Week 1: Basics of Data Analysis

☝️ Before you dive into the theory, a word of caution. This page tries to expand on the theory required to understand the underlying mathematics for this project, which, without sufficient practice on your end, wouldn't turn out to be an effective resource at all. Try to solve the problems that are mentioned at the end of each topic. Once again, in case of doubts, just let us know :)

Foundations of Probability Theory

Basic Probability Concepts and Definitions

Sample Space and Events

- **Sample Space (Ω)**: The set of all possible outcomes of an experiment
- **Event (A)**: A subset of the sample space
- **Elementary Event**: A single outcome from the sample space

Example: Rolling a fair six-sided die

- Sample Space: Ω = {1, 2, 3, 4, 5, 6}

- Event A (rolling an even number): A = {2, 4, 6}
- Elementary Event: Rolling a 3

Probability Axioms (Kolmogorov Axioms)

1. **Non-negativity**: P(A) ≥ 0 for all events A
2. **Normalization**: P(Ω) = 1
3. **Countable Additivity**: For mutually exclusive events A_1, A_2, ...:
   $P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ...$

Some Properties

- **Complement Rule**: $P(A^c) = 1 - P(A)$
- **Addition Rule**: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Empty Set**: $P(\emptyset) = 0$

https://youtu.be/1uW3qMFA9Ho

Conditional Probability and Independence

Conditional Probability

P(A|B) = P(A ∩ B) / P(B), provided P(B) > 0

The probability of event A occurring given that event B has occurred is the probability

of A and B both happening divided by the probability that event B happens which makes sense

intuitionally as well .

Law of Total Probability

If $B_1, B_2, \ldots, B_n$ form a partition of Ω, then:
$P(A) = \Sigma P(A|B_i) \times P(B_i)$

Bayes' Theorem for conditional probability

$P(A|B) = [P(B|A) \times P(A)]/P(B)$

Independence

Events A and B are independent if:
$P(A \cap B) = P(A) \times P(B)$

Equivalently: $P(A|B) = P(A)$ and $P(B|A) = P(B)$

https://youtu.be/B5y6fy5iUtg

Random Variables

Definition

A random variable X is a function from the sample space Ω to the real numbers ℝ:
X: Ω → ℝ

Types of Random Variables

1. **Discrete Random Variables**: Countable range
2. **Continuous Random Variables**: Uncountable range

Probability Mass Function (PMF)

For discrete random variable X:
$p_x(x) = P(X = x)$

Properties:

- $p_x(x) \geq 0$ for all x
- $\Sigma\ p_x(x) = 1$ (sum over all possible values)

Probability Density Function (PDF)

For continuous random variable X:

- $P(a \leq X \leq b) = \int_a^b f_x(x)dx$

Properties:

- $f_x(x) \geq 0 \forall x$
- $\int_{-\infty}^{\infty} f_x(x)\, dx = 1$

https://youtu.be/k2BB0p8byGA?si=zVLS0WxcwNHvK1D1

Some Common Probability Distributions

Discrete Distributions

Bernoulli Distribution

- Parameter: p (success probability)
- PMF: $P(X = 1) = p, P(X = 0) = 1 - p$
- Mean: μ = p
- Example is a single coin toss with Heads being probability p and tails 1-p

Binomial Distribution

- Parameters: n (trials), p (success probability)

$$PMF : P(X = k) = C(n, k) \times p^k \times (1 - p)^{n-k}$$

- Example is multiple coin tosses where each heads has probability p and tails 1-p

Poisson Distribution

- Parameter: λ (rate)
- PMF: $P(X = k) = (\lambda^k \times e^-\lambda)/k!$

https://youtu.be/3MOahpLxj6A?si=JYXmbhuNjDW_amtO	https://youtu.be/-qCEoqpwjf4?si=e5xj77eM1wrYMeIJ

Continuous Distributions

Distribution	PDF / PMF	Variance
Uniform(a, b)	$f(x) = \frac{1}{b-a} \ldots x \in [a, b]$	$\frac{(b-a)^2}{12}$
Normal (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	σ^2
Gamma (α, β)	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$\frac{\alpha}{\beta^2}$
Chi-Square (χ^2_n)	Sum of n squared standard normals	$2n$

https://youtu.be/mHfn_7ym6to?si=tW63IMzFhO_r7ISu

Practice Problems

Try to solve problems from the following chapters from Sheldon Ross until you reach a sense of familiarity with these topics.

