

# Session 1

Tuesday, 25 November 2025 23.17

# Bivariate Exercises

Sunday, 16 November 2025 14.50

**12.5 (c)** In this problem we estimate  $\pi$  using probability arguments. Let  $X \sim U(-1, 1)$  and  $Y \sim U(-1, 1)$  for  $X$  and  $Y$  independent. First relate  $P[X^2 + Y^2 \leq 1]$  to the value of  $\pi$ . Then generate realizations of  $X$  and  $Y$  and use them to estimate  $\pi$ .

## 12.5 (c) – Estimating $\pi$ with probability

We're told:

- $X \sim U(-1, 1)$
- $Y \sim U(-1, 1)$
- $X$  and  $Y$  are independent.

### 1. Relate $P[X^2 + Y^2 \leq 1]$ to $\pi$

The pair  $(X, Y)$  is uniformly distributed over the square  $[-1, 1] \times [-1, 1]$ .

- Area of the square:  $2 \times 2 = 4$ .
- The condition  $X^2 + Y^2 \leq 1$  is the unit disk (circle of radius 1) centered at the origin.
- Area of that circle:  $\pi \cdot 1^2 = \pi$ .

Because the distribution is uniform over the square,

$$P[X^2 + Y^2 \leq 1] = \frac{\text{area of circle}}{\text{area of square}} = \frac{\pi}{4}.$$

So

$$\pi = 4 P[X^2 + Y^2 \leq 1].$$

### 2. Simulation idea + pseudocode

We can estimate this probability by simulation, then turn it into an estimate of  $\pi$ .

Algorithm idea:

1. Pick a large integer  $N$  (e.g. 100,000).
2. For each  $i = 1, \dots, N$ :
  - Generate  $X_i \sim U(-1, 1)$ .
  - Generate  $Y_i \sim U(-1, 1)$ .
  - If  $X_i^2 + Y_i^2 \leq 1$ , increase a counter `inside` by 1.
3. Estimate the probability as  $\hat{p} = \text{inside}/N$ .
4. Estimate  $\pi$  as  $\hat{\pi} = 4\hat{p}$ .

Pseudocode:

```
pseudo
N := 100000
inside := 0

for i from 1 to N do
    X := uniform_random(-1, 1)
    Y := uniform_random(-1, 1)
    if X*X + Y*Y <= 1 then
        inside := inside + 1
    end_if
```

Copy code

```

    i := uniform(1, 1)
    if X*X + Y*Y <= 1 then
        inside := inside + 1
    end if
end for

p_hat := inside / N
pi_hat := 4 * p_hat

print("Estimated pi =", pi_hat)

```

**12.11 (c) (f)** If  $(X, Y)$  has a standard bivariate Gaussian PDF, find  $P[X^2 + Y^2 = 10]$ .

### 12.11 (f) – Standard bivariate Gaussian and $X^2 + Y^2$

"Standard bivariate Gaussian" here means

- $X \sim N(0, 1)$ ,
- $Y \sim N(0, 1)$ ,
- $X$  and  $Y$  independent.

Then the joint density is

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right).$$

The random variable

$$S = X^2 + Y^2$$

has a chi-square distribution with 2 degrees of freedom, i.e.

$$S \sim \chi_2^2.$$

For  $\chi_2^2$ , the pdf is

$$f_S(s) = \frac{1}{2} e^{-s/2}, \quad s \geq 0,$$

and the CDF is

$$P(S \leq s) = 1 - e^{-s/2}.$$

The exact inequality sign in your scan is a bit hard to see, so here are both useful forms:

- If the problem is  $P[X^2 + Y^2 \leq 10]$ :

$$P[X^2 + Y^2 \leq 10] = 1 - e^{-10/2} = 1 - e^{-5}.$$

- If the problem is  $P[X^2 + Y^2 \geq 10]$ :

$$P[X^2 + Y^2 \geq 10] = e^{-10/2} = e^{-5}.$$

(And if it literally said " $= 10$ ", then for a continuous distribution the probability is 0.)

**12.15 (c) (f)** If

$$p_{X,Y}(x,y) = \begin{cases} 2 & 0 < x < 1, 0 < y < x \\ 0 & \text{otherwise} \end{cases}$$

find the marginal PDFs.

### 12.15 (f) – Marginal PDFs

We are given the joint pdf

$$p_{X,Y}(x,y) = \begin{cases} 2, & 0 < x < 1, 0 < y < x, \\ 0, & \text{otherwise.} \end{cases}$$

This describes a uniform density over the triangular region

$$\{(x,y) : 0 < x < 1, 0 < y < x\}.$$

#### 1. Marginal of $X$ : $p_X(x)$

Use

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y) dy.$$

For a fixed  $x$ , the allowed  $y$ -values must satisfy  $0 < y < x$ . So

- If  $0 < x < 1$ :

$$p_X(x) = \int_{y=0}^x 2 dy = 2x.$$

- Otherwise,  $p_X(x) = 0$ .

So

$$p_X(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

(Quick check:  $\int_0^1 2x dx = x^2|_0^1 = 1$ . Good.)

## 2. Marginal of $Y$ : $p_Y(y)$

Now

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x, y) dx.$$

From the support, we have  $0 < y < x < 1$ . For a fixed  $y$ , this means:

- $y$  must be between 0 and 1.
- $x$  must range from  $x = y$  up to  $x = 1$ .

So:

- If  $0 < y < 1$ :

$$p_Y(y) = \int_{x=y}^1 2 dx = 2(1 - y).$$

- Otherwise,  $p_Y(y) = 0$ .

Thus

$$p_Y(y) = \begin{cases} 2(1 - y), & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

(Again,  $\int_0^1 2(1 - y) dy = 1$ , so it's a proper pdf.)

**12.38 (c) (f)** If  $X$  and  $Y$  have the joint PDF

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\right)$$

find the joint PDF of the transformed random vector

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

### 12.38 (f) – Linear transform of a bivariate normal

We're given

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\right).$$

Define the new vector

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

Call the transformation matrix  $A$ :

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}, \quad \mu = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

#### 1. Mean of $(W, Z)$

For a linear transform of a normal vector:

$$\begin{bmatrix} W \\ Z \end{bmatrix} = A \begin{bmatrix} X \\ Y \end{bmatrix} \implies \mu_{WZ} = A\mu.$$

Compute:

$$\mu_{WZ} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 1 \cdot 2 \\ 2 \cdot 1 + 3 \cdot 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \end{bmatrix}.$$

So  $E[W] = 3$ ,  $E[Z] = 8$ .

#### 2. Covariance of $(W, Z)$

Again, for a linear transform:

$$\Sigma_{WZ} = A\Sigma A^T.$$

First, compute  $A\Sigma$ :

$$A\Sigma = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 1 \cdot (-1) & 1 \cdot (-1) + 1 \cdot 2 \\ 2 \cdot 2 + 3 \cdot (-1) & 2 \cdot (-1) + 3 \cdot 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}.$$

Then

$$\Sigma_{WZ} = (A\Sigma)A^T = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 14 \end{bmatrix}.$$

So

$$\Sigma_{WZ} = \begin{bmatrix} \text{Var}(W) & \text{Cov}(W, Z) \\ \text{Cov}(W, Z) & \text{Var}(Z) \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 14 \end{bmatrix}.$$

### 3. Distribution and joint pdf of $(W, Z)$

Because a linear transform of a multivariate normal is also multivariate normal,

$$\begin{bmatrix} W \\ Z \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 3 \\ 8 \end{bmatrix}, \begin{bmatrix} 2 & 5 \\ 5 & 14 \end{bmatrix}\right).$$

The determinant of  $\Sigma_{WZ}$  is

$$|\Sigma_{WZ}| = 2 \cdot 14 - 5 \cdot 5 = 28 - 25 = 3.$$

The inverse is

$$\Sigma_{WZ}^{-1} = \frac{1}{3} \begin{bmatrix} 14 & -5 \\ -5 & 2 \end{bmatrix}.$$

So the joint pdf of  $(W, Z)$  is

$$f_{W,Z}(w, z) = \frac{1}{2\pi\sqrt{|\Sigma_{WZ}|}} \exp\left(-\frac{1}{2} [w-3 \ z-8] \Sigma_{WZ}^{-1} \begin{bmatrix} w-3 \\ z-8 \end{bmatrix}\right),$$

i.e.

$$f_{W,Z}(w, z) = \frac{1}{2\pi\sqrt{3}} \exp\left(-\frac{1}{2} [w-3 \ z-8] \frac{1}{3} \begin{bmatrix} 14 & -5 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} w-3 \\ z-8 \end{bmatrix}\right).$$

That completely specifies the transformed joint pdf.

# Conditional Probability Densities and Expectation Exercises

Tuesday, 25 November 2025 21.33

**13.2 (c) (w)** Determine if the proposed conditional PDF

$$p_{Y|X}(y|x) = \begin{cases} c \exp(-y/x) & y \geq 0, x > 0 \\ 0 & \text{otherwise} \end{cases}$$

is a valid conditional PDF for some  $c$ . If so, find the required value of  $c$ .

## 13.2 – Validity of a conditional PDF

Proposed conditional pdf:

$$p_{Y|X}(y|x) = \begin{cases} ce^{-y/x}, & y \geq 0, x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

For a fixed  $x > 0$ , a conditional pdf in  $y$  must integrate to 1:

$$\int_{-\infty}^{\infty} p_{Y|X}(y|x) dy = \int_0^{\infty} ce^{-y/x} dy = 1.$$

Compute the integral:

$$\int_0^{\infty} e^{-y/x} dy = \left[ -xe^{-y/x} \right]_0^{\infty} = 0 - (-x) = x.$$

So

$$c \cdot x = 1 \Rightarrow c = \frac{1}{x}.$$

But  $c$  is supposed to be a *constant*, the same for all  $x > 0$ .

There is no single constant  $c$  that satisfies  $cx = 1$  for all  $x > 0$ .

- So with constant  $c$ : this is not a valid family of conditional pdfs.
- If we were allowed to let  $c$  depend on  $x$ , we would need

$$c(x) = \frac{1}{x},$$

and then it would be valid.

**13.4 (c) (f)** If

$$p_{X,Y}(x,y) = \begin{cases} 2 \exp[-(x+y)] & 0 \leq y \leq x, x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

find  $p_{Y|X}(y|x)$ .

### 13.4(f) – Find $p_{Y|X}(y|x)$

Joint pdf:

$$p_{X,Y}(x, y) = \begin{cases} 2e^{-(x+y)}, & 0 \leq y \leq x, x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

#### 1. Find the marginal $p_X(x)$

For  $x \geq 0$ , allowed  $y$  are  $0 \leq y \leq x$ :

$$\begin{aligned} p_X(x) &= \int_{-\infty}^{\infty} p_{X,Y}(x, y) dy = \int_0^x 2e^{-(x+y)} dy = 2e^{-x} \int_0^x e^{-y} dy. \\ &\quad \int_0^x e^{-y} dy = 1 - e^{-x}. \end{aligned}$$

So

$$p_X(x) = 2e^{-x}(1 - e^{-x}), \quad x \geq 0,$$

and 0 otherwise.

#### 2. Conditional pdf $p_{Y|X}(y|x)$

By definition

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

For  $0 \leq y \leq x, x > 0$ :

$$p_{Y|X}(y|x) = \frac{2e^{-(x+y)}}{2e^{-x}(1 - e^{-x})} = \frac{e^{-y}}{1 - e^{-x}}, \quad 0 \leq y \leq x.$$

So

$$p_{Y|X}(y|x) = \begin{cases} \frac{e^{-y}}{1 - e^{-x}}, & 0 \leq y \leq x, x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

This is an exponential density in  $y$ , truncated to  $[0, x]$  and then normalized.

## 13.5 (w) Plot the joint PDF

$$p_{X,Y}(x, y) = \begin{cases} 2x & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Next determine by inspection the conditional PDF  $p_{Y|X}(y|x)$ . Recall that the conditional PDF is just the normalized cross-section of the joint PDF.

## 13.5 – Plot the joint pdf and find $p_{Y|X}(y|x)$

Joint pdf:

$$p_{X,Y}(x,y) = \begin{cases} 2x, & 0 < x < 1, 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

### 1. Shape / plot

- Support is the unit square:  $0 < x < 1, 0 < y < 1$ .
- For each fixed  $x$ ,  $p_{X,Y}(x,y) = 2x$  does **not depend on  $y$** , so it's a vertical "curtain" that increases linearly with  $x$ .

If you wanted to plot this on a computer:

pseudo

 Copy code

```
# Pseudocode to plot p_{X,Y}
for x from 0 to 1 step small_step:
    for y from 0 to 1 step small_step:
        z := 2 * x           # value of joint pdf
        plot_point(x, y, z) # e.g. surface or heatmap
```

## 2. Conditional pdf $p_{Y|X}(y|x)$

For fixed  $x \in (0, 1)$ , the cross-section as a function of  $y$  is:

- $p_{X,Y}(x,y) = 2x$  for  $0 < y < 1$ .

Normalize over  $y$ :

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{\int_0^1 p_{X,Y}(x,t) dt} = \frac{2x}{\int_0^1 2x dt}.$$
$$\int_0^1 2x dt = 2x.$$

So

$$p_{Y|X}(y|x) = \frac{2x}{2x} = 1, \quad 0 < y < 1.$$

Thus

$$p_{Y|X}(y|x) = \begin{cases} 1, & 0 < y < 1, 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

This means  $Y$  is uniform on  $(0, 1)$  even after conditioning on  $X$ ; in fact,  $X$  and  $Y$  are independent.

**13.8 (c) (f)** If  $X \sim \mathcal{U}(0, 1)$  and  $Y|(X = x) \sim \mathcal{U}(0, x)$ , find the joint PDF for  $X$  and  $Y$  and also the marginal PDF for  $Y$ .

### 13.8(f) – Joint pdf and marginal of $Y$

Given:

- $X \sim U(0, 1)$ .
- $Y | (X = x) \sim N(0, x)$ .

Here  $N(0, x)$  means mean 0, variance  $x$  (so standard deviation  $\sqrt{x}$ ).

#### 1. Joint pdf $p_{X,Y}(x, y)$

Use

$$p_{X,Y}(x, y) = p_{Y|X}(y|x) p_X(x).$$

- $p_X(x) = 1$  for  $0 < x < 1$ , 0 otherwise.
- For variance  $x$ , the conditional normal pdf is

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{y^2}{2x}\right), \quad x > 0, -\infty < y < \infty.$$

So

$$p_{X,Y}(x, y) = \begin{cases} \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{y^2}{2x}\right), & 0 < x < 1, -\infty < y < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

#### 2. Marginal pdf $p_Y(y)$

Integrate out  $x$ :

$$p_Y(y) = \int_0^1 p_{X,Y}(x, y) dx = \int_0^1 \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{y^2}{2x}\right) dx.$$

This integral does not simplify to a basic elementary function; it's perfectly acceptable to leave it in this form:

$$p_Y(y) = \int_0^1 \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{y^2}{2x}\right) dx, \quad -\infty < y < \infty.$$

(You can also note that  $p_Y(y)$  is an even function:  $p_Y(y) = p_Y(-y)$ , since the integrand depends on  $y^2$ .)

**13.16 (c) (w)** A resistor is chosen from a bin of 10 ohm resistors whose distribution satisfies  $R \sim \mathcal{N}(10, 0.25)$ . A  $i = 1$  amp current source is applied to the resistor and the subsequent voltage  $V$  is measured with a voltmeter. The voltmeter has an error  $E$  that is modeled as  $E \sim \mathcal{N}(0, 1)$ . Find the probability that  $V > 10$  volts if an 11 ohm resistor is chosen. Note that  $V = iR + E$ . What assumption do you need to make about the dependence between  $R$  and  $E$ ?

## 13.16 – Probability that measured voltage exceeds 10 V

We have:

- True resistance  $R \sim N(10, 0.25)$ :
  - mean  $E[R] = 10$ ,
  - variance  $\text{Var}(R) = 0.25$  (sd 0.5).
- Measurement error  $E \sim N(0, 1)$ .
- Applied current  $i = 1$  amp.
- Measured voltage:

$$V = iR + E = R + E.$$

We're asked:  $P(V > 10)$  using a high precision voltmeter (modeled here by the given  $E$ ) and what dependence assumption is needed between  $R$  and  $E$ .

### 1. Distribution of $V$

Assume  $R$  and  $E$  are independent (this is the key assumption they ask for).

Sum of independent normals is normal:

- Mean:

$$E[V] = E[R] + E[E] = 10 + 0 = 10.$$

- Variance:

$$\text{Var}(V) = \text{Var}(R) + \text{Var}(E) = 0.25 + 1 = 1.25.$$

So

$$V \sim N(10, 1.25).$$

### 2. Probability $P(V > 10)$

Standardize:

$$P(V > 10) = P\left(\frac{V - 10}{\sqrt{1.25}} > 0\right).$$

But

$$Z = \frac{V - 10}{\sqrt{1.25}} \sim N(0, 1).$$

For a standard normal,

$$P(Z > 0) = 0.5.$$

So

$$P(V > 10) = 0.5.$$

### 3. Assumption about dependence

To combine the variances as we did and assert that  $V$  is normal with variance  $0.25 + 1$ , we must assume:

- The measurement error  $E$  is independent of the true resistance  $R$  (equivalently,  $\text{Cov}(R, E) = 0$ ).

That's the required assumption about their dependence.

# N-Dimensional Random Vectors Exercises

Tuesday, 25 November 2025 21.36

**14.1 (∴) (w,f)** If  $Y = X_1 + X_2 + X_3$ , where  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$  and

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$
$$\mathbf{C} = \begin{bmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1 \end{bmatrix}$$

find the mean and variance of  $Y$ .

## 14.1 – Mean and variance of $Y = X_1 + X_2 + X_3$

We're given a 3-D Gaussian random vector

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}),$$

with

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1 \end{bmatrix}.$$

Define  $Y = X_1 + X_2 + X_3$ . In vector form,

$$Y = \mathbf{a}^\top \mathbf{X}, \quad \mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

### Mean of $Y$

For a linear function of a random vector,

$$E[Y] = \mathbf{a}^\top E[\mathbf{X}] = \mathbf{a}^\top \boldsymbol{\mu}.$$

So

$$E[Y] = [1 \ 1 \ 1] \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 + 2 + 3 = 6.$$

## Mean of $Y$

For a linear function of a random vector,

$$E[Y] = a^\top E[X] = a^\top \mu.$$

So

$$E[Y] = [1 \ 1 \ 1] \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 + 2 + 3 = 6.$$

## Variance of $Y$

For a Gaussian vector,

$$\text{Var}(Y) = a^\top C a.$$

Compute:

$$Ca = \begin{bmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 + 1/2 + 1/4 \\ 1/2 + 1 + 1/2 \\ 1/4 + 1/2 + 1 \end{bmatrix} = \begin{bmatrix} 1.75 \\ 2 \\ 1.75 \end{bmatrix}.$$

Then

$$a^\top (Ca) = [1 \ 1 \ 1] \begin{bmatrix} 1.75 \\ 2 \\ 1.75 \end{bmatrix} = 1.75 + 2 + 1.75 = 5.5 = \frac{11}{2}.$$

Answer:

$$E[Y] = 6, \quad \text{Var}(Y) = \frac{11}{2} = 5.5.$$

- 14.6 (∴) (w)** A large group of college students have weights that can be modeled as a  $\mathcal{N}(150, 30)$  random variable. If 4 students are selected at random, what is the probability that they will all weigh more than 150 lbs?

## 14.6 – Probability 4 students all weigh more than 150 lb

Weights of students are modeled as

$$W \sim \mathcal{N}(150, 30),$$

so mean 150, variance 30 (standard deviation  $\sqrt{30}$ ). Assume weights of different students are independent.

We want

$$P(\text{all } 4 > 150) = (P(W > 150))^4.$$

Standardize one student:

$$Z = \frac{W - 150}{\sqrt{30}} \sim \mathcal{N}(0, 1).$$

Then

$$P(W > 150) = P(Z > 0) = 0.5.$$

Therefore

$$P(\text{all } 4 > 150) = (0.5)^4 = \frac{1}{16} = 0.0625.$$

- 14.11 (c)** Simulate a single realization of a random vector composed of IID random variables with PDF  $X_i \sim \mathcal{N}(1, 2)$  for  $i = 1, 2, \dots, N$ . Do this by repeating an experiment that successively generates  $X \sim \mathcal{N}(1, 2)$ . Then, find the outcome of the sample mean random variable and discuss what happens as  $N$  becomes large.

### 14.11 (c) – Sample mean of IID $\mathcal{N}(1, 2)$ variables (simulation idea)

We have IID random variables

$$X_i \sim \mathcal{N}(1, 2), \quad i = 1, 2, \dots, N.$$

Define the sample mean

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i.$$

#### Theory

- Each  $X_i$  has mean 1 and variance 2.
- Since they're IID,

$$E[\bar{X}_N] = 1,$$

$$\text{Var}(\bar{X}_N) = \frac{2}{N}.$$

As  $N$  becomes large,  $\text{Var}(\bar{X}_N) \rightarrow 0$ .

So  $\bar{X}_N$  concentrates around 1 and converges to 1 in probability (Law of Large Numbers).

## Pseudocode for simulation

For a single realization of the  $N$ -vector and its sample mean:

pseudo

 Copy code

```
N := 1000      # choose any N
sum := 0

for i from 1 to N do
    Xi := normal_random(mean = 1, variance = 2)
    sum := sum + Xi
end for

Xbar := sum / N
print("Sample mean =", Xbar)
```

To see what happens as  $N$  gets large, repeat this for different  $N$  (e.g. 10, 100, 1000, 10000) and observe that the printed  $\bar{X}_N$  gets closer and closer to 1 and varies less.

**14.12 (S) (w,c)** An  $N \times 1$  random vector  $\mathbf{X}$  has  $E_{X_i}[X_i] = \mu$  and  $\text{var}(X_i) = i\sigma^2$  for  $i = 1, 2, \dots, N$ . The components of  $\mathbf{X}$  are independent. Does the sample mean random variable converge to  $\mu$  as  $N$  becomes large? Carry out a computer simulation for this problem and explain your results.

## 14.12 (w,c) – Does the sample mean converge to $\mu$ ?

We have a random vector of length  $N$  whose components satisfy:

- $E[X_i] = \mu$ ,
- $\text{Var}(X_i) = i\sigma^2$ , for  $i = 1, 2, \dots, N$ ,
- the components  $X_i$  are independent.

Define the sample mean:

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i.$$

**Mean of  $\bar{X}_N$**

$$E[\bar{X}_N] = \frac{1}{N} \sum_{i=1}^N E[X_i] = \frac{1}{N} \sum_{i=1}^N \mu = \mu.$$

So the sample mean is **unbiased**.

**Variance of  $\bar{X}_N$**

Because the  $X_i$  are independent, variances add:

$$\text{Var}(\bar{X}_N) = \frac{1}{N^2} \sum_{i=1}^N \text{Var}(X_i) = \frac{1}{N^2} \sum_{i=1}^N i\sigma^2 = \frac{\sigma^2}{N^2} \cdot \frac{N(N+1)}{2} = \frac{\sigma^2}{2} \cdot \frac{N+1}{N}.$$

As  $N \rightarrow \infty$ ,

$$\text{Var}(\bar{X}_N) \rightarrow \frac{\sigma^2}{2} \neq 0.$$

Because the variance does **not** go to 0, the sample mean does **not** converge to  $\mu$  in mean square (and hence not in probability). So the sample mean is *not* a consistent estimator of  $\mu$  in this setting.

## Pseudocode for simulation

A way to simulate such  $X_i$ :

- Write  $X_i = \mu + \sigma\sqrt{i} Z_i$ , where  $Z_i \sim N(0, 1)$  independent.

pseudo

 Copy code

```
N := 1000
mu := 0           # pick some values
sigma := 1

sum := 0
for i from 1 to N do
    Zi := normal_random(mean = 0, variance = 1)
    Xi := mu + sigma * sqrt(i) * Zi
    sum := sum + Xi
end for

Xbar := sum / N
print("Sample mean =", Xbar)
```

If you repeat this simulation many times and plot a histogram of  $\bar{X}_N$  for large  $N$ , you'll see that its spread (standard deviation) stays around  $\sqrt{\sigma^2/2} = \sigma/\sqrt{2}$ , not shrinking to zero. That matches the theoretical variance above.

**14.19 (..) (w,c)** Find the mean and variance of the random variable

$$Y = \sum_{i=1}^{12} (U_i - 1/2)$$

where  $U_i \sim \mathcal{U}(0, 1)$  and the  $U_i$ 's are IID. Estimate the PDF of  $Y$  using a computer simulation and compare it to a standard Gaussian PDF. See Section 15.5 for a theoretical justification of your results.

## 14.19 (w,c) – Mean/variance of $Y$ and simulation of its PDF

Random variable:

$$Y = \sum_{i=1}^{12} (U_i - 1/2),$$

where the  $U_i$  are IID  $U(0, 1)$ .

### Mean of $Y$

For a single uniform,

$$E[U_i] = \frac{1}{2}.$$

So

$$E[U_i - 1/2] = \frac{1}{2} - \frac{1}{2} = 0.$$

Since expectation is linear and the terms are identical,

$$E[Y] = \sum_{i=1}^{12} E[U_i - 1/2] = \sum_{i=1}^{12} 0 = 0.$$

### Variance of $Y$

For  $U_i \sim U(0, 1)$ ,

$$\text{Var}(U_i) = \frac{1}{12}.$$

Subtracting a constant does not change variance, so

$$\text{Var}(U_i - 1/2) = \frac{1}{12}.$$

The  $U_i$  are independent, so the variances add:

$$\text{Var}(Y) = \sum_{i=1}^{12} \text{Var}(U_i - 1/2) = 12 \cdot \frac{1}{12} = 1.$$

So:

$$E[Y] = 0, \quad \text{Var}(Y) = 1.$$

By the Central Limit Theorem, the sum of many independent terms with finite variance is approximately normal, so  $Y$  should be roughly distributed like a standard Gaussian  $N(0, 1)$ .

## Pseudocode to estimate the PDF and compare to standard normal

pseudo

 Copy code

```
num_trials := 100000      # large number for a good approximation
Y_samples := empty_list

for k from 1 to num_trials do
    sum := 0
    for i from 1 to 12 do
        Ui := uniform_random(0, 1)
        sum := sum + (Ui - 0.5)
    end for
    append sum to Y_samples
end for

# Now estimate the PDF of Y (e.g., via a histogram)
plot_histogram(Y_samples, normalized = true)

# For comparison, also plot the standard normal pdf N(0,1)
# over the same x-range on the same figure
for x in some_range:
    normal_pdf := (1 / sqrt(2 * pi)) * exp(-x^2 / 2)
    plot_curve(x, normal_pdf)
```

You should see the histogram of  $Y$  closely follow the bell shape of the standard normal curve, which matches the fact that  $E[Y] = 0$  and  $\text{Var}(Y) = 1$  and that a sum of many small independent terms tends toward a Gaussian.

# Discrete Random Variables, PMF and CDF

Tuesday, 25 November 2025 22.01

**5.4 (♂) (w)** Consider a random experiment for which  $\mathcal{S} = \{s_i : s_i = -3, -2, -1, 0, 1, 2, 3\}$  and the outcomes are equally likely. If a random variable is defined as  $X(s_i) = s_i^2$ , find  $S_X$  and the PMF.

## Problem 5.4

Consider a random experiment for which  
 $S = \{s_i : s_i = -3, -2, -1, 0, 1, 2, 3\}$   
and the outcomes are equally likely.  
If a random variable is defined as  $X(s_i) = s_i^2$ , find  $S_X$  and the PMF.

### 1. Understand the sample space

The sample space is

$$S = \{-3, -2, -1, 0, 1, 2, 3\}.$$

There are 7 possible outcomes, and they are **equally likely**, so

$$P(s_i) = \frac{1}{7} \quad \text{for each } s_i \in S.$$

### 2. Apply the random variable $X$

The random variable is defined as  $X(s_i) = s_i^2$ .

So we square each possible outcome:

- $X(-3) = 9$
- $X(-2) = 4$
- $X(-1) = 1$
- $X(0) = 0$
- $X(1) = 1$
- $X(2) = 4$
- $X(3) = 9$

### 3. Find the set $S_X$ (the range of $X$ )

List the **distinct** values that  $X$  can take:

$$S_X = \{0, 1, 4, 9\}.$$

#### 4. Find the PMF of $X$

The PMF is  $p_X(x) = P(X = x)$ .

We group outcomes that give the same value of  $X$ :

- $X = 0$  only when  $s = 0$ .

$$P(X = 0) = P(s = 0) = \frac{1}{7}.$$

- $X = 1$  when  $s = -1$  or  $s = 1$ .

$$P(X = 1) = P(s = -1) + P(s = 1) = \frac{1}{7} + \frac{1}{7} = \frac{2}{7}.$$

- $X = 4$  when  $s = -2$  or  $s = 2$ .

$$P(X = 4) = P(s = -2) + P(s = 2) = \frac{1}{7} + \frac{1}{7} = \frac{2}{7}.$$

- $X = 9$  when  $s = -3$  or  $s = 3$ .

$$P(X = 9) = P(s = -3) + P(s = 3) = \frac{1}{7} + \frac{1}{7} = \frac{2}{7}.$$

Check they sum to 1:

$$\frac{1}{7} + \frac{2}{7} + \frac{2}{7} + \frac{2}{7} = \frac{7}{7} = 1.$$

Final answer:

- $S_X = \{0, 1, 4, 9\}$
- PMF:

$$p_X(x) = \begin{cases} \frac{1}{7}, & x = 0, \\ \frac{2}{7}, & x = 1, \\ \frac{2}{7}, & x = 4, \\ \frac{2}{7}, & x = 9, \\ 0, & \text{otherwise.} \end{cases}$$

5.31 (c) The arrival rate of calls at a mobile switching station is 1 per second.

The probability of  $k$  calls in a  $T$  second interval is given by a Poisson PMF with  $\lambda = \text{arrival rate} \times T$ . What is the probability that there will be more than 100 calls placed in a 1-minute interval?

### Problem 5.31

The arrival rate of calls at a mobile switching station is 1 per second.

The probability of  $k$  calls in a  $T$ -second interval is given by a Poisson PMF with

$$\lambda = (\text{arrival rate}) \times T.$$

What is the probability that there will be **more than 100 calls** placed in a 1-minute interval?

#### 1. Identify the Poisson parameter

- Arrival rate = 1 call per second.
- Time interval = 1 minute = 60 seconds.

So Poisson parameter:

$$\lambda = (\text{arrival rate}) \times T = 1 \times 60 = 60.$$

Let  $N$  be the number of calls in 1 minute. Then

$$N \sim \text{Poisson}(\lambda = 60).$$

The Poisson PMF is:

$$P(N = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

#### 2. We want $P(N > 100)$

This is

$$P(N > 100) = 1 - P(N \leq 100) = 1 - \sum_{k=0}^{100} \frac{e^{-60} 60^k}{k!}.$$

That's the exact expression.

Numerically this probability is **very small**. Evaluating (with a calculator or software) gives approximately

$$P(N > 100) \approx 8.7 \times 10^{-7}.$$

That is

$$P(N > 100) \approx 0.00000087,$$

less than one in a million.

So it's **extremely unlikely** to see more than 100 calls in one minute given an average rate of 1 call per second.

### 3. Pseudocode to compute this on a computer

If you wanted to compute this probability programmatically:

pseudo

 Copy code

```
lambda := 60
sum_cdf := 0

for k from 0 to 100 do
    term := exp(-lambda) * lambda^k / factorial(k)
    sum_cdf := sum_cdf + term
end for

prob_more_than_100 := 1 - sum_cdf
print("P(N > 100) =", prob_more_than_100)
```

Most languages (Python, R, MATLAB, etc.) also have built-in Poisson CDF functions, so you could do something like:

pseudo

 Copy code

```
lambda := 60
prob_more_than_100 := 1 - poisson_cdf(100, lambda)
```

---

If you have more exercises from that sheet, feel free to send them in the same way and I'll walk through those too.

