Probability Theory and Stochastic Processes
Lecture Notes – Part 2

İstanbul Teknik Üniversitesi

Mustafa Kamasak, PhD

v2018.12.31

Multivariate Random Variables

Joint Distributions of Discrete RV

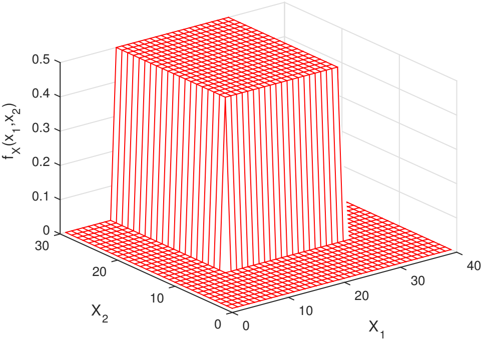
- Suppose there are k (≥ 2) discrete rv. $X_1, X_2, \cdots, X_k$ where X_i can take $x_i \in \mathcal{X}_i$.
- Joint probability of $\mathbf{X} = (X_1, X_2, \cdots, X_k)$ is

$$f(\mathbf{X}) = P(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k) \qquad \forall x_i \in \mathcal{X}_i \ .$$

- $f(\mathbf{X})$ is a joint pmf iff
  a) $f(\mathbf{X}) \geq 0$ for $\forall \mathbf{x} = (x_1, x_2, \cdots, x_k) \in \prod_{i=1}^{k} \mathcal{X}_i$
  b) $\sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{X}) = 1$

Joint pmf Example

Is this a valid pmf?



Marginal Distribution of Discrete RV

- Sometimes we are not interested in some of the random variable
- From $\mathbf{X} = (X_1, X_2, \cdots, X_k)$, we are not interested in X_1, then

$$f(X_2, \cdots, X_k) = \sum_{x_1 \in \mathcal{X}_1} f(X_1, X_2, \cdots, X_k)$$

- If we are only interested in rv X_1 from $\mathbf{X} = (X_1, X_2, \cdots, X_k)$

$$f(X_1) = \sum_{x_2 \in \mathcal{X}_2} \cdots \sum_{x_k \in \mathcal{X}_k} f(X_1, X_2, \cdots, X_k)$$

- $f(X_1)$ is called the marginal density of X_1

Conditional Distribution of Discrete RV

- The conditional density of X_1 given $X_2 = x_2$ is

$$f_{i|j}(x_i) = \frac{f(x_i, x_j)}{f_j(x_j)}$$

  $\forall x_i \in \mathcal{X}_i$

- Conditional mean

$$E(X_i|X_j = x_j) = \sum_{x_i \in \mathcal{X}_i} x_i f_{i|j}(x_i)$$

  for any fixed value of $x_j \in \mathcal{X}_j$.

- Notice that the conditional mean is a function of X_j. Hence the conditional mean is also a random variable.

Multinomial Distribution

- Let $\mathbf{X} = (X_1, X_2, \cdots, X_k)$ has multinomial distr. with $Mult(N, p_1, \cdots, p_k)$

$$P(X_1 = x_1, \cdots, X_r = x_r) = \frac{N!}{\prod_{i=1}^{r} x_i!} \prod_{i=1}^{r} (p_i)^{x_i}$$

- Any subset of $\mathbf{X}$ of size $r \leq k$ has also multinomial distr.

$$P(X_1, \cdots, X_r | \sum_{j \neq \{1, \cdots, r\}} x_j = T) = \frac{(N-T)!}{\prod_{i=1}^{r} x_i!} \prod_{i=1}^{r} (p_i)^{x_i}$$

Multinominal Distribution – Example

- Assume a fair dice is tossed 20 times
- Let N_i be the number of dice lands with i on top $i \in \{0, \cdots, 6\}$
- Joint probability is:

$$f(X_1, \cdots, X_6) = Mult(20, \frac{1}{6}, \cdots, \frac{1}{6})$$

- The marginal distribution of X_i is Binomial $(20, \frac{1}{6})$ $\forall i$

Multinominal Distribution – Example

- Assume a fair dice is tossed 20 times
- Let N_i be the number of dice lands with i on top
  $i \in \{0, \cdots, 6\}$
- if $X_2 + X_3 + X_4 + X_6 = 10$ find
  $P(X_1, X_5 | X_2 + X_3 + X_4 + X_6 = 10)$
  $P(X_1 = x_1, X_5 = x_5 | X_2 + X_3 + X_4 + X_6 = 10) = Mult(10, \frac{1}{2}, \frac{1}{2})$

$$P(X_1 = x_1, X_5 = x_5 | X_2 + X_3 + X_4 + X_6 = 10) = \frac{10!}{x_1! x_5!} (\frac{1}{2})^{x_1} (\frac{1}{2})^{x_5}$$

Continuous Multivariate Distributions

- If the support of continuous rv $X_1 \in \mathcal{X}_1$ and $X_2 \in \mathcal{X}_2$, then the support of $(X_1, X_2) \in \mathcal{X}_1 \times \mathcal{X}_2$
- Hence $f(x_1, x_2) > 0$ if $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. Otherwise $f(x_1, x_2) = 0$.
- $f(x_1, x_2)$ is called joint pdf of (X_1, X_2) if

$$\int_{\mathcal{X}_1} \int_{\mathcal{X}_2} f(x_1, x_2) dx_1 dx_2 = 1$$

- Marginal pdf of X_1 is

$$f_1(x_1) = \int_{\mathcal{X}_2} f(x_1, x_2) dx_2$$

Continuous Multivariate Distributions

- Conditional pdf of X_1 given $X_2 = x_2$ is

$$f_{1|2}(x_1|X_2 = x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}$$

- Conditional mean and conditional variance are

$$\mu_{1|2} = E(X_1|X_2 = x_2) = \int_{\mathcal{X}_1} x_1 f_{1|2}(x_1|X_2 = x_2) dx_1$$

$$\sigma^2_{1|2} = E(X_1^2|X_2 = x_2) - \mu^2_{1|2}$$

Continuous Multivariate Distributions

- Expected value of $g(X_1, X_2)$ is

$$E(g(X_1, X_2)) = \int_{\mathcal{X}_1} \int_{\mathcal{X}_2} g(x_1, x_2) f(x_1, x_2) dx_1 dx_2$$