- Ch 1: Combinatorics (counting problems)
- Ch 2: Basic probability axioms
- Ch 3: Conditional probability and independence

Expectation, Variance, and Joint Distributions

Expectation and Moments

Expected Value (Mean)

Discrete Random Variable: $E[X] = \sum_x x \cdot P(X = x)$
Continuous Random Variable: $E[X] = \int_{-\infty}^{\infty} x f(x) dx$

The expected value is ***the mean of the possible values a random variable can take***, weighted by the probability of those outcomes.

Properties of Expectation

1. **Linearity**: $E[aX + bY] = aE[X] + bE[Y]$
2. **Constant**: $E[c] = c$ for any constant c
3. **Independence**: If X and Y are independent, $E[XY] = E[X]E[Y]$

Law of the Unconscious Statistician (LOTUS)

For function g(X):
Discrete: $E[g(X)] = \Sigma_x g(x) \cdot P(X = x)$
Continuous: $E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$ where f is pdf

We are simply just extending the standard definition of expectation to functions

Variance and Standard Deviation

Variance: $Var(X) = E[(X - \mu)^2] = E[X^2] - (E[X])^2$

The proof for this is very simple to do, try to do it by self.

Standard Deviation: $\sigma(X) = \sqrt{Var(X)}$

Properties of Variance

1. **Constant**: $Var(aX + b) = a^2 Var(X)$
2. **Independence**: If X and Y are independent, $Var(X + Y) = Var(X) + Var(Y)$
3. **General case**: $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

Moments:

Moments in statistics are **quantitative measures (or a set of statistical parameters) that describe the specific characteristics of a probability distribution**. In simple terms, the moment measures how spread out or concentrated the number in a dataset is around the central value, such as the mean.

Higher Moments

k-th moment: $E[X^k]$
k-th central moment: $E[(X-\mu)^k]$

The following are terms you should remember:

Skewness: γ_1 = $E[(X-\mu)^3]/\sigma^3$
Kurtosis: γ_2 = $E[(X-\mu)^4]/\sigma^4$

Moment Generating Functions

Definition

The **moment generating function (MGF)** of random variable X is:
$M_X(t) = E[e^{tX}]$ for t in some neighborhood of 0

Properties of MGF

1. **Uniqueness**: MGF uniquely determines the distribution
2. **Moments**: $E[X^n] = M_X^{(n)}(0)$ (n-th derivative at 0)
3. **Independence**: If X and Y are independent, $M_{X+Y}(t) = M_X(t)M_Y(t)$
4. **Linear transformation**: $M_{(aX+b)}(t) = e^{(bt)} \cdot M_X(at)$

Common MGFs

Distribution	MGF
Bernoulli(p)	$1-p+pe^t$
Binomial(n,p)	$(1-p+pe^t)^n$
Poisson(λ)	$exp(\lambda(e^t-1))$

Normal(μ,σ²)	$exp(\mu t + \sigma^2 t^2/2)$
Exponential(λ)	$\lambda/(\lambda - t), t < \lambda$

Characteristic Functions

Definition: $\varphi_X(t) = E[e^{i*t*X}]$ (always exists)
Properties: Similar to MGF but using complex exponentials

Joint Distributions and Dependence

We can also define probability mass and distribution functions for more than 1 random variable:

Joint Probability Mass Function (Discrete)

For discrete random variables X and Y:
$p_{X,Y}(x, y) = P(X = x, Y = y)$

Joint Probability Density Function (Continuous)

