenvectors. So $D = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$ and $P^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$.

$$\begin{aligned}A^n = P \begin{bmatrix} 1^n & 0 \\ 0 & \frac{1}{2}^n \end{bmatrix} P^{-1} &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.5^n \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0.5^n & -(0.5^n) \end{bmatrix}\end{aligned}$$

$$a_{12}^n = \frac{1}{2} - \frac{1}{2} \cdot (-(0.5^n)) \to \frac{1}{2}.$$

This question is straightforward in my opinion! ■

1.2.3 Problem 3

3: 2021 Sept Exam, #6

Let $A \in \mathbb{R}^{m \times n}$ be an $m \times n$ matrix with $n < m$. Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_n$ and $\vec{v}_1, \ldots, \vec{v}_n$ are the eigenvalues and eigenvectors of $A^T A$. What can we say about ALL the eigenvalues and eigenvectors of AA^T. Justify your answer.

Solution. When it comes AA^T, especially when A is non-symmetric or even non-square, we should think of Singular Value Decomposition SVD! Let $A = U\Sigma V^{-1}$ be its SVD. Then $A^T = V\Sigma^T U^{-1}$. The singular values of A are the square root of the eigenvalues of AA^T, and we see that A and A^T share the same singular values. Note that U is composed of orthonormal eigenvectors of AA^T and V is composed of orthonormal eigenvectors of $A^T A$. $AA^T \vec{v}_i = \lambda \vec{v}_i$ ■

1.2.4 Problem 4

4: Eigenvalue of Orthogonal Matrix

Let A be a 3×3 real-valued matrix such that $A^T A = AA^T = I_3$ and $\det(A) = 1$. Prove that 1 is an eigenvalue of A.

Solution. Since the problem wants to tell us that A is orthogonal, we should be thinking of the length-preserving property. Let λ be an eigenvalue of A and $\vec{v}$ be a corresponding unit eigenvector. Then

$$\|A\vec{v}\| = \sqrt{\vec{v}^T A^T A \vec{v}} = 1 = \|\lambda \vec{v}\| = |\lambda|$$

.

The determinant is the product of the eigenvalues and -1 cannot be the only eigenvalue of A because $(-1)^3 = -1 \neq 1 = \det A$. ■

1.2.5 Problem 5

5: (Straightforward) Trace of the square of a symmetric matrix is zero means zero matrix

Let A be an $n \times n$ symmetric matrix such that $\mathrm{Tr}(A^2) = 0$. Show that $A = 0_{n \times n}$. Hint: Use the fact that $\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB)$.

Solution. The hint apparently wants us to apply the spectral theorem to obtain a diagonalization $A = Q\Lambda Q^T$.

$$\mathrm{Tr}\, A^2 = \mathrm{Tr}\left(Q\Lambda^2 Q^T\right) = \mathrm{Tr}\left(Q^T Q \Lambda^2\right) = \mathrm{Tr}\big(\Lambda^2\big) = 0.$$

The trace is equal to the sum of the eigenvalues (to be honest, with this fact, we don't really need the hint), i.e. the diagonal of Λ^2 is zero. Since the entries of Λ^2 are non-zero, $\Lambda^2 = 0$ and hence $\Lambda = 0$. Therefore $A = 0$. ■

1.2.6 Problem 6

6: Eigenvectors are the same iff Multiplication commutes

Let $A, B \in \mathbb{R}^{n\times n}$ have respective eigendecompositions $Q_1 D_1 Q_1^T$ and $Q_2 D_2 Q_2^T$ (recall this means each D_i is a diagonal matrix of eigenvalues and each Q_i is an orthogonal matrix). Prove that $Q_1 = Q_2$ if and only if $AB = BA$. You may assume that A, B do not have any repeated eigenvalues.

Solution. Suppose $AB = BA$, consider an eigenpair λ and $\vec{v}$ of A.

$$BA\vec{v} = \lambda B\vec{v}. = AB\vec{v}.$$

This means that $B\vec{v}$ is an eigenvector of A corresponding to the eigenvalue λ. This then imply to A and B share the same set of eigenvalues λ_i with corresponding eigenvectors $\vec{v}_i$ and $B\vec{v}_i$. For $Q_1 = Q_2$, we need to show that $\vec{v}_i \propto B\vec{v}$:

$$AB\vec{v} = \lambda B\vec{v}, \implies B\vec{v} \propto \vec{v}$$

since the eigenspaces of A are all one-dimensional.
The other direction is easier. Suppose $Q_1 = Q_2$, then

$$AB = Q_1 D_1 Q_1^T Q_2 D_2 Q_2^T = Q_2 D_2 Q_2^T Q_1 D_1 Q_1^T = BA$$

■

1.2.7 Problem 7

7: (Straightforward) Eigenvalue of uv^T

Let $A = uv^T \in \mathbb{R}^{n\times n}$ be a rank-one matrix, i.e. $u, v \in \mathbb{R}^n$. Suppose $u, v \neq 0_n$. Find, with proof, all the eigenvalues of A.

Solution. Let λ be an eigenvalue of A and $\vec{x}$ be a corresponding eigenvector, then

$$A\vec{x} = uv^T\vec{x} = \lambda\vec{x}$$

Note that

$$uv^T x = u\langle v, x\rangle = \lambda\vec{x}$$

This means that $\vec{x}$ and $\vec{u}$ share the same direction. So

$$A\vec{u} = uv^T u = \lambda u$$

Therefore,

$$\lambda = \vec{v}^T\vec{u}$$

There can be no other eigenvalues because A has rank-one.
Comment: Should find this problem straightforward. ■

1.2.8 Problem 8

Heisenberg Uncertainty Principle

Suppose $A, B \in \mathbb{R}^{n\times n}$ are symmetric matrices satisfying $AB + BA = Id$. Show that for all vectors $v \in \mathbb{R}^n \backslash \{0_n\}$,

$$\max\{\frac{\|Av\|_2}{\|v\|_2}, \frac{\|Bv\|_2}{\|v\|_2}\} \geq 1/\sqrt{2}.$$

Solution. This seems to be an interesting problem. Here's my thought process (which turned out to be misleading tho, because the canonical solution is simple): since the problem mentioned A, B are symmetric and we are apparently dealing with Rayleigh quotient, and a maximization problem, so we should definitely look at the eigenvalue of A, B.