- Conditional expected value of $h(X_1)$ is

$$E(h(X_1)|X_2 = x_2) = \int_{\mathcal{X}_1} h(x_1) f_{i|j}(x_1|X_2 = x_2) dx_1$$

Probability of Events with Continuous Multivariate Distributions

- The probability of event A is:

$$P(A) = \int\int_{A \cap (\mathcal{X}_1 \times \mathcal{X}_2)} f(x_1, x_2) dx_1 dx_2$$

- The conditional probability of evet B is:

$$P(B|X_2 = x_2) = \int_{B \cap \mathcal{X}_1} f_{1|2}(x_1|X_2 = x_2) dx_1$$

Covariance

- Covariance aims to capture joint dependence between two rv.

$$
\begin{aligned}
Cov(X_1, X_2) &= E((X_1 - \mu_{X_1})(X_2 - \mu_{X_2})) \\
&= E(X_1 X_2) - \mu_{X_1}\mu_{X_2}
\end{aligned}
$$

- $Cov(X_1, X_2) = Cov(X_2, X_1)$. Hence Covariance matrix is symmetric.
- $Cov(X_1, X_1) = \sigma^2_{X_1}$
- $Cov(X_1, k) = 0$ when $k \in \mathbb{R}$ is a constant real number
- $Cov(X_1 + X_2, Y_1 + Y_2) = Cov(X_1, Y_1) + Cov(X_1, Y_2) + Cov(X_2, Y_1) + Cov(X_2, Y_2)$

Covariance Matrix

Covariance matrix is

$$
\begin{aligned}
C &= E((\mathbf{X} - \mu_X)(\mathbf{X} - \mu_X)^T) \\
&= \begin{bmatrix} C(X_1, X_1) & C(X_1, X_2) & \cdots & C(X_1, X_k) \\ C(X_2, X_1) & C(X_2, X_2) & \cdots & C(X_2, X_k) \\ \cdots & \cdots & \cdots & \cdots \\ C(X_k, X_1) & C(X_k, X_2) & \cdots & C(X_k, X_k) \end{bmatrix}
\end{aligned}
$$

- $Cov(X_1, X_1) = \sigma^2_{X_1} \rightarrow$ Diagonals of covariance matrix is individual variances of rv.
- $Cov(X_1, X_2) = Cov(X_2, X_1) \rightarrow$ Covariance matrix is symmetric.
- Covariance matrix is positive semidefinite ie. $y^T C y \geq 0 \; \forall \mathbf{y}$

$$
\begin{aligned}
y^T C y &= y^T E((\mathbf{X} - \mu_X)(\mathbf{X} - \mu_X)^T) y \\
&= E(y^T (\mathbf{X} - \mu_X)(\mathbf{X} - \mu_X)^T y) \\
&= E((y^T (\mathbf{X} - \mu_X))^2) \geq 0
\end{aligned}
$$

Covariance Matrix

Hence, a valid covariance matrix

- Should have nonnegative diagonal entries
- Should be symmetrical
- Should have nonnegative eigenvalues (covers the first item)

Covariance Matrix

```
function val = isValidCovariance(C)

% returns 1 if C matrix is a valid
% covariance matrix. If not 0 is returned

% check if all diagonals are positive
d = diag(C);
if sum(d < 0) ~= 0
    val = 0;
    return
end

% check if symmetric
if C' ~= C
    val = 0;
    return
end

% check if positive semidefinite
[d,v] = eig(C);
d = diag(d);
if sum(d < 0) ~= 0
    val = 0;
    return
end

val = 1;
```

Correlation Coefficient

- Correlation coefficient ρ is a value in [-1,1] range.
- Definition:
$$\rho_{12} = \frac{Cov(X_1, X_2)}{\sigma_1 \sigma_2}$$
- $-1 \leq \rho_{12} \leq 1$ and $|\rho_{12}| \leq 1$
- $|\rho_{12}| = 1$ when X_1 and X_2 are linearly related.
- $|\rho_{12}| = 0$ when X_1 and X_2 are uncorrelated.

Properties related to Multivariate RV and Covariance

For fixed $a_i, b_i \in \mathbb{R}$:

- $E(\sum_i a_i X_i) = \sum_i a_i E(X_i)$
- $Cov(X_i, X_j) = E(X_i, X_j) - E(X_i)E(X_j)$
- Variance of $\sum_i a_i X_i$ is:

$$Var(\sum_i a_i X_i) = \sum_i a_i^2 \sigma_i^2 + 2 \sum_i \sum_{j, i \neq j} a_i a_j Cov(X_i, X_j)$$

$$Cov(\sum_i a_i X_i, \sum_j b_j Y_j) = \sum_i \sum_j a_i b_j Cov(X_i, Y_j)$$

Independence of Multivariate RV

- $X_1, X_2, \cdots, X_k$ form a collection of independent rv iff

$$f(x_1, x_2, \cdots, x_k) = \prod_{i=1}^{k} f_i(x_i)$$

- In order to show that $X_1, X_2, \cdots, X_k$ are dependent, it is sufficient to show a particular value of $x_1, x_2, \cdots, x_k$ such that

$$f(x_1, x_2, \cdots, x_k) \neq \prod_{i=1}^{k} f_i(x_i)$$

- For independent $X_1, X_2, \cdots, X_k$ and real valued $g_i(x_i)$

$$E(\prod_{i=1}^{k} g_i(x_i)) = \prod_{i=1}^{k} E(g_i(x_i))$$

Independence and Correlation

- If X_i and X_j are independent rv then

$$\begin{aligned} Cov(X_i, X_j) &= E(X_i X_j) - \mu_i \mu_j \\ &= E(X_i) E(X_j) - \mu_i \mu_j \\ &= 0 \end{aligned}$$

- Independent rv are uncorrelated
- Uncorrelated rv are not independent
- For example: $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 = X_1^2$, then

$$\begin{aligned} Cov(X_1 X_2) &= E(X_1 X_2) - \mu_1 \mu_2 \\ &= E(X_1^3) - 0\mu_2 \end{aligned}$$