For continuous random variables X and Y:
$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(X, Y)(x, y) dx dy$

Marginal Distributions

This really refers to distributions of 1 random variable given no restriction/constraint on other

Discrete: $p_X(x) = \Sigma_y p_{(X,Y)}(x, y)$
Continuous: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$

Conditional Distributions

Discrete: $P(Y = y|X = x) = p_(X, Y)(x, y)/p_X(x)$
Continuous: $f_(Y|X)(y|x) = f_(X, Y)(x, y)/f_X(x)$

Independence of Random Variables

X and Y are said to be independent iff the following condition holds:
Discrete: $p_{(X,Y)}(x, y) = p_X(x) \cdot p_Y(y)$ for all x, y

Continuous: $f_{(X,Y)}(x,y) = f_X(x) \cdot f_Y(y)$ for all x,y

---

Covariance and Correlation

Covariance

$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$ (due to linearity of expectation)

Properties of Covariance

1. $Cov(X,X) = Var(X)$
2. $Cov(X,Y) = Cov(Y,X)$ (symmetric)
3. $Cov(aX+b, cY+d) = ac \cdot Cov(X,Y)$
4. If X and Y are independent, $Cov(X,Y) = 0$ (converse not always true)

Correlation Coefficient

$\rho(X,Y) = Cov(X,Y)/(\sigma_X \cdot \sigma_Y)$ where σ_X and σ_Y refer to the standard deviations of random variables X and Y , NOT the variance keep this in mind.

Properties of Correlation

1. $-1 \leq \rho(X,Y) \leq 1$
2. $|\rho(X,Y)| = 1$ if and only if $Y = aX + b$ for some constants a,b
3. $\rho(X,Y) = 0$ implies X and Y are uncorrelated (but not necessarily independent)

Advanced Distributions

Multivariate Normal Distribution

For random vector $X = (X_1, X_2, ..., X_n)$:
$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)$

where **μ** is the mean vector and Σ is the covariance matrix.

Bivariate Normal

For $X\ N(\mu_1, \sigma_1^2)$ and $Y\ N(\mu_2, \sigma_2^2)$ with correlation ρ:

$f(x, y) = 1/(2\pi\sigma_1\sigma_2\sqrt(1-\rho^2))exp(-Q/2(1-\rho^2))$

where $Q = (x-\mu_1)^2/\sigma_1^2 - 2\rho(x-\mu_1)(y-\mu_2)/(\sigma_1\sigma_2) + (y-\mu_2)^2/\sigma_2^2$

Gamma Distribution

X ~ Gamma(α, β) has PDF:
$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$

Mean: $E[X] = \alpha/\beta$
Variance: $Var(X) = \alpha/\beta^2$
MGF: $M(t) = (\beta/(\beta - t))^\alpha$ for $t < \beta$

Beta Distribution

X ~ Beta(α, β) has PDF:
$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1$ for 0 < x < 1

Mean: $E[X] = \alpha/(\alpha + \beta)$
Variance: $Var(X) = \alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)]$

Chi-Square Distribution

If $Z_1, Z_2, ..., Z_n$ are independent $N(0, 1)$, then $X = Z_1^2 + Z_2^2 + ... + Z_n^2 \sim \chi^2(n)$

Mean: $E[X] = n$
Variance: $Var(X) = 2n$
MGF: $M(t) = (1 - 2t)$ raised to the $(-n/2)$ for $t < 1/2$

Transformations of Random Variables

Function of One Random Variable

If Y = g(X) where g is monotonic:
$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$

Linear Transformation

If $Y = aX + b$:

- $E[Y] = aE[X] + b$
- $Var(Y) = a^2 Var(X)$
- $M_Y(t) = e^(bt) M_X(at)$ where M is moment generating function

Sum of Independent Random Variables

If X and Y are independent:

- $E[X + Y] = E[X] + E[Y]$
- $Var(X + Y) = Var(X) + Var(Y)$
- $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

Special Cases

Sum of Normals: If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent, then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Sum of Poissons: If $X \sim Poisson(\lambda_1)$ and $Y \sim Poisson(\lambda_2)$ are independent, then $X + Y \sim Poisson(\lambda_1 + \lambda_2)$

Practice Problems

Try to solve problems from the following chapters from Sheldon Ross until you reach a sense of familiarity with these topics.