1. Since A and B are real symmetric, so is

$$(A - B)^2.$$

But for any symmetric X,

$$X^2 \succeq 0 \qquad \big(\text{i.e. } v^T X^2 v = \|Xv\|^2 \geq 0\big).$$

Hence

$$(A-B)^2 \succeq 0 \implies A^2-(AB+BA)+B^2 \succeq 0 \implies A^2+B^2 \succeq (AB+BA) = I.$$

2. Loewner order $A^2 + B^2 \succeq I$ means all eigenvalues of $A^2 + B^2$ are ≥ 1. Equivalently, for every v

$$v^T(A^2 + B^2)\, v \geq v^T v \implies \|Av\|^2 + \|Bv\|^2 \geq \|v\|^2.$$

3. Finally,

$$\max\{\|Av\|^2, \|Bv\|^2\} \geq \frac{\|Av\|^2 + \|Bv\|^2}{2} \geq \frac{\|v\|^2}{2},$$

so

$$\max\left\{\frac{\|Av\|}{\|v\|}, \frac{\|Bv\|}{\|v\|}\right\} \geq \frac{1}{\sqrt{2}},$$

The canonical clean solution is actually to consider the inner product

$$\begin{aligned}\|v\|_2^2 = v^T v = v^T I v &= v^T(AB + BA)v \\ &= v^T ABv + v^T BAv \\ &\leq 2|\langle Av, Bv\rangle| \\ &\leq 2\|Av\|_2\|Bv\|_2\end{aligned}$$

where the last inequality is by Cauchy schwarz.
Finally, one of $\frac{\|Av\|_2}{\|v\|_2}$ and $\frac{\|Bv\|_2}{\|v\|_2}$ must be greater than $\frac{1}{\sqrt{2}}$.
Comment: This teaches us a lesson that whenever we see the 2-norm, consider playing with $v^T v$. I was too obsessed with eigenvalues... ■

1.2.9 Problem 9

Invertibility

Let $A = (a_{ij})$ be an $n \times n$ real matrix whose diagonal entries a_{ii} satisfy $a_{ii} \geq 1$ for all i. Suppose also $\sum_{i \neq j} a_{ij} < 1$. Prove that the inverse matrix A^{-1} exists.

Solution. I gave myself the hint that I can use proof by contradiction since the statement is boolean.

1. Suppose by contradiction that A is not invertible. Then the kernel of A is non-trivial. (I did this step correctly).

2. We know that the diagonal entries are all greater than 1 and the sum of all non-diagonal entries is less than 1.

3. This means that there exists $\vec{v} \neq \vec{0}$ such that $A\vec{v} = \vec{0}$ and assume that $\|\vec{v}\| = 1$. Then

$$\begin{aligned}
&\sum_{j=1}^{n} a_{ij} v_j = 0 \\
&\implies a_{ii} v_i = -\sum_{j:j \neq i} a_{ij} v_j \\
&\implies \sum_{i=1}^{n} a_{ii}^2 v_i^2 = \sum_{i=1}^{n} (\sum_{j:j \neq i} a_{ij} v_j)^2
\end{aligned}$$

4. However,

$$\sum_{i=1}^{n} a_{ii}^2 v_i^2 \geq 1$$

but

$$\sum_{i=1}^{n} (\sum_{j:j \neq i} a_{ij} v_j)^2 \leq \sum_{i=1}^{n} \sum_{j:j \neq i} (a_{ij}^2) \sum_{j:j \neq i} (v_j)^2 = \sum_{i=1}^{n} \sum_{j:j \neq i} a_{ij}^2 < 1$$

■

1.2.10 Problem 10

9: Greshgorin Circle Problem

Let $A \in \mathbb{C}^{n \times n}$ with entries a_{ij}. Let

$$R_i := \sum_{j \neq i} |a_{ij}|$$

Let $D(a_{ii}, R_i)$ be a closed disc centered at a_{ii} with radius R_i, called the Greshgorin disc. Show that every eigenvalue of A lies at least in one of the Goshgorin discs of A.

Solution. This appears to be a hard and interesting problem. My first intuition is proof by contradiction.

1. Suppose there exists λ (with eigenvector $\vec{v}$) that doesn't lie in any Goshgorin disc: for all i
$$|\lambda - a_{ii}| > R_i = \sum_{j\neq i} |a_{ij}|$$

2. We need to reach a contradiction out of this. Let me try considering eigen vectors $\vec{v}$ such that the absolute value of the entries is upper bounded by 1. Then
$$\begin{aligned} &Av = \lambda v \\ &\sum_{j=1}^{n} a_{ij} v_j = \lambda v_i \\ &\sum_{j:j\neq i} a_{ij} v_j + a_{ii} v_i = \lambda v_i \end{aligned}$$

   Then we get
$$\sum_{j:j\neq i} |a_{ij}| \geq |\sum_{j:j\neq i} a_{ij}| = |v_i(\lambda - a_{ii})| = |\lambda - a_{ii}|$$
   which is a contradiction.

■

1.2.11 Problem 11

10: Limit of A^n

Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$. Find $\lim_{n\to\infty} A^n$. Hint: the eigenvalues of a lower triangular matrix are its diagonal entries.