- Correlation is a measure of linear relation between rv

Functions of Multivariate rv

- Let X and Y ve rv with joint pdf $f_{XY}(x, y)$.
- Consider 2 functions: $z = g(x, y)$ and $w = h(x, y)$.
- Find $f_{ZW}(z, w)$
- Define Jacobian of this transform

$$J = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial x} g(x, y) & \frac{\partial}{\partial y} g(x, y) \\ \\ \frac{\partial}{\partial x} h(x, y) & \frac{\partial}{\partial y} h(x, y) \end{vmatrix}$$

- Define $J(x_i, y_i) = J|_{x=x_i, y=y_i}$

Functions of Multivariate rv

- Assume (x_i, y_i) for $i = 1, 2, \cdots, n$ are the solutions of $g(.)$ and $h(.)$ such that $g(x_i, y_i) = z_i$ and $h(x_i, y_i) = w_i$.
- Then

$$f_{ZW}(z_i, w_i) = \sum_{i=1}^{n} \frac{f_{XY}(x_i, y_i)}{J(x_i, y_i)}$$

Functions of Multivariate rv – Example

- For jointly distributed rv X and Y, consider the following functions:

$$z = ax + by$$
$$w = cx + dy$$

- $J(x, y) = |ad - bc|$
- if $ad - bc \neq 0$, this system has a single solution:

$$x = Az + Bw$$
$$y = Cz + Dw$$

where A, B, C, D can be written in terms of a, b, c, d.

$$f_{zw} = \frac{1}{|ad - bc|} f_{XY}(Az + Bw, Cz + Dw)$$

Functions of Multivariate rv

$$z = ax + by$$
$$w = cx + dy$$

$$f_{zw} = \frac{1}{|ad - bc|} f_{XY}(Az + Bw, Cz + Dw)$$

- For transformations of linear systems the type of the joint distribution does not change.
- For example, if X and Y are jointly normal, their linear combinations will generate rv that are also jointly normal.

Single Function of Multivariate rv

What happens when we are interested in the output of a single function?

- Let X and Y ve rv with joint pdf $f_{XY}(x, y)$.
- Consider a single function: $z = g(x, y)$.
- Find $f_Z(z)$

Define a second dummy function such as $w = h(x, y) = x$.
Find $f_{Z,W}(z, w)$.
Find $f_Z(z)$ using marginalization ie.

$$f_Z(z) = \int_{w \in \mathcal{W}} f_{Z,W}(z, w) dw$$

Single Function of Multivariate rv

- When $W = X$ Jacobian becomes:

$$J = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial x} g(x,y) & \frac{\partial}{\partial y} g(x,y) \\ 1 & 0 \end{vmatrix} = \left| \frac{\partial}{\partial y} g(x,y) \right|$$

- When $W = Y$ Jacobian becomes:

$$J = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial x} g(x,y) & \frac{\partial}{\partial y} g(x,y) \\ 0 & 1 \end{vmatrix} = \left| \frac{\partial}{\partial x} g(x,y) \right|$$

- Choose whichever produces simpler result.

Single Function of Multivariate rv – Example

- Let X, Y have joint distribution of $f_{XY}(xy)$
- Consider $Z = X + Y$, find $f_Z(z)$
- Introduce dummy function $W = X$
- Jacobian is $|J| = 1$, then

$$f_{ZW}(z, w) = f_{XY}(w, z - w)$$

- Find marginal density of Z

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{ZW}(z, w) dw \\ &= \int_{-\infty}^{\infty} f_{XY}(w, z - w) dw \\ &= \int_{-\infty}^{\infty} f_{XY}(x, z - x) dx \end{aligned}$$

Addition of Independent RV

- From previous example

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z - x)dx$$

- If X and Y are independent rv, $f_{XY}(x, y) = f_X(x)f_Y(y)$
- Then

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{XY}(x, z - x)dx \\ &= \int_{-\infty}^{\infty} f_X(x)f_Y(z - x)dx \end{aligned}$$

- Remember signal processing course. This is convolution integral.
- **When two independent rv. are added, their pdf are convolved.**

Single Function of Multivariate rv – Example

- Let X, Y have joint distribution of $f_{XY}(xy)$
- Consider $Z = X/Y$, find $f_Z(z)$
- Introduce dummy function $W = Y$ (this is prefered to obtain a simpler jacobian, $W = X$ would also work.)
- Jacobian is $|J| = |1/Y| = |1/W|$, then

$$f_{ZW}(z, w) = f_{XY}(zw, w)|w|$$

- Find marginal density of Z

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{XY}(zw, w)|w| dw \\ &= \int_{-\infty}^{\infty} f_{XY}(zy, y)|y| dy \end{aligned}$$

Bivariate Normal Distribution

- Bivariate Normal rv are of specific interest
- Let X_1 and X_2 are jointly normal

$$f(X_1, X_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp\left\{-\frac{u_1^2 - 2\rho_{12}u_1u_2 + u_2^2}{2(1-\rho_{12}^2)}\right\}$$

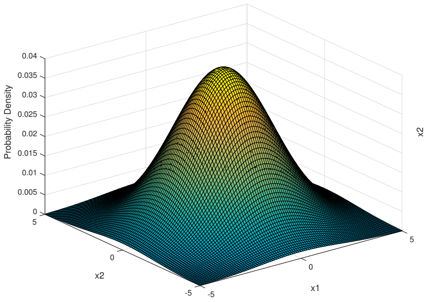
where $u_1 = \frac{x_1 - \mu_{X_1}}{\sigma_{X_1}}$, $u_2 = \frac{x_2 - \mu_{X_2}}{\sigma_{X_2}}$, and ρ_{12} is the correlation coefficient.

