EC304: Probability Theory and Stochastic Process

Module 5: Covariance and Corelation in Joint Probability Density Function

1 Covariance and Corelation

Covariance and correlation are two measures used to understand how two random variables vary together. They are essential in statistics, signal processing, communication systems, machine learning, and sensor fusion.

1.1 Covariance

For two random variables X and Y, covariance is defined as:

$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

Equivalent and often useful form:

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]\,E[Y]$$

Interpretation

$$\mathrm{Cov}(X,Y) \begin{cases} > 0 & \text{Variables increase together (positive relation)} \\ < 0 & \text{One increases while the other decreases} \\ = 0 & \text{No linear relationship (independent or uncorrelated)} \end{cases}$$

Units

Covariance has *combined units*. Example: Temperature (°C) and Humidity (

Thus covariance is not scale-free.

Example 1

Let

$$X = \{1,2,3\}, \quad Y = \{2,4,6\}$$

These follow a perfect linear relationship $Y = 2X$.

$$E[X] = 2, \quad E[Y] = 4$$

$$E[XY] = \frac{1}{3}(1\cdot 2 + 2\cdot 4 + 3\cdot 6) = \frac{28}{3}$$

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{28}{3} - (2)(4) = \frac{4}{3}$$

Covariance $> 0 \rightarrow$ strong positive relation.

1.2 Correlation

Definition

Correlation is the **normalized** covariance:

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Properties

$$-1 \le \rho_{XY} \le 1$$
$$|\rho| = 1 \Rightarrow \text{perfect linear relationship}$$
$$\rho = 0 \Rightarrow \text{no linear relationship}$$

Unitless

Because correlation divides by standard deviations, it has no units.

Example 2

From Example 1:

$$\sigma_X = \sqrt{\tfrac{2}{3}}, \quad \sigma_Y = 2\sqrt{\tfrac{2}{3}}$$

Correlation:

$$\rho_{XY} = \frac{\frac{4}{3}}{\left(\sqrt{\frac{2}{3}}\right)\left(2\sqrt{\frac{2}{3}}\right)} = 1$$

So correlation detects the perfect linear relationship.

Example 3: (Temperature and Humidity)

A weather node measures:

- Temperature X in °C
- Humidity Y in %

Suppose over several readings we obtain the simplified dataset:

Reading #	1	2	3	4
X (Temp °C)	24	25	27	30
Y (Humidity %)	60	58	55	50

Humidity decreases as temperature increases.
Means are computed as,

$$E[X] = 26.5, \quad E[Y] = 55.75$$

Covariance,

$$E[XY] = \frac{1}{4}(24 \cdot 60 + 25 \cdot 58 + 27 \cdot 55 + 30 \cdot 50) = 1484.75$$

$$\mathrm{Cov}(X,Y) = 1484.75 - (26.5)(55.75) = -43.375$$

Note:

Covariance is *negative* $\rightarrow$ as temperature increases, humidity decreases (evaporation effect).

Correlation

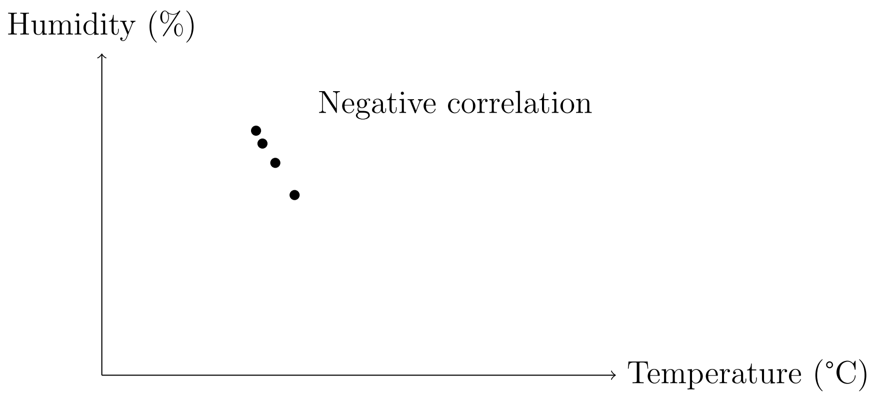
Compute standard deviations:

$$\sigma_X = 2.217, \qquad \sigma_Y = 4.031$$

$$\rho_{XY} = \frac{-43.375}{(2.217)(4.031)} = -0.486$$

Meaning: Moderate negative correlation $\rightarrow$ typical in many sensor environments.

Scatter plot (conceptual)



Summary

- Covariance measures *joint variability*. It has units.
- Correlation is normalized covariance. It is unitless.
- Covariance shows direction (positive/negative); correlation shows strength.
- Sensor data often show correlation due to physics (e.g., temperature–humidity).
- Zero covariance does not imply independence, but independence implies zero covariance.