Solution. This is a typical diagonalization problem. From the hint, we know that the eigenvalues are 1, 1/2, and 1/3.

- Solving $\begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, We get that $\frac{1}{2}x_1 + \frac{1}{2}x_2 = x_2 \implies x_1 = x_2$.

  From the last equation, we know that $x_1 = x_2 = x_3$. An eigenvector is just $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

- Solving $\begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

■

1.2.12 Problem 12

11: AB and BA have the same eigenvalues

Show that AB and BA have the same eigenvalues if A is invertible.

Solution. The proof is quite short. Let λ be an eigenvalue of AB and $\vec{v}_i$ be corresponding eigenvectors. Then

$$\begin{aligned} AB\vec{v} &= \lambda\vec{v} \\ \implies B\vec{v} &= \lambda A^{-1}\vec{v} \\ \implies BA(A^{-1}\vec{v}) = B\vec{v} &= \lambda(A^{-1}\vec{v}) \end{aligned}$$

This shows that λ is an eigenvalue of BA with the same multiplicity since A^{-1} is also an invertible linear transformation.
An alternative solution shows that AB and BA share the same characteristic polynomial:

$$\det(AB-\lambda I) = \det\big(A^{-1}A(AB-\lambda I)\big) = \det\big(A^{-1}(AB-\lambda I)A\big) = \det(BA-\lambda I)$$

■

1.2.13 Problem 13

Invertibility Exercise

Let $A = (a_{ij})$ be a 2×2 real matrix such that

$$a_{11}^2 + a_{12}^2 + a_{21}^2 + a_{22}^2 < \frac{1}{1000}$$

Show that $I + A$ is invertible.

Solution. My intuition is to look at the determinant of $I + A$ since the determinant of a two by two matrix is very easy to compute.

$$\begin{aligned} \det(I+A) &= (1+a_{11})(1+a_{22}) - a_{12}a_{21} \\ &= 1 + a_{11} + a_{22} + a_{11}a_{22} - a_{12}a_{21} \end{aligned}$$

One thing we can do is to use the fact that $|a_{ij}| \leq \frac{1}{\sqrt{1000}} \leq \frac{1}{10}$. The statement in the problem is true even if

$$a_{11}^2 + a_{12}^2 + a_{21}^2 + a_{22}^2 < 1$$

An alternative solution: Suppose by contradiction that $I+A$ has a non-trivial kernel, i.e. there exists $\vec{v}$ such that $(I+A)\vec{v} = 0$. Then

$$\begin{cases} v_1 + a_{11}v_1 + a_{12}v_2 = 0 \\ a_{21}v_1 + v_2 + a_{22}v_2 = 0 \end{cases}$$

The next key step is to look at the norm of $\vec{v}$, which is

$$\|\vec{v}\|_2^2 = (a_{11}v_1 + a_{12}v_2)^2 + (a_{21}v_1 + a_{22}v_2)^2 \leq (a_{11}^2 + a_{12}^2 + a_{21}^2 + a_{22}^2)(v_{11}^2 + v_{22}^2) < \frac{1}{1000}\|\vec{v}\|_2^2$$

The first inequality is true because

$$(ax + by)^2 \leq (a^2 + b^2)(x^2 + y^2)$$

Note that this is actually Cauchy schwarz. ∎

1.2.14 Problem 14

12: Eigenvalue and Optimization Exercise

Let $A \in \mathbb{R}^{n \times n}$ real symmetric matrix and let $\lambda_1 \geq \cdots \geq \lambda_n$ be its eigenvalues in descending order. Show that

$$\lambda_k \leq \max_{U:\dim(U)=k} \min_{x \in U:\|x\|_2=1} \langle Ax, x \rangle$$

The maximum above is over all k-dimensional subspaces U of $\mathbb{R}^n$. Hint: form an orthonormal basis of eigenvectors to make U.

Solution. Following the solution, let $\{\vec{v}_i\}$ be an orthonormal basis of eigenvectors of U, corresponding to the eigenvalues λ_i. The existence of such basis is by the Spectral Theorem. Let U be the direct sum of the first K eigenspaces. For any $x \in U$, $x = \sum_{i=1}^k a_i v_i$. Let $\|x\|_2 = 1$. Then

$$\langle Ax, x \rangle = \langle \sum_{i=1}^k a_i \lambda_i v_i, \sum_{i=1}^k a_i v_i \rangle = \sum_{i=1}^k \lambda_i a_i^2 \geq \sum_{i=1}^k a_i^2 \lambda_k = \lambda_k$$

∎

1.2.15 Problem 15

13: Projection Exercise

For a vector $\vec{v} \in \mathbb{R}^n \backslash \{0\}$, define the map $F : \mathbb{R}^n \to \mathbb{R}^n$ by

$$F(x) = \operatorname{argmin}_{z \in \mathrm{Span}(\vec{v})} \|z - x\|_2.$$

Compute F explicitly in terms of $\vec{v}$. Is $F : \mathbb{R}^n \to \mathbb{R}^n$ a linear transformation?

Solution.

$$F(x) = \operatorname{argmin}_a \|a\vec{v} - \vec{x}\|_2$$

We want to minimize the distance between $a\vec{v}$ and $\vec{x}$. We should just project x onto the Span of $\vec{v}$ and F should return

$$F(x) = \frac{v^T x}{\|v\|_2 \|x\|_2} \|x\|_2 \frac{\vec{v}}{\|v\|_2} = \frac{v^T x}{\|v\|_2^2} v$$

This is apparently a linear transformation because projection is linear.
Some words on simplification:

$$v^T x v = v(v^T x) = (vv^T)x$$

since scalar-vector multiplication is commutative. ∎

1.2.16 Problem 16

14: (Straightforward)Square Root of a Symmetric Matrix

Suppose $A \in \mathbb{R}^{n\times n}$ and $A = A^T$ with all eigenvalues of A being positive. Show there exists a matrix B such that $B^2 = A$.