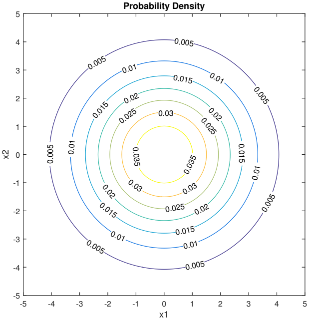
Bivariate Normal Distribution – Matlab

```
mu = [0 0];
Sigma = [4 0;0 4];

N=5;
x1 = -N:.1:N; x2 = -N:.1:N;
[X1,X2] = meshgrid(x1,x2);

F = mvnpdf([X1(:) X2(:)],mu,Sigma);
F = reshape(F,length(x2),length(x1));

[c,h]=contour(x1,x2,F);
clabel(c,h)
axis('image')
xlabel('x1'); ylabel('x2');
title('Probability Density');
```

Bivariate Normal Distribution – Matlab

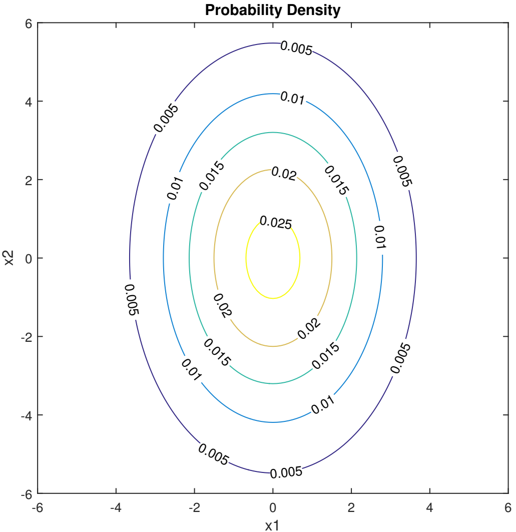
$$\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$





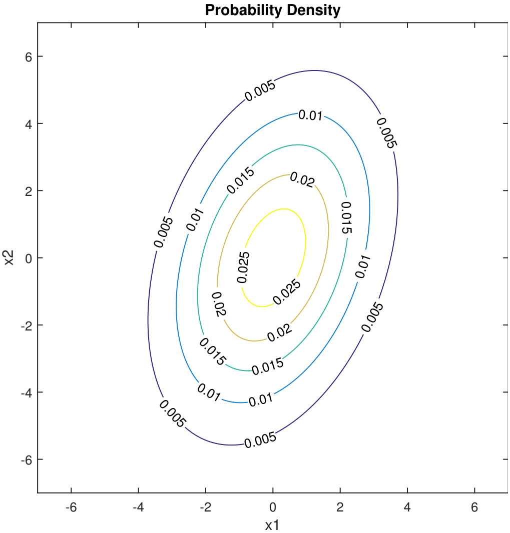
Bivariate Normal Distribution – Matlab

$$\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$$



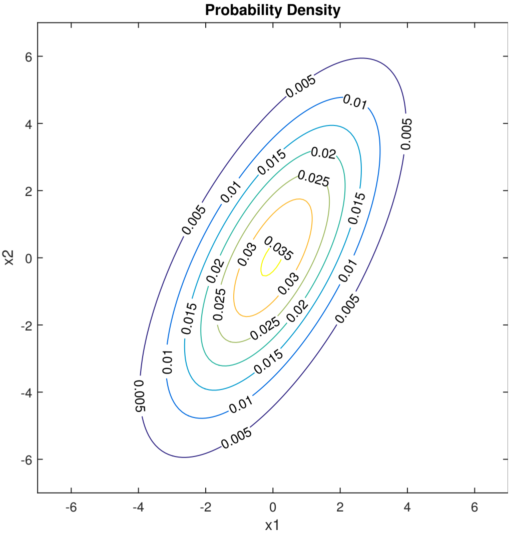
Bivariate Normal Distribution – Matlab

$$\Sigma = \begin{bmatrix} 4 & 2 \\ 2 & 9 \end{bmatrix}$$



Bivariate Normal Distribution – Matlab

$$\Sigma = \begin{bmatrix} 4 & 4 \\ 4 & 9 \end{bmatrix}$$



Correlation vs Independence for Bivariate Normal rv

- Uncorrelated bivariate Normal rv are also independent.
- This is an exception of "Uncorrelated rv are not independent"
- Why? If X_1 and X_2 are jointly normal

$$f(X_1, X_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp\left\{-\frac{u_1^2 - 2\rho u_1 u_2 + u_2^2}{2(1-\rho_{12}^2)}\right\}$$

- If X_1 and X_2 are uncorrelated, it means $\rho_{12} = 0$, then $f(X_1, X_2)$ can be factorized:

$$\begin{aligned} f(X_1, X_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{u_1^2 + u_2^2}{2}\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{u_1^2}{2}\right\} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{u_2^2}{2}\right\} \end{aligned}$$

which implies independence.

Multivariate Normal Distribution

- Let $X_1, X_2, \cdots, X_k$ are jointly normal
- Switch to vector notation for compact formulae
  $\mathbf{X} = (X_1, X_2, \cdots, X_k)$, $\mu_{\mathbf{X}} = (\mu_1, \mu_2, \cdots, \mu_k)$ and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & Cov(X_1, X_2) & \cdots & Cov(X_1, X_k) \\ Cov(X_1, X_2) & \sigma_2^2 & \cdots & Cov(X_2, X_k) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(X_1, X_k) & Cov(X_2, X_k) & \cdots & \sigma_k^2 \end{bmatrix}$$

- Then

$$f(\mathbf{X}) = \frac{1}{(2\pi)^{(k/2)}\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{X} - \mu_{\mathbf{X}})^T \Sigma^{-1} (\mathbf{X} - \mu_{\mathbf{X}})\right\}$$

Marginal Densities for Multivariate Normal Distribution

- If $X_1, X_2, \cdots, X_k$ are jointly normal, then the marginal distribution of X_i is $\mathcal{N}(\mu_i, \sigma_i^2)$ $\forall i \in \{1, 2, \cdots, k\}$
- If X_i $i \in \{1, 2, \cdots, k\}$ has marginal distribution of $\mathcal{N}(\mu_i, \sigma_i^2)$ does not imply $X_1, X_2, \cdots, X_k$ are jointly normal.
- If $X_1, X_2, \cdots, X_k$ are uncorrelated, then
  - Σ is a diagonal matrix, ie. $Cov(X_i, X_j) = 0$ if $i \neq j$
  - $X_1, X_2, \cdots, X_k$ are independent rv

$$f(\mathbf{X}) = \frac{1}{(2\pi)^{(k/2)}\sqrt{\sigma_1^2\sigma_2^2\cdots\sigma_k^2}} \exp\left\{-\sum_{i=1}^{k}\frac{(x-\mu_i)^2}{2\sigma_i^2}\right\}$$

Linear Transformations of Multivariate Normal Distribution

- Let $\mathbf{X} = (X_1, X_2, \cdots, X_k) \sim \mathcal{N}(\mu_X, \Sigma_X)$
- Consider a linear transformation

$$\mathbf{Y} = \begin{bmatrix} \mathrm{Y}_1 \\ \mathrm{Y}_2 \\ \vdots \\ \mathrm{Y}_k \end{bmatrix} = G \begin{bmatrix} \mathrm{X}_1 \\ \mathrm{X}_2 \\ \vdots \\ \mathrm{X}_k \end{bmatrix} = G\mathbf{X}$$

- Then $\mathbf{Y} \sim \mathcal{N}(G\mu, G\Sigma_X G^T)$
- Need proof?