2 Joint Probability Density Function

In many engineering applications, especially in electronics and communication systems, random variables (RVs) represent uncertain signal characteristics such as amplitude, duration, noise level, etc. When two such signal parameters vary jointly, their behaviour can be described by a **joint probability density function (PDF)**.

Definition

Consider two continuous random variables:
- X: Pulse peak amplitude (in volts), $0 \leq X \leq 2$ - Y: Normalized pulse width (unitless), $0 \leq Y \leq 1$

The joint PDF is defined as:

$$f_{X,Y}(x,y) = \frac{3}{4}x^2 y, \quad 0 \leq x \leq 2, \quad 0 \leq y \leq 1.$$

Outside this region, $f_{X,Y}(x,y) = 0$.

Normalization Check

$$\int_0^2 \int_0^1 \frac{3}{4}x^2 y \, dy \, dx = \frac{3}{4}\left(\int_0^2 x^2 dx\right)\left(\int_0^1 y dy\right) = 1.$$

So this is a valid PDF.

Marginal PDFs

Marginal of X

$$f_X(x) = \int_0^1 \frac{3}{4}x^2 y \, dy = \frac{3}{8}x^2.$$

Marginal of Y

$$f_Y(y) = \int_0^2 \frac{3}{4}x^2 y \, dx = 2y.$$

Since $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, X and Y are **independent**.

Expectations

Expected Amplitude $E[X]$

$$E[X] = \int_0^2 x f_X(x) dx = \frac{3}{8}\int_0^2 x^3 dx = 1.5 \text{ V}.$$

Expected Width $E[Y]$

$$E[Y] = \int_0^1 y f_Y(y) dy = 2\int_0^1 y^2 dy = \frac{2}{3}.$$

Expected Pulse Area $E[XY]$

Since X and Y are independent:

$$E[XY] = E[X]E[Y] = 1.5 \times \frac{2}{3} = 1.$$

This quantity corresponds to the expected pulse "area" (amplitude $\times$ width).

Variance Calculations

Variance of X

$$E[X^2] = \frac{3}{8}\int_0^2 x^4 dx = \frac{12}{5}, \qquad \mathrm{Var}(X) = 0.15.$$

Variance of Y

$$E[Y^2] = 2\int_0^1 y^3 dy = 0.5, \qquad \mathrm{Var}(Y) = \frac{1}{18}.$$

3 Examples

Example A — Independent Uniform Sensor Model

Let

$$X \sim \mathrm{Uniform}(20°\mathrm{C}, 30°\mathrm{C}), \qquad Y \sim \mathrm{Uniform}(40\%, 60\%),$$

and assume independence. The joint PDF on the rectangle $20 \le x \le 30,\ 40 \le y \le 60$ (here units: °C for x, % for y) is

$$f_{X,Y}(x,y) = \frac{1}{(30-20)(60-40)} = \frac{1}{200}, \qquad 20 \le x \le 30,\ 40 \le y \le 60.$$

Variance of a Uniform Random Variable

For a continuous uniform random variable $U \sim \mathrm{Uniform}(a,b)$, the variance is given by the standard formula:

$$\mathrm{Var}(U) = \frac{(b-a)^2}{12}.$$

This result is derived from the definition of variance,

$$\mathrm{Var}(U) = E[U^2] - (E[U])^2,$$

where, for a uniform distribution:

$$E[U] = \frac{a+b}{2}, \qquad E[U^2] = \frac{a^2+ab+b^2}{3}.$$

Substituting into the variance definition:

$$\mathrm{Var}(U) = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

Moments and spread

$$E[X] = \frac{20+30}{2} = 25 \ (^\circ\mathrm{C}).$$
$$E[Y] = \frac{40+60}{2} = 50 \ (\%).$$

Independence implies

$$E[XY] = E[X]\,E[Y] = 25 \times 50 = 1250 \quad (\text{units: } {}^\circ\mathrm{C} \times \%).$$

Variances (uniform formula $\mathrm{Var}(U(a,b)) = \frac{(b-a)^2}{12}$):

$$\mathrm{Var}(X) = \frac{(30-20)^2}{12} = \frac{100}{12} = \frac{25}{3} \approx 8.3333 \ (^\circ\mathrm{C}^2).$$

$$\mathrm{Var}(Y) = \frac{(60-40)^2}{12} = \frac{400}{12} = \frac{100}{3} \approx 33.3333 \ (\%^2).$$

Covariance and correlation:

$$\mathrm{Cov}(X,Y) = 0, \qquad \rho_{XY} = 0.$$

Interpretation

- The *typical* temperature is 25°C and typical relative humidity is 50%. - Because sensors are independent here, knowing temperature gives no information about humidity. - $E[XY] = 1250$ is the expected product: if some performance metric depends multiplicatively on temperature and humidity (e.g., a simple empirical index proportional to $X \cdot Y$), use this number for expected behaviour. - Variances tell about sensor spread: temperature fluctuates less (std $\approx \sqrt{8.333} = 2.886$°C) than humidity (std $\approx \sqrt{33.333} = 5.774\%$).

