If a process involves "continuous functions", that process is called continuous. RP

Definitions:

Probability space: $\sigma$-algebra $\rightarrow$ random variables

(Borel field) $\rightarrow$ random processes

Probability space:

Random experiment: An experiment whose outcome is not known in advance

Outcome: An experiment result

Sample space $(\Omega)$: set of all outcomes

Event: A subset of sample space

**EX: Coin Toss:**

$n = \{ H, T \}$, $A = \{ H \}$

Finite sample space.
Ex: Countably Infinite Sample Space
exp: random integer pick
\[ n = \{-\infty, -\infty, -3, -2, -1, 0, 1, 2, 3, \ldots, \infty\} \]
\[ A = \{0, 2, 4, \ldots, \infty\} \]

Ex: Uncountably Infinite Sample Space
exp: pick a real number in \([0, 1] \)
\[ n = \{x : 0 \leq x \leq 1\} \]
\[ A = \{x : 0 \leq x \leq 1/2\} \]

Ex: Gender and height
\[ n = \{(M, h) : 150 \leq h \leq 250\} \cup \{(F, h) : 130 \leq h \leq 225\} \]
\[ A = \{(M, h) : h \geq 200\} \]
A: How to assign probabilities to Events.

Probabilities are assigned to sets and probability as a function is from sets to real number.

The assignment procedure should satisfy following:

\[
\begin{align*}
A.1. & \quad P\{\Omega\} = 1 \\
A.2. & \quad P\{\emptyset\} = 0 \\
A.3.1. & \quad \text{If disjoint events } A_1, A_2, \ldots, A_n \quad \text{then} \quad P\left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} P\{A_n\} \\
A.3.2. & \quad \text{Law of total probability} \quad \text{for } \Omega = \{H_1, HT, TH, TT\} \\
\end{align*}
\]

Komogorov's Axioms (1930)

Ex: 2-coin tossing

\[
\Omega = \{H_1, HT, TH, TT\}
\]

\[
P\{HH\} = \alpha_1 \\
P\{HT\} = \alpha_2 \\
P\{TH\} = \alpha_3 \\
P\{TT\} = \alpha_4
\]

\[
(A3) \quad P\{HHUHTVTHUUTTS\} = 1 \\
\Rightarrow P\{HH\} + P\{HT\} + P\{TH\} + P\{TT\} = 1
\]

\[
\sum_{k=1}^{4} \alpha_k = 1 \quad \alpha_k > 0
\]

Any other event

\[
A = \{HH, TT\} \\
P\{A\} = P\{HH\} + P\{TT\} = \alpha_1 + \alpha_4
\]
To do some algebra with sets, we need to have consistency in our operations. The algebra for sets is done by union and intersection operations.

We consider the set events satisfying the following conditions as a valid field:

Borel field (σ-algebra) ① \( \Omega \) is an event
(σ-algebra) ② \( A_1, A_2, \ldots \) are events
\[ \Omega \bigcup \bigcup_{k=1}^{n} A_k \] is also an event.
(σ-algebra) ③ \( A_i \) is an event, \( A_i^c \) is also an event.

\[ \begin{align*}
\Omega &= \{ HH, HT, TH, TT \} \\
F_1 &= \{ \emptyset, \Omega, HH, HT, TH, TT \} \bigcup \{ HH, HT \} \bigcup \{ HH, HT, TH \} \bigcup \{ HH, HT, TT \} \\
&\quad \bigcup \{ HH, TH \} \bigcup \{ HH, TH, TT \} \bigcup \{ HH, TT \} \bigcup \{ HH, TH, TT \} \\
&\quad \bigcup \{ HT, TH \} \bigcup \{ HT, TH, TT \} \\
&\quad \bigcup \{ HT, TT \} \bigcup \{ HT, TH, TT \} \\
&\quad \bigcup \{ TH, TT \} \\
F_2 &= \{ \emptyset, \Omega, HH, TT \} \bigcup \{ TH, HT \} \bigcup \{ HT, TH \} \\
F_3 &= \{ \emptyset, \Omega \}
\end{align*} \]

\[ \frac{1}{2^1} = \frac{1}{2} = 1 \text{ subsets} \quad \text{power set} \]
EX: \( n = \frac{3}{2} \times 0.4 \times 41 \frac{3}{4} \times 62 \)

We construct events from half-open intervals, \( [a, b] \).

\[ f_1 = \left( \frac{1}{2}, 2, \phi \right) \cap \left( a, \frac{1}{2}, \frac{1}{2} \right) \]

A field can be constructed using all half-open intervals in \( \mathbb{A} \).

Note: Event of \( \left\{ \frac{1}{2}, \left[ \frac{1}{4}, \frac{1}{2} \right] \right\} \) is also in the field constructed.

\( \sum_{k=2}^{\infty} \left( \frac{1}{2} - \frac{1}{k}, \frac{1}{2} \right) \]

Since \( A \cap B = (A \cup B)^c \)

\( \left\{ \frac{1}{4}, \frac{1}{2} \right\} \cup \left[ \frac{1}{4}, \frac{1}{2} \right] \]

Consequences of Kolmogorov's Axioms:

1. \( P[\emptyset] = 0 \)

2. \( P[\cup_{n=1}^{\infty} A_n] = \lim_{n \to \infty} P[\cap_{n=1}^{\infty} A_n] \)

3. \( P[A^c] = 1 - P[A] \)

4. \( P[A \cap B] = P[A] \cap P[B] \)

5. \( \sum_{n=1}^{\infty} P[A_n] = 1 \) - \( A_n \) disjoint

6. \( P[\cup_{n=1}^{\infty} A_n] = \lim_{n \to \infty} P[\cap_{n=1}^{\infty} A_n] \)

7. \( P[\cup_{n=1}^{\infty} A_n] = \lim_{n \to \infty} P[\cap_{n=1}^{\infty} A_n] \), \( A_1 \subset A_2 \subset A_3 \)

8. \( P[\cup_{n=1}^{\infty} A_n] = \lim_{n \to \infty} P[\cap_{n=1}^{\infty} A_n] \), \( A_1 \supset A_2 \supset A_3 \)

Sec 1.2.2

1. \( P[\emptyset] = 0 \)

Apply (A1)

\( P[\cup_{n=1}^{\infty} A_n] = \lim_{n \to \infty} P[\cap_{n=1}^{\infty} A_n] \)

\( P[\emptyset] = 0 \)

Apply (A1)

\( P[\emptyset] = 0 \)

2. \( A_1 \subset A_2 \subset A_3 \subset \emptyset \)

Apply (A2) and use \( P[\emptyset] = 0 \)

3. \( P[A^c] = 1 - P[A] \)

Apply (A1)
4. \[ A \cup B \Rightarrow p(A^2) \subset p(B^2) \]

\[ A \cup B \]

\[ p(A^2) \subset p(B^2) \]

\[ A \cup B \]

\[ p(B^2) \supset p(A^2) \]

5. \[ p(A_n) \leq 1 \] \( A_n \) disjoint

\[ \bigcup_{n=1}^{\infty} A_n = \text{from previous statement} \]

\[ p\left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} p(A_n) \]

\[ p\left( \bigcup_{n=1}^{\infty} A_n \right) \leq p\left( \bigcup_{n=1}^{\infty} A_n \right) \]

6. \[ p\left( \bigcup_{n=1}^{\infty} A_n \right) = \lim_{m \to \infty} p\left( \bigcup_{n=1}^{m} A_n \right) \]

\[ A \cup (A_2 \cap A_1^c) = (A \cup A_2) \cap (A \cup A_1^c) \]

\[ B_1 = A_1 \quad B_2 = A_2 \cap A_1^c \Rightarrow B_1 \text{ and } B_2 \text{ are disjoint} \]

\[ B_3 = A_2 \cap B_1 = A_2 \cap B_1 \]

\[ B_k \text{ are disjoint} \]

Then, \[ p\left( \bigcup_{n=1}^{\infty} A_n \right) \leq \lim_{m \to \infty} p\left( \bigcup_{n=1}^{m} B_n \right) \]

\[ = \lim_{m \to \infty} \sum_{n=1}^{m} p(B_n) \]

\[ = \lim_{m \to \infty} p\left( \bigcup_{n=1}^{m} B_n \right) \]

\[ = \lim_{m \to \infty} p\left( \bigcup_{n=1}^{\infty} A_n \right) \]
7. Apply (6) \[ \]  

8. \( A_1 \rightarrow A_1' \rightarrow A_2 \rightarrow A_2' \) then apply (6) \[ \]

Union Bound:\[ P \left( \bigcup_{k=1}^N A_k \right) \leq \sum_{k=1}^N P(A_k) \]

- \( A_k \) : Event \( k \)

\[ P(A_1 \cup A_2) = P(A_1 \cup \overline{(A_2 \cap A_1')}) \]

- Since \( (A_2 \cap A_1') \subset A_2 \)

\[ P(A_2 \cap A_1') \leq P(A_2) \]

Using this fact:\[ \left( \bigcup_{k=2}^N A_k \right) \leq p(A_1) + p \left( \bigcup_{k=2}^N A_k \right) \leq p(A_1) + p(A_2) + p \left( \bigcup_{k=3}^N A_k \right) \]

Conditional Probability: \[ P(A|B) \] event A and B occur at the same time, \[ \frac{p(A \cap B)}{p(B)} \]

event A, event B

\[ p = \left\{ \begin{array}{ll} \text{A occurs given B occurs} & \text{B occurs} \\ \text{A does not occur given B occurs} & \text{B does not occur} \end{array} \right. \]

\[ B: \text{reduced sample space after conditioning} \]

\[ A|B: \text{sample space after conditioning} \]
Boyle's Rule:

\[ P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \quad \text{since} \quad P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A) \]

\[ \text{from def:} \]

\[ P(A \cap B) = P(B \cap A) \]

Note: \( A \cap B \) event is related to \( B \cap A \) event by Boyle's Rule.

Total Probability Theorem: \( B_i \) disjoint sets covering sample space (partition):

1. \( B_i \cap B_j = \emptyset \quad i \neq j \)
2. \( \bigcup_{i=1}^{n} B_i = \Omega \)

\[ P(A) = \sum_{i=1}^{n} P(A \cap B_i) \cdot P(B_i) \]

\[ \frac{P(A \cap B_i)}{P(A \cap B_i)} \]

Independence if \( P(A \cap B) = P(A) \cdot P(B) \), \( A \) and \( B \) are independent.

Also, equivalent to: \( P(A \mid B) = P(A) \)

\[ P(B \mid A) = P(B) \]

If \( A_1, A_2, A_3, \ldots \) are independent:

1. If they are pairwise independent, that is

\[ P(A_k \cap A_l) = P(A_k) \cdot P(A_l) \quad k \neq l \quad \forall k, l \]

2. They should be independent in triplets, that is

\[ P(A_k \cap A_l \cap A_m) = P(A_k) \cdot P(A_l) \cdot P(A_m) \quad k \neq l \neq m \]

3. Independent in quartets (and so on)
Conditional Independence:

\[ p(A \cap B \mid C) = p(A \mid C) p(B \mid C) \]

Borel-Cantelli Lemma (Papoulis, 4th Edition)

1. \( A_1, A_2, A_3, \ldots \) are a sequence of events

\[ p_k = p(A_k) \]

\[ \sum_{k=1}^{\infty} p_k < \infty \implies \sum_{k=1}^{\infty} I(A_k) < \infty \text{ with prob. } 1 \]

\[ I(A_k) = \begin{cases} 1 & \text{if } A_k \text{ occurs} \\ 0 & \text{if } A_k \text{ does not occur} \end{cases} \]

Indicator function

Proof: \( B = \bigcap_{k=1}^{\infty} A_k \)

Then, \( B = \bigcap_{n=1}^{\infty} B_n \) if \( B \) occurs \( \iff \) \( B_n \) occurs for \( n = 1, 2, 3, \ldots \)

Since, if \( \exists \in B \implies \exists \in B_n \) \( \forall n \) and then \( \exists \in \bigcap_{n=1}^{\infty} B_n \)

on event occurring infinitely many times

\[ X \in \bigcap_{n=1}^{\infty} B_n \implies X \text{ belongs to infinite number of } A_k \text{ events} \]
\[ P(B) = P \bigg( \bigcap_{n=1}^{\infty} A_{B_n} \bigg) \]

\[ B_1 \Rightarrow B_2 \Rightarrow B_3 \Rightarrow \cdots \quad (*) \]

\[ = P \bigg( \lim_{n \to \infty} B_n \bigg) \]

\[ = P \bigg\{ \lim_{n \to \infty} B_n \bigg\} \quad \text{using (*)} \]

\[ = \lim_{n \to \infty} P\{B_n\} \]

\[ P\{B_n\} = \sum_{k=n}^{\infty} P(A_k) \to 0 \quad \text{union bound} \]

Since claim is:

\[ \sum_{k=1}^{\infty} P(A_k) < \infty \]

\[ P \bigg( \bigcap_{n=1}^{\infty} B_n \bigg) = 0 \]

\[ P \bigg( B^c \bigg) = 1 \to \text{Eventually many } A_k \text{ occurs.} \]

**B.C. Lemma 2:**

If \( A_1, A_2, \ldots \) are independent events

and if \( \sum_{k=1}^{\infty} P(A_k) \) diverges \( \Rightarrow \) infinitely many \( A_k \) events occur.

**Proof:** Papoulis 4th ed

\[ \text{EX: } P(\text{success}) = p \quad \text{You repeat success/fail trials infinitely many times. Each trial is independent} \]

\[ P(\text{fail}) = 1 - p \]

\[ \text{Q: Can there be infinite number } n \text{-successes in a row?} \]

\[ S \not\subseteq SS \not\subseteq \ldots \]

10, 11, 4, 8, 6, 9, 7, 4

\[ \text{not, not, not} \]
\[ P(\text{Fe} = n - \text{successes}) = p_1^k \times \frac{1}{2}, 2, 3, \ldots \] and each frame n-success probability is independent from others.

Then by 3.C lemma 2:

\[ \sum_{k=1}^{\infty} p(F_k = \text{success}) \rightarrow \infty \]

So, there are infinitely many n-successes in a row.

Random Variables:

\[ \Omega \] is a mapping from \( \Omega \) to \( \mathbb{R} \), \( \mathbb{R} \) is real number (real line).

\[ X = X(w) \]

sample space \( \Omega \) \( \rightarrow \) real line \( \mathbb{R} \)

Required Condition For Valid R.V:

\[ \{ w : X(w) \leq x \} \] should be a valid event.

\[ \{ w : x(w) \leq x \} \]

\[ \{ w : x(w) \leq x^+ \} \]

Ex:

\[ \Omega = \{ a_1, a_2, a_3, a_4 \} \]

\[ F = \{ \emptyset, \Omega, \{ a_1, a_2 \}, \{ a_3, a_4 \} \} \]

\[ X = x(w) \]

\[ w = a_1 \]

\[ x(w) = \]

\[ F \]

Q: Is \( X \) a valid random variable?

A:

\[ \{ w : X(w) < x^+ \} = \emptyset \in F \]

\[ \{ w : X(w) \leq x^+ \} \notin F \] not a valid R.V.
CDF: Cumulative Distribution Function

\[ F_X(x) \triangleq \Pr \{ \omega : X(\omega) \leq x \} \]

Properties:

1. \( \lim_{x \to -\infty} F_X(x) = 0 \) \( \lim_{x \to +\infty} F_X(x) = 1 \)

2. \( F_X(x) \) is non-decreasing.

3. \( F_X(x) = F_X(x^+) \), i.e., CDF is cont. from right.

PDF: Probability Density Function

\[ f_X(x) \triangleq \frac{d}{dx} F_X(x) \]

\[ F_X(x) \triangleq \int_{-\infty}^{x} f_X(x) \, dx \]

Example:

If \( X \) and \( Y \) are independent r.v.'s:

Let \( Z = X + Y \) \( f_Z(z) = ? \)
Solution:

\[ F_2(z) = \frac{d}{dz} F_1(z) \]

\[ F_2(z) = \mathbb{P}(X + Y \leq z) = \mathbb{P}(X \leq z - Y) \]

\[ = \int_{-\infty}^{\infty} \mathbb{P}(X \leq z - Y) f_Y(y) \, dy \]

\[ = \int_{-\infty}^{\infty} F_X(z - y) f_Y(y) \, dy \]

\[ = f_X(z) \ast f_Y(z) \]

Remember:

A set

Then:

\[ \mathbb{P}(A \cup B) = \sum_{k=1}^{2} \mathbb{P}(A \cap B_k) \]

since \( X \) and \( Y \) are independent.

\[ \mathbb{P}(X + Y \leq z) = \mathbb{P}(X \leq z - Y) \]

\[ = \int_{-\infty}^{\infty} F_X(z - y) f_Y(y) \, dy \]

Total prob.
Ex: \( X_1, X_2, \ldots, X_N \) independent \( \Gamma \) distributed with pdf \( f_{X_k}(x) \). 

a) \( Z = \max(X_1, X_2, \ldots, X_N) \). \( f_Z(z) \) and \( f_X(z) \).

b) \( P(Z = X_1^2) = ? \)

Solution:

\( a) \ E_Z(\bar{z}) = \bar{z} \text{ s.t. } P(\bar{z} \leq \bar{z}) \) \( A \subset B \) \( \{ A = B \} \)

\( P(Z < \bar{z}) \)

\( = P(\bar{z} \leq X_1, X_2, \ldots, X_N) \)

\( = P(X_1 < \bar{z}, \ldots, X_N < \bar{z}) \)

\( = \prod_{i=1}^{N} P(X_i < \bar{z}) \)

\( = [E_X(\bar{z})]^{-N} \)

\( f_Z(\bar{z}) = \frac{1}{\bar{z}} f_{X}(\bar{z}) = N [E_X(\bar{z})]^{-N} \)

\( b) P(\bar{z} = X_1^2) = P(\bar{z} = X_1^2) \) \( \text{ independent} \)

\( = \prod_{i=1}^{N} P(X_i < \bar{z}) \)

\( = \prod_{i=1}^{N} \int_{0}^{\bar{z}} f_{X_i}(x_i) \) \( dx_i \)

\( = N \int_{0}^{\bar{z}} \left( \frac{1}{\bar{z}} \right)^{N-1} \) \( dx_i \)

\( = N \bar{z}^{-N} \int_{0}^{\bar{z}} \) \( dx_i \)

\( = N \bar{z}^{-N} \) \( \left[ x \right]_{0}^{\bar{z}} \)

\( = N \bar{z}^{-N} \) \( \bar{z} \)

\( = N \bar{z}^{-N} \) \( \bar{z} \)

\( = N \bar{z}^{-N} \) \( \bar{z} \)
Expectation:

\[ E(x^2) = \int_{-\infty}^{\infty} x^2 f(x) \, dx \]

**Note:** If \( x > 0 \), that is \( f(x) = 0 \) for \( x < 0 \), then

\[ E(x^2) = \int_{-\infty}^{\infty} F^c(x) \, dx \]

**Complementary cdf:** \( F^c(x) = P(X > x) = 1 - F(x) \)

**Proof (1):**

\[ \int_{-\infty}^{\infty} x^2 f(x) \, dx = \int_{-\infty}^{\infty} x^2 \left( \int_{0}^{x} f(x) \, dx \right) \, dx = \int_{0}^{\infty} \left( \int_{0}^{x} f(x) \, dx \right) \, dx \]

**Proof (2):** For discrete r.v.

\[ F^c(x) = P(X > x) \]

\[ E(x^2) = \sum (x^2 p(x)) \]
Note 2: If \( X \) takes both \( + \)ve and \( - \)ve values,
then,
\[
E[X^2] = -\int_{-\infty}^{0} F_X(x)dx + \int_{0}^{\infty} F_X(x)dx
\]
See the picture in Fig. 1.4 from book and convince yourself.

