RANDOM PROCESSES

Lecture Notes

Ahmet Ademoglu, *PhD*
Bogazici University
Institute of Biomedical Engineering

Discrete Random Variables

A *random variable* is a number whose value depends upon the outcome of a random experiment.

Mathematically, a random variable X is a real-valued function on the space of outcomes which maps a probabilistic event to a real number:

$$X : \Omega \longrightarrow \mathcal{R} \tag{1}$$

A discrete random variable X has finitely or countably many values x_i, $i = 1, 2, \ldots$ and $p(x_i) = P(X = x_i)$ with $i = 1, 2, \ldots$ is called the probability mass function (*pmf*) of X.

Flipping a coin experiment

A real valued variable X is defined as

- If the flip is heads then $\omega_1 = \{H\} \longrightarrow X = 1$
- If the flip is tails then $\omega_2 = \{T\} \longrightarrow X = -1$

Assume that X is a discrete random variable with possible values $x_i, i = 1, 2 \ldots$.

Expected value, also called expectation, average, or mean, of X is

$$E\{X\} = \sum_i x_i P(X = x_i) = \sum_i x_i p(x_i)$$

For any function, $g : \mathcal{R} \longrightarrow \mathcal{R}$,

$$E\{g(X)\} = \sum_i g(x_i) P(X = x_i)$$

Variance, σ^2 of X is

$$\sigma^2 = E\left\{[X - E\{X\}]^2\right\} = E\{X^2\} - E\{X\}^2$$

Rolling a dice N times

$E\{X\} = \frac{1 \cdot n_1 + 2 \cdot n_2 + 3 \cdot n_3 + 4 n_4 + 5 \cdot n_5 + 6 \cdot n_6}{N} = 1 \cdot \frac{n_1}{N} + 2 \cdot \frac{n_2}{N} + \cdots + 6 \cdot \frac{n_6}{N}$

For very large N, $\frac{n_1}{N} = \frac{n_2}{N} \cdots = \frac{n_6}{N} \approx \frac{1}{6}$

$E\{X\} = \frac{1}{6} \cdot (1 + 2 + \cdots + 6) = \frac{3}{2}$

Continuous Random Variables

A random variable X is continuous if there exists a nonnegative function f so that, for every interval B,

$$P(X \in B) = \int_B f(x)dx$$

The function $f = f_X$ is called the density of X.

The function $F = F_X$ given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)ds$$

is called the distribution function of X.

On an open interval where f is continuous,

$$F'(x) = f(x).$$

Expected Value

By analogy with discrete random variables, we define,

$$E\{X\} = \int_{-\infty}^{\infty} xf(x)dx,$$

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x)f(x)dx$$

and variance is computed by the same formula:

$$Var(X) = E\{X^2\} - (E\{X\})^2.$$

$$f(x) = \begin{cases} cx & 0 < x < 4, \\ 0 & otherwise. \end{cases}$$

(a) Determine c. (b) Compute $P(1 \leq X \leq 2)$. (c) Determine $E\{X\}$ and $Var\{X\}$.

(a) $\int_{-\infty}^{\infty} f(x)dx = 1 = \int_{0}^{4} cxdx = c\frac{x^2}{2}\Big|_0^4 = c\frac{16}{2} \longrightarrow c = \frac{1}{8}$ (b)

$P(1 \leq X \leq 2) = \int_{1}^{2} \frac{1}{8}xdx = \frac{x^2}{16}\Big|_1^2 = \frac{4-1}{16} = \frac{3}{16}$

(c) $E\{X\} = \int_{0}^{4} \frac{1}{8}x^2dx = \frac{x^3}{24}\Big|_0^4 = \frac{64}{24} = \frac{8}{3}$, $E\{X^2\} = \int_{0}^{4} \frac{1}{8}x^3dx = \frac{x^4}{32}\Big|_0^4 = \frac{256}{32} = 8$, $Var\{X\} = 8 - \frac{64}{9} = \frac{8}{9}$

Assume that X has density $f(x) = \begin{cases} 3x^2 & \textit{if } x \in [0,1], \\ 0 & \textit{otherwise}. \end{cases}$

Compute the density f_Y of $Y = 1 - X^4$.

$x \in [0,1]$ and $y \in [0,1]$.

$F_Y(y) = P(Y \leq y) = P(1 - X^4 \leq y) = P(1 - y \leq X^4) = P((1-y)^{\frac{1}{4}} \leq X)$

$= 1 - P(X \leq (1-y)^{\frac{1}{4}}) = 1 - \int_0^{(1-y)^{\frac{1}{4}}} 3x^2 dx = 1 - x^3 \Big|_0^{(1-y)^{\frac{1}{4}}} = 1 - (1-y)^{\frac{3}{4}}$

$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{3}{4}(1-y)^{-\frac{1}{4}}$ for $y \in [0,1]$ and $f_y(Y) = 0$, otherwise.

