Stochastic Noise & Methods Workshop

Presented by IEEE Signal Processing Society

Applications of Stochastic Noise & Methods

- **Noise in SDR:** AWGN/thermal, Phase Noise/Carrier Frequency Offset, IQ imbalance/DC offset, Impulsive/Co-Channel Interference, and Power Amplifier Nonlinearity/Clipping
- **Diagnose Fast:** Signal to Noise Ratio, Error Vector Magnitude, Bit Error Rate, Power Spectral Density, Constellation Diagrams; Check Stationarity and Bandwidth.
- **Classical + Probabilistic:** Matched Filtering, Pilot-Aided CFO Estimation, PLL/Kalman, Equalization (MMSE), Denoising (median/wavelet), and Automatic Gain Control/Calibration.
- **AI Components:** Learned Denoisers/Equalizers/Detectors, Deep Beamforming/CSI compression, RL for Adaptive Gains; Train with Realistic Fading/Phase-Noise/CFO/Jitter Sims + Domain Randomization, Leverage Pilots/Self-Supervision, Optimize for EVM/BER, PCA.
- **Practical:** Use Pilots and Over-the-Air Calibration, Collect I/Q across Operating Conditions, Enforce Latency/Causality, Monitor Distribution Shift, Validate on Real Captures, and Finding Signals Hidden in AWGN.

How Do We Apply Stochastic Noise & Methods?

- **Axioms of Probability:** Bayes Rule, Chain Rule, Conditional Probability, Law of Total Probability, Mutual Exclusivity, Probability, Relative Frequencies, and Statistical Independence.

- **Decision Rules:** Likelihood Ratios, Maximum A Posteriori Rule (MAP), Maximum Likelihood Estimation Rule (ML).

- **PMFs, PDFs, and CDFs:** Random Variables, Probability Density Functions, Probability Mass Functions, Cumulative Distribution Functions, Joint Probability Density & Distribution Functions of Multiple Random Variables.

- **Moments and Dependence:** Expectation, Conditional Expectation, Covariance, Correlation Coefficient, Variance; Distinguish Independence vs Uncorrelated Random Variables.

- **Estimation Theory:** MMSE for Parameter Estimation/Equalization

Densities, Distributions, RVs, and Use Cases

<u>Continuous:</u>

- **Gaussian:** (AWGN, proper/improper) baseband noise; improper if IQ imbalance leakage.
- **Rayleigh:** NLOS fading amplitude; also the envelope of I/Q Gaussian noise.
- **Uniform:** Quantization noise and random source defaults.
- **Chi-square/Gamma:** Noise power estimates, energy detector outputs.
- **Exponential:** Inter-arrival times and Rayleigh power.
- **Laplacian:** Used in computer analysis of speech and images.

<u>Discrete</u>

- **Bernoulli/Binomial:** Bit/packet error counts (BER/BLER)
- **Geometric:** Trials until first success (retransmissions, lock acquisition).
- **Poisson:** Arrivals per interval (packets, interferers, impulsive hits).

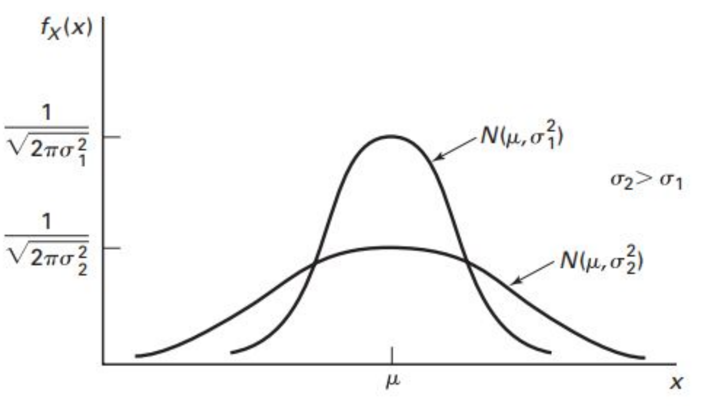


Figure 2.4-1 The Normal pdf.