# Continuous Random Variables, PDF and CDF

Tuesday, 25 November 2025 22.05

- 10.2 (..) (w)** The temperature in degrees Fahrenheit is modeled as a uniform random variable with  $T \sim U(20, 60)$ . If  $T$  is rounded off to the nearest  $1/2^\circ$  to form  $\hat{T}$ , what is  $P[\hat{T} = 30^\circ]$ ? What can you say about the use of a PDF versus a PMF to describe the probabilistic outcome of a physical experiment?

## 10.2 – Uniform temperature, rounded to nearest $0.5^\circ$

$T \sim U(20, 60)$  (degrees F).

$T$  is rounded to the nearest  $0.5^\circ$  to form  $\hat{T}$ .

Find  $P[\hat{T} = 30^\circ]$ .

What can you say about using a PDF vs a PMF for a physical experiment?

### Step 1: What values of $T$ round to $30^\circ$ ?

Rounding to the nearest  $0.5^\circ$ :

- Numbers from **29.75 up to 30.25** round to **30**.

(At the endpoints the probability is 0 anyway for a continuous variable, so we don't care if we include/exclude them.)

So we want

$$P(\hat{T} = 30) = P(29.75 \leq T < 30.25).$$

### Step 2: Use the uniform distribution

For  $T \sim U(20, 60)$ , the pdf is

$$f_T(t) = \frac{1}{60 - 20} = \frac{1}{40}, \quad 20 \leq t \leq 60.$$

For any interval  $[a, b] \subset [20, 60]$ ,

$$P(a \leq T \leq b) = \frac{b - a}{60 - 20} = \frac{b - a}{40}.$$

Here the interval length is  $30.25 - 29.75 = 0.5$ , so

$$P(\hat{T} = 30) = P(29.75 \leq T < 30.25) = \frac{0.5}{40} = \frac{1}{80} = 0.0125.$$

Answer:  $P(\hat{T} = 30^\circ) = 1/80$ .

### Step 3: PDF vs PMF for physical measurements

- $T$  is modeled as **continuous**, so any exact value (like  $T = 30$ ) has probability 0. A **PDF** is the right tool for that idealized model.
- Once we **round** to the nearest  $0.5^\circ$ ,  $\hat{T}$  can only take discrete values (20.0, 20.5, ..., 60.0). Now each value has a **positive** probability, so  $\hat{T}$  is described by a **PMF**.
- In real experiments, measurements are always limited by instrument resolution, so what we actually *observe* is discrete. The continuous PDF is a convenient approximation to model the underlying "true" quantity.

- 10.4 (..) (w)** Which of the functions shown in Figure 10.33 are valid PDFs? If a function is not a PDF, why not?

## 10.4 – Which functions are valid PDFs?

"Which of the functions shown in Figure 10.33 are valid PDFs? If a function is not a PDF, why not?"

I don't have access to Figure 10.33 itself, so I can't label specific curves. But here is exactly how you decide for each function in that figure:

A function  $f_X(x)$  is a **valid pdf** iff:

1. **Non-negativity:**

$$f_X(x) \geq 0 \quad \text{for all } x.$$

If the graph dips below the x-axis anywhere, it's **not** a valid pdf.

2. **Total area = 1:**

$$\int_{-\infty}^{\infty} f_X(x) dx = 1.$$

- If the total area is **less than 1**, probabilities would sum to less than 1 → invalid.
- If the total area is **greater than 1**, probabilities would sum to more than 1 → invalid.

Other notes:

- Discontinuities or piecewise shapes are OK as long as those two conditions hold.
- A spike of "infinite height" can be okay if its area is finite (like a delta), but in standard intro problems they usually avoid that.

So, for each function in your figure:

- Check that it never goes below 0.
- Compute (or reason about) its total area.

Those that satisfy both criteria are the valid PDFs; others are not, and the reason will be one (or both) of these conditions failing.

**10.12 (..) (w)** A constant or DC current source that outputs 1 amp is connected to a resistor of nominal resistance of 1 ohm. If the resistance value can vary according to  $R \sim \mathcal{N}(1, 0.1)$ , what is the probability that the voltage across the resistor will be between 0.99 and 1.01 volts?

### 10.12 – Normal resistance, probability voltage in a tight range

A 1-A DC current source is connected to a resistor with nominal resistance 1 Ω.

The resistance varies as  $R \sim \mathcal{N}(1, 0.1)$ .

What is  $P(0.99 \leq V \leq 1.01)$  for the voltage  $V$  across the resistor?

Ohm's law:  $V = IR$ .

Here  $I = 1 \text{ A}$ , so

$$V = 1 \cdot R = R.$$

Thus  $V \sim \mathcal{N}(1, 0.1)$  as well.

Let  $\mu = 1$  and  $\sigma^2 = 0.1$ . Then  $\sigma = \sqrt{0.1}$ .

We want:

$$P(0.99 \leq V \leq 1.01) = P(0.99 \leq R \leq 1.01).$$

**Standardize**

Define

$$Z = \frac{R - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

Then

$$P(0.99 \leq R \leq 1.01) = P\left(\frac{0.99 - 1}{\sigma} \leq Z \leq \frac{1.01 - 1}{\sigma}\right).$$

Compute the bounds:

$$\frac{0.99 - 1}{\sigma} = \frac{-0.01}{\sqrt{0.1}} \approx -0.0316,$$

$$\frac{1.01 - 1}{\sigma} = \frac{+0.01}{\sqrt{0.1}} \approx +0.0316.$$

So

$$P(0.99 \leq V \leq 1.01) = P(-0.0316 \leq Z \leq 0.0316) = \Phi(0.0316) - \Phi(-0.0316) = 2\Phi(0.0316) - 1.$$

Numerically (using a standard normal table or software),

$$\Phi(0.0316) \approx 0.5126 \Rightarrow P \approx 2(0.5126) - 1 = 0.0252.$$

**Answer:**

$P(0.99 \leq V \leq 1.01) \approx 0.0252$  (about 2.5%).

Symbolic form (often how the book wants it):

$$P = 2\Phi\left(\frac{0.01}{\sqrt{0.1}}\right) - 1.$$

**10.21 (f,c)** If  $X \sim \mathcal{N}(0, 1)$ , determine the number of outcomes out of 1000 that you would expect to occur within the interval [1, 2]. Next conduct a computer simulation to carry out this experiment. How many outcomes actually occur within this interval?

### 10.21 – Expected count in [1,2] and a simulation

If  $X \sim \mathcal{N}(0, 1)$ , determine how many outcomes out of 1000 you'd expect in the interval [1, 2].

Then simulate this experiment on a computer; how many actually fall in this interval?

#### Step 1: Probability that one sample is in [1,2]

For standard normal  $Z$ ,

$$P(1 \leq X \leq 2) = \Phi(2) - \Phi(1),$$

where  $\Phi$  is the standard normal CDF.

Using common values:

- $\Phi(1) \approx 0.8413$
- $\Phi(2) \approx 0.9772$

So

$$P(1 \leq X \leq 2) \approx 0.9772 - 0.8413 \approx 0.1359.$$

#### Step 2: Expected count out of 1000

Expected number =  $1000 \times P(1 \leq X \leq 2)$ :

$$E[\text{count}] = 1000 \times 0.1359 \approx 135.9.$$

So you'd expect about 136 samples in the interval [1, 2].

### Step 3: Pseudocode for the simulation

You can simulate 1000 samples from  $N(0, 1)$ , count how many lie in [1,2]:

pseudo

 Copy code

```
N := 1000
count := 0

for i from 1 to N do
    X := normal_random(mean = 0, variance = 1)
    if X >= 1 and X <= 2 then
        count := count + 1
    end if
end for

print("Number in [1,2] =", count)
print("Proportion in [1,2] =", count / N)
```

If you run this many times, the count should hover around 136, sometimes a bit lower, sometimes higher.

**10.22 (S) (w)** If  $X \sim \mathcal{N}(\mu, \sigma^2)$ , find the CDF of  $X$  in terms of  $\Phi(x)$ .

### 10.22 – CDF of $X \sim \mathcal{N}(\mu, \sigma^2)$ in terms of $\Phi$

| If  $X \sim \mathcal{N}(\mu, \sigma^2)$ , find the CDF  $F_X(x)$  in terms of  $\Phi(x)$ .

By definition,

$$F_X(x) = P(X \leq x).$$

Standardize to convert to a standard normal:

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

Then

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

So the CDF is

$$\boxed{F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)}.$$

That's exactly how a general normal CDF relates to the standard normal CDF.

# Transformation of Random Variables

Tuesday, 25 November 2025 22.07

**5.20 (♂) (f)** If  $X \sim \text{Pois}(\lambda)$ , find the PMF for  $Y = 2X$ .

## 5.20 – PMF of $Y = 2X$ when $X \sim \text{Pois}(\lambda)$

We know the PMF of a Poisson random variable:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Define

$$Y = 2X.$$

### 1. What values can $Y$ take?

Since  $X$  is  $0, 1, 2, \dots$ , then

$$Y = 2X \in \{0, 2, 4, 6, \dots\}.$$

So  $Y$  only takes even nonnegative integers.

### 2. PMF of $Y$

To find  $P(Y = y)$ , write it in terms of  $X$ :

- If  $y$  is even, say  $y = 2k$  with  $k = 0, 1, 2, \dots$ ,

$$P(Y = y) = P(2X = 2k) = P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}.$$

- If  $y$  is odd, then there is no integer  $X$  such that  $2X = y$ , so

$$P(Y = y) = 0.$$

We can write the PMF as:

$$p_Y(y) = \begin{cases} \frac{e^{-\lambda} \lambda^{y/2}}{(y/2)!}, & y = 0, 2, 4, \dots, \\ 0, & \text{otherwise.} \end{cases}$$

**10.40 (♂) (f)** Find the PDF of  $Y = X^4 + 1$  if  $X \sim \exp(\lambda)$ .

### 10.40 – PDF of $Y = X^4 + 1$ if $X \sim \exp(\lambda)$

For an exponential with parameter  $\lambda > 0$ ,

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Define

$$Y = X^4 + 1.$$

#### 1. Relationship between $Y$ and $X$

Because  $X \geq 0$ , we have:

- Minimum  $X = 0 \Rightarrow Y = 0^4 + 1 = 1$ .
- As  $X$  increases,  $Y$  increases (function is strictly increasing for  $x > 0$ ).

So  $Y \in [1, \infty)$ , and the transformation is one-to-one.

Inverse transformation:

$$Y = X^4 + 1 \quad \Rightarrow \quad X = (Y - 1)^{1/4}.$$

Derivative:

$$\frac{dX}{dY} = \frac{1}{4}(Y - 1)^{-3/4}.$$

#### 2. Use the change-of-variables formula

For a monotone transformation:

$$f_Y(y) = f_X(x(y)) \left| \frac{dx}{dy} \right|.$$

Plug in  $x = (y - 1)^{1/4}$ :

$$f_Y(y) = \lambda e^{-\lambda(y-1)^{1/4}} \cdot \frac{1}{4}(y - 1)^{-3/4}, \quad y \geq 1.$$

So:

$$f_Y(y) = \frac{\lambda}{4}(y - 1)^{-3/4} \exp[-\lambda(y - 1)^{1/4}], \quad y \geq 1; \quad f_Y(y) = 0 \text{ for } y < 1.$$

**10.41 (w)** Find the constants  $a$  and  $b$  so that  $Y = aX + b$ , where  $X \sim \mathcal{U}(0, 1)$ , yields  $Y \sim \mathcal{U}(2, 6)$ .

### 10.41 – Linear transform of a uniform: $Y = aX + b$

Find  $a$  and  $b$  so that  $X \sim \mathcal{U}(0, 1)$  and  $Y = aX + b$  gives  $Y \sim \mathcal{U}(2, 6)$ .

If  $X \sim \mathcal{U}(0, 1)$  and  $Y = aX + b$  with  $a > 0$ , then:

- When  $X = 0$ ,  $Y = b$  = left endpoint.
- When  $X = 1$ ,  $Y = a + b$  = right endpoint.

We want the support of  $Y$  to be  $(2, 6)$ , so:

$$b = 2, \quad a + b = 6.$$

Then  $a = 6 - b = 6 - 2 = 4$ .

Answer:

$$a = 4, \quad b = 2,$$

so  $Y = 4X + 2 \sim \mathcal{U}(2, 6)$ .

**10.46 (..) (f,c)** If  $Y = X^2$ , where  $X \sim U(0, 1)$ , determine the PDF of  $Y$ . Next perform a computer simulation using the realizations of  $Y$  (obtained as  $y_m = x_m^2$ , where  $x_m$  is the  $m$ th realization of  $X$ ) to estimate the PDF  $p_Y(y)$ . Do your theoretical results match the simulated results?

### 10.46 – PDF of $Y = X^2$ when $X \sim U(0, 1)$ , plus simulation

Given  $X \sim U(0, 1)$ , so

$$f_X(x) = 1, \quad 0 \leq x \leq 1.$$

Define

$$Y = X^2.$$

#### 1. Range and inverse

- For  $0 \leq X \leq 1$ , we get  $0 \leq Y \leq 1$ .
- The transformation is increasing on  $[0, 1]$ .
- Inverse:  $X = \sqrt{Y}$ , for  $0 \leq Y \leq 1$ .

Derivative:

$$\frac{dX}{dY} = \frac{1}{2\sqrt{Y}}.$$

#### 2. PDF of $Y$

Again using the change-of-variables formula:

$$f_Y(y) = f_X(x(y)) \left| \frac{dx}{dy} \right|.$$

Here:

$$f_Y(y) = 1 \cdot \frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}, \quad 0 < y < 1.$$

(We usually just leave it for  $0 < y < 1$ ; the behavior at 0 doesn't change probabilities.)

So

$$\boxed{f_Y(y) = \frac{1}{2\sqrt{y}}, \quad 0 < y < 1; \quad f_Y(y) = 0 \text{ otherwise.}}$$

This has more mass near  $y = 0$  (because small  $x$  values square to even smaller  $y$ ), which matches intuition.

#### 3. Pseudocode for the simulation

The problem also asks for a computer simulation to estimate the PDF and compare.

Basic idea:

1. Generate many samples  $x_m$  from  $U(0, 1)$ .
2. Compute  $y_m = x_m^2$ .
3. Use the  $y_m$  to make a histogram or kernel density estimate.
4. Overlay the theoretical pdf  $1/(2\sqrt{y})$ .

Here's pseudocode:

```

pseudo
Copy code

N := 100000      # number of samples
Y_samples := empty_list

for m from 1 to N do
    X := uniform_random(0, 1)  # sample from U(0,1)
    Y := X^2
    Y_samples.append(Y)

```

```

for m from 1 to N do
    X := uniform_random(0, 1)    # sample from U(0,1)
    Y := X * X                  # Y = X^2
    append Y to Y_samples
end for

# Now estimate the PDF of Y using a histogram
num_bins := 50
histogram := make_histogram(Y_samples, num_bins, range = [0, 1],
                            normalize = true)  # area ≈ 1

plot_histogram(histogram)      # simulated pdf

# On the same plot, draw the theoretical pdf  $f_Y(y) = 1 / (2 \sqrt{y})$ 
for y from small_positive to 1 step small_step do
    f_y := 1 / (2 * sqrt(y))
    plot_point(y, f_y)          # or plot_curve for a smooth line
end for

```

If you do this, the histogram should hug the curve  $1/(2\sqrt{y})$ , especially as  $N$  gets large—showing excellent agreement between theory and simulation.

# Expected value, Variance

Tuesday, 25 November 2025 22.07

**6.2 (c) (f)** For the discrete random variable with PMF

$$p_X[k] = \frac{1}{10} \quad k = 0, 1, \dots, 9$$

find the expected value of  $X$ .

## 6.2 – Expected value of a discrete uniform variable

PMF:

$$p_X[k] = \frac{1}{10}, \quad k = 0, 1, \dots, 9.$$

The expected value is

$$E[X] = \sum_{k=0}^9 k p_X[k] = \sum_{k=0}^9 k \cdot \frac{1}{10} = \frac{1}{10} \sum_{k=0}^9 k.$$

Sum of 0 through 9:

$$\sum_{k=0}^9 k = \frac{9 \cdot 10}{2} = 45.$$

So

$$E[X] = \frac{1}{10} \cdot 45 = 4.5.$$

Answer:  $E[X] = 4.5$ .

**6.13 (c) (t)** Can  $E[X^2]$  ever be equal to  $E^2[X]$ ? If so, when?

## 6.13 – When can $E[X^2] = (E[X])^2$ ?

Recall the variance formula:

$$\text{Var}(X) = E[X^2] - (E[X])^2 \geq 0.$$

Equality  $E[X^2] = (E[X])^2$  happens exactly when

$$\text{Var}(X) = 0.$$

Variance is zero only if  $X$  is constant with probability 1 (no randomness). That is,  $P(X = c) = 1$  for some constant  $c$ .

Answer: Yes, when  $X$  is deterministic (always equal to the same value).

**6.17 (w)** If  $Y = aX + b$ , what is the variance of  $Y$  in terms of the variance of  $X$ ?

### 6.17 – Variance of $Y = aX + b$

We want  $\text{Var}(Y)$  in terms of  $\text{Var}(X)$ .

Mean of  $Y$ :

$$E[Y] = E[aX + b] = aE[X] + b.$$

Then

$$Y - E[Y] = (aX + b) - (aE[X] + b) = a(X - E[X]).$$

Variance:

$$\text{Var}(Y) = E[(Y - E[Y])^2] = E[a^2(X - E[X])^2] = a^2 E[(X - E[X])^2] = a^2 \text{Var}(X).$$

The constant  $b$  does not affect variance.

Answer:  $\text{Var}(Y) = a^2 \text{Var}(X)$ .

**9.10 (..) (w)** The components of a random vector  $\mathbf{X} = [X_1 X_2 \dots X_N]^T$  all have the same mean  $E_X[X]$  and the same variance  $\text{var}(X)$ . The “sample mean” random variable

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

is formed. If the  $X_i$ 's are independent, find the mean and variance of  $\bar{X}$ . What happens to the variance as  $N \rightarrow \infty$ ? Does this tell you anything about the PMF of  $\bar{X}$  as  $N \rightarrow \infty$ ?

### 9.10 – Mean and variance of the sample mean

We have a random vector  $\mathbf{X} = [X_1, \dots, X_N]^T$ , with each  $X_i$  having

- mean  $E[X_i] = E_X[X] = \mu$ ,
- variance  $\text{Var}(X_i) = \text{Var}(X) = \sigma^2$ .

The sample mean is

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i.$$

Assume the  $X_i$  are independent.

#### Mean of $\bar{X}$

$$E[\bar{X}] = E\left[\frac{1}{N} \sum_{i=1}^N X_i\right] = \frac{1}{N} \sum_{i=1}^N E[X_i] = \frac{1}{N} \cdot N\mu = \mu.$$

So the sample mean is **unbiased**.

#### Variance of $\bar{X}$

Independence  $\Rightarrow$  variances add:

$$\text{Var}\left(\sum_{i=1}^N X_i\right) = \sum_{i=1}^N \text{Var}(X_i) = N\sigma^2.$$

Thus

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{N} \sum_{i=1}^N X_i\right) = \frac{1}{N^2} \cdot N\sigma^2 = \frac{\sigma^2}{N}.$$

What happens as  $N \rightarrow \infty$ ?

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{N} \rightarrow 0.$$

So the distribution (pmf/pdf) of  $\bar{X}$  becomes more and more concentrated around  $\mu$ ; in the limit it collapses to a spike at  $\mu$ . This is the idea behind the **law of large numbers**.

**11.10 (c) (w)** The signal-to-noise ratio (SNR) of a random variable quantifies the accuracy of a measurement of a physical quantity. It is defined as  $E^2[X]/\text{var}(X)$  and is seen to increase as the mean, which represents the true value, increases and also as the variance, which represents the power of the measurement error, i.e.,  $X - E[X]$ , decreases. For example, if  $X \sim \mathcal{N}(\mu, \sigma^2)$ , then  $\text{SNR} = \mu^2/\sigma^2$ . Determine the SNR if the measurement is  $X = A + U$ , where  $A$  is the true value and  $U$  is the measurement error with  $U \sim \mathcal{U}(-1/2, 1/2)$ . For an SNR of 1000 what should  $A$  be?

### 11.10 – SNR for $X = A + U$ , with $U \sim U(-1/2, 1/2)$

SNR is defined as

$$\text{SNR} = \frac{E[X]^2}{\text{Var}(X)}.$$

We're told the measurement is

$$X = A + U,$$

where

- $A$  is the true value (a constant),
- $U \sim U(-1/2, 1/2)$  is the error.

#### Mean and variance of $U$

For a uniform on  $[a, b]$ :

- $E[U] = \frac{a+b}{2}$ ,
- $\text{Var}(U) = \frac{(b-a)^2}{12}$ .

Here  $a = -1/2, b = 1/2$ :

$$E[U] = 0, \quad \text{Var}(U) = \frac{(1)^2}{12} = \frac{1}{12}.$$

#### Mean and variance of $X$

$$E[X] = E[A + U] = A + 0 = A,$$

$$\text{Var}(X) = \text{Var}(A + U) = \text{Var}(U) = \frac{1}{12}$$

(because a constant doesn't add variance).

#### SNR

$$\text{SNR} = \frac{E[X]^2}{\text{Var}(X)} = \frac{A^2}{1/12} = 12A^2.$$

For  $\text{SNR} = 1000$ :

$$12A^2 = 1000 \Rightarrow A^2 = \frac{1000}{12} = \frac{250}{3},$$

$$A = \sqrt{\frac{250}{3}} \approx 9.13.$$

Answer:

- $\text{SNR} = 12A^2$ .
- For  $\text{SNR} = 1000$ ,  $A = \sqrt{250/3} \approx 9.13$ .

### 11.23 (f) Determine $E[X^2]$ for a $\mathcal{N}(\mu, \sigma^2)$ random variable.

#### 11.23 – $E[X^2]$ for $X \sim \mathcal{N}(\mu, \sigma^2)$

We know

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

But  $E[X] = \mu$ , so

$$\sigma^2 = E[X^2] - \mu^2 \Rightarrow E[X^2] = \sigma^2 + \mu^2.$$

Answer:  $E[X^2] = \mu^2 + \sigma^2$ .

### 11.24 (f) Determine $E[(2X + 1)^2]$ for a $\mathcal{N}(\mu, \sigma^2)$ random variable.

## 11.24 – $E[(2X + 1)^2]$ for $X \sim \mathcal{N}(\mu, \sigma^2)$

Let

$$Y = 2X + 1.$$

One way is to expand the square; another is to treat  $Y$  as a linear transform of a normal and use mean/variance. I'll show both quickly.

### Method 1: Expand

$$(2X + 1)^2 = 4X^2 + 4X + 1.$$

Take expectation:

$$E[(2X + 1)^2] = 4E[X^2] + 4E[X] + 1.$$

From 11.23:

$$E[X^2] = \mu^2 + \sigma^2, \quad E[X] = \mu.$$

So

$$E[(2X + 1)^2] = 4(\mu^2 + \sigma^2) + 4\mu + 1 = 4\mu^2 + 4\sigma^2 + 4\mu + 1.$$

### Method 2: Use mean and variance of $Y$

Since  $Y = 2X + 1$ ,

- $E[Y] = 2\mu + 1$ ,
- $\text{Var}(Y) = 4\sigma^2$ .

And

$$E[Y^2] = \text{Var}(Y) + (E[Y])^2 = 4\sigma^2 + (2\mu + 1)^2 = 4\sigma^2 + 4\mu^2 + 4\mu + 1,$$

same result.

Answer:

$$E[(2X + 1)^2] = 4\sigma^2 + 4\mu^2 + 4\mu + 1.$$

# Joint distributions, covariance, independent random variables

Tuesday, 25 November 2025 22.07

**7.10 (..) (f)** The sample space for a random vector is  $S_{X,Y} = \{(i,j) : i = 1, 2, 3, 4, 5; j = 1, 2, 3, 4\}$ . If the outcomes are equally likely, find  $P[(X, Y) \in A]$ , where  $A = \{(i,j) : 1 \leq i \leq 2; 3 \leq j \leq 4\}$ .

## 7.10

Sample space:  $S_{X,Y} = \{(i,j) : i = 1, 2, 3, 4, 5; j = 1, 2, 3, 4\}$ .

Outcomes equally likely.

Find  $P[(X, Y) \in A]$  where

$A = \{(i,j) : 1 \leq i \leq 2; 3 \leq j \leq 4\}$ .

Total number of outcomes in the sample space:

- $i$  has 5 possibilities (1–5),
- $j$  has 4 possibilities (1–4),

so  $|S_{X,Y}| = 5 \times 4 = 20$ .

Now count the outcomes in  $A$ :

- $i$  can be 1 or 2 → 2 choices,
- $j$  can be 3 or 4 → 2 choices,

so  $|A| = 2 \times 2 = 4$ .

All outcomes are equally likely, so

$$P[(X, Y) \in A] = \frac{|A|}{|S_{X,Y}|} = \frac{4}{20} = \frac{1}{5}.$$

**7.19 (..) (w)** A joint PMF is given by

$$p_{X,Y}[i,j] = \begin{cases} \frac{1}{4} & (i,j) = (0,0) \\ \frac{1}{4} & (i,j) = (1,1) \\ \frac{1}{4} & (i,j) = (1,0) \\ \frac{1}{4} & (i,j) = (1,-1) \end{cases}$$

Are  $X$  and  $Y$  independent?

### 7.19

Joint PMF

$$p_{X,Y}[i,j] = \frac{1}{4} \text{ for } (i,j) \in \{(0,0), (1,1), (1,0), (1,-1)\},$$

and 0 otherwise.

Are  $X$  and  $Y$  independent?

First find the marginals.

Marginal of  $X$ :

- $P(X = 0) = P(0,0) = 1/4$ .
- $P(X = 1) = P(1,1) + P(1,0) + P(1,-1) = 1/4 + 1/4 + 1/4 = 3/4$ .

Marginal of  $Y$ :

- $P(Y = 1) = P(1,1) = 1/4$ .
- $P(Y = 0) = P(0,0) + P(1,0) = 1/4 + 1/4 = 1/2$ .
- $P(Y = -1) = P(1,-1) = 1/4$ .

For independence we'd need

$$p_{X,Y}(i,j) = p_X(i)p_Y(j) \text{ for every pair.}$$

Check one pair, say  $(0,0)$ :

- Actual joint:  $p_{X,Y}(0,0) = 1/4$ .
- Product of marginals:  $p_X(0)p_Y(0) = (1/4)(1/2) = 1/8$ .

Since  $1/4 \neq 1/8$ , the factorization fails.

Conclusion:  $X$  and  $Y$  are not independent.

**7.27 (c) (f)** If  $X \sim \text{Ber}(1/2)$ ,  $Y \sim \text{Ber}(1/2)$ , and  $X$  and  $Y$  are independent, find the PMF for  $Z = X + Y$ . Why does the width of the PMF increase? Does the variance increase?

### 7.27

$X \sim \text{Ber}(1/2)$ ,  $Y \sim \text{Ber}(1/2)$ , independent.

Find the PMF of  $Z = X + Y$ .

Why does the width increase? Does the variance increase?

Possible pairs  $(X, Y)$  and the resulting  $Z$ :

- $(0,0) \rightarrow Z = 0$ , prob  $1/2 \cdot 1/2 = 1/4$ .
- $(1,0) \rightarrow Z = 1$ , prob  $1/4$ .
- $(0,1) \rightarrow Z = 1$ , prob  $1/4$ .
- $(1,1) \rightarrow Z = 2$ , prob  $1/4$ .

So PMF of  $Z$ :

$$P(Z = 0) = 1/4, \quad P(Z = 1) = 1/4 + 1/4 = 1/2, \quad P(Z = 2) = 1/4.$$

Width:

- $X$  (or  $Y$ ) takes values  $\{0, 1\} \rightarrow$  width 1.
- $Z$  takes values  $\{0, 1, 2\} \rightarrow$  width 2.

Adding two variables extends the range because you can get both lower and higher totals.

Variance:

Each Bernoulli( $1/2$ ) has variance

$$\text{Var}(X) = p(1 - p) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

For independent  $X, Y$ ,

$$\text{Var}(Z) = \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

So the variance does increase (from  $1/4$  to  $1/2$ ).

**7.33 (c) (f)** Find the covariance for the joint PMF given in Table 7.5.

### 7.33

"Find the covariance for the joint PMF given in Table 7.5."

I don't have Table 7.5, so I can't give the numerical value. But here's exactly how you'd compute it from any joint PMF  $p_{X,Y}(i, j)$ :

**1.** Compute

$$E[X] = \sum_i \sum_j i p_{X,Y}(i, j), \quad E[Y] = \sum_i \sum_j j p_{X,Y}(i, j).$$

**2.** Compute

$$E[XY] = \sum_i \sum_j ij p_{X,Y}(i, j).$$

**3.** Covariance:

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y].$$

If you send a snapshot of Table 7.5 I can plug in the numbers and finish the arithmetic.

#### ✓ Exercise 7.33 — Find the covariance for Table 7.5

The joint PMF from Table 7.5 is:

$X = i$

$Y = 0$

$Y = 1$

0

3/8

1/8

1

1/8

3/8

### ✓ Exercise 7.33 — Find the covariance for Table 7.5

The joint PMF from Table 7.5 is:

$X = i$	$Y = 0$	$Y = 1$
0	3/8	1/8
1	1/8	3/8

Marginals:

- $p_X(0) = 1/2, p_X(1) = 1/2$
- $p_Y(0) = 1/2, p_Y(1) = 1/2$

We need:

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y].$$

#### Step 1 — Compute $E[X]$

$$E[X] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

#### Step 2 — Compute $E[Y]$

$$E[Y] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

#### Step 3 — Compute $E[XY]$

Use the joint PMF:

$$E[XY] = \sum_{i=0}^1 \sum_{j=0}^1 ij p_{X,Y}(i,j).$$

Compute term-by-term:

- $(i, j) = (0, 0): 0 \cdot 0 \cdot \frac{3}{8} = 0$
- $(0, 1): 0 \cdot 1 \cdot \frac{1}{8} = 0$
- $(1, 0): 1 \cdot 0 \cdot \frac{1}{8} = 0$
- $(1, 1): 1 \cdot 1 \cdot \frac{3}{8} = \frac{3}{8}$

So:

$$E[XY] = \frac{3}{8}.$$

#### Step 4 — Combine to find covariance

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{3}{8} - \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{3}{8} - \frac{1}{4}.$$

Compute:

$$\frac{3}{8} - \frac{1}{4} = \frac{3}{8} - \frac{2}{8} = \frac{1}{8}.$$

#### ★ Final Answer

$$\boxed{\text{cov}(X, Y) = \frac{1}{8}}$$

The covariance is **positive**, meaning that  $X$  and  $Y$  tend to increase together (which makes sense because the high-probability entries are  $(0,0)$  and  $(1,1)$ ).

**7.34 (t)** Prove the following properties of the covariance:

$$\begin{aligned}
 \text{cov}(X, Y) &= E_{X,Y}[XY] - E_X[X]E_Y[Y] \\
 \text{cov}(X, X) &= \text{var}(X) \\
 \text{cov}(Y, X) &= \text{cov}(X, Y) \\
 \text{cov}(cX, Y) &= c[\text{cov}(X, Y)] \\
 \text{cov}(X, cY) &= c[\text{cov}(X, Y)] \\
 \text{cov}(X, X + Y) &= \text{cov}(X, X) + \text{cov}(X, Y) \\
 \text{cov}(X + Y, X) &= \text{cov}(X, X) + \text{cov}(Y, X)
 \end{aligned}$$

for  $c$  a constant.