Assume that X is uniform on $[0,1]$. What is the probability that the binary expansion of X starts with 0.010?

Smallest $X = 0.010 = \frac{1}{4}$ and largest $X = 0.011 = \frac{3}{8}$,

$P(\frac{1}{4} \leq X < \frac{3}{8}) = \frac{3}{8} - \frac{1}{4} = \frac{1}{8}$

Let X be a continuous random variable with probability density function on $0 \leq x \leq 1$, $f(x) = 3x^2$ What is the pdf of $Y = X^2$.

$P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = \int_0^{\sqrt{y}} 3t^2 dt = y^{\frac{3}{2}} = F_Y(y)$, $0 \leq y \leq 1$

$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{3}{2} y^{\frac{1}{2}}$ for $0 \leq y \leq 1$.

Theorem

If X is a continuous random variable, then the pdf of $y = g(X)$ is

$$f_Y(y) = \sum_{i=1}^{k} \frac{f_x(x^i)}{|g'(x^i)|}$$

where $x^1, \ldots, x^k$ are the roots of the equation $y = g(x)$.

Suppose X has a Gaussian distribution with a mean of 0 and variance of 1 and $Y = X^2 + 4$. Find the pdf of Y.

$y = g(x) = x^2 + 4$, $g^{-1}(y) = \pm\sqrt{y-4}$, $g'(x) = 2x$

$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-4)} \cdot \frac{1}{2\sqrt{y-4}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-4)} \cdot \frac{1}{2\sqrt{y-4}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-4)} \cdot \frac{1}{\sqrt{y-4}}$

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-4)} \cdot \frac{1}{\sqrt{y-4}} & 4 \leq y \\ 0 & otherwise \end{cases}$$

Let X be a normal distributed, $\mathcal{N}(\mu, \sigma^2)$ random variable and let $Y = \alpha X + \beta$, with $\alpha > 0$. How is Y distributed?

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

$$F_Y(y) = P(Y \leq y) = P(\alpha X + \beta \leq y) = P(X \leq \frac{y-\beta}{\alpha}) = \int_{-\infty}^{\frac{y-\beta}{\alpha}} f_X(x)dx$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \int_{-\infty}^{\frac{y-\beta}{\alpha}} f_X(x)dx = f_Y(\frac{y-\beta}{\alpha})\frac{1}{\alpha}$$

$$f_Y(y) = \frac{1}{(2\pi\sigma^2\alpha^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2\alpha^2}(y-\alpha\mu-\beta)^2} = f_Y(y) = \mathcal{N}(\alpha\mu+\beta, \sigma^2\alpha^2)$$

Standard Normal Distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma}$ is *standard normal*

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

and its distribution function, also called as *error function* is

$$\Phi(z) = F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}x^2} dx$$

Jointly Distributed Random Variables

Flipping two coins experiment :

A real valued variable $x(i)$ with $i = 1, 2$ is defined as

- If the flip is heads then $\omega_1 = \{H\} \longrightarrow x(i) = 1$
- If the flip is tails then $\omega_2 = \{T\} \longrightarrow x(i) = -1$

Joint distribution function

$$F_{x(1),x(2)}(\alpha_1, \alpha_2) = Pr\{x(1) \leq \alpha_1, x(2) \leq \alpha_2\}$$

Joint density function

$$f_{x(1),x(2)} = \frac{\partial^2}{\partial\alpha_1 \partial\alpha_2} F_{x(1),x(2)}(\alpha_1, \alpha_2)$$

Joint Moments

Cross-correlation

$$r_{xy} = E\{xy^*\}$$

Cross-covariance

$$c_{xy} = Cov(x, y) = E\Big\{[x - m_x][y - m_y]^*\Big\} = E\Big\{xy^*\Big\} - m_x m_y^*$$

Correlation Coefficient

$$\rho_{xy} = \frac{E\Big\{xy^*\Big\} - m_x m_y^*}{\sigma_x \sigma_y}$$

$|\rho_{xy}| \leq 1$

Independent, Uncorrelated, Orthogonal Random Variables

Two random variables are statistically independent if the joint probability density function is separable;

$$f_{xy}(\alpha, \beta) = f_x(\alpha) f_y(\beta)$$

Two random variables are uncorrelated if

$$E\left\{xy^*\right\} = E\{x\}E\{y^*\}$$

If x and y are uncorrelated, their covariance c_{xy} is 0.

$$Var\{x + y\} = Var\{x\} + Var\{y\}$$

If $r_{xy} = 0$ then x and y are orthogonal.