Example B — Dependent Model (Humidity constrained by Temperature)

Rationale and physical mapping

In some simplified models, humidity-related measures (e.g., dew point fraction or humidity potential) cannot exceed a value tied to temperature. To capture one-sided dependence, model the joint support as the triangular region

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

where we use normalized units:

- X : normalized temperature (0 to 1) — map to real temperature by $T = T_{\min} + X(T_{\max} - T_{\min})$.
- Y : normalized humidity-related fraction (0 to 1) but constrained to be $\le X$.

This enforces Y is always at most X (a simple model of a physical constraint).

Joint PDF (triangular)

Define

$$f_{X,Y}(x,y) = 2, \qquad 0 \le y \le x \le 1,$$

and $f_{X,Y} = 0$ elsewhere. Check normalization:

$$\iint_{0\le y\le x\le 1} 2\,dy\,dx = \int_{x=0}^{1} 2\left(\int_{y=0}^{x} dy\right)dx = \int_0^1 2x\,dx = x^2\Big|_0^1 = 1.$$

Marginal PDFs

$$f_X(x) = \int_0^x 2\,dy = 2x, \qquad 0 \le x \le 1.$$

$$f_Y(y) = \int_{x=y}^{1} 2\,dx = 2(1-y), \qquad 0 \le y \le 1.$$

Since $f_{X,Y}(x,y) \ne f_X(x)f_Y(y)$ (check: $2 \ne 4x(1-y)$ generally), X and Y are dependent.

Expectations and product

Compute step by step (all integrals over $[0,1]$ with triangular support):

Expected temperature $E[X]$

$$E[X] = \int_0^1 x f_X(x)\,dx = \int_0^1 x(2x)\,dx = 2\int_0^1 x^2 dx = 2\cdot\frac{1}{3} = \frac{2}{3} \approx 0.6667.$$

Expected humidity-related fraction $E[Y]$

$$E[Y] = \int_0^1 y f_Y(y)\,dy = \int_0^1 y\cdot 2(1-y)\,dy = 2\int_0^1 \left(y-y^2\right)dy = 2\left(\tfrac{1}{2}-\tfrac{1}{3}\right) = 2\cdot\tfrac{1}{6} = \tfrac{1}{3} \approx 0.3333.$$

Expected product $E[XY]$

$$E[XY] = \iint_{0\le y\le x\le 1} xy\cdot 2\,dy\,dx = 2\int_0^1 x\left(\int_0^x y\,dy\right)dx = 2\int_0^1 x\cdot\frac{x^2}{2}\,dx = \int_0^1 x^3 dx = \frac{1}{4} = 0.25.$$

Note $E[X]E[Y] = \frac{2}{3}\cdot\frac{1}{3} = \frac{2}{9} \approx 0.2222$. So

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{1}{4} - \frac{2}{9} = \frac{1}{36} \approx 0.02778 > 0.$$

A small positive covariance: larger X tends to allow larger Y.

Variances and correlation

$$E[X^2] = \int_0^1 x^2 f_X(x)\,dx = \int_0^1 x^2(2x)\,dx = 2\int_0^1 x^3 dx = 2\cdot\tfrac{1}{4} = \tfrac{1}{2}.$$

Thus

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \tfrac{1}{2} - \left(\tfrac{2}{3}\right)^2 = \tfrac{1}{2} - \tfrac{4}{9} = \tfrac{1}{18} \approx 0.05556.$$

$$E[Y^2] = \int_0^1 y^2 \cdot 2(1-y)\,dy = 2\int_0^1 (y^2 - y^3)dy = 2\Big(\tfrac{1}{3} - \tfrac{1}{4}\Big) = 2 \cdot \tfrac{1}{12} = \tfrac{1}{6}.$$

So

$$\mathrm{Var}(Y) = \tfrac{1}{6} - \left(\tfrac{1}{3}\right)^2 = \tfrac{1}{6} - \tfrac{1}{9} = \tfrac{1}{18} \approx 0.05556.$$

Standard deviations are equal here: $\sigma_X = \sigma_Y = \sqrt{1/18}$. Correlation:

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{1/36}{1/18} = \frac{1}{2} = 0.5.$$

So there is a moderate positive correlation.

