Kalman Filtering

Let \( X, Y_1, Y_2, \ldots \) be random vectors with finite second moments.

**Notion of an innovation sequence**

Consider \( X(Y_1, \ldots, Y_n) = E[X | Y_1, Y_2, \ldots, Y_n] \)

We know that
\[
E[X] = E[X] + \text{Cov}(X, Y) \text{Cov}(Y, Y)^{-1} (Y - E[Y])
\]

Can matrix inversion be avoided?

**Special case of interest:**

The vectors \( Y_i \) are such that

\( E[Y_i] = 0 \) \( \forall i \), \( E[Y_i Y_j^T] = 0 \) \( \forall i \neq j \)

In this case
\[
E[X | Y_1, Y_2, \ldots, Y_n] = E[X] + \sum_{i=1}^{n} \hat{E}[X - E[X] | Y_i]
\]

Let us verify the orthogonality principle under \( X_{1}, \ldots, X_{n} \). Start with error

\[
e = X - E[X] - \sum_{i=1}^{n} \hat{E}[X - E[X] | Y_i]
\]

Need to show that \( E[e] = 0 \) and \( E[e Y_i^T] = 0 \) \( \forall i \)

First, note that \( \hat{E}[X - E[X] | Y_i] \) is linear in \( Y_i \)'s, and since \( E[X - E[X]] = 0 \)

we may write \( \hat{E}[X - E[X] | Y_i] = B_i Y_i \) for some matrix \( B_i \in \mathbb{R}^{m \times n} \)

Hence,

\[
E[e] = E[X - E[X]] - E\left[ \sum_{i=1}^{n} B_i Y_i \right] = 0 \quad \forall i
\]

\[
E[e Y_i^T] = E[X - E[X] - \sum_{i=1}^{n} \hat{E}[X - E[X] | Y_i] Y_i^T]
\]

\[
= E[X - \sum_{i=1}^{n} \hat{E}[X - E[X] | Y_i] Y_i^T] = 0
\]

\[
= E[X - E[X] Y_i^T] - \sum_{i=1}^{n} E[0 Y_i Y_i^T] = 0
\]
Problem is that we require orthogonality of $\gamma_i$. Think of Gram-Schmidt.

Solution: Orthogonalize the sequence of random vectors $Y_1, Y_2, \ldots, Y_n$.

Start with

$$1 \perp [Y_i - \bar{Y}] = \tilde{Y}_i$$

Take $Y_{12} = Y_k - \bar{Y}$ and $Y_{12} = Y_k - \bar{Y} = \text{Cov}(Y_k, Y_1^{k-1})^{-1}$

$$E[\tilde{Y}_2] = 0 \quad (\text{clear from above})$$

$$E[\tilde{Y}_k, \tilde{Y}_j] = 0 \quad \forall k, j$$

Easy to see

$$\tilde{Y}_2 = Y_2 - \bar{Y} = \text{Cov}(Y_2, Y_1) \text{Cov}(Y_1, Y_1)^{-1} (Y_1 - \bar{Y})$$

$$\tilde{Y}_n = Y_n - \bar{Y}$$

$$E[\tilde{Y}_2, Y_1] = \text{Cov}(Y_2, Y_1) - \text{Cov}(Y_2, Y_1) \text{Cov}(Y_1, Y_1)^{-1} \text{Cov}(Y_1, Y_1)$$

$$= \text{Cov}(Y_2, Y_1) - \text{Cov}(Y_2, Y_1) = 0$$

Induction also establishes that RV obtained through a linear transfer of $\tilde{Y}_k$ is the same as that obtained from a linear transfer of $Y_k$.

Hence,

$$E[X|Y_1, \ldots, Y_n] = E[X] + \sum_{i=1}^{n} E[X - E[X|Y_i]] \tilde{Y}_i$$

$$= E[X] + \sum_{i=1}^{n} \text{Cov}(X, Y_i) \text{Cov}(Y_i, Y_i)^{-1} \tilde{Y}_i$$

i.o.

$$\tilde{Y}_1 = Y_1 - E[Y]$$

$$\tilde{Y}_2 = Y_2 - E[Y_2] = \sum_{i=1}^{k} \text{Cov}(X, \tilde{Y}_i) \text{Cov}(\tilde{Y}_i, \tilde{Y}_i)^{-1} \tilde{Y}_i$$

$$\tilde{Y}_k$$ is called the innovation sequence.
Gram-Schmidt Orthogonalization

