

# Mathematics Notes

Junchi Wang  
*Imperial College London*

## 1 Mathematics

### 1.1 Taylor's Series

If  $x \rightarrow a$ ,

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k$$

Specially,

$$e^x = \sum_{k=0}^{\infty} \frac{a^k}{k!} (x-a)^k = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (x \rightarrow 0)$$

### 1.2 Combination

$$C_n^k = \binom{n}{k} = \frac{n!}{(n-k)!k!} \quad (\text{Combinations})$$

$$A_n^k = \frac{n!}{(n-k)!} \quad (\text{Permutations})$$

### 1.3 Differentiate

$$\frac{d}{dx}(\sin x) = \cos x$$

$$\frac{d}{dx}(\cos x) = -\sin x$$

$$\frac{d}{dx}(\tan x) = 1 + \tan^2 x$$

### 1.4 Integral

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

Specially, for  $\mu = 0$  and  $\sigma^2 = 1$ :

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$

### 1.5 Series

$$\sum_{n=1}^{\infty} \frac{1}{n} \rightarrow +\infty$$

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

## 1.6 Trigonometric Functions

$$\cos(A + B) = \cos A \cos B - \sin A \sin B$$

$$\sin(A + B) = \sin A \cos B + \cos A \sin B$$

$$\cos(2A) = \cos^2 A - \sin^2 A = 1 - 2 \sin^2 A = 2 \cos^2 A - 1$$

$$\sin(2A) = 2 \sin A \cos A$$

## 2 Lecture 1: Probability & Stochastic Processes

### 2.1 Normal Distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mathbb{E}[X] = \mu \quad \text{Var}(X) = \sigma^2$$

### 2.2 Poisson Distribution

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$\mathbb{E}[X] = \lambda \quad \text{Var}(X) = \lambda$$

### 2.3 Function of a Random Variable

Let  $y = g(x)$ , then:

$$F_Y(y) = \mathbb{P}(Y \leq y_0) = \mathbb{P}(g(X) \leq y_0) = \int_D f_X(x) dx$$

$$f_Y(y) = \frac{dF_Y(y)}{dy}$$

### 2.4 Mean, Variance, Moments, Characteristic Function

$$\mathbb{E}[X] = \int f_X(x) \cdot x dx = \sum_x x \cdot p(x)$$

$$\text{Var}(X) = \mathbb{E}[(X - \bar{X})^2]$$

$$m_n = \mathbb{E}[X^n]$$

$$\varphi_X(\omega) = \mathbb{E}[e^{j\omega X}] = \int e^{j\omega x} f_X(x) dx$$

### 2.5 Other Distributions

#### Uniform Distribution

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

#### Exponential Distribution

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases}$$

### Bernoulli Distribution

$$P(X = 0) = p, \quad P(X = 1) = 1 - p$$

### Binomial Distribution

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

## 3 Lecture 2: Joint and Marginal Distributions

### 3.1 Joint and Marginal

$$F_{XY}(x, y) = \mathbb{P}(X \leq x_0, Y \leq y_0)$$

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} \quad (\text{Joint})$$

Marginal distributions:

$$f_X(x) = \int f_{XY}(x, y) dy \quad f_Y(y) = \int f_{XY}(x, y) dx$$

### 3.2 Function of Two Random Variables

Let  $Z = g(X, Y)$ , then:

$$F_Z(z) = \mathbb{P}(Z \leq z_0) = \mathbb{P}(g(X, Y) \leq z_0) = \iint_D f_{XY}(x, y) dx dy$$

$$f_Z(z) = \frac{dF_Z(z)}{dz}$$

**Note:** The range of  $Z$  needs to be considered. If  $1 > X, Y > 0$ , then  $1 > Z = XY > 0$ . The integral area will be affected.

### 3.3 Two Functions of Two Random Variables

Let:

$$Z = g(X, Y), \quad W = h(X, Y)$$

Then:

$$f_{ZW}(z, w) = \sum_i \frac{1}{|J(x_i, y_i)|} f_{XY}(x_i, y_i) \quad (\text{joint p.d.f.})$$

Where the Jacobian matrix is:

$$J(x, y) = \det \begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{bmatrix}$$

*This concept can be extended to multiple functions of multiple random variables.*

### 3.4 Covariance and Correlation

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

$$\text{Correlation} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} \in [-1, 1]$$