### 7.34 – Covariance properties

Recall the definition:

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

(1)  $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$

Expand the product:

$$(X - E[X])(Y - E[Y]) = XY - XE[Y] - E[X]Y + E[X]E[Y].$$

Take expectation and use linearity:

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y].$$

(2)  $\text{cov}(X, X) = \text{var}(X)$

$$\text{cov}(X, X) = E[(X - E[X])^2] = \text{Var}(X).$$

(3)  $\text{cov}(Y, X) = \text{cov}(X, Y)$

From the symmetric definition:

$$\text{cov}(Y, X) = E[(Y - E[Y])(X - E[X])] = E[(X - E[X])(Y - E[Y])] = \text{cov}(X, Y).$$

(4)  $\text{cov}(cX, Y) = c \text{ cov}(X, Y)$

Compute:

$$\text{cov}(cX, Y) = E[(cX - E[cX])(Y - E[Y])] = E[(cX - cE[X])(Y - E[Y])] = cE[(X - E[X])(Y - E[Y])] = c \text{ cov}(X, Y).$$

(5)  $\text{cov}(X, cY) = c \text{ cov}(X, Y)$

Same logic:

$$\text{cov}(X, cY) = E[(X - E[X])(cY - cE[Y])] = cE[(X - E[X])(Y - E[Y])] = c \text{ cov}(X, Y).$$

$$(6) \text{cov}(X, X + Y) = \text{cov}(X, X) + \text{cov}(X, Y)$$

Using linearity in the second slot:

$$\text{cov}(X, X + Y) = E[(X - E[X])(((X + Y) - E[X + Y]))].$$

But  $E[X + Y] = E[X] + E[Y]$ , so

$$(X + Y) - E[X + Y] = (X - E[X]) + (Y - E[Y]).$$

Then

$$\text{cov}(X, X + Y) = E[(X - E[X])(X - E[X])] + E[(X - E[X])(Y - E[Y])] = \text{cov}(X, X) + \text{cov}(X, Y).$$

$$(7) \text{cov}(X + Y, X) = \text{cov}(X, X) + \text{cov}(Y, X)$$

Same idea, now linear in the first slot:

$$\begin{aligned} \text{cov}(X + Y, X) &= E[((X + Y) - E[X + Y])(X - E[X])] \\ &= E[(X - E[X])(X - E[X])] + E[(Y - E[Y])(X - E[X])] = \text{cov}(X, X) + \text{cov}(Y, X). \end{aligned}$$

**12.2 (c)** Reproduce Figure 12.2a by letting  $X \sim \mathcal{U}(-1, 1)$  and  $Y \sim \mathcal{U}(-1, 1)$ , where  $X$  and  $Y$  are independent. Omit any realizations of  $(X, Y)$  for which  $\sqrt{X^2 + Y^2} > 1$ . Explain why this produces a uniform distribution of points in the unit circle. See also Problem 13.23 for a more formal justification of this procedure.

### 12.2(c) – Simulating a uniform distribution in the unit circle

Let  $X \sim U(-1, 1)$  and  $Y \sim U(-1, 1)$ , independent.

Omit any realization with  $\sqrt{X^2 + Y^2} > 1$ .

Explain why the remaining points are uniform in the unit circle, and give pseudocode to reproduce the figure.

The joint density of  $(X, Y)$  over the square  $[-1, 1] \times [-1, 1]$  is constant:

$$f_{X,Y}(x, y) = \frac{1}{4}, \quad -1 \leq x \leq 1, \quad -1 \leq y \leq 1.$$

When we **reject** points outside the unit circle ( $x^2 + y^2 > 1$ ), the conditional distribution of the accepted points (those with  $x^2 + y^2 \leq 1$ ) has density proportional to the original one **restricted** to the disk. Since the original density is constant, the restricted one is also constant  $\rightarrow$  **uniform over the disk**.

#### Pseudocode

pseudo Copy code

```

N := 10000           # number of accepted points desired
points := empty_list

while length(points) < N do
    X := uniform_random(-1, 1)
    Y := uniform_random(-1, 1)

    if X*X + Y*Y <= 1 then
        append (X, Y) to points
    end if
end while

plot_points(points)      # scatter plot approximates uniform disk

```

**12.56 (c)** Simulate on the computer realizations of the random vector

$$\begin{bmatrix} W \\ Z \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}\right).$$

Plot these realizations as well as the contours of constant PDF on the same graph.

### 12.56(c) – Simulate correlated Gaussian vector and plot with contours

We want to simulate

$$\begin{bmatrix} W \\ Z \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}\right).$$

#### Constructing samples

One way: use independent standard normals and a Cholesky factor.

Let

$$\Sigma = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}.$$

A convenient lower-triangular  $L$  with  $\Sigma = LL^T$  is

$$L = \begin{bmatrix} 1 & 0 \\ -0.9 & \sqrt{1 - 0.9^2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -0.9 & \sqrt{0.19} \end{bmatrix}.$$

If  $U, V$  are independent  $N(0, 1)$ , then

$$\begin{bmatrix} W \\ Z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + L \begin{bmatrix} U \\ V \end{bmatrix}$$

has the desired distribution.

## Pseudocode

pseudo

 Copy code

```
N := 10000

muW := 1
muZ := 1
L11 := 1.0
L21 := -0.9
L22 := sqrt(0.19)

W_samples := empty_list
Z_samples := empty_list

for k from 1 to N do
    U := normal_random(0, 1)    # independent N(0,1)
    V := normal_random(0, 1)

    W := muW + L11 * U          # 1*U
    Z := muZ + L21 * U + L22 * V # -0.9*U + sqrt(0.19)*V + 1

    append W to W_samples
    append Z to Z_samples
end for

# Scatter plot of samples
plot_points(W_samples, Z_samples)

# Plot contours of the bivariate normal PDF on same axes
# (usually using a grid and evaluating the bivariate normal formula)
draw_bivariate_normal_contours(mean = (1,1),
                                covariance = [[1, -0.9], [-0.9, 1]])
```

The scatter cloud will be elongated and tilted, matching the contours of the correlated Gaussian.

# Conditional Probability Mass Functions and Expectation

Tuesday, 25 November 2025 22.46

**8.2 (c) (w)** A fair coin is tossed with the outcome mapped into  $X = 1$  for a head and  $X = 0$  for a tail. If it comes up heads, then a fair die is tossed. The outcome of the die is denoted by  $Y$  and is set equal to the number of dots observed. If the coin comes up tails, then we set  $Y = 0$ . Find the conditional PMF  $p_{Y|X}[j|i]$  for all values of  $i$  and  $j$ . Next, compute  $P[Y = 1]$ .

## 8.2

Fair coin  $\rightarrow X = 1$  for head,  $X = 0$  for tail.

If head, toss a fair die and set  $Y$  = dots (1–6).

If tail, set  $Y = 0$ .

(a) Find  $p_{Y|X}[j|i]$  for all  $i, j$ .

(b) Compute  $P(Y = 1)$ .

(a) Conditional PMF  $p_{Y|X}(j|i)$

- Case  $X = 1$  (coin = head):

We toss a fair die, so

$$P(Y = j | X = 1) = \begin{cases} 1/6, & j = 1, 2, 3, 4, 5, 6, \\ 0, & \text{otherwise.} \end{cases}$$

- Case  $X = 0$  (coin = tail):

We set  $Y = 0$  deterministically, so

$$P(Y = j | X = 0) = \begin{cases} 1, & j = 0, \\ 0, & j \neq 0. \end{cases}$$

(b)  $P(Y = 1)$

Use the law of total probability:

$$P(Y = 1) = P(Y = 1 | X = 1)P(X = 1) + P(Y = 1 | X = 0)P(X = 0).$$

The coin is fair, so  $P(X = 1) = P(X = 0) = 1/2$ . Also  $P(Y = 1 | X = 1) = 1/6$ ,  $P(Y = 1 | X = 0) = 0$ . Hence

$$P(Y = 1) = \frac{1}{6} \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{12}.$$

**8.6 (c) (f)** If

$$p_{X,Y}[i,j] = \begin{cases} \frac{1}{6} & i = 0, j = 0 \\ \frac{1}{3} & i = 0, j = 1 \\ \frac{1}{3} & i = 1, j = 0 \\ \frac{1}{6} & i = 1, j = 1 \end{cases}$$

and  $p_{Y|X}$  and  $p_{X|Y}$ .

## 8.6

Joint PMF:

$$p_{X,Y}(i,j) = \begin{cases} \frac{1}{6}, & (i,j) = (0,0), \\ \frac{1}{3}, & (0,1), \\ \frac{1}{3}, & (1,0), \\ \frac{1}{6}, & (1,1), \\ 0, & \text{otherwise.} \end{cases}$$

Find  $p_{Y|X}$  and  $p_{X|Y}$ .

### Marginals

$$p_X(0) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}, \quad p_X(1) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.$$

$$p_Y(0) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}, \quad p_Y(1) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.$$

### Conditional $p_{Y|X}(j | i)$

- For  $i = 0$ :

$$P(Y = 0 | X = 0) = \frac{p(0,0)}{p_X(0)} = \frac{1/6}{1/2} = \frac{1}{3},$$

$$P(Y = 1 | X = 0) = \frac{p(0,1)}{p_X(0)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

- For  $i = 1$ :

$$P(Y = 0 | X = 1) = \frac{p(1,0)}{p_X(1)} = \frac{1/3}{1/2} = \frac{2}{3},$$

$$P(Y = 1 | X = 1) = \frac{p(1,1)}{p_X(1)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

So

$$p_{Y|X}(j | i) = \begin{array}{c|cc} & j = 0 & j = 1 \\ \hline i = 0 & 1/3 & 2/3 \\ i = 1 & 2/3 & 1/3 \end{array}$$

### Conditional $p_{X|Y}(i | j)$

By symmetry you get the same pattern:

- For  $j = 0$ :

$$P(X = 0 | Y = 0) = \frac{1/6}{1/2} = 1/3, \quad P(X = 1 | Y = 0) = \frac{1/3}{1/2} = 2/3.$$

- For  $j = 1$ :

$$P(X = 0 | Y = 1) = 2/3, \quad P(X = 1 | Y = 1) = 1/3.$$

So

$$p_{X|Y}(i | j) = \begin{array}{c|cc} & j = 0 & j = 1 \\ \hline i = 0 & 1/3 & 2/3 \\ i = 1 & 2/3 & 1/3 \end{array}$$

**8.13 (c) (w)** A random vector  $(X, Y)$  has a sample space shown in Figure 8.13 with the sample points depicted as solid circles. The four points are equally probable. Note that the points in Figure 8.13b are the corners of the square shown in Figure 8.13a after rotation by  $+45^\circ$ . For both cases compute  $p_{Y|X}$  and  $p_Y$  to determine if the random variables are independent.

### 8.13 (conceptual)

A random vector  $(X, Y)$  has four equally probable points shown in Fig. 8.13.

(a) 8.13a: points are corners of an axis-aligned square.

(b) 8.13b: those same corners after a  $+45^\circ$  rotation.

For each case, compute  $p_{Y|X}$  and  $p_Y$  and decide if  $X$  and  $Y$  are independent.

We don't need exact numerical coordinates; only how many points share each  $X$  and  $Y$  value.

#### Case (a): axis-aligned square

Think of the four equally likely points as

$$(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2),$$

each with probability  $1/4$ .

- Possible  $X$ -values:  $x_1, x_2$  with probabilities  $1/2$  each.
- Possible  $Y$ -values:  $y_1, y_2$  with probabilities  $1/2$  each.

Joint probabilities:

$$P(X = x_i, Y = y_j) = 1/4 = (1/2)(1/2) = P(X = x_i)P(Y = y_j)$$

for all  $i, j$ . Thus

- $p_Y(y_1) = p_Y(y_2) = 1/2$ .
- $p_{Y|X}(y_j | x_i) = \frac{1/4}{1/2} = 1/2$  for all  $i, j$ .

Since  $p_{X,Y} = p_X p_Y$ ,  $X$  and  $Y$  are independent in case (a).

#### Case (b): rotated by $45^\circ$

After rotation, the four points are no longer of the form

$\{x_1, x_2\} \times \{y_1, y_2\}$ . A typical situation (e.g. starting from corners  $(\pm 1, \pm 1)$ ) gives points like

$$(\sqrt{2}, 0), (0, \sqrt{2}), (-\sqrt{2}, 0), (0, -\sqrt{2}),$$

each with probability  $1/4$ .

- $X$ -values:  $-\sqrt{2}, 0, \sqrt{2}$  with probabilities  $1/4, 1/2, 1/4$ .
- $Y$ -values:  $-\sqrt{2}, 0, \sqrt{2}$  with the same probabilities.

Conditionals (example):

- Given  $X = -\sqrt{2}$ , the only point is  $(-\sqrt{2}, 0)$ , so  $P(Y = 0 | X = -\sqrt{2}) = 1$ .
- But  $P(Y = 0) = 1/2$ .

Since  $P(Y = 0 | X = -\sqrt{2}) \neq P(Y = 0)$ ,  $X$  and  $Y$  are not independent in case (b). The conditional PMF depends on the value of  $X$ .

**8.21 (c) (f)** Find the conditional mean of  $Y$  given  $X$  if the joint PMF is uniformly distributed over the points  $\mathcal{S}_{X,Y} = \{(0,0), (1,0), (1,1), (2,0), (2,1), (2,2)\}$ .

**8.21**

Joint PMF is uniform over

$$S_{X,Y} = \{(0,0), (1,0), (1,1), (2,0), (2,1), (2,2)\}.$$

Find the conditional mean  $E[Y | X]$ .

Each of the 6 points has probability 1/6.

**Step 1: Marginal of  $X$** 

- $X = 0$ : only  $(0,0) \rightarrow P(X = 0) = 1/6$ .
- $X = 1$ :  $(1,0), (1,1) \rightarrow P(X = 1) = 2/6 = 1/3$ .
- $X = 2$ :  $(2,0), (2,1), (2,2) \rightarrow P(X = 2) = 3/6 = 1/2$ .

**Step 2: Conditional distribution of  $Y$  given  $X = x$** 

- Given  $X = 0$ : only  $Y = 0$ .  
So  $P(Y = 0 | X = 0) = 1$ .
- Given  $X = 1$ :  $Y = 0$  or 1.  
Joint probs:  $P(1,0) = P(1,1) = 1/6$ .

Conditional:

$$P(Y = 0 | X = 1) = \frac{1/6}{1/3} = \frac{1}{2}, \quad P(Y = 1 | X = 1) = \frac{1/6}{1/3} = \frac{1}{2}.$$

- Given  $X = 2$ :  $Y = 0, 1, 2$ .

Each joint prob is 1/6, so conditionally

$$P(Y = 0 | X = 2) = P(Y = 1 | X = 2) = P(Y = 2 | X = 2) = \frac{1/6}{1/2} = \frac{1}{3}.$$

**Step 3: Conditional mean  $E[Y | X = x]$** 

- $E[Y | X = 0] = 0$ .
- $E[Y | X = 1] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$ .
- $E[Y | X = 2] = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} = 1$ .

So the conditional mean function is

$$E[Y | X] = \begin{cases} 0, & X = 0, \\ \frac{1}{2}, & X = 1, \\ 1, & X = 2. \end{cases}$$

(Notice this is  $E[Y | X = x] = x/2$  for the values  $x = 0, 1, 2$ .)

## M-Dimensional (discrete) Random Vectors

Tuesday, 25 November 2025 22.47

**9.12 (f)** If the covariance matrix of a  $3 \times 1$  random vector  $\mathbf{X}$  is

$$\mathbf{C}_X = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 2 \\ 1 & 2 & 4 \end{bmatrix}$$

find the correlation coefficients  $\rho_{X_1, X_2}$ ,  $\rho_{X_1, X_3}$ , and  $\rho_{X_2, X_3}$ .

9.12

Covariance matrix of  $X = [X_1 \ X_2 \ X_3]^T$  is

$$C_X = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 2 \\ 1 & 2 & 4 \end{bmatrix}.$$

Find  $\rho_{X_1, X_2}$ ,  $\rho_{X_1, X_3}$ ,  $\rho_{X_2, X_3}$ .

Recall

$$\rho_{X_i, X_j} = \frac{\text{cov}(X_i, X_j)}{\sqrt{\text{var}(X_i)} \sqrt{\text{var}(X_j)}}.$$

From  $C_X$ :

- $\text{var}(X_1) = 1 \rightarrow \sigma_1 = 1$
- $\text{var}(X_2) = 2 \rightarrow \sigma_2 = \sqrt{2}$
- $\text{var}(X_3) = 4 \rightarrow \sigma_3 = 2$

Covariances:

- $\text{cov}(X_1, X_2) = 0$
- $\text{cov}(X_1, X_3) = 1$
- $\text{cov}(X_2, X_3) = 2$

So

$$\rho_{X_1, X_2} = \frac{0}{1 \cdot \sqrt{2}} = 0,$$

$$\rho_{X_1, X_3} = \frac{1}{1 \cdot 2} = \frac{1}{2},$$

$$\rho_{X_2, X_3} = \frac{2}{\sqrt{2} \cdot 2} = \frac{1}{\sqrt{2}}.$$

**9.13 (c) (w)** A  $2 \times 1$  random vector is given by

$$\mathbf{X} = \begin{bmatrix} U \\ 2U \end{bmatrix}$$

where  $\text{var}(U) = 1$ . Find the covariance matrix for  $\mathbf{X}$ . Next find the correlation coefficient  $\rho_{X_1, X_2}$ . Finally, compute the determinant of the covariance matrix. Is the covariance matrix positive definite? Hint: A positive definite matrix must have a positive determinant.

**9.13**

$$X = \begin{bmatrix} U \\ 2U \end{bmatrix} \text{ with } \text{var}(U) = 1.$$

- (a) Find covariance matrix of  $X$ .
- (b) Find  $\rho_{X_1, X_2}$ .
- (c) Compute determinant of the covariance matrix and decide if it is positive definite.

Let  $X_1 = U$ ,  $X_2 = 2U$ .

- $\text{var}(X_1) = \text{var}(U) = 1$ .
- $\text{var}(X_2) = \text{var}(2U) = 4 \text{ var}(U) = 4$ .
- $\text{cov}(X_1, X_2) = \text{cov}(U, 2U) = 2 \text{ var}(U) = 2$ .

So

$$C_X = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

Correlation:

$$\rho_{X_1, X_2} = \frac{\text{cov}(X_1, X_2)}{\sigma_1 \sigma_2} = \frac{2}{1 \cdot 2} = 1.$$

They are perfectly positively correlated (in fact  $X_2 = 2X_1$ ).

Determinant:

$$\det(C_X) = 1 \cdot 4 - 2 \cdot 2 = 0.$$

A covariance matrix must be **positive semidefinite**; positive definite would require strictly positive determinant and no linear dependence. Since the determinant is 0 and  $X_2 = 2X_1$ , the matrix is *not* positive definite, only semidefinite.

**9.17 (c) (w)** Which of the following matrices are not valid covariance matrices and why?

- a.  $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$
- b.  $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$
- c.  $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$
- d.  $\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$

**9.17**

Which of these are **not** valid covariance matrices, and why?

- a.  $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$
- b.  $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$
- c.  $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$
- d.  $\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$

Requirements for a covariance matrix:

1. Symmetric.
2. Diagonal entries (variances)  $\geq 0$ .
3. Positive semidefinite (for  $2 \times 2$  this essentially means determinant  $\geq 0$  and diagonals  $\geq 0$ ).

Check each:

(a)  $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

Symmetric, diagonals positive, but determinant

$$\det = 1 \cdot 1 - 2 \cdot 2 = -3 < 0.$$

Negative determinant  $\Rightarrow$  not positive semidefinite  $\Rightarrow$  **not valid**.

(b)  $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$

Symmetric, determinant  $= 1 > 0$ , but diagonals are negative, implying negative variances – impossible. So **not valid**.

(c)  $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

Symmetric, diagonals positive, determinant  $= 4 - 1 = 3 > 0$ . This is positive definite, so **valid**.

(d)  $\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$

Not symmetric ( $\text{cov}(X_1, X_2) = 1$  but  $\text{cov}(X_2, X_1) = 0$ ). Covariance matrices must be symmetric. So **not valid**.

## 9.19 (f) If a random vector $\mathbf{X}$ is transformed according to

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_1 + X_2 \end{aligned}$$

and the mean of  $\mathbf{X}$  is

$$E_{\mathbf{X}}[\mathbf{X}] = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

find the mean of  $\mathbf{Y} = [Y_1 \ Y_2]^T$ .

**9.19**

$$Y_1 = X_1, \quad Y_2 = X_1 + X_2.$$

$$E_X[X] = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

Find the mean of  $\mathbf{Y} = [Y_1 \ Y_2]^T$ .

Write the transform as  $\mathbf{Y} = A\mathbf{X}$  with

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Then

$$E[\mathbf{Y}] = A E[\mathbf{X}].$$

Compute:

$$E[\mathbf{Y}] = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \cdot 3 + 0 \cdot 4 \\ 1 \cdot 3 + 1 \cdot 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

So  $E[Y_1] = 3$ ,  $E[Y_2] = 7$ .

**9.20 (c) (f)** If the random vector  $\mathbf{X}$  given in Problem 9.19 has a covariance matrix

$$\mathbf{C}_X = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

find the covariance matrix for  $\mathbf{Y} = [Y_1 \ Y_2]^T$ .

**9.20**

Same setup as 9.19, and covariance of  $X$  is

$$\mathbf{C}_X = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Find covariance of  $\mathbf{Y}$ .

For a linear transform  $\mathbf{Y} = A\mathbf{X}$ ,

$$\mathbf{C}_Y = A\mathbf{C}_X A^T.$$

$$\text{We already have } A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

First compute  $A\mathbf{C}_X$ :

$$A\mathbf{C}_X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 3 \end{bmatrix}.$$

$$\text{Now multiply by } A^T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}:$$

$$\mathbf{C}_Y = (A\mathbf{C}_X)A^T = \begin{bmatrix} 2 & 1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix}.$$

So

$$\mathbf{C}_Y = \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix}.$$

**9.29 (f)** For the random walk described in Example 9.5 find the mean and the variance of  $X_n$  as a function of  $n$  if  $p = 3/4$ . What do they indicate about the probable outcomes of  $X_1, X_2, \dots, X_N$ ?

**9.29**

Random walk of Example 9.5. Find mean and variance of  $X_n$  for  $p = 3/4$ . What do they say about probable outcomes?

In the usual 1-D random walk model:

- At each step  $k$ , you move  $+1$  with probability  $p$  and  $-1$  with probability  $1 - p$ .
- Let the step  $R_k$  be  $+1$  or  $-1$ , and

$$X_n = \sum_{k=1}^n R_k,$$

with  $X_0 = 0$ .

For one step:

$$E[R_k] = (+1)p + (-1)(1 - p) = 2p - 1,$$

$$R_k^2 = 1 \Rightarrow E[R_k^2] = 1,$$

$$\text{Var}(R_k) = E[R_k^2] - (E[R_k])^2 = 1 - (2p - 1)^2 = 4p(1 - p).$$

Steps are independent, so

$$E[X_n] = n E[R_1] = n(2p - 1),$$

$$\text{Var}(X_n) = n \text{Var}(R_1) = 4np(1 - p).$$

Now plug in  $p = 3/4$ :

- $2p - 1 = 2 \cdot \frac{3}{4} - 1 = \frac{1}{2}$ .
- $4p(1 - p) = 4 \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{4}$ .

So

$$E[X_n] = \frac{n}{2}, \quad \text{Var}(X_n) = \frac{3}{4}n.$$

Interpretation:

- The **mean position** grows linearly: the walker drifts upward like  $n/2$ .
- The **standard deviation** grows like  $\sqrt{\frac{3}{4}n} \propto \sqrt{n}$ , much slower than the mean.

So for large  $n$ , most paths will be somewhere around  $n/2$  with fluctuations of order  $\sqrt{n}$ ; i.e., strongly biased to the positive side.

**9.30 (c)** For the random walk of Problem 9.29 simulate several realizations of the random vector  $\mathbf{X} = [X_1 \ X_2 \ \dots \ X_N]^T$  and plot these as  $x_n$  versus  $n$  for  $n = 1, 2, \dots, N = 50$ . Does the appearance of the outcomes corroborate your results in Problem 9.29? Also, compare your results to those shown in Figure 9.3b.

### 9.30 (simulation of the random walk)

Simulate several realizations of the random walk from 9.29 up to  $N = 50$ . Plot  $x_n$  vs  $n$ . Does the appearance match the mean/variance behavior?

Here's pseudocode to simulate  $M$  trajectories:

```
pseudo Copy code
N := 50           # number of steps
p := 0.75         # probability of +1 step
M := 10          # number of sample paths to plot

for m from 1 to M do
    X := 0
    path := empty_list
    append X to path  # X_0 = 0

    for n from 1 to N do
        U := uniform_random(0, 1)
        if U < p then
            step := +1
        else
            step := -1
        end if

        X := X + step
        append X to path
    end for

    plot_line(0..N, path, label="path m") # overlay all paths
end for

add_title("Random walk paths, p = 0.75")
add_xlabel("n")
add_ylabel("X_n")
```

On the plot, each path will roughly climb upward with slope about 0.5, with random wiggles around that trend. The spread between paths should increase like  $\sqrt{n}$ , matching the variance formula from 9.29.

**9.34 (t)** In this problem we explain how to generate a computer realization of a random vector with a given covariance matrix. This procedure was used to produce the realizations shown in Figure 9.4a. For simplicity the desired  $N \times 1$  random vector  $\mathbf{X}$  is assumed to have a zero mean vector. The procedure is to first generate an  $N \times 1$  random vector  $\mathbf{U}$  whose elements are zero mean, uncorrelated random variables with unit variances so that its covariance matrix is  $\mathbf{I}$ . Then transform  $\mathbf{U}$  according to  $\mathbf{X} = \mathbf{B}\mathbf{U}$ , where  $\mathbf{B}$  is an appropriate  $N \times N$  matrix. The matrix  $\mathbf{B}$  is obtained from the  $N \times N$  matrix  $\sqrt{\Lambda}$  whose elements are obtained from the eigenvalue matrix  $\Lambda$  of  $\mathbf{C}_X$  by taking the square root of the elements of  $\Lambda$ , and  $\mathbf{V}$ , where  $\mathbf{V}$  is the eigenvector matrix of  $\mathbf{C}_X$ , to form  $\mathbf{B} = \mathbf{V}\sqrt{\Lambda}$ . Prove that the covariance matrix of  $\mathbf{B}\mathbf{U}$  will be  $\mathbf{C}_X$ .

### 9.34 (theoretical – generating correlated vectors)

We have a desired zero-mean random vector  $X$  with covariance  $C_X$ .

Procedure:

1. "Generate  $U$  with covariance  $I$  (zero mean, uncorrelated, unit variance components)."
2. "Let  $X = BU$ , where  $B = V\sqrt{\Lambda}$ .  
Here  $C_X = V\Lambda V^T$  is the eigen-decomposition,  
and  $\sqrt{\Lambda}$  has the square roots of the eigenvalues on the diagonal.  
Prove that the covariance of  $BU$  is  $C_X$ ."

We want to show:

$$\text{Cov}(X) = C_X.$$

Given:

- $X = BU$ ,
- $E[U] = 0$ ,
- $\text{Cov}(U) = E[UU^T] = I$ ,
- $B = V\sqrt{\Lambda}$ ,
- $C_X = V\Lambda V^T$ .

Because  $E[X] = 0$ , covariance is simply

$$\text{Cov}(X) = E[XX^T].$$

Compute:

$$\text{Cov}(X) = E[(BU)(BU)^T] = E[BUU^TB^T] = B E[UU^T] B^T = BIB^T = BB^T.$$

Now plug in  $B = V\sqrt{\Lambda}$ :

$$BB^T = (V\sqrt{\Lambda})(V\sqrt{\Lambda})^T = V\sqrt{\Lambda}\sqrt{\Lambda}^T V^T.$$

Since  $\Lambda$  is diagonal with non-negative entries,  $\sqrt{\Lambda}\sqrt{\Lambda}^T = \Lambda$ . Hence

$$BB^T = V\Lambda V^T = C_X.$$

Therefore  $\text{Cov}(X) = C_X$ , as required.

# Session 2

Wednesday, 26 November 2025 10.13

# Mean and Covariance

Wednesday, 26 November 2025 10.04

## 1.2 Means and Covariances

The first exercise is for you to practice computation with means and covariances. Consider two random vectors:

$$\mathbf{X} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}\right) \quad \text{and} \quad \mathbf{Y} \sim \mathcal{N}\left(\begin{bmatrix} -3 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\right) \quad (1)$$

Assume that  $\mathbf{X}$  and  $\mathbf{Y}$  are independent and compute the following entities:

$$E[10\mathbf{X}], \quad E[\mathbf{X} + \mathbf{Y}], \quad E[\mathbf{X}^T \mathbf{Y}], \quad E[\mathbf{X} \mathbf{Y}^T],$$

$$E\left[\begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix} \mathbf{X}\right], \quad \text{Cov}\left[\begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix} \mathbf{X}\right] \quad \text{Cov}(\mathbf{X} + \mathbf{Y})$$

*Hint:* Keep your derivation in vector/matrix form for as long as you possible can. This saves a lot of writing, time, and potential errors. In fact the above exercise can be done without splitting vectors into its entries at all!

### 1.2 – Means and Covariances

We're given two independent 2D Gaussian random vectors

$$\mathbf{X} \sim \mathcal{N}\left(\mu_X = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \Sigma_X = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}\right), \quad \mathbf{Y} \sim \mathcal{N}\left(\mu_Y = \begin{bmatrix} -3 \\ 5 \end{bmatrix}, \Sigma_Y = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\right).$$

Use linearity of expectation and rules:

- $E(AX) = A E(X)$
- $\text{Cov}(AX) = A \Sigma_X A^T$
- If  $X, Y$  independent:  $E(XY^T) = E(X) E(Y)^T$ ,  
and  $\text{Cov}(X + Y) = \Sigma_X + \Sigma_Y$ .

1)  $E[10X]$

$$E[10X] = 10 \mu_X = 10 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 10 \\ 20 \end{bmatrix}.$$

2)  $E[X + Y]$

$$E[X + Y] = \mu_X + \mu_Y = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} -2 \\ 7 \end{bmatrix}.$$

$$3) E[X^T Y]$$

This is a scalar. For independent vectors:

$$E[X^T Y] = E[X]^T E[Y] = \mu_X^T \mu_Y.$$

Compute:

$$\mu_X^T \mu_Y = [1 \ 2] \begin{bmatrix} -3 \\ 5 \end{bmatrix} = 1 \cdot (-3) + 2 \cdot 5 = -3 + 10 = 7.$$

So  $E[X^T Y] = 7$ .

$$4) E[XY^T]$$

This is a  $2 \times 2$  matrix. For independent vectors:

$$E[XY^T] = \mu_X \mu_Y^T.$$

$$\mu_X \mu_Y^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} [-3 \ 5] = \begin{bmatrix} 1 \cdot (-3) & 1 \cdot 5 \\ 2 \cdot (-3) & 2 \cdot 5 \end{bmatrix} = \begin{bmatrix} -3 & 5 \\ -6 & 10 \end{bmatrix}.$$

$$5) E\left[\begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix} X\right]$$

$$\text{Let } A = \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix}.$$

$$E[AX] = A \mu_X = \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \cdot 1 + 2 \cdot 2 \\ 0 \cdot 1 + 1 \cdot 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}.$$

$$6) \text{Cov}\left(\begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix} X\right)$$

$$\text{Cov}(AX) = A \Sigma_X A^T.$$

First  $A \Sigma_X$ :

$$A \Sigma_X = \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 8 & 6 \\ 1 & 2 \end{bmatrix}.$$

Now multiply by  $A^T$ :

$$C = (A \Sigma_X) A^T = \begin{bmatrix} 8 & 6 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 28 & 6 \\ 6 & 2 \end{bmatrix}.$$

So

$$\text{Cov}\left(\begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix} X\right) = \begin{bmatrix} 28 & 6 \\ 6 & 2 \end{bmatrix}.$$

$$7) \text{Cov}(X + Y)$$

Since  $X, Y$  are independent:

$$\text{Cov}(X + Y)$$

Since  $X, Y$  are independent:

$$\text{Cov}(X + Y) = \Sigma_X + \Sigma_Y = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}.$$

## 1.5 An Average is not the Mean!

Let  $X_1, \dots, X_N$  be independent Gaussian random variables with zero mean and variance one and define the average

$$S_N = \frac{1}{N} \sum_{n=1}^N X_n$$

The average  $S_N$  can be used as an estimate of the mean of a random variable.