Note 3: The original def. for \( E[X^3] \) is still very valuable;
the "new" relation for \( E[X^3] \) is also useful in some cases.

Note 4: \( E[g(X)] = \int g(x)f_X(x)dx \)
function of a single r.v.

Note 5: \( E[X^3] = 0 \rightarrow \text{mean} \)
\( \text{var}(x) = E[(x-\bar{x})^2] = E[X^2] - (\bar{x})^2 \rightarrow \text{variance} \)

When we have 2 or more r.v.'s,
\[
E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy
\]
\[
\text{cov}(X,Y) = E[(X-\bar{X})(Y-\bar{Y})] = E[XY] - \bar{X}\bar{Y}
\]
covariance of \( X \) and \( Y \)
Note that \( \text{var}(X) = \text{cov}(X,X) \)
Definition:
\[ \text{Cov}(X,Y) = 0 \implies X \text{ and } Y \text{ are uncorrelated r.v.'s} \]

Note 6: If \( X \) and \( Y \) are independent
\[
\begin{align*}
\int \left[ g(x)h(y) + xg(x)y(x,y) \right] dx dy &= \left\{ \int g(x)h(x) dx \right\} \left\{ \int h(y) dy \right\} \\
\text{independent} &\implies \left\{ \int g(x)h(x) dx \right\} \left\{ \int h(y) dy \right\} \\
\text{EF}(x) \cdot \text{EF}(y) &= \text{EF}(x) \cdot \text{EF}(y) \\
\text{EF}(x) \cdot \text{EF}(y) &= \text{EF}(x) \cdot \text{EF}(y)
\end{align*}
\]

Note again that if \( X, Y \) are independent \( \implies X, Y \) are also uncorrelated.

Uncorrelatedness \( \implies \text{Cov}(X,Y) = 0 \implies \text{EF}\{(X-\text{EF}(X))(Y-\text{EF}(Y))\} = 0 \)

\[
\begin{align*}
\text{by} &\implies \text{EF}\{(X-\text{EF}(X))\text{EF}\{(Y-\text{EF}(Y))\} = 0 \\
\text{EF}(X) = 0 &\implies \text{EF}(Y) = 0
\end{align*}
\]

Note: The converse is not true in general.

\( X, Y \) uncorrelated

\( X, Y \) independent

Note 7: \( \text{EF}\{X_1 + X_2 + \ldots + X_N\} = \text{EF}\{X_1\} + \text{EF}\{X_2\} + \ldots + \text{EF}\{X_N\} \)

This relation is valid when \( X_k \)'s are independent or dependent, correlated or uncorrelated.

Note 8: \( I_A \) \{ \text{event A occurs} \}
\( I_A ^{\text{c}} \) \{ \text{event A does not occur} \} etc.

Indicator function.
\[ E \text{In } \mathbb{Y} = 1 \cdot P(\text{A happening}) + 0 \cdot P(\text{A does not happening}) = P(\text{A occurring}) \]

Note 9: Herded Expectation

\[ E_{X,Y} \mathbb{g}(x,y) = \int Y \{ E_{Y \mid X} \mathbb{g}(x,y) \mid y \} f_X(x) dy \]

\[ = \int (\int \mathbb{g}(x,y) f_{XY}(x,y) dx) f_Y(y) dy \]

\[ = \int \mathbb{g}(x,y) f_{XY}(x,y) f_Y(y) dy \]

Note 8

(Ex) In data communications, some bit patterns are reserved for signaling, for example, 011111 can denote end-of-transmission and once receivers decodes this sequence of bits, it stops listening.

If transmitter send 0110011111100 bits to the receiver payload, then

\[ 01100111111100 \]

Additional 0 inserted to avoid false termination in payload.

If you sending \( n \) bits and let \( P(\text{"1"}) = p \),

then what is the expected number of stuffed bits?
Ak: Event of bit stuffing after bk

\[ I_{Ak} = \begin{cases} 1 & \text{Ak occurs} \\ 0 & \text{other} \end{cases} \]

\[ \# \text{bits} \overset{\text{stuffed}}{\rightarrow} \sum_{k=1}^{n} I_{Ak} = \sum_{k=1}^{n} \text{stuffed} \]

\[ p(A_k \text{ happening}) = \sum_{k=5}^{\infty} (1-p) + \frac{1}{p} \]

\[ p(A_1) = 0 \]
\[ p(A_2) = 0 \]
\[ p(A_3) = 0 \]
\[ p(A_4) = 0 \]
\[ p(A_5) = (1-p) q^4 + 0 \]
\[ p(A_6) = (1-p) q^4 \]
\[ p(A_n) = (1-p) q^4 \]

Ex: Note 9.

2 hours to exit

MINZ-

5 hours trip and you come back to original point

7 hours trip and come back to original point

Weirs always select one of these with equal prob. of 1/3.
A1. What is expected time for mine to go out.

A2. Y be his first choice


\[ E[3X^2] = 3 \cdot \frac{1}{3} + (E[3X^2] + 5) \cdot \frac{1}{3} + (E[3X^2] + 7) \cdot \frac{1}{3} \]

\[ E[3X^2] = 15 \]

E[X] = \text{uniform dist. over region A}

b) \( f_{x|y}(x|y) = ? \)

c) \( f_{y|x}(y|x) \)

e) Are X and Y ind? \( \text{cov}(X, Y) = ? \)
a) \( f_{X,Y}(x,y) = \mathbb{1}_{\mathbb{R}^2} \) for \( 0 < x < 1, 0 < y < 1 \) and \( 0 \) otherwise.

\[
\int_0^\infty 2dx = 2(1-x) \quad \text{for } 0 < x < 1
\]

\[
f_Y(y) = \int f_{X,Y}(x,y) \, dx = \begin{cases} \frac{1-y}{2} & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}
\]

b) \( f_X(x) = \int f_{X,Y}(x,y) \, dy = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}
\)

f_Y(y) = \int f_{X,Y}(x,y) \, dx = \begin{cases} \frac{2}{3}(1-y) & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}

\[
c) \quad f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{2}{3}(1-y) & 0 < x < 1, \quad 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}
\]

\[
d) \quad E[X|Y = y] = \frac{1-y}{2}
\]

\[
E[Y|X = x] = \frac{1-x}{2}
\]

\[
E[X] = E\{E[X|Y]\} = E\left\{\frac{1-Y}{2}\right\} = \frac{1}{2} - \frac{1}{2} = \frac{1}{2}
\]

\[
e) \quad f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \text{not possible}
\]

\[
\text{they are not independent}
\]

\[
\text{cov}(X,Y) = E[X \cdot Y] - E[X] \cdot E[Y] = E[X \cdot Y] - \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{12} - \frac{1}{4} = -\frac{1}{12}
\]

\[
E[X^2] = \iint f_{X,Y}(x,y) \, dx \, dy = \int_0^1 \int_0^1 x^2 \, dx \, dy = \int_0^1 \left[ \frac{x^3}{3} \right]_0^1 \, dy = \frac{1}{12}
\]
Moment Generating Functions:

\[ \phi_X(r) = \mathbb{E}(e^{rX}) = \int_{-\infty}^{\infty} e^{rx} f_X(x) \, dx, \quad r \in \mathbb{R} \]

\[ g_X(r) = \sum_{k=0}^{\infty} e^{rk} \mathbb{P}(X = k) \]

\[ \mathcal{L}\text{-transform} \]

\[ \phi_X(r) = e^{\sum k \mathbb{P}(k)} \]

\[ \phi_X(r) = e^{\sum k \mathbb{P}(X = k)} \]

\[ \mathbb{P}(X = k) \]

\[ k \rightarrow \mathbb{P}(X = k) \]

\[ \mathcal{L}\text{-transform} \]

\[ \phi_X(r) = \sum_{n=0}^{\infty} \frac{e^{nk}}{n!} \]

\[ \sum_{n=0}^{\infty} \frac{e^{nk}}{n!} \]

\[ e^{\sum_{n=0}^{\infty} \frac{e^{nk}}{n!}} \]

Remember:

1. \( g_X(0) = 1 \rightarrow r = 0 \in \mathbb{R} \)

2. \( E[X^r] = \frac{d^r}{dx^r} g_X(x) \bigg|_{x=0} = g_X^{(r)}(0) \rightarrow \) assuming that an interval around \( r = 0 \) is in \( \mathbb{R} \)!

\[ \mathbb{E}[X^r] = \left. \frac{d}{dx} e^{rx} \right|_{x=0} = e^{0} \mathbb{E}[X^r] \rightarrow \mathbb{E} \frac{e^{rx}}{r^r} \]

Ex:

\( f_X(x) = \lambda e^{-\lambda x} u(x) \), exponential dist.

\[ f_X(x) \]

\[ \phi_X(r) = \left. \frac{\lambda e^{-\lambda x}}{s + \lambda} \right|_{s=-r} = \frac{\lambda}{\lambda - r} \]
\[ E\{X\} = \frac{\lambda}{(\lambda-r)^2} = \frac{1}{\lambda} \]

\[ \text{Var}\{X\} = E\{(X-\mu)^2\} = \frac{2}{\lambda^2} \]

Then, exp. distribution with parameter \( \lambda \) has

\[ E\{X\} = \frac{1}{\lambda} \]

\[ \text{Var}\{X\} = E\{X^2\} - (E\{X\})^2 = \frac{2}{\lambda^2} \]

Probability Inequalities: non-negative rv (e.g. exp. dist.)

1. Markov Inequality \( \Rightarrow \forall \gamma \geq 0 \): \( P\{Y \geq \gamma\} \leq \frac{E[Y]}{\gamma} \)

2. Chebyshev Inequality \( \Rightarrow P\{|X-\mu| \geq \sigma\} \leq \frac{1}{\sigma^2} \text{Var}[X] \)

3. One-sided Chebyshev Inequality \( \Rightarrow P\{X \geq a\} \leq \frac{E[X] - E[X]}{a} \)

4. Chernoff Bound \( \Rightarrow \text{for any } b > 0 \)

5. Markov Inequality \( \Rightarrow P\{Y \geq a\} \leq \frac{E[Y]}{a} \)

\[ E[Y] = \int_{-\infty}^{\infty} y F_Y(dy) \geq \int_{a}^{\infty} y F_Y(dy) \geq P\{Y \geq a\} \cdot a \]

\[ E[Y^2] = \int_{-\infty}^{\infty} y^2 F_Y(dy) \geq \int_{a}^{\infty} y^2 F_Y(dy) \geq P\{Y^2 \geq a\} \cdot a^2 \]

\[ P\{Y \geq a\} \leq \frac{E[Y]}{a} \]

\[ F_X(\infty) = 0 \]

\[ F_X(a) = \frac{1}{\lambda} \]

\[ F_Y(a) = 0 \]

\[ F_Y(a) = a \]

\[ F_Y(a) = 0 \]
2. Chebyshev's Inequality

Let \( z = (Y - \mu_y)^2 \) and apply Markov's Inequality.

\[
P_z(z > a^2) \leq \frac{E_z z}{a^2} = \frac{\text{Var}(Y)}{a^2}
\]

\[
P_z((Y - \mu_y)^2 > a^2) \leq \frac{E_z (Y - \mu_y)^2}{a^2} = \frac{\text{Var}(Y)}{a^2}
\]

\[
P_z((Y - \mu_y)^2 > a^2) = \frac{\text{Var}(Y)}{a^2}
\]

\[
\{Y: (Y - \mu_y)^2 > a^2\} \Rightarrow \{Y: \text{Var}(Y) > a^2\}
\]

\[
P_z(\text{Var}(Y) > a^2) \leq \frac{\text{Var}(Y)}{a^2}
\]

3. One-Sided Chebyshev Inequality

\[
Y - \mu_y \geq a \Leftrightarrow (Y - \mu_y) + b > a + b \Rightarrow ((Y - \mu_y) + b)^2 > (a + b)^2
\]

\[
P_z(Y - \mu_y > a) = P_z((Y - \mu_y) + b > a + b)
\]

\[
\leq P_z((Y - \mu_y) + b)^2 > (a + b)^2
\]

\[
\text{Markov: } \frac{E_z ((Y - \mu_y) + b)^2}{(a + b)^2} \leq \frac{\text{Var}(Y) + b^2 + 2E_z (Y - \mu_y) b}{(a + b)^2}
\]

\[
P_z(Y - \mu_y > a) \leq \frac{\text{Var}(Y) + b^2}{(a + b)^2} \text{ for any } b > 0
\]
I can take derivative of \( \frac{a^2 + b^2}{(a+b)^2} \) and select \( b \) such that \( \frac{a^2 + b^2}{(a+b)^2} \) is minimized.

\[ \Rightarrow \text{If I do that } b_k = \frac{\sigma^2}{a} \Rightarrow \rho_{x,y} > a^2 \frac{\sigma^2}{\sigma^2 + a^2} \]

Chernoff Bound:

\[ P_{\{Y > a^2 \} \leq e^{-a^2} \sigma^2(r) \quad r > 0, \; r \in \Re^+ \}

Moment generating function of \( Y \):

\[ P_{\{Y > a^2 \} \leq e^{-a^2} \sigma^2(r) \quad r > 0, \; r \in \Re^+ \}

Proof:

\[ Y > a^2 \Rightarrow e^{Y} > e^{a^2} \]

\[ e^{Y} \sim \text{Exp}(\lambda) \]

\[ P_{\{Y > a^2 \} \leq e^{-a^2} \sigma^2(r) \quad r > 0, \; r \in \Re^+ \}

Note: This is valid for \( a > 0 \).

Example: Exponential distribution \( f_Y(y) = \lambda e^{-\lambda y} \quad y \geq 0 \)

\[ \mu_Y = \frac{1}{\lambda} \]

\[ \sigma^2_Y = \frac{1}{\lambda^2} \]

\[ \sigma^2_Y(r) = \frac{1}{\lambda} \cdot \frac{1}{1+r} \]

\[ P_{\{Y > \mu_Y \} \leq e^{-a^2} \sigma^2(r) \quad r > 0, \; r \in \Re^+ \}

Exact Result:

\[ P_{\{Y > \mu_Y \} \leq e^{-a^2} \sigma^2(r) \quad r > 0, \; r \in \Re^+ \}

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]

\[ = e^{-\lambda (1+1/r)} \]
1. Markov Inequality
\[ P[X > k\mu_Y] \leq \frac{E[Y]}{k\mu_Y} \]

2. Chebyshev Inequality
\[ P[|Y - \mu_Y| > (k-1)\mu_Y] \leq \frac{1}{(k-1)^2} \left( \frac{\sigma^2}{(k-1)\mu_Y^2} \right) \]
\[ \equiv (Y > k\mu_Y) \]
\[ \text{or} \]
\[ Y < (2-k)\mu_Y \]

3. One Sided Chebyshev Inequality
\[ P[Y - \mu_Y > (k-1)\mu_Y] \leq \frac{1}{1 + (k-1)^2} \left( \frac{\sigma^2}{\sigma^2 + (k-1)\mu_Y^2} \right) \]

4. Chernoff
\[ P[Y > k\mu_Y] \leq e^{-\frac{k\mu_Y^2}{g_Y(r)}} \cdot \text{minimize RHS wrt } r \]
\[ L = e^{-\frac{k\mu_Y^2}{g_Y(r)}} \cdot \text{RHS} \]
\[ \delta \left( \frac{\sigma^2}{2}, \frac{\lambda}{r} \right) = 0 \text{ (4)} \]
\[(4) \text{ is satisfied at } r = \lambda \left( 1 - \frac{1}{k} \right) \]
by substituting \( r \) in RHS
\[ P[Y > k\mu_Y] \leq e^{-\frac{k\mu_Y^2}{g_Y(r)}} \cdot \text{Chernoff} \]

STOCHASTIC CONVERGENCE:
\( X_1, X_2, \ldots \) a sequence of R.V.
By: Does \( X_k \) as \( k \to \infty \) converge in some sense to a R.V.?
Remember: 1. Convergence in real numbers:
\[ \lim_{n \to \infty} a_n = A \quad \text{for any } \varepsilon, \text{ there exists } n \in N(\varepsilon) \text{ such that } |a_n - A| < \varepsilon, \quad n > N(\varepsilon) \]

Ex: \[ \lim_{n \to \infty} \frac{1}{n} = 0, \text{ given } \varepsilon, \quad \frac{1}{n} < \varepsilon \quad \Rightarrow \quad n > \frac{1}{\varepsilon} \\
N(\varepsilon) = \left\lceil \frac{1}{\varepsilon} \right\rceil \text{ (ceiling function)} \]

2. Convergence of Functions

1) Pointwise Convergence

If \( f_n(x) \to f(x) \),

then for a given \( x \), the sequence of numbers \( f_n(x) \to f(x) \),

that is, the convergence of real numbers for a fixed \( x \).

2) Uniform Convergence

\[ f_k(x) \to f(x), \text{ given } \varepsilon, \text{ there exists } n \in N(\varepsilon) \text{ such that} \]

\[ \sup \{|f_k(x) - f(x)| < \varepsilon \text{ for } x \in [x, x+\varepsilon] \} \]

Ex: \[ f_k(x) = e^{-kx}, \quad x > 0 \]

Claim: \[ \frac{1}{e^k} \to 0 \text{ as } k \to \infty \]

Choose \( k > k_{\text{upper bound}} = \ln(\frac{1}{\varepsilon}) \) for \( x > 1 \)

and \( \frac{1}{e^k} < \varepsilon \text{ is } k > \ln(\frac{1}{\varepsilon}) \)
So, by choosing $k \in \mathbb{N}$ ($k \geq \frac{1}{\varepsilon}$)
we satisfy uniform convergence condition for a given $\varepsilon$.