- Ch 4: Random variables and expectation
- Ch 5: Continuous random variables

Probability Inequalities and Convergence

Fundamental Probability Inequalities

Markov's Inequality

Statement: For non-negative random variable X and a > 0:
$P(X \geq a) \leq E[X]/a$

Proof:
$E[X] = \int_0^\infty xf(x)dx \geq \int_a^\infty xf(x)dx \geq a\int_a^\infty f(x)dx = aP(X \geq a)$

Applications:

- Provides upper bound when only mean is known
- Foundation for other inequalities
- Used in algorithm analysis

https://www.youtube.com/watch?v=vjYanZ1nsZg&pp=ygU3cHJvYmFiaWxpdHkgaW5lcXVhbGl0aWVzIGFuZCBjb252ZXJnZW5jZSBhbmQgZGl2ZXJnZW5jZQ%3D%3D

Chebyshev's Inequality

Statement: For any random variable X with finite mean μ and variance σ²:
$P(|X - \mu| \geq k\sigma) \leq 1/k^2 \quad for \quad k > 0$

Equivalent forms:

- $P(|X - \mu| < k\sigma) \geq 1 - 1/k^2$
- $P(|X - \mu| \geq \varepsilon) \leq \sigma^2/\varepsilon^2 \quad for \quad \varepsilon > 0$

Proof: Apply Markov's inequality to $Y = (X - \mu)^2$ with $a = k^2\sigma^2$

Key Insights:

- At least 75% of data within 2 standard deviations
- At least 89% of data within 3 standard deviations
- Works for ANY distribution with finite variance

One-Sided Chebyshev (Cantelli's Inequality)

Statement: For random variable X with mean μ and variance σ²:
$P(X - \mu \geq k\sigma) \leq 1/(1 + k^2) \quad for \quad k > 0$

Advantage: Tighter bound for one-sided deviations

Chernoff Bounds

Statement: For random variable X with MGF $M_X(t)$:
$P(X \geq a) \leq e^{(-ta)} M_X(t)$

Method: Choose t to minimize the bound
$P(X \geq a) \leq inf_{(t>0)} e^{(-ta)} M_X(t)$

Applications: Exponentially decreasing bounds for sums of independent random variables

> 📎 **Just like NT, inequalities like Cauchy-Schwartz and Jensen's also creep into inequalities using expectation of R.Vs. As a fun exercise you can try proving them :)**

Convergence

Convergence describes how a sequence of random variables X_1, X_2, X_3, ... behaves as n approaches infinity. We want to know if and how these random variables approach some target value or distribution.

https://www.youtube.com/watch?v=Ajar_6MAOLw&t=402s&pp=ygU3cHJvYmFiaWxpdHkgaW5lcXVhbGl0aWVzIGFuZCBjb252ZXJnZW5jZSBhbmQgZGl2ZXJnZW5jZQ%3D%3D

Laws of Large Numbers

Weak Law of Large Numbers (WLLN)

Statement: Let X_1, X_2, ... be independent random variables with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2 < \infty$.
Then $\bar{X}_n = (X_1 + ... + X_n)/n \to \mu$

Proof using Chebyshev:

- $E[\bar{X}_n] = \mu$
- $Var(\bar{X}_n) = \sigma^2/n$
- $P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \sigma^2/(n\varepsilon^2) \to 0$ as $n \to \infty$

Strong Law of Large Numbers (SLLN)

Statement: Under same conditions as WLLN:
$\bar{X}_n \to \mu$

Kolmogorov's SLLN: $IfE[X_i] = \mu$ and $\Sigma_i Var(X_i)/i^2 < \infty$, then $\bar{X}_n \to \mu$