- Simulate  $S_N$  for  $N = 1, \dots, 1000$  and plot the result as a function of  $N$ . Notice that this can be done in a simple manner using the Matlab function `mean` in a `for` loop. Repeat it multiple times (for instance 10) and plot all the results in the same figure.
- Calculate –with pen and paper!– the mean and variance of the random variable  $S_N$  (recall that  $\text{Var}(V + W) = \text{Var}(V) + \text{Var}(W)$  for independent  $V$  and  $W$ ). How do these depend on the number of samples  $N$ ?
- Discuss the result. Do you think that the command name `mean` is appropriate?

We have i.i.d. Gaussian r.v.'s

$$X_1, \dots, X_N \sim \mathcal{N}(0, 1),$$

and define their average

$$S_N = \frac{1}{N} \sum_{n=1}^N X_n.$$

### 1) Mean and variance of $S_N$ (pen & paper)

#### Mean

Use linearity of expectation:

$$E[S_N] = E\left[\frac{1}{N} \sum_{n=1}^N X_n\right] = \frac{1}{N} \sum_{n=1}^N E[X_n] = \frac{1}{N} \cdot N \cdot 0 = 0.$$

So the **expected value of the average** is the true mean (0).

#### Variance

Because the  $X_n$  are independent and identically distributed with  $\text{Var}(X_n) = 1$ :

$$\text{Var}(S_N) = \text{Var}\left(\frac{1}{N} \sum_{n=1}^N X_n\right) = \frac{1}{N^2} \text{Var}\left(\sum_{n=1}^N X_n\right).$$

For independent variables, variances add:

$$\text{Var}\left(\sum_{n=1}^N X_n\right) = \sum_{n=1}^N \text{Var}(X_n) = N \cdot 1 = N.$$

Thus

$$\text{Var}(S_N) = \frac{1}{N^2} \cdot N = \frac{1}{N}.$$

So:

- $E[S_N] = 0$
- $\text{Var}(S_N) = \frac{1}{N}$ .

As  $N$  increases, the variance (and therefore the typical fluctuation size) shrinks like  $1/N$ , and the standard deviation shrinks like  $1/\sqrt{N}$ . So the averages become less noisy as we use more samples.

## 2) Simulation pseudocode

We want to simulate one realization of  $S_N$  for  $N = 1, \dots, 1000$ , and then repeat this experiment several times (say 10) and plot all curves.

Here's language-agnostic pseudocode:

```
pseudo  
  
maxN := 1000  
numRuns := 10  
  
for run from 1 to numRuns do  
    S := 0                      # running sum  
    curve := empty_list  
  
    for N from 1 to maxN do  
        X := normal_random(mean=0, std=1)  # one  $N(0,1)$  sample  
        S := S + X                      # update sum  
        S_N := S / N                    # current average  
        append S_N to curve  
    end for  
  
    plot_line(1..maxN, curve, label="run " + run)  
end for  
  
add_horizontal_line(y=0)          # true mean  
add_xlabel("N")  
add_ylabel("Sample average S_N")  
add_title("Realizations of sample mean S_N vs N")  
show_plot()
```

What you'd see:

- Each curve starts noisy for small  $N$ , maybe far from 0.
- As  $N$  grows, all curves wobble closer and closer to 0, and their spread decreases – that's the law of large numbers in action.

### 3) Discussion – “An Average is not the Mean!”

Key points:

- For each fixed  $N$ ,  $S_N$  is a **random variable**. A particular computed value is **one realization** of that random variable, not the mathematical mean itself.
- On average,  $S_N$  is correct:  $E[S_N] = 0$ ; we say it's an **unbiased estimator** of the mean.
- But because  $\text{Var}(S_N) = 1/N$ , any *single* sample mean can deviate from the true mean, especially for small  $N$ .
- As  $N \rightarrow \infty$ , the variance tends to 0, and by the **law of large numbers**,  $S_N$  converges to the true mean with high probability.

So: **an average (sample mean)** is a **random estimate of the mean**, not the mean itself.

---

### 4) Is the command name `mean` appropriate?

Conceptually, the function `mean` in MATLAB (or NumPy, etc.) computes the **sample mean**, i.e. one realization of  $S_N$ . The true mean is a property of the underlying distribution and is not known in practice.

- From a strict probabilistic viewpoint, `sampleMean` might have been a more precise name.
- In practice, “`mean`” is widely understood to mean “the average of these data”, so the name is acceptable — as long as you remember it returns a *random quantity* when the data are random.

So the moral is: *what `mean` returns is an estimator of the theoretical mean, not the theoretical mean itself.*

# Helpful matlab commands

Wednesday, 26 November 2025 10.11

## 1.3 Ten useful Matlab Commands for Monte Carlo Simulation

Familiarize yourself with the following Matlab commands:

```
help      rand      randn     mean      var  
plot     histogram  stem      scatter   cdfplot
```

Read the help and try the commands out. Explain what the commands do and how they work to your colleagues. These are just a few of the relevant commands, but you will get far with these!

### 1.3 – Ten useful MATLAB commands (conceptual)

Very briefly what each does:

- **help** – shows documentation for a command, e.g. `help rand`.
- **rand** – generates uniform random numbers in (0,1).
- **randn** – generates standard normal (Gaussian) random numbers with mean 0, variance 1.
- **mean** – computes sample mean of a vector/matrix.
- **var** – computes sample variance of data.
- **plot** – makes a 2D line plot.
- **histogram** – plots a histogram (empirical distribution) of data.
- **stem** – plots discrete data with “sticks” (good for PMFs).
- **scatter** – scatterplot of x-y points.
- **cdfplot** – plot empirical CDF of data.

## 1.4 Random Numbers?

Restart Matlab (really, close it and open it again). Run `randn(1)` and write down the result. Then restart Matlab and run `randn(1)` once more. Compare and discuss the two results. When could the observed behavior be wanted or unwanted? Read about “seeding” the random number generators in the documentation.

## 1.4 – Random Numbers? (seeding idea)

If you:

1. Close MATLAB, reopen.
2. Run `randn(1)` and note the value.
3. Close MATLAB again, reopen.
4. Run `randn(1)` again,

you often get **the same value** both times. That's because MATLAB's random number generator starts from a default **seed** when you start MATLAB.

- This behavior is **wanted** when you want reproducible experiments: same code → same "random" results.
- It's **unwanted** if you expect different randomness each time; then you can reseed with something like the current time.

In general: "seeding" means setting the starting state of the pseudorandom generator.

# Law of Large Numbers

Wednesday, 26 November 2025 10.13

## 1.6 Law of Large Numbers — Monte Carlo Simulation

The law of large numbers tell us that if an expectation exists, we can approximate this by averaging a large number of independent samples<sup>1</sup>:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N g(X_n) = E[g(X)]$$

This can be used to approximate expected values by computer simulation considering a finite  $N$ . This approach is called Monte Carlo simulation. Monte Carlo simulation is particularly relevant if expectation is difficult or impossible to carry out. To see how this works out, we test the idea using a toy example, where we actually can compute the expectation by hand and compare this to the simulation.

- Let  $X \sim U(-1, 1)$  and compute the expectation  $E[X^3]$  by direct integration.
- Now make a Monte Carlo simulation by drawing independent realizations of  $X^3$  and taking the appropriate average. How large should  $N$  be to achieve a sufficient accuracy?
- Try to make a similar simulation for another function  $g$  for which you cannot compute the mean directly. Try also to redo the experiment with a different distribution for  $X$ .
- The term “Monte Carlo simulation” is used for a broad range of methods, including the type of method described above. Wikipedia has a long and informative article about the topic — check this article out!

### 1.6 – Law of Large Numbers / Monte Carlo

#### (a) Analytical $E[X^3]$ for $X \sim U(-1, 1)$

Density:

$$f_X(x) = \frac{1}{2}, \quad -1 < x < 1.$$

$$E[X^3] = \int_{-1}^1 x^3 f_X(x) dx = \frac{1}{2} \int_{-1}^1 x^3 dx = \frac{1}{2} \left[ \frac{x^4}{4} \right]_{-1}^1 = \frac{1}{2} \cdot \frac{1^4 - (-1)^4}{4} = 0.$$

So the true expectation is 0.

### (b) Monte Carlo simulation (pseudocode)

We approximate  $E[X^3]$  by averaging many samples:

pseudo

 Copy code

```
N := 100000          # number of samples
sum := 0

for k from 1 to N do
    U := uniform_random(-1, 1)  # sample from U(-1,1)
    sum := sum + U^3
end for

estimate := sum / N
print("Monte Carlo estimate of E[X^3] =", estimate)
```

As  $N$  gets large, this estimate will get closer to 0.

The typical error shrinks like  $1/\sqrt{N}$ , so:

- $N = 1,000 \rightarrow$  rough estimate
- $N = 100,000 \rightarrow$  much more accurate.

You can repeat this with another function  $g(X)$  and/or a different distribution for  $X$  in exactly the same way: just replace `U^3` with `g(U)` and change how `U` is sampled.

# Simulation of dice

Wednesday, 26 November 2025 10.14

## 1.7 Simulation of a Dice

Write a program to simulate a fair dice. This can be done in one line of code using the `rand` function!

### 1.7 – Simulation of a Dice

We want integers 1–6 with equal probability. Using a `uniform(0,1)`:

pseudo

 Copy code

```
U := uniform_random(0, 1)      # in (0,1)
die := floor(6 * U) + 1         # result is 1,2,...,6
```

That's the logic behind a one-line MATLAB solution like `die = randi(6)` or `die = floor(6*rand)+1`.

# Joint, Conditional, Marginal PDS

Wednesday, 26 November 2025 10.15

## 1.8 Joint, marginal and conditional pdf

In this exercise, we highlight the relation between the joint, marginal and conditional pdfs. We consider three scalar random variables:

$$X \sim \mathcal{U}(0, 3), \quad Y \sim \mathcal{U}(0, 1), \quad \text{and} \quad Z = X + Y$$

where  $X$  and  $Y$  are independent.

- Draw a large number of joint realizations of  $X$ ,  $Y$  and  $Z$ . Investigate scatter-plots of the pairs  $[X, Y]$  and  $[Z, X]$ .
- Write up and sketch the following pdfs:  $p_X$ ,  $p_Y$ ,  $p_{X,Y}$ , and  $p_{X,Z}$ .
- Write up the conditional pdfs  $p_{Y|X}$  and  $p_{X|Y}$  discuss their form.
- Derive the conditional pdf  $p_{Z|X}$ . The easiest way to do this is to consider the sketch the joint pdf  $p_{X,Z}$  and then consider the graphical interpretation of the conditional random variable  $Z|X = x$ , namely that we are “given a particular value  $x$  of  $X$ ”.
- Compute the conditional expectation  $E[Z|X]$  (it is best to do this graphically first and then to formalize). This is an intuitive guess (or estimate) of the value of  $Z$  based upon an observation of  $X$ .
- Now repeat the previous two questions with  $p_{X|Z}$ . Note that the form the conditional pdf is in this case more complicated, however, the conditional expectation can be obtained graphically by inspection of your drawing of the joint pdf.

## 1.8 – Joint, marginal and conditional pdf

Given independent scalar random variables:

- $X \sim U(0, 3)$   
 $\Rightarrow f_X(x) = \frac{1}{3}, 0 < x < 3.$
- $Y \sim U(0, 1)$   
 $\Rightarrow f_Y(y) = 1, 0 < y < 1.$
- $Z = X + Y.$

Simulation pseudocode (first bullet)

pseudo

 Copy code

```
N := 100000
Xs := empty_list
Ys := empty_list
Zs := empty_list

for k from 1 to N do
    X := uniform_random(0, 3)
    Y := uniform_random(0, 1)
    Z := X + Y

    append X to Xs
    append Y to Ys
    append Z to Zs
end for

scatter_plot(Xs, Ys)           # for pair (X,Y)
scatter_plot(Zs, Xs)           # for pair (Z,X)
```

- $(X, Y)$  scatter: uniform cloud over the rectangle  $(0, 3) \times (0, 1)$ .
- $(Z, X)$  scatter: points lie in a tilted strip since  $Z = X + Y$  with  $Y \in (0, 1)$ .

PDFs  $f_X, f_Y, f_{X,Y}, f_{X,Z}$

We already have:

$$f_X(x) = \begin{cases} \frac{1}{3}, & 0 < x < 3, \\ 0, & \text{otherwise,} \end{cases} \quad f_Y(y) = \begin{cases} 1, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Independence  $\Rightarrow$  joint:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) = \begin{cases} \frac{1}{3}, & 0 < x < 3, 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

To get  $f_{X,Z}$ , use the transformation  $(X, Y) \mapsto (X, Z)$  with  $Z = X + Y$ , so

$Y = Z - X$ . Jacobian = 1.

$$f_{X,Z}(x, z) = f_{X,Y}(x, z - x) = \begin{cases} \frac{1}{3}, & 0 < x < 3, 0 < z - x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

The condition  $0 < z - x < 1$  means  $x < z < x + 1$ . So on the  $(x,z)$ -plane,  $f_{X,Z} = 1/3$  in the slanted strip defined by  $0 < x < 3$  and  $x < z < x + 1$ .

## Conditional pdfs $f_{Y|X}$ and $f_{X|Y}$

By independence:

$$f_{Y|X}(y | x) = f_Y(y), \quad f_{X|Y}(x | y) = f_X(x)$$

on their supports.

So:

$$f_{Y|X}(y | x) = \begin{cases} 1, & 0 < y < 1, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$f_{X|Y}(x | y) = \begin{cases} \frac{1}{3}, & 0 < x < 3, \\ 0, & \text{otherwise.} \end{cases}$$

They do **not** depend on the conditioning variable at all (hallmark of independence).

## Conditional pdf $f_{Z|X}$

$$f_{Z|X}(z | x) = \frac{f_{X,Z}(x, z)}{f_X(x)}.$$

For  $0 < x < 3$ ,  $f_X(x) = 1/3$ , and  $f_{X,Z}(x, z) = 1/3$  when  $x < z < x + 1$ . Thus:

$$f_{Z|X}(z | x) = \begin{cases} 1, & x < z < x + 1, \\ 0, & \text{otherwise.} \end{cases}$$

So conditioned on  $X = x$ ,  $Z$  is uniform on the interval  $(x, x + 1)$ .

Conditional expectation:

Mean of a uniform on  $(x, x + 1)$  is the midpoint:

$$E[Z | X = x] = x + \frac{1}{2}.$$

## Marginal pdf of $Z$

We'll need this to talk about  $f_{X|Z}$ .

$$f_Z(z) = \int f_{X,Z}(x, z) dx = \frac{1}{3} \text{ length of the x-interval where } 0 < x < 3, 0 < z - x < 1.$$

Equivalently,  $x$  must satisfy

$$0 < x < 3, \quad z - 1 < x < z.$$

So

$$x \in [\max(0, z - 1), \min(3, z)].$$

The length of this interval gives:

- For  $0 < z < 1$ : interval  $(0, z)$ , length  $z$ .
- For  $1 \leq z \leq 3$ : interval  $(z - 1, z)$ , length 1.
- For  $3 < z < 4$ : interval  $(z - 1, 3)$ , length  $4 - z$ .

Thus

$$f_Z(z) = \begin{cases} \frac{z}{3}, & 0 < z < 1, \\ \frac{1}{3}, & 1 \leq z \leq 3, \\ \frac{4-z}{3}, & 3 < z < 4, \\ 0, & \text{otherwise.} \end{cases}$$

This is a "trapezoid / triangular" shape: up, flat, down.

## Conditional pdf $f_{X|Z}$ and $E[X | Z = z]$

$$f_{X|Z}(x | z) = \frac{f_{X,Z}(x, z)}{f_Z(z)}.$$

Recall  $f_{X,Z}(x, z) = 1/3$  when  $x$  is in the valid interval and 0 otherwise. Therefore, for each fixed  $z$ ,  $X|Z = z$  is uniform on that interval:

- If  $0 < z < 1$ :  $x \in (0, z) \Rightarrow$  uniform on  $(0, z)$ .
- If  $1 \leq z \leq 3$ :  $x \in (z - 1, z) \Rightarrow$  uniform on  $(z - 1, z)$ .
- If  $3 < z < 4$ :  $x \in (z - 1, 3) \Rightarrow$  uniform on  $(z - 1, 3)$ .

So the conditional expectation is the midpoint of each interval:

$$E[X | Z = z] = \begin{cases} \frac{z}{2}, & 0 < z < 1, \\ z - \frac{1}{2}, & 1 \leq z \leq 3, \\ 1 + \frac{z}{2}, & 3 < z < 4. \end{cases}$$

Graphically this is a piecewise-linear function of  $z$ .

# Session 3

Wednesday, 26 November 2025 16.27

# Stochastic processes

Wednesday, 26 November 2025 16.27

**16.1 (..) (w)** Describe a random process that you are likely to encounter in the following situations:

- listening to the daily weather forecast
- paying the monthly telephone bill
- leaving for work in the morning

Why is each process a random one?

16.1 — Describe a random process in each situation

(a) Listening to the daily weather forecast

Let

$$X[n] = \text{forecasted temperature on day } n.$$

Why random?

The weather is uncertain and varies day to day.

(b) Paying the monthly telephone bill

Let

$$X[n] = \text{telephone bill for month } n.$$

Why random?

Your phone usage changes unpredictably.

(c) Leaving for work in the morning

Let

$$X[n] = \text{departure time on day } n.$$

Why random?

You may leave earlier or later depending on traffic, meetings, mood, etc.

**16.2 (w)** A single die is tossed repeatedly. What are  $S$  and  $S_X$ ? Also, can you determine the joint PMF for any  $N$  sample times?

16.2 — Repeatedly tossing a single die

- Sample space of each toss:

$$S = \{1, 2, 3, 4, 5, 6\}.$$

- Sample space of the random process:

$$S_X = S^\infty = \text{all infinite sequences of die values.}$$

- Joint PMF for  $N$  samples, using independence:

$$p_{X(1), \dots, X(N)}(x_1, \dots, x_N) = \left(\frac{1}{6}\right)^N.$$

for any  $x_i \in \{1, \dots, 6\}$ .

**16.4 (c) (w)** For a Bernoulli random process determine the probability that we will observe an alternating sequence of 1's and 0's for the first 100 samples with the first sample being a 1. What is the probability that we will observe an alternating sequence of 1's and 0's for all  $n$ ?

16.4 — Alternating Bernoulli sequence

Assume

$$P(U[n] = 1) = p, \quad P(U[n] = 0) = 1 - p.$$

(a) Probability of observing alternating sequence for first 100 samples, starting with 1

The required sequence is:

$$1, 0, 1, 0, 1, 0, \dots$$

Probability:

$$P = p(1-p)p(1-p)\cdots$$

There are 50 ones and 50 zeros:

$$P = p^{50}(1-p)^{50}$$

(b) Probability that alternation continues for all  $n$

Infinite product:

- If  $0 < p < 1$ , then

$$p^\infty(1-p)^\infty = 0.$$

So:

$$P = 0$$

There is zero probability an infinite random sequence alternates forever.

**16.7 (c) (f)** A *biased* random walk process is defined as  $X[n] = \sum_{i=0}^n U[i]$ , where  $U[i]$  is a Bernoulli random process with

$$p_U[k] = \begin{cases} \frac{1}{4} & k = -1 \\ \frac{3}{4} & k = 1. \end{cases}$$

What is  $E[X[n]]$  and  $\text{var}(X[n])$  as a function of  $n$ ? Next, simulate on a computer a realization of this random process. What happens as  $n \rightarrow \infty$  and why?

## 16.7 — Biased random walk

Given:

$$X[n] = \sum_{i=0}^n U[i]$$

where

$$P(U = 1) = \frac{3}{4}, \quad P(U = -1) = \frac{1}{4}.$$

### Mean

$$E[U] = 1 \cdot \frac{3}{4} + (-1) \cdot \frac{1}{4} = \frac{1}{2}.$$

$$E[X[n]] = (n+1)E[U] = \left\lceil \frac{n+1}{2} \right\rceil.$$

### Variance

$$\text{Var}(U) = E[U^2] - E[U]^2$$

$$E[U^2] = 1^2 \cdot \frac{3}{4} + (-1)^2 \cdot \frac{1}{4} = 1.$$

$$\text{Var}(U) = 1 - \frac{1}{4} = \frac{3}{4}.$$

Because  $U[i]$  are independent:

$$\text{Var}(X[n]) = (n+1)\text{Var}(U) = \left\lceil \frac{3(n+1)}{4} \right\rceil.$$

### Simulation pseudo-code

java

Copy code

```
for n = 0 to N:  
    generate U[n] = +1 w.p. 0.75, -1 w.p. 0.25  
    X[n] = X[n-1] + U[n]  
plot(X)
```

As  $n \rightarrow \infty$ :

- Mean  $\rightarrow +\infty$  (positive drift)
- Variance  $\rightarrow +\infty$
- Trajectory tends to drift upward.

**16.14 (c)** Plot 50 realizations of a WGN random process  $X[n]$  with  $\sigma^2 = 1$  for  $n = 0, 1, \dots, 49$  using a scatter diagram (see Figure 16.15 for an example). Use the MATLAB commands `plot(x,y, '.')` and `hold on` to plot each realization as dots and to overlay the realizations on the same graph, respectively. For a fixed  $n$  can you explain the observed distribution of the dots?

### 16.14 — Scatter plot of 50 realizations of WGN

For WGN with variance 1:

$$X[n] \sim \mathcal{N}(0, 1).$$

#### Expected scatter behavior

For each fixed  $n$ :

- Realizations  $X_k[n]$  are i.i.d. Gaussian.
- Plot of 50 points will form a vertical cloud centered at 0 with spread  $\approx \pm 3$ .

#### Pseudo-code

vbnnet

Copy code

```
for k = 1 to 50:
    for n = 0 to 49:
        X[k,n] = randn()
    plot(n, X[k,n], '.')
    hold on
```

**16.15 (f)** Prove that

$$\frac{1}{(2\pi)^{N/2} \det^{1/2}(\mathbf{C})} \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}\right)$$

where  $\mathbf{x} = [x_1 \ x_2 \dots \ x_N]^T$  and  $\mathbf{C} = \sigma^2 \mathbf{I}$  for  $\mathbf{I}$  an  $N \times N$  identity matrix, reduces to (16.4).

### 16.15 — Prove multivariate Gaussian reduces to product of independent PDFs

Given:

$$C = \sigma^2 I$$

$$C^{-1} = \frac{1}{\sigma^2} I, \quad \det(C) = (\sigma^2)^N.$$

Plug into general formula:

$$\frac{1}{(2\pi)^{N/2} (\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \mathbf{x}^T \mathbf{x}\right).$$

But

$$\mathbf{x}^T \mathbf{x} = \sum_{i=1}^N x_i^2.$$

Thus

$$\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x_i^2/(2\sigma^2)},$$

which is the product of independent univariate Gaussians.

**16.16 (c) (f)** A “white” uniform random process is defined to be an IID random process with  $X[n] \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$  for all  $n$ . Determine the mean and covariance sequences for this random process and compare them to those of the WGN random process. Explain your results.

16.16 “White” uniform random process

### 16.16 — “White” uniform random process

Given:

$$X[n] \sim U(-\sqrt{3}, \sqrt{3})$$

This has:

- mean 0
- variance 1 (since  $\frac{(b-a)^2}{12} = \frac{(2\sqrt{3})^2}{12} = 1$ )

#### Mean sequence

$$m_X[n] = 0.$$

#### Covariance sequence

IID  $\rightarrow 0$  for  $k \neq 0$ .

$$R_X[k] = \begin{cases} 1, & k = 0 \\ 0 & k \neq 0 \end{cases}$$

#### Comparison to WGN

Both are:

- white
- zero-mean
- variance 1

But distributions differ:

- WGN  $\rightarrow$  Gaussian
- This one  $\rightarrow$  Uniform

**16.24 (c) (w)** A random process is IID with samples  $X[n] \sim \mathcal{N}(\mu, 1)$ . It is desired to remove the mean of the random process by forming the new random process  $Y[n] = X[n] - X[n-1]$ . First determine the mean sequence of  $Y[n]$ . Next find  $\text{cov}(Y[0], Y[1])$ . Is  $Y[n]$  an IID random process with a zero mean sequence?

### 16.24 — Removing mean via first difference

Given IID:

$$X[n] \sim \mathcal{N}(\mu, 1)$$

Define:

$$Y[n] = X[n] - X[n-1].$$

#### Mean sequence

$$E[Y[n]] = \mu - \mu = 0.$$

#### Covariance

Compute:

$$Y[0] = X[0] - X[-1]$$

$$Y[1] = X[1] - X[0]$$

Assuming independence:

$$\text{cov}(Y[0], Y[1]) = \text{cov}(X[0] - X[-1], X[1] - X[0])$$

Only term with shared variable is  $X[0]$ :

$$\text{cov}(Y[0], Y[1]) = -\text{Var}(X[0]) = -1.$$

Only term with shared variable is  $X[0]$ :

$$\text{cov}(Y[0], Y[1]) = -\text{Var}(X[0]) = -1.$$

Is  $Y[n]$  IID?

No — they are correlated:

$$\text{cov}(Y[n], Y[n+1]) = -1 \neq 0.$$

**16.26 (c) (f)** If a sum random process is defined as  $X[n] = \sum_{i=0}^n U[i]$  for  $n \geq 0$ , where  $E[U[i]] = 0$  and  $\text{var}(U[i]) = \sigma_U^2$  for  $i \geq 0$  and the  $U[i]$  are IID, find the mean and covariance sequences of  $X[n]$ .

### 16.26 — Sum random process

Given:

$$X[n] = \sum_{i=0}^n U[i], \quad E[U] = 0, \quad \text{Var}(U) = \sigma_U^2$$

Mean

$$E[X[n]] = 0.$$

Covariance

For  $k \geq 0$ :

$$R_X[k] = E[X[n]X[n+k]].$$

Since  $X[n]$  contains  $n+1$  terms of variance  $\sigma_U^2$ :

$$\text{Var}(X[n]) = (n+1)\sigma_U^2.$$

For  $k > 0$ :

$$R_X[k] = (n+1)\sigma_U^2.$$

Thus:

$$R_X[k] = \sigma_U^2(n+1)$$

This process is non-stationary.

**16.30 (f)** A random process is defined as  $X[n] = As[n]$  for all  $n$ , where  $A \sim \mathcal{N}(0, 1)$  and  $s[n]$  is a deterministic signal. Find the mean and covariance sequences.

### 16.30 — Process $X[n] = As[n]$

Given:

- $A \sim \mathcal{N}(0, 1)$
- $s[n]$  deterministic

#### Mean

$$E[X[n]] = E[As[n]] = 0.$$

#### Covariance

$$\text{cov}(X[n], X[m]) = E[A^2 s[n] s[m]] = 1 \cdot s[n] s[m].$$

Thus:

$$R_X[n, m] = s[n] s[m]$$

**16.31 (..) (f)** A random process is defined as  $X[n] = AU[n]$  for all  $n$ , where  $A \sim \mathcal{N}(0, \sigma_A^2)$  and  $U[n]$  is WGN with variance  $\sigma_U^2$ , and  $A$  is independent of  $U[n]$  for all  $n$ . Find the mean and covariance sequences. What type of random process is  $X[n]$ ?

#### 16.31

$X[n] = AU[n]$ ,  $A \sim \mathcal{N}(0, \sigma_A^2)$ ,  $U[n]$  is WGN with variance  $\sigma_U^2$ , and  $A$  is independent of all  $U[n]$ .

#### Mean sequence

$$m_X[n] = E[X[n]] = E[AU[n]] = E[A] E[U[n]] = 0 \cdot 0 = 0.$$

(Usual WGN is zero-mean.)

#### Covariance (ACS)

$$\begin{aligned} R_X[n, m] &= E[X[n]X[m]] \\ &= E[A^2 U[n] U[m]] = E[A^2] E[U[n] U[m]] \end{aligned}$$

(independence of  $A$  and  $U$ ).

For WGN with variance  $\sigma_U^2$ ,

$$E[U[n] U[m]] = \begin{cases} \sigma_U^2, & n = m, \\ 0, & n \neq m. \end{cases}$$

Also  $E[A^2] = \text{Var}(A) = \sigma_A^2$ .

So with lag  $k = m - n$ :

$$R_X[k] = \begin{cases} \sigma_A^2 \sigma_U^2, & k = 0, \\ 0, & k \neq 0. \end{cases}$$

#### Type of process

- Mean doesn't depend on  $n$ .
- ACS depends only on lag  $k$ .

So  $X[n]$  is **WSS** and **white**.

It is **not Gaussian** (a product of Gaussians is not Gaussian) and samples are not independent (they all share the same random scale  $A$ ), only uncorrelated.

# Session 5

Wednesday, 26 November 2025 17.34

# WSS Random Processes

Wednesday, 26 November 2025 17.34

**17.1 (..) (w)** A Bernoulli random process  $X[n]$  for  $-\infty < n < \infty$  consists of independent random variables with each random variable taking on the values  $+1$  and  $-1$  with probabilities  $p$  and  $1-p$ , respectively. Is this random process WSS? If it is WSS, find its mean sequence and autocorrelation sequence.

## 17.1 — Bernoulli $\pm 1$ process

Each  $X[n] \in \{+1, -1\}$  with

$$P(X[n] = 1) = p, \quad P(X[n] = -1) = 1 - p,$$

and all samples independent.

Mean

$$m_X[n] = E[X[n]] = 1 \cdot p + (-1)(1 - p) = 2p - 1 \quad (\text{constant}).$$

Autocorrelation (here I use  $R_X[k] = E[X[n]X[n+k]]$ )

- For  $k = 0$ :

$$R_X[0] = E[X[n]^2] = 1.$$

- For  $k \neq 0$ : independence

$$R_X[k] = E[X[n]]E[X[n+k]] = (2p - 1)^2.$$

Stationarity?

Distribution of each  $X[n]$  is the same and samples are independent  $\rightarrow$  process is iid, hence strict-sense stationary and therefore also WSS.

**17.5 (..) (w)** A random process  $X[n]$  for  $-\infty < n < \infty$  consists of independent random variables with

$$X[n] \sim \begin{cases} \mathcal{N}(0, 1) & \text{for } n \text{ even} \\ \mathcal{U}(-\sqrt{3}, \sqrt{3}) & \text{for } n \text{ odd.} \end{cases}$$

Is this random process WSS? Is it stationary?