## 4 Lecture 3: Inequalities and Limit Theorems

### 4.1 Inequality

**Markov Inequality:**

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

**Generalized Inequality:**

$$\mathbb{P}(g(X) \geq g(a)) \leq \frac{\mathbb{E}[g(X)]}{g(a)}$$

**Chebyshev Inequality:** Let  $g(x) = (X - \mu)^2$ , then:

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\mathbb{E}[(X - \mu)^2]}{a^2} = \frac{\sigma^2}{a^2}$$

**Chernoff Bound:** Let  $g(x) = e^{\lambda x}$ , then:

$$\mathbb{P}(e^{\lambda x} \geq e^{\lambda a}) \leq \frac{\mathbb{E}[e^{\lambda x}]}{e^{\lambda a}} = \frac{\int e^{\lambda x} f_X(x) dx}{e^{\lambda a}}$$

So,

$$\mathbb{P}(X \geq a) \leq \min_{\lambda > 0} e^{-\lambda a} \varphi(\lambda)$$

where  $\varphi(\lambda) = \mathbb{E}[e^{\lambda X}]$

### 4.2 Law of Large Numbers

For  $N$  observations of a random variable  $X$ , with  $\mathbb{E}[X] = \mu$ , define:

$$\hat{X} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

Then:

$$\hat{X} \rightarrow \mu \quad \text{as } N \rightarrow \infty$$

### 4.3 Central Limit Theorem

For  $N$  i.i.d. random variables with  $\mu = 0$ , define:

$$Y = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

Then:

$$Y \sim \mathcal{N}(0, \sigma^2)$$

## 5 Lecture 4: Parameters Estimation

### 5.1 Maximum Likelihood

$$\hat{\theta} = \arg \max_{\theta} f_X(x_1, x_2, \dots, x_n; \theta)$$

1. Log-likelihood:

$$\mathcal{L}(x_1, x_2, \dots, x_n; \theta) = \log f_X(x_1, x_2, \dots, x_n; \theta)$$

2. Solve:

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0$$

## 5.2 Cramér-Rao Bound

$$\text{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

$$I(\theta) = \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta} \ln f_X(x_1, x_2, \dots, x_n; \theta) \right)^2 \right]$$

## 5.3 MMSE Estimator

$$y(n) = \mathbf{w}^T(n) \mathbf{x}(n)$$

where:

$$\mathbf{w}(n) = [w(n) \quad w(n-1) \quad \dots \quad w(n-M+1)]^T, \quad \mathbf{x}(n) = [x(n) \quad x(n-1) \quad \dots \quad x(n-M+1)]^T$$

$$\mathbf{w}^*(n) = \mathbf{R}^{-1} \mathbf{P}$$

with:

$$\mathbf{R} = \mathbb{E}[\mathbf{x}(n) \mathbf{x}^H(n)], \quad \mathbf{P} = \mathbb{E}[d(n) \mathbf{x}^*(n)]$$

If  $y(n) = x(n+1)$ , it becomes an MSE predictor.

**MSE Expression:**

$$\begin{aligned} \text{MSE} &= \mathbb{E}[(x(n+1) - \hat{x}(n+1))^2] = \mathbb{E}[(x(n+1) - \mathbf{w}^T(n) \mathbf{x}(n))^2] \\ &= R(0) - 2\mathbf{w}^T \mathbf{P} + \mathbf{w}^T \mathbf{R} \mathbf{w} \\ &= R(0) - \mathbf{w}^T \mathbf{P} \end{aligned}$$

## 6 Lecture 5: Stochastic Processes

### 6.1 Stochastic Process

Assume a stochastic process  $X(t)$ . For fixed  $t$ ,  $X(t)$  is a random variable.

$$F_X(x, t) = \mathbb{P}(X(t) \leq x)$$

$$f_X(x, t) = \frac{d}{dx} F_X(x, t)$$

### 6.2 Mean, Autocorrelation, Covariance

$$\mu(t) \triangleq \int_{-\infty}^{\infty} x_t f_X(x, t) dx$$

$$R_{XX}(t_1, t_2) \triangleq \mathbb{E}[X_{t_1} X_{t_2}^*] = \iint X_{t_1} X_{t_2}^* f(x_{t_1}, x_{t_2}) dx_{t_1} dx_{t_2}$$

$$\text{Cov}(t_1, t_2) \triangleq \mathbb{E}[(X_{t_1} - \mu(t_1))(X_{t_2} - \mu(t_2))^*]$$

Properties:

$$R_{XX}(t_2, t_1) = R_{XX}^*(t_1, t_2)$$

### 6.3 Stationarity

**Strict-Sense Stationary:** A process is  $n$ -th order strictly stationary if:

$$f(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = f(x_1, x_2, \dots, x_n; t_1 + c, t_2 + c, \dots, t_n + c), \quad \forall c$$

i.e., the distribution depends only on the **relative time differences**.