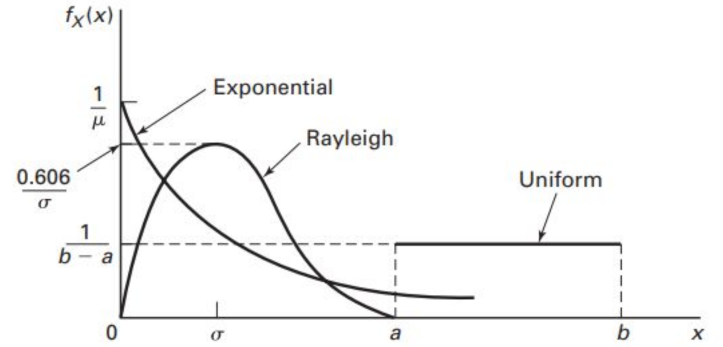


Figure 2.4-4 The Rayleigh, exponential, and uniform pdf's.

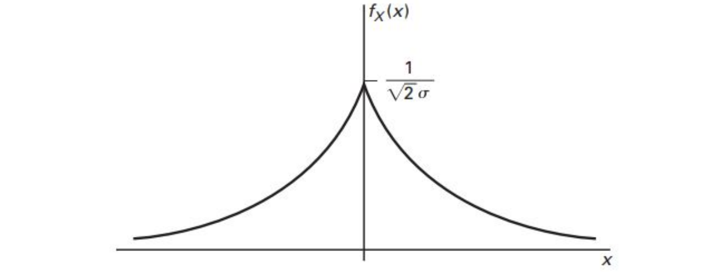


Figure 2.4-5 The Laplacian pdf used in computer analysis of speech and images.

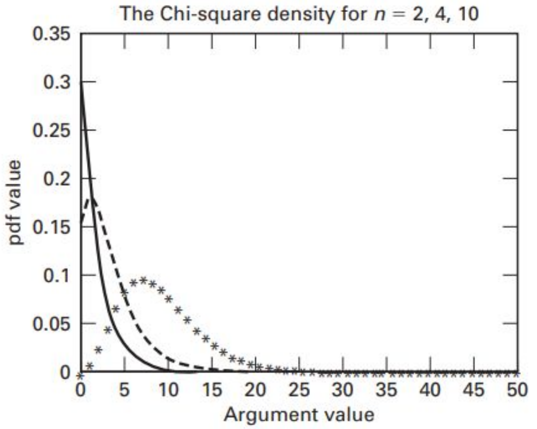


Figure 2.4-6 The Chi-square probability density function for $n = 2$ (solid), $n = 4$ (dashed), and $n = 10$ (stars). Note that for larger values of n, the shape approaches that of a Normal pdf with a positive mean-parameter μ.

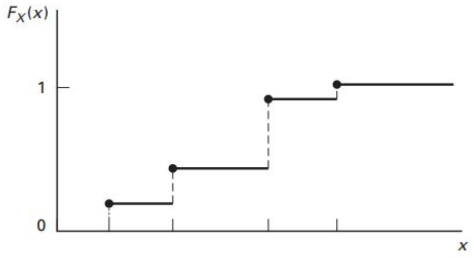


Figure 2.5-1 The cumulative distribution function for a discrete random variable.

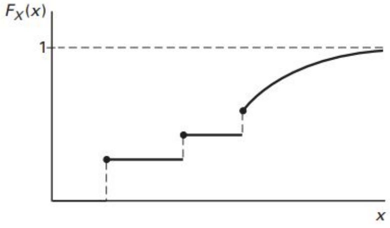


Figure 2.5-2 The CDF of a mixed RV.

Table 2.4-2 Common Continuous Probability Densities and Distribution Functions

Family	pdf $f_X(x)$	CDF $F_X(x)$
Uniform $U(a,b)$	$\frac{1}{b-a}\left[u(x-a)-u(x-b)\right]$	$\begin{cases} 0, & x<a, \\ \frac{x-a}{b-a}, & a \le x < b \\ 1, & b \le x \end{cases}$,
Exponential $\mu > 0$	$\frac{1}{\mu}e^{-x/\mu}\,u(x)$	$\begin{cases} 0, & x<0, \\ 1-e^{-x/\mu}, & x \ge 0 \end{cases}$
Gaussian $N(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2]$	$\frac{1}{2}+\mathrm{erf}(\frac{x-\mu}{\sigma})$
Laplacian $\sigma > 0$	$\frac{1}{\sqrt{2}\sigma}\exp[-\sqrt{2}\lvert x\rvert/\sigma]$	$\frac{1}{2}[1+\mathrm{sgn}(x)(1-\exp(-\sqrt{2}\lvert x\rvert/\sigma))]$
Rayleigh $\sigma > 0$	$\frac{x}{\sigma^2}e^{-x^2/2\sigma^2}u(x)$	$\left[1-e^{-x^2/2\sigma^2}\right]u(x)$

Table 2.5-1 Table Common Discrete RVs, PMFs, and CDFs

Family	PMF $P_K(k)$	CDF $F_K(k)$
Bernoulli p, q	$q\delta(k)+p\delta(k-1)$	$qu(k)+pu(k-1)$
Binomial n, k	$\binom{n}{k}p^k q^{n-k}\left[u(k)-u(n-k)\right]$	$\begin{cases} 0, & k<0, \\ \sum_{l=0}^{k}\binom{n}{l}p^l q^{n-l}, & 0 \le k < n \\ 1, & k \ge n. \end{cases}$
Poisson $\mu > 0$	$\frac{\mu k}{k!}e^{-\mu}u(k)$	$\frac{\gamma(k+1,\mu)^{\dagger}}{k!}\times u(k)$
Geometric p, q	$pq^k u(k)$	$p\left(\frac{1-q^{k+1}}{1-q}\right)u(k)$

MAP & ML Rule

Prior: Probability Value Given (Ex: P(C) = .25)

Likelihood: Conditional Density (Ex: fy(Y I X=x))

Used all the time in making important decisions.

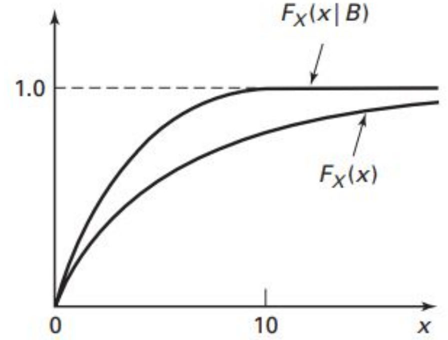


Figure 2.6-1 Conditional and unconditional CDFs of X.

Conditional Distribution: This is a distribution given some event, B, that happened same axiom applies for densities.

Likelihood Ratio $$\Lambda(x) = \frac{p(x|H_1)}{p(x|H_2)}$$

MAP Likelihood Ratio $$\boxed{\Lambda(x) \underset{H_2}{\overset{H_1}{\gtrless}} \frac{P(H_2)}{P(H_1)}}$$

ML Likelihood Ratio $$\boxed{\Lambda(x) \underset{H_2}{\overset{H_1}{\gtrless}} 1}$$

ML Decision Rule

$$\boxed{\text{Decide } H_1 \text{ if } P(H_1|x) > P(H_2|x)}$$

$$p(x|H_1) \underset{H_2}{\overset{H_1}{\gtrless}} p(x|H_2)$$

MAP Decision Rule

$$\boxed{\text{Decide } H_1 \text{ if } P(H_1|x) > P(H_2|x)}$$

Likelihood $$P(H_i|x) = \frac{p(x|H_i)P(H_i)}{p(x)}$$ Prior

$$p(x|H_1)P(H_1) \underset{H_2}{\overset{H_1}{\gtrless}} p(x|H_2)P(H_2)$$

Problem #1

A binary bit b ∈ {0,1} is sent over a channel with additive white gaussian noise. The receiver observes a scalar x.

Conditional distributions:

- If b=0, x ~ Gaussian(2, 1) (mean = u, variance = σ^2)
- If b=1, x ~ Gaussian(−2, 1) (mean = u, variance = σ^2)

Priors:

- P(0) = 0.3
- P(1) = 1 - P(0) = 0.7

Derive the MAP and ML decision rules for parameter x.

A *Gaussian random variable* X has density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

with parameters (mean) μ and (variance) $\sigma^2 \geq 0$.

Likelihood Ratio $\Lambda(x) = \frac{p(x|H_1)}{p(x|H_2)}$

MAP Likelihood Ratio $\boxed{\Lambda(x) \underset{H_2}{\overset{H_1}{\gtrless}} \frac{P(H_2)}{P(H_1)}}$

ML Likelihood Ratio $\boxed{\Lambda(x) \underset{H_2}{\overset{H_1}{\gtrless}} 1}$

Hint:

Prior: Probability Value Given (Ex: P(C) = .25)

Likelihood: Conditional Density (Ex: fy(Y | X=x))

Problem #2

(a) A dual-tone sonar is used to detect an underwater vehicle. Two closely spaced frequencies are transmitted, and the echoes are processed into two channel outputs X and Y. Under H0 (no vehicle), X and Y are independent Gaussian random variables with mean 0 and variance 1/4. Under H1 (vehicle present), X and Y are jointly Gaussian with mean 0, equal marginal variance 1, and correlation coefficient ρ = sqrt(¾). The prior probability of H1 is 0.40. Write down the maximum a posteriori (MAP) rule for deciding between H0 and H1 given the observation X = x, Y = y. Hint: Express your decision as a comparison of the posterior probabilities, or equivalently as a likelihood-ratio test weighted by the prior odds.

(b) A different sonar processes two channels X and Y that are always uncorrelated and zero-mean Gaussian. Under H0 (no vehicle), Var(X) = Var(Y) = 1; under H1 (vehicle present), Var(X) = Var(Y) = 2. The prior probability of H1 is 0.40. Derive the maximum likelihood (ML) decision rule for deciding whether H1 is true given X = x, Y = y. State the rule explicitly as an inequality in terms of x and y.

A *Gaussian random variable* X has density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

with parameters (mean) μ and (variance) $\sigma^2 \geq 0$.

Definition (zero-mean, unit-variance). Random variables X and Y with means 0, variances 1, and correlation coefficient ρ are jointly Gaussian if their joint density is

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right), \quad x, y \in \mathbb{R} \text{ [1]}.$$

Likelihood Ratio $\Lambda(x) = \dfrac{p(x|H_1)}{p(x|H_2)}$

MAP Likelihood Ratio $\boxed{\Lambda(x) \underset{H_2}{\overset{H_1}{\gtrless}} \dfrac{P(H_2)}{P(H_1)}}$

ML Likelihood Ratio $\boxed{\Lambda(x) \underset{H_2}{\overset{H_1}{\gtrless}} 1}$

Statistical Independence

X and Y are (statistically) independent iff

$$F_{X,Y}(x,y) = F_X(x)\,F_Y(y) \quad \text{for all } x, y.$$

Implications:

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

Discrete: $P_{X,Y}(x,y) = P_X(x)\,P_Y(y)$.

$P(\bar{A}) = 1 - P(A)$

$P(A) = 1 - P(\bar{A})$

RVs X and Y are *uncorrelated* if $\rho_{XY} = 0$.

MMSE Estimators

Estimator as a Random Variable: An estimator is any function of observed random variables: $\hat{X} = g(Y) \quad \text{or} \quad \hat{q} = g(Y)$ and itself is a random variable. Used for signal denoising, channel estimation, and image restoration; in Bayesian settings, they equal the posterior mean **(MMSE = Minimum Mean Square Error)**.

2. MMSE Estimators

General MMSE problem.

$$g^\star \;=\; \arg\min_g \; \mathbb{E}\Big[\big(X - g(Y)\big)^2\Big]\,,$$

possibly with constraints on the form of g (e.g., constant, affine, or unconstrained)

2.1 Constant MMSE Estimator

If $g(Y) = a$ is constant, the minimizer is

$$a^\star = \mathbb{E}[X]\,, \qquad \mathrm{MSE} = \mathrm{Var}(X).$$

Bayesian estimation setting. Given observation Y, choose a "good" estimator for the random quantity X by minimizing the expected cost of the error $e = X - g(Y)$ under a cost function $C(\cdot)$:

$$\text{minimize} \quad \mathbb{E}\{C[X - g(Y)]\}\,.$$

The most common case takes the *mean-squared error* (MSE) cost:

$$\mathrm{MSE}(g) \;=\; \mathbb{E}\Big[\big(X - g(Y)\big)^2\Big]\,,$$

and seeks the minimum-MSE (MMSE) estimator [1].

2.2 Linear (Affine) MMSE Estimator

If $g(Y) = aY + b$ is affine, the minimizer is

$$a^\star = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}\,, \qquad b^\star = \mathbb{E}[X] - a^\star\,\mathbb{E}[Y].$$

The resulting MSE is

$$\mathrm{MSE} \;=\; \mathrm{Var}(X) - \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(Y)}.$$

If X, Y are zero-mean with unit variances and correlation ρ, then $a^\star = \rho$, $b^\star = 0$, and $\mathrm{MSE} = 1 - \rho^2$.

2.3 Best (Unconstrained) MMSE Estimator

The MMSE-optimal estimator over all measurable functions is the conditional expectation:

$$g^\star(Y) \;=\; \mathbb{E}[X \mid Y].$$

If (X, Y) are jointly Gaussian, this conditional expectation is affine:

$$\mathbb{E}[X \mid Y] \;=\; \mathbb{E}[X] + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}\big(Y - \mathbb{E}[Y]\big).$$

In the zero-mean unit-variance case with correlation ρ, this reduces to $\mathbb{E}[X \mid Y] = \rho Y$ and $\mathrm{Var}(X \mid Y) = 1 - \rho^2$.

Problem #3

Let X and Y be jointly Gaussian random variables with:

E[X] = 7 Var[X] = 1

E[Y] = 0 Var[Y] = .5

Cov[X, Y] = - .125

2. MMSE Estimators

General MMSE problem.

$$g^\star = \arg\min_g \mathbb{E}\left[(X - g(Y))^2\right],$$

possibly with constraints on the form of g (e.g., constant, affine, or unconstrained)

2.1 Constant MMSE Estimator

If $g(Y) = a$ is constant, the minimizer is

$$a^\star = \mathbb{E}[X], \qquad \mathrm{MSE} = \mathrm{Var}(X).$$

2.2 Linear (Affine) MMSE Estimator

If $g(Y) = aY + b$ is affine, the minimizer is

$$a^\star = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}, \qquad b^\star = \mathbb{E}[X] - a^\star\,\mathbb{E}[Y].$$

The resulting MSE is

$$\mathrm{MSE} = \mathrm{Var}(X) - \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(Y)}.$$

If X, Y are zero-mean with unit variances and correlation ρ, then $a^\star = \rho$, $b^\star = 0$, and $\mathrm{MSE} = 1 - \rho^2$.

2.3 Best (Unconstrained) MMSE Estimator

The MMSE-optimal estimator over all measurable functions is the conditional expectation:

$$g^\star(Y) = \mathbb{E}[X \mid Y].$$

If (X, Y) are jointly Gaussian, this conditional expectation is affine:

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X] + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}(Y - \mathbb{E}[Y]).$$

In the zero-mean unit-variance case with correlation ρ, this reduces to $\mathbb{E}[X \mid Y] = \rho Y$ and $\mathrm{Var}(X \mid Y) = 1 - \rho^2$.

a) Find the best MMSE estimator for Y given X.

b) Calculate MSE achieved by the best MMSE estimator for Y given X.

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

HINT: SS-1. Show that for optimal linear prediction of X given Y, the mean-square error is $\sigma_X^2(1-\rho^2)$.

Problem #4

X and Y are random variables with joint density:

f_XY(x,y) = { 2, for 0 ≤ x ≤ 1, 0 ≤ y ≤ x
{ 0 otherwise

(a) Find E[Y|X = x].

(b) Find E{E[Y|X]}.

(c) Find the best MMSE estimator of Y given X.

(d) Find the mean-squared error achieved by the best MMSE estimator for Y given X.

The *conditional expected value of* Y *given* $X = x$ is

$$E[Y|X=x] = E[Y|x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\,dy.$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

$$\mathrm{MSE}(g) = \mathbb{E}\left[(X - g(Y))^2\right]$$

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{when } f_Y(y) > 0$$

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}. \quad [1]$$

Continuous X: $E[X] = \int_{-\infty}^{\infty} x\, f_X(x)\,dx$

2.3 Best (Unconstrained) MMSE Estimator

The MMSE-optimal estimator over all measurable functions is the conditional expectation:

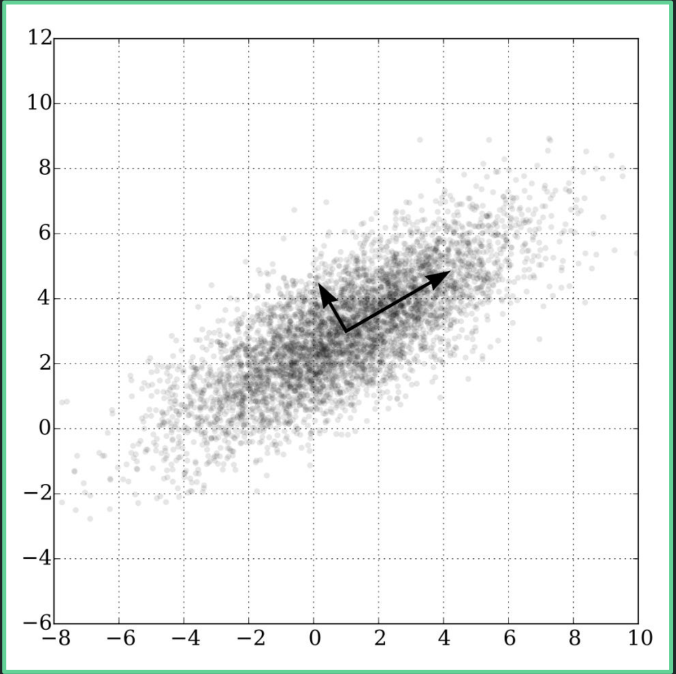
$$g^\star(Y) = \mathbb{E}[X \mid Y].$$

Principal Component Analysis (PCA)

PCA: Is a perpendicular linear transformation that on a space with real inner products which data into a new coordinate system that the greatest variance by some constant scalar projection of the data comes to lie on the first coordinate, and then the second greatest variance follows onto the second coordinate and so on.

Use Case: PCA is usually used for dimensionality reduction (technique used in machine learning to reduce the number of input variables in a dataset)

Graph (On the Right): PCA of a multivariate Gaussian distribution centered at (1, 3) with a standard deviation of 3 in roughly the (0.866, 0.5) direction and of 1 in the orthogonal direction. The vectors shown are the eigenvectors of the covariance matrix scaled by the square root of the corresponding eigenvalue, and shifted so their tails are at the mean.



Reference: https://en.wikipedia.org/wiki/Principal_component_analysis

Example Problem

Feature	EX1	EX2	EX3	EX4
X1	4	8	13	7
X2	11	4	5	14
X3	-4.3052	3.7361	5.6928	-5.1238

Step 1 : Calculate Mean

In step 1 calculate the mean for each feature that is x1 and x2

$$\overline{x1} = \frac{1}{4}(4+8+13+7) = 8$$

$$\overline{x2} = \frac{1}{4}(11+4+5+14) = 8.5$$

$$\overline{x1} = 8$$

$$\overline{x2} = 8.5$$

The calculate mean for feature x1 is 8 and x2 is 8.5

Step 2 : Calculation of Covariance Matrix

In this step we calculate covariance Matrix.The formula to calculate covariance matrix is

$$s = \begin{bmatrix} cov(x1,x1) & cov(x1,x2) \\ cov(x2,x1) & cov(x2,x2) \end{bmatrix}$$

The cov(x1,x1) can be calculated by

$$cov(x_1,x_1) = \frac{1}{N-1}\sum_{K=1}^{N}(x_{1k}-\overline{x1})(x_{1k}-\overline{x1})$$

$$cov(x_1,x_1) = \frac{1}{3}((4-8)^2+(8-8)^2+(13-8)^2+(7-8)^2)$$

$$= 14$$

$$cov(x_1,x_2) = \frac{1}{N-1}\sum_{K=1}^{N}(x_{1k}-\overline{x1})(x_{2k}-\overline{x2})$$

$$cov(x_1,x_2) = \frac{1}{3}((4-8)(11-8.5)+(8-8)(4-8.5)+(13-8)(5-8.5)+(7-8)(14-8.5))$$

$$= -11$$

Step 3: Eigenvalues of the covariance Matrix

The characteristic Equation of the covariance matrix is

$$0 = det(s - \lambda I)$$

$$= \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix}$$

$$= (14-\lambda)(23-\lambda) - (-11)(-11)$$

$$= \lambda^2 - 37\lambda + 201$$

The roots of this equation are

$$\lambda = \frac{1}{2}(37 \pm \sqrt{565})$$

$$\lambda 1 = 30.3849$$

$$\lambda 2 = 6.6151$$

Step 4: Computation of the Eignevectors

$$u = \begin{bmatrix} u1 \\ u2 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = (s - \lambda I)u$$

$$= \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix}\begin{bmatrix} u1 \\ u2 \end{bmatrix}$$

$$= (14-\lambda)u1 - 11u2$$

$$= -11u1 + (23-\lambda)u2$$

$$(14-\lambda)u1 = 11u2 = t$$

$$\frac{u1}{11} = \frac{u2}{14-\lambda} = t$$

$$u1 = 11t$$

$$u2 = (14-\lambda)t$$

Assume t=1

$$u1 = \begin{bmatrix} 11 \\ 14-\lambda \end{bmatrix}$$

To find a unit Eigenvector we compute the length of u1 which is given by

$$||U|| = \sqrt{11^2 + (14-\lambda)^2}$$

$$= \sqrt{11^2 + (14-30.3849)^2}$$

$$= 19.7348$$

$$e1 = \begin{bmatrix} \frac{11}{||U||} \\ \frac{(14-\lambda)}{||U||} \end{bmatrix}$$

$$e1 = \begin{bmatrix} \frac{11}{19.7348} \\ \frac{(14-30.3849)}{19.7348} \end{bmatrix}$$

$$e1 = \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix}$$

Similarly

$$e2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$$

Step 5: Computation of First Principal Components

$$e_1^T \begin{bmatrix} x_{1k} - \overline{x1} \\ x_{2k} - \overline{x2} \end{bmatrix}$$

$$= \begin{bmatrix} 0.5574 - 0.8303 \end{bmatrix} \begin{bmatrix} x_{11} - \overline{x1} \\ x_{21} - \overline{x2} \end{bmatrix}$$

$$= 0.5574(x_{11} - \overline{x1}) - 0.8303(x_{21} - \overline{x2})$$

$$= 0.5574(4-8) - 0.8303(11-8.5)$$

$$= -4.30535$$

Problem #5

A bike-share planner records two features for four stations: average morning checkouts x = [13, 8, 6, 9] and nearby foot traffic y = [20, 3, 11, 10]. To understand the main pattern in these 2D data, use PCA:

Solve the characteristic equation, $\det(\Sigma - \lambda I) = 0$, where Σ is the covariance matrix, λ represents the eigenvalues, and I is the identity matrix.

Identity Matrices

$1 \times 1 \quad [1]$

$2 \times 2 \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$3 \times 3 \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

etc.

$$A\vec{x} = \lambda\vec{x}$$

$n \times n$ Matrix

Eigenvector

Eigenvalue

Calculate the principal component scores

Purpose: To transform the original data into the new principal component space.

Formula: $Y = ZW$ or $Y = Z \cdot E^T$

- Z is the standardized data matrix.
- W is the matrix of eigenvectors, or the matrix of principal components.
- Y is the matrix of principal component scores, with each column being a principal component.
- For a specific data point and the first principal component, the formula is: $PC_1 = e_{11}(x_1 - \bar{x}_1) + e_{12}(x_2 - \bar{x}_2) + \ldots + e_{1p}(x_p - \bar{x}_p)$.

Formula for the covariance between two features, x_1 and x_2:

$$cov(x_1, x_2) = \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{n-1}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the means of features x_1 and x_2.
- n is the number of data points.

Problem #5 Solutions

Mean (x, y):

[9, 11]

Principal components (rows = PC1, PC2):

[[0.31112297 0.95036966],

[0.95036966 -0.31112297]]

(PC1 = u1, PC2 = -u2)

Sign doesn't matter same axis.