**STOCHASTIC CONVERGENCE**

1. **Convergence in Distribution**
   \[
   \mathbb{F}_{X_k}(x) \rightarrow \mathbb{F}_Z(x)
   \]

   $X_k$, $Z$ are r.v.'s
   
   $Z$ is the limiting r.v.
   \[
   X_k \rightarrow Z \text{ in distribution}
   \]

   if $\mathbb{F}_{X_k}(x) \rightarrow \mathbb{F}_Z(x)$ at each $x$ for which $\mathbb{F}_Z(x)$ is continuous.

*Ex.: Central Limit Theorem*

Let $X_1, X_2, \ldots$ i.i.d. with finite mean $\mu$ and variance $\sigma^2$.

Then,
\[
\mathbb{F}_{S_n}(x) = \phi \left( \frac{x - n\mu}{\sigma \sqrt{n}} \right)
\]

\[
\lim_{n \to \infty} \mathbb{F}_{S_n}(x) = \phi \left( \frac{x}{\sigma} \right)
\]

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

**NOTE:**
\[
\bar{X} \equiv \frac{S_n}{n} = \frac{\sum X_i}{n}
\]

$E(\bar{X}) = 0$, since $E[X_i] = \mu$.

$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$, since $\text{Var}(X_i) = \sigma^2$.

and $\text{Var}(\bar{X})^2 = \frac{\text{Var}(S_n)^2}{n^2} = \frac{\sigma^2}{n}$.

*Ex.*

$X_1, X_2, \ldots$ i.i.d.

\[
X_k \sim N(\mu, \sigma^2)
\]

\[
\frac{S_n}{\sigma \sqrt{n}} \rightarrow N(0, 1)
\]

\[
\begin{array}{c|c|c}
\hline
i & x_i & \mathbb{F}(x_i) \\
\hline
1 & 1/2 & 1/4 \\
2 & 1/2 & 1/2 \\
\hline
\end{array}
\]

\[
S_n = X_1 + X_2 + \cdots + X_n
\]

$S_1 = X_1$
\[ S_2 = x_1 + x_2 \]

Remember \( f_{S_2}(s_2) = f_X(x_1) \cdot f_X(x_2) \)

\[ S_3 = x_1 + x_2 + x_3 \]

(2) Convergence in Probability

A sequence of r.v's converges to \( \xi \) in probability if

\[ \lim_{n \to \infty} P \left( |X_n - \xi| > \varepsilon \right) = 0 \quad \text{for any } \varepsilon > 0 \]

Notes:

1. Let's call

\[ \hat{X}_n = X_n - \xi \]

Another r.v

We can show that if \( \hat{X}_n \to 0 \) in probability then \( X_n \to \xi \) in probability.

2. Call \[ a_k = P \left( \left| \hat{X}_k \right| > \varepsilon \right) \]

Convergence in probability is \( \lim_{k \to \infty} a_k = 0 \) for every \( \varepsilon \).
STOCHASTIC CONVERGENCE

1. Conv. in Distribution: EX. Central Limit Theorem

2. Conv. in Probability

$Z_1, Z_2, Z_3, \ldots$ a sequence of r.v.

$Z_k \xrightarrow{k \to \infty} Z$

if converges in what sense?

Convergence in Probability:

$$\lim_{n \to \infty} P(|Z_n - \mu| > \varepsilon) = 0 \quad \forall \varepsilon > 0 \quad \text{in probability}$$

Assume $Z_n$'s are independent.

Let's show $Z_n \xrightarrow{p} 0$

$$P(|Z_n - 0| > \varepsilon) \leq \frac{1}{\varepsilon^2}$$

$P(|Z_n - 0| > \varepsilon) \leq \frac{1}{\varepsilon^2}$
\[ \lim_{n \to \infty} P \left( \frac{\mid X_n - 0 \mid}{\sqrt{n}} > \epsilon \right) = 0 \]

converges in probability

2.1, 2.2, 2.3, 2.4

Sample: 1, 0.5, 1.5, 1.0, 1.5, 3.0, 1.5

Path: \( \frac{1}{\sqrt{n}} \text{prob.} \)

2n \rightarrow m.s.

\[ \frac{1}{\sqrt{n}} \text{prob.} \]

(3) Convergence in Mean Square:

If \( \lim_{n \to \infty} E \left( (Z_n - 2)^2 \right) = 0 \), then \( Z_n \xrightarrow{m.s.} 2 \) in mean-square.

Ex: Let's apply earlier example to check conv. in M.S.

\[ E \left( \left( Z_n - 0 \right)^2 \right) = \frac{1}{n} P \left( Z_n = \frac{1}{n} \right) = \frac{1}{n} P \left( Z_n = 0 \right) \]

\[ \xrightarrow{0} \text{so}\ Z_n \xrightarrow{m.s.} 0 \text{ in mean-square sense also} \]

Ex: Let's modify earlier example as:

\[ Z_n = \begin{cases} \frac{1}{\sqrt{n}} \text{ with prob. } \frac{1}{n} \\ 0 \text{ with prob. } 1 - \frac{1}{n} \end{cases} \]
\[
E \left( (Z_n - c) \right) = \frac{1}{\sqrt{n}} \sum_p^2 \left( Z_n - n \right) + \frac{1}{n^2} \sum_{\mathbb{R}}^2 Z_n = o^2
\]

\[
= 0
\]

\[
\lim_{n \to \infty} E \left( (Z_n - c) \right) ^2 = 0
\]

Claim: If \( Z_n \to c \) in mean square sense,
then \( Z_n \to c \) in probability.

Since:
\[
P \left( |Z_n - c| > \epsilon \right) \leq \frac{E \left( (Z_n - c)^2 \right)}{\epsilon^2}
\]

Chebychev inequality.

So, taking limit as \( n \to \infty \)

\[
\lim_{n \to \infty} P \left( |Z_n - c| > \epsilon \right) \leq \lim_{n \to \infty} \frac{E \left( (Z_n - c)^2 \right)}{\epsilon^2} = 0
\]
Relationships Between Convergence Modes

Convergence in mean-square

Conv. in probability

Conv. in distribution

A random experiment is said to converge with probability if

\[ P\{ \lim_{n \to \infty} X_n(w) = 2(w) \} = 1 \]

for a fixed \( w \) this is an ordinary limit operation definition of \( n \to \infty \)

Ex: \( n = [0,1] \)

Sample space

For a fixed \( w \) this is an ordinary limit operation definition of \( n \to \infty \)

\[ X \text{ is two-valued r.v.} \]

\[ X = \{0, 1\} \]
\[ P_{\frac{1}{n}}(\tau(w)) = 0 \quad \tau = 1 - \frac{1}{n} \]

\[ P_{\frac{1}{n}}(\tau(w)) = 1 \quad \tau = 1/n \]

Q: Do I have \( 2^n \to 0 \) with prob. 1?

\[ P_{\frac{1}{n}}(\tau(w)) = 0 \]

Yes, apart from \( w = 0 \), all outcomes become exactly 0 for sufficiently large \( n \).

Almost-sure convergence (cont'd.)

\[ 2^n \to 2 \quad \text{as} \quad n \to \infty \]

\[ \lim_{n \to \infty} 2^{n}(w) = 2(w)^n = 1 \]

ordinary limit for a fixed \( w \)

**Example:**

- \( 2^{n+1}(w) = \frac{1}{2} \cdot 2^n(w) \)
- \( 2^{n+2}(w) = \frac{1}{2} \cdot 2^{n+1}(w) \)
- \( 2^{n+3}(w) = \frac{1}{2} \cdot 2^{n+2}(w) \)

\( \lambda = \{0,1\} \)

\( 2^{n}(w) \) is defined as

1. \( 2^{n}(w) = 1 \quad \text{if} \ w = 1 \)
2. \( 2^{n}(w) \) is a "rectangle" function of length \( \frac{1}{2^n} \) and starting at \( \left( \frac{1}{2} \cdot \frac{1}{2^n} \right) \mod \frac{1}{2} \)
3. \( 2^{n}(w) \) may do a circular shift

\( (5.25 \mod 1 = 0.25) \)
\[ P \{ \exists (w) = 1 \} = \frac{1}{2} \]
\[ P \{ \exists 2(w) = 0 \} = \frac{1}{4} \]
\[ P \{ \exists (w) = 1, \exists 2(w) = 0 \} = \frac{1}{4} \] \( \text{we} \{ \frac{1}{2}, \frac{1}{8} \} \)
\[ P \{ \exists 3 = 1 \} = \frac{1}{3} \]
\[ P \{ \exists 2 = 0, \exists 3 = 1 \} = \frac{2}{3} \]
\[ P \{ \exists 2n = 1 \} = \frac{1}{n} \]
\[ \Theta : 2^n \rightarrow 0 \]
\[ \Omega = \{ 0.99, 2(0.99), 3(0.99), \ldots \} \]

Since the series diverges, the rectangular function gets thinner but it rotates an infinite number of times in \([0,1]\) interval \( \Rightarrow S_n \)

\(2n\) does not converge to 0 with prob 1.

Convergence with prob 1 is difficult to check and it requires going back to the definition of rv. There are some sufficient conditions which are simpler to use and if satisfied guarantees a.s. convergence.

\( \sum \frac{1}{2n} \rightarrow 0 \) a.s. (Boyle-Cantelli lemma)

(2) \(2n\)'s r.v.'s with finite expectation

\[ \sum_{n=1}^{\infty} 12n \rightarrow 0 \quad \rightarrow 2n \quad a.s. \] (Textbook Lemma 5.2.1)
EX: $\sum_{n=1}^{\infty} 1/n$ with prob. $1/n^2$

$\sum_{n=1}^{\infty} 1/n$ a.s.

Apply St. $P\left(1/e - 1 \leq \sum_{n}^{\infty} 1/n \leq 1/e \right) = \frac{1}{e} \Rightarrow \frac{1}{n^2} \Rightarrow P(\varepsilon) \leq 1$

\[
\sum_{n=1}^{\infty} P\left(1/e - 1 \leq \sum_{n}^{\infty} 1/n \leq 1/e \right) = 1 \\
= \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \approx 1.6
\]

\[
\Rightarrow \sum_{n}^{\infty} 1/n \to 0
\]

Weak Law of Large Numbers:

Let $S_n = X_1 + X_2 + \cdots + X_n$ where $X_k$'s are i.i.d. with finite variance $\sigma^2$, then

\[
\lim_{n \to \infty} P\left(\frac{|S_n - \mu|}{\sigma} > \varepsilon \right) = 0 \quad \forall \varepsilon
\]

\[
\Rightarrow \frac{S_n}{n} \xrightarrow{P} \mu \Rightarrow \frac{1}{n} \sum_{i=1}^{n} X_i = \mu
\]

Proof: Remember mean square convergence guarantees convergence in prob.

Do I have $L^2$ convergence?

\[
\lim_{n \to \infty} E\left(\frac{(S_n - \bar{X})^2}{\sigma^2} \right) = 0
\]

\[
\frac{S_n}{n} = \frac{S_n}{n} - \frac{n\bar{X}}{n} + \frac{n\bar{X}}{n}
\]

\[
= \frac{1}{n} \left( (X_1 - \bar{X}) + (X_2 - \bar{X}) + \cdots + (X_n - \bar{X}) \right)
\]
\[
\begin{align*}
E \left\{ \left( \frac{S_n - x}{\sigma} \right)^2 \right\} = & \, E \left\{ \left( x_1 - x \right)^2 + \left( x_2 - x \right)^2 + \cdots + \left( x_n - x \right)^2 \right\} / n^2 \\
= & \, E \left\{ \left( x_1 - x \right)^2 + \left( x_2 - x \right)^2 + \cdots + \left( x_n - x \right)^2 \right\} / n^2 \\
= & \, \sigma x_2^2 + \sigma x_2^2 + \cdots + \sigma x_2^2 + 0 + 0 + \cdots + \sigma x_2^2 = \sigma x_2^2
\end{align*}
\]

\begin{itemize}
\item \[\lim_{n \to \infty} E \left\{ \left( \frac{S_n - x}{\sigma} \right)^2 \right\} = 0 \implies \text{ conclusion:} \]
\item \[
\frac{\sigma x_2^2}{n^2} \to 0 \implies \text{ We have } \mu S \text{ conv. to } x \text{ and therefore } \text{ conv. in prob. to } x.
\end{itemize}

Comments:
1. \[S_n \text{ is nothing but sample mean } \left( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \right) \]

and \[\bar{x} \overset{d}{=} \frac{1}{n} \sum_{i=1}^{n} x_i \]

is the basis of

computer experiments or monte-carlo trials conducted to 'calculate' \( \bar{x} \).

\[\bar{x} = \int_{-\infty}^{\infty} x f_{\bar{x}}(x) \, dx \]

2. \[X \overset{a}{=} 1 \text{ if } A \text{ happens, } 0 \text{ otherwise } \]

\[\implies \text{ relative frequency of event } A \]

\[E[X] = P[A \text{ happening}] \]

\[= P[A \text{ not happening}] \]

\[= \text{ relative frequency interpretation of probability} \]

3. Weak law of large numbers can be extended to dependent \( x_i \)'s and \( x_i \)'s having infinite variance.

4. There is a strong form for law of large numbers, for \( x_i \) i.i.d

and \[E \{ X \} < \infty, \]

\[S_n \overset{a.s.}{\to} \bar{x} \] *strong* law of large numbers.
Poisson Processes

- A member of "arrival" processes (such as arrival times of customers to a shop.)

\[ N(t) \]: # of customers arrived until time \( t \).

\[ X_t \]: arrival times of customers.

\[ S_t \]: arrival time / epoch

\[ X_t = S_t - S_{t-1} \]

**Note:** \( S_n = \sum_{k=1}^{n} X_k \)

Hence,

\[ S_1, S_2, \ldots, S_n \leftrightarrow X_1, X_2, \ldots, X_n \]

Arrival Times

Interarrival Times

So, knowing either of one of them gives the other one. So joint pdf of either one is should be sufficient to find joint pdf of other one.

\[ N(t) \]: Counting r.v., \( N(t) \) starts from 0 at \( t=0 \) and incremented by 1 at different "t" values.

\[ N(t) \]: # arrivals until time \( t \) and including time "t".

\[ \{ S_n \geq t \} = \{ N(t) \geq n \} \]

\[ \text{\( n \)th customer arrives before \( t \) \} \]

\[ \text{has arrived before \( t \) have \( n \) or more customers} \]

Complement \[ \{ S_n > t \} = \{ \text{no} \} \neq \{ \text{N(t) > n} \} \]
Renewal Process: An arrival process with i.i.d. interarrival times (x's) are called renewal process.

Poisson Process: A renewal process with exponential PDF, i.e.: 
\[ f_X(x) = \lambda e^{-\lambda x} \]  \((\lambda): rate \ of \ the \ process\)

Properties of Poisson Processes

1. Memoryless: A r.v. is memoryless if:
\[ P \{ X > t + x \} = P \{ X > x \} \]  \( \text{or} \)
\[ P \{ X > t + x \} = P \{ X > t \} P \{ X > x \} \]

Exponential distribution is memoryless, since:
\[ P \{ X > t + x \} = e^{-\lambda x} \]
(\exp(\lambda))

Note that only exp. r.v. satisfies memoryless property.
Poisson Process:

Remember, Poisson process is a renewal process with exponential interarrival distribution. (Exp. dist. is memoryless, i.e. 
\[ P(\bar{X} > t + z | \bar{X} > t) = P(\bar{X} > z) \])

Theorem 2.2.5: For a Poisson process at any time, \( t \), the first arrival after \( t \) (waiting time) is independent of \( N(t) \) and all arrival epochs before \( t \). (It's also independent of r.v.'s \( X(t_1), X(t_2), \ldots \), \( t_1 + t_2 < t \))

Proof: Case

(a) \( N(t) = 0 \)

\[ P(\bar{X} > t + z | N(t) = 0) = P(\bar{X} > t + z) \]

\[ = P(\bar{X}_1 > t + z) = e^{-\lambda t} \]

\[ = P(\bar{X}_1 > t + z | \bar{X} > t) \cdot P(\bar{X} > t) \]

(b) \( N(t) = 1 \)

\[ P(\bar{X} > t + z | N(t) = 1) = P(\bar{X} > t + z) \]

\[ = P(\bar{X}_1 > t + z) = e^{-\lambda t} \]
\[ (*) \quad \Pr \{ \tau > t \} = \Pr \left\{ \tau > t \mid \tau \leq t \right\} = \Pr \left\{ \tau > t \mid \tau = \tau^* \right\} \]

\[ \Pr \{ \tau > t \} = \Pr \{ \tau > t \mid \tau = \tau^* \} \]

\[ = \Pr \{ \tau > t \mid \tau = \tau^* \} \]

\[ = \Pr \{ \tau > t \mid \tau = \tau^* \} \]

\[ = \Pr \{ \tau > t \mid \tau = \tau^* \} = e^{-\lambda t} \]

**Note:** The same argument \((*)\) holds when conditioning is not only on \( \tau \), but also on \( \tau^* \).

The \( \tau^* \) information is equivalent to \( \tau(t) \) for \( t > 0 \).

\[ \Pr \{ \tau > t \} = \Pr \{ \tau > t \mid \tau = \tau^* \} = e^{-\lambda t} \]

So, additional waiting time \( \tau \) is independent of \( N(t) \) for \( \tau > 0 \).

**Definition:** Stationary increment.

A counting process is called stationary increment if

\[ N(t') - N(t) = N(t' - t) \quad \text{for all} \quad 0 < t < t' \]

Poisson process is stationary increments since

\[ \# \text{ arrivals in } t' - t = N(t') - N(t) \]

Is independent of \( t' \) and \( t \), but depends only on \( t' - t \) (waiting period).
Definition: Independent Increments

\{ N(t) : t \geq 0 \} is independent increments if for every \( k \)

\[ N(t_1), N(t_2), \ldots, N(t_k) \]

are independent from each other.

Conclusion: Poisson process is stationary and independent increments: \( \text{IP} \) process.