Solution. $A = A^T$ means that A is symmetric, then by the spectral theorem, A can be diagonalized into

$$Q\Lambda Q^T$$

where the digonal of Λ consists of the eigenvalues of A, which are all posiitive and thus we can take square roots of them. Therefore, let $\sqrt{\Lambda}$ denote the diagonal matrix with the square root of the eigenvalues of A on the diagonal. Since the square of a diagonal matrix can be obtained by squaring the diagonal entries, we see that $\sqrt{A}^2 = A$. We can define the matrix B as $Q\sqrt{\Lambda}Q^T$ and

$$B^2 = Q\sqrt{\Lambda}Q^T Q\sqrt{\Lambda}Q^T = Q\Lambda Q^T = A$$

■

1.2.17 Problem 17

15: $|\det|$ is equal to product of singular values

Let $A \in \mathbb{R}^{n\times n}$ and let $\{\sigma_i\}_{i=1}^n$ be the singular values of A. Show that $|\det(A)| = \prod_{i=1}^n \sigma_i$

Solution. Don't work with characteristic polynomial for this one. We should definitely consider the SVD of A which is of the form

$$A = U\Sigma V^T$$

where U and V^T are orthonormal matrices. Then

$$\det A = \det\left(U\Sigma V^T\right) = \det U \det \Sigma \det V^T = \pm \det(\Sigma)$$

Note that the determinant of orthonormal matrices is either $+1$ or -1, since $\det(Q)\det\left(Q^T\right) = \det(I)$ and transposition does not change the determinant. ■

1.2.18 Problem 18

16: Low-Rank Matrix Approximation

Let $A \in \mathbb{R}^{m\times n}$ and for a positive integer $p < rank(A)$, define $A_p = \sum_{i=1}^p \sigma_i u_i v_i^T$ where σ_i is the ith largest singular value of A, and u_i, v_i are respective left and right singular vectors, i.e. the SVD is $A = U\Sigma V^T$. Then prove

$$\sup_{x:\|x\|_2=1} \|(A - A_p)x\|_2 = \sigma_{p+1}$$

Solution.

$$\begin{aligned}
A - A_p &= \sum_{i=1}^{n} \sigma_i u_i v_i^T - \sum_{i=1}^{p} \sigma_i u_i v_i^T \\
&= \sum_{i=p+1}^{n} \sigma_i u_i v_i^T \\
\|(A - A_p)\vec{x}\|_2^2 &= \left(\sum_{i=p+1}^{n} \sigma_i u_i v_i^T \vec{x} \right)^T \left(\sum_{j=p+1}^{n} \sigma_j u_j v_j^T \vec{x} \right) \\
&= \left(\sum_{i=p+1}^{n} \sigma_i x^T v_i u_i^T \right) \left(\sum_{j=p+1}^{n} \sigma_j u_j v_j^T x \right) \\
&= \sum_{i=p+1}^{n} \sigma_i^2 x^T v_i u_i^T u_i v_i^T \vec{x} \quad \text{by orthonormality of } \vec{u}, \vec{v} \\
&= \sum_{i=p+1}^{n} \sigma_i^2 x^T v_i v_i^T x \\
&= \sum_{i=p+1}^{n} \sigma_i^2 (v_i^T x)^2
\end{aligned}$$

Note that this is a convex combination of σ_i^2. To maximize it, we can let $\|v_{p+1}^T x_{p+1}\| = 1$ Alternatively, we can argue that $U^T(A - A_p)V = diag(0, \ldots, 0, \sigma_{p+1}, \ldots)$. By orthgonal invariance of the 2-norm, we have

$$\|(A - A_p)x\|_2 = \|U^T(A - A_p)Vx\|_2$$

which is equal to the biggest singular value of $U^T(A - A_p)V$, i.e. σ_{p+1}. ■

1.2.19 Problem 19

17: Eigenvalues of Symmetric Idempotent Matrix

Suppose $P \in \mathbb{R}^{n \times n}$ symmetric matrix that satisfies $P^2 = P$, a so-called idempotent matrix. Find all the eigenvalues of P with their algebraic multiplicities in terms of P.

Solution. Since P is symmetric, we can decompose it into $Q\Lambda Q^T$ by the spectral theorem. Then we know that

$$P^2 = Q\Lambda Q^T Q\Lambda Q^T = \Lambda^2 = P$$

Note that the eigenvalues of P are either 1 or 0.
We can also consider an eigenpair $\lambda, \vec{v}$ of P, then

$$P\vec{v} = P^2\vec{v} = \lambda\vec{v} = \lambda^2\vec{v}.$$

The eigenvalue 1 of P has multiplicity of $rank(\Lambda) = rank(P)$ since multiplication by an invertible matrix does not change the rank. Also, we then have that the multiplicity of the eigenvlue $\lambda = 0$ is $n - rank(P)$. ■

1.2.20 Problem 20

18: Singular Covariance Matrix means Linearly Dependent a.s.

Suppose that Σ is the covariance matrix of k zero-mean random variables $X_1, \ldots, X_k$, i.e. if $X = (X_1, \ldots, X_k)$, then $\Sigma := \mathbb{E}[XX^T]$. Prove that if Σ is singular, then $X_1, \ldots, X_k$ are linearly dependent almost everywhere.