Example:

$$f(x_1, x_2, x_3; t_1, t_2, t_3) = f(x_1, x_2, x_3; t_1 + c, t_2 + c, t_3 + c) = f(x_1, x_2, x_3; 0, t_2 - t_1, t_3 - t_1)$$

**Wide-Sense Stationary (WSS): 1st Order:**

$$f(x; t_1) = f(x, 0) \quad \text{and} \quad \mathbb{E}[X(t)] = \mu(t) \text{ doesn't depend on } t$$

**2nd Order:**

$$f(x_1, x_2; t_1, t_2) = f(x_1, x_2; 0, t_2 - t_1)$$

$$\mathbb{E}[X(t_1)X(t_2)] = \mu(t_1, t_2) \text{ depends only on } \tau = t_2 - t_1$$

## 7 Lecture 6: Power Spectrum

### 7.1 ARMA Model

$$x(n) = - \sum_{k=1}^p a_k x(n-k) + \sum_{k=0}^q b_k y(n-k)$$

Taking the Z-transform:

$$X(z) + \sum_{k=1}^p a_k X(z) z^{-k} = \sum_{k=0}^q Y(z) b_k z^{-k}$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1 + \sum_{k=1}^p a_k z^{-k}}{\sum_{k=0}^q b_k z^{-k}}$$

### 7.2 Wiener-Khinchin Theorem

$$R_{XX}(\tau) \xleftrightarrow{\text{FT}} S_{XX}(\omega)$$

$$S_{XX}(\omega) = \int R_{XX}(\tau) e^{-j\omega\tau} d\tau$$

$$R_{XX}(\tau) = \frac{1}{2\pi} \int S_{XX}(\omega) e^{j\omega\tau} d\omega$$

### 7.3 Matched Filter

Let:

$$r(t) = s(t) + w(t)$$

$$y(t) = y_s(t) + n(t), \quad y(t) = h(t) * r(t)$$

$$y_s(t) = s(t) * h(t), \quad n(t) = h(t) * w(t)$$

Maximum output SNR at time  $t_0$ :

$$\text{SNR} = \frac{P_s}{P_n} = \frac{|Y_s(t_0)|^2}{\mathbb{E}[|n(t_0)|^2]} = \frac{|Y_s(t_0)|^2}{R_{nn}(0)}$$

In frequency domain:

$$Y_s(\omega) = S(\omega)H(\omega)$$

$$\begin{aligned} \text{SNR} &= \frac{1}{2\pi} \cdot \frac{|\int S(\omega)H(\omega)e^{j\omega t_0} d\omega|^2}{\int S_{nn}(\omega)|H(\omega)|^2 d\omega} \\ &= \frac{1}{2\pi} \cdot \frac{|\int S(\omega)H(\omega)e^{j\omega t_0} d\omega|^2}{\int N_0|H(\omega)|^2 d\omega} = \frac{E_s}{N_0} \end{aligned}$$

$$H(\omega) = S^*(\omega)e^{-j\omega t_0} \Rightarrow h(t) = s(t_0 - t)$$

## 8 Lecture 8: Markov Chain

### 8.1 Markov Chain Definition

Discrete time:  $1, 2, \dots, n, n+1$

Discrete state:  $i_0, i_1, \dots, i_n$

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

A \*\*homogeneous\*\* Markov chain satisfies:

$$P_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i) \quad (\text{independent of } n)$$

### 8.2 Markov Matrix

$$P = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

### 8.3 Probability Distribution

$\pi^{(0)}$  — probability distribution at  $t = 0$   $\pi^{(n)}$  — probability distribution at  $t = n$

$$\pi^{(n)} = \pi^{(0)} P^n$$

### 8.4 Limiting Probability Distribution

$$\pi = \lim_{n \rightarrow \infty} \pi^{(n)} = \lim_{n \rightarrow \infty} \pi^{(0)} P^n$$

Assume convergence:

$$\pi = \pi P \Rightarrow \pi \cdot \mathbf{1} = \pi P \quad (\text{stationary distribution})$$

## 8.5 Stationary Distribution and Eigen Decomposition

Recall eigenvalue relation:

$$\lambda \mathbf{v} = A \mathbf{v}$$

Stationary distribution  $\pi$  is the eigenvector corresponding to  $\lambda = 1$ .

Example:

$$P = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \Rightarrow \pi = [0.3047 \quad 0.3905 \quad 0.3048]$$

## Lecture 8 (Continued): Random Walker / Gambler's Ruin

### 1. Random Walker

Transition matrix  $P$  for a random walk with a left barrier:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ q & 0 & p & 0 & \cdots \\ 0 & q & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix}$$

There is a barrier on the left (at 0) that prevents moving further left.