Joint Distributions and Independence

Discrete Case

Assume that you have a pair (X, Y) of discrete random variables X and Y. Their joint probability mass function is given by $p(x, y) = P(X = x, Y = y)$ so that

$$P((X, Y) \in A) = \sum_{(x,y)\in A} p(x, y).$$

The marginal probability mass functions are the *pmf*'s of X and Y, given by

$$P(X = x) = \sum_{y} P(X = x, Y = y) = \sum_{y} p(x, y)$$

$$P(Y = y) = \sum_{x} P(X = x, Y = y) = \sum_{x} p(x, y)$$

Independence

Two random variables X and Y are independent if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

for all intervals A and B. In the discrete case, X and Y are independent exactly when

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all possible values x and y of X and Y, that is, the joint *pmf* is the product of the marginal *pmf*'s.

Continuous Case

We say that (X, Y) is a jointly continuous pair of random variables if there exists a joint density $f(x, y) \geq 0$ so that

$$P((X, Y) \in S) = \iint_S f(x, y) dxdy,$$

where S is some nice (say, open or closed) subset of $\mathcal{R}^2$.

Marginal Densities

If f is the joint density of (X, Y), then the two marginal densities, which are the densities of X and Y, are computed by integrating out the other variable:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Two jointly continuous random variables X and Y are independent exactly when the joint density is the product of the marginal ones:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

for all x and y.

$$f(x,y) = \begin{cases} cx^2y & \text{if } x^2 \leq y \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

Determine *(a)* the constant c, *(b)* $P(X \geq Y)$, *(c)* $P(X = Y)$, and (d) $P(X = 2Y)$.

(a) $\int_{-1}^{1}\int_{x^2}^{1} cx^2ydxdy = 1,\ c = \frac{21}{4},$

(b) $P(X \geq Y) = \int_{x=0}^{1}\int_{y=x^2}^{x} \frac{21}{4}x^2ydxdy = \frac{3}{20} \leftrightarrow$ Area between $y = x$ and $y = x^2$.

(c) $= 0.$

(d) $= 0.$

Compute marginal densities and determine whether X and Y are independent.

$f_X(x) = \int_{x^2}^{1} \frac{21}{4}x^2ydy = \frac{21}{8}x^2(1 - x^4)$ for $x \in [-1, 1]$ and 0 elsewhere.

$f_X(x) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4}x^2ydx = \frac{7}{2}y^{\frac{5}{2}}$ for $y \in [0, 1]$ and 0 elsewhere.

$f_{X,Y}(x, y) \neq f_X(x)f_Y(y)$, so X and Y are not independent.

Conditional distributions

The conditional *pmf* of X given $Y = y$ is, in the discrete case, given simply by

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

For a jointly continuous pair of random variables X and Y , the conditional density of X given $Y = y$ is

$$f_X(x|Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

where $f_{X,Y}(x, y)$ is the joint density of (X, Y).

Suppose (X,Y) has joint density $f(x,y) = \begin{cases} \frac{21}{4}x^2y & x^2 \le y \le 1, \\ 0 & \textit{otherwise} \end{cases}$

Compute $f_X(x|Y=y)$.

$$f_X(x|Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)},\ f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4}x^2y dx = \frac{7}{2}y^{\frac{5}{2}}$$

$$f_X(x|Y=y) = \frac{\frac{21}{4}x^2y}{\frac{7}{2}y^{\frac{5}{2}}} = \frac{3}{2}x^2y^{-\frac{3}{2}}$$

Transformation of Random Variables

Theorem

Let $f(\mathbf{x})$ be the value of the probability density of the multivariate continuous random variable $\mathbf{X}$. If the vector-valued function $\mathbf{y} = U(\mathbf{x})$ is differentiable and invertible for all values within the domain of $\mathbf{x}$, then for corresponding values of $\mathbf{y}$, the probability density of $\mathbf{Y} = U(\mathbf{Y})$ is given by

$$f(\mathbf{y}) = f_x(U^{-1}(\mathbf{y})) \cdot \left| \frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) \right|$$

Consider a bivariate random variable $\mathbf{X}$ with states $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and probability density function $f(\mathbf{x}) = \frac{1}{2\pi} \exp(-\frac{1}{2}\mathbf{x}^T\mathbf{x})$. If $\mathbf{A}$ is a transformation matrix defined as $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, what is the probability density function of $\mathbf{y} = \mathbf{A}\mathbf{x}$?

$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \mathbf{y}$, $f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}) = \frac{1}{2\pi} \exp(-\frac{1}{2}\mathbf{y}^T\mathbf{A}^{-T}\mathbf{A}\mathbf{y})$

$\frac{\partial}{\partial \mathbf{y}}\mathbf{A}^{-\mathbf{1}}\mathbf{y} = \mathbf{A}^{-T}$, $\left|\frac{\partial}{\partial \mathbf{y}}\mathbf{A}^{-\mathbf{1}}\mathbf{y}\right| = \left|\mathbf{A}^{-\mathbf{T}}\right| = \frac{1}{|\mathbf{A}|} = \frac{1}{ad-bc}$