Solution. Σ is a symemtric positive definite matrix. If Σ is singular, then it has zero eigenvalue. Let's $\vec{v}$ be the corresponding eigenvector. Since $\Sigma\vec{v} = \vec{0}$, $v^T\Sigma v = 0$. Then

$$
\begin{aligned}
&v^T\Sigma v = 0 \implies v^T\mathbb{E}[XX^T]v = 0 \\
&\mathbb{E}[v^TXX^Tv] = 0 \\
&= \mathbb{E}((X^Tv)^2) = 0
\end{aligned}
$$

$(X^Tq)^2$ is a non-negative random variable with mean zero, it must be zero almost surely for its expectation to be zero.
Another arguemnt we can use is that a random variable has variance zero iff it's almost surely constant. ∎

Random Variables and Transformations

2.1 Fundamentals

2.1.1 PDF Transformation Law for Monotone Transformations

Theorem 2.1.3

Let X have cdf $F_X(x)$, let $Y = g(X)$,

- If g is an increasing function, then $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.
- If g is a decreasing function and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$

Uniform-Exponential Relationship Part 1

Suppose $X \sim f_X(x) = 1$ for $0 < x < 1$. Let $Y = -\log(X)$.

$$\frac{\mathrm{d}}{\mathrm{d}x} g(x) = -\frac{1}{x}$$

which is decreasing on $[0, 1]$. $-\log(x)$ ranges from 0 to ∞ on this interval. Note that

$$g^{-1}(y) = e^{-y}.$$

Therefore,

$$F_Y(y) = 1 - F_X(e^{-y}) = 1 - e^{-y}$$

This is the exponential distribution.

Theorem 2.1.5

Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ and that g^{-1} has a continuous derivative on $\mathcal{Y}$. Then the pdf of Y is given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)| & \text{if } y = g(x) \quad \text{for some } x \in \mathcal{X}\} \\ 0 & \text{otherwise} \end{cases}$$

Square Transformation

Suppose X is a continuous random variable. For $y > 0$, the cdf of $Y = X^2$ is

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}).$$

By continuity, this is equal to

$$F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

whence differentiating gives us the pdf:

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y})$$

Normal-chi squared relationship

Let X have the standard normal distribution,

General pdf Transformation Law

Let X have pdf $f_X(x)$ and let $Y = g(X)$. Suppose there exists a partition $\{A_i\}$ of $\mathcal{X}$ such that $\mathbb{P}(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Furthermore, suppose there exist functions $\{g_i\}$ on A_i such that

1. $g(x) = g_i(x)$ on A_i
2. $g_i(x)$ is monotone on A_i.
3. The set $\mathcal{Y} = \{y : y = g_i(x) \quad \text{for some } x \in A_i\}$ is the same for each $i = 1, \ldots, k$.
4. $g_i^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$

Then the pdf of Y is given by

$$f_Y(y) = \sum_{i=1}^{k} f_X(g_i^{-1}(y)|\frac{\mathrm{d}}{\mathrm{d}y} g_i^{-1}(y))$$

2.1.2 Expected Values

Uniform-Exponential Relationship-II

Let X have a uniform (0,1) distribution. Define a new random variable $g(X) = -\log(X)$. Then

$$\mathbb{E}g(X) = \mathbb{E}(-\log X) = \int_0^1 -\log x \mathrm{d}x = x - x\log x|_0^1 = 1$$

But we also saw that $Y = -\log X$ has cdf $1 - e^{-y}$

Cauchy Distribution does not have finite expected value

Let X be a Cauchy random variable, i.e.

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad -\infty < x < \infty$$

- $$\int_0^\infty \frac{1}{1+x^2}\mathrm{d}x = \arctan(\infty) - \arctan(0) = \frac{\pi}{2}$$
  so it's straighforward to check that this is a probability distribution.
- However, the $\mathbb{E}|X| = \infty$.
  $$\begin{aligned}\mathbb{E}|X| &= \int_{-\infty}^\infty \frac{|x|}{\pi}\frac{1}{1+x^2}\mathrm{d}x \\ &= \frac{2}{\pi}\int_0^\infty \frac{x}{1+x^2}\mathrm{d}x\end{aligned}$$
  For any positive number M,
  $$\int_0^M \frac{x}{1+x^2}\mathrm{d}x = \left.\frac{\log(1+x^2)}{2}\right|_0^M = \frac{\log(1+M^2)}{2}$$
  Then
  $$\mathbb{E}|X| = \lim_{M\to\infty}\frac{1}{\pi}\log\big(1+M^2\big) = \infty.$$

2.1.3 Moment Generating Function

Moment Generating Function

The moment generating function of the random variable X is

$$M_X(t) = \mathbb{E}\left[e^{Xt}\right]$$

provided that the expectation exists for t in some neighborhood of 0.

MGF generates the moments

$$\frac{\mathrm{d}^n}{\mathrm{d}t^n} M_X(t)|_{t=0} = \mathbb{E}[X^n]$$

Gamma Moment Generating Function

Gamma Moment Generating Function

Consider the gamma pdf

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad 0 < x < \infty, \alpha > 0, \beta > 0$$

where $\Gamma(\alpha)$ denotes the gamma function evaluated at α. The mgf is then

$$M_X(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty e^{tx} x^{\alpha-1} e^{-x/\beta} \mathrm{d}x = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x/(\frac{\beta}{1-\beta t})} \mathrm{d}x$$

Note that the integrand is the kernel (i.e. the part disregarding the normalizing constant) of another gamma pdf. Therefore, the integral, assuming it exists, evaluates to $\Gamma(\alpha)(\frac{\beta}{1-\beta t})^\alpha$. Then

$$M_X(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \Gamma(\alpha) \left(\frac{\beta}{1-\beta t}\right)^\alpha = \left(\frac{1}{1-\beta t}\right)^\alpha,$$

if $t < 1/\beta$. Otherwise, the integral diverges so to the mgf doesn't exist. As an application of the mgf, the mean of the gamma distribution can be found by

$$\mathbb{E}[X] = \frac{\mathrm{d}}{\mathrm{d}t} M_X(t)\bigg|_{t=0} = -\alpha(1-\beta t)^{-\alpha-1} \cdot (-\beta) = \alpha\beta.$$

2.2 Core Competency Exam Questions

2.2.1 Problem 3

3: 2018 Summer Practice #9

Suppose X_1, X_2 are i.i.d. random variables from a distribution F with mean 0 and variance 1.