Interpretation (sensor context)

- The normalized mean temperature is $E[X] = 2/3$ and normalized mean humidity measure is $E[Y] = 1/3$. If you map normalization back to real ranges, multiply accordingly.
- $E[XY] = 0.25$ is the expected multiplicative effect (e.g., expected humidity contribution scaled by temperature), and because of dependence, it is larger than $E[X]E[Y]$ by the covariance amount.
- $\mathrm{Cov}(X,Y) > 0$ and $\rho_{XY} = 0.5$ indicate that when the (normalized) temperature is higher, the humidity-related fraction tends to be higher too — plausible in many microclimates or sensor models.
- Equal variances here are a quirk of this triangular model; the key takeaway is how dependence changes $E[XY]$ relative to the independent case.

4 Additional Examples of Joint PDFs in Electronics

4.1 Example 1: Noise Voltage and Noise Current in an Amplifier

Consider an analog amplifier with two input noise sources:

- Noise voltage V_n (in mV)
- Noise current I_n (in μA)

Assume their joint PDF is:

$$f_{V_n,I_n}(v,i) = \frac{1}{9}vi, \qquad 0 \le v \le 2,\ 0 \le i \le 3.$$

Marginal PDFs

Noise voltage:

$$f_{V_n}(v) = \int_0^3 \frac{1}{9}vi\,di = \frac{v}{2}.$$

Noise current:

$$f_{I_n}(i) = \int_0^2 \frac{1}{9}vi\,dv = \frac{2i}{9}.$$

Means

$$E[V_n] = \int_0^2 v \cdot \frac{v}{2}\,dv = \frac{4}{3}, \qquad E[I_n] = \int_0^3 i \cdot \frac{2i}{9}\,di = 2.$$

Second Moment

$$E[V_nI_n] = \int_0^2 \int_0^3 vi\left(\frac{1}{9}vi\right)di\,dv = \frac{1}{9}\left(\int_0^2 v^2\,dv\right)\left(\int_0^3 i^2\,di\right) = 2.$$

Covariance and Correlation

$$\mathrm{Cov}(V_n, I_n) = E[V_nI_n] - E[V_n]E[I_n] = 2 - \left(\frac{4}{3}\right)(2) = -\frac{2}{3}.$$

Variances:

$$\mathrm{Var}(V_n) = \frac{2^2}{12} = \frac{1}{3}, \qquad \mathrm{Var}(I_n) = \frac{3^2}{12} = \frac{9}{12} = \frac{3}{4}.$$

Correlation coefficient:

$$\rho = \frac{\mathrm{Cov}(V_n, I_n)}{\sqrt{\mathrm{Var}(V_n)\,\mathrm{Var}(I_n)}} = \frac{-\frac{2}{3}}{\sqrt{\frac{1}{3}\cdot\frac{3}{4}}} = -\frac{2}{3}.$$

Significance

Noise voltage and current are negatively correlated, meaning that as one increases, the other tends to decrease. This is useful in low-noise amplifier design and noise budgeting.

4.2 Example 2: Light Intensity and Photodiode Current

A photodiode produces a current proportional to light intensity.

Let:

- L: Light intensity (mW/cm^2)
- I_p: Photocurrent (μA)

Joint PDF:

$$f_{L,I_p}(l,i) = \frac{3}{8} l^2 i, \qquad 0 \le l \le 1,\ 0 \le i \le 4.$$

Marginal PDFs

Light intensity:

$$f_L(l) = \int_0^4 \frac{3}{8} l^2 i \, di = 3l^2.$$

Photocurrent:

$$f_{I_p}(i) = \int_0^1 \frac{3}{8} l^2 i \, dl = \frac{i}{8}.$$

Means

$$E[L] = \int_0^1 l \cdot 3l^2 \, dl = \frac{3}{4}, \qquad E[I_p] = \int_0^4 i \cdot \frac{i}{8} \, di = \frac{8}{3}.$$

Second Moment

$$E[LI_p] = \int_0^1 \int_0^4 li \left(\frac{3}{8} l^2 i\right) di \, dl = \frac{3}{8} \left(\int_0^1 l^3 \, dl\right) \left(\int_0^4 i^2 \, di\right) = 2.$$

Covariance and Correlation

$$\mathrm{Cov}(L, I_p) = E[LI_p] - E[L]E[I_p] = 2 - \left(\frac{3}{4}\right)\left(\frac{8}{3}\right) = 0.$$

Variances:

$$\mathrm{Var}(L) = \frac{1}{12}, \qquad \mathrm{Var}(I_p) = \frac{16}{3}.$$

Correlation:

$$\rho = \frac{\mathrm{Cov}(L, I_p)}{\sqrt{\mathrm{Var}(L)\,\mathrm{Var}(I_p)}} = 0.$$

Significance

The photocurrent depends on light intensity, but the statistical model chosen makes them uncorrelated. Such models are useful in optical receiver noise, fading analysis, and optical link performance evaluation.