Probability Density of \( n \)th arrival:

\[ S_n = X_1 + X_2 + \ldots + X_n \]

\( X_i \)'s i.i.d. and exp-distr.: \( f_s(t) \)

\[ f_{S_n}(t) = \frac{n!}{t^{n-1}} \frac{e^{-t}}{(n-1)!} \]

\( n \)th arrival time

Notes: \( f_{X_1|S_2}(x_1|s_2) = f_{S_2}(s_2) f_{X_1|S_2}(s_2|x_1) \)

\[ = \lambda e^{-x_1} \lambda e^{-s_2} \lambda (s_2 - x_1) \]

\[ = \lambda e^{-s_2} \lambda s_2 \]

\( 0 < x_1 \leq s_2 \)}
\[ f_{s_2}(s_2) = \int_{-\infty}^{s_2} f_{x_1, x_2}(x_1, s_2) \, dx_1 = \int_0^{s_2} \lambda^2 e^{-\lambda x_1} \, dx_1 = \frac{\lambda^2}{\lambda} s_2 \cdot \lambda e^{-\lambda s_2} \text{ for } s_2 > 0. \]

\[ s_2 = x_1 + x_2 \]

\[ s_1 = s + x_2 \text{ given } f_{x_2}(x_2) \]

Find \( f_{s_2}(s_2) \)

\[ f_{s_2}(s_2) = f_{x_2}(s_2 - s) \]

Review:
- 1 func of 1 r.v.
- 1 func of 2 r.v. \( p^2 = x^2 + y^2 \)
- 2 func of 2 r.v. \( u^2 = x^2 + y^2 \)
- \( \theta = \tan^{-1}(y/x) \)
Poisson Process:

\[ X_t \sim \text{i.i.d. exp. distributed with parameter } \lambda \]

\[ f_{X_t}(x_t) = \lambda e^{-\lambda x_t}, \quad x_t > 0 \]

\[ P_{X_t > z | N(t) = 3} = \frac{e^{-\lambda z}}{3!} \]

Properties:
1. iid Increments

\[ \tilde{N}(t_1, t_2) \text{ and } \tilde{N}(t_3, t_4) \to \text{equivalent to } \tilde{N}(0, t_4 - t_3) \]

- # arrivals
  - in \((t_1, t_2]\)
  - in \((t_3, t_4]\)

2. Stationary Increments

\[ f_{S_n}(t) = \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} \]

\[ S_n = X_1 + X_2 + \cdots + X_n \]

- Erlang distribution (Chi-square with even degrees of freedom)

Probability Mass Function of \(N(t)\):

Th. 2.2.10: For a Poisson process with rate \(\lambda\), \(N(t)\) (number of arrivals in \([0, t]\))

Textbook gives by Poisson r.v.

\[ P_{N(t) = n} = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \]
Proof: We know \( f_{S_M(t)}(x) = \frac{e^{-x}}{(n-1)!} \)

I will calculate \( P\{t < S_{M(t)} + \delta \} \) in two different ways.

1. \( P\{t < S_{M(t)} + \delta \} = \int_{t}^{t+\delta} f_{S_{M(t)}}(x) dx = e^{-t} + o(\delta) \) as is small quantity.

Remember!

* \( o(\delta) \) represents functions of \( \delta \)

\[ \lim_{\delta \to 0} \frac{f(t+\delta) - f(t)}{\delta} = \frac{f(t+\delta) - f(t)}{(t+\delta) - t} \]

**Ex.** \( \delta^2 \) is \( o(\delta) \)

**Ex.** \( \sin(\delta) \) is not \( o(\delta) \)

2. \( P\{t < S_{M(1)} + \delta \} = e^{-t} + o(\delta) \) and \( \{ \text{1 arrival in } (t, t+\delta) \} \)

\[ = P\{t < S_{M(1)} + \delta \} \]

\[ \text{and } \{ \text{1 arrival in } (t, t+\delta) \} \]

\[ = P\{t < S_{M(t)} + \delta \} \]

\[ \text{and } \{ \text{1 arrival in } (t, t+\delta) \} \]

\[ = P\{n(t) = 0, \text{1 arrival in } (t, t+\delta) \} \]

\[ = P\{n(t) = n-1, \text{2 arrivals in } (t, t+\delta) \} \]

\[ = P\{n(t) = n \} \]

\[ \text{from independent property} \]

\[ \text{and } \{ \text{2 arrivals in } (0, \delta) \} \]

\[ = P\{n(t) = n-1, \text{2 arrivals in } (t, t+\delta) \} \]

\[ = P\{n(t) = n \} \]

\[ \text{from independent property} \]

\[ f_{S_M}(x) = \frac{e^{-x}}{(n-1)!} \]

\[ f_{S_M}(x) = \lambda e^{-\lambda t} - \lambda (1 - \lambda t + \frac{(\lambda t)^2}{2}) \]

\[ = \lambda e^{-\lambda t} \]

\[ = \lambda e^{-\lambda t} \]

\[ = \lambda e^{-\lambda t} \]
\( \text{Poisson R.V. Properties:} \)

\[ \text{PMF: } P_N(N = n) = \lambda^n e^{-\lambda} \]

\( n = 0, 1, 2, \ldots \)

mean: \( E[N] = \lambda \)

variance: \( \text{var}(N) = \lambda \)

M.G.F.: \( M_N(t) = e^{\lambda(e^t - 1)} \)
Notes:

1. For poisson process with rate \( \lambda \) (1 unit is arrivals/sec).

   then \( \Delta t \) has the unit of arrivals.

   then: \[ P(N(\Delta t) = n) = (\lambda \Delta t)^n e^{-\lambda \Delta t} \]

   \[ E[N(\Delta t)] = \lambda \Delta t \] rate per unit of poisson process (arrival/sec)

   (rate discussion, next lecture)

   When two independent Poisson r.v. are added, the resultant r.v. is also

   Poisson with rate \( \lambda_1 + \lambda_2 \)

   \[ \text{see fig. } \Rightarrow g_\lambda (r) = e^{-\lambda} \lambda^r r! \]

   \( \Delta t \) arrivals/sec \( \Rightarrow \lambda t \equiv \# \text{ arrivals} \)

   \[ E[N(t)] = \lambda t \]

   \[ \text{var}[N(t)] = \lambda t \]

\[ N = N_1 + N_2 \]

Poisson has rate \( \lambda_1 + \lambda_2 \)

Poisson r.v. is in some ways "discrete analog" of Gaussian r.v.,

that is under mild condition several counting processes (not necessarily

Poisson) when summed approach to a Poisson process (similar to

CLT for Gaussian r.v.)
Def #1

Renewal Process
with
Exp interarrival times

\[ N(t) \text{; Poisson } \text{ w/ rate } \lambda t \]

\[ \text{ex 2.4} \]

\[ \text{ex 2.7} \]

Def #2

\[ P \{ N(t, t+\delta) = 0 \} \]

\[ = 1 - e^{-\lambda \delta + o(\delta)} \]

\[ P \{ N(t, t+\delta) = 1 \} \]

\[ = \lambda \delta + o(\delta) \]

\[ P \{ N(t, t+\delta) = 2 \} = o(\delta) \]

\[ N(t) \text{ is independent and stationary increments.} \]

Def #3

N_{1}(t) \text{ and } N_{2}(t) \text{ are independent Poisson processes.}

\[
\begin{align*}
\text{Two counting processes are said if } \forall N > 0 \\
0 < t_{1} < t_{2} < t_{N} \\
N(t_{1}), N(t_{2}), \ldots, N(t_{N}) & \text{ are independent} \\
N_{2}(t_{1}), N_{2}(t_{2}), \ldots, N_{2}(t_{N}) & \text{ are independent} \\
\text{then } N(t) = N_{1}(t) + N_{2}(t) \text{ is Poisson with rate } \lambda_{1} + \lambda_{2}
\end{align*}
\]

\[ \text{Combined process} \]

\[ N(t) \]

\[ N_{1}(t) \rightarrow N_{2}(t) \rightarrow N(t) \]

Combined process
\[
\begin{align*}
\mathbb{P} \left[ N(t, t + \delta) = 0 \right] & = e^{-\lambda \delta + o(\delta)} \\
\mathbb{P} \left[ N_1(t, t + \delta) = 0, N_2(t, t + \delta) = 0 \right] & = e^{-\lambda_1 \delta + o(\delta)} \\
& \quad \times \left( 1 - \lambda_2 \delta + o(\delta) \right)
\end{align*}
\]
Since $\mathcal{N}_1(t)$ and $\mathcal{N}_2(t)$ are Poisson dist CV and i.i.d.
their sum is also Poisson dist CV.

Let $X$ be the interarrival time of combined process.

$p \mathbb{P}(X > x) = p \mathbb{P}(\mathcal{N}_1(t+x) = 0, \mathcal{N}_2(t+x) = 0) = 0 \forall x \geq 0$

Waiting time for next arrival for the combined process is $\mathbb{E}[X] = \frac{1}{\lambda} + \frac{1}{\mu}$

So, waiting time is exp. distributed for combined process.

\[ N(t) \]
\[ \text{rate: } \lambda \beta \]
\[ N_1(t) \]
\[ \text{rate: } (\lambda - \beta) \]
\[ N_2(t) \]

$p$: prob. of switching input to $\mathcal{N}_1(t)$

$(1-p)$; prob. of $\text{prob. Head}$ and $\text{coin toss experiment}$ operates the switch

$\mathcal{N}_1(t)$ and $\mathcal{N}_2(t)$ are Poisson Process with rate $\lambda = \lambda \beta$ and $\lambda_2 = (1-p)\lambda$.

Furthermore, $\mathcal{N}_1(t)$ and $\mathcal{N}_2(t)$ are independent.
Let's show $N(t+\varepsilon)$ is Poisson

$$P_{\varepsilon} N(t,t+\varepsilon) = 1 \Rightarrow P_{\varepsilon} N(t,t+\varepsilon) = 1, \text{ switched to } 0 \varepsilon +$$

$$P_{\varepsilon} N(t,t+\varepsilon) > 2, \text{ one of them switched to } 1 \varepsilon$$

$$= \left[p + o(\varepsilon)\right] \cdot p + o(\varepsilon)$$

$$= (\lambda \varepsilon) p + o(\varepsilon)$$

$$P_{\varepsilon} N(t,t+\varepsilon) = 0 \varepsilon +$$

$$P_{\varepsilon} N(t,t+\varepsilon) > 2, \text{ all of them switched to } 2 \varepsilon$$

$$= \left[p + o(\varepsilon)\right] + o(\varepsilon) + (1 - \lambda \varepsilon + o(\varepsilon))$$

$$= 1 - (\lambda \varepsilon) p + o(\varepsilon) \quad \lambda \varepsilon$$

$$P_{\varepsilon} N(t,t+\varepsilon) > 2 \Rightarrow o(\varepsilon) \Rightarrow \text{ then } N(t) \text{ is Poisson with rate } \lambda_{1} = \lambda p$$

(similarly for $N_{2}(t)$)

Proof for $N(t)$ and $N_{2}(t)$ are independent:

$$P_{\varepsilon} N(t) = m, N_{2}(t) = k \mid N(t) = m+k \Rightarrow \binom{m+k}{m} p^{m} (1-p)^{k}$$
\[ P^2 N_1(t+1) = m \text{, } N_1(t+1) = k \text{, } N_1(t) = m + k \]
\[ = P^2 N_1(t) = m \text{, } N_1(t) = k \text{, } N_1(t) = m + k \frac{x^k}{(m+k)!} \left( -\frac{x^k}{m!} \right) - \frac{x^k}{(m+k)!} \]
\[ = \frac{(\lambda t)^m}{m!} e^{\lambda t} \frac{1}{(1-p)^m} \left( -\frac{x^k}{(m+k)!} \right) - \frac{x^k}{(m+k)!} \]
\[ = P^2 [N_1(t+1) = m \text{, } N_1(t) = k \text{, } \text{ independent!}] \]

\[ N_1(t) \text{ is also independent of } N_2(t+1) \text{, } t \leq t' \]
\[ (0, t] = (0, t'] \cup (t', t] \]
and from proof \( X_1(t+1) \) is ind. \( N_2(t+1) \).

\( N_1(t+1, t] \) is ind. from \( N_2(t+1) \), since \( N_2 \) is a Poisson process with ind. increments.

You are at car wash. There are two lines generating "clean cars" with rate \( d_1 \) and \( d_2 \).
The processes are Poisson and independent.
You join line 1. There are many cars in both lines.
Let \( S_i \) be the departure time of \( i \)-th car from line 1.
\[ P \{ X_1 \leq X_1^0 \} = \int P \{ X_1 \leq x \} \, dx = \frac{1}{\lambda_1 + \lambda_2} \]

You drive at the post office. Two clerks are busy and no other clients waiting. Clerks operate at rate \( \lambda_1 \) customers/hour and \( \lambda_2 \) customers/hour. Processes are Poisson and independent.

Find the expected amount of time that you spent in post office until your task is completed.
Also see Ross P. Sos for other solutions.

\[ E_x(t) = \frac{1}{\lambda} e^{-\lambda t} \]

\[ P_x(1) = e^{-\lambda t} \]

\[ \frac{1}{\lambda} + \frac{1}{\mu} + \frac{1}{\alpha} = 1 \]

\[ \alpha = \frac{1}{\lambda} + \frac{1}{\mu} \]

\[ E[X] = E[X_1] + E[X_2] \]

\[ E[X_1] = \frac{1}{\lambda} \]

\[ E[X_2] = \frac{1}{\mu} \]

\[ E[X] = \frac{1}{\lambda} + \frac{1}{\mu} \]

\[ T = W + P \]

\[ \text{total waiting time for request} \]

\[ \text{processing time} \]

\[ \text{arrival rate} \]

\[ \text{Service rate} \]

\[ \text{Poisson arrival rate} \]

\[ \text{Poisson service rate} \]

\[ \text{waiting time} \]

\[ \text{time to be available} \]
\[ N(t) = a_1 \quad \text{if} \quad \sum_{i=1}^{a_1} 1(X_i = n) = n \]

Let \( Y_1, Y_2, \ldots, Y_N \) be a rv's (iid) dist by distribution \( f(y) \).

\[ Y_{(1)} = \min \{ Y_1, \ldots, Y_N \} \]

\[ Y_{(2)} = \text{second minimum} \{ Y_1, \ldots, Y_N \} = \min \{ Y_2, Y_1, \ldots, Y_N \} \]

\[ Y_{(3)} = \text{3rd smallest in the list} \]

\[ Y_{(n)} = \text{max} \{ Y_1, \ldots, Y_N \} \]

\[ f_{Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}}(y_{(1)}, y_{(2)}, \ldots, y_{(n)}) = \]

Joint density of ordered rv's

Let's remember

\[ f_{Y_1, Y_2, \ldots, Y_N}(y_1, y_2, \ldots, y_N) = \prod_{k=1}^{n} f(y_k) \quad \text{(rvs are independent with density } f(y)) \]

not ordered.
\[
\mathbf{Y} = \begin{bmatrix}
  2 \\
  -2 \\
  3
\end{bmatrix} \quad \rightarrow \quad \mathbf{Y}(1) = \begin{bmatrix}
  -2 \\
  4 \\
  3
\end{bmatrix}
\]

ordered realization

\[
\mathbf{Y} = \begin{bmatrix}
  3 \\
  2 \\
 -2
\end{bmatrix}
\]

\[
\mathbf{Y} = \begin{bmatrix}
  3 \\
 -2 \\
  1
\end{bmatrix}
\]

all orderings of realization gives the same ordered realization

\[
f_{Y(y_1, y_2, y_3)}(y_1, y_2, y_3) = \]

\[
P \left\{ y_1 \leq y_1 \leq y_3 + \delta \right\} + P \left\{ y_2 \leq y_2 \leq y_2 + \delta \right\} + P \left\{ y_3 \leq y_3 \leq y_3 + \delta \right\} =
\]

\[
\begin{bmatrix}
  1 & 2 & 3 \\
  2 & 1 & 3 \\
  3 & 2 & 1 \\
  1 & 3 & 2 \\
  2 & 3 & 1 \\
  3 & 1 & 2
\end{bmatrix}
\]

\[
= f_Y(y_1) f_Y(y_2) f_Y(y_3) \delta + f_2 \delta^2 + \delta^3
\]

\[
= \left( 3! \prod f_Y(y_k) \right) \delta + f_2 \delta^2 + \delta^3
\]
\[ P \{ A_k = n \} = \frac{n}{n!} \prod_{k=1}^{n} f(y_k) f_{y_k} f_y (n = 3) \]

\[ f_{y(1), y(2), y(3)} (y_1, y_2, y_3) = \frac{1}{n!} \prod_{k=1}^{n} f(y_k) f_{y_k} f_y (n = 3) \]

Joint density of the ordering before ordering

\[ y_1, y_2, \ldots, y_n \text{ be uniform in } (0, 1) \]

\[ f_{y(1), y(2), \ldots, y(n)} = \frac{n!}{e^n} y_1 y_2 \cdots y_n \]

Conditional arrival joint density is

\[ f_{S_1, S_2, \ldots, S_n} (s_1, s_2, \ldots, s_n) = \frac{n!}{e^n} s_1 s_2 \cdots s_n \]

\[ P \{ t_k < S_k < t_k + S_k, k = 1, 2, \ldots, n^2 \} = \frac{n^2}{e^n} \]

\[ L \{ A_k, N(1) = n \} = P \{ \text{arrival in } t_k, t_k + S_k, k = 1, 2, \ldots, n^2 \} \]

\[ P \{ N(1) = n^2 \} = \frac{n^2}{e^n} \]
\[
\left( e^{-\lambda t_1} \right) \left( e^{-\lambda t_2} \right) \cdot \ldots \cdot \left( e^{-\lambda t_n} \right) = e^{-\sum_{i=1}^{n} \lambda t_i}
\]

\[
\left. \frac{e^{\lambda t}}{n!} \right|_{0 \text{ arrivals}}
\]

\[
\frac{n! - f_1 f_2 - \ldots - f_n}{n!}
\]

\[
P\{N(t) = n\} = \frac{f_1 f_2 \ldots f_n}{n!} N(t) = n + f_1 f_2 \ldots f_n.
\]

Proof is completed by cancelling \( f_1 f_2 \ldots f_n \) from both parts.

Joint dist of \( s_1, s_2, \ldots, s_n \) given \( N(t) = n \) is nothing but ordered statistics of \( n \) i.i.d. rv's with unif dist in \( [0, t] \).

The result is not surprising since Poisson process is stationary and independent increments.
Conditional Waiting Times

Given $N(t) = n$,

\[
X_1 \leq X_2 \leq \cdots \leq X_n \leq t
\]

\[
X_1 < X_2 < \cdots < X_{n-1} < X_n < t
\]

Ordered list of uniform picks in $(0,t]$ interval.

$x_n$: the waiting time from $(n-1)^{th}$ arrival to $n^{th}$ arrival.

\[
x_1 = s_1
\]

\[
x_2 = s_2 - s_1
\]

\[
x_3 = s_3 - s_2
\]

\[
x_n = s_n - s_{n-1}
\]

\[
\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{bmatrix} =
\begin{bmatrix}
1 \\
1 & 1 \\
\vdots \\
1 & 1 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
s_1 \\
s_2 \\
\vdots \\
s_n
\end{bmatrix}
\]

Joint distribution of $x_k$'s given $N(t) = n$ (LHS of *).