$v_1, v_2$ linearly independent, but not orthogonal

\begin{align*}
\text{normalized to have length 1} & \quad \text{normalized to have length 1} \\
\text{If we have n linearly independent vectors } v_1, v_2, \ldots, v_n & \\
\text{start with } \quad \text{proj}_u (v) = \frac{\langle u, v \rangle}{\langle u, u \rangle} u & \quad \text{and} \\
U_1 = v_1 & \quad e_1 = \frac{u_1}{\| u_1 \|} \\
U_2 = v_2 - \text{proj}_{U_1} (v_2) & \quad e_2 = \frac{u_2}{\| u_2 \|} \\
U_3 = v_3 - \text{proj}_{U_1} (v_3) - \text{proj}_{U_2} (v_3) & \quad e_3 = \frac{u_3}{\| u_3 \|} \\
\ldots & \quad \text{etc}
\end{align*}
Estimate state sequence given observation sequence

\[ x_{k+1} = F_k x_k + w_k, \quad k \geq 0 \]

\[ y_k = H_k x_k + v_k, \quad k \geq 0 \]

(Application: tracking, network dynamics, etc.)

\( x_0, y_0, \ldots, x_k, y_k, \ldots \) are pairwise uncorrelated

\( F_k, H_k \) known

\[ E[x_0] = \bar{x}_0, \quad \text{cov}(x_0) = P_0 \]

\[ E[w_k] = 0, \quad \text{cov}(w_k) = Q_k \]

\[ E[v_k] = 0, \quad \text{cov}(v_k) = R_k \]

\( P_0, Q_k, R_k \) known

\[ y_k \]

\[ \begin{array}{c}
  \text{Notation} \\
  \overline{x}_k = E[x_k], \quad P_k = \text{cov}(x_k) \\
  x_{k+1} = F_k \overline{x}_k \quad P_{k+1} = \text{cov}(F_k x_k + w_k) \\
  = F_k P_k F_k^T + Q_k
  \end{array} \]

Let \( y_k \) as before denote \((y_0, \ldots, y_k)\)

Let's write

\[ \hat{x}_{k|k} = E[x_k| y_k] \quad \text{covariance matrix of the error} \]

\[ \Sigma_{k|k} = \text{cov}(x_k - \hat{x}_{k|k}) \]

We have

\[ \hat{x}_{k|k} = \left[ F_k - K_k H_k \right] \hat{x}_{k-1} + K_k y_k \]

\[ \text{incorporate "innovation" from} \]

\[ \text{where} \]

\[ K_k = F_k \Sigma_{k|k-1} H_k \left( H_k \Sigma_{k|k-1} H_k + R_k \right)^{-1} \]

\[ \Sigma_{k|k} = F_k \Sigma_{k|k-1} F_k^T + R_k \]

Kalman gain

\[ \Sigma_{k|k} = \left( H_k \Sigma_{k|k-1} H_k + R_k \right)^{-1} \]
Initial conditions
\[ x_{01-1} = x_0, \quad z_{01-1} = z_0 \]
Steps: Decompose the influence of $y_{k-1}$, $y_k$ on $x_{k|k}$

Use innovation sequence

$$\tilde{y}_k = y_k - \mathbb{E}[y_k | y_{k-1}] \quad \text{instead of } y_k$$

Have $\tilde{y}_k = y_k - \mathbb{H} \tilde{x}_{k|k-1}$ and $x_{k|k} = \mathbb{E}[x_k | y_{k-1}, \tilde{y}_k] \ast$

will

$$x_{k|k} = \hat{x}_k + \sum_{i=1}^{k-1} \mathbb{E}[x_k - x_k | y_i]$$

Recall that $\mathbb{E}[x_k] = \mathbb{E}[x_k] + \text{Cov}(x_k, y_k) \text{Cov}(y_k)^{-1} (y_k - \mathbb{E}[y_k])$

$\ast$ becomes

$$x_{k|k} = x_{k|k-1} + \text{Cov}(x_k, \tilde{y}_k) \text{Cov}(\tilde{y}_k)^{-1} \tilde{y}_k$$

as $\mathbb{E}[\tilde{y}_k] = 0$

$$z_{k|k} = z_{k|k-1} - \text{Cov}(x_k, \tilde{y}_k) \text{Cov}(\tilde{y}_k)^{-1} \tilde{y}_k$$