$f(\mathbf{y}) = \frac{1}{2\pi} \exp(-\frac{1}{2}\mathbf{y}^T\mathbf{A}^{-T}\mathbf{A}\mathbf{y}) \cdot \frac{1}{ad-bc}$

Let two resistors, having independent resistances, X_1 and X_2, uniformly distributed between 9 and 11 ohms, be placed in parallel. Find the probability density function of resistance Y_1 of the parallel combination.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad y_1 = \frac{x_1 x_2}{x_1+x_2},\ y_2 = x_2 \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{x_1 x_2}{x_1+x_2} \\ x_2 \end{bmatrix}$$

$$\mathbf{y} = U(\mathbf{x}),\ U^{-1}(\mathbf{y}) = \begin{bmatrix} \frac{y_1 y_2}{y_2 - y_1} \\ y_2 \end{bmatrix},\ \frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) = \begin{bmatrix} \frac{y_2^2}{(y_2-y_1)^2} & \frac{-y_1^2}{(y_2-y_1)^2} \\ 0 & 1 \end{bmatrix}$$

$$\left| \frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) \right| = \frac{y_2^2}{(y_2-y_1)^2}$$

$$9 \leq x_2 \leq 11 \longrightarrow 9 \leq y_2 \leq 11, \qquad 9 \leq x_1 \leq 11 \longrightarrow \frac{9 \cdot 9}{9+9} = \frac{9}{2} \leq y_1 \leq \frac{11 \cdot 11}{11+11} = \frac{11}{2}$$

$$f_Y(y) = \begin{cases} \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{y_2^2}{(y_2-y_1)^2} & \frac{9}{2} \leq y_1 \leq \frac{11}{2},\ 9 \leq y_2 \leq 11 \\ 0 & \text{otherwise} \end{cases}$$

Determine the pdf of $Y = X_1 + X_2$ where X_1 and X_2 are independent random variables.

$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\quad y_1 = x_1 + x_2,\ y_2 = x_2 \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{Ax}$

$\mathbf{y} = U(\mathbf{x}), \quad U^{-1}(\mathbf{y}) = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \left| \frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) \right| = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \mathbf{A}^{-1}$

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{y}) \left| \frac{\partial}{\partial \mathbf{y}} U^{-1}(\mathbf{y}) \right| = f_{\mathbf{X}}(y1 - y2, y2) \cdot 1 = f_{X_1}(y2 - y1) f_{X_2}(y2)$$

$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(y_1, y_2) = f_{X_1}(y1 - y2) f_{X_2}(y2)$

$$f_{Y_1}(y_1) = f_Y(y) = \int f_{\mathbf{X}}(y1 - y2, y2) dy_2 = \int f_{X_1}(y_2 - y_1) \cdot f_{X_2}(y_2) dy_2$$

Bias, Variance and Consistency of Estimation

Bias

$\theta - E[\hat{\theta}_N] \longrightarrow$ If an estimator is unbiased then $\theta = E[\hat{\theta}_N]$.

Variance

$$Var[\hat{\theta}_N] = E[|\hat{\theta}_N - E[\hat{\theta}_N]|^2] = 0.$$

Consistency

$$\lim_{N\to\infty} = Var[\hat{\theta}_N] = \lim_{N\to\infty} E[|\hat{\theta}_N - E[\hat{\theta}_N]|^2] = 0$$

Bias-Variance Trade-off

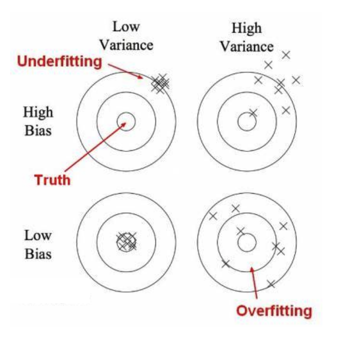
Let's assume we make a measurement such that the independent parameter x is related to our measurements as

$$y = f(x) + \epsilon$$

where the error term often has a normal distribution as $\mathbf{N}(\epsilon|0, \sigma^2)$.

$$Error = E\{(y - \hat{f}(x))^2\} = \underbrace{[f(x) - E\{\hat{f}(x)\}]^2}_{Bias} + \underbrace{f(x) - E\{\hat{f}(x)\}}_{Variance} + \underbrace{\sigma_e^2}_{Irreducible\ Error}$$

$\hat{f}(x)$: Estimate of y based on a model estimator and the observation x.