Linear Transformations of Multivariate Normal Distribution

- Jacobian becomes $|J| = |G|$
- Pdf of $\mathbf{Y}$ is

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{|G|} f_{\mathbf{X}}\left(G^{-1}\mathbf{y}\right) \\
&= \frac{1}{(2\pi)^{(k/2)}\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(G^{-1}\mathbf{y} - \mu_{\mathbf{X}})^T \Sigma_X^{-1} (G^{-1}\mathbf{y} - \mu_{\mathbf{X}})\right\} \\
&= \frac{1}{(2\pi)^{(k/2)}\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(G^{-1}(\mathbf{y} - G\mu_{\mathbf{X}}))^T \Sigma_X^{-1} (G^{-1}(\mathbf{y} - G\mu_{\mathbf{X}}))\right\} \\
&= \frac{1}{(2\pi)^{(k/2)}\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{y} - G\mu_{\mathbf{X}})^T {G^{-1}}^T \Sigma_X^{-1} G^{-1} (\mathbf{y} - G\mu_{\mathbf{X}})\right\}
\end{aligned}
$$

- $\Sigma_Y^{-1} = {G^{-1}}^T \Sigma_X^{-1} G^{-1} \rightarrow \Sigma_Y = G \Sigma_X G^T$
- $\mu_{\mathbf{Y}} = G\mu_{\mathbf{X}}$

Decorrelation of Multivariate Normal Distribution

- We would like to find a linear transform such that $\mathbf{Y}$ has uncorrelated rvs $\rightarrow \Sigma_Y = G\Sigma_X G^T$ is a diagonal matrix
- Find eigenvalues ($\Lambda = diag\lambda_X$) and eigenvectors (V) of Σ_X such that

$$\Sigma_X = V\Lambda V^T$$

- Let $G = V^T$, then

$$\Sigma_Y = G\Sigma_X G^T = V^T(V\Lambda V^T)V = \Lambda$$

- Then rv in $\mathbf{Y}$ will be uncorrelated with eachother

Decorrelation of Multivariate Normal Distribution

- Let $\mathbf{X} = (X_1, X_2, \cdots, X_k) \sim \mathcal{N}(\mu_X, \Sigma_X)$
- Consider the following linear transformation

$$\mathbf{Y} = G\mathbf{X} - \mu_X$$

  where $G = \Lambda^{-1/2} V^T$
- Then $\mathbf{Y} \sim \mathcal{N}(0, I)$

Decorrelation of Multivariate Normal Distribution

- How to find eigenvalues and eigenvectors of Σ_X
- $\Sigma_X v = \lambda v$
- To find eigenvalues, solve $det(\Sigma_X - \lambda I) = 0$
- To find eigenvectors, solve $(\Sigma_X - \lambda I)v = 0$

Decorrelation of Multivariate Normal Distribution – Example

- Consider the previous example with

$$\Sigma_X = \begin{bmatrix} 4 & 4 \\ 4 & 9 \end{bmatrix}$$

- $det(\Sigma_X - \lambda I) = det(\begin{bmatrix} 4-\lambda & 4 \\ 4 & 9-\lambda \end{bmatrix}) = 0 \rightarrow$
  $\lambda^2 - 13\lambda + 20 = 0$
- Eigenvalues are 1.78 and 11.22
- Then, $\begin{bmatrix} 2.22 & 4 \\ 4 & 7.22 \end{bmatrix} v_1 = 0$ and $\begin{bmatrix} -7.22 & 4 \\ 4 & -2.22 \end{bmatrix} v_2 = 0$
- Eigenvectors are

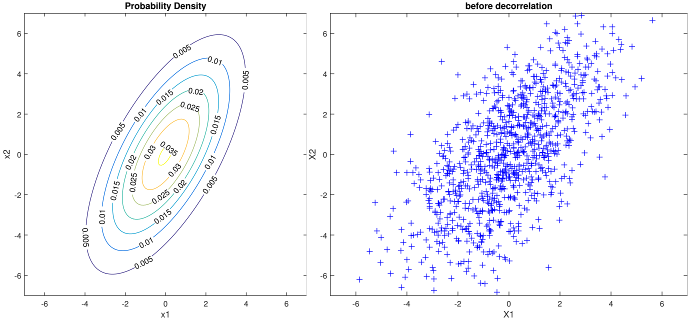
$$\begin{bmatrix} -0.8746 & 0.4848 \\ 0.4848 & 0.8746 \end{bmatrix}$$

Decorrelation of Multivariate Normal Distribution – Example

```
% before decorrelation
mu = [0 0];
Sigma = [4 4;4 9];
X=mvnrnd(mu,Sigma,1000)';
figure;plot(X(1,:),X(2,:),'b+');
axis([-7 7 -7 7]);

% after decorrelation
[V,Lambda]=eig(Sigma);
G = V';
Y=G*X;
figure;plot(Y(1,:),Y(2,:),'ro');
axis([-7 7 -7 7]);
```