From joint dist. of $s_k$'s given $N(t) = n$ (RHS of *).

\[
\prod_{k=1}^{n} f(x_1, x_2, \ldots, x_n | N(t) = n) = \frac{1}{(s_1 s_2 \cdots s_n)^{n-1}} \frac{1}{(n-1)!} \det(M^{-1})
\]

\[
\prod_{k=1}^{n} f(x_1, x_2, \ldots, x_n | N(t) = n) = \frac{1}{(s_1 s_2 \cdots s_n)^{n-1}} \frac{1}{(n-1)!} \det(M)
\]

12/11/2014
\[
\frac{f_{X_1}(x_1 | N(t) = n)}{f_{X_1}(x_1) \cdot f_{N(t)}(n)} = \frac{f_{X_1}(x_1 | N(t) = n)}{f_{X_1}(x_1) \cdot \frac{e^{-\lambda t} \cdot \lambda^{n} \cdot n!}{n!}}
\]

Second way to get \( P_{X_1}(x_1 | N(t) = n) \)

\[
P_{X_1}(x_1 | N(t) = n) = \frac{P_{X_1}(x_1 > x_1 \text{ and } N(t) = n)}{P_{N}(N(t) = n)}
\]

\[
= \frac{P_{X_1}(x_1 > x_1 \text{ and } N(t) = n)}{P_{X_1}(x_1 > x_1 \text{ and } N(t) = n)} \cdot \frac{e^{-\lambda t} \cdot \lambda^{n} \cdot n!}{n!}
\]

\[
= \left( \frac{t-x_1}{t} \right)^n
\]

\[
f_{X_1}(x_1 | N(t) = n) = \frac{d}{dx_1} \left( \frac{1 - \left( \frac{t-x_1}{t} \right)^n}{cd \cdot \Phi \left( \frac{x_1 - N(t) \div n}{n} \right) \right)
\]

\[
= n \cdot \frac{(t-x_1)}{t^n}
\]
Non-Homogeneous Poisson Process:

If the rate \( \dot{\lambda}(t) \) of the Poisson process varies by time \( t \), i.e \( \lambda(t) \),
then the resultant process is called non-homogeneous Poisson process.

\[
\begin{align*}
\Pr(N(t) = 0) &= e^{-\int_0^t \lambda(s) \, ds} \\
\Pr(N(t) = 1) &= \int_0^t \lambda(s) \, ds \\
\Pr(N(t) > 2) &= e^{-\int_0^t \lambda(s) \, ds}
\end{align*}
\]

Theorem: Let \( \lambda(t) \) be rate of non-homogeneous Poisson process, then

\[
\Pr(N(t) = n) = \frac{(m(t, \Omega))^n}{n!} e^{-m(t, \Omega)}
\]

where

\[
m(t, \Omega) = \int_0^t \lambda(t') \, dt'
\]

 equivalent of Def #2 defined earlier.

Ex: A barber shop operates as follows:

00:00 - 06:00 : Closed
06:00 - 12:00 : All hours, typically 1 customer in morning hours
12:00 - 18:00 : Peak hours, typically 2 customers in afternoon hours
18:00 - 24:00 : Closed

Assume arrivals are Poisson distributed.

a) \( \Pr(2 \text{ customers in all hours}) = ? \)
b) \( \Pr(2 \text{ customers in 24 hours}) = ? \)
c) \( \Pr(2 \text{ customers in all hours} | 2 \text{ customers in 24 hours}) = ? \)
Solution:

\[ \Lambda(t) \]

\[ \frac{1}{3} \]

\[ \frac{1}{6} \]

0 6 12 18 24 30 36 42 48 (p.2)

\[ E \{ N(t) \} = \Lambda t \]

a) \[ P \{ 2 \text{ customers in all hours} \} \]

\[
M_{NN} = \int_{0}^{12} \lambda(t) \, dt = \left. \int_{0}^{12} \lambda(t) \, dt \right|_{0}^{12} = e^{-M_{NN}} = e^{-\lambda \cdot 12} = e^{-1}\]

\[ N_{NN} = 1 \]

b) \[ P \{ 2 \text{ customers in 24 hours} \} = ? \]

\[
M_{NN} = \int_{0}^{24} \lambda(t) \, dt = 3
\]

\[
P \{ N(0, 24) = 2 \} = \frac{1}{2} \cdot e^{-\frac{3}{2}} = \frac{1}{2} \cdot 0.6797 = 0.3399
\]
Solt 2: P(2 customers in 24 hours)

\[ P(2 \text{ in AM, 0 in PM}) + P(1 \text{ in AM + 1 in PM}) + P(0 \text{ in AM + 2 in PM}) \]

\[ = e^{- \frac{3}{2}} \cdot \frac{3!}{2!} + e^{- \frac{3}{2}} \cdot \frac{2!}{1!} \cdot e^{- \frac{3}{2}} \cdot \frac{2!}{2!} + e^{- \frac{3}{2}} \cdot \frac{2!}{1!} \cdot e^{- \frac{3}{2}} \cdot \frac{2!}{2!} \]

\[ = e^{- \frac{3}{2}} \cdot \left( \frac{1}{2} + \frac{2}{2} + \frac{2}{2} \right) = e^{- \frac{3}{2}} \cdot \frac{3}{2} \]

(c) P(2 AM customers | 2 customers in 24 hours)

\[ = \frac{P(2 \text{ AM, 0 PM})}{P(2 \text{ in 24 hours})} = \frac{e^{- \frac{3}{2}}}{e^{- \frac{3}{2}} \cdot \frac{3}{2}} = \frac{1}{\frac{3}{2}} = \frac{2}{3} \]

Claim: P(3 AM arrival | 2 PM arrival in 24 hours) = \(1/3\).

Since

\[ P(1 \text{ AM}) = 1, \text{ P(MA)} = 0.7 \]

\[ P(1 \text{ AM} + \text{PM}) = \frac{1}{e^{\frac{3}{2}}} \]

\[ P(1 \text{ AM} + \text{PM}) = \frac{1}{e^{\frac{3}{2}}} \]

\[ \frac{P(1 \text{ AM} + \text{PM})}{P(1 \text{ AM})} = \frac{e^{-\frac{3}{2}}}{\frac{3}{2}} \]

\[ \frac{P(2 \text{ AM} + \text{PM})}{P(2 \text{ in 24 hours})} = \frac{e^{-\frac{3}{2}}}{\frac{3}{2}} \]

\[ \frac{P(3 \text{ AM})}{P(3 \text{ AM})} = \frac{1}{3} \]

\[ \frac{P(2 \text{ AM})}{P(2 \text{ in 24 hours})} = \frac{1}{3} \]
Two arrivals in an hour gives only 2 arrivals in 24 hours. 

is then \( \frac{1}{2} \times \frac{1}{3} = \frac{1}{6} \)

Compound Poisson Process:

\[ x(t) = \sum_{i=1}^{N(t)} y_i \]

\( N(t) \) : Homogenous Poisson process with rate \( \lambda \). \( y_i \)'s are i.i.d. \( \text{dist} \) \( \mathcal{N} \) \( y \)

\( y_i \) : i.i.d. \( \text{dist} \) \( y \)

\( y_i \) are independent.

\[ E \left\{ x(t)^2 \right\} = ? \]

\[ = \sum_{n=0}^{\infty} E \left\{ x(t)^2 \mid N(t) = n \right\} P \left\{ N(t) = n \right\} \]

\[ = \sum_{n=0}^{\infty} \left( \sum_{i=1}^{n} y_i^2 \right) P \left\{ N(t) = n \right\} \]

\[ = \sum_{n=0}^{\infty} n \cdot \bar{y}^2 P \left\{ N(t) = n \right\} \]

\[ = E \left\{ N(t) \bar{y}^2 \right\} \]

\[ = (\lambda t) \bar{y}^2 \]

\[ E \left\{ x(t) z t \right\} = E \left\{ \sum_{i=1}^{N(t)} y_i \right\} \bar{y}^2 \]

\[ = E \left\{ N(t) \bar{y}^2 \right\} \]

\[ = (\lambda t) \bar{y}^2 \]
2. \( \text{Var} \{ x(t) \} = (At) E^2 X^2 \) \( \quad \text{[MIT]} \)

Check Wikipedia (Compound Poisson Process)

Ex: 

\[ X \sim N(0, .5) \]
\[ Y \sim N(0, .8) \]

\( N(0, t) \) is a Poisson process with rate \( \lambda \).

\[ E^2 N(0, t)^2 = 5 \lambda \]

\[ \text{Cov}(x, y) = E^2 (x - \mu_x)(y - \mu_y)^2 = E^2 xy - E^2 x - E^2 y \]

\[ = 5 \lambda + 6 \lambda \]

\[ E^2 N(0, 0.5) N(0, 0.8)^2 = E^2 N(0.5) [N(0.5) + N(0.8)]^2 \]

\[ = E^2 N(0.5)^2 + E^2 N(0.5) E^2 N(0.8)^2 \]

\[ = 5 \lambda + 3 \lambda^2 + 5 \lambda \]

\[ \text{Var} N^2 = E^2 N^2 = \lambda t \]

\[ \text{poisson rv} \]

\[ \text{Cov}(x, y) = 5 \lambda + 30 \lambda^2 = 30 \lambda^2 = 5 \lambda \]
Random Vectors:

A random vector is completely defined by its joint pdf of its components:

\[ x = [x_1, x_2, \ldots, x_N] \text{ with joint pdf } f_X(x_1, x_2, \ldots, x_N) \]

**Example:** Let \( X \) be a vector whose entries are iid \( N(\mu, \sigma^2) \)

Find joint pdf \( f_X(X) \)

\[
f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}
\]

\[
||x||^2 = \sum_{i=1}^{N} x_i^2
\]

\[
\frac{1}{(2\pi\sigma^2)^N} e^{-(||x||^2 - \mu x)^2/2\sigma^2}
\]

Gaussian Vector: A random vector \( \mathbf{x} \) which can be expressed as

\[ \mathbf{x} = A \mathbf{w} \]

where \( A \) is a real valued matrix

and \( f_{\mathbf{x}}(\mathbf{w}) \) is \( N(\mu \mathbf{w}, \Sigma) \)

\[
\frac{1}{(2\pi)^{N/2}} e^{-\frac{1}{2}||\mathbf{w} - \mu||^2}
\]
Special Cases:

\[ A \] in the definition is invertible

\[ x = A^{-1} w \]

\[ f_\mathbf{x}(x) = \frac{f_{\mathbf{w}}(A^{-1} x)}{|\det(A)|} \quad \text{N function of} \]

\[ \mathbf{w} \] cy

\[ \mathbf{w} \]

\[ f_\mathbf{w}(\mathbf{w}) = \frac{1}{(\sqrt{2\pi})^k |\det(S)|} e^{-\frac{1}{2} \| \mathbf{w} - \mu_w \|^2} \]

\[ \mathbf{x} = A^{-1} \mathbf{w} \]

\[ \| \mathbf{x} - \mu_x \|^2 = \| A^{-1} (\mathbf{x} - A \mu_w) \|^2 \]

\[ \frac{1}{\sqrt{\det(S)}} \frac{1}{|\det(A)|} e^{-\frac{1}{2} \| \mathbf{x} - A \mu_w \|^2} \]

\[ \mathbf{x} = A^{-1} \mathbf{w} \quad \text{condition} \]

\[ \text{matrix} \]

\[ \mathbf{z} = \mathbf{w} \]

\[ f_\mathbf{w}(\mathbf{w}) = \frac{1}{(\sqrt{2\pi})^k |\det(S)|} e^{-\frac{1}{2} (\mathbf{w} - \mu_w)^T S^{-1} (\mathbf{w} - \mu_w)} \]

Joint pdf of Gaussian vector

Notes: 1. Joint pdf only depends on \[ \mu_x \] and \[ \Sigma_{xx} \] mean vector covariance matrix

2. \[ n-1 \] we get

\[ f_\mathbf{z}(\mathbf{z}) = \frac{1}{\sqrt{2\pi} \sigma_z} e^{-\frac{1}{2\sigma_z^2} (\mathbf{z} - \mu_z)^2} \]
2\textsuperscript{nd} Moment Descriptions of Random Vectors:

1. Mean:
   \[ \mu_X = E\{X\} \]

2. Covariance:
   \[ K_X = E\left\{ (X - \mu_X)(X - \mu_X)^T \right\} \]
   Covariance matrix:
   \[
   K_X = E\left\{ \begin{bmatrix}
   X_1 - \mu_{X_1} \\
   X_2 - \mu_{X_2} \\
   \vdots \\
   X_n - \mu_{X_n}
   \end{bmatrix} \begin{bmatrix}
   X_1 - \mu_{X_1} \\
   X_2 - \mu_{X_2} \\
   \vdots \\
   X_n - \mu_{X_n}
   \end{bmatrix}^T \right\}
   \]
   
   \[
   \text{cov}(X_i, X_j) = E\{X_iX_j\} - \mu_i \mu_j
   \]

\[ \text{cov}(X_1, X_2) = E\{(X_1 - \mu_1)(X_2 - \mu_2)\} \]

Change in 2\textsuperscript{nd} Order Descriptions After a Linear Mapping:

\[ X \xrightarrow{A} Y \]

Random vector\n\[ \mu_X, \text{cov}(X) \]

New random vector\n\[ \mu_Y, \text{cov}(Y) \]
\[ y = \frac{1}{\lambda} x \rightarrow 1. \mu y = \mathbb{E} y \bar{y} = \mathbb{E} y (Ax) = A \mathbb{E} \bar{x} = \lambda \mu x \]

\[ \Rightarrow 2. \bar{y} = \mathbb{E} (y - \mu y)(y - \mu y)^T \]

\[ = \mathbb{E} \bar{y} \bar{y}^T - \lambda \mu \bar{y} \mu \bar{y}^T \]

\[ = \mathbb{E} \left( A \mu (A^T \lambda)^T \right) \bar{y} = \mu x \mu x^T \lambda^T \]

\[ = A \left( \mathbb{E} \bar{x} \bar{x}^T (y - \mu y)(y - \mu y)^T \right) \lambda^T = \mu x x^T \lambda^T \]

\[ \left[ \lambda x \right]_j = \frac{1}{N} \sum_1^n x_k \]

\[ \{ \frac{1}{N} \} \]

After a linear mapping:

1. Mean changes to \( \mu x \rightarrow \lambda \mu x \)
2. Cov changes to \( \bar{y} \rightarrow A \mu x \lambda^T \)

Note: Cov of 2 r.v's \( x \) and \( y \) are not a function of \( \mu x \) and \( \mu y \)

\[ \text{Cov}(x, y) = \mathbb{E} \left( (x - \mu x)(y - \mu y) \right) \]

\[ \text{mean subtracted, } \lambda \text{ mean subtracted} \]

Because of this, in cov. matrix and some similar calculations, there is no harm in assuming that all vectors have zero mean.
\[ X_k = \begin{bmatrix} x_1 \\ 1 \\ \vdots \\ x_N \end{bmatrix} \]

The joint pdf of all rv’s is sufficient to char. the random vector.

The measurement signal of the \( k \)-th measurement is of interest:

\[ \begin{align*}
X_k & \sim N(0, \sigma^2 I) \\
\end{align*} \]

\[ \Gamma = \frac{1}{2} s + \Theta \]

\[ \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \sim N(0, \sigma^2 I) \]

\[ \rightarrow \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \] is a non-random constant.

\[ \gamma_k = \text{indep. from measurement} \]

\[ \sum_{k=1}^{N} \gamma_k = \sum_{k=1}^{N} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = N \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \]

\[ \Delta X = \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_N \end{bmatrix} \]

\[ \begin{align*}
\hat{X}_k &= \hat{\Delta} X_k A^T b \\
\end{align*} \]

\[ \mu = \frac{1}{N} \sum_{k=1}^{N} x[k] \\
\]

\[ \sigma^2 = \frac{1}{N-1} \sum_{k=1}^{N} (x[k] - \hat{\mu})^2 \]

1. Mean vector \[ E_{X} (X) \]
2. Cov. matrix \[ E_{X} (X - \mu)(X - \mu)^T \]

\[ \text{cov. matrix does not depend on } \mu \text{ and we can assume that } \mu = 0. \text{ For } \text{cov. calculations: } \rightarrow E_{X} X_{2m} X_{2m}^T \]
\[ \mathbf{v} \rightarrow \mathbf{y} \rightarrow \mathbf{x} \]

\[ \text{covariance matrix of } \mathbf{y} \]

\[ \text{cov}(x, y) = \text{cov}(y, x) \]

\[ \text{cov}(x+y, z) = \text{cov}(x, z) + \text{cov}(y, z) \]

\[ \text{var}(x) = \text{cov}(x, x) \]

\[ \text{cov}(x, y) \neq x \text{ cov}(x, y) \]

\[ \text{var} \left( \frac{1}{N} \sum_{k=1}^{N} x_k \right) = ? \]

\[ (1) \quad \text{cov} \left( \frac{1}{N} \sum_{k=1}^{N} x_k, \frac{1}{N} \sum_{k=1}^{N} y_k \right) \]

\[ (2) \quad \sum_{k=1}^{N} \text{cov} \left( x_{k1}, \sum_{l=1}^{N} y_{l2} \right) \]

\[ (3) \quad \sum_{k=1}^{N} \text{cov} \left( x_{k1}, y_{k2} \right) \]
\[ N \sum_{k=1}^{N} \text{var}(x_k) + \sum_{k=1}^{N} \sum_{k_2 \neq k_1}^{N} \text{cov}(x_{k_1}, x_{k_2}) \rightarrow \frac{1}{2} N \sum_{k=1}^{N-1} \sum_{k_2 = k+1}^{N} \text{cov}(x_{k_1}, x_{k_2}) \]

\[ \frac{1}{N} X \begin{pmatrix} x_1 & \cdots & x_N \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \text{var}(x_1) & \text{cov}(x_1, x_2) & \cdots \end{pmatrix} \begin{pmatrix} 1 \end{pmatrix} \]

\[ \text{cov}(x) = \begin{pmatrix} \text{var}(x) & \text{var}(x) & \cdots \end{pmatrix} \begin{pmatrix} 1 \end{pmatrix} \]

\[ \text{var}(x) = \begin{pmatrix} \text{var}(x) & \text{var}(x) & \cdots \end{pmatrix} \begin{pmatrix} 1 \end{pmatrix} \]

(see document on the web with the same title)

\[ p_{xy} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}} \]

\[ |p_{xy}| \leq 1, \text{ i.e. } -1 \leq p_{xy} \leq 1 \]

\[ p_{xy} = 0 \Leftrightarrow \text{cov}(x, y) = 0, \text{ and } x \text{ and } y \text{ are uncorrelated} \]

\[ p_{xy} = \pm 1 \Leftrightarrow y = x \pm c \]

\[ \text{fully correlated random variable} \rightarrow \text{non-random constant} \]

\[ p_{xy} = 1 \rightarrow y \rightarrow x \]
$K_2$ is symmetric matrix

$K_2 = E \{ \frac{e_1}{2} \frac{e_2}{2} + \}$

$K_2 = W_2^T W_2$ (zero one many vectors)

$\Rightarrow \Rightarrow$ Eigen decomposition

$W_2 = \begin{bmatrix} e_1 & e_2 \end{bmatrix}$

$D = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$

$K_2$ $W_2 = \lambda \lambda W_2$

$K_2 \begin{bmatrix} e_1 & -e_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \ 0 \ & \lambda_2 \end{bmatrix}$

$K_2 W_2 = \lambda \lambda W_2$

$W_2 = \begin{bmatrix} e_1 & e_2 \end{bmatrix}$

Since $K_2$ is symmetric,

$\lambda$s are real valued

$e_1 \perp e_2$ for $\lambda \neq \lambda$

and

$e_1 \perp e_2$ for $\lambda_1 = \lambda_2$ can be also orthogonализed

$K_2 = W_2 W_2^T = \begin{bmatrix} e_1 & e_2 \end{bmatrix}$

diagonal matrix of real numbers
$x^T$ is positive semi-definite

A symmetric matrix $A$ is positive semi-definite if

$$x^T A x > 0, \quad \forall x \in \mathbb{R}^N$$

The quadratic form

$$\begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$$

is $> 0$.

$$a_{11} x_1^2 + a_{22} x_2^2 + \cdots + a_{NN} x_N^2 > 0$$

$A > 0 \Rightarrow A$ is positive semi-definite

$A > 0 \Rightarrow A$ is positive definite

$A < 0 \Rightarrow$ negative definite

Show $k^T x > 0$

$k^T x^2 > 0$

$x^T E x \geq 0$

$$E = \left\{ \frac{(x^T)^2}{x^T x} \right\} \geq 0$$

$$E \left\{ (x^T)^2 \right\} > 0 \Leftrightarrow$$

If these two properties are satisfied by matrix $k$,

then can I be sure that $k$ is a covariance matrix?