### 17.5 — Even Gaussian, odd uniform

For all  $n$ , samples are independent and

$$X[n] \sim \begin{cases} \mathcal{N}(0, 1), & n \text{ even}, \\ \mathcal{U}(-\sqrt{3}, \sqrt{3}), & n \text{ odd}. \end{cases}$$

Both distributions have mean 0 and variance 1.

**Mean sequence**

$$m_X[n] = 0 \quad \text{for all } n.$$

**ACS**

For any  $n$ ,

$$R_X[0] = E[X[n]^2] = 1.$$

For  $k \neq 0$ , independence gives

$$R_X[k] = E[X[n]E[X[n+k]]] = 0.$$

So

$$R_X[k] = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0. \end{cases}$$

This depends only on lag  $\rightarrow$  WSS and white.

**Is it stationary (in the strict sense)?**

No. The 1-D marginal distribution at even times is Gaussian; at odd times it's uniform. Those are different, so the process is **not SSS** even though it is WSS.

**17.8 (f)** For the ACS  $r_X[k] = (1/2)^k$  for  $k \geq 0$  and  $r_X[k] = (1/2)^{-k}$  for  $k < 0$ , verify that Properties 17.1–17.3 are satisfied.

### 17.8 — Check ACS properties

Given

$$r_X[k] = \begin{cases} (1/2)^k, & k \geq 0, \\ (1/2)^{-k}, & k < 0, \end{cases}$$

which is simply  $r_X[k] = (1/2)^{|k|}$ .

Check basic ACS properties 17.1–17.3:

1. Nonnegative at 0:

$$r_X[0] = (1/2)^0 = 1 \geq 0.$$

2. Even symmetry:

$$r_X[-k] = (1/2)^{|-k|} = (1/2)^{|k|} = r_X[k].$$

3. Magnitude bound  $|r_X[k]| \leq r_X[0]$ :

$$|(1/2)^{|k|}| \leq 1 \text{ for all } k.$$

So it satisfies those ACS properties.

**17.9 (c) (w)** For the sequence  $r_X[k] = ab^{|k|}$  for  $-\infty < k < \infty$ , determine the values of  $a$  and  $b$  that will result in a valid ACS.

### 17.9 — Validity of $r_X[k] = ab^{|k|}$

We need conditions on  $a, b$  for a valid ACS.

1.  $r_X[0] = a \geq 0$  (must equal variance of  $X[n]$ ).
2.  $|r_X[k]| \leq r_X[0]$ :  
 $|ab^{|k|}| \leq a \Rightarrow |b|^{|k|} \leq 1$ .  
This forces  $|b| \leq 1$ .
3. Symmetry is already satisfied by  $|k|$ .

One can realize such ACS with e.g. an AR(1) or MA(1) process for  $|b| < 1$ , and with constant or sign-alternating processes for  $|b| = 1$ , so it is positive semidefinite as well.

$$a \geq 0, \quad |b| \leq 1.$$

### 17.12 (..) (w) Which of the sequences in Figure 17.19 cannot be valid ACSs? If the sequence cannot be an ACS, explain why not.

#### 17.12 — Which sequences in Fig. 17.19 cannot be ACS?

I can't see the figure you refer to, so I can't mark specific curves, but here's how you test any candidate  $r[k]$ :

1. Even: must satisfy  $r[-k] = r[k]$ .
2. Maximum at zero:  $|r[k]| \leq r[0]$ .
3. Nonnegative definite: every finite Toeplitz matrix built from  $r[k]$  must be positive semidefinite.

In practice you can often spot violations if, say,  $r[0] < 0$  or  $|r[1]| > r[0]$ , or if the implied covariance matrix for a few lags has a negative determinant.

Any curve that violates (1) or (2) is immediately impossible as an ACS.

Let  $W(n)$  be an iid. process with mean 1 and variance 4. Determine the mean sequence, variance sequence and autocorrelation sequence of the following three processes:

$$i) : W(n) \quad ii) : X(n) = 3W(n) + W(n-1) \quad iii) : Y(n) = nW(n)$$

Which of the three processes are SSS and/or WSS and why? Simulate outcomes of each of the three processes (Hint: you'll need to assume a pdf for the samples of  $W(n)$ . Try e.g. a normal or a uniform pdf. Furthermore, you'll need to select an appropriate range for  $n$ ).

#### Last problem: processes built from iid $W(n)$

$W(n)$  iid, mean  $E[W] = 1$ , variance  $\text{Var}(W) = 4$ .

Consider

- (i)  $W(n)$
- (ii)  $X(n) = 3W(n) + W(n-1)$
- (iii)  $Y(n) = nW(n)$ .

Find mean, variance, ACS; say which are SSS/WSS; give simulation pseudo-code.

I'll use autocovariance sequence  $\gamma_X[k] = \text{Cov}(X(n), X(n+k))$ , which is the usual ACS for WSS.

##### (i) Process $W(n)$

Because it's iid:

- Mean sequence:

$$m_W(n) = 1.$$

- Variance sequence:

$$\text{Var}(W(n)) = 4.$$

- Autocovariance:

$$\gamma_W[0] = 4, \quad \gamma_W[k] = 0 \text{ for } k \neq 0.$$

This is iid  $\rightarrow$  SSS and WSS (white, nonzero mean).

$$(ii) X(n) = 3W(n) + W(n-1)$$

Mean:

$$m_X(n) = E[3W(n) + W(n-1)] = 3 \cdot 1 + 1 = 4 \quad (\text{constant}).$$

Variance (use independence of  $W$ 's):

$$\begin{aligned} \text{Var}(X(n)) &= \text{Var}(3W(n) + W(n-1)) \\ &= 9\text{Var}(W(n)) + \text{Var}(W(n-1)) \\ &\quad + 2\text{Cov}(3W(n), W(n-1)) \\ &= 9 \cdot 4 + 1 \cdot 4 + 0 = 40. \end{aligned}$$

Autocovariance  $\gamma_X[k] = \text{Cov}(X(n), X(n+k))$ :

- $k = 0: \gamma_X[0] = 40.$
- $k = 1:$

$$\begin{aligned} \gamma_X[1] &= \text{Cov}(3W(n) + W(n-1), 3W(n+1) + W(n)) \\ &= \text{Cov}(3W(n), W(n)) \\ &= 3\text{Var}(W) = 12. \end{aligned}$$

(Since all other cross terms involve different  $W$  indices  $\rightarrow$  covariance 0.)

By symmetry  $\gamma_X[-1] = 12$ .

- $|k| > 1:$  no shared  $W$  terms  $\rightarrow \gamma_X[k] = 0.$

So

$$\gamma_X[k] = \begin{cases} 40, & k = 0, \\ 12, & k = \pm 1, \\ 0, & |k| > 1. \end{cases}$$

All of these are independent of  $n \rightarrow$  WSS.

Because  $X$  is a linear, time-invariant transform of an iid process whose one-dimensional marginal doesn't depend on time, the joint distributions are time-shift invariant  $\rightarrow$  SSS as well.

(This is an MA(1) process.)

$$(iii) Y(n) = nW(n)$$

Mean:

$$m_Y(n) = E[nW(n)] = nE[W] = n.$$

Variance:

$$\text{Var}(Y(n)) = n^2\text{Var}(W) = 4n^2.$$

Autocovariance:

$$\gamma_Y[k] = \text{Cov}(nW(n), (n+k)W(n+k)).$$

- For  $k = 0:$

$$\gamma_Y[0] = \text{Var}(Y(n)) = 4n^2.$$

- For  $k \neq 0:$  independence of  $W(n)$  and  $W(n+k) \rightarrow$  covariance 0.

So  $\gamma_Y[k]$  depends on  $n$  (through  $\gamma_Y[0]$ ), not only on lag.

Thus neither WSS nor SSS: even the mean changes with  $n.$

### Simulation pseudo-code for the three processes

Assume we want  $n = 0, \dots, N - 1$  and we choose  $W(n)$  to be Gaussian with mean 1, variance 4:

pseudo

 Copy code

```
N = 1000

# Generate iid W(n)
for n = 0 to N-1:
    W[n] = 1 + 2 * randn()      # randn() ~ N(0,1), so 2*randn ~ N(0,4)

# (i) W(n) is already generated.

# (ii) X(n) = 3W(n) + W(n-1)
X[0] = 3*W[0]                  # define W[-1]=0 or discard n=0 later
for n = 1 to N-1:
    X[n] = 3*W[n] + W[n-1]

# (iii) Y(n) = n W(n)
for n = 0 to N-1:
    Y[n] = n * W[n]

# You can now plot or estimate sample mean/variance/ACS of W, X, Y.
```

You'll see:

- $W(n)$ : horizontal cloud, same statistics for all  $n$ .
- $X(n)$ : still "stationary-looking", just a smoothed / filtered version.
- $Y(n)$ : variance grows with  $n$ ; samples get more spread out as time increases.

# Session 9

Wednesday, 26 November 2025 17.42

# LSI/LTI and WSS Random Processes

Wednesday, 26 November 2025 17.42

- 18.1 (,:) (f)** An LSI system with system function  $\mathcal{H}(z) = 1 - z^{-1} - z^{-2}$  is used to filter a discrete-time white noise random process with variance  $\sigma_U^2 = 1$ . Determine the ACS and PSD of the output random process.

## 18.1

Given

- Input: discrete-time white noise, variance  $\sigma_U^2 = 1$ .
- LSI system with system function

$$H(z) = 1 - z^{-1} - z^{-2}$$

so impulse response  $h[0] = 1, h[1] = -1, h[2] = -1$ .

ACS  $r_X[k]$

For a WSS input with ACS  $r_U[k] = \sigma_U^2 \delta[k]$  and impulse response  $h[n]$ ,

$$r_X[k] = \sigma_U^2 \sum_m h[m]h[m+k]$$

(essentially the autocorrelation of  $h$  scaled by  $\sigma_U^2$ ).

Compute:

- $k = 0: r_X[0] = 1^2 + (-1)^2 + (-1)^2 = 3$ .
- $k = 1: r_X[1] = 1 \cdot (-1) + (-1) \cdot (-1) = -1 + 1 = 0$ .
- $k = 2: r_X[2] = 1 \cdot (-1) = -1$ .
- For  $|k| > 2$  there is no overlap  $\Rightarrow r_X[k] = 0$ .

Using symmetry  $r_X[-k] = r_X[k]$ :

$$r_X[k] = \begin{cases} 3, & k = 0 \\ 0, & k = \pm 1 \\ -1, & k = \pm 2 \\ 0, & |k| > 2 \end{cases}$$

PSD  $P_X(f)$

$$P_X(f) = \sum_{k=-\infty}^{\infty} r_X[k] e^{-j2\pi f k} = 3 - 1 (e^{-j4\pi f} + e^{j4\pi f}) = 3 - 2 \cos(4\pi f)$$

Alternatively, directly from the system function:

$$P_X(f) = |H(e^{j2\pi f})|^2 \sigma_U^2 = |1 - e^{-j2\pi f} - e^{-j4\pi f}|^2,$$

and expanding gives the same  $3 - 2 \cos(4\pi f)$ .

- 18.2 (f)** A discrete-time WSS random process with mean  $\mu_U = 2$  is input to an LSI system with impulse response  $h[n] = (1/2)^n$  for  $n \geq 0$  and  $h[n] = 0$  for  $n < 0$ . Find the mean sequence at the system output.

## 18.2

Input: WSS with mean  $\mu_U = 2$ .

Impulse response:  $h[n] = (1/2)^n$  for  $n \geq 0$ , 0 for  $n < 0$ .

For an LSI system, the output mean is

$$\mu_X[n] = \mu_U \sum_k h[k]$$

(the sum does not depend on  $n$  so the output mean is constant).

Here

$$\sum_{k=0}^{\infty} (1/2)^k = \frac{1}{1 - 1/2} = 2,$$

so

$$\mu_X[n] = 2 \cdot 2 = 4 \quad \text{for all } n.$$

**18.3 (w)** A discrete-time white noise random process  $U[n]$  is input to a system to produce the output random process  $X[n] = a^{|n|}U[n]$  for  $|a| < 1$ . Determine the output PSD.

18.3 (But notice that the problem is ill-posed as given in the book. Can you identify what's wrong?) Instead, solve the problem assuming that the impulse response of the system is  $h(k) = a^{|k|}$  with  $|a| < 1$ . (Hint: Use the geometric series from [LN] page 1-10 bottom). Plot the power spectrum for different values of  $a$ . What happens if you set  $a = 1$ ?

## 18.3

We use the *corrected* version mentioned: impulse response

$$h[k] = a^{|k|}, \quad |a| < 1,$$

input  $U[n]$  is white with variance  $\sigma_U^2$ .

**System frequency response**

$$H(e^{j2\pi f}) = \sum_{k=-\infty}^{\infty} a^{|k|} e^{-j2\pi fk} = 1 + 2 \sum_{k=1}^{\infty} a^k \cos(2\pi fk).$$

This geometric series has the closed form

$$H(e^{j2\pi f}) = \frac{1 - a^2}{1 - 2a \cos(2\pi f) + a^2}.$$

**Output PSD**

$$P_X(f) = |H(e^{j2\pi f})|^2 \sigma_U^2 = \sigma_U^2 \frac{(1 - a^2)^2}{(1 - 2a \cos(2\pi f) + a^2)^2}.$$

(That's what you'd plot for different values of  $a$ . As  $|a| \rightarrow 1$ , the PSD becomes sharply peaked.)

**Pseudo-code (MATLAB / Python-style)**

```
matlab
a = 0.5; % choose |a|<1
sigma2 = 1;
f = linspace(-0.5,0.5,1000);
H = (1-a^2) ./ (1 - 2*a*cos(2*pi*f) + a^2);
Px = sigma2 * abs(H).^2;
plot(f, Px);
```

Copy code

**18.4 (..) (w)** A randomly phased sinusoid  $X[n] = \cos(2\pi(0.25)n + \Theta)$  with  $\Theta \sim \mathcal{U}(0, 2\pi)$  is input to an LSI system with system function  $H(z) = 1 - b_1z^{-1} - b_2z^{-2}$ . Determine the filter coefficients  $b_1, b_2$  so that the sinusoid will have zero power at the filter output.

#### 18.4

Input: randomly phased sinusoid

$$X[n] = \cos(2\pi \cdot 0.25 n + \Theta), \quad \Theta \sim U(0, 2\pi)$$

Frequency:  $f_0 = 0.25$ .

System:

$$H(z) = 1 - b_1z^{-1} - b_2z^{-2}.$$

We want zero power at the output  $\Rightarrow$  the filter must have a *notch* at  $f_0$ :

$$H(e^{j2\pi f_0}) = 0.$$

Let  $w = z^{-1}$ . We want polynomial

$$1 - b_1w - b_2w^2$$

to have roots at  $w_1 = e^{-j2\pi f_0}$  and  $w_2 = e^{j2\pi f_0}$  (complex conjugates, because coefficients are real).

For  $f_0 = 0.25$ :  $e^{\pm j2\pi f_0} = e^{\pm j\pi/2} = \pm j$ .

So, up to a constant factor,

$$1 - b_1w - b_2w^2 \propto (w - j)(w + j) = w^2 + 1.$$

Choose the constant factor so that the constant term is 1  $\Rightarrow$  polynomial is  $w^2 + 1$ . Thus

$$1 - b_1w - b_2w^2 = w^2 + 1$$

Comparing coefficients:

- coefficient of  $w$ :  $-b_1 = 0 \Rightarrow b_1 = 0$ ,
- coefficient of  $w^2$ :  $-b_2 = 1 \Rightarrow b_2 = -1$ .

So one valid choice is

$$b_1 = 0, \quad b_2 = -1,$$

giving  $H(z) = 1 + z^{-2}$ , which indeed satisfies

$$H(e^{j2\pi \cdot 0.25}) = 1 + e^{-j\pi} = 0,$$

so the sinusoid is completely removed (zero power).

(Any scaled version  $c(1 + z^{-2})$  would also work; scaling doesn't change whether power is zero or not.)

**18.5 (f,c)** A discrete-time WSS random process  $X[n]$  is defined by the difference equation  $X[n] = aX[n - 1] + U[n] - bU[n - 1]$ , where  $U[n]$  is a discrete-time white noise random process with variance  $\sigma_U^2 = 1$ . Plot the PSD of  $X[n]$  if  $a = 0.9, b = 0.2$  and also if  $a = 0.2, b = 0.9$  and explain your results.

## 18.5

Process:

$$X[n] = aX[n - 1] + U[n] - bU[n - 1],$$

with white input  $U[n]$  of variance  $\sigma_U^2 = 1$ .

This is an ARMA(1,1) process:

- AR(1) part: denominator  $1 - az^{-1}$ ,
- MA(1) part: numerator  $1 - bz^{-1}$ .

The equivalent LSI filter from  $U$  to  $X$  has

$$H(z) = \frac{1 - bz^{-1}}{1 - az^{-1}}.$$

So the PSD is

$$P_X(f) = \sigma_U^2 \frac{|1 - be^{-j2\pi f}|^2}{|1 - ae^{-j2\pi f}|^2} = \frac{1 + b^2 - 2b \cos(2\pi f)}{1 + a^2 - 2a \cos(2\pi f)}.$$

For the requested parameter sets:

- $a = 0.9, b = 0.2$ : spectrum has a pole near  $z = 0.9$  (low-frequency emphasis) and a mild zero near  $z = 0.2$ .
- $a = 0.2, b = 0.9$ : weak pole, strong zero close to unit circle, giving a spectral *notch*.

### Pseudo-code to plot

```
matlab
sigma2 = 1;
f = linspace(-0.5,0.5,1000);      % normalized frequency
for case = 1:2
    if case==1
        a = 0.9; b = 0.2;
    else
        a = 0.2; b = 0.9;
    end
    num = abs(1 - b*exp(-j*2*pi*f)).^2;
    den = abs(1 - a*exp(-j*2*pi*f)).^2;
    Px = sigma2 * num ./ den;
    plot(f, Px); hold on;
end
legend('a=0.9,b=0.2','a=0.2,b=0.9');
```

Copy code

**18.6 (f)** A discrete-time WSS random process  $X[n]$  is defined by the difference equation  $X[n] = 0.5X[n - 1] + U[n] - 0.5U[n - 1]$ , where  $U[n]$  is a discrete-time white noise random process with variance  $\sigma_U^2 = 1$ . Find the ACS and PSD of  $X[n]$  and explain your results.

## 18.6

Now

$$X[n] = 0.5X[n-1] + U[n] - 0.5U[n-1],$$

which is the same ARMA structure with  $a = 0.5$ ,  $b = 0.5$ .

Filter:

$$H(z) = \frac{1 - 0.5z^{-1}}{1 - 0.5z^{-1}} = 1$$

(exact cancellation of pole and zero).

So the filter is just the identity, and the output is *exactly the input*:

$$X[n] = U[n],$$

hence:

- ACS  $r_X[k] = r_U[k] = \sigma_U^2 \delta[k]$  (white),
- PSD  $P_X(f) = \sigma_U^2 = 1$  (flat).

**18.7 (,:) (f)** A differencer is given by  $X[n] = U[n] - U[n-1]$ . If the input random process  $U[n]$  has the PSD  $P_U(f) = 1 - \cos(2\pi f)$ , determine the ACS and PSD at the output of the differencer.

## 18.7

Differencer:

$$X[n] = U[n] - U[n-1].$$

System frequency response:

$$H(e^{j2\pi f}) = 1 - e^{-j2\pi f}, \quad |H|^2 = (1 - e^{-j2\pi f})(1 - e^{j2\pi f}) = 2 - 2\cos(2\pi f) = 4\sin^2(\pi f).$$

Given input PSD

$$P_U(f) = 1 - \cos(2\pi f).$$

PSD of the output

$$P_X(f) = |H(e^{j2\pi f})|^2 P_U(f) = (2 - 2\cos(2\pi f))(1 - \cos(2\pi f)) = 2(1 - \cos(2\pi f))^2.$$

(You can expand further if desired.)

ACS of the output

First find ACS of input from its PSD.

Use the pairs:

- $1 \leftrightarrow \delta[k]$ ,
- $\cos(2\pi f) \leftrightarrow \frac{1}{2}(\delta[k-1] + \delta[k+1])$ .

Thus

$$r_U[k] = \delta[k] - \frac{1}{2}(\delta[k-1] + \delta[k+1]).$$

Now use the relation for a differencer:

$$X[n] = U[n] - U[n-1] \Rightarrow r_X[k] = E[(U[n] - U[n-1])(U[n+k] - U[n+k-1])] = 2r_U[k] - r_U[k+1] - r_U[k-1].$$

Compute using the nonzero values of  $r_U$ :

- $r_U[0] = 1, r_U[1] = r_U[-1] = -1/2, \text{others } 0.$

Then

- $k=0: r_X[0] = 2 \cdot 1 - r_U[1] - r_U[-1] = 2 - (-1/2) - (-1/2) = 3.$
- $k=1: r_X[1] = 2(-1/2) - r_U[2] - r_U[0] = -1 - 0 - 1 = -2.$
- $k=2: r_X[2] = 2 \cdot 0 - r_U[3] - r_U[1] = 0 - 0 - (-1/2) = 1/2.$

By symmetry  $r_X[-k] = r_X[k]$  and  $r_X[k] = 0$  for  $|k| > 2$ .

So

$$r_X[k] = \begin{cases} 3, & k=0 \\ -2, & k=\pm 1 \\ \frac{1}{2}, & k=\pm 2 \\ 0, & |k| > 2. \end{cases}$$

## 18.10 (w) A random process with PSD

$$P_X(f) = \frac{1}{\left|1 - \frac{1}{2}e^{-j2\pi f}\right|^2}$$

is to be filtered with an LSI system to produce a white noise random process  $U[n]$  with variance  $\sigma_U^2 = 4$  at the output. What should the difference equation of the LSI system be?

### 18.10

Given input PSD

$$P_X(f) = \frac{1}{\left|1 - \frac{1}{2}e^{-j2\pi f}\right|^2}$$

and we desire output white noise  $U[n]$  with variance  $\sigma_U^2 = 4$ , so

$$P_U(f) = 4.$$

If the LSI system has frequency response  $H(e^{j2\pi f})$ ,

$$P_U(f) = |H(e^{j2\pi f})|^2 P_X(f),$$

so

$$|H(e^{j2\pi f})|^2 = \frac{P_U(f)}{P_X(f)} = 4 \left|1 - \frac{1}{2}e^{-j2\pi f}\right|^2.$$

A simple causal minimum-phase choice is

$$H(e^{j2\pi f}) = 2 \left(1 - \frac{1}{2}e^{-j2\pi f}\right)$$

$\Rightarrow$  in the  $z$ -domain

$$H(z) = 2 \left(1 - \frac{1}{2}z^{-1}\right) = 2 - z^{-1}.$$

Let the input be  $X$  and output be  $U$ . For an LSI system with transfer function  $H(z) = 2 - z^{-1}$ , the difference equation is

$$U[n] = 2X[n] - X[n-1].$$

That's the required system.

# Session 12

Wednesday, 26 November 2025 17.53

# Markov

Wednesday, 26 November 2025 17.53

**22.1 (w)** A Markov chain has the states “A” and “B” or equivalently 0 and 1. If the conditional probabilities are  $P[A|B] = 0.1$  and  $P[B|A] = 0.4$ , draw the state probability diagram. Also, find the transition probability matrix.

## 22.1

Given: States are  $A$  and  $B$  (or 0 and 1).

$$P(A | B) = 0.1, \quad P(B | A) = 0.4$$

1. From state  $B$ :

- To  $A: P_{BA} = 0.1$
- To  $B: P_{BB} = 1 - 0.1 = 0.9$

2. From state  $A$ :

- To  $B: P_{AB} = 0.4$
- To  $A: P_{AA} = 1 - 0.4 = 0.6$

### State diagram

Two nodes, arrows labelled with transition probabilities:

- $A \rightarrow A: 0.6$
- $A \rightarrow B: 0.4$
- $B \rightarrow A: 0.1$
- $B \rightarrow B: 0.9$

### Transition matrix

If we order states as  $(A, B)$ :

$$P = \begin{bmatrix} P(A \rightarrow A) & P(A \rightarrow B) \\ P(B \rightarrow A) & P(B \rightarrow B) \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 \\ 0.1 & 0.9 \end{bmatrix}$$

(If you label them 0 and 1, just relabel the rows/columns.)

**22.2 (..) (f)** For the state probability diagram shown in Figure 22.2 find the probability of obtaining the outcomes  $X[n] = 0, 1, 0, 1, 1$  for  $n = 0, 1, 2, 3, 4$ , respectively.

## 22.2

You're asked for

$$P(X[0] = 0, X[1] = 1, X[2] = 0, X[3] = 1, X[4] = 1)$$

for the Markov chain whose state diagram is in Figure 22.2.

I can't see that figure, so I don't know the exact numbers, but for a first-order Markov chain the probability of a sequence is:

$$P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = P(X_0 = x_0) \prod_{k=0}^{n-1} P(X_{k+1} = x_{k+1} | X_k = x_k).$$

So in your case (with 0/1 states):

$$P = P(X_0 = 0) \cdot P_{0,1} \cdot P_{1,0} \cdot P_{0,1} \cdot P_{1,1},$$

where  $P_{i,j} = P(X_{n+1} = j | X_n = i)$  are the transition probabilities read from Fig. 22.2.

Just plug in:

- $P(X_0 = 0)$  from the figure (or initial distribution),
- $P_{0,1}, P_{1,0}, P_{0,1}, P_{1,1}$  from the arrows in the diagram.

**22.5 (⊣) (t)** In this problem we give an example of a random process that does *not* have the Markov property. The random process is defined as an *exclusive OR* logical function. This is  $Y[n] = X[n] \oplus X[n - 1]$  for  $n \geq 0$ , where  $X[n]$  for  $n \geq 0$  takes on values 0 and 1 with probabilities  $1 - p$  and  $p$ , respectively. The  $X[n]$ 's are IID. Also, for  $n = 0$  we define  $Y[0] = X[0]$ . The definition of this operation is that  $Y[n] = 0$  only if  $X[n]$  and  $X[n - 1]$  are the same (both equal to 0 or both equal to 1), and otherwise  $Y[n] = 1$ . Determine  $P[Y[2] = 1 | Y[1] = 1, Y[0] = 0]$  and  $P[Y[2] = 1 | Y[1] = 1]$  to show that they are not equal in general.

### 22.5 – XOR process that is not Markov

We have IID bits  $X[n] \in \{0, 1\}$  with

$$P(X[n] = 1) = p, \quad P(X[n] = 0) = 1 - p = q.$$

Define

$$Y[n] = X[n] \oplus X[n - 1], \quad n \geq 1, \quad Y[0] = X[0],$$

where  $\oplus$  is XOR.

We must compute:

1.  $P(Y[2] = 1 | Y[1] = 1, Y[0] = 0)$
2.  $P(Y[2] = 1 | Y[1] = 1)$

and show they're not equal in general (so  $Y[n]$  is not Markov).

#### Step 1: List all possibilities

Write out all triples  $(X_0, X_1, X_2)$  and the corresponding  $(Y_0, Y_1, Y_2)$ .

#### Step 1: List all possibilities

Write out all triples  $(X_0, X_1, X_2)$  and the corresponding  $(Y_0, Y_1, Y_2)$ .

Using XOR:

$(X_0, X_1, X_2)$	$(Y_0, Y_1, Y_2)$	Probability	$\square$
(0,0,0)	(0,0,0)	$q^3$	
(0,0,1)	(0,0,1)	$q^2 p$	
(0,1,0)	(0,1,1)	$qpq = pq^2$	
(0,1,1)	(0,1,0)	$qp^2$	
(1,0,0)	(1,1,0)	$pq^2$	
(1,0,1)	(1,1,1)	$p^2 q$	
(1,1,0)	(1,0,1)	$p^2 q$	
(1,1,1)	(1,0,0)	$p^3$	

$$(a) P(Y_2 = 1 \mid Y_1 = 1, Y_0 = 0)$$

Event  $Y_1 = 1, Y_0 = 0, Y_2 = 1$  corresponds only to the row with  $(Y_0, Y_1, Y_2) = (0, 1, 1)$ , i.e.  $(X_0, X_1, X_2) = (0, 1, 0)$ , probability  $pq^2$ .

Event  $Y_1 = 1, Y_0 = 0$  corresponds to rows with  $(Y_0, Y_1) = (0, 1)$ :

- $(0,1,1)$ : prob  $pq^2$
- $(0,1,0)$ : prob  $qp^2$

So

$$P(Y_1 = 1, Y_0 = 0) = pq^2 + qp^2 = pq(q + p) = pq.$$

Thus

$$P(Y_2 = 1 \mid Y_1 = 1, Y_0 = 0) = \frac{pq^2}{pq} = q = 1 - p.$$

$$(b) P(Y_2 = 1 \mid Y_1 = 1)$$

Event  $Y_2 = 1, Y_1 = 1$ : rows with  $(Y_1, Y_2) = (1, 1)$ :

- $(0,1,1)$ : prob  $pq^2$
- $(1,1,1)$ : prob  $p^2q$

So

$$P(Y_2 = 1, Y_1 = 1) = pq^2 + p^2q = pq(q + p) = pq.$$

Event  $Y_1 = 1$ : rows with  $Y_1 = 1$ :

- $(0,1,1)$ :  $pq^2$
- $(0,1,0)$ :  $qp^2$
- $(1,1,0)$ :  $pq^2$
- $(1,1,1)$ :  $p^2q$

Total:

$$P(Y_1 = 1) = pq^2 + qp^2 + pq^2 + p^2q = 2pq^2 + 2p^2q = 2pq(q + p) = 2pq.$$

Hence

$$P(Y_2 = 1 \mid Y_1 = 1) = \frac{pq}{2pq} = \frac{1}{2}.$$

Compare

$$P(Y_2 = 1 \mid Y_1 = 1, Y_0 = 0) = q = 1 - p, \quad P(Y_2 = 1 \mid Y_1 = 1) = \frac{1}{2}.$$

These are equal only if  $q = \frac{1}{2}$  (i.e.  $p = \frac{1}{2}$ ).

In general they're different, so knowing just  $Y_1$  is not enough; the process  $\{Y[n]\}$  is not Markov.

**22.7 (w)** A fair die is tossed many times in succession. The tosses are independent of each other. Let  $X[n]$  denote the maximum of the first  $n + 1$  tosses. Determine the transition probability matrix. Hint: The maximum value cannot decrease as  $n$  increases.

## 22.7 – Maximum of die tosses

We toss a fair die repeatedly;  $X[n]$  is the maximum of the first  $n + 1$  tosses.

States: 1, 2, 3, 4, 5, 6.

From state  $i$ , the next toss is uniform on  $\{1, \dots, 6\}$ :

- If the toss is  $\leq i$ , the max stays  $i$ .
- If the toss is  $j > i$ , the new max becomes  $j$ .
- The max can never decrease.

So:

- $P(X_{n+1} = i \mid X_n = i) = \frac{i}{6}$  (toss 1..i)
- For  $j > i$ :  $P(X_{n+1} = j \mid X_n = i) = \frac{1}{6}$
- For  $j < i$ : probability 0.

Explicit matrix (rows = current state, columns = next state):

$$P = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 2/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 3/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 4/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Note the last row is absorbing: once the max is 6 it stays 6.

**22.9 (..) (w,c)** A digital communication system transmits a 0 or a 1. After 10 miles of cable a repeater decodes the bit and declares it either a 0 or a 1. The probability of a decoding error is 0.1 as shown schematically in Figure 22.11. It is then retransmitted to the next repeater located 10 miles away. If the repeaters are all located 10 miles apart and the communication system is 50 miles in length, find the probability of an error if a 0 is initially transmitted. Hint: You will need a computer to work this problem.