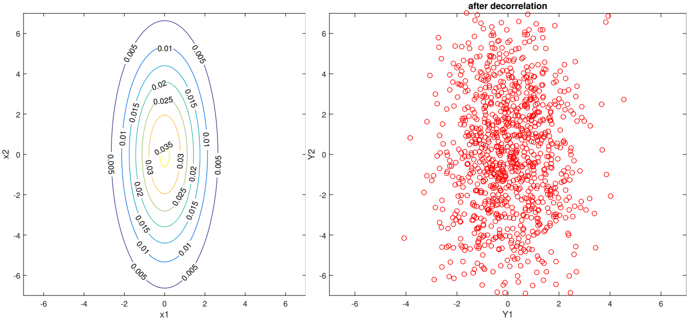
Decorrelation of Multivariate Normal Distribution – Example

Before Decorrelation



Decorrelation of Multivariate Normal Distribution – Example

After Decorrelation



Prediction of rv from another rv by Linear Transform

- Assume we would like to predict Y from X with $\hat{Y} = aX + b$
- Find best a, b such that $E((Y - \hat{Y})^2)$ is minimized.
- Take partial derivative wrt to a and b

$$\frac{\partial}{\partial a} E((Y - aX - b)^2) = 0$$
$$\frac{\partial}{\partial b} E((Y - aX - b)^2) = 0$$

Prediction of rv from another rv by Linear Transform

- Skipping derivation

$$a = \frac{Cov(X,Y)}{\sigma_X^2}$$
$$b = \mu_Y - \frac{Cov(X,Y)}{\sigma_X^2}\mu_X$$

- If $Cov(X,Y) = 0$ then $\hat{Y} = \mu_Y$. Hence X provides no information about Y.

Central Limit Theorem – CLT

- Let $X_1, X_2, \cdots, X_k$ be independent and identically distributed (**iid**) rv with expected value μ and variance σ^2
- Their summation $Z = \sum_{i=1}^{k} X_i \sim \mathcal{N}(k\mu, k\sigma^2)$ regardless of their pdf with sufficient k.
- Identical statements

$$\sum_{i=1}^{k} X_i \sim \mathcal{N}(k\mu, k\sigma^2)$$

$$\frac{1}{k}\sum_{i=1}^{k} X_i \sim \mathcal{N}(\mu, \frac{1}{k}\sigma^2)$$

$$\frac{1}{\sqrt{k}\sigma}\sum_{i=1}^{k} X_i - \mu \sim \mathcal{N}(0, 1)$$

- Need larger k for more skewed distributions.
- As a rule of thumb $k \geq 20$ is sufficient.

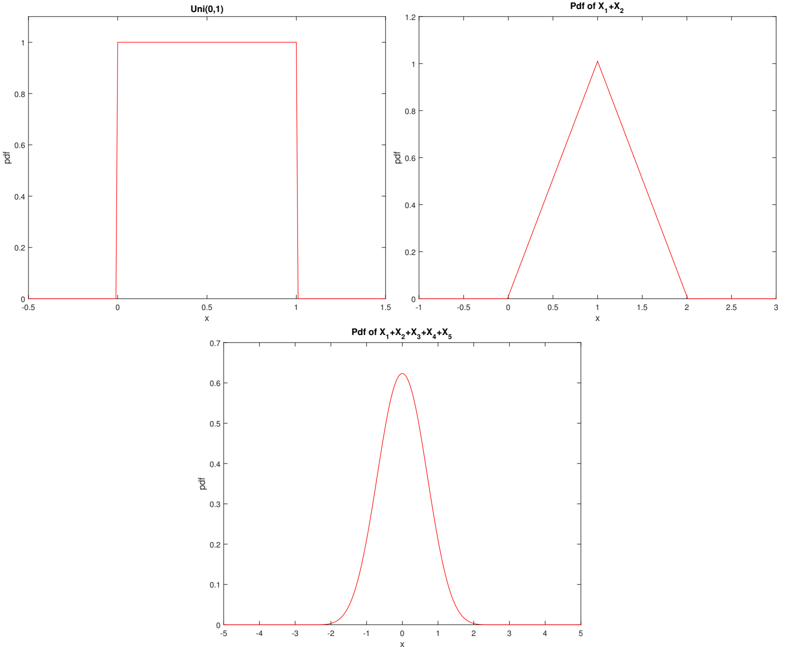
Central Limit Theorem – CLT

- Remember slide 29
- When two independent rv. are added, their pdf are convolved.
- Consider $X_i \sim Uni(0,1)$, then
  - $X_1 + X_2 \sim Uni(0,1) * Uni(0,1)$
  - $X_1, X_2, \cdots, X_k \sim Uni(0,1) * Uni(0,1) * \cdots * Uni(0,1)$

CLT – Example

```
t=-.5:.01:1.5;
f=zeros(size(t));
f(t>=0 & t<=1)=1;
plot(t,f,'r-');

% convolve once
f2 = 0.01*conv(f,f)
t = ((0:400)-100)*.01;
figure; plot(t,f2,'r-')

% convolve 3 more times
fconv = f2;
for i=1:3
    fconv = 0.01*conv(fconv,f)
end
t = ((0:5*200)-500)*.01; figure;
plot(t,fconv,'r-')
```

CLT – Example



Random Processes

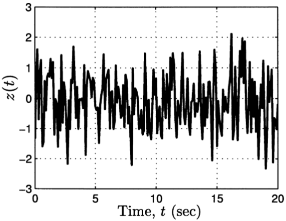
Random Processes

- A random process maps an event (or outcome of an experiment) to a function of time (continous) or a sequence (discrete).
- For example: Central Bank has a meeting on interest rates. Let $X(t)$ be the temporal behaviour of USD TRL currency rate after the decision of meeting result (new interest rate) is announced. Each outcome of the meeting will have different $X(t)$.
  Asssume, the meeting has 3 outcomes:
  - O1 Increase interest rates $\rightarrow X_1(t)$
  - O2 Do not change interest rates $\rightarrow X_2(t)$
  - O3 Decrease interest rates $\rightarrow X_3(t)$
- $X(t)$ is a random process
- $X_i(t)$ is called a realization of this random process.
- The set $\{X_1(t), X_2(t), X_3(t)\}$ form the ensemble of this random process.