(Is two properties sufficient to generate a valid covariance matrix?)

Yes!

I will construct a gaussian vector with given $k$ matrix

and covariance matrix of gaussian vector will be equal to $k$ matrix.
Since $K$ is symmetric, I can decompose it as follows:

\[ K = \Theta \Lambda \Theta^T \]

\[ K^{1/2} = \Theta \Lambda^{1/2} \Theta^T \]

(applicable for $N \geq 1$)

\[ K^{1/2} K^{1/2} = \Theta \Lambda^{1/2} \Theta^T \Theta \Lambda^{1/2} \Theta^T = \Theta \Lambda \Theta^T = K \]

Then, the Gaussian vector definition says that

\[ \frac{z}{\sqrt{2}} = \Lambda^{1/2} w \]

is a Gaussian vector, provided that $w \sim N(0, 1)$

\[ \frac{z}{\sqrt{2}} = \Lambda^{1/2} w \]

So, I can create a Gaussian vector $z$ with a given $K$ as a covariance matrix provided that $K$ is symmetric and positive-semidefinite.
\[
\mathbf{a} = \begin{bmatrix} 1 & 0 \\ -\alpha & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 & 0 \\ -\alpha & 1 \end{bmatrix}, \quad \alpha > 0, \quad 2 - \alpha^2 > 0, \quad 10(\alpha^2)
\]

\[
N = 2, \quad \mathbf{z} = \begin{bmatrix} X \\ Y \end{bmatrix}
\]

\[
\mathbf{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho_{XY} \sigma_X \sigma_Y \\ \rho_{XY} \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix})
\]

\[
\det(\cdot) = \sigma_X^2 \sigma_Y^2 - \rho_{XY}^2 \sigma_X^2 \sigma_Y^2 > 0.
\]

\[
|\rho_{XY}| < 1
\]

\[
\mathbf{v}_0 = \mathbf{e} \begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} \mathbf{v}_X \\ \mathbf{v}_Y \end{bmatrix} = \begin{bmatrix} \mathbf{v}_X \\ \mathbf{v}_Y \end{bmatrix} = \begin{bmatrix} \mathbf{cov}(X,Y) \\ \mathbf{cov}(Y,X) \end{bmatrix} \quad \mathbf{cov}(X,Y) = \rho_{XY} \sigma_X \sigma_Y
\]

\[
\mathbf{cov}(X,Y) = \rho_{XY} \sigma_X \sigma_Y
\]

\[
f_2(z) = \frac{1}{(2\pi)^{N/2} \sqrt{\det(\Sigma_2)^{1/2}}} e^{-\frac{1}{2} \mathbf{z}^T \Sigma_2^{-1} \mathbf{z}} = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1-\rho_{XY}^2}} e^{-\frac{\mathbf{z}^T \mathbf{z}}{2(1-\rho_{XY}^2)}}
\]
\[ f_{xy} = 0 \quad \Rightarrow \quad \sigma_x^2 = \sigma_y^2 = \sigma^2 \quad \text{(uncorrelated } X \text{ and } Y \text{)} \]

\[ f_{xy}(x,y) = f_X(x) f_Y(y) \]

Only for Gaussian \( X \) and \( Y \) uncorrelatedness:

\[ f_{xy} = 0 \quad \Rightarrow \quad \sigma_x^2 \neq \sigma_y^2 \]

If \( \sigma_x^2 > \sigma_y^2 \):

- Larger area: Larger variation
- Closer to the origin: More concentrated

\[ \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} = \frac{1}{c^2} \]

\[ f_{xy} > 0 \quad \Rightarrow \quad \sigma_x^2 > \sigma_y^2 \]

\[ f_{xy} < 0 \]
\[ k_1 x^2 + k_2 y^2 + k_{12} xy = c^2 \]

\[ \exp \left( -\frac{1}{2} z \mathbf{z}^T \right) \]

Gaussian processes:

A stochastic process with process variable \( t \) is called Gaussian if its samples \( x(t_1), \ldots, x(t_n) \) is jointly Gaussian distributed for all \( n \) (number of samples), \( t_1, t_2, \ldots, t_n \).

\[ \mathbb{E} \{ x(t) \} = \mu_x(t) \quad \text{mean function} \]

\[ k_x(t_1, t_2) = \mathbb{E} \left\{ [x(t_1) - \mu_x(t_1)] [x(t_2) - \mu_x(t_2)] \right\} = \text{cov. function} \]

\[ k_x(t_1, t_2) = 3 e^{-|t_1 - t_2|} \]

\[ \mathbb{E} \{ x(s) \} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \quad \mathbb{E} \{ x(s) x(s') \} = \begin{bmatrix} 3 & 3 e^{-5} \\ 3 e^{-5} & 3 \end{bmatrix} \]

Note: Gaussian process is completely characterized by mean function and covariance function.

\[ \mathbb{E} \{ x(t) \} = 2t + 3 = \mu_x(t) \]

\[ k_x(t_1, t_2) = 3 e^{-|t_1 - t_2|} \]

a) Find pdf \( x(5) \), \( x(5) \sim \mathcal{N}(\mu_x(5), \text{var}_x(x(5), x(5))) = \mathcal{N}(13, 3) \)

b) Find joint pdf \( x(5), x(10) \)

\[ \begin{bmatrix} x(5) \\ x(10) \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu_x(5) \\ \mu_x(10) \end{bmatrix}, \begin{bmatrix} k_x(5, 5) & k_x(5, 10) \\ k_x(10, 5) & k_x(10, 10) \end{bmatrix} \right) \]

\[ = \begin{bmatrix} 13 \\ 23 \end{bmatrix}, \quad \begin{bmatrix} 3 & 3 e^{-5} \\ 3 e^{-5} & 3 \end{bmatrix} \]

c) \[ \mathbb{E} \{ (x(5) - x(6))^2 \} = \mathbb{E} \{ (x(5))^2 \} + \mathbb{E} \{ (x(6))^2 \} - 2 \mathbb{E} \{ x(5) x(6) \} \]

\[ = (k_x(5, 5) + \mu_x(5)^2) + (k_x(6, 6) + \mu_x(6)^2) - 2(k_x(5, 6) + \mu_x(5) \mu_x(6)) \]
Ex: $X_n = X_n^{(n-1)} + zn$, $n > 1$; $X_0 \sim N(\mu, \sigma^2)$

Given is: stable and linear

$X_n \sim N(\mu, \sigma^2)$

Initial condition:

Weakly Gaussian with dist: $N(0, \sigma^2)$ and $w_i$s are independent of $X_0$

1) Find pdf of $X_n$, $n \geq 0$

$X_n = X_0 \cdot a^n + \sum_{k=1}^{n-1} a^{n-k} w_k$, $n > 0$

Zero-input solution: $X_n = 0^n + \sum_{k=1}^{n-1} a^{n-k} w_k$

$X_0 = 0^n$

$x_1 = a \cdot x_0 + w_1$

$x_2 = a \cdot (a \cdot x_0 + w_1) + w_2$

Solution: $X_n = a^k \cdot x_0 + \sum_{k=1}^{n-1} a^{n-k} w_k$.

So, $X_n$ is Gaussian distributed.

$E[X_n] = \mu a^n$

$\text{Var}(X_n) = \text{Var}(a^n X_0) + \sum_{k=1}^{n} a^{n-k} \text{Var}(w_k)$ (since $X_0$ and $w_k$s are independent)

$= a^{2n} \sigma_0^2 + \sum_{k=1}^{n} a^{n-k} \text{Var}(w_k)$ (and $w_k$s are independent)

$= a^{2n} \sigma_0^2 + \sum_{k=1}^{n} \sigma^2 a^{2(n-k)} \text{Var}(w_k)$

$= a^{2n} \sigma_0^2 + \left[ \sum_{k=1}^{n} a^{2(n-k)} \right] \sigma^2$

$= a^{2n} \sigma_0^2 + \frac{a^{2n} - 1}{a^2 - 1} \sigma^2$

$= \frac{a^{2n} \sigma_0^2 + (1 - a^{2n}) \sigma^2}{1 - a^2}$
\[
\begin{align*}
\text{a)} & \quad \text{As } n \to \infty, \quad \text{var}\{x_n\} \to 0, \\
\text{b)} & \quad kX[n, k] = E\left\{ (X_n - \bar{X}_n)(X_k - \bar{X}_k) \right\} \\
& \quad = E\left\{ (X_n - \mu_0)^2 + 2 \sum_{k=1}^{n-k} w_{k-1} \right\} \\
& \quad = \bar{\sigma}^2 + \sum_{k=1}^{n-k} w_{k-1} \\
& \quad = \bar{\sigma}^2 + \sum_{k=1}^{n-k} \frac{k(k-1)}{2} \sigma^2 \\
& \quad = \bar{\sigma}^2 + \frac{n-k-1}{2} \sigma^2 \\
& \quad = \bar{\sigma}^2 + \frac{n-k}{2} \sigma^2 \\
& \quad \text{Assume that } \quad k \leq n \\
& \quad \min(k, n) = k \\
& \quad \text{without any loss of generality.}
\end{align*}
\]
c) Joint pdf

\[
\begin{bmatrix}
X_3 \\
X_5
\end{bmatrix} \sim N
\begin{bmatrix}
\mu_{X3} \\
\mu_{X5}
\end{bmatrix},
\begin{bmatrix}
\Sigma_{X3} & \Sigma_{X3X5} \\
\Sigma_{X5X3} & \Sigma_{X5}
\end{bmatrix}
\]

Stationary Process:

1st order stationarity:

A process \( X(t) \) is 1st order stationary if

\[
f_X(x(t_1) | t_1) = f_X(x(t_1 + A) | t_1 + A)
\]

density for \( X(t) \)

2nd order stationarity:

A process is 2nd order stationary if

\[
f_{X(t_1), X(t_2)}(x_1, x_2) = f_{X(t_1 + A), X(t_2 + A)}(x_1, x_2)
\]

joint pdf for \( X(t_1) \) and \( X(t_2) \)

\( N \)th order stationarity:

\[
f_{X(t_1), X(t_2), \ldots, X(t_N)}(x_1, x_2, \ldots, x_N) = f_{X(t_1 + A), X(t_2 + A), \ldots, X(t_N + A)}(x_1, x_2, \ldots, x_N)
\]

A process is strict sense stationary (SSS) if it is \( N \)th order stationary for all \( N \).

Let's focus on Gaussian process and examine the conditions for which Gaussian process is stationary.
**1st order:**

\[ x(t) \sim \mathcal{N}(\mu_x(t), \sigma_x^2(t)) \]

\[ x(t) \sim \mathcal{N}(\mu_x(t), \sigma_x^2(t)) \quad (t \neq t_1) \]

So, 1st order stationarity requires:

1. \( \mu_x(t) = \mu_x(t_1) \) since equalities
2. \( \sigma_x^2 = \sigma_x^2(t_1) \)

Then:

\[ \mu_x(t) = \text{constant} \]

\[ \sigma_x^2(t) = \text{constant} \]

**2nd order:**

\[ \begin{bmatrix} x(t) \\ x(t_A) \end{bmatrix} \sim \mathcal{N}( \begin{bmatrix} \mu_x(t) \\ \mu_x(t_A) \end{bmatrix}, \begin{bmatrix} \text{cov}(x(t), x(t)) & \text{cov}(x(t), x(t_A)) \\ \text{cov}(x(t_A), x(t)) & \text{cov}(x(t_A), x(t_A)) \end{bmatrix} ) \]

\[ \begin{bmatrix} x(t+\Delta) \\ x(t_2+\Delta) \end{bmatrix} \sim \mathcal{N}( \begin{bmatrix} \mu_x(t+\Delta) \\ \mu_x(t_2+\Delta) \end{bmatrix}, \begin{bmatrix} \text{cov}(x(t+\Delta), x(t+\Delta)) & \text{cov}(x(t+\Delta), x(t_2+\Delta)) \\ \text{cov}(x(t_2+\Delta), x(t+\Delta)) & \text{cov}(x(t_2+\Delta), x(t_2+\Delta)) \end{bmatrix} ) \]

Then 2nd order stationarity:

1. \( \mu_x(t) = \text{constant} \quad \forall t, \Delta \)
2. \( \text{cov}(x(t), x(t_2+\Delta)) = \text{cov}(x(t), x(t_2+\Delta)) \quad \forall t \)

Then set \( \Delta = -t_1 \)

| Then \( \text{cov}(x(t), x(t_2+\Delta)) = \text{cov}(x(t), x(t_2-t_1)) \)
| Then \( \text{cov}(x(t), x(t_2+\Delta)) \) can be written as a function of \( u = t_2-t_1 \) (function of single variable)
\[ X(t) \sim N \left( \mu_x(t), \sigma_x^2(t) \right) \]

\[ X(t) \sim N \left( \mu_x(t+1), \sigma_x^2(t+1) \right) \]

So, for \( N \)th order stationarity:

1. \( \mu_x(t) = \text{constant} \)

2. \( \text{cov}(x(t), x(t+1)) = \text{cov}(x(t+1), x(t+2)) \quad \forall t \left[ \text{since } \text{cov}(x(t), x(t+1)) = \text{cov}(x(t), x(t+2)) \right] \)

Wide sense stationarity:

In many applications, stationarity in the strict sense can not be checked or guaranteed, so we use a relaxed form of stationarity which is more practical in many applications:

\text{WSS:}

1. \( E[X(t)^2] = \text{constant} \quad \forall t \) if \( 1 \) and \( 2 \)

2. \( \text{cov}(x(t_2), x(t_3)) = \text{func}(t_2-t_1) \quad \forall t \)

\text{Note:} 1. A 1\textsuperscript{st} order stationary process is also stationary in the mean.

\[ \mu_x(t) = \text{constant} \quad (\text{1\textsuperscript{st} check of WSS}) \]
2. A 2nd order stationary process is also stationary in the covariance function.

So from 1 and 2 we can say that 2nd order stationarity implies WSS.

A process that is 2nd order stationary is guaranteed to be first order stationary since
\[ f(x_1, x_2) = \int \int f(x_{1+h}, x_{2+h}) \, dx_1 \, dx_2 \]
can be marginalized wrt to \( x_1 \) \[ \int_{-\infty}^{\infty} f(x_{2+h}) \, dx_1 \] then we have 1st order stationarity.

Note: WSS does not imply 1st order stationarity.

Since WSS is about moments but not joint pdf's.

Note: An important special case is Gaussian processes.
Stationarity requires for joint pdf sense coincides with WSS checks.
so, \( X(t) \) Gaussian process and WSS \( \Rightarrow \) \( X(t) \) Gaussian, SSS.

Implies 03/12/2014

STATIONARY PROCESS

1) Stationarity in joint-pdf (1st order, 2nd order, .., nth order -- SSS):

2) Wide sense stationarity (WSS):

- 1. \( E[X(t)] = \text{constant} \Rightarrow \text{Stationary in the mean} \)
- 2. \( E[X(t)X(t-\tau)] = \text{func}(\tau) \Rightarrow \text{Stationary in autocorrelation} \)

* 2nd order stationarity \( \Rightarrow \) WSS
* WSS not even 1st order stationarity
  about moments  about pdf
* If process is Gaussian \( \Rightarrow \) SSS
  and WSS
  expected value & cov are invariant of the t
Ex: \( p(t) \)

\[
x(T) = \sum_{k=0}^{\infty} a_k p(t - kT) = x(t)
\]

\( a_k \) is i.i.d \( N(0, \sigma^2) \)

\( T = 1 \), \( 2T = 3T \)

\( \Rightarrow \)

\( x(t) \)

\( x(T) \)

\( x(t) \)

\( x(t) \)

\( \text{Is } x(t) \text{ stationary?} \)

1st order stationary:

\[
f_{x(t)}(x_1) \sim N(x_1, \sigma^2)
\]

\[
f_{x(t)}(x_1) \sim N(x_1, \sigma^2)
\]

So, \( x(t) \) is 1st order stationary

\( l(x) = \text{floor} \) (least integer less than \( x \))

2nd order stationary:

\[
f_{x(t_1), x(t_2)}(x_1, x_2) = \begin{cases} f_{X(t_1)}(x_1) f_{X(t_2)}(x_2), & \text{if } |t_1 - t_2| = 0 \\ f_{X(t_1)}(x_1) f_{X(t_2)}(x_2) \cdot \text{other}, & \text{if } |t_1 - t_2| > 0 \\ f_{X(t_1)}(x_1) f_{X(t_2)}(x_2), & \text{if } |t_1 - t_2| = 0 \\
\end{cases}
\]

Ex: 3

\[
\begin{align*}
\text{Ex:} & \\
\Delta &= T/3 \\
& & T/3/1_e \quad T/2/1_e
\end{align*}
\]
So this example shows that $x(t)$ is not 2nd order stationary.

No need to check 3rd or higher order stationarities.

Q: Is $x(t)$ WSS?