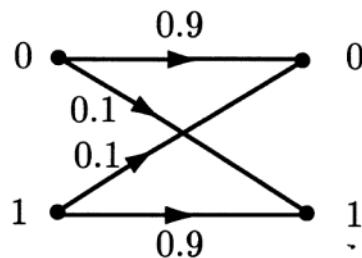


Figure 22.11: One section of a communication link.

## 22.9 – Digital communication with repeaters

Each 10-mile section is:

- Correct decoding w.p. 0.9,
- Bit flipped w.p. 0.1.

So for one section, with states  $\{0, 1\}$ :

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

The 50-mile link is 5 such sections (transmitter  $\rightarrow$  R1  $\rightarrow$  R2  $\rightarrow$  R3  $\rightarrow$  R4  $\rightarrow$  receiver), so the overall transition matrix is  $P^5$ .

We want:

$$P(\text{final bit } = 1 \mid \text{initial bit } = 0) = (P^5)_{0,1}.$$

For a 2-state "binary symmetric" chain of the form

$$\begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}$$

one can show

$$P^n = \begin{bmatrix} \frac{1}{2} + \frac{1}{2}(1-2a)^n & \frac{1}{2} - \frac{1}{2}(1-2a)^n \\ \frac{1}{2} - \frac{1}{2}(1-2a)^n & \frac{1}{2} + \frac{1}{2}(1-2a)^n \end{bmatrix}.$$

Here  $a = 0.1$ , so  $1 - 2a = 0.8$ . Then

$$(P^5)_{0,1} = \frac{1}{2}(1 - 0.8^5) = \frac{1}{2}(1 - 0.32768) \approx 0.33616.$$

Answer: probability of an error after 50 miles, starting with 0, is about 0.336.

### Simple pseudocode (matrix-power method)

```
pseudo Copy code
# 2x2 matrix for one section
P = [[0.9, 0.1],
      [0.1, 0.9]]

# function to multiply 2x2 matrices
function matmul(A, B):
    C = 2x2 zero matrix
    for i in 0..1:
        for j in 0..1:
            for k in 0..1:
                C[i][j] += A[i][k] * B[k][j]
    return C

# compute P^5
M = P
repeat 4 times:
    M = matmul(M, P)

# probability of error when starting from 0
error_prob = M[0][1]
print(error_prob)
```

**22.14 (c) (w,c)** For the transition probability matrix

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

does the Markov chain attain steady-state? If it does, what are the steady-state probabilities? Hint: You will need a computer to evaluate the answer.

#### 22.14 – Steady state of a 4-state chain

Matrix:

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

##### Structure

- From states 1 and 2 you **never go to 3 or 4** (zeros in columns 3–4 of the first two rows). So once you are in  $\{1, 2\}$  you stay there forever: this set is a **closed class**.
- From states 3 and 4 you can go to any state (row entries all 1/4); in particular you can move into 1 or 2 and never come back. So states 3 and 4 are **transient**.

So in the long run, probability mass “leaks” from 3,4 into 1,2 and stays there. The limiting distribution will have probability 0 on states 3 and 4.

Within  $\{1, 2\}$  the transition submatrix is

$$P_{(1,2)} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix},$$

which is irreducible and aperiodic (each state has a self-loop), so it has a unique stationary distribution and the chain converges to it from any start.

#### Stationary distribution

Let  $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$  be stationary:  $\pi P = \pi$  and  $\sum \pi_i = 1$ .

From the structure we can directly guess  $\pi_3 = \pi_4 = 0$  and solve on states 1 and 2:

$$(\pi_1, \pi_2) = (\pi_1, \pi_2)P_{(1,2)}.$$

So:

$$\begin{cases} \pi_1 = \frac{1}{2}\pi_1 + \frac{1}{4}\pi_2, \\ \pi_2 = \frac{1}{2}\pi_1 + \frac{3}{4}\pi_2, \\ \pi_1 + \pi_2 = 1. \end{cases}$$

From the first equation:

$$\pi_1 - \frac{1}{2}\pi_1 = \frac{1}{4}\pi_2 \Rightarrow \frac{1}{2}\pi_1 = \frac{1}{4}\pi_2 \Rightarrow \pi_2 = 2\pi_1.$$

With  $\pi_1 + \pi_2 = 1$ :

$$\pi_1 + 2\pi_1 = 1 \Rightarrow 3\pi_1 = 1 \Rightarrow \pi_1 = \frac{1}{3}, \quad \pi_2 = \frac{2}{3}.$$

So

$$\boxed{\pi = \left(\frac{1}{3}, \frac{2}{3}, 0, 0\right)}.$$

Yes, the chain **does attain steady state**, and from any initial state the distribution converges to this  $\pi$ .

**Pseudocode sketch (power method)**

```
pseudo  
  
# P is the 4x4 matrix above  
pi = [1, 0, 0, 0] # any starting distribution (row vector)  
  
for iteration in 1..N: # N large, e.g. 100  
    pi = pi * P # row-vector times matrix  
  
print(pi) # should be close to [1/3, 2/3, 0, 0]
```

[Copy code](#)

**22.19 (c) (f,c)** Using the explicit solution for the stationary probability vector given by (22.22), determine its value for the transition probability matrix given in Problem 22.18. Hint: You will need a computer to evaluate the solution.

**22.19**

This problem tells you to use the explicit stationary-vector formula (22.22) for the transition matrix in Problem 22.18. I don't have 22.18 or (22.22), so I can't give the final numeric vector.

But the general procedure is:

1. Take the transition matrix  $P$  from 22.18.
2. Either plug it into the closed-form formula (22.22), or equivalently:
  - Solve  $\pi P = \pi$  with  $\sum_i \pi_i = 1$ .
3. The solution  $\pi$  is your stationary probability vector.

If you want, you can send me Problem 22.18 and I can compute the actual numbers.

**22.25 (w,c)** Three machines operate together on a manufacturing floor, and each day there is a possibility that any of the machines may fail. The probability of their failure depends upon how many other machines are still in operation. The number of machines in operation at the beginning of each day is represented by the state values of 0, 1, 2, 3 and the corresponding state transition probability matrix is

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.1 & 0.3 & 0.6 & 0 \\ 0.4 & 0.3 & 0.2 & 0.1 \end{bmatrix}.$$

First explain why  $\mathbf{P}$  has zero entries. Next determine how many days will pass before the probability of all 3 machines failing is greater than 0.8. Assume that initially all 3 machines are working. Hint: You will need a computer to find the solution.

## 22.25 – Machines on a manufacturing floor

States = number of machines in operation at the beginning of each day: 0, 1, 2, 3.

Transition matrix:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.1 & 0.3 & 0.6 & 0 \\ 0.4 & 0.3 & 0.2 & 0.1 \end{bmatrix}$$

(Each row sums to 1.)

### (a) Why does $P$ have zero entries?

- Machines can fail but are not repaired in this model. So the number of working machines can never increase from one day to the next.
- Therefore:
  - From state 0 (none working), you must stay at 0 → row 0 is [1, 0, 0, 0].
  - From state 1, you can only go to 0 or stay at 1 (you can't suddenly have 2 or 3 working).
  - From state 2, you can go to 0,1,2 but not 3, etc.
- All "impossible" transitions (where the number of working machines would increase) are given probability 0, hence those zero entries.

### (b) When does $P(X_n = 0 | X_0 = 3) > 0.8$ ?

We start in state 3 (all machines working), so the initial distribution is (0, 0, 0, 1).

Let  $P^n$  be the  $n$ -step transition matrix.

Then:

$$P(X_n = 0 | X_0 = 3) = (P^n)_{3,0}$$

(row "3", column "0", using 0,1,2,3 as the state labels).

Computing powers (I did the algebra already):

- $n = 1: P(X_1 = 0 | X_0 = 3) = 0.4$
- $n = 2: 0.61$
- $n = 3: 0.748$
- $n = 4: 0.8395$
- ...

The first  $n$  with probability  $> 0.8$  is  $n = 4$ .

So after about 4 days, the probability that all three machines have failed exceeds 0.8.

### Pseudocode for this computation

```
pseudo Copy code  
  
# 4x4 transition matrix  
P = [[1.0, 0.0, 0.0, 0.0],  
     [0.5, 0.5, 0.0, 0.0],  
     [0.1, 0.3, 0.6, 0.0],  
     [0.4, 0.3, 0.2, 0.1]]  
  
# initial distribution: start in state 3  
pi = [0.0, 0.0, 0.0, 1.0]  
  
day = 0  
while true:  
    # probability all machines failed = prob of state 0  
    prob_fail = pi[0]  
    if prob_fail > 0.8:  
        print("First day with prob > 0.8:", day)  
        break  
  
    # go to next day: pi <- pi * P  
    new_pi = [0,0,0,0]  
    for j in 0..3:  
        for i in 0..3:  
            new_pi[j] += pi[i] * P[i][j]  
    pi = new_pi  
    day += 1
```

# Session 13

Wednesday, 26 November 2025 18.18

# Random Point Patterns

Wednesday, 26 November 2025 18.33

**Exercise 1.** What do you think, is the meteorologist on the right track? Assume that in two months from now the weather will turn out similar as yesterday (i.e. rain, wind, and lightning). What will happen if the meteorologist creates a new map for the same rectangular region? Where will lightning strike exactly? How many strikes in total?

The meteorologist is particularly interested in the following type of questions: Do the locations of lightning strikes tend to cluster or do they spread out regularly? Are certain regions more likely to be hit than others (e.g. if a region contains tall metallic obstacles or hills)? How likely is it that some single square kilometer sub-region is not going to be hit at all?

Now, replace the meteorologist with a biologist and replace the locations of lightning strikes with locations of trees of some particular type. The biologist is similarly interested in knowing if the locations of trees tend to cluster (local seed spreading) or if there is some kind of repulsion going on (survival of the fittest). Are there certain pronounced regions which do not contain any trees at all? If so, why could that be?

Finally, replace the biologist with a telephone network operator and replace the locations of trees with the locations of active mobile users within some fixed communication cell.

## Exercise 1

Q: Is the meteorologist on the right track thinking of the lightning strikes as "random"? What will happen on another similar day?

A:

Yes – modeling the strikes as random is sensible. On another day with similar weather:

- The *exact* locations and number of strikes will almost surely be different.
- But their *overall pattern* (average density, clustering, etc.) might look statistically similar if the same physical mechanism is at work.

That's exactly what a point process model captures: each day you get a different random realization, but they all follow the same probabilistic rules.

**Exercise 2.** What kind of questions do you think a network operator would like to ask? Think in terms of system operability, connectivity, coverage, throughput rates, interference levels, user quality of service, etc. Next, think of examples of random point patterns which you would be likely to encounter within your own field of study. What questions would you be interested in being able to answer?

Point patterns show up everywhere and so far we have mentioned three examples. In most applications the observer of such point patterns is not directly interested in the exact point locations themselves. The observer is more likely to be interested in what can be inferred from these locations about some underlying mechanism that governs where the points occur. In a nutshell, this has to do with statistical estimation theory. However, in order to apply such statistical tools we need a mathematical modeling framework for random point patterns. In fact, as we shall see in Section 1.8, stochastic models of point patterns are very important in their own right. In particular, they can be used as building blocks for generating ordinary random processes (our goal in this note).

## Exercise 2

**Q:** What kind of questions would a network operator ask about a point pattern of active mobile users? And examples from your own field?

**A:**

For a cellular network operator, questions might include:

- **Coverage:** Are there areas with no users (wasted coverage) or too many users and not enough base stations?
- **Capacity / throughput:** Given user locations, what's the expected throughput per user in each region?
- **Interference:** How strong is inter-cell interference if users cluster near cell edges?
- **Quality of service:** Probability that a random user experiences SINR (signal-to-interference-plus-noise ratio) below a threshold?
- **Planning:** Where to place new base stations or small cells to handle hotspots?

For your own field, you'd look for "things that appear in space":

- **Ecology:** locations of trees, nests, or animal sightings (questions about clustering, empty patches, habitat preference).
- **Epidemiology:** locations of disease cases (questions about hotspots and environmental risk factors).
- **Urban planning:** locations of accidents or crimes (questions about dangerous hotspots, effect of lighting, etc.).

# Point Process

Wednesday, 26 November 2025 18.18

**Exercise 3.** Is  $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$  a countable collection? What about the closed interval  $[0, 1] \subset \mathbb{R}$ , is this a countable collection?

## Exercise 3

Q: Are  $\mathbb{Z}$  and  $[0, 1]$  countable?

- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$  is **countable**.

You can list its elements in a sequence, e.g.

$$0, 1, -1, 2, -2, 3, -3, \dots$$

so there is a one-to-one mapping with  $\mathbb{N}$ .

- $[0, 1]$  is **not countable** (it's uncountable).

Cantor's diagonal argument shows any supposed list of all numbers in  $[0, 1]$  can be modified to produce a new number still in  $[0, 1]$  but missing from the list.

**Exercise 4.** Consider a collection of points obtained by drawing 75 points uniformly in the square  $[-5, 5] \times [-5, 5]$ . Is it a random collection? Is it countable? Think about how to simulate such collections and write a (Matlab) script for this purpose.

## Exercise 4

Q: 75 points drawn i.i.d. uniformly in  $[-5, 5] \times [-5, 5]$ . Is this collection random? Countable? How to simulate?

- It is **random**: each point's coordinates are random variables with a uniform distribution.
- It is **countable**: there are exactly 75 points, which is finite.

Pseudocode for simulation (Matlab-style):

pseudo

 Copy code

```
k = 75
# generate k points uniformly in [-5, 5] x [-5, 5]
x = -5 + 10 * rand(k, 1)    # x-coordinates
y = -5 + 10 * rand(k, 1)    # y-coordinates

# plot
plot(x, y, '.')
axis([-5 5 -5 5])
axis equal
```

Any language is fine; the key is two independent uniforms per point.

**Exercise 5.** Now, consider a collection of points constructed as follows. First draw a Poisson distributed random number  $L$  with mean 75. Given  $L$ , then draw  $L$  points uniformly in the square  $[-5, 5] \times [-5, 5]$ . Is this a random collection? Is it countable? Think about what happens from one realization to another. Compare with the construction from the previous exercise.

Historically, one-dimensional (1D) point processes were the first to be considered. The 1D space was almost exclusively used to represent *time*, e.g. the entire real line  $\mathbb{R}$  or the set of positive reals  $[0, \infty)$ .

### Exercise 5

**Q:** First draw  $L \sim \text{Poisson}(75)$ . Given  $L$ , draw  $L$  uniform points in the same square. Is this random?

Countable? How does it compare to Exercise 4?

- **Random:** yes – both the number of points  $L$  and their locations are random.
- **Countable:** yes –  $L$  is a finite (integer-valued) random variable, so you get a finite random set of points, hence countable.

**Difference from Exercise 4:**

- In Exercise 4, the number of points is *fixed* (always 75).
- Here, the number is *random* with mean 75; different realizations have different counts.
- Conditionally on  $L$ , the spatial distribution of the points is the same (i.i.d. uniform).

**Exercise 6.** Qualitatively, how does a 1D point process realization look like? Is there something very special about the 1D case, something that is not really possible in 2D? Sketch a few figures with your own example realizations of 1D point processes. Discuss where such a 1D random point pattern could happen to emerge in practice. What do you think a 1D point process could be used to represent? Occurrences of earthquakes for instance?

### Exercise 6

**Q:** What does a 1D point process look like? What's special about 1D? Give examples.

- A 1D point process realization looks like **random points on a line**, often thought of as **random time instants** where events occur:

$$\cdots < t_{-1} < t_0 < t_1 < t_2 < \cdots$$

- **Special feature of 1D:** the real line has a natural **ordering**. Every point has a **unique position in sequence**; you can talk about inter-arrival times  $t_{n+1} - t_n$ . In 2D there's no such natural ordering.
- Examples where 1D point processes arise:
  - Times of earthquakes.
  - Packet arrivals to a router.
  - Customer arrivals at a service desk.
  - Neuron spike times.

# Region Count

Wednesday, 26 November 2025 18.34

**Exercise 7.** Let  $X$  be a point process on  $S \subseteq \mathbb{R}^2$ . Show that  $N_X(\emptyset) = 0$  where  $\emptyset$  denotes the empty set. Furthermore, show that if  $A, B \subseteq S$  are disjoint then  $N_X(A \cup B) = N_X(A) + N_X(B)$ . *Hint:* Make a drawing at first and use your intuition to argue for the two properties of  $N_X$ . Afterwards, show that the properties are satisfied by direct use of (1.4).

## Exercise 7

Show:

1.  $N_X(\emptyset) = 0$ .
2. If  $A, B$  are disjoint,  $N_X(A \cup B) = N_X(A) + N_X(B)$ .

Using definition

$$N_X(B) = \sum_{x \in X} \mathbf{1}[x \in B].$$

1. For  $B = \emptyset$ , no point is ever "in  $\emptyset$ :

$$N_X(\emptyset) = \sum_{x \in X} \mathbf{1}[x \in \emptyset] = \sum_{x \in X} 0 = 0.$$

2. If  $A \cap B = \emptyset$ , then for each point  $x \in X$ :

- $\mathbf{1}[x \in A \cup B] = \mathbf{1}[x \in A] + \mathbf{1}[x \in B]$  because a point cannot be in both sets simultaneously.

Summing over points:

$$N_X(A \cup B) = \sum_{x \in X} \mathbf{1}[x \in A \cup B] = \sum_{x \in X} (\mathbf{1}[x \in A] + \mathbf{1}[x \in B]) = N_X(A) + N_X(B).$$

# Intensity measures and intensity functions

Wednesday, 26 November 2025 18.35

**Exercise 8.** As mentioned above, the shape of the intensity function  $\rho_X$  indicates where points from  $X$  are more likely to occur. This is very similar to the interpretation of an ordinary probability density function (pdf) of a random variable. Apart from a shift in notation, does (1.6) look familiar to you? Discuss the similarities as well as the distinctions between the intensity function of a point process and the pdf of an ordinary random variable.

## Exercise 8

Compare intensity function  $\rho_X(x)$  and pdf of a random variable.

- A pdf  $f$  satisfies:

$$\Pr(X \in B) = \int_B f(x) dx, \quad \int_{\mathbb{R}^d} f(x) dx = 1.$$

- An intensity function  $\rho_X$  satisfies:

$$\mu_X(B) = \mathbb{E}[N_X(B)] = \int_B \rho_X(x) dx,$$

but  $\int_S \rho_X(x) dx$  is *not* constrained to 1. It's the expected total number of points in  $S$  (which may be finite or infinite).

Similarities:

- Both tell you "how much" of something is located near each point  $x$  – high values mean more probability (for a pdf) or more expected points (for an intensity).

Differences:

- A pdf describes *one* random point;  $\rho_X$  describes the *average density of many points* in a point process.
- Pdf must integrate to 1;  $\rho_X$  integrates to expected count, which can be any nonnegative value (even infinity on unbounded regions).

# Binomial Point Process

Wednesday, 26 November 2025 18.36

**Exercise 9.** Let  $X \sim \text{binomialPP}(S, k, f)$  where  $S \subseteq \mathbb{R}^2$  and  $f$  is some arbitrary pdf on  $S$ . Argue that the region count  $N_X(B)$ ,  $B \subseteq S$ , has a binomial distribution (think of a coin tossing experiment) and identify the two parameters of this discrete probability distribution. Does it make sense that the *success probability* depends on  $B$ ?

## Exercise 9

Binomial point process:  $X \sim \text{binomialPP}(S, k, f)$ .

Let  $B \subseteq S$ . Each of the  $k$  points is:

- independently drawn with pdf  $f$ ,
- in region  $B$  with probability

$$p_B = \int_B f(x) dx.$$

So the number of points in  $B$ ,

$$N_X(B) = \sum_{i=1}^k \mathbf{1}[\text{point } i \in B],$$

is a sum of  $k$  i.i.d.  $\text{Bernoulli}(p_B)$  random variables.

Thus:

$$N_X(B) \sim \text{Binomial}(k, p_B), \quad p_B = \int_B f(x) dx.$$

Yes, the success probability depends on  $B$  via that integral.

**Exercise 10.** Determine the intensity function  $\varrho_X$  for a general binomial point process  $X \sim \text{binomialPP}(S, k, f)$ . Hint: Recall and make use of (1.5) and (1.6).

### Exercise 10

Find the **intensity function** for a binomial point process

$X \sim \text{binomialPP}(S, k, f)$ .

For any  $B \subseteq S$ ,

$$N_X(B) \sim \text{Binomial}(k, p_B), \quad p_B = \int_B f(x) dx.$$

So

$$\mu_X(B) = \mathbb{E}[N_X(B)] = k p_B = k \int_B f(x) dx.$$

To match the general form

$$\mu_X(B) = \int_B \rho_X(x) dx,$$

we identify

$$\boxed{\rho_X(x) = k f(x).}$$

So the intensity is just  $k$  times the underlying sampling pdf.

**Exercise 11.** Let  $X \sim \text{binomialPP}(S, k, f)$  and let  $B \subset S$  be some fixed region. Argue whether or not the two region counts  $N_X(B)$  and  $N_X(S \setminus B)$  are independent random variables (draw it). Intuitively, are  $N_X(B)$  and  $N_X(S \setminus B)$  positively correlated, negatively correlated or uncorrelated?

### Exercise 11

Are  $N_X(B)$  and  $N_X(S \setminus B)$  independent? Correlated?

Recall: total number of points is fixed  $k$ . And

$$N_X(B) + N_X(S \setminus B) = k.$$

So if you know  $N_X(B)$ , you *completely* know  $N_X(S \setminus B) = k - N_X(B)$ .

- Therefore, they are **not independent**.
- In fact, they are **perfectly negatively correlated**: if one is large, the other must be small.

Intuitively, you are “splitting” a fixed number of points into two bins; more in one bin means fewer in the other.

# Poisson Point Process

Wednesday, 26 November 2025 18.36

**Exercise 12.** Consider part *i*) in Definition 6. What happens with the region count  $N_X(B)$  if  $\mu_X(B) = 0$ ? Will it affect part *ii*) and how?

## Exercise 12

For a Poisson point process, what if  $\mu_X(B) = 0$ ?

From part (i) of the definition:

$$\Pr\{N_X(B) = k\} = e^{-\mu_X(B)} \frac{\mu_X(B)^k}{k!}.$$

If  $\mu_X(B) = 0$ :

- $\Pr\{N_X(B) = 0\} = e^0 \frac{0^0}{0!} = 1$  (by convention in Poisson pmf).
- For  $k \geq 1$ ,  $\Pr\{N_X(B) = k\} = 0$ .

So almost surely there are no points in  $B$ ; it is empty with probability 1.

Effect on (ii): condition "given  $N_X(B) = k$ " is irrelevant because the only possible  $k$  is 0. So part (ii) becomes trivial: the conditional binomial structure never activates for positive  $k$ .

**Exercise 13.** Let  $X \sim \text{PoissonPP}([0, 2] \times [0, 1], \rho_0)$  for some constant  $\rho_0 > 0$ . That is,  $X$  is a homogeneous Poisson point process on the bounded rectangle  $S = [0, 2] \times [0, 1]$ . What is the expected number of points from  $X$  falling in the region  $B = [0, 1] \times [0, 1]$ ? What is the probability that  $X$  has no points in  $S$  at all?

## Exercise 13

$X \sim \text{PoissonPP}([0, 2] \times [0, 1], \rho_0)$ , homogeneous with constant intensity  $\rho_0 > 0$ .

1. Expected number of points in  $B = [0, 1] \times [0, 1]$ :

Area of  $B$  is  $1 \cdot 1 = 1$ . For homogeneous intensity,

$$\mathbb{E}[N_X(B)] = \mu_X(B) = \rho_0 \cdot \text{area}(B) = \rho_0.$$

2. Probability of no points in  $S = [0, 2] \times [0, 1]$ :

Area of  $S$  is  $2 \cdot 1 = 2$ .

$N_X(S) \sim \text{Poisson}(\mu_X(S))$  with

$$\mu_X(S) = \rho_0 \cdot 2 = 2\rho_0.$$

So

$$\Pr\{N_X(S) = 0\} = e^{-2\rho_0}.$$

**Exercise 14.** Let's think in terms of computer simulation. In principle, what steps would you need to carry out if you wanted to simulate the point process  $X \sim \text{PoissonPP}(\mathbb{R}^2, 1)$ ? What is  $\mu_X(S)$  in this case?

#### Exercise 14

Simulating  $X \sim \text{PoissonPP}(\mathbb{R}^2, 1)$  in principle.

- On the *whole*  $\mathbb{R}^2$  there are infinitely many points, so

$$\mu_X(S) = \int_{\mathbb{R}^2} 1 dx = \infty.$$

So you can't simulate *all* of  $\mathbb{R}^2$ . In practice you simulate on a bounded window.

Principle for a bounded window  $W \subset \mathbb{R}^2$ :

1. Compute  $\mu_X(W) = \int_W \rho_X(x) dx$ . For intensity 1, this is just the area of  $W$ .
2. Draw  $L \sim \text{Poisson}(\mu_X(W))$ .
3. Given  $L$ , draw  $L$  points i.i.d. uniformly in  $W$ .

Pseudocode for a window  $W = [-A, A] \times [-A, A]$ :

```
pseudo
lambda = 1.0
A = 5.0
areaW = (2*A) * (2*A)

# 1. number of points
L = PoissonRandom(lambda * areaW)

# 2. locations
x = -A + 2*A * rand(L, 1)
y = -A + 2*A * rand(L, 1)

plot(x, y, '.')
axis([-A A -A A])
axis equal
```

Copy code

**Exercise 15.** Sketch a few different example realizations of the 1D point process  $Y$ . Sketch the corresponding realizations of  $Z(\cdot)$ , e.g. in the range  $t \in [0, 20]$ . Explain what will happen if  $\varrho_0$  is selected larger. Express  $Z(t)$  as a certain region count  $N_Y(B_t)$  for some suitably chosen region  $B_t$  and argue that  $Z(t)$  is Poisson distributed with mean parameter  $\varrho_0 t$ . Is  $Z(\cdot)$  a wide-sense stationary (WSS) process?

### Exercise 15 – Poisson counting process

$Y$  is 1D homogeneous Poisson PP on  $[0, \infty)$  with rate  $\rho_0$ .

Define

$$Z(t) = \sum_{y \in Y} \mathbf{1}[y \leq t], \quad t \geq 0.$$

1. Sketch: Realizations of  $Y$  are scattered points on the time axis.  $Z(t)$  is a staircase function that increases by 1 at each point and is constant between points.
2. Express as region count:

Let  $B_t = [0, t] \subset [0, \infty)$ . Then

$$Z(t) = N_Y(B_t).$$

3. Distribution of  $Z(t)$ :

For homogeneous 1D Poisson rate  $\rho_0$ ,

$$N_Y(B_t) \sim \text{Poisson}(\mu_Y(B_t)), \quad \mu_Y(B_t) = \rho_0 t.$$

So

$$Z(t) \sim \text{Poisson}(\rho_0 t).$$

4. Is  $Z(t)$  WSS (wide-sense stationary)?

- Mean:  $\mathbb{E}[Z(t)] = \rho_0 t$ , which is linear in  $t$ , not constant  $\rightarrow$  not stationary in mean, so not WSS.
- $Z(t)$  has stationary increments (distribution of  $Z(t+s) - Z(s)$  depends only on  $t$ ), but  $Z(t)$  itself is not stationary.

# Queuing theory

Wednesday, 26 November 2025 18.37

**Exercise 16.** What kind of questions do you think are typically sought to be answered in applications involving queues? *Hint:* Think of concepts such as queue lengths and customer waiting times.

## Exercise 16

Typical questions in queuing applications:

- Average queue length (mean of  $Z(t)$ ).
- Average waiting time per customer.
- Probability that the system is empty (no customers).
- Probability that queue length exceeds some threshold.
- Distribution of busy/idle periods of the server(s).
- System stability: does the queue length remain bounded over time?

All of these help in dimensioning the system (how many servers, expected delays, etc.).

**Exercise 17.** For each random point  $\mathbf{x} = (x_1, x_2) \in X$ , what is the interpretation of the random time instance  $x_1 + x_2$ ?

## Exercise 17

In the 2D queueing model, each point is  $\mathbf{x} = (x_1, x_2)$ :

- $x_1$  = arrival time of a customer.
- $x_2$  = service time required.

Then

$$x_1 + x_2 = \text{departure time of that customer.}$$

So  $x_1 + x_2$  is the time when service for that customer finishes (assuming service starts immediately at arrival, as in the  $M/G/\infty$  model).

**Exercise 18.** Sketch one example realization of the 2D point process  $X$ . Sketch the corresponding realization of  $Z(t)$ , e.g. in the range  $t \in [-20, 20]$ . Look carefully at (1.8) and express  $Z(t)$  as a certain region count  $N_X(B_t)$  for some<sup>1</sup> appropriately chosen region  $B_t \subset \mathbb{R} \times [0, \infty)$ . Recall Definition 6 and argue that  $Z(t)$  is a Poisson distributed random variable. Finally, try to calculate

$$\mathbb{E}[Z(t)] = \mathbb{E}[N_X(B_t)] = \mu_X(B_t) = \int_{B_t} \varrho_X(\mathbf{x}) d\mathbf{x} = \iint_{B_t} \lambda p(x_2) dx_1 dx_2,$$

and discuss whether you find it reasonable that this mean function does not depend on time  $t$ .

*Final remark:* It can be shown that the random process  $Z(\cdot)$  in (1.8) is in fact strict-sense stationary (SSS).

#### Exercise 18 – M/G/ $\infty$ queue via PPP

We have  $X \sim \text{PoissonPP}(\mathbb{R} \times [0, \infty), \rho_X)$  with

$$\rho_X(x_1, x_2) = \lambda p(x_2),$$

where  $p$  is a pdf on  $[0, \infty)$ .

Define

$$Z(t) = \sum_{x \in X} \mathbf{1}[x_1 \leq t, x_1 + x_2 > t], \quad t \in \mathbb{R}.$$

1. **Interpretation:** A customer contributes 1 to  $Z(t)$  if they have arrived ( $x_1 \leq t$ ) but not yet finished service ( $x_1 + x_2 > t$ ). So  $Z(t)$  is the **number of customers in service at time  $t$**  (queue length for this infinite-server system).
2. **Express  $Z(t)$  as region count  $N_X(B_t)$ :**

The condition  $\{x_1 \leq t, x_1 + x_2 > t\}$  defines a region

$$B_t = \{(x_1, x_2) : x_1 \leq t, x_1 + x_2 > t, x_2 \geq 0\}.$$

Then

$$Z(t) = N_X(B_t).$$

**3. Show  $Z(t)$  is Poisson distributed:**

For a PoissonPP, any region count  $N_X(B_t)$  is Poisson with mean

$$\mu_X(B_t) = \int_{B_t} \rho_X(x_1, x_2) dx_1 dx_2 = \int_{B_t} \lambda p(x_2) dx_1 dx_2.$$

**4. Compute the mean and show it doesn't depend on  $t$ :**

The region  $B_t$  can be described as:

- $x_2 \geq 0,$
- $x_1 \in (t - x_2, t].$

So for any fixed  $x_2$ , the length of the allowed  $x_1$ -interval is  $x_2$ .

Thus,

$$\mu_X(B_t) = \int_0^\infty \int_{t-x_2}^t \lambda p(x_2) dx_1 dx_2 = \int_0^\infty \lambda p(x_2) (x_2) dx_2 = \lambda \mathbb{E}[X_2],$$

where  $X_2$  is a random variable with pdf  $p$ .

This expression does not depend on  $t$ . Therefore,

- $Z(t) \sim \text{Poisson}(\lambda \mathbb{E}[X_2]),$
- $\mathbb{E}[Z(t)] = \lambda \mathbb{E}[X_2]$  for all  $t$ .