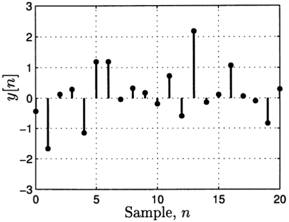
Random Processes

- In random processes there is a order of values.
- Order is typically given in time. However, it can be spatial (eg. image).
- A random process can be
  - Continous time – continous value (CTCV)
  - Discrete time – continous value (DTCV)
  - Continous time – discrete value (CTDV)
  - Discrete time – discrete value (DTDV)

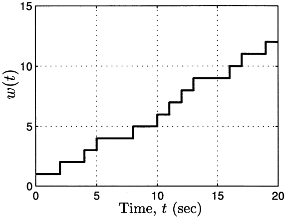
Random Processes



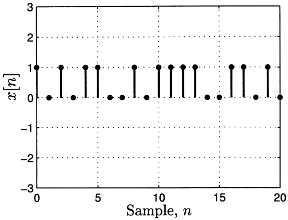
CTCV



DTCV



CTDV



DTDV

Samples are taken from Kay, S. Intutive Probability and Random Processes Using Matlab.

Statistics of Random Process

- A random process has
  - a distribution for each realization $X_i(t)$,
  - a distribution for a particular time τ between the ensemble $X_*(\tau) = \{X_1(\tau), X_2(\tau), \cdots, X_k(\tau)\}$
- Hence, the statistics of a random process can be computed
  - in time
  - in ensemble

Mean

- Mean in time is the expected value for a particular realization

$$\mu_i = E(X_i(t))$$

  Note that μ_i is not a function of time.
- Mean in ensemble is expected value of the ensemble at a particular time

$$\mu_X(t) = E([X_1(t), X_2(t), \cdots, X_k(t)])$$

  Note that $\mu_X(t)$ is a function of time.

Variance

- Variance in time for a particular realization

$$\sigma_i^2 = E((X_i(t) - \mu_i)^2)$$

  σ_i^2 is not a function of time.

- Variance in ensemble for a particular time
  Let $X_*(t) = \{X_1(t), X_2(t), \cdots, X_k(t)\}$

$$\sigma_X^2(t) = E((X_*(t) - \mu_X(t))^2)$$

  Note that $\sigma_X^2(t)$ is a function of time.

Autocorrelation and Autocovariance

- Autocorrelation is the correlation of a realization at different points in time.
- Given as a function of the two times or of the time difference.

$$
\begin{aligned}
R_{ii}(t_1, t_2) &= R_{ii}(t_1, t_1 + \tau) \\
&= E(X_i(t_1)X_i(t_2))
\end{aligned}
$$

  where $\tau = t_2 - t_1$
- Note that $R_{ii}(t_1, t_2) = R_{ii}(t_2, t_1)$
- Autocovariance is a measure of a random process co-vary at different times.

$$
\begin{aligned}
C_{ii}(t_1, t_2) &= E(\ (X_i(t_1) - \mu(t_1))(X_i(t_2) - \mu(t_2))\ ) \\
&= R_{ii}(t_1, t_2) - \mu(t_1)\mu(t_2)
\end{aligned}
$$

- Note that $C_{ii}(t_1, t_2) = C_{ii}(t_2, t_1)$
- Note that $\sigma_X^2(t_1) = C_{ii}(t_1, t_1)$

Autocorrelation Coefficient

- Autocorrelation coefficient is

$$r_{ii}(t_1, t_2) = \frac{C_{ii}(t_1, t_2)}{\sigma_X(t_1)\sigma_X(t_2)}$$

Cross-correlation and Cross-covariance

- The correlation and covariance between 2 random processes are called cross-correlation and cross-covariance
- Cross-correlation

$$R_{ij}(t_1, t_2) = E(X_i(t_1)X_j(t_2))$$

- Cross-variance

$$\begin{aligned} C_{ij}(t_1, t_2) &= E(\,(X_i(t_1) - \mu_i(t_1))(X_j(t_2) - \mu_j(t_2))\,) \\ &= R_{ij}(t_1, t_2) - \mu_i(t_1)\mu_j(t_2) \end{aligned}$$

Cross-correlation Coefficient

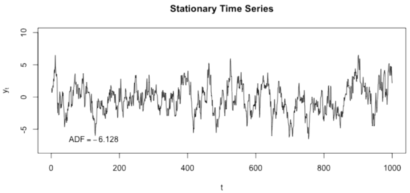
- Cross-correlation coefficient is

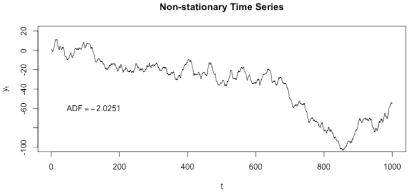
$$r_{ij}(t_1, t_2) = \frac{C_{ij}(t_1, t_2)}{\sigma_i(t_1)\sigma_j(t_2)}$$

- Two random processes are called
  - Uncorrelated if $C_{ij}(t_1, t_2) = 0$ for all $t_1 \neq t_2$.
  - Orthogonal if $R_{ij}(t_1, t_2) = 0$ for all $t_1 \neq t_2$.
  - Independent if for all t_i, x_i, x_j
    $P(X_i(t_i) \leq x_i, X_j(t_i) \leq x_j) = P(X_i(t_i) \leq x_i)P(X_j(t_i) \leq x_j)$
- Independence implies uncorrelation. Uncorrelation does not imply indepdence.

Stationarity

- Stationarity is the time invariance of a random process.
- Stationarity has two forms
  - Strict sense stationarity (SSS)
  - Wide sense stationarity (WSS)





https://en.wikipedia.org/wiki/Stationary_process

Strict Sense Stationarity (SSS)

- A random process is called SSS if

$$P(X(t_1) \leq x_1, \cdots, X(t_n) \leq x_n) = P(X(t_1+\tau) \leq x_1, \cdots, X(t_n+\tau) \leq x_n)$$

  for all values of t_i, τ, n.
- The distribution of the random process should not change in time (at τ units later) at any given time.
- This is a very strict requirement.
- If a random process satisfies this condition up to N but not $N+1$, it is called Nth order stationary process.
- First order stationarity

$$P(X(t_1) \leq x_1) = P(X(t_1+\tau) \leq x_1)$$

  for all τ values.

Wide Sense Stationarity (WSS)

- A random process is called WSS if
  - Mean is constant in time $\mu_x(t) = K$.
  - Autocorrelation depends only on time difference. $R_{ii}(t_1, t_2) = R_{ii}(t_1 - t_2)$
- SSS implies WSS.
- WSS does not imply SSS (except for Gaussian process).