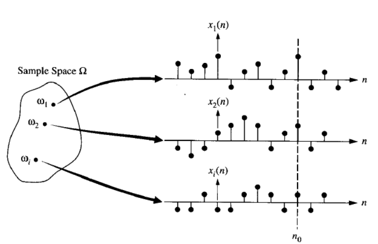


Random Processes

$$x[n] = A\cos[\omega_0 n]$$

$$\ldots, x[-2], x[-1], x[0], x[1], x[2], \ldots$$

$$f_{x[n]}(\alpha) = Pr\{x[n] \leq \alpha\},\ f_{x[n]}(\alpha) = \frac{d}{d\alpha} F_{x[n]}(\alpha)$$



$$m_x[n] = E\Big\{x[n]\Big\}$$

$$\sigma_x^2(n) = E\Big\{|x[n] - m_x(n)|^2\Big\}$$

Crosscovariance between $x[n]$ and $y[n]$:

$$c_{xy(k,l)} = E\Big\{x[k]y[l]^*\Big\} - E\{x(k)\}E\{y[l]^*\}$$

Crosscorrelation between $x[n]$ and $y[n]$:

$$r_{xy(k,l)} = E\Big\{x[k]y[l]^*\Big\}$$

The Harmonic Process

$$x[n] = A\sin[n\omega_0 + \phi], \quad f_\Phi(\phi) = \begin{cases} \frac{1}{2\pi} & |\phi| \le \pi \\ 0 & otherwise \end{cases}$$

$m_x[n] = 0$ and $r_x(k,l) = \frac{|A|^2}{2}\cos((k-l)\omega)$

The autocorrelation matrix of a random vector $\mathbf{x}$ is

$$E\{\mathbf{x}\mathbf{x}^H\}.$$

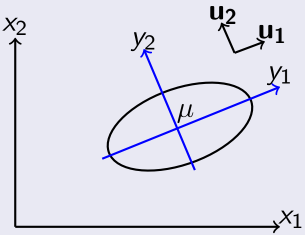
and its autocovariance matrix is

$$E\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})^H\} = E\{\mathbf{x}\mathbf{x}^H\} - E\{\mathbf{x}\}E\{\mathbf{x}\}^H.$$

Multivariate Gaussian Distribution

$$\mathcal{N}(\mathbf{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}$$

Geometry of Multivariate Gaussian



$$\Delta^2 = (\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)$$

$$\Sigma^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^T$$

$$y_i = \mathbf{u}_i^T(\mathbf{x}-\mu)$$

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}$$

Show that $E\{\mathbf{x}\} = \mu$ for a Multivariate Gaussian.

$$E\{\mathbf{x}\} = \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}\mathbf{x}d\mathbf{x} = \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}}(\mathbf{z}+\mu)d\mathbf{z}$$

$$= \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}}\mathbf{z}d\mathbf{z} + \mu\int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}}d\mathbf{z} = 0 + \mu\cdot 1 = \mu$$

Show that $Var\{\mathbf{x}\} = \Sigma$ for a Multivariate Gaussian.

$\Sigma = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^T, \quad \mathbf{x} - \mu = \mathbf{U}\mathbf{y} = \sum_{i=1}^{D} \mathbf{u}_i y_i \quad \frac{\partial x_i}{\partial y_j} = \mathbf{J}_{ij} = \mathbf{U}_{ji}, \quad |\mathbf{J}| = |\mathbf{U}^T|$ since $\mathbf{U}^T = \mathbf{U}^{-1}$

$\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $|\mathbf{U}^T\mathbf{U}| = |\mathbf{U}^T||\mathbf{U}| = |\mathbf{U}^T||\mathbf{U}^T| = 1$, $|\mathbf{J}| = 1$

$dx_i = |\frac{\partial x_i}{\partial y_j}| dy_j$ or $d\mathbf{x} = |\mathbf{J}| d\mathbf{y} = d\mathbf{y}$

and remembering that for $\mathbf{y} = g(\mathbf{x})$ $f(\mathbf{y}) = f_x(g^{-1}(\mathbf{y})) \cdot \left| \frac{\partial}{\partial \mathbf{y}} g^{-1}(\mathbf{y}) \right|$

$$E\{(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T\} = \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)} (\mathbf{x}-\mu)(\mathbf{x}-\mu)^T d\mathbf{x}$$

$$= \frac{1}{(2\pi)^{D/2}\prod_{p=1}^{D}\lambda_p^{1/2}} \int \sum_{i,j} \mathbf{u}_i y_i y_j \mathbf{u}_j^T e^{-\frac{1}{2}\sum_{k=1}^{D}\frac{y_k^2}{\lambda_k}} d\mathbf{y}$$

$$= \frac{1}{(2\pi)^{D/2}\prod_{p=1}^{D}\lambda_p^{1/2}} \left[\mathbf{u}_1\mathbf{u}_1^T \int y_1^2 e^{-\frac{1}{2}\frac{y_1^2}{\lambda_1}} dy_1 \int e^{-\frac{1}{2}\frac{y_2^2}{\lambda_2}} dy_2 \ldots \right] +$$

$$\left[\mathbf{u}_2\mathbf{u}_2^T \int e^{-\frac{1}{2}\frac{y_1^1}{\lambda_1}} dy_1 \int y_2^2 e^{\frac{y_2^2}{\lambda_2}} dy_2 \ldots \right] + \cdots +$$

$$\left[\mathbf{u}_D\mathbf{u}_D^T \int e^{-\frac{1}{2}\frac{y_1^1}{\lambda_1}} dy_1 \ldots \int y_D^2 e^{-\frac{1}{2}\frac{y_D^2}{\lambda_D}} dy_D \right] = \mathbf{u}_1\mathbf{u}_1^T\lambda_1 + \mathbf{u}_2\mathbf{u}_2^T\lambda_2 + \cdots = \sum_{i=1}^{D}\lambda_i\mathbf{u}_i\mathbf{u}_i^T = \Sigma$$