- If $F = N(0,1)$, show that
$$\frac{X_1 + X_2}{\sqrt{2}} \stackrel{d}{=} X_1$$
- If
$$\frac{X_1 + X_2}{\sqrt{2}} \stackrel{d}{=} X_1,$$
show that $F = N(0,1)$.

Solution. For part a, we know that X_1, X_2 are independent standard normal variables. Therefore, their sum follows that $N(0+0, 1+1)$ distribution. By scaling the sum by $\sqrt{2}$, we get an $N(0,1)$ random variable.
For the other direction, we can use the fact that MGF characterizes distributions.

$$\frac{X_1 + X_2}{\sqrt{2}} \stackrel{d}{=} X_1,$$

means that

$$M_{\frac{X_1+X_2}{\sqrt{2}}}(t) = M_{X_1}(t).$$

By independence, we know that

$$M_{\frac{X_1+X_2}{\sqrt{2}}} = M_{X_1}(\frac{1}{\sqrt{2}}t)M_{X_2}(\frac{1}{\sqrt{2}}t) = M_{X_1}(\sqrt{2}t)^2 = M_{X_1}$$

Next, the trick is to take log of the above functional equation!! ■

2.2.2 Problem 4

4: 2018 Summer Practice #14

Suppose $f : [0, \infty) \to \mathbb{R}$ is a function such that $f(x+y) = f(x)f(y)$.

- Show that $f9x) \geq 0$ for all real $x \geq 0$.
- Show that $f(0) \in \{0, 1\}$.
- Show that for any nonnegative rational number r one has $f(r) = c^r$ where $c \in [0, \infty)$
- ...

Solution. An easy problem in my opinion. ■

2.2.3 Problem 5

5: 2018 Summer #15

Suppose X is a random variable taking values in $[0, 1]$.

- Show that $\text{Var}(X) \leq \frac{1}{4}$.
- Find a random variable for which equality holds.

Solution. $1/4$ is a special number and is related to $\frac{1}{2}$. First note that $\mathbb{E}X$ is the best constant estimator in the MSE sense, so

$$\mathbb{E}[(X - \mathbb{E}X)^2] \leq \mathbb{E}[(X - \frac{1}{2})^2] = \mathbb{E}[X^2] - \mathbb{E}[X] + \frac{1}{4} \leq \frac{1}{4}$$

since $X^2 \leq X$. The Bernoulli rv that takes 0 with probability 0.5 and 1 with probability 0.5 achieves equality for this variance bound. ■

2.2.4 Problem 7

7: 2018 September # 4

- Let X be a random variable and $a \in \mathbb{R}$. Show that (using Markov's inequality or otherwise):
$$\mathbb{P}[X \geq a] \leq \inf_{s \geq 0} e^{-sa}\mathbb{E}[e^{sX}].$$
- Let N be a Poisson random variable with parameter $\lambda > 0$, i.e.
$$\mathbb{P}[N = n] = e^{-\lambda}\frac{\lambda^n}{n}, \quad n \geq 0$$
Show that $\mathbb{E}[e^{sN} = e^{\lambda(e^s - 1)}]$ for all $s \in \mathbb{R}$.
- Let N be as in ii and let $m \geq \lambda$ be an integer. Use i and ii to show that
$$P(N \geq m) \leq (\frac{\lambda}{m})^m e^{m-\lambda}$$

Solution.

- By the Markov inequality,
$$P(X \geq a) = P(e^{sX} \leq e^{sa}) \leq e^{-sa}\mathbb{E}[e^{sX}]$$
for any positive s. Therefore,
$$\mathbb{P}(X \geq a) \leq e^{-sa}\mathbb{E}[e^{sX}]$$

- $$\begin{aligned}\mathbb{E}[e^{sN}] &= \sum_{n=0}^{\infty} e^{-sn}\, e^{-\lambda} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} e^{-sn} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(e^{-s}\lambda)^n}{n!} \\ &= e^{-\lambda} e^{e^{-s}\lambda} \\ &= e^{\lambda(e^s-1)}\end{aligned}$$

- $$P(N \geq m) \leq \inf_{s \leq 0} e^{-sm} e^{\lambda(e^s-1)}$$
  Let's find the minimizing s. Letting the derivative be 0, we get
  $$-me^{-sm}e^{\lambda(e^s-1)} + e^{-sm}\lambda\, e^s e^{\lambda(e^s-1)} = 0$$
  This simplifies to
  $$\lambda e^s - m = 0 \iff e^{-s} = \frac{\lambda}{m}$$
  Therefore $\inf_{s\geq 0} e^{-sa}\mathbb{E}[e^{sX}] = \left(\frac{\lambda}{m}\right)^m e^{\lambda(\frac{m}{\lambda}-1)} = \left(\frac{\lambda}{m}\right)^m e^{m-\lambda}$.

■

2.2.5 Problem 8

2019 May # 4

We model the lifetime of a device as a random variable $T \geq 0$ with cdf $F(t)$ and density $f(t)$. Suppose that $f(t)$ is continuous for $t \geq 0$ and define the intensity of failure as
$$\lambda(t) = \lim_{h \downarrow 0} \frac{P(t \leq T \leq t+h | T \geq t)}{h}, \quad \text{for } t \geq 0.$$

- Express $\lambda(t)$ through $f(t)$ and $F(t)$.
- Compute the intensity of failure when $T \sim \exp(\alpha), \alpha > 0$.
- Show that $F(t) = 1 - \exp\left(-\int_0^t \lambda(s)\mathrm{d}s\right)$ for $t \geq 0$.
- Determine $F(t)$ and $f(t)$ in the case that $\lambda(t) = \alpha t^\gamma$ for some $\alpha > 0$ and $\gamma > 0$.