### 2. Gambler's Ruin

Transition matrix  $P$  for Gambler's Ruin problem:

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ q & 0 & p & \cdots & 0 & 0 \\ 0 & q & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & p & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

The gambler stops gambling if they either: - Lose all their money (reach state 0), or - Reach the target amount  $N$ .

**Ruin Probability:** Given initial capital  $i$ , the probability of ruin (reaching state 0) is:

$$P_i = \begin{cases} \frac{1 - (\frac{p}{q})^{N-i}}{1 - (\frac{p}{q})^N} & \text{if } p \neq \frac{1}{2} \\ \frac{N-i}{N} & \text{if } p = \frac{1}{2} \end{cases} = \mathbb{P}(S_T = 0)$$

**Notes:** -  $N$ : Target capital (absorbing boundary at the top) -  $\mathbb{P}(S_{n+1} = S_n + 1) = p$  -  $\mathbb{P}(S_{n+1} = S_n - 1) = q$  -  $p + q = 1$

## 9 Lecture 9: Continuous-Time Processes

### 9.1 1. Poisson Processes

Define  $X(t) = n(t_1, t_2)$ , the number of arrivals in the interval  $(t_1, t_2)$ . Let  $t = t_2 - t_1$ . This is a Poisson process if:

$$\mathbb{P}(n(t_1, t_2) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$



Expected value:

$$\mathbb{E}[n(t_1, t_2)] = \lambda t$$

Where  $\lambda$  is the expected number of events per unit time. Example: Average one car accident in 5 minutes implies  $\lambda = 0.2$  per minute.

## 9.2 2. Poisson Meets Bernoulli

Suppose: - Number of car accidents follows a Poisson distribution. - Whether a car accident is fatal or not follows a Bernoulli distribution.

Let: -  $n$ : total number of accidents -  $k$ : number of fatal accidents

**Step 1: Poisson distribution for total accidents:**

$$\mathbb{P}(n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

**Step 2: Binomial distribution for  $k$  fatal out of  $n$ :**

$$\mathbb{P}(k | n) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n \geq k$$

**Step 3: Total probability of  $k$  fatal accidents:**

$$\begin{aligned} \mathbb{P}(k) &= \sum_{n=k}^{\infty} \mathbb{P}(k | n) \mathbb{P}(n) \\ &= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{(\lambda t)^n e^{-\lambda t}}{n!} \\ &= \sum_{n=k}^{\infty} \frac{[(1-p)\lambda t]^{n-k}}{(n-k)!} \cdot \frac{(\lambda t p)^k e^{-\lambda t}}{k!} \\ &= \frac{(\lambda t p)^k e^{-\lambda t}}{k!} \sum_{n=0}^{\infty} \frac{[(1-p)\lambda t]^n}{n!} \\ &= \frac{(\lambda t p)^k}{k!} e^{-(1-p)\lambda t} e^{-p\lambda t} = \frac{(\lambda p t)^k e^{-\lambda p t}}{k!} \end{aligned}$$

Hence, the number of fatal car accidents also follows a Poisson distribution with parameter  $\lambda p t$ .

## 9.3 Inter-arrival Interval

The time interval between any two arrivals  $\tau_n$  follows exponential distribution, i.e.

$$f_{\tau_n}(t) = \lambda e^{-\lambda t}$$

# 10 Lecture 10: Martingales

## 10.1 1. Martingales

A sequence  $X_n$  is a martingale if:

$$\mathbb{E}[X_{n+1} | X_n, \dots, X_1] = X_n$$

This represents a form of "stability."

## 10.2 2. Generalized Martingales

A sequence  $S_n$  with finite mean is a martingale with respect to the sequence  $X_n$  if:

$$\mathbb{E}[S_{n+1} \mid X_n, \dots, X_1] = S_n$$

**Submartingale:**

$$\mathbb{E}[S_{n+1} \mid X_1, X_2, \dots, X_n] \geq S_n$$

**Supermartingale:**

$$\mathbb{E}[S_{n+1} \mid X_1, X_2, \dots, X_n] \leq S_n$$

## 10.3 3. Stopping Time $T$

Stopping time is the time at which a stochastic process stops due to some condition or outcome. For example, when a gambler runs out of money, the stochastic process of his money stops.

If  $X_n$  is a martingale, then the following property holds:

$$\mathbb{E}[X_T] = \mathbb{E}[X_0]$$