1. $E\{x(t)\} = ? \implies E\{x(t)\} = 0$

2. $E\{x(t)x(t-\tau)\} = \text{func}(\tau)$
   $E\{x(t_1)x(t_2)\} = \text{func}(t_1-t_2)$

\[E_{\tau_{12}}(t_1 = \frac{\tau}{2}, t_2 = \frac{3\tau}{4})\text{ pulse}\]

\[-E\{x(t_1)x(t_2)\} = E\{x_0^2\} = \sigma^2\]

\[t_{12} = \frac{\tau}{2} + \Delta, \quad t_{22} = \frac{3\tau}{4} + \Delta\]

\[A = \frac{\tau}{3}, \quad \Delta = \frac{\tau}{12}\]

\[E\{x(t_1)x(t_2)\} = E\{x_0^2\}E\{\sigma_2^2\} = 0, 0 = 0\]

$x(t)$ is stationary in the mean.

$x(t)$ is not stationary in the autocorrelation.

$x(t)$ is not WSS.

Properties of Auto-correlation Function for WSS Processes:

For zero-mean processes, autocorrelation function $E\{x(t_1)x(t_2)\}$ is identical to covariance function $E\{x(t_1) - \mu)(x(t_2) - \mu)\}$.

\[r_x[k] = E\{x[n]x^*[n-k]\} = \text{autocorrelation sequence}\]

\[r_x[k] = r_x[-k]\]

Proof:

\[r_x[k] = E\{x[n]x^*[n-k]\} = E\{x[n+k]x[n]\}^* = r_x[k]^*\]
So, if $x[n]$ is real-valued, $r_x[k]$ is an even sequence: $r_x[k] = r_x[-k]

x[n]$ is complex valued, $r_x[k]$ is Hermitian symmetric: $r_x[k] = r_x^*[k]$.

2. $r_x[0] > |r_x[k]| \quad \forall k$

$r_x[0] = \sum_{n=-\infty}^{\infty} x[n] x[n]^*$

$\sum_{n=-\infty}^{\infty} x[n] x[n]^* = \sum_{n=0}^{\infty} x[n]^2$

Ensemble power of the process $x[n]$.

Proof: $r_x[0] = \sum_{n=-\infty}^{\infty} x[n] x[n]^*$

$\omega = x[n-k]$  \quad ($\text{Pauli model}$)

$\varphi = \frac{1}{2} \log \left( \frac{E_x^2 + \omega^2}{E_x^2} \right)$  \quad (only valid for $\omega \neq 0$)

$\omega = \frac{\sqrt{r_x[0] + \varphi}}{}$

$r_x[0] > |r_x[k]|$ \quad $\forall k$

3. If $r_x[0] = r_x[\pi] \quad \exists N \neq 0$

$r_x[k]$ is periodic by $N$: $r_x[k] = r_x[k+N] \quad \forall k$

Proof: $\omega = x[n-k] - x[n-k+N]$

$2\pi k \quad \Rightarrow \omega = x[n]$.

If $|\omega| = 1$, then $(E_x^2 \omega^2) < \sum_{k=0}^{N-1} (E_x^2 \omega^2) \Rightarrow 2(r_x[0] - r_x[N]) = 0$

$(r_x[k] - r_x[k+\pi])^2 \leq \left( \sum_{k=0}^{N-1} r_x[k] \right)^2$

$= E_x^2 \sum_{k=0}^{N-1} x[n-k]^2 - 2x[n-k]x[n-k+N] + x[n-k+\pi]^2$

$= E_2 [\sum_{n=0}^{N-1} x[n-k] - 2x[n-k]x[n-k+N] + x[n-k+\pi]]$

So, this shows that $(r_x[k] - r_x[k-N])^2 < 0$

$r_x[k] = r_x[k-N] \quad \forall k$
1. **Autocorrelation Matrix for WSS \( x[n] \)**

\[
\begin{align*}
\tilde{\mathbf{R}} &= \begin{bmatrix} x[n] \\
x[n-1] \\
x[n-2] \end{bmatrix} \\
&= \mathbf{E} \begin{bmatrix} x[n] \\
x[n-1] \\
x[n-2] \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 & \mathbf{r}_2 \\
\mathbf{r}_1^T & \mathbf{r}_0 & \mathbf{r}_1 \\
\mathbf{r}_2^T & \mathbf{r}_1^T & \mathbf{r}_0 \\
\end{bmatrix}
\end{align*}
\]

2. **Hermitian symmetric matrix**

\[2 \Rightarrow \mathbf{r} \geq 0\]

3. **Toeplitz Structure**

Over diagonal and sub/super diagonals, we have the same value in the matrix.

4. **A sequence is a valid autocorrelation sequence if and only if**

\[
\mathbf{r} = \mathbf{E} \begin{bmatrix} x[n] \\
x[n-1] \\
x[n-2] \end{bmatrix} \quad (\mathbf{r} \geq 0)
\]

\[2 \Rightarrow \mathbf{r} \geq 0 \text{ for all } N. \ (\mathbf{r} \in \mathbb{R}^{n \times n})\]
Filtering of WSS Processes:

\[ y[n] = \sum_{k=-\infty}^{\infty} h[k] x[n-k] \]

If \( x[n] \) is WSS, what can we say about the stationarity of \( y[n] \)?

Let's check whether \( y[n] \) satisfies WSS conditions:

1. \( \mathbb{E}[y[n]] = \text{constant} \)

\[ \mathbb{E}[y[n]] = \mu \sum_{k=-\infty}^{\infty} h[k] = \mu \mathcal{H}(j\omega) \]
\[ \mathcal{H}(j\omega) = \sum_{k=-\infty}^{\infty} h[k] e^{-j\omega k} \]

Provided that \( \mathcal{H}(j\omega) \) is finite (i.e., no poles at \( \omega = 1 \)), then \( \mu = 1 \).

\[ \mathbb{E}[y[n]] = \text{constant} \]

So, 1st condition for WSS is satisfied.

2. \( \mathbb{E}[y[n] y[n-k]] = \text{func}(k) \) \forall n \)

\[ \mathbb{E}[y[n] y[n-k]] = \mathbb{E}[\sum_{k=0}^{\infty} h[k] x[n-k] x[n-k]] \]

Since \( x[n] \) is WSS:
\[ \sum_{k=0}^{\infty} h[k] \mathbb{E}[x[n-k] x[n-k]] = h[k] \mathbb{E}[x[n] x[n-k]] \]

Then, \( \mathbb{E}[y[n] y[n-k]] = \mathbb{E}[y[n] x[n-k]] = h[k] \mathbb{E}[x[n] x[n-k]] \)

depends only on \( h[k] \)
\[ E\{y[n]y[n-k]\} = \sum_{k=0}^{\infty} h[k]y[n-k] \]

\[ = E\{y[n]\} \sum_{k=0}^{\infty} h[k]x[n-k] \]

\[ = \sum_{k=0}^{\infty} h[k] \cdot E\{x[n-k]\} \]

Then \[ E\{y[n]y[n-k]\} = h[-k] \star r_{xx}[k] \]

Joint WSS: Two processes \( x[n], y[n] \) are called jointly WSS if:

1. \( x[n] \) is WSS
2. \( y[n] \) is WSS
3. \( E\{x[n]y[n-k]\} = h[k] \star r_{xx}[k] \)

Comment: If \( x[n] \) is WSS, then \( x[n], y[n] \) are jointly WSS.
Power Spectral Density:

\[ r_y[k] \stackrel{\text{DFT}}{\leftrightarrow} S_y(e^{j\omega}) \]

Auto-correlation of a WSS sequence \( y[n] \)

\[ S_y(e^{j\omega}) = \mathcal{F}\{r_y[k]\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} r_y[k] e^{-j\omega k} \, dk \]

\[ r_y[k] = \mathcal{F}\{S_y(e^{j\omega})\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} S_y(e^{j\omega}) e^{j\omega k} \, d\omega \]

\[ S_y(e^{j\omega}) = H(e^{j\omega}) H(e^{j\omega})^* S_x(e^{j\omega}) \]

\[ S_y(e^{j\omega}) = |H(e^{j\omega})|^2 S_x(e^{j\omega}) \]

PSD of the output

Note: Cross power spectral density

\[ S_{yx}(e^{j\omega}) = \mathcal{F}\{r_{yx}[k]\} \]
Properties of \( S_y(e^{j\omega}) \)

1. \( S_y(e^{j\omega}) \) is real-valued: \( S_y(e^{j\omega}) = S_y(e^{-j\omega}) \)

2. \( S_y(e^{j\omega}) \geq 0 \)

Proof: Assume \( S_y(e^{j\omega}) \) is the output of a (TI) system with input \( S_X(e^{j\omega}) = 1 \) (white noise), white noise \( x[k] = \delta[k] \).

Then \( S_y(e^{j\omega}) = \lim_{\tau \to \infty} \int S_x(e^{j\omega}) \phi_k(e^{j\omega}) \, d\omega \)

3. Any non-negative function say \( S(e^{j\omega}) \geq 0 \) is a valid power spectral density?

Yes. For proof see 503 notes or Papulis book on prob and random variables.

Power Spectral Density (Cont'd)

\( S_y(e^{j\omega}) = |DTFT\{r_y[k]\}|^2 \)

Note:
\( r_y[n] = DTFT \{ S_y(e^{j\omega}) \} \)

\[ E\left\{ (y[n] - \bar{y})^2 \right\} = \frac{1}{2\pi} \int S_y(e^{j\omega}) \, dw \]

Ensemble power: \( \frac{1}{2\pi} \int S_y(e^{j\omega}) \, dw \)

Power of random sequence \( y[n] \):

\[ \frac{1}{2\pi} \int_0^\pi S_y(e^{j\omega}) \, dw \]

Properties:

1. \( S_y(e^{j\omega}) \) real valued.
2. \( S_y(e^{j\omega}) \geq 0 \)
3. Any non-negative function with finite "area" can be considered as a PSD of a process.
**EA:**

\[ S_x(e^{j\omega}) \]

**E:**

\[ S_x(e^{j\omega}) = A \quad \text{if } \omega \in (-\pi, \pi] \]

\[ S_x(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{power}^2 \]

\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} (2\pi \sigma^2) = \sigma_w^2 \]

\[ \text{white noise} \]

\[ 2\pi \quad \omega \]

\[ \text{If PSD is flat} \]

\[ (i.e. \text{constant}) \]

\[ \text{for all } \omega, \text{ then} \]

\[ \text{the associated process} \]

\[ \text{is called white noise} \]

\[ \text{(Equally in all definitions} \]

\[ \text{we assume noise is zero-mean} \]

**Ex:**

\[ H(z) = \frac{1}{1 - e^{-j\omega} z^{-1}} \]

\[ x[n] \]

\[ y[n] \]

\[ \text{white noise} \]

\[ \text{single-pole LTI system} \]

\[ \text{with variance} \]

\[ \sigma_w^2 \]

\[ S_x(e^{j\omega}) = |H(e^{j\omega})|^2 \cdot S_x(e^{j\omega}) \]

\[ H(e^{j\omega}) = H(z) = \frac{1}{1 - e^{-j\omega} z^{-1}} \]

\[ H(e^{j\omega}) \]

\[ \omega = \omega_0, \ \omega = -\omega_0 \]

\[ \text{low-pass filter} \]

\[ |H(e^{j\omega})|^2 \]

\[ \text{high-pass filter} \]

\[ |H(e^{j\omega})|^2 \]

\[ \omega = 0, \ \omega = \pm \pi \]

**Note:** The image contains handwritten mathematical equations and diagrams related to signal processing and Fourier transforms. The text includes explanations of white noise, single-pole LTI systems, and various frequency domain representations of signal characteristics. The diagrams illustrate frequency responses and transformations.
In the figure, \( x(t) \) is a low-pass filter. \( y(t) \) is called a low-pass process.

**Interpretation for PSD:**

\[
H(e^{j\omega}) \rightarrow \alpha(n) \rightarrow \{H(2\omega) \}
\]

\( H(e^{j\omega}) \) is a low-pass filter centered around \( w = w_x \).

**What's output process variance?**

\[
E\{\alpha(n)^2\} = \frac{\pi}{2\pi} \int_{-\pi}^{\pi} S_H(e^{j\omega}) d\omega = \frac{1}{2\pi} \int_{0}^{\pi} S_H(e^{j\omega}) d\omega = \frac{1}{2\pi} \int_{0}^{\pi} \left( \frac{2\pi}{\epsilon} \right)^2 S_H(e^{j\omega}) d\omega = \frac{1}{2\pi} \int_{0}^{\pi} S_H(e^{j\omega}) d\omega \quad \left( \frac{\epsilon}{2} < \pi \right)
\]
Markov Chains:

Let \( X_n, n = 0, 1, 2, \ldots \) be a random process taking finite or countable number of possible values, \( X_n \in \{1, 2, 3, \ldots \} \).

We call \( X_n \) as the state of the process at time \( n \) and consider that process jumps from state to state with some probabilities at any time instant.

The process is said to be Markov, if

\[
\mathbb{P}\{X_{n+1} = j | X_n = i, X_{n-1} = X_{n-2}, \ldots, X_1 = X_0, X_0 = x_0\} = \]

\[
P_{ij} = \mathbb{P}\{X_{n+1} = j | X_n = i\} = P_{ij} \quad \text{state transition probability from state } i \text{ to state } j.
\]

So, given the present state \( (X_n) \), the future state \( (X_{n+1}) \) is independent from states in history \((X_0, X_1, \ldots, X_{n-1})\).

This kind of independence is called conditional independence.

So, future is independent from past given present sample.

\[
\mathbb{P}\{X_{n+1} = A | X_n = B, X_{n-1} = X_{n-2}, \ldots, X_1 = X_0, X_0 = x_0\} = \]

\[
\mathbb{P}_{n+1} = \mathbb{P}_{ij} = \mathbb{P}\{X_{n+1} = j | X_n = i\} = P_{ij}
\]

\[\]
\[ P_2^2 x_n = A^2 \quad P_2^2 x_n = B \quad x_{n+1} = A \]

So, \( P_2^2 x_n \rightarrow A \) (\( x_n = B \)), \( x_{n+1} = x_{n+2} \) does only depend \( x_n \) (current time) but not on future samples (F).

So, this says that a Markov chain in "reverse time" is also a Markov chain with different state transition probabilities.

The Markov chains are denoted as:

\[ a \rightarrow b \rightarrow c \]

(these \( a,b,c \) are r.v.'s)

\( a = x_{n+1}, \quad b = x_n, \quad c = x_{n+2} \)

So, if I have \( a \rightarrow b \rightarrow c \),

then, \( a \rightarrow b \rightarrow c \) is also a Markov chain with different transition probabilities.

In some books: \( a \rightarrow b \rightarrow c \)

A Markov chain whose transition prob. does not change by time "a" are called "homogeneous" Markov chains. We will mostly focus on homogeneous Markov chains.

**Ex:** Spider and Fly

A spider is located at A, a fly moves randomly between A, B, C, D positions without any knowledge of spider.

The moves are according to some assigned probabilities:

- \( A \rightarrow B \rightarrow C \rightarrow D \rightarrow A \)
- \( C \rightarrow D \rightarrow C \)
- \( D \rightarrow A \rightarrow D \)
- \( B \rightarrow D \rightarrow B \)
Notes:

1. Row sums of $P$ matrix is equal to 1 (5.1 continuity of states)

   $\sum_{j=1}^{s} P_{i,j} = 1$

   (since $\sum_{j=1}^{s} P_{i,j} = 1$)

   Such $P$ matrices are called stochastic matrices.

2. If both row and column sum is equal to 1

   such matrices are called doubly stochastic matrices.

3. What is $P\{\text{Fly captured} | X_o = 10\}$?

   $P\{\text{Fly captured} | X_o = 10\} = \begin{pmatrix}
   1 & 10 & 1 \n   7 & 10 & 3 \n   0 & 10 & 4
\end{pmatrix}$
the value for \( Z \) is not uncertain, we need to calculate it, but we can see that safety is reached since in the first transition, \( A \) can reach states 3 and 4 (reaches safety) with 0.6 prob.

Q1. What is 2-step transition prob?

\[ P^2_{X_{n+2} = A | X_n = B} = ? \]

\[ A: P^2_{X_{n+3} = A | X_n = B} = \sum_{k=1}^{15} P^2_{X_{n+2} = A, X_{n+1} = s | X_n = B} \]

\[ = \sum_{k=1}^{15} P_{k} \cdot P_{k+1} = \sum_{k=1}^{15} P_{k} \cdot \frac{N_{k}}{k+1} \]

\[ [A, B]_{ij} = \frac{N_{i}}{i} \cdot \frac{N_{j}}{j} \]

So, two-step transition matrix is

\[ P^2 = P \cdot P \]

1-step transition matrix:

\[ P^{1A} \]

\[ P^{2A} \]

\[ X_{n+1} \]

\[ X_{n+2} \]
So, 3-step transition matrix $P^3$

$n$-step transition matrix $P^n$

Chapman-Kolmogorov Equation:

\[ P^{n+m} = P^n P^m \]

\[ P_{ij}^{(n+m)} = \sum_{k=1}^{m} P_{ik}^{(n)} P_{kj}^{(m)} \]

Chapman-Kolmogorov Equation

For a Markov chain, discrete

1 m-step transition probability $k \rightarrow j$

Gaussain process, increment small, soребк

Take-home vertex - Hunter - chain computer assignment

Markov Chains (cont'd.):

Fly-spider

\[
P = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0.3 & 0.3 & 0.4 \\
0 & 0.5 & 0.5 & 0 \\
0 & 0.5 & 0.5 & 0.5
\end{bmatrix}
\]

$p_{ij} = \text{Prob of moving } i \rightarrow j$

Ω. $P_{ij}$ fly moves to state 1 for the first time $X_n = 2$ ?

(meets the spider in time $n$)

Initial state

\[ P_{ij} X_0 = 1 \rightarrow X_{n+1} \rightarrow X_{n+2} = 1 \rightarrow X_n + 1 \rightarrow X_0 + 2 \]

= $P_{ij}$ fly remains in state 2 for $3, 4, 5, \ldots$ $q_{ij}$ if $X_0 = 2$

and $X_n = 1,$

= $0.5 \cdot 0.5 = 0.25$
D. \( P_{\frac{3}{2}} \) fly meets spider at any time \( x_0 = 2 \)

\[
= \sum_{n=1}^{\infty} P_{\frac{3}{2}} \text{ fly meets spider first time at instant } n \ | \ x_0 = 2
\]

\[
= \sum_{n=1}^{\infty} (0.3)^{n-1} \cdot 0.1 \cdot \frac{A_{F_{\frac{3}{2}}(n)}}{1 - 0.3} = \frac{0.6}{1 - 0.3} = F_{\frac{3}{2}}(0)
\]

\[ F_{\frac{3}{2}}(n) = F_{\frac{3}{2}}(0) \]

E. \( P_{\frac{3}{2}} \) fly escapes to safety at time \( n \ | \ x_0 = 2 \)

\[
= (0.3)^{n-1} \cdot 0.6 \cdot \frac{A_{F_{\frac{3}{2}}(n)}}{1 - 0.3} = F_{\frac{3}{2}}(0)
\]

\[ F_{\frac{3}{2}}(n) = F_{\frac{3}{2}}(0) \]

F. \( P_{\frac{3}{2}} \) fly escapes at any time and remains safe

\[
= \sum_{n=1}^{\infty} (0.3)^{n-1} \cdot 0.6 \cdot \frac{A_{F_{\frac{3}{2}}(n)}}{1 - 0.3} = \frac{0.6}{1 - 0.3} = F_{\frac{3}{2}}(0)
\]

\[ F_{\frac{3}{2}}(n) = F_{\frac{3}{2}}(0) \]

G. \( P_{\frac{3}{2}} \) fly remains in state 2 at time \( n \ | \ x_0 = 2 \) \( = (0.3)^n \)

H. \( P_{\frac{3}{2}} \) fly returns back to state 2 at any time \( x_0 = 2 \) \( = 0.3 \)

State 2: "transient state"

State 1: Danger state (fly remains there!)