Stationarity: Statistical version of time invariance

1st order stationarity :

$f_{x(n)}(\alpha) = f_{x(n+k)}(\alpha) \longrightarrow m_x(n) = m_x$ and $\sigma_x^2(n) = \sigma_x^2$

2nd order stationarity :

$$f_{x_1(n),x_2(n)}(\alpha_1, \alpha_2) = f_{x_1(n+k),x_2(n+k)}(\alpha_1, \alpha_2)$$

$$r_x(k, l) = \int_{-\infty}^{\infty} \alpha\beta f_{x(k),x(l)}(\alpha, \beta) d\alpha d\beta$$

$$r_x(k+n, l+n) = \int_{-\infty}^{\infty} \alpha\beta f_{x(k+n),x(l+n)}(\alpha, \beta) d\alpha d\beta = r_x(k, l) = r_x(k-l, 0)$$

Wide sense stationarity = 1st order + 2nd order stationarity

For Gaussian Processes :

WS stationarity $\leftrightharpoons$ Strict Sense Stationarity

Joint WS Stationarity between two processes

$r_{xy}(k,l) = r_{xy}(k+n, l+n) = r_{xy}(k-l) = E\left\{x(k)y(l)^*\right\}$

Properties

1. $r_x(k) = r_x(-k)^*$
2. $r_x(0) = E\left\{|x(k)|^2\right\} \geq 0$
3. $r_x(0) \geq |r_x(k)|$
4. If $r_x(k)$ is periodic with period k_0 and $r_x(k_0) = r_x(0)$ then $x(n)$ is *mean square periodic i.e.*, $E\left\{|x(n) - x(n-k_0)|^2\right\} = 0$. As an example random phase harmonic $x(n) = A\cos(n\omega_0 + \phi)$ with $r_x(k) = \frac{1}{2}A^2\cos(k\omega_0)$.

Autocorrelation Matrix

For a random vector $\mathbf{x} = [x(0), x(1), \ldots, x(p)]^T$

$$\mathbf{R}_x = E\left\{\mathbf{x}\mathbf{x}^H\right\} = \begin{bmatrix} r_x(0) & r_x^*(1) & r_x^*(2) & \cdots & r_x^*(p) \\ r_x(1) & r_x(0) & r_x^*(1) & \cdots & r_x^*(p-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_x(p) & r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \end{bmatrix}$$

Crosscorrelation Matrix

$$\mathbf{C}_x = E\left\{(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^H\right\} = \mathbf{R}_x - \mathbf{m}_x \mathbf{m}_y^H$$

Toeplitz Matrix

Diagonal terms of the matrix are identical.

Properties

1. $\mathbf{R}_x$ of a WS Stationary process is Toeplitz and Hermitian.
2. $\mathbf{R}_x$ of a WS Stationary process is nonnegative, $\mathbf{R}_x > 0$.
3. λ_k of $\mathbf{R}_x$ of a WS stationary process are real and nonnegative.

Sample Mean Ensemble averaging over realizations: $\hat{m}_x(n) = \frac{1}{L}\sum_{i=1}^{L} x_i(n)$

Time averaging over a single realization : $\hat{m}_x(n) = \frac{1}{N}\sum_{i=1}^{N} x(n)$

x(n) = Acos($n\omega_0$) where A is a r.v. having values 1 or 2 equally likely.

$E\{x(n)\} = E\{A\}\cos(n\omega_0) = 1.5\cos(n\omega_0)$
For large N, $\hat{m}_x(N) \approx 0 \neq E\{x(n)\}$.

$x(n) = A$ where A is a r.v. assuming 1 if *Heads* and -1 if *Tails* with equal likelihood.

$m_x(n) = E\{x(n)\} = E\{A\} = 0$
The sample mean :
$\hat{m}_x(N) = 1$ and $\hat{m}_x(N) = -1$ with an equal probability of 0.5.
$\hat{m}_x(N)$ does not converge to true mean m_x.

$x(n)$ is a *Bernouilli* sequence assuming 1 if *Heads* and -1 if *Tails* th a process.

$m_x(n) = E\{x(n)\} = E\{A\} = 0$

$$\hat{m}_x(N) = \frac{1}{N}\sum_{i=0}^{N} x(n) = \frac{n_1}{N} - \frac{n_2}{N} \longrightarrow 0 \text{ as } N \longrightarrow \infty$$

$\hat{m}_x(N)$ converges to true mean m_x.