Solution. For part a, let's just rewrite the expression from the definition using f and F.

$$P(t \leq T \leq t+h|T \geq t) = \frac{F(t+h) - F(t)}{1 - F(t)}$$
$$\lim_{h\downarrow 0} \frac{F(t+h) - F(t)}{h} = f(t)$$

Therefore

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

For part b,

$$\begin{aligned}\lambda(t) &= \frac{f(t)}{1 - F(t)} \\ &= \frac{\alpha e^{-\alpha t}}{e^{-\alpha t}} \mathbf{1}_{t \geq 0} \\ &= \alpha \mathbf{1}_{t \geq 0}\end{aligned}$$

For part c, first of all

$$\int_0^t \lambda(s) \mathrm{d}s = \int_0^t \frac{f(s)}{1 - F(s)} \mathrm{d}s.$$

Let $y = F(t)$, then $\frac{\mathrm{d}y}{\mathrm{d}t} = f(t)$ and thus $\mathrm{d}y = f(t)\mathrm{d}t$.
Now the above integral becomes

$$\int_{F(0)}^{F(t)} \frac{1}{1-y} \mathrm{d}y = \ln(1-y)|_{F(t)}^{F(0)} = -\ln(1 - F(t))$$

Now

$$1 - \exp\{\ln(1 - F(t))\} = 1 - 1 + F(t) = F(t).$$

For part d, we should go from part c and compute $F(t)$.

$$F(t) = 1 - \exp\{-\int_0^t \lambda(s) \mathrm{d}s\}$$

$$\int_0^t \alpha s^\gamma \mathrm{d}s = \frac{\alpha}{\gamma + 1} t^{\gamma+1}$$

Then

$$F(t) = 1 - \exp\{-\frac{\alpha}{\gamma + 1} t^{\gamma+1}\}$$

Then by differentiating $F(t)$, we get

$$f(t) = -\exp\{-\frac{\alpha}{\gamma + 1} t^{\gamma+1}\}(-\alpha t^\gamma)$$

■

2.2.6 Problem 9

2019 May #5

You are working as a TA in the help room for a duration t. The number of students arriving during that period is Poisson distributed with parameter $t\lambda$. For each student, the time T to answer their questions is exponentially distributed with parameter α and this time is independent of all other students. Prove that the distribution of the number X of students that arrive while you are busy with one fixed (randomly chosen) student is geometric with some parameter p and determine

p in terms of α and λ.
Hint: The formula
$$\int_0^\infty s^k e^{-s} \mathrm{d}s = s!$$
for $k = 0, 1, 2, \ldots$ can be used without proof.

Solution. This problem is all about computing integral.

- First of all, we know
$$\mathbb{P}(X = k | T = t) \sim \text{Poisson}(\lambda t)$$
- Now
$$P(X = n) = \int_0^\infty P(X = n | T = t) f(t) \mathrm{d}t$$
-
$$\begin{aligned} P(X = n) &= \int_0^\infty \frac{e^{-t\lambda}(t\lambda)^n}{n!} \alpha e^{-\alpha t} \mathrm{d}t \\ &= \frac{\alpha}{n!} \frac{1}{\alpha + \lambda} \int_0^\infty e^{-s} \left(\frac{\lambda}{\alpha + \lambda} s \right)^n \mathrm{d}s \\ &= \frac{\alpha \lambda^n}{(\alpha + \lambda)^{n+1}} \end{aligned}$$

  This is a geometric random variable with parameter $p = \frac{\lambda}{\alpha + \lambda}$

∎

2.2.7 Problem 10

2019 May # 7

Suppose you have n red balls and one blue ball. We will do two experiments. In the first experiment, you first drop n red balls uniformly on the interval $[0, 1]$, independent of each other. Having done this, now you drop the blue ball uniformly in the interval, independent of all previous ball drops. Let X denote the number of red balls to the left of the blue ball. Find $P(X = k)$, for $k = 0, ..., n$.
In the second, experiment, you drop all the $(n + 1)$ balls uniformly on $[0, 1]$, independent of each other. Let Y denote the number of red balls to the left of the blue ball as before. Find $\mathbb{P}(Y = k)$.

Solution. Let $R_1, \ldots, R_n$ be the positions of the red balls, which are i.i.d. Uniform(0,1). Let B be the position of the blue ball. Let
$$X = \mathbf{1}_{B \le U_1} + \cdots + \mathbf{1}_{B \le U_n}.$$
Given $B = b$, X is a binomial(b, n) random variables. Then
$$P(X = k) = \int_0^1 \binom{n}{k} b^k (1 - b)^{n-k} \mathrm{d}b = \binom{n}{k} \beta(k + 1, n - k + 1)$$

This simplifies to

$$\frac{n!}{k!(n-k)!}\frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}.$$

This is my original and somewhat canonical idea, which differs from the official solution:

$$P(X=k) = \mathbb{E}_{U_1,\dots,U_n}[P(X=k|U_1,\dots,U_n)] = \mathbb{E}[U_{(k+1)} - U_{(k)}] = \frac{k+1}{n+1} - \frac{k}{n+1} = \frac{1}{n+1}.$$

For this to work, we need to know that the order statistics $U_{(k)} \sim \text{Beta}(k, n-k+1)$. ■

2.2.8 Problem 11

S

uppose that X is a non-negative random variable. Show that

2.3 Supplementary Questions

2.3.1 Median Minimizes the Absolute Error

Median Minimizes the Absolute/L^1 Error

Let X be a random variable. Show that the median of X is the constant a that minimizes $\mathbb{E}|X-a|$.