Interpretation:

- Sample averages converge to population mean
- Foundation of statistical inference
- Justifies Monte Carlo methods

Applications of LLN

1. **Casino profits**: Long-run house edge realization
2. **Quality control**: Sample proportion → true defect rate
3. **Insurance**: Claim frequency → expected frequency
4. **Polling**: Sample proportion → population proportion

https://youtu.be/3eiio3Tw7UQ?si=cDdWqwfsjrlcGdlR

Practice Problems

Try to solve problems from the following chapters from Rohatgi & Saleh until you reach a sense of familiarity with these topics.

- Ch 3 - Sec 3.4: Moments and Generating functions - Some Moment Inequalities

---

Central Limit Theorem (CLT)

But what is CLT?

The CLT states that when you average many independent random variables, the distribution of that average approaches a normal distribution, regardless of the original distribution shape.

Key insight: The "magic" of CLT is that it works for ANY original distribution - uniform, exponential, discrete, continuous, even weird distributions.

---

https://www.youtube.com/watch?v=YAIJCEDH2uY

There's a really nice 3B1B video on CLT too. Do check it out :)

Standard Form of CLT

Setup: Let $X_1, X_2, ..., X_n$ be independent, identically distributed random variables with:

- Mean: $E[X_i] = \mu$
- Variance: $Var[X_i] = \sigma^2 < \infty$

Sample mean: $\bar{X}_n = (X_1 + X_2 + ... + X_n)/n$

Standardized form: $Z_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$

CLT Statement: As n → ∞, Z_n converges in distribution to N(0,1)

Practical form: For large n, $\bar{X}_n \approx N(\mu, \sigma^2/n)$

Key Components Explained

Why the √n factor?

- $Var[\bar{X}_n] = \sigma^2/n$

- Standard deviation of $\bar{X}_n$ is $\sigma/\sqrt{n}$
- To standardize, we multiply by √n to get unit variance

Why σ²/n for the variance?

- $Var[X_1 + ... + X_n] = n\sigma^2$ (for independent variables)
- $Var[(X_1 + ... + X_n)/n] = n\sigma^2/n^2 = \sigma^2/n$

What "large n" means

- Depends on original distribution
- Symmetric distributions: n ≥ 30 often sufficient
- Skewed distributions: may need n ≥ 100+
- Already normal: any n works

Few Examples

Example 1: Chebyshev Application

A machine produces items with mean weight 100g and standard deviation 5g. What can we say about the probability that a randomly chosen item weighs between 85g and 115g?

Solution:
$\mu = 100, \sigma = 5$, interval is $\mu \pm 3\sigma (k = 3)$

Using Chebyshev: $P(|X - 100| < 15) \geq 1 - 1/3^2 = 1 - 1/9 = 8/9 \approx 89$

If normally distributed: $P(|X - 100| < 15) \approx 99.7$

Key insight: Chebyshev gives conservative bound for any distribution

Example 2: Law of Large Numbers Verification

Simulate coin flips to verify WLLN. $Let X_1, X_2$, ... be Bernoulli(0.5).

Theoretical: $E[X_i] = 0.5, so \bar{X}_n \to^p 0.5$

Simulation approach:

- Generate n coin flips
- Calculate running average $\bar{X}_k$ for k = 1, 2, ..., n
- Plot $\bar{X}_k$ vs k to show convergence to 0.5

Chebyshev bound: $P(|\bar{X}_n - 0.5| \geq \varepsilon) \leq 0.25/(n\varepsilon^2)$

This simulation can be done by simple code

Practice Problems

Try to solve problems from the following chapters from Rohatgi & Saleh until you reach a sense of familiarity with these topics.

- Ch 7 - Sec 7.6: Basic Asymptotes: Large Sample Theory

Multivariate Gaussian Distribution

Representation

The Multivariate Gaussian (or Normal) distribution is a generalisation of the one-dimensional (univariate) normal distribution to higher dimensions.