That matches the intuition: average number in system = arrival rate  $\times$  mean service time (Little's law for this simple model).

# Campbell

Wednesday, 26 November 2025 18.38

**Exercise 19.** Use Campbell's Theorem to calculate the mean  $\mathbb{E}[Z(t)]$  of the shot-noise random processes defined in (1.7) and in (1.8).

## Exercise 19 – Campbell's Theorem

Campbell's Theorem says:

$$\mathbb{E}\left[\sum_{x \in X} g(x)\right] = \int_S g(x)\rho_X(x) dx.$$

### (a) Process in (1.7) – Poisson counting process

Here  $Y$  is 1D Poisson PP on  $[0, \infty)$  with constant intensity  $\rho_0$ , and

$$Z(t) = \sum_{y \in Y} \mathbf{1}[y \leq t].$$

View this as a point process on  $S = [0, \infty)$  and set

$$g(y) = \mathbf{1}[y \leq t].$$

Then

$$\mathbb{E}[Z(t)] = \int_0^\infty \mathbf{1}[y \leq t] \rho_0 dy = \rho_0 \int_0^t dy = \rho_0 t,$$

which matches the earlier result.

### (b) Process in (1.8) – M/G/ $\infty$ queue

Here  $X$  is a 2D PoissonPP on  $\mathbb{R} \times [0, \infty)$  with

$$\rho_X(x_1, x_2) = \lambda p(x_2),$$

and

$$Z(t) = \sum_{x \in X} \mathbf{1}[x_1 \leq t, x_1 + x_2 > t].$$

Let

$$g(x_1, x_2) = \mathbf{1}[x_1 \leq t, x_1 + x_2 > t].$$

Then by Campbell:

$$\mathbb{E}[Z(t)] = \int_{\mathbb{R}} \int_0^\infty g(x_1, x_2) \lambda p(x_2) dx_2 dx_1 = \int_{B_t} \lambda p(x_2) dx_1 dx_2,$$

where  $B_t$  is as in Exercise 18. As we already computed:

$$\mathbb{E}[Z(t)] = \lambda \mathbb{E}[X_2],$$

independent of  $t$ .

So Campbell's Theorem neatly reproduces the mean functions for both models.

# Eksam 2023

Wednesday, 26 November 2025 19.09

## Problem 1:

The process  $U(n)$  is an i.i.d. random process with samples drawn identically and independently from the following probability mass function (pmf):

$$p_U(u) = \begin{cases} 1/3 & \text{for } u = 1 \\ 1/3 & \text{for } u = 0 \\ 1/3 & \text{for } u = -1 \\ 0 & \text{otherwise} \end{cases}.$$

**1.1** Calculate the mean function and the autocorrelation function (ACF) of  $U(n)$ .

### Problem 1

We have an i.i.d. process  $U(n)$  with pmf

$$p_U(u) = \begin{cases} 1/3, & u = 1, 0, -1, \\ 0, & \text{otherwise.} \end{cases}$$

#### 1.1 Mean and ACF of $U(n)$

Mean:

$$\mathbb{E}[U(n)] = \sum_u u p_U(u) = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + (-1) \cdot \frac{1}{3} = 0.$$

So the mean function is

$$m_U(n) = \mathbb{E}[U(n)] = 0 \quad \forall n.$$

Autocorrelation:

Because  $U(n)$  is i.i.d.:

- For  $k = 0$ :

$$R_U(0) = \mathbb{E}[U(n)^2] = 1^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + (-1)^2 \cdot \frac{1}{3} = \frac{2}{3}.$$

- For  $k \neq 0$ :  $U(n)$  and  $U(n+k)$  are independent, so

$$R_U(k) = \mathbb{E}[U(n)]\mathbb{E}[U(n+k)] = 0 \cdot 0 = 0.$$

So

$$R_U(k) = \begin{cases} \frac{2}{3}, & k = 0, \\ 0, & k \neq 0. \end{cases}$$

**1.2** Is  $U(n)$  a wide-sense stationary (WSS) process? Is it strict-sense stationary (SSS)? Justify your answers.

## 1.2 Is $U(n)$ WSS? SSS?

- **WSS:**

Mean is constant (0) and ACF depends only on lag  $k$ , not on  $n$ .

$\Rightarrow U(n)$  is WSS.

- **SSS (strict-sense stationary):**

All samples are independent and identically distributed, so any finite collection

$(U(n_1), \dots, U(n_m))$  has the same joint pdf as  $(U(n_1 + k), \dots, U(n_m + k))$ .

$\Rightarrow U(n)$  is SSS as well.

Next, consider the following process built from the samples of  $U(n)$ :

$$X(n) = \sum_{i=0}^n U(i), \quad n = 0, 1, \dots$$

**1.3** Draw a sketch of a 10-sample long realization of  $U(n)$  and of the corresponding realization of  $X(n)$ .

Now define the partial-sum process

$$X(n) = \sum_{i=0}^n U(i), \quad n = 0, 1, \dots$$

(so  $X(n)$  is a random walk built from the  $U(i)$ ).

## 1.3 Sketch realizations of $U(n)$ and $X(n)$

- A realization of  $U(n)$  is just a sequence of values  $-1, 0, 1$  in time.
- The corresponding  $X(n)$  starts at  $X(0) = U(0)$  and then moves up by  $+1$ , stays, or moves down by  $-1$  at each step, accumulating the values.

So in a sketch:

- First row: bars or points at -1, 0, 1.
- Second row: a "zig-zagging" cumulative sum curve.

(No formulas to derive here, just the concept.)

**1.4** Calculate the mean and variance functions of  $X(n)$ .

## 1.4 Mean and variance of $X(n)$

We have

$$X(n) = \sum_{i=0}^n U(i),$$

a sum of  $n + 1$  independent, identically distributed random variables.

**Mean:**

$$\mathbb{E}[X(n)] = \sum_{i=0}^n \mathbb{E}[U(i)] = (n + 1) \cdot 0 = 0.$$

**Variance:**

Because the  $U(i)$  are independent and each has variance  $2/3$ ,

$$\text{Var}(X(n)) = \sum_{i=0}^n \text{Var}(U(i)) = (n + 1) \cdot \frac{2}{3} = \frac{2}{3}(n + 1).$$

So

$$m_X(n) = 0, \quad \sigma_X^2(n) = \frac{2}{3}(n + 1).$$

**1.5** Based on the results of the previous item, is  $X(n)$  WSS or SSS? Why?

## 1.5 Is $X(n)$ WSS or SSS?

- For **WSS**, we need constant mean and variance, and ACF depending only on lag.  
Here, mean is constant (0) but variance grows with  $n$ :  $\frac{2}{3}(n + 1)$ .  
 $\Rightarrow X(n)$  is not WSS.
- For **SSS**, all finite-dimensional distributions must be invariant to shifts in time.  
But the variance of  $X(n)$  clearly depends on  $n$ ; e.g.,  $\text{Var}(X(1)) \neq \text{Var}(X(10))$ .  
That already shows the distributions change with  $n$ .  
 $\Rightarrow X(n)$  is not SSS either.

Finally, consider a WSS continuous-time process  $Y(t)$  with an ACF reading  $R_Y(\tau) = \cos(\pi\tau)$ .

**1.6** Calculate and sketch the power spectral density (PSD) of the process.

Now consider a continuous-time WSS process  $Y(t)$  with ACF

$$R_Y(\tau) = \cos(\pi\tau).$$

### 1.6 PSD of $Y(t)$

Recall: for a WSS continuous-time process, the PSD is the Fourier transform of the ACF:

$$S_Y(f) = \int_{-\infty}^{\infty} R_Y(\tau) e^{-j2\pi f\tau} d\tau.$$

But we can use the known transform:

$$\cos(2\pi f_0\tau) \leftrightarrow \frac{1}{2} [\delta(f - f_0) + \delta(f + f_0)].$$

Here,

$$R_Y(\tau) = \cos(\pi\tau) = \cos\left(2\pi \cdot \frac{1}{2}\tau\right),$$

so  $f_0 = \frac{1}{2}$ . Therefore,

$$S_Y(f) = \frac{1}{2} \delta\left(f - \frac{1}{2}\right) + \frac{1}{2} \delta\left(f + \frac{1}{2}\right).$$

**Sketch:** two impulses (spikes):

- One at  $f = +0.5$  with weight  $1/2$ .
- One at  $f = -0.5$  with weight  $1/2$ .

### 1.7 Calculate the covariance matrix of the random vector $\mathbf{Y} = [Y(0), Y(1), Y(5/2)]^T$ .

#### 1.7 Covariance matrix of $\mathbf{Y} = [Y(0), Y(1), Y(5/2)]^T$

Assuming mean zero, the covariance entries are

$$\text{Cov}(Y(t_i), Y(t_j)) = R_Y(t_i - t_j) = \cos(\pi(t_i - t_j)).$$

Let  $t_1 = 0, t_2 = 1, t_3 = 5/2$ .

Compute needed ACF values:

- $R_Y(0) = \cos(0) = 1$ .
- $R_Y(1) = \cos(\pi) = -1$ .
- $R_Y(2.5) = \cos(2.5\pi) = \cos(2\pi + \frac{\pi}{2}) = \cos(\frac{\pi}{2}) = 0$ .
- $R_Y(1.5) = \cos(1.5\pi) = \cos(3\pi/2) = 0$ .

Using symmetry  $R_Y(-\tau) = R_Y(\tau)$ :

The covariance matrix is

$$\Sigma_Y = \begin{bmatrix} R(0) & R(1) & R(2.5) \\ R(1) & R(0) & R(1.5) \\ R(2.5) & R(1.5) & R(0) \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

**Problem 2:**

A white Gaussian noise (WGN) process with variance  $\sigma^2$  is input to a linear time invariant (LTI) system with impulse response

$$h(n) = \delta(n) + a\delta(n-2)$$

where  $a$  is an unknown constant. The output process,  $X(n)$ , is known to have the following PSD and ACF properties:

$$S_X(1/8) = 4, \quad R_X(2) = -2. \quad (1)$$

**2.1** Find the values of the noise variance  $\sigma^2$  and the coefficient  $a$  so that the conditions in (1) are fulfilled.

**2.1 Find  $\sigma^2$  and  $a$**

The discrete-time frequency response:

$$H(f) = 1 + ae^{-j4\pi f}$$

(because shift by 2 samples gives factor  $e^{-j2\pi f \cdot 2} = e^{-j4\pi f}$ ).

For WGN of variance  $\sigma^2$ , input PSD is  $S_{\text{in}}(f) = \sigma^2$ ; output PSD

$$S_X(f) = |H(f)|^2 \sigma^2.$$

At  $f = 1/8$ , we have  $e^{-j4\pi(1/8)} = e^{-j\pi/2} = -j$ . So

$$H(1/8) = 1 + a(-j) = 1 - ja,$$

$$|H(1/8)|^2 = 1 + a^2.$$

Given  $S_X(1/8) = 4$ ,

$$\sigma^2(1 + a^2) = 4. \quad (1)$$

Now use the output ACF. For WGN input, the output ACF is

$$R_X(k) = \sigma^2 \sum_n h(n)h(n+k).$$

Here  $h(0) = 1$ ,  $h(2) = a$ , others 0.

For  $k = 2$ ,

$$R_X(2) = \sigma^2[h(0)h(2) + h(2)h(4)] = \sigma^2(1 \cdot a + a \cdot 0) = a\sigma^2.$$

Given  $R_X(2) = -2$ ,

$$a\sigma^2 = -2. \quad (2)$$

From (2):  $a = -2/\sigma^2$ . Plug into (1):

$$\sigma^2 \left(1 + \frac{4}{\sigma^4}\right) = 4 \quad \Rightarrow \quad \sigma^4 - 4\sigma^2 + 4 = 0.$$

Treat  $x = \sigma^2$ :

$$x^2 - 4x + 4 = 0 \Rightarrow (x - 2)^2 = 0 \Rightarrow x = 2.$$

So  $\sigma^2 = 2$  and then  $a = -2/\sigma^2 = -1$ .

$$\boxed{\sigma^2 = 2, \quad a = -1.}$$

## 2.2 What type of ARMA process is $X(n)$ ?

### 2.2 Type of ARMA process of $X(n)$

With those parameters,

$$h(n) = \delta(n) - \delta(n - 2),$$

$$H(z) = 1 - z^{-2}.$$

This is a **finite impulse response** filter (only taps at 0 and 2), with no feedback part, so the output is:

- AR order  $p = 0$
- MA order  $q = 2$  (with one coefficient zero between 0 and 2, but still MA(2))

So  $X(n)$  is an MA(2) process (in ARMA notation: ARMA(0,2)).

## 2.3 Calculate and sketch the full ACF and PSD of $X(n)$ .

### 2.3 Full ACF and PSD of $X(n)$

We already know how to get the ACF:

$$R_X(k) = \sigma^2 \sum_n h(n)h(n+k).$$

Nonzero values of  $h$ :  $h(0) = 1$ ,  $h(2) = -1$ ,  $\sigma^2 = 2$ .

- $k = 0$ :

$$R_X(0) = \sigma^2(h(0)^2 + h(2)^2) = 2(1 + 1) = 4.$$

- $k = 2$ :

$$R_X(2) = \sigma^2(h(0)h(2)) = 2(1 \cdot -1) = -2.$$

- $k = -2$ : for real sequences,  $R_X(-2) = R_X(2) = -2$ .
- All other lags: no overlapping nonzero taps  $\rightarrow$  sum is 0.

So

$$R_X(k) = \begin{cases} 4, & k = 0, \\ -2, & k = \pm 2, \\ 0, & \text{otherwise.} \end{cases}$$

PSD: either:

$$S_X(f) = \sum_{k=-\infty}^{\infty} R_X(k)e^{-j2\pi fk} = 4 - 2e^{-j4\pi f} - 2e^{j4\pi f} = 4 - 4\cos(4\pi f),$$

or via  $S_X(f) = \sigma^2|1 - e^{-j4\pi f}|^2$  with  $\sigma^2 = 2$ :

$$|1 - e^{-j4\pi f}|^2 = 2 - 2\cos(4\pi f),$$

$$S_X(f) = 2(2 - 2\cos(4\pi f)) = 4 - 4\cos(4\pi f) = 8\sin^2(2\pi f).$$

Sketch: periodic in  $f$  with period 0.5, zero at  $f = 0, \pm 0.5, \dots$ , and peaks 8 at  $f = \pm 0.25, \dots$

Next, let  $Y(n)$  be an autorregressive process given by

$$Y(n) = Z(n) + \frac{1}{3}Y(n-1) + \frac{2}{9}Y(n-2)$$

where  $Z(n)$  is WGN with unit variance.

**2.4** Use the Yule-Walker equations to pose a system of equations to calculate the ACF of  $Y(n)$  for time lags  $k = 0, 1, 2$ . (Just pose the system of equations, you don't need to solve it).

Now another process:

$$Y(n) = Z(n) + \frac{1}{3}Y(n-1) + \frac{2}{9}Y(n-2),$$

where  $Z(n)$  is WGN with unit variance.

This is an AR(2) process with coefficients

$$\phi_1 = 1/3, \phi_2 = 2/9.$$

## 2.4 Yule–Walker equations for $R_Y(0), R_Y(1), R_Y(2)$

For an AR(2) process:

- For  $k \geq 1$ :

$$R_Y(k) = \phi_1 R_Y(k-1) + \phi_2 R_Y(k-2).$$

- For  $k = 0$ :

$$R_Y(0) = \phi_1 R_Y(1) + \phi_2 R_Y(2) + \sigma_Z^2,$$

where  $\sigma_Z^2$  is the variance of  $Z(n)$  (here 1).

So with numbers:

1.  $k = 0$ :

$$R_Y(0) = \frac{1}{3}R_Y(1) + \frac{2}{9}R_Y(2) + 1.$$

2.  $k = 1$ :

$$R_Y(1) = \frac{1}{3}R_Y(0) + \frac{2}{9}R_Y(1).$$

3.  $k = 2$ :

$$R_Y(2) = \frac{1}{3}R_Y(1) + \frac{2}{9}R_Y(0).$$

That's the required system of three equations in the three unknowns  $R_Y(0), R_Y(1), R_Y(2)$ .

- 2.5** Assuming that a 100-sample long realization of  $Y(n)$  is available, explain with pseudocode how it can be used to **estimate** the ACF of the process for time lags  $k = -99, -98, \dots, 0, \dots, 98, 99$ .

## 2.5 Pseudocode to estimate the ACF of $Y(n)$ from 100 samples

We have one realization  $y[0], \dots, y[99]$ , length  $N = 100$ .

A common (unbiased) estimator for lag  $k$  is

$$\hat{R}_Y(k) = \frac{1}{N - |k|} \sum_{n=0}^{N-1-|k|} y[n] y[n + k], \quad k \geq 0,$$

and for negative lags,  $\hat{R}_Y(-k) = \hat{R}_Y(k)$  since the process is real and WSS.

Pseudocode:

```
pseudo Copy code
N = 100
# y[0..99] contains the realization of Y(n)

# allocate array for lags -99..99
for k = -99 to 99:
    Rhat[k] = 0

# compute for non-negative lags
for k = 0 to 99:
    sum = 0
    for n = 0 to N-1-k:
        sum = sum + y[n] * y[n + k]
    Rhat[k] = sum / (N - k)

# mirror for negative lags
for k = 1 to 99:
    Rhat[-k] = Rhat[k]
```

After this,  $Rhat[k]$  approximates  $R_Y(k)$  for all  $k$  from  $-99$  to  $99$ .

## Problem 3:

The position of sensor nodes in an Internet of Things (IoT) network is modeled using a two-dimensional Poisson point process  $X \sim \text{PoissonPP}(S, \varrho)$  defined in a rectangular area  $S = [-2, 2] \times [-3, 3]$  and with intensity function

$$\varrho(x, y) = \begin{cases} \frac{1}{4}x^2 & \text{if } (x, y) \in S \\ 0 & \text{otherwise.} \end{cases}$$

**3.1** Calculate the expected number of sensor nodes within the area  $S$ .

### Problem 3 – 2D Poisson point process

We have  $X \sim \text{PoissonPP}(S, \lambda(x, y))$  on

$$S = [-2, 2] \times [-3, 3], \quad \lambda(x, y) = \frac{1}{4}x^2 \text{ on } S, \text{ 0 outside.}$$

#### 3.1 Expected number of sensors in $S$

For a Poisson point process, the expected number of points in region  $B$  is

$$\mu_X(B) = \int_B \lambda(x, y) dx dy.$$

So for all of  $S$ :

$$\begin{aligned}\mathbb{E}[N_X(S)] &= \int_{-2}^2 \int_{-3}^3 \frac{1}{4}x^2 dy dx \\ &= \int_{-2}^2 \frac{1}{4}x^2 \cdot (6) dx = \frac{6}{4} \int_{-2}^2 x^2 dx \\ &= \frac{3}{2} \left[ \frac{x^3}{3} \right]_{-2}^2 = \frac{3}{2} \cdot \frac{8 - (-8)}{3} = \frac{3}{2} \cdot \frac{16}{3} = 8.\end{aligned}$$

So the expected number is

$$\boxed{\mathbb{E}[N_X(S)] = 8.}$$

#### 3.2 Sketch a possible realization of the point process $X$ in $S$ . Annotate your sketch to explain in which areas you expect the density of nodes to be highest.

##### 3.2 Sketch a realization and where density is highest

$\lambda(x, y) = \frac{1}{4}x^2$  depends only on  $x$ , and increases with  $|x|$ .

- Near  $x = 0$ :  $\lambda \approx 0 \rightarrow$  few nodes.
- Near  $x = \pm 2$ :  $x^2$  is largest  $\rightarrow$  highest node density in two vertical stripes along the left and right edges of the rectangle.
- Distribution is symmetric in  $x$  and uniform in  $y$  (for each fixed  $x$ ).

So in a sketch, you'd draw:

- Sparse points around the vertical line  $x = 0$ ,
- Dense clusters near  $x = -2$  and  $x = 2$ , spread evenly in the  $y$  direction.

Next, we define two equally sized subregions of  $S$  as:

$$\begin{aligned}B_1 &= (0, 2] \times (0, 3] \\ B_2 &= [-2, 0) \times [-3, 0)\end{aligned}$$

#### 3.3 What is the probability that region $B_1$ contains more than 1 sensor?

### 3.3 Probability that $B_1$ contains more than 1 sensor

Regions:

$$B_1 = (0, 2] \times (0, 3].$$

First find its mean number of points:

$$\begin{aligned}\mu_X(B_1) &= \int_0^2 \int_0^3 \frac{1}{4}x^2 dy dx \\ &= \int_0^2 \frac{1}{4}x^2 \cdot 3 dx = \frac{3}{4} \int_0^2 x^2 dx \\ &= \frac{3}{4} \left[ \frac{x^3}{3} \right]_0^2 = \frac{3}{4} \cdot \frac{8}{3} = 2.\end{aligned}$$

Thus  $N_X(B_1) \sim \text{Poisson}(2)$ .

We want  $P(N_X(B_1) > 1)$ :

$$\begin{aligned}P(N > 1) &= 1 - P(N = 0) - P(N = 1) \\ &= 1 - e^{-2} \frac{2^0}{0!} - e^{-2} \frac{2^1}{1!} \\ &= 1 - e^{-2}(1 + 2) \\ &= 1 - 3e^{-2}.\end{aligned}$$

So

$$P(N_X(B_1) > 1) = 1 - 3e^{-2}.$$

**3.4** Assume we know that region  $B_2$  contains 7 sensors. Under this condition, what is now the probability that  $B_1$  contains more than 1 sensor?

### 3.4 Given $B_2$ contains 7 sensors, what about $B_1$ ?

$$B_2 = [-2, 0) \times [-3, 0),$$

and  $B_1$  and  $B_2$  are disjoint.

For a Poisson point process, counts in disjoint regions are independent. So knowing that  $B_2$  contains 7 points does not change the distribution of the number of points in  $B_1$ .

Therefore, conditionally on " $B_2$  contains 7 sensors", we still have

$N_X(B_1) \sim \text{Poisson}(2)$ , and

$$P(N_X(B_1) > 1 | N_X(B_2) = 7) = 1 - 3e^{-2}.$$

To conclude, each of the sensors transmits an amount of data packets per unit time that depends on its coordinates according to the function  $f(x, y) = \frac{3y}{x}$ . Thus, the total number of data packets transmitted per unit time in the whole network is

$$N_p = \sum_{(x,y) \in X} \frac{3y}{x}$$

**3.5** Calculate the expected value of the random variable  $N_p$ . (*Hint: use Campbell's theorem*).

### 3.5 Expected total number of packets per unit time, $N_p$

Each sensor at position  $(x, y)$  transmits  $f(x, y) = \frac{3y}{x}$  packets per unit time, and

$$N_p = \sum_{(x,y) \in X} \frac{3y}{x}.$$

Campbell's theorem for a Poisson point process says:

$$\mathbb{E} \left[ \sum_{(x,y) \in X} g(x, y) \right] = \int_S g(x, y) \lambda(x, y) dx dy.$$

Here  $g(x, y) = \frac{3y}{x}$ ,  $\lambda(x, y) = \frac{1}{4}x^2$ . So

$$g(x, y)\lambda(x, y) = \frac{3y}{x} \cdot \frac{1}{4}x^2 = \frac{3}{4}xy.$$

Thus

$$\begin{aligned} \mathbb{E}[N_p] &= \int_{-2}^2 \int_{-3}^3 \frac{3}{4}xy dy dx \\ &= \frac{3}{4} \int_{-2}^2 x \left( \int_{-3}^3 y dy \right) dx. \end{aligned}$$

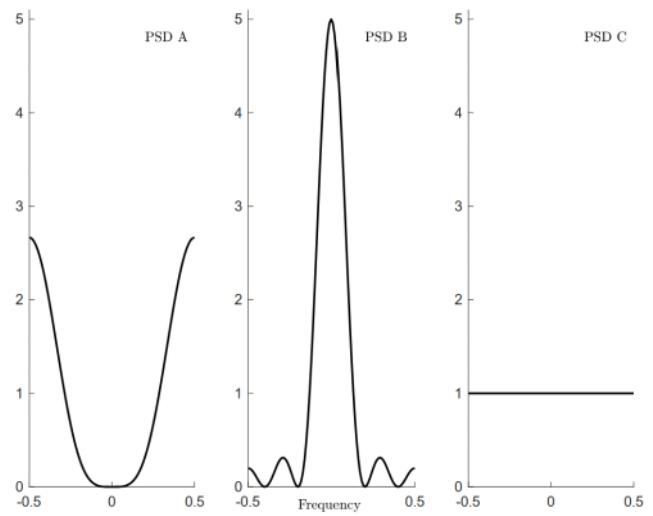
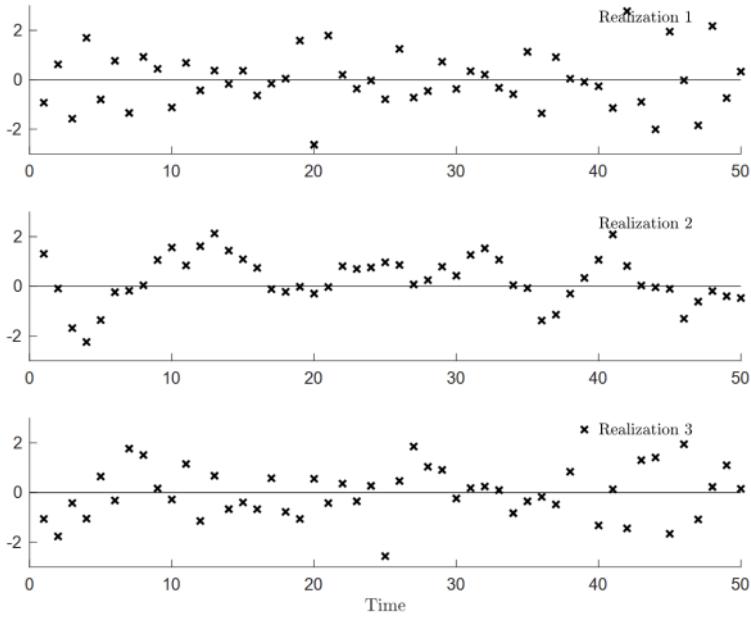
But

$$\int_{-3}^3 y dy = 0$$

(because the integrand is odd over a symmetric interval). Hence

$$\boxed{\mathbb{E}[N_p] = 0.}$$

Intuitively, positive- $y$  sensors contribute positive packets, negative- $y$  sensors contribute negative packets in a perfectly symmetric way, so they cancel out in expectation.



- 1.1** Match each of the three realizations to one of the three PSD's. For each pair, explain why you have grouped together that particular realization with that particular PSD.

#### Problem 1

We have three zero-mean, unit-variance WSS processes with three different PSDs (A,B,C), and one discrete-time autocorrelation  $R_x(k)$ .

##### 1.1 Matching realizations and PSDs

Use the rule of thumb:

- Low-pass PSD (most energy at low frequencies) → realization is smooth / slowly varying.
- High-pass PSD (energy away from DC) → realization is fast oscillatory, alternates sign a lot.
- Flat PSD (white noise) → realization is rough, but with no obvious smooth trend or oscillation pattern.

From the plots (page 2–3):

- PSD B: tall narrow peak at  $f = 0 \rightarrow$  strongly low-pass.  
⇒ must correspond to the smoothest realization → Realization 2.
- PSD A: "U-shaped": low near  $f = 0$ , higher near  $|f| = 0.5 \rightarrow$  high-pass.  
⇒ corresponds to the realization with clear fast sign changes / oscillations → Realization 3.
- PSD C: constant level for all  $f \rightarrow$  white noise.  
⇒ corresponds to realization that looks "pure random" but not especially smooth nor especially oscillatory → Realization 1.

So the matching is:

- Realization 1 ↔ PSD C
- Realization 2 ↔ PSD B
- Realization 3 ↔ PSD A

- 1.2** It can be identified directly from PSD A, PSD B and PSD C, that all three of them correspond to processes with zero mean. Explain how.

#### 1.2 How do we see that the mean is zero from the PSD?

For a discrete-time WSS process,

$$\text{non-zero mean } m \neq 0 \Rightarrow S_x(f) \text{ has a Dirac impulse } m^2 \delta(f) \text{ at } f = 0.$$

All three PSDs A,B,C on page 3 are continuous curves without an impulse spike at  $f = 0$ .

Hence none of the processes has such a DC impulse → all have zero mean.

- 1.3** Which of the power spectral densities A, B or C corresponds to the autocorrelation function  $R_x(k)$  shown in the figure below? Justify your answer.

### 1.3 Which PSD gives the shown autocorrelation $R_X(k)$ ?

The figure (page 3) shows

$$R_X(0) = 1, \quad R_X(1) = R_X(-1) = -\frac{1}{4}, \quad R_X(k) = 0 \text{ for } |k| \geq 2.$$

The PSD is the discrete-time Fourier transform

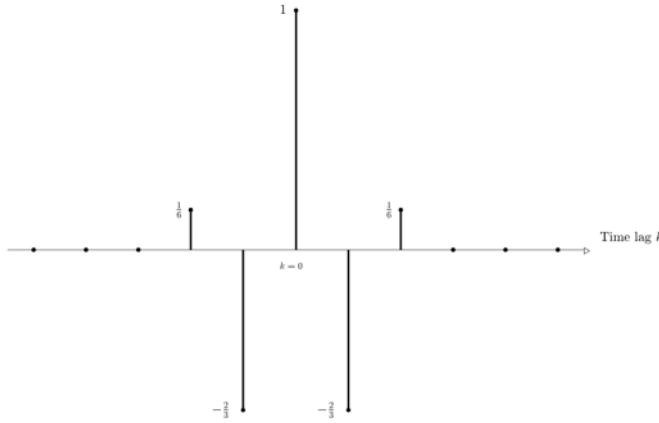
$$S_X(f) = \sum_{k=-\infty}^{\infty} R_X(k) e^{-j2\pi fk} = 1 - \frac{1}{4}e^{-j2\pi f} - \frac{1}{4}e^{j2\pi f} = 1 - \frac{1}{2}\cos(2\pi f).$$

At  $f = 0$ :  $S_X(0) = 1 - \frac{1}{2} = 0.5$ ;

at  $f = \pm 0.5$ :  $S_X(0.5) = 1 - \frac{1}{2}\cos(\pi) = 1 + \frac{1}{2} = 1.5$ .

So the PSD is small near  $f = 0$  and larger towards  $|f| = 0.5$  (a high-pass shape).

That is exactly PSD A.



### 1.4 Assume now that $\{X(n)\}$ is a zero mean Gaussian process with autocorrelation function $R_x(k)$ as given in the above figure. State the pdf of the vector $\mathbf{X}$ defined as

$$\mathbf{X} = \begin{bmatrix} X(23) \\ X(24) \end{bmatrix}.$$

#### 1.4 Pdf of $X = [X(23) \ X(24)]^T$

Assume  $\{X(n)\}$  is zero-mean Gaussian with the above  $R_X(k)$ .

- $E[X(23)] = E[X(24)] = 0$ .
- $\text{Var}(X(23)) = \text{Var}(X(24)) = R_X(0) = 1$ .
- $\text{Cov}(X(23), X(24)) = R_X(1) = -1/4$ .

So  $\mathbf{X}$  is bivariate Gaussian:

$$X \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma_X\right), \quad \Sigma_X = \begin{bmatrix} 1 & -\frac{1}{4} \\ -\frac{1}{4} & 1 \end{bmatrix}.$$

The pdf is

$$f_X(\mathbf{x}) = \frac{1}{2\pi\sqrt{\det\Sigma_X}} \exp\left(-\frac{1}{2}\mathbf{x}^T \Sigma_X^{-1} \mathbf{x}\right), \quad \mathbf{x} \in \mathbb{R}^2.$$

(You get full credit already by identifying it as a zero-mean Gaussian with that covariance.)

### 1.5 Calculate the mean and covariance of the vector $\mathbf{Y}$ defined as

$$\mathbf{Y} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \mathbf{X}.$$

#### 1.5 Mean and covariance of $Y$

Given

$$Y = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} X \quad \Rightarrow \quad \begin{cases} Y_1 = X(23), \\ Y_2 = X(23) + X(24). \end{cases}$$

- Mean:

$$E[Y] = A E[X] = A \cdot 0 = 0, \quad A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

- Covariance:

$$\Sigma_Y = A \Sigma_X A^T.$$

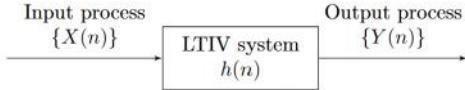
Using  $\Sigma_X$  above,

$$\Sigma_Y = \begin{bmatrix} 1 & 3/4 \\ 3/4 & 3/2 \end{bmatrix}.$$

So

- $\text{Var}(Y_1) = 1$ ,
- $\text{Var}(Y_2) = 3/2$ ,
- $\text{Cov}(Y_1, Y_2) = 3/4$ .

Consider the linear time-invariant (LTIV) system



where  $X(n)$  is a discrete-time WSS process with zero mean and autocorrelation function  $R_x(k)$ . The transfer function of the system is

$$H(f) = 1 - \frac{3}{4} \exp(-j4\pi f), \quad -\frac{1}{2} \leq f < \frac{1}{2}.$$

**2.1** Compute the mean of the output process  $\{Y(n)\}$ .

#### Problem 2 – LTIV system

Input  $\{X(n)\}$ : zero-mean WSS with autocorrelation  $R_x(k)$ .

LTI system with frequency response

$$H(f) = 1 - \frac{3}{4} e^{-j4\pi f}, \quad -\frac{1}{2} \leq f < \frac{1}{2}.$$

Output is  $\{Y(n)\}$ .

#### 2.1 Mean of $Y(n)$

For an LTI system, the output mean is

$$\mathbb{E}[Y(n)] = \left( \sum_k h(k) \right) \mathbb{E}[X(n)].$$

Here  $\mathbb{E}[X(n)] = 0$ , so

$$\boxed{\mathbb{E}[Y(n)] = 0.}$$

**2.2** Justify without calculations that  $\{Y(n)\}$  is WSS.

#### 2.2 Why is $\{Y(n)\}$ WSS?

Standard result: the output of a linear time-invariant system with WSS input is also WSS (provided  $h(n)$  is absolutely summable).

Because:

- $Y(n)$  is a linear combination of shifted versions of  $X(n)$ .
- Time invariance makes the second-order statistics of  $Y(n)$  depend only on time differences, not absolute time.

Thus  $R_Y(k)$  depends only on the lag  $k \rightarrow Y(n)$  is WSS.

**2.3** Express the autocorrelation function  $R_Y(k)$  of the output process  $\{Y(n)\}$  in terms of  $R_X(k)$ .

#### 2.3 Autocorrelation $R_Y(k)$ in terms of $R_X(k)$

Let

$$y(n) = \sum_{\ell} h(\ell) x(n-\ell).$$

Then

$$R_Y(k) = \mathbb{E}[y(n+k)y(n)] = \sum_{\ell} \sum_m h(\ell) h(m) R_X(k+\ell-m).$$

This is the standard expression:

$$\boxed{R_Y(k) = \sum_{\ell} \sum_m h(\ell) h(m) R_X(k+\ell-m).}$$

(Equivalently:  $R_Y = R_X * (h * h^*)$ , with  $h^*(n) = h(-n)$ .)

**2.4** Let  $\delta$  be the Kronecker delta. State at least two reasons why  $R_Y(k)$  cannot be of the form

## 2.4 Why can't $R_Y(k) = -3\delta(k+1) - \delta(k) + \delta(k-2)$ ?

Two simple reasons:

1. Variance must be non-negative:

$$R_Y(0) = \mathbb{E}[Y(n)^2] \geq 0.$$

But plugging  $k=0$ :

$$R_Y(0) = -3\delta(1) - \delta(0) + \delta(-2) = -1 < 0$$

which is impossible.

2. Autocorrelation of a real WSS process is even:

For real processes,  $R_Y(-k) = R_Y(k)$ .

Check:

- $R_Y(-1) = -3\delta(0) - \delta(-1) + \delta(-3) = -3$ ,
- $R_Y(1) = -3\delta(2) - \delta(1) + \delta(-1) = 0$ .

So  $R_Y(-1) \neq R_Y(1) \rightarrow$  not even  $\rightarrow$  cannot be a valid autocorrelation.

(One could also mention that a valid PSD must be non-negative; the Fourier transform of that function would violate that.)

$$R_Y(k) = -3\delta(k+1) - \delta(k) + \delta(k-2)$$

Notice that the above equation is *not* true.

## 2.5 Suppose that the input $\{X(n)\}$ is a Gaussian process with

### 2.5 PSD of $Y(n)$ when $R_X(k) = 2\delta(k)$

Here the input is white noise with variance 2:

$$S_X(f) = 2, \quad -\frac{1}{2} \leq f < \frac{1}{2}.$$

For an LTI system:

$$S_Y(f) = |H(f)|^2 S_X(f).$$

First find  $|H(f)|^2$ :

$$\begin{aligned} |H(f)|^2 &= \left(1 - \frac{3}{4}e^{-j4\pi f}\right)\left(1 - \frac{3}{4}e^{j4\pi f}\right) \\ &= 1 + \left(\frac{3}{4}\right)^2 - \frac{3}{4}(e^{j4\pi f} + e^{-j4\pi f}) \\ &= 1 + \frac{9}{16} - \frac{3}{2}\cos(4\pi f) \\ &= \frac{25}{16} - \frac{3}{2}\cos(4\pi f). \end{aligned}$$

Thus

$$S_Y(f) = 2 \left( \frac{25}{16} - \frac{3}{2}\cos(4\pi f) \right) = \frac{25}{8} - 3\cos(4\pi f), \quad -\frac{1}{2} \leq f < \frac{1}{2}.$$

Sketch: a cosine-shaped curve, periodic in  $f$  with period 0.5, minimum  $S_Y^{\min} = 1/8$ , maximum  $S_Y^{\max} = 49/8$ .

## 2.6 Identify whether the output is an AR( $p$ ), MA( $q$ ) or ARMA( $p, q$ ) process. Give the values for $p$ and $q$ .

### 2.6 Is the output AR, MA, or ARMA?

In the time domain,

$$H(z) = 1 - \frac{3}{4}z^{-4} \Rightarrow y(n) = x(n) - \frac{3}{4}x(n-4),$$

when  $x(n)$  is the input (here, white noise for this question).

This is a finite impulse response (FIR) filter with taps at delays 0 and 4 only, no feedback in terms of  $y(n)$ .

So, with white-noise input,  $\{Y(n)\}$  is an MA(4) process (moving-average of order 4, with intermediate coefficients equal to 0):

- MA order  $q = 4$ .
- AR order  $p = 0$ .

### Problem 3:

Consider the discrete time, discrete valued process  $\{X(n)\}$  defined as

$$X(0) = \begin{cases} 1 & \text{with probability } 0 \leq b \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

for  $n = 0$  and for  $n = 1, 2, 3, \dots$ , conditioned on  $X(n-1)$ , as

$$X(n) = \begin{cases} 1 & \text{with probability } a + (1-2a)X(n-1) \text{ with } 0 \leq a \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

### 3.1 Explain using pseudo-code how 3 independent realization of $X(n)$ of each 100 samples can be simulated.

### 3.1 Pseudocode to simulate 3 realizations of length 100

Let  $U()$  generate a Uniform(0,1) random number.

```
pseudo
# Parameters
a = ...      # given in the problem
b = ...      # given in the problem
N = 100       # samples per realization
R = 3         # number of realizations

for r = 1 to R:
    # --- initial state ---
    u = U()
    if u < b:
        X[0] = 1
    else:
        X[0] = 0
    # --- generate the rest ---
    for n = 1 to N-1:
        u = U()
        if X[n-1] == 0:
            # Then P(X(n)=1) = a
            if u < a:
                X[n] = 1
            else:
                X[n] = 0
        else: # X[n-1] == 1
            # Then P(X(n)=1) = 1 - a
            if u < 1 - a:
                X[n] = 1
            else:
                X[n] = 0
    # store or plot X[0..N-1] for realization r
end
```

### 3.2 Verify that $\{X(n)\}$ is a 2-state, discrete time Markov process.

#### 3.2 Show $\{X(n)\}$ is a 2-state Markov chain

From the definition:

$$P(X(n) = 1 | X(n-1), X(n-2), \dots, X(0)) = a + (1-2a)X(n-1),$$

and the probability of 0 is just 1 minus that.

So the conditional distribution of  $X(n)$  depends only on  $X(n-1)$ , not on earlier values.

This is exactly the Markov property:

$$P(X(n) | X(n-1), X(n-2), \dots, X(0)) = P(X(n) | X(n-1)).$$

Since the state space is  $\{0, 1\}$ , it's a 2-state discrete-time Markov chain.

### 3.3 Draw the state-diagram for $\{X(n)\}$ and give its state transition probability matrix.

#### 3.3 State diagram and transition matrix

Compute the transition probabilities explicitly:

- If  $X(n-1) = 0$ :

$$P(X(n) = 1 | X(n-1) = 0) = a + (1-2a) \cdot 0 = a,$$

$$P(X(n) = 0 | X(n-1) = 0) = 1 - a.$$

- If  $X(n-1) = 1$ :

$$P(X(n) = 1 | X(n-1) = 1) = a + (1-2a) \cdot 1 = 1 - a,$$

$$P(X(n) = 0 | X(n-1) = 1) = a.$$

State diagram:

- From state 0: loop to 0 with prob  $1 - a$ , arrow to 1 with prob  $a$ .
- From state 1: arrow to 0 with prob  $a$ , loop to 1 with prob  $1 - a$ .

Transition matrix (rows = current, columns = next; order of states: 0,1):

$$P = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}.$$

### 3.4 Show that the steady-state probability vector $\pi$ equals

$$\pi = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.$$

### 3.4 Steady-state probability vector $\pi$

Steady state satisfies

$$\pi^T = \pi^T P, \quad \pi_0 + \pi_1 = 1.$$

With

$$P = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix},$$

solve:

$$\begin{cases} \pi_0 = \pi_0(1-a) + \pi_1 a, \\ \pi_1 = \pi_0 a + \pi_1(1-a). \end{cases}$$

But the two equations are the same when combined with  $\pi_0 + \pi_1 = 1$ .

Because the chain is symmetric, the solution is

$$\boxed{\pi = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}}.$$

You can verify  $\pi^T P = \pi^T$  for any  $0 \leq a \leq 1$ .

**3.5** Now, assume  $b = \frac{1}{2}$  and  $a = 0.9$ . Compute the mean of  $X(n)$ . Furthermore, evaluate the autocorrelation function at time-lags  $k = 0, 1$ , and  $2$ .

### 3.5 Mean and autocorrelation for $b = \frac{1}{2}$ , $a = 0.9$

#### 1. Mean $\mathbb{E}[X(n)]$

Initial distribution:

$$P(X(0) = 1) = b = \frac{1}{2}, \quad P(X(0) = 0) = \frac{1}{2}.$$

This equals the steady-state distribution, so the chain is stationary from  $n = 0$  onward:

$$P(X(n) = 1) = \pi_1 = \frac{1}{2}, \quad \forall n.$$

Since  $X(n) \in \{0, 1\}$ :

$$\boxed{\mathbb{E}[X(n)] = P(X(n) = 1) = \frac{1}{2}, \quad \forall n.}$$

#### 2. Autocorrelation $R_X(k) = \mathbb{E}[X(n)X(n+k)]$

For a stationary chain,

$$R_X(k) = P(X(n) = 1, X(n+k) = 1) = \pi_1 \cdot (P^k)_{1,1},$$

where  $(P^k)_{1,1}$  is the prob. of being in state 1  $k$  steps later starting from 1.

Our  $P$  for  $a = 0.9$  is

$$P = \begin{bmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{bmatrix}.$$

- Lag 0:

$$R_X(0) = \mathbb{E}[X(n)^2] = \mathbb{E}[X(n)] = \frac{1}{2}.$$

- Lag 1:

$$P(X(n) = 1, X(n+1) = 1) = P(X(n) = 1) P(X(n+1) = 1 | X(n) = 1) = \frac{1}{2} \cdot (1-a) = \frac{1}{2} \cdot 0.1 = 0.05.$$

$$\text{So } \boxed{R_X(1) = 0.05.}$$

- Lag 2:

First compute  $P^2$ :

$$P^2 = \begin{bmatrix} 41/50 & 9/50 \\ 9/50 & 41/50 \end{bmatrix}.$$

Thus  $(P^2)_{1,1} = 41/50 = 0.82$ , and

$$R_X(2) = \pi_1 (P^2)_{1,1} = \frac{1}{2} \cdot \frac{41}{50} = \frac{41}{100} = 0.41.$$

Summary:

$$\boxed{R_X(0) = 0.5, \quad R_X(1) = 0.05, \quad R_X(2) = 0.41.}$$

**3.6** Let  $Y(n)$  be defined as  $X(n)$

$$Y(n) = 3 - 2X(n).$$

Which type of process is  $Y(n)$  and what can you say about it?

### 3.6 Process $Y(n) = 3 - 2X(n)$

Since  $X(n) \in \{0, 1\}$ ,  $Y(n)$  takes values:

- If  $X(n) = 0$ ,  $Y(n) = 3$ ,
- If  $X(n) = 1$ ,  $Y(n) = 1$ .

The mapping  $X \leftrightarrow Y$  is one-to-one and deterministic:

$$X(n) = \frac{3 - Y(n)}{2}.$$

An invertible deterministic transform of a Markov chain is again a **Markov chain**, with the same transition probabilities, just relabeled states  $\{3, 1\}$  instead of  $\{0, 1\}$ .

So:

- $Y(n)$  is a 2-state, time-homogeneous Markov chain with states 3 and 1.
- Its steady-state probabilities are also  $(1/2, 1/2)$ .
- Mean and autocorrelation of  $Y(n)$  are easily obtained from those of  $X(n)$  using linearity:

$$\mathbb{E}[Y(n)] = 3 - 2\mathbb{E}[X(n)] = 3 - 2 \cdot \frac{1}{2} = 2,$$

etc.

# Eksame2025

Wednesday, 26 November 2025 18.40

Consider a Gaussian random vector  $\mathbf{X}$  defined as

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim \mathcal{N}(\mu_{\mathbf{X}}, \Sigma_{\mathbf{X}}) \quad (1)$$

with mean and covariance

$$\mu_{\mathbf{X}} = \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \Sigma_{\mathbf{X}} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}. \quad (2)$$

**1.1** Identify the mean and variance of the stochastic variable  $X_2$ . What is the covariance of  $X_1$  and  $X_2$ ?

## 1.1 Mean / variance of $X_2$ and covariance of $X_1, X_2$

- Mean of  $X_2$ : second entry of  $\mu_X$ :

$$\mathbb{E}[X_2] = -1.$$

- Variance of  $X_2$ : element (2,2) of  $\Sigma_X$ :

$$\text{Var}(X_2) = \Sigma_{22} = 2.$$

- Covariance of  $X_1, X_2$ : element (1,2) (or (2,1)) of  $\Sigma_X$ :

$$\text{Cov}(X_1, X_2) = \Sigma_{12} = 1.$$

**1.2** State the probability density function (pdf) for  $X_1$ .

## 1.2 Pdf of $X_1$

Any component of a Gaussian vector is Gaussian with corresponding mean and variance:

$$X_1 \sim \mathcal{N}(4, 1).$$

So the pdf is

$$f_{X_1}(x) = \frac{1}{\sqrt{2\pi \cdot 1}} \exp\left(-\frac{(x-4)^2}{2 \cdot 1}\right), \quad x \in \mathbb{R}.$$



**1.3** Provided a multiple realizations of the vector  $\mathbf{X}$ , sketch how would scatter plots for  $(X_1, X_2)$   $(X_1, X_3)$  and  $(X_2, X_3)$ , would look like. Explain your reasoning.

### 1.3 Sketch / description of scatter plots

Think of many realizations of  $X$ .

#### (a) $(X_1, X_2)$

These two are jointly Gaussian with covariance matrix

$$\Sigma_{12} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

Correlation:

$$\rho_{12} = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \text{Var}(X_2)}} = \frac{1}{\sqrt{1 \cdot 2}} = \frac{1}{\sqrt{2}} \approx 0.71.$$

So the scatter plot is an **elongated ellipse with positive slope**, centered at the mean  $(4, -1)$ : points tend to increase in  $X_2$  when  $X_1$  increases.

---

#### (b) $(X_1, X_3)$

Covariance is  $\Sigma_{13} = 0$ . In a multivariate Gaussian, **zero covariance implies independence**. So  $X_1$  and  $X_3$  are independent. Their variances:

- $X_1$ : var 1,
- $X_3$ : var 3.

Scatter plot: a “roundish” cloud aligned with the axes, centered at  $(4, 0)$ . No tilt.

#### (c) $(X_2, X_3)$

Similarly  $\Sigma_{23} = 0$ , so  $X_2$  and  $X_3$  are also independent.

Scatter plot: axis-aligned cloud centered at  $(-1, 0)$ . No visible correlation.

We define a new random variable  $Y$  as a linear transformation of  $\mathbf{X}$  according to

$$Y = X_1 - 3X_3. \quad (3)$$

**1.4** Derive the mean and variance of  $Y$ .



#### 1.4 Mean and variance of $Y = X_1 - 3X_3$

Write  $Y = c^T X$  with  $c = (1, 0, -3)^T$ .

- Mean:

$$\mathbb{E}[Y] = c^T \mu_X = 1 \cdot 4 + 0 \cdot (-1) + (-3) \cdot 0 = 4.$$

- Variance:

$$\text{Var}(Y) = c^T \Sigma_X c.$$

Quick way:

$$Y = X_1 - 3X_3 \Rightarrow \text{Var}(Y) = \text{Var}(X_1) + 9\text{Var}(X_3) - 6\text{Cov}(X_1, X_3).$$

Here  $\text{Var}(X_1) = 1$ ,  $\text{Var}(X_3) = 3$ ,  $\text{Cov}(X_1, X_3) = 0$ , so

$$\text{Var}(Y) = 1 + 9 \cdot 3 = 1 + 27 = 28.$$

Thus

$$Y \sim \mathcal{N}(4, 28).$$

#### 1.5 Calculate the covariance of $Y$ and $X_2$ .

##### 1.5 Covariance of $Y$ and $X_2$

$$\text{Cov}(Y, X_2) = \text{Cov}(X_1 - 3X_3, X_2) = \text{Cov}(X_1, X_2) - 3\text{Cov}(X_3, X_2).$$

From  $\Sigma_X$ :  $\text{Cov}(X_1, X_2) = 1$ ,  $\text{Cov}(X_3, X_2) = 0$ .

So

$$\boxed{\text{Cov}(Y, X_2) = 1.}$$

#### 1.6 Suppose that we are to simulate 100 realizations of $\mathbf{X}$ and $Y$ based on a random number generator which returns independent Gaussian random variables with zero mean and variance one. Write an algorithm in pseudo code to do that.

*Hint:* Note that  $\Sigma_{\mathbf{X}} = A^T A$  with  $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

##### 1.6 Pseudocode to simulate 100 realizations of $X$ and $Y$

We can use that a Gaussian vector can be generated as

$$X = \mu_X + BZ,$$

where  $Z \sim \mathcal{N}(0, I_3)$  (3 independent standard normals) and  $B$  satisfies  $BB^T = \Sigma_X$ .

One valid choice is

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix}$$

(since  $BB^T = \Sigma_X$ ).



Pseudocode (Matlab / Python style):

```
pseudo Copy code

# Input: nSamples = 100
# Output: arrays X[3, nSamples], Y[nSamples]

mu = [4, -1, 0]^T
B = [[1, 0, 0],
      [1, 1, 0],
      [0, 0, sqrt(3)]] 

for n = 1 to 100:
    # 1. Generate 3 independent N(0,1) variables
    Z = [randn(), randn(), randn()] # randn() ~ N(0,1)

    # 2. Form one realization of X
    X_col = mu + B * Z           # matrix-vector multiply
    X1 = X_col[1]
    X2 = X_col[2]
    X3 = X_col[3]

    # 3. Compute corresponding Y
    Y[n] = X1 - 3 * X3

    # 4. Store X if needed
    X[:, n] = X_col
end
```

This uses only independent  $N(0, 1)$  and linear transforms.

- 1.7 Consider the sample average of 100 independent realizations of  $X_1$ , defined as let the sample average be defined according to

$$\hat{\mu}_{X_1} = \frac{1}{100} \sum_{n=1}^{100} X_1^{(n)},$$

where  $X_1^{(n)}$  is the  $n^{\text{th}}$  realization of the random variable  $X_1$ . Calculate the expectation and variance of  $\hat{\mu}_{X_1}$ . How is sample average  $\hat{\mu}_{X_1}$  distributed and why?



## 1.7 Sample average of 100 independent realizations of $X_1$

Sample average:

$$\hat{\mu}_{X_1} = \frac{1}{100} \sum_{n=1}^{100} X_1^{(n)},$$

where the  $X_1^{(n)}$  are independent copies of  $X_1 \sim \mathcal{N}(4, 1)$ .

- Expectation:

$$\mathbb{E}[\hat{\mu}_{X_1}] = \frac{1}{100} \sum_{n=1}^{100} \mathbb{E}[X_1^{(n)}] = \frac{1}{100} \cdot 100 \cdot 4 = 4.$$

- Variance (for independent variables):

$$\text{Var}(\hat{\mu}_{X_1}) = \frac{1}{100^2} \sum_{n=1}^{100} \text{Var}(X_1^{(n)}) = \frac{1}{100^2} \cdot 100 \cdot 1 = \frac{1}{100}.$$

- Distribution: A linear combination of jointly Gaussian variables is again Gaussian. The vector  $(X_1^{(1)}, \dots, X_1^{(100)})$  is multivariate Gaussian, so  $\hat{\mu}_{X_1}$  is exactly

$$\hat{\mu}_{X_1} \sim \mathcal{N}\left(4, \frac{1}{100}\right).$$

(We don't even need the CLT here; it's exact because everything is Gaussian.)

Consider the discrete-time random process

$$X(n) = \cos\left(\frac{n\pi}{2}\right) + W(n), \quad (4)$$

where  $W(n)$  is a zero mean Gaussian process with autocorrelation function

$$R_W(k) = \begin{cases} \frac{1}{2}, & k = 0 \\ 0, & k \neq 0. \end{cases} \quad (5)$$

## 2.1 Sketch a few realizations of $X(n)$

Deterministic part:

$$\cos\left(\frac{n\pi}{2}\right) = \begin{cases} 1, & n = 0 \pmod{4}, \\ 0, & n = 1 \pmod{4}, \\ -1, & n = 2 \pmod{4}, \\ 0, & n = 3 \pmod{4}. \end{cases}$$

So the "skeleton" is  $1, 0, -1, 0, 1, 0, -1, 0, \dots$

Then you add independent Gaussian noise  $W(n) \sim \mathcal{N}(0, 1/2)$ . In a sketch you'd show samples scattered around those values with some randomness each time step.

## 2.1 Sketch a few realizations of $X(n)$



## 2.2 Mean of $X(n)$

$$\mathbb{E}[X(n)] = \cos\left(\frac{n\pi}{2}\right) + \mathbb{E}[W(n)] = \cos\left(\frac{n\pi}{2}\right) + 0 = \cos\left(\frac{n\pi}{2}\right).$$

So the mean is periodic with period 4.

## 2.2 Compute the mean $\mathbb{E}[X(n)]$ .

### 2.3 Stationarity properties of $X(n)$ and $W(n)$

- $W(n)$ :

Zero mean, and its autocorrelation depends only on lag  $k$ :

$$R_W(k) = \frac{1}{2}\delta[k],$$

where  $\delta[k]$  is the discrete Kronecker delta.

$\Rightarrow W(n)$  is **wide-sense stationary (WSS)** and, since it is also Gaussian, it is strictly stationary.

- $X(n)$ :

Mean  $\mathbb{E}[X(n)] = \cos(n\pi/2)$  clearly depends on  $n$  and is not constant.

$\Rightarrow X(n)$  is **not WSS** (and not strictly stationary either).

## 2.3 Comment on the stationarity properties of the two processes $X(n)$ and $W(n)$ .



## 2.4 Show that $a = 4$ gives WSS for $Y(n)$ , but $a = 1$ does not

Define

$$Y(n) = X(n) - X(n-a).$$

First compute the mean:

$$\mathbb{E}[Y(n)] = \mathbb{E}[X(n)] - \mathbb{E}[X(n-a)] = \cos\left(\frac{n\pi}{2}\right) - \cos\left(\frac{(n-a)\pi}{2}\right).$$

- If  $a = 4$ :

$$\cos\left(\frac{(n-4)\pi}{2}\right) = \cos\left(\frac{n\pi}{2} - 2\pi\right) = \cos\left(\frac{n\pi}{2}\right),$$

so

$$\mathbb{E}[Y(n)] = \cos\left(\frac{n\pi}{2}\right) - \cos\left(\frac{n\pi}{2}\right) = 0,$$

constant in  $n$ .

Moreover,

$$Y(n) = X(n) - X(n-4) = [\cos(\cdot) + W(n)] - [\cos(\cdot) + W(n-4)] = W(n) - W(n-4).$$

This is a linear filter of the WSS process  $W(n)$  with constant coefficients; such linear time-invariant filtering preserves WSS.

$\Rightarrow$  For  $a = 4$ ,  $Y(n)$  is WSS.

- If  $a = 1$ :

$$\mathbb{E}[Y(n)] = \cos\left(\frac{n\pi}{2}\right) - \cos\left(\frac{(n-1)\pi}{2}\right),$$

which cycles through several different values as  $n$  changes; it is not constant.

$\Rightarrow$  For  $a = 1$ ,  $Y(n)$  is not WSS.

Now, consider a third process  $Y(n)$  obtained as

$$Y(n) = X(n) - X(n-a), \quad (6)$$

where  $a$  is some positive integer chosen such that  $Y(n)$  is WSS.

## 2.4 Show that choosing $a = 4$ results in $Y(n)$ being wide-sense stationary (WSS), but that this is not true for the case $a = 1$

From now on set  $a = 4$ , so  $Y(n) = W(n) - W(n-4)$ .

### 2.5 Type of process $Y(n)$

- $Y(n)$  is a zero-mean, WSS Gaussian process, obtained by linearly filtering white Gaussian noise.
- More specifically, it's a moving-average (MA) process of order 1 with lag 4 (two taps at delays 0 and 4 with coefficients 1 and  $-1$ ).



From now on, we let  $a = 4$  and hence  $Y(n)$  is WSS.

## 2.5 Which type of process is $Y(n)$ ?

From now on set  $a = 4$ , so  $Y(n) = W(n) - W(n - 4)$ .

### 2.5 Type of process $Y(n)$

- $Y(n)$  is a zero-mean, WSS Gaussian process, obtained by linearly filtering white Gaussian noise.
- More specifically, it's a moving-average (MA) process of order 1 with lag 4 (two taps at delays 0 and 4 with coefficients 1 and  $-1$ ).

## 2.6 Compute the auto-correlation of $Y(n)$ .

### 2.6 Autocorrelation of $Y(n)$

$$Y(n) = W(n) - W(n - 4).$$

Autocorrelation:

$$R_Y(k) = \mathbb{E}[Y(n+k)Y(n)].$$

Compute:

$$\begin{aligned} R_Y(k) &= \mathbb{E}[(W(n+k) - W(n+k-4))(W(n) - W(n-4))] \\ &= \mathbb{E}[W(n+k)W(n)] - \mathbb{E}[W(n+k)W(n-4)] \\ &\quad - \mathbb{E}[W(n+k-4)W(n)] + \mathbb{E}[W(n+k-4)W(n-4)]. \end{aligned}$$

Using  $\mathbb{E}[W(m)W(n)] = R_W(m-n) = \frac{1}{2}$  if  $m = n$ , else 0:

- $\mathbb{E}[W(n+k)W(n)] = R_W(k)$ .
- $\mathbb{E}[W(n+k)W(n-4)] = R_W(k+4)$ .
- $\mathbb{E}[W(n+k-4)W(n)] = R_W(k-4)$ .
- $\mathbb{E}[W(n+k-4)W(n-4)] = R_W(k)$ .

Thus

$$R_Y(k) = 2R_W(k) - R_W(k+4) - R_W(k-4).$$

Since  $R_W(k) = \frac{1}{2}$  for  $k = 0$  and 0 otherwise, nonzero values happen only when the arguments are 0:

- $k = 0 : R_Y(0) = 2 \cdot \frac{1}{2} - 0 - 0 = 1$ .
- $k = 4 : R_Y(4) = 2 \cdot 0 - 0 - \frac{1}{2} = -\frac{1}{2}$ .
- $k = -4 : R_Y(-4) = 2 \cdot 0 - \frac{1}{2} - 0 = -\frac{1}{2}$ .
- All other  $k : R_Y(k) = 0$ .

So

$$R_Y(k) = \begin{cases} 1, & k = 0, \\ -\frac{1}{2}, & k = \pm 4, \\ 0, & \text{otherwise.} \end{cases}$$

## 2.7 Compute the power spectral density (PSD) for $Y(n)$ .



## 2.7 Power spectral density (PSD) of $Y(n)$

The PSD is the DTFT of the autocorrelation:

$$S_Y(\omega) = \sum_{k=-\infty}^{\infty} R_Y(k)e^{-j\omega k}.$$

Only  $k = 0, \pm 4$  contribute:

$$\begin{aligned} S_Y(\omega) &= 1 \cdot e^{-j\omega \cdot 0} - \frac{1}{2}e^{-j\omega \cdot 4} - \frac{1}{2}e^{-j\omega \cdot (-4)} \\ &= 1 - \frac{1}{2}e^{-j4\omega} - \frac{1}{2}e^{j4\omega} \\ &= 1 - \frac{1}{2}(e^{j4\omega} + e^{-j4\omega}) \\ &= 1 - \cos(4\omega). \end{aligned}$$

So

$$S_Y(\omega) = 1 - \cos(4\omega).$$

(You can also get this from the filter viewpoint: filter  $H(z) = 1 - z^{-4}$  driven by white noise of variance  $1/2$ . Then  $S_Y(\omega) = |H(e^{j\omega})|^2 \cdot \frac{1}{2} = (2 - 2 \cos 4\omega) \cdot \frac{1}{2} = 1 - \cos 4\omega$ .)

Shape: periodic in  $\omega$  with period  $2\pi$ , zero where  $\cos(4\omega) = 1$  (e.g.  $\omega = m\pi/2$ ), and maximum 2 where  $\cos(4\omega) = -1$ .

## 2.8 Sketch the PSD for the process $Z(n)$ defined as

$$Z(n) = 3Y(n) + 7.$$

Compare it to the PSD for  $Y(n)$ .

### 2.8 PSD of $Z(n) = 3Y(n) + 7$ and comparison

- Adding a **constant** (7) changes only the mean; it does **not** change the autocorrelation for nonzero lags, hence does not change the PSD.
- Multiplying by 3 scales the PSD by  $3^2 = 9$  (because PSD is related to power, and power scales with the square of amplitude).

So

$$S_Z(\omega) = 9S_Y(\omega) = 9(1 - \cos(4\omega)).$$

Sketch: exactly the same shape as  $S_Y(\omega)$ , but every value is multiplied by 9. Peaks are 9× higher; zeros are at the same frequencies.