Solution. 1. To make progress, we need to write something down. Let f be the probability density function corresponding to X.

$$E|X-a| = \int_{-\infty}^{\infty} |x-a|f(x)\mathrm{d}x$$

2. By linearity,

$$E|X-a| = \int_{-\infty}^{\infty} |x-a|f(x)\mathrm{d}x = \int_{-\infty}^{a} (a-x)f(x)\mathrm{d}x + \int_{a}^{\infty} (x-a)f(x)\mathrm{d}x$$

3. Differentiating with respect to a and set the derivative to zero, we get that

$$F(a) + af(a) - af(a) - af(a) - (\frac{\mathrm{d}}{\mathrm{d}a}aF(\infty) - aF(a)) = 0$$

4. Simplify to get

$$F(a) - af(a) - 1 + af(a) + F(a) = 0 \implies F(a) = \frac{1}{2}$$

5. By definition, the minimizer a is the median of X.

■

2.3.2 A Tight Bound of Variance of Bounded Random Variables

A Tight Bound of Variance of Bounded Random Variables

Let X be a random variable taking values in the interval $[0, 1]$.

- Show that the $\text{Var}X \le \frac{1}{4}$.
- Show that this bound is tight by finding a X that achieves this bound.

Solution. I provide two approaches to solve the first part.

1. The first approach starts by noting the fact that $X^2 \le X$ on $[0, 1]$. Then we have that
$$\mathbb{E}[X^2] \le \mathbb{E}[X].$$
2. Then it's natural for us to consider the decomposition of variance
$$\text{Var}X = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \le \mathbb{E}[X] - \mathbb{E}[X^2].$$
3. Applying calculus to maximize $\mu - \mu^2$ on $[0, 1]$, we get
$$\frac{\mathrm{d}}{\mathrm{d}\mu}(\mu - \mu^2) = 1 - 2\mu = 0 \quad \implies \mu = \frac{1}{2}. \quad \frac{\mathrm{d}^2}{\mathrm{d}\mu^2}(\mu - \mu^2) = -2 < 0.$$
4. Finally, $\text{Var}X \le \frac{1}{2} - \frac{1}{2}^2 = \frac{1}{4}$.

1. The second approach uses the fact that the expectation is the single constant-predictor that minimizes the mean square error. In particular, it's at least as good as the constant-predictor $\frac{1}{2}$:
$$\mathbb{E}\left[(X - \mu)^2\right] \le \mathbb{E}\left[(X - \frac{1}{2})^2\right].$$
2. The maximum distance between X and $\frac{1}{2}$ is $\frac{1}{2}$ since $X \in [0, 1]$. Therefore,
$$\mathbb{E}\left[(X - \frac{1}{2})^2\right] \le \frac{1}{4}.$$

For the second part of the problem, it was immediate for me to think of the random variable
$$X = \begin{cases} 0 & \text{w.p. } \frac{1}{2} \\ 1 & \text{w.p. } 1 \end{cases}.$$

■

Multiple Random Variables Exercises

3.1 Fundamentals

3.1.1 Joint and Marginal Distributions

Joint CDF

3.2 Core Competency Exam Questions

Summer 2018 Practice # 12

Suppose that $U_1, U_2 \overset{iid}{\sim} U(0,1)$. Let $V_1 = \max(U_1, U_2)$ and $V_2 = \min(U_1, U_2)$.

- Find
$$\Pr[V_1 \geq x, V_2 \leq y]$$
where $x, y \in [0,1]$.
- Hence or otherwise find the joint density for (V_1, V_2)
- Hence or otherwise compute $\mathbb{E}(V_1^2 + V_2^2)$

Solution.

- Let's use de Morgan's law to decompose the event:

$$\begin{aligned}
P(V_1 \geq x, V_2 \leq y) &= P(U_1 \geq x, U_2 \leq y \quad || \quad U_1 \leq y, U_2 \geq x) \\
&= P(U_1 \geq x, U_2 \leq y) + P(U_1 \leq y, U_2 \geq x) - P(x \leq U_1 \leq y, \quad x \leq U_2 \leq y) \\
&= \begin{cases} 2y - y^2 - x^2 & \text{if } x \leq y \\ 2y - 2xy & \text{otherwise} \end{cases}
\end{aligned}$$

To find the joint density, we take derivative of

$$
\begin{aligned}
f(x,y) =& \frac{\partial^2 P(V_1 < x, V_2 < y)}{\partial x \partial y} = \frac{\partial^2}{\partial x \partial y} P(V_1 < x) - P(V_1 < x, V_2 > y) \\
=& \frac{\partial^2}{\partial x \partial y} \begin{cases} x^2 - (x-y)^2 & \text{if } y \in [0,1] \text{ and } x \geq y \\ x^2 & \text{otherwise} \end{cases} \\
x^2 =& \begin{cases} 2 & \text{if } y \in [0,1] \text{ and } x \geq y \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

■

Multivariate Gaussians

4.1 Fundamentals

Random Samples

5.1 Fundamentals

Stochastic Convergence

6.1 Fundamentals

Point Estimation

7.1 Fundamentals

Hypothesis Testing

8.1 Fundamentals

Asymptotics of Tests and Estimators

9.1 Fundamentals

9.1.1 Consistency of Estimators

Consistent Estimators

A sequence of estimators $W_n = W_n(X_1, \ldots, X_n)$ is a consistent sequence of estimators of parameter θ if $W_n(X_1, \ldots, X_n) \xrightarrow{p} \theta$. That is for every $\epsilon > 0$,

$$P(|W_n(X_1, \ldots, X_n) - \theta| < \epsilon) \to 1$$

as $n \to \infty$.