Autocorrelation Properties of WSS Processes

- $R_{ii}(0) = E(X_i^2(t)) \geq 0$. This is the average power of the process.
- $R_{ii}(\tau) = R_{ii}(-\tau)$
- $|R_{ii}(\tau)| \leq R_{ii}(0)$
- If $\lim_{\tau\to\infty} R_{ii}(\tau) = c$ then $c = \mu_X^2$

Cross-correlation Properties of WSS Processes

- $R_{ij}(\tau) = R_{ij}(-\tau)$
- $|R_{ij}(\tau)| \leq \sqrt{R_{ii}(0)R_{jj}(0)}$
- $|R_{ij}(\tau)| \leq 0.5(R_{ii}(0) + R_{jj}(0))$
- $R_{ij}(\tau) = 0$ if processes are orthogonal.
- $R_{ij}(\tau) = \mu_i\mu_j$ if processes are independent.

Wiener-Khinchine Relation

Power spectral density of a random process is the Fourier transform of its autocorrelation function.

$$
\begin{aligned}
S_{ii}(f) &= \mathcal{F}(R_{ii}(\tau)) \\
&= \int_{-\infty}^{\infty} R_{ii}(\tau) e^{-j2\pi f \tau} d\tau
\end{aligned}
$$

Then

$$
\begin{aligned}
R_{ii}(\tau) &= \mathcal{F}^{-1}(S_{ii}(f)) \\
&= \int_{-\infty}^{\infty} S_{ii}(f) e^{j2\pi f \tau} df
\end{aligned}
$$

Markov Process

- A random process $X(t)$ is called first order Markov random process if for all sequences of times $t_1 < t_2 < \cdots < t_k$

$$P(X(t_k) \leq x_k | X(t_{k-1}, ..., X(t_1)) = P(X(t_k) \leq x_k | X(t_{k-1}))$$

- Conditional probability density distribution of $X(t_k)$ given all past values $X(t_{k-1} = x_{k-1}, ..., X(t_1) = x_1$ depends **only** upon the most recent value $X(t_{k-1})$
- Special cases of Markov processes
  - DTDV $\rightarrow$ Random walk
  - CTDV $\rightarrow$ Poisson process
  - CTCV $\rightarrow$ Brownian motion

Independent Increments

A random process $X(t)$ is said to have independent increments of for all times $t_1 < t_2 < \cdots < t_k$, the random variables $X(t_2) - X(t_1)$, $X(t_3) - X(t_2)$, $\cdots$ are mutually independent.

Martingale Process

A random process $X(t)$ is called Martingale process if

$$E(|X(t)|) < \infty \quad \forall t$$

and

$$E(X(t_2)|X(t_1), t_1 \leq t_2) = X(t_1)$$

Gaussian Process

A random process $X(t)$ is called a Gaussian process if all its n th order distributions $F_{X_1 X_2 \cdots X_n}$ are n-variate Gaussian distributions. If a Gaussian process is also Markov, it is called a Gaussian Markov process.

Random Walk

Let U_i for $i = 1, 2, \cdots, N$ be independent random variable with

$$P_U[k] = \begin{cases} 1 - p & k = -d \\ p & k = d \end{cases}$$

and

$$X[n] = \sum_{i=1}^{n} U_i$$

This random process $X[n]$ is called random walk.

Random Walk

If $p = 1 - p = 0.5$ then $E(U_i) = 0$ and $E(U_i^2) = d^2$.

$$P(X[n] = kd) = \left(\begin{array}{c} n \\ k \end{array} \right) (.5)^k (.5)^{n-k}$$

The autocorrelation function is

$$\begin{aligned} R_{XX}(n_1, n_2) &= E(X[n_1]X[n_2]) \\ &= E(X[n_1]\{X[n_2] - X[n1] + X[n1]\}) \\ &= E(X[n_1]^2) - E(X[n_1])E(X[n_2] - X[n_1]) \\ &= n_1 d^2 \end{aligned}$$

as $X[n_1]$ and $X[n_2]$ are independent when $n_2 > n_1$.
Hence $X[n]$ is a Markov and Martingale process.

Wiener Process

- Let $Y(t)$ be a continous time random process such that

$$Y(t) = \begin{cases} 0 & t = 0 \\ X[n] & (n-1)T \leq t \leq T \end{cases}$$

- Then $E(Y(t)) = 0$ and $E(Y^2(t)) = \frac{td^2}{T} = nd^2$
- Wiener process is obtained from $Y(t)$ by letting T and d approach to zero with $d^2 = \alpha T$ to ensure its finite and nonzero variance.

Properties of Wiener Process

- $W(t)$ is CTCV, independent increment
- $E(W(t)) = 0$ and $E(W^2(t)) = \alpha t$
- $W(t)$ will have Gaussian distribution as total displacement of position can be regarded as sum of large number of small independent increments (CLT)
- $\forall t'$ such that $0 \leq t' < t$ the increment $W(t) - W(t')$ has a Gaussian pdf with zero mean and variance of $\alpha(t - t')$.
- $R_{WW}(t_1, t_2) = \alpha \min(t_1, t_2)$
- Wiener process is Markov and Martingale.

Poisson Process

- CTDV random process
- Define $Q(t)$ as the number of events occured from 0 to t. Assume $Q(0) = 0$
  1. For t_1, t_2 and $t_2 > t_1$, $Q(t_2) - Q(t_1)$ is Poisson distributed and

     $$P(Q(t_2) - Q(t_1) = k) = \frac{(\lambda(t_2 - t_1))^k}{k!} \exp(-\lambda(t_2 - t_1))$$

  2. Number of events that occur in any interval of time is independent of number of events that occur in other nonoverlapping time intervals.

     $$P(Q(t) = k) = \frac{(\lambda t)^k}{k!} \exp(-\lambda t)$$

     Hence $E(Q(t)) = \lambda t$ and $\sigma^2 = \lambda t$.
- $R_{QQ}(t_1, t_2) = \lambda^2 t_1 t_2 + \lambda \min(t_1, t_2)$
- Poisson process is Markov but not Martingale