Ergodicity

A process is ergodic if the ensemble average is replaced by time average.

$$\hat{m}_x(N) = \frac{1}{N}\sum_{n=0}^{N-1} x(n) = \hat{m}_x \text{ and } \hat{r}_x(k,N) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)x^*(n-k)$$

Ergodic in the mean

$$\lim_{N\to\infty} \hat{m}_x(N) = m_x$$

Sample mean has convergence in the mean square sense if

Asymptotically unbiased : $\lim_{N\to\infty} E\{\hat{m}_x(N)\} = m_x$

Variance goes to zero as $N \longrightarrow \infty$: $\lim_{N\to\infty} Var\{\hat{m}_x(N)\} = 0$

Mean Ergodic Theorem 1:$x(n)$ is ergodic in the mean if $\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} c_x(k) = 0$

Mean Ergodic Theorem 2 : $x(n)$ is ergodic in the mean if $\lim_{k\to\infty} c_x(k) = 0$

Ergodic in the autocorrelation

If $\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} c_x^2(k) = 0 \longrightarrow \lim_{N\to\infty} \hat{r}_x(k,N) = r_x(k)$

White Noise

A WS stationary process with $c_v(k) = \sigma_v^2 \delta(k)$

Power Spectrum or Power Spectral Density

$$P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_x(k) e^{-j\omega} \longleftrightarrow r_x(k) = \int_{-\pi}^{\pi} P_x(e^{j\omega}) e^{jk\omega} d\omega$$

Alternatively, $P_x(z) = \sum_{k=-\infty}^{\infty} r_x(k) z^{-k}$

Properties

1. A WS stationary process $x(n)$ has real valued power spectrum, $P_x(e^{j\omega}) = P_x^*(e^{j\omega})$ and if $x(n)$ is real, $P_x(e^{j\omega})$ is even.
2. The power spectrum of a WS stationary process is nonnegative, $P_x(e^{j\omega}) \geq 0$.
3. The power of a zero mean WS stationary process is

$$r_x(0) = E\left\{|x(n)|^2\right\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) d\omega$$

Power Spectrum of White Noise

$P_v(e^{j\omega}) = \sigma_v^2$

Power Spectrum of sinusoid with random phase

$r_x(k) = \frac{1}{2}A^2 cos(k\omega_0)$

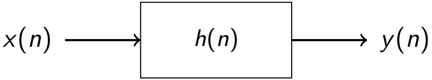
$P_x(e^{j\omega}) = \frac{1}{2}\pi A^2 \left[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)\right]$

Property

The eigenvalues of a $n \times n$ autorcorrelation matrix of a zero mean WS stationary random process are bounded by

$$\min_{\omega} P_x(e^{j\omega}) \leq \lambda_i \leq \max_{\omega} P_x(e^{j\omega})$$

Filtering random processes



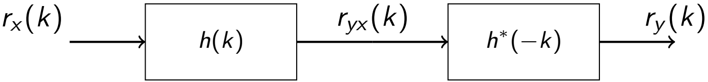
$$E\left\{y(n)\right\} = E\left\{\sum_{k=-\infty}^{\infty} h(k)x(n-k)\right\} = \sum_{k=-\infty}^{\infty} h(k)E\left\{x(n-k)\right\}$$

$$= m_x \sum_{k=-\infty}^{\infty} h(k) = m_x H(e^{j0})$$

$$r_{yx}(k) = r_x(k) * h(k)$$

$$r_y(k) = r_{yx}(k) * h^*(-k) = \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h(l) r_x(m-l+k) h^*(m)$$

$$= r_x(k) * h(k) * h^*(-k)$$



$$r_h(k) = h(n) * h^*(-k) \text{ and } r_y(k) = r_x(k) * r_h(k)$$

$$r_y(0) = \sigma_y^2 = \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h(l) r_x(m-l) h^*(m) = \mathbf{h}^H \mathbf{R}_x \mathbf{h}$$

$$P_y(e^{j\omega}) = P_x(e^{j\omega}) |H(e^{jw})|^2 \text{ and } P_y(z) = P_x(z) H(z) H^*(1/z^*)$$

Filtering White Noise $w(n)$ with $\sigma_w^2 = 1$

$$H(z) = \frac{1}{1-0.25z^{-1}}$$
$$P_x(z) = \sigma_w^2 H(z) H^*(z^{-1}) = \frac{1}{(1-0.25z^{-1})(1-0.25z)}$$
$$P_x(z) = \frac{16/15}{1-0.25z^{-1}} - \frac{4/5}{1-4z^{-1}}$$
$$r_x(k) = \frac{16}{15}\left(\frac{1}{4}\right)^k u(k) + \frac{16}{15}4^k u(-k-1) = \frac{16}{15}\left(\frac{1}{4}\right)^{|k|}$$