States 3-4: Safety state (fly reaches safety states and remains there!)

Classification of States:

Recurrent/Transient: state j is recurrent if and only if \( \text{Pr}(F_{j} = 1) \)

\[
\text{prob visiting state } j \ | \ x_0 = 2 \text{ at any time }
\]
State $j$ is transient if $P_{ij} < 1$, that is, not returning to state $j$ (starting from state $j$) has a non-zero probability.

**Absorbing State (Trapping State):**

State $j$ is absorbing if $P_{jj} = 1$ (no way out!)

Clearly, absorbing state is a recurrent state.

**Periodic State:**

State $j$ is periodic if there exists an integer $N (N > 2)$ for which $P_{jj}^{(N)} = 1$.

**Definitions:**

1. $i \rightarrow j$: state $j$ is accessible from state $i$.
   
   (there exists a sequence of transitions connecting state $i$ to state $j$)

2. $i \leftrightarrow j$: states $i$ and $j$ communicate, i.e.

   $i \rightarrow j$ and $j \rightarrow i$.

   By definition, every state communicates with itself.

**Class:** A set of communicating states.

**Example:**

- $A \leftrightarrow B$
- $D \leftrightarrow E$

Classes = \{ $A, B$, $D, E$, $C$ \}
Observations: Classes partition the states into sets.

- Each state is in one and only one class.
- There are no states without classes.

Fact: States of a class are all recurrent or transient (all states of a class have the same property).

Proof: Theorem 4.2.6, page 5 of textbook.

- \[ P \text{Ever reaching } 1 \mid X_0 = 2 \frac{5}{4} \]

\[ = \sum_{k=1}^{4} P \text{Ever reaching } 1 \mid X_k = k, X_0 = 2 \frac{5}{4} \]

\[ = \frac{5}{4} P \text{Ever reaching } 1 \mid X_1 = 1, X_0 = 2 \frac{5}{4} \]

\[ = \frac{5}{4} P(A \mid B) = P(A \mid B) \cdot P(B) \]

\[ P(A, B, C) = P(A \mid B, C) \cdot P(B, C) \]

\[ = \sum_{k=1}^{4} P \text{Ever reaching } 1 \mid X_k = k \frac{5}{4} \]

\[ = \sum_{k=1}^{4} \text{F}_k \cdot P \text{Ever reaching } 1 \mid X_k = k \]

\[ = \sum_{k=1}^{4} F_{k1} P_{21} = \sum_{k=1}^{4} P_{2k} F_{k1} \]

\[ F_{21} = \begin{bmatrix} 0.3 & 0.4 & 0.2 & 0.2 \\ 0.1 & 0.5 & 0.5 & 0.5 \\ 0.2 & 0.5 & 0.4 & 0.1 \\ 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix} \]

- Second row of P matrix.
\[ F_{1i} = 0.1 + 0.3 F_{1i} \]

\[ F_{1i} = \frac{0.1}{0.7} = \frac{1}{7} \]

For this problem, \( F_{1i} \) is written very simply from "boundary" conditions.

**Notes:**

- For this example, the boundary conditions are written by simply noting the class of the recurrent states.

  \[
  \text{Classes: } 1, 2, 3, 4, 5, 8, 12, 17
  \]

  \[
  \text{Transient, recurrent, recurrent}
  \]

1. \( F_{ij} = 1 \) if states \( i \) and \( j \) belong to the same recurrent class.

2. \( F_{ij} = 0 \) if states \( i \) and \( j \) do not belong to the same recurrent class.

\( F_{ij} \) is not trivially written for transient states, we need to make calculations to find

\[ P \text{ ever reaching a state} \]

\[ \text{as a recurrent state, or possibly another transient state} \]

For example, \( F_{11} = \frac{1}{7} \) is the transient to a recurrent class ever reaching probability.
In Matrix Calculation (Transient to Recurrent states):

\[ F_{ij}^{(n)} = \sum_{\tau=1}^{n} p_{i_\tau j_\tau} \cdot \mathbb{I}_{X_{\tau+1}=j_\tau, X_{\tau+2}=j_\tau, \ldots, X_{\tau+n}=j_\tau} \]

\[ F_{ij} = \lim_{n \to \infty} F_{ij}^{(n)} \]

This lecture we focus on \( F_{ij} \) calculation from i to j (Transient states) to \( j \) (Recurrent states).

**Example (Cyclic):**

\[
\begin{bmatrix}
0 & 0.5 & 0.5 \\
0.1 & 0.2 & 0.7 \\
0 & 0 & 1
\end{bmatrix}
\]

\[
P_{ij} = \text{Prob}\{ X_n = j | X_0 = i \}
\]

**Classes:**

\[
\{1, 2\}, \{3, 4, 5\}, \{6, 7\}
\]

- Recurrent
- Recurrent
- Transient

**Goal:** \( F_{61} = ? \), \( F_{93} = ? \)

**Approach:** lump states of each class into a "big" state:

\[ A = \{1, 2\} \]

\[ B = \{3, 4, 5\} \]

\[ C \]
Write the new state transition matrix $P$ by defining transient states as the last states.

$$P = \begin{bmatrix}
1 & 0 \\
0 & 0
\end{bmatrix}$$

$$P = \begin{bmatrix}
0.1 & 0.3 & 0.6 & 0 \\
0.2 & 0.2 & 0.2 & 0.4 \\
0.1 & 0.5 & 0 & 0.3 \\
0 & 0 & 0.1 & 0
\end{bmatrix}$$

\[
\begin{pmatrix}
P^n
\end{pmatrix}
= \begin{bmatrix}
I & 0 \\
B_0
\end{bmatrix}
\Rightarrow B_n = (I + P + P^2 + \ldots + P^{n-1}) B_0
\]

$n$-step transition matrix

\[
F_{6A}^{(1)} = P_{6A}^5 \text{ reads } A \text{ for the first time at time } n \mid X_0 = A
\]

\[
F_{6A}^{(2)} = \begin{bmatrix}
0.1 & 0.9 \\
0.2 & 0.8
\end{bmatrix}
\]

\[
F_{6A} = \begin{bmatrix}
0.1 \\
0.2
\end{bmatrix}
\Rightarrow 0.2 = F_{6A}
\]

\[
P_2 X_1 \in \{X_0 = A\}
\]

\[
P_2 X_1 = \begin{bmatrix}
0.3 & 0.1 \\
0.2 & 0.8
\end{bmatrix}
\]

\[
F_{6A}^{(2)} = \begin{bmatrix}
0.1 & 0.9 \\
0.2 & 0.8
\end{bmatrix}
\Rightarrow 0.8 = F_{6A}^{(2)}
\]

\[
P_2 X_1 = \begin{bmatrix}
0.3 & 0.1 \\
0.2 & 0.8
\end{bmatrix}
\Rightarrow 0.8 = F_{6A}^{(2)}
\]

\[
P_2 X_1 \in \{X_0 = B\}
\]

\[
P_2 X_1 = \begin{bmatrix}
0.3 & 0.1 \\
0.2 & 0.8
\end{bmatrix}
\Rightarrow 0.8 = F_{6A}^{(2)}
\]

\[
P_2 X_1 \in \{X_0 = C\}
\]

\[
P_2 X_1 = \begin{bmatrix}
0.3 & 0.1 \\
0.2 & 0.8
\end{bmatrix}
\Rightarrow 0.8 = F_{6A}^{(2)}
\]

\[
P_2 X_1 \in \{X_0 = D\}
\]

\[
P_2 X_1 = \begin{bmatrix}
0.3 & 0.1 \\
0.2 & 0.8
\end{bmatrix}
\Rightarrow 0.8 = F_{6A}^{(2)}
\]
\[ F(n) = \left[ \begin{array}{cc} \Sigma F_{6A}(n) & \Sigma F_{6B}(n) \\ F_{7A}(n) & F_{7B}(n) \end{array} \right] = G \cdot B \]

\[ F_{6A} = \sum_{n=1}^{\infty} F_{6A}(n) \]

\[ F = \left[ \begin{array}{cc} F_{6A} & F_{6B} \\ F_{7A} & F_{7B} \end{array} \right] = \left[ \begin{array}{cc} \Sigma F_{6A}(n) & \Sigma F_{6B}(n) \\ \Sigma F_{7A}(n) & \Sigma F_{7B}(n) \end{array} \right] = \sum_{n=1}^{\infty} F_{6A}(n) \cdot \sum_{n=1}^{\infty} F_{6B}(n) = (\sum_{n=1}^{\infty} G_n) \cdot B \]

\[ \sum_{k=0}^{\infty} r^k = \frac{1}{1 - r}, \quad |r| < 1 \]

\[ \sum_{n=0}^{\infty} r^n = \frac{1}{1 - r} \]

\[ \sum_{n=0}^{\infty} G_n = S \]

\[ I + G + G^2 + \cdots = S \]

\[ I + B \left[ I + G + G^2 + \cdots \right] = S \]

\[ I + B \cdot S = S \]

\[ I - (I - B) \cdot S = S \]

\[ S = (I - B)^{-1} \cdot S \]

Then, for example, given:

\[ F = (I - G) \cdot B = \left[ \begin{array}{cc} 1.5 & 0.15 \\ 0.5 & 0.75 \end{array} \right] \cdot \left[ \begin{array}{cc} 0.1 & 0.5 \\ 0.2 & 0.2 \end{array} \right] = \left[ \begin{array}{cc} 0.21 & 0.16 \\ 0.14 & 0.39 \end{array} \right] \]
Expected Number of Steps Until Absorption

\[ \mathbb{E} \text{ steps until absorption } | X_0 = 6 \]

Event: $E_v$

After first transition (Step $X(1)$)

\[ \mathbb{E} \mathbb{E} | X_1 = 3, X_0 = 6 \ p \{ \mathbb{E} | X_1 = 3, X_0 = 6 \} \]

\[ \mathbb{E} \mathbb{E} | X_1 = 7, X_0 = 6 \ p \{ \mathbb{E} | X_1 = 7, X_0 = 6 \} \]

We know the transition probabilities

\[ X \]

A: $E_v = \mathbb{E}^2$ steps until absorption $| X_0 = 6$ for $t = 0$

B: What is the expected number of steps from $t = 0$ until absorption into either A or B given that $X_0$ is a transient state?
\[ R_6^{\text{abs}} = 1 + R_6 \cdot (0.3) + R_7^{\text{abs}} (0.1) \]

\[ R_7^{\text{abs}} = 1 + R_6 \cdot (0.2) + R_9^{\text{abs}} (0.4) \]

\[
\begin{bmatrix}
R_6^{\text{abs}} \\
R_7^{\text{abs}}
\end{bmatrix}
= \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.3 & 0.1 \\ 0.2 & 0.4 \end{bmatrix}
\begin{bmatrix}
R_6^{\text{abs}} \\
R_9^{\text{abs}}
\end{bmatrix}
\]

\[
(\mathbf{I} - \Theta) \begin{bmatrix}
R_6^{\text{abs}} \\
R_7^{\text{abs}}
\end{bmatrix}
= \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\]

\[
\Rightarrow
\begin{bmatrix}
R_6^{\text{abs}} \\
R_7^{\text{abs}}
\end{bmatrix}
= (\mathbf{I} - \Theta)^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}
= \begin{bmatrix} 4 \\ 2.25 \end{bmatrix}
\]
Steady-state Probabilities of Markov Chains

Assume that we have a Markov chain with \( N \) states.

Let

\[
p_0^T = [p_1, p_2, \ldots, p_N]^T
\]

\( P \) is the \( (N \times N) \) transition matrix:

\[
P = \begin{bmatrix}
p_{11} & p_{12} & \cdots & p_{1N} \\
p_{21} & p_{22} & \cdots & p_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
p_{N1} & p_{N2} & \cdots & p_{NN}
\end{bmatrix}
\]

So

\( p_0^T : 1 \times N \) vector (row vector) indicating prob. of each state at time \( 0 \).

Then

\[
p_1^T = p_0^T P
\]

\( \Rightarrow p_1^T = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \end{bmatrix} = p_0^T P
\]

Similarly,

\[
p_n^T = p_0^T P^n
\]

\( N \)-step prob. transition matrix.

\( p_0^T \) is an \( N \times 1 \) vector.

If I wait long enough, does \( p_n^T \) converge to something?

Assume it converges:

\[
\Pi = \lim_{N \to \infty} p_n^T = \lim_{N \to \infty} p_0^T P^n
\]

\( \Pi \) exists due to the assumption.

Then

\[
\Pi^T P = \Pi^T
\]

So, \( \Pi \) is an "invariant" distribution for 1-step transition matrix \( P \).
It is a left eigenvector of $\Pi$ with eigenvalue of 1.

Note: 1. $\Lambda e^\pi = \lambda e^\pi \implies e^\pi$ is a right eigenvector of $A$ with eigenvalue $\lambda e^\pi$.

      (Transpose)

      $e^\pi A^\top = A^\top e^\pi \implies$ then if $e^\pi$ is a right eigenvector of $A$

      then $e^\pi$ is a left eigenvector of $A^\top$

      with the same eigenvalue.

(2) Since $\det(A^\top - \lambda I) = \det((A - \lambda I)^\top) = \det(A - \lambda I^\top)$

the eigenvalues of $A$ and $A^\top$ are the same.

Q4: Do I have always a solution for

$\Pi^T \bar{p} = \Pi^T \pi$

An: $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

Then $\bar{p}$ is a right eigenvector with eigenvalue 1.

1-step transition matrix

(row sum = 1)

So, there's a left eigenvector with eigenvalue 0. From the notes 1 and 2

that means, we always have a solution for

$\Pi^T \bar{p} = \Pi^T$
Example: \[ \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.6 & 0 & 0.4 \\ 0 & 0.4 & 0.6 \end{bmatrix} \]

Classes = \{1, 2, 3\}, no transient classes

A recurrent class.

\[ P^N = P \times P^{N-1} \]

\[ \pi^T = \pi^T P = \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 \end{bmatrix} \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 \end{bmatrix}^T \]

\[ \pi_1 = 0.3 \pi_1 + 0.6 \pi_2 \\
\pi_2 = 0.5 \pi_1 + 0.4 \pi_3 \\
\pi_3 = 0.2 \pi_1 + 0.4 \pi_2 + 0.6 \pi_3 \]

Let's set \( \pi_1 = 1 \) \( \Rightarrow \) \[ \begin{align*}
\pi_2 &= 70 \\
\pi_3 &= 100
\end{align*} \]

\[ \pi^T [I - P] = 0 \]

\[ \pi \times \begin{bmatrix} 60 & 70 \\ 100 & 100 \end{bmatrix} \Rightarrow \pi = \begin{bmatrix} 60/123 \\ 100/123 \end{bmatrix} \]

\[ \pi = \begin{bmatrix} \frac{423}{123} \\ \frac{577}{123} \end{bmatrix} \]

Prob. dist. for \( X_N \) for large \( N \).

Q2: Is the steady-state prob. dist. unique? (In other words, can there be a multiplicity of eigenvalues of \( L \)?)

A2: The steady-state prob. dist. is unique for each recurrent class, if there are multiple recurrent classes, then there are multiple steady-state prob. dist. for each class.
By 3: Since $\Pi$ is independent from initial probability assignment at $n=0$.

and

$$\Pi = \Pi^0 \lim_{n \to \infty} P_n$$

and then

$$\Pi^\infty = \left[ \begin{array}{c} \Pi^T \\ \Pi^T \\ \vdots \\ \Pi^T \end{array} \right]$$

Srows of $\Pi^\infty$ are all $\Pi^T$ vectors.

Under what conditions is this result correct?

A 3: This result is correct for a finite-state Markov chain with a

dense recurrent class and no transient classes.

$$\Pi^\infty = \text{E} \Lambda^{-1}$$

$$E = [e_1 e_2 ... e_N]$$

$e_i$'s are eigenvectors with eigenvalue $\lambda_i$.

Eigenvalue decomposition

of $\Pi$ matrix

$$\Pi^n = E \Lambda^n E^{-1}$$

Fact: $P = [e_1 e_2 ... e_N] \begin{bmatrix} \Lambda & 0 \\ 0 & \Lambda \end{bmatrix} \begin{bmatrix} \Pi^T \\ \Pi^T \end{bmatrix}$

$e_i$ and $\Pi^T$ are right and left eigenvectors for the eigenvalue $\lambda_i$. 

(since left and right eigenvectors are related with \( P \) and \( P^T \) matrices)

In homework #4, it has been noted that eigenvalues of \( P \) matrix is less than or equal to \( 1 \) in magnitude.

\[ |\lambda_k| \leq 1 \quad \forall k \]

Then let \( e_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \) and \( \Pi_1 \) be the associated left eigenvector corresponding to the steady-state prob. distribution.

If there is only one recurrent class, then there is only one eigenvalue with the value \( 1 \); i.e., all others are less than \( 1 \) in magnitude.

\[
P^N = [e_1 \ldots e_m] \begin{bmatrix} \lambda_1^N & 0 & \cdots & 0 \\ 0 & \lambda_2^N & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m^N \end{bmatrix} \begin{bmatrix} \Pi_1^T \\ \Pi_2^T \\ \vdots \\ \Pi_m^T \end{bmatrix}
\]

\[
= \sum_{k=1}^{M} \lambda_k^N e_k \Pi_k^T = \left( e_1 \Pi_1^T + \sum_{k=2}^{M} \lambda_k e_k \Pi_k^T \right)
\]

\[
\lim_{N \to \infty} P^N = e_1 \Pi_1^T + 0 = \frac{1}{\Pi_1^T} \begin{bmatrix} \Pi_1^T \\ \Pi_2^T \\ \vdots \\ \Pi_m^T \end{bmatrix}
\]

\[
= \begin{bmatrix} \frac{1}{\Pi_1} \\ \frac{1}{\Pi_2} \\ \vdots \\ \frac{1}{\Pi_m} \end{bmatrix}
\]