It is denoted as $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

- $\mathbf{x} \in \mathbb{R}^D$ is a D-dimensional random vector
- $\boldsymbol{\mu} \in \mathbb{R}^D$ is the m**ean vector**
- $\boldsymbol{\Sigma} \in \mathbb{R}^{D \times D}$ is the **covariance matrix**

Probability Density Function (PDF)

The probability density function for a D-dimensional vector $\mathbf{x}$ is:

$p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$ where

1. **Normalization Constant is** $\frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}}$ which ensures the integral of the PDF over the entire space equals 1. $|\boldsymbol{\Sigma}|$ is the determinant of the covariance matrix.
2. **Mahalanobis Distance (The Quadratic Form):** $\Delta^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ This represents the "distance" of point $\mathbf{x}$ from the mean $\boldsymbol{\mu}$, corrected for the

variance and correlations defined by $\mathbf{\Sigma}$

Properties of the Covariance Matrix

The covariance matrix determines the geometric shape of the density. For a valid valid Gaussian distribution, $\mathbf{\Sigma}$ must be:

- **Symmetric:** $\mathbf{\Sigma} = \mathbf{\Sigma}^T$ $(\Sigma_{ij} = \Sigma_{ji})$.
- **Positive Semi-Definite (PSD):** $\mathbf{z}^T \mathbf{\Sigma} \mathbf{z} \geq 0$ for any vector $\mathbf{z}$
    - *Note:* If $\mathbf{\Sigma}$ is strictly Positive Definite (all eigenvalues > 0), it is invertible (non-degenerate)

Linear Transformations (Affine Property)

One of the most useful properties of the Multivariate Gaussian is its closure under linear transformations. If you apply a matrix multiplication and a shift to a Gaussian vector, the result is **still Gaussian**.

Given a random vector $\mathbf{x} \in \mathbb{R}^D$ we have $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ and if we define a transformed variable $\mathbf{y} \in \mathbb{R}^M$ via the linear equation $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ then

- $\mathbf{A} \in \mathbb{R}^{M \times D}$this gives us the transformation matrix (rotation/scaling)
- $\mathbf{b} \in \mathbb{R}^M$is the bias vector (shift)

The Resulting Distribution

The variable $\mathbf{y}$ is distributed as $\mathbf{y} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b},\ \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T)$

Important result:

1. Applying linearity of expectation we get $\mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{A}\mathbf{x} + \mathbf{b}] = \mathbf{A}\mathbb{E}[\mathbf{x}] + \mathbf{b} = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$
2. Using the property that $\mathrm{Var}(\mathbf{A}\mathbf{x}) = \mathbf{A}\mathrm{Var}(\mathbf{x})\mathbf{A}^T$ we get $\mathrm{Cov}[\mathbf{y}] = \mathbb{E}[(\mathbf{y} - \mathbb{E}[\mathbf{y}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^T] = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T$

> Why is this important?
> This property is the foundation of the Kalman Filter in the "prediction" step.

Multivariate Gaussians

Multivariate Gaussians Ajit Rajwade CS 215, Data Analysis and Interpretation 1

https://docs.google.com/presentation/d/1ElqLbDMwPWuXatmeCx6bSM8anVnkjPTxYdiOKxslkZw/edit?usp=sharing

TRY TO FOLLOW THESE SLIDES (TILL SILDE#56). You can skip the derivations if they're too hard to follow.

Topics for extra reading:

NOTE: These are not relevant for the purpose of the project and so were excluded to be part of this page. However, if you want to read and study them in depth do reach out to us 😇 (we can provide the necessary theory and practice material)

1. Stochastic Processes and Calculus
2. Markov Chains
3. Chapman-Kolmogorov Equations
4. Brownian Motion and Martingales
5. Itô's Lemma & SDEs
6. Black-Scholes (and its derivation 😈)

🔥 Credits to MIT OCW, Stanford Uni theory notes, CMU Statistics 36-705, Sheldon Ross' book and countless other resource pages out there.