Next, consider the time update

$$\hat{x}_{k+1|k} = \mathbb{E}[F_{k+1} x_k + w_k | y_k]$$

$$= F_{k+1} \mathbb{E}[x_k | y_k] + \mathbb{E}[w_k | y_k]$$

(2) $= F_{k+1} \hat{x}_{k|k}$

$$z_{k+1|k} = \text{Cov}(x_{k+1} - \hat{x}_{k+1|k})$$

$$= \text{Cov}(F_{k+1} x_k + w_k - \hat{x}_{k+1|k}) = \text{Cov}(F_{k+1} x_k - F_{k+1} \hat{x}_{k|k} + w_k)$$

$$= \text{Cov}(F_{k+1} (x_k - \hat{x}_{k|k}) + w_k) = F_{k+1} \text{Cov}(x_k - \hat{x}_{k|k}) + Q_k$$

Combining (1), (2) gives

$$\hat{x}_{k+1|k} = F_{k+1} \hat{x}_{k|k} + F_{k+1} \text{Cov}(x_k, \tilde{y}_k) \text{Cov}(\tilde{y}_k)^{-1} \tilde{y}_k$$

$$= F_{k+1} \hat{x}_{k|k} + F_{k+1} \text{Cov}(x_k, \tilde{y}_k) \text{Cov}(\tilde{y}_k)^{-1} (y_k - \mathbb{H} \hat{x}_{k|k-1})$$

If we set $k_x = F_{k+1} \text{Cov}(x_k, \tilde{y}_k) \text{Cov}(\tilde{y}_k)^{-1} (y_k - \mathbb{H} \hat{x}_{k|k-1})$

$$\text{Cov}(x_k, \tilde{y}_k) = \text{Cov}(x_k, \mathbb{H}^T (x_k - \hat{x}_{k|k-1}))$$

$$= \text{Cov}(x_k - \hat{x}_{k|k-1}, \mathbb{H}^T (x_k - \hat{x}_{k|k-1}))$$

$$= z_{k|k-1} \mathbb{H}^T$$
\[ \text{Cov}(\hat{y}_e) = \text{Cov}(H\bar{e}^T(y_e - \hat{x}_e) + \bar{e}) \]
\[ = \text{Cov}(H\bar{e}^T(x_e - \bar{x}_e) - \bar{e}) \]
\[ = H\bar{e}^T \Sigma_{e_1 e_1} H\bar{e} + \Sigma_{e_1 e_1} \]
How do we get from \((X_{k|k-1})\) to \((\hat{X}_{k+1|k})\)?

Step 1) Information update from \(y_k\)
Step 2) Time update

Instead of considering \(y_k\) as new information, we use \(\tilde{y}_k = y_k - \hat{E}_k\hat{y}_k\)

\(\tilde{y}_k, \tilde{y}_k, \ldots\) is the innovation sequence

It is easy to see that \(\hat{E}_k[y_k|y_{k-1}] = \hat{E}_k[\hat{X}_k]\hat{X}_{k|k-1}\)

as \(y_k, y_{k-1}\) are uncorrelated.
Hence \(\tilde{y}_k = y_k - \hat{X}_{k|k-1}\)

Furthermore, \(\hat{X}_{k|k} = \hat{E}_k[\hat{X}_k|\tilde{y}_k]\)

Now, recall that since the \(\tilde{y}_k\)'s are orthogonal,

\[\hat{E}_k[\tilde{y}_k, \ldots, \tilde{y}_n] = \sum_{i=1}^{n} \hat{E}_k[\tilde{y}_i]\]

\[\hat{X}_{k|k} = \hat{E}_k[\hat{X}_k|\tilde{y}_k] + \hat{E}_k[\hat{X}_k - \hat{X}_{k|k-1}][\hat{y}_k] + \hat{X}_{k|k-1}\]

\[= \hat{X}_{k|k-1} + \text{Cov}(\hat{X}_k, \tilde{y}_k)\text{Cov}(\tilde{y}_k)^{-1} \tilde{y}_k\]

and \(\tilde{X}_{k|k} = \tilde{X}_{k|k-1} - \text{Cov}(\tilde{X}_k, \tilde{y}_k)\text{Cov}(\tilde{y}_k)^{-1} \tilde{y}_k\)

Similarly, \(\hat{X}_{k+1|k} = \hat{E}_k[\hat{X}_{k+1}|\tilde{y}_k]\)

\[= F_k \hat{E}_k[\tilde{y}_k] + \hat{E}_k[w_k|\tilde{y}_k]\]

\[= F_k \hat{E}_k[\tilde{y}_k] + \hat{E}_k[w_k|\tilde{y}_k] + \hat{X}_{k|k-1}\]

Hence \(\tilde{X}_{k+1|k} = \text{Cov}(\hat{X}_{k+1} - \hat{X}_{k+1|k})\)

\[= \text{Cov}(F_k\tilde{X}_k + w_k - \hat{F}_k \hat{X}_{k|k})\]

\[= \text{Cov} (F_k (\hat{X}_k - \hat{X}_{k|k}) + w_k) = F_k \tilde{E}_k \tilde{F}_k^T + Q_k\]
Random Processes

\[ X = (X_t : t \in \mathbb{T}) \quad \text{RS on} \quad (\mathcal{X}, \mathcal{F}, \mathbb{P}) \]

\[ \mathbb{T} = \mathbb{Z} \quad \text{Discrete RP} \]

\[ \mathbb{T} = \mathbb{R} \quad \text{Continuous RP} \]

\[ X \text{ is a function on } \mathbb{T} \times \Omega \]

\[ X_t(\omega) \]

\[ \omega \text{ fixed} - X_t(\omega) \text{ function of } \omega \text{ only} \]

\[ t \text{ fixed} - X_t(\omega) \text{ is a function of } t \]

\[ = \text{Sample path (corresponding to sample } \omega) \]

\[ \mu_X(t) = \mathbb{E}[X_t] \quad \text{mean function} \]

\[ R_X(s,t) = \mathbb{E}[X_s X_t] \quad \text{correlation function (autocov)} \]

\[ P(X_t \leq x, \omega) = X_{t_n} \leq x_n \quad n \text{th order CDF} \]

Second order RP : \[ \mathbb{E}[X_t] < \infty \quad \forall t \in \mathbb{T} \]

Gaussian process : \[ X_t : t \in \mathbb{T} \] are jointly Gaussian

(need to know mean and autocorrelation function)

Some important examples

Random walks (gambler's ruin)

Let \[ W_i, i=1,2,3,... \text{ be iid}, P(W_i = 1) = p, P(W_i = -1) = 1-p \]

Define \[ X_n = X_0 + W_1 + ... + W_n, \] where \[ X_0 \] is a RU

independent on the \[ W_i \]'s

Clearly:

\[ \mathbb{E}[X_n] = n(2p-1) + \mathbb{E}[X_0] \]

\[ \text{var}(X_n) = \text{var}(X_0) + 4np(1-p) \]

\[ \lim_{n \to \infty} \frac{X_n}{n} = 2p-1 \quad \text{a.s., m.s.} \text{ for fixed value of } X_0 \]
Gamblers ruin: boundary conditions dictate termination at
index n when \( X_n = b \) or \( X_n = 0 \)

Initial wealth \( k \) (minimizes \( X_0 \))

Let's event that \( b \) is reached before \( 0 \)

\[ S_k = \text{Pr}(S_b) = \text{prob. reaching } b \text{ before } 0, \text{ given that starting capital is } k \]

\[ S_0 = 0, \quad S_b = 1; \quad 1 \leq k \leq b-1, \text{ we have} \]

\[ S_k = p \cdot W_2 = p \cdot S_k \cdot (S_0 \mid W_2 = 1) + \text{p} \cdot W_4 = \text{p} \cdot S_k \cdot (S_0 \mid W_4 = 1) \]

\[ = \text{p} \cdot S_{k+1} + (1-\text{p})S_{k-1} \quad 1 \leq k \leq b-1 \]

Let's take \( p = 1/2 \) first. Then

\[ S_k = \frac{1}{2} (S_{k+1} + S_{k-1}) \]

and clearly, \( S_k = a \cdot k + c \)

The boundary conditions give

\[ S_0 = ak + c = 0 \implies c = 0 \]

\[ S_b = ab + c = 1 \implies a = \frac{1}{b} \]

so that \( S_k = \frac{k}{b} \)

\[ p \neq 1/2, \text{ we seek a solution for the characteristic equation} \]

\[ \theta = p \cdot \theta^2 + (1-p) \cdot 1 \]

\[ p \cdot \theta^2 - \theta + (1-p) = 0 \]

\[ \theta_{1,2} = \frac{1 \pm \sqrt{1-4p} \cdot (1-p)}{2p} = \frac{1 \pm \sqrt{1-4p+4p^2}}{2p} = \frac{1 \pm (1-2p)}{2p} \]

\[ \frac{1-2p}{p} \]

so that \( S_k = a \cdot \left( \frac{1-p}{p} \right)^k + c \)

With the boundary conditions, get

\[ S_k = \frac{1 - \left( \frac{1-p}{p} \right)^k}{1 - \left( \frac{1-p}{p} \right)^b} \quad 0 \leq k \leq b \]
\( p > 1/2 \)

Note that \( S_0 \)'s are nested events

\[ S_1 \succ S_2 \succ \ldots \]

\[ S = \text{event that gambler's wealth converge to } \infty \text{ without bankruptcy} \]

\[ \lim_{b \to 0} \Pr(S_b) \]

\[ \Pr(S) = \lim_{b \to 0} \Pr(S_b) = \lim_{b \to 0} \Pr(S_b) = 1 - \left( \frac{1-p}{p} \right)^k \]

Processes with independent increments

\textbf{Markingales}

\( X = (X_t : t \in \mathbb{T}) \)

Increment over \([a, b]\) is \( X_b - X_a \)

Independent increments: any \( n \geq 0 \) integer and \( t_0 < t_1 \ldots < t_n \in \mathbb{T} \)

\( X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}} \) are

mutually independent

\( X = (X_t : t \in \mathbb{T}) \) is a martingale if \( \mathbb{E}[X_t] < \infty \) \( \forall t \)

and for any integer \( n > 0 \), \( t_1 < t_2 < \ldots < t_{n+1} \)

\[ \mathbb{E}[X_{t_{n+1}} | X_{t_1}, \ldots, X_{t_n}] = X_{t_n} \]

\[ \text{i.e. } \mathbb{E}[X_{t_{n+1}} - X_{t_n} | X_{t_1}, \ldots, X_{t_n}] = 0 \]

Claim: If \( X = (X_t : t \in \mathbb{T}) \) is an independent increment process with increments of mean zero, and \( X_0 = \text{const} \), then \( X \) is a martingale

Example: Gambler run will \( p = 1/2 \) is a martingale
Polya urn model: Iterative drawing of a marble

For any given colour, \( p \) of marbles in urn is a martingale

Say, \( r \) red balls and \( g \) green balls

Sample, add one more ball of the same color

\[
R_n = \# \text{ red balls at time } n \\
G_n = \# \text{ green balls at time } n
\]

1) \( R_n + G_n = r + g + n \)

2) \( H_n = \frac{R_n}{R_n + G_n} = \frac{R_n}{r + g + n} \)

\[
R_{n+3} = (R_{n+2}) I \left\{ \text{Until } U_{n+1} < H_n r \right\} + R_n I \left\{ \text{Until } U_{n+1} > H_n r \right\}
\]

\( U_1, U_2, \ldots \) sequence of uniform on \([0,1]\) iid RVs

\[
E \left[ H_{n+4} | H_0, H_1, ..., H_n \right] = \frac{R_{n+1}}{r + g + n + 1} E \left[ I \left\{ \text{Until } U_{n+1} < H_n r \right\} | H_0, ..., H_n \right] + \frac{R_n}{r + g + n + 1} E \left[ I \left\{ \text{Until } U_{n+1} > H_n r \right\} | H_0, ..., H_n \right]
\]

\[
= \frac{R_{n+1}}{r + g + n + 1} \left( \frac{R_n}{r + g + n} \right) + \frac{R_n}{r + g + n + 1} \left( \frac{G_n}{r + g + n} \right) = \frac{R_n}{r + g + n} \left[ \frac{R_{n+1}}{r + g + n + 1} + \frac{G_n}{r + g + n} \right] = \frac{R_n}{r + g + n} \cdot H_n
\]

The martingale is non-negative and bounded

Galton-Watson branching process (related to constructing family tree of European Royals)

Each generation contains a number of living members of the Royal family
At each unit of time, each member of the population produced offspring
with a distribution \( \{ p(n) : n \geq 0 \} \)

(sampled independently, everyone considered equally fit)

A normalized version of total # of royals is a Martingale
The Doob Martingale

A, Z1, Z2, ...RVs over the same probability space

\[ X_i = E[A | Z_1, ..., Z_i] \]

is called a Doob martingale

\[ E[X_i | Z_1, ..., Z_{i-1}] = E[E[X_i | Z_1, ..., Z_i] | Z_1, ..., Z_{i-1}] = E[A | Z_1, ..., Z_{i-1}] = X_{i-1} \]

Random graphs: edge exposure martingales, etc.

Azuma's inequality

Let \( \{X_i\} \) be a martingale sequence with respect to the filtration \( \mathcal{F}_i \