Generating a process with $P_x(e^{j\omega}) = \frac{5+4\cos(\omega)}{10+6\cos(\omega)}$

$$P_x(z) = \frac{5+2e^{j2\omega}+2e^{-j2\omega}}{10+3e^{j\omega}+3e^{-j\omega}} \longrightarrow P_x(z) = \frac{2z^2+1}{3z+1}\frac{2z^{-2}+1}{3z^{-1}+1}$$
$$P_x(z) = H(z)H(z^{-1}) \longrightarrow H(z) = \frac{2z^2+1}{3z+1} = z\frac{2}{3}\frac{1+\frac{1}{2}z^{-2}}{1+\frac{1}{3}z^{-1}}$$
$$h(n) = \frac{2}{3}\left(-\frac{1}{3}\right)^n u(n) + \frac{1}{3}\left(-\frac{1}{3}\right)^{n-2} u(n-2)$$

Spectral Factorization

Assume that $x(n)$ is a WS stationary random process with continuous $P_x(e^{j\omega})$ *i.e.* which contains no periodic components;

$$P_x(z) = \sigma_0^2(z)Q(z)Q^*(1/z^*)$$

Also assume that $\ln P_x(z)$ is *analytic* which means all its derivatives are continuous so that it can be expressed as a series expansion in $\rho < |z| < 1/\rho$ that contains the unit circle.

$$\ln P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} c(k)e^{-jk\omega} \overset{\mathcal{F}}{\longleftrightarrow} c(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln P_x(e^{j\omega})e^{jk\omega}d\omega$$

$$c(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln P_x(e^{j\omega})d\omega$$

$c(-k) = c^*(k)$ since $P_x(e^{j\omega}) > 0$.

$$P_x(z) = \exp\left\{\sum_{k=-\infty}^{\infty} c(k) z^{-k}\right\} =$$

$$\exp\{c(0)\}\exp\left\{\sum_{k=1}^{\infty} c(k) z^{-k}\right\}\exp\left\{\sum_{k=-\infty}^{-1} c(k) z^{-k}\right\}$$

Define $Q(z) = \exp\left\{\sum_{k=1}^{\infty} c(k) z^{-k}\right\}$; $\quad |z| > \rho$

$q(k)$: a causal and stable sequence. $Q(z) = q(0) + q(1)z^{-1} + \dots$

$q(0) = \lim_{z\to\infty} Q(z) = 1$

$Q(z)$ is analytic for $|z| > \rho$, $Q(z)$ is a *minimum phase* filter (no poles/zeros outside of unit circle).

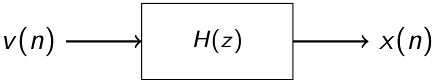
$Q(z)$ has a stable and causal inverse $1/Q(z)$

$$\exp\left\{\sum_{k=-\infty}^{-1} c(k) z^{-k}\right\} = \exp\left\{\sum_{k=1}^{\infty} c^*(k) z^{k}\right\} = \exp\left\{\sum_{k=1}^{\infty} c(k)(1/z^*)^{-k}\right\}^*$$

$= Q^*(1/z^*)$ so that $P_x(z)$ can be spectrally factorized.

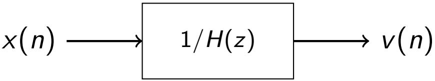
$$P_x(z) = \sigma_0^2 Q(z) Q^*(1/z^*)$$

Random process modeled as the output of a minimum phase filter to white noise input



$P_v(z) = \sigma_0^2$ $\qquad$ $P_x(z) = \sigma_0^2 H(z) H^*(1/z^*)$

A whitening filter for the process $x(n)$



$P_x(z) = \sigma_0^2 H(z) H^*(1/z^*)$ $\qquad$ $P_v(z) = \sigma_0^2$

When $P_x(z) = \frac{N(z)}{D(z)}$ is spectrally factorized

$$P_x(z) = \sigma_0^2 Q(z) Q^*(1/z^*) = \sigma_0^2 \left[\frac{B(z)}{A(z)}\right] \left[\frac{B^*(1/z^*)}{A^*(1/z^*)}\right]$$

$$B(z) = 1 + b(1)z^{-1} + \ldots + b(q)z^{-q}$$

$$A(z) = 1 + a(1)z^{-1} + \ldots + a(p)z^{-p}$$

Both $B(z)$ and $A(z)$ are monic with all roots inside unit circle.

Wold Decomposition Theorem

Any WS stationary random process $x(n)$ can be decomposed into a totally predictable signal and a random process orthogonal to it *i.e.*

$$x(n) = x_p(n) + x_r(n)$$

where $x_p(n) = \sum_{k=1}^{\infty} a(k) x_p(n-k)$ and $E\left\{x_r(m) x_p^*(n)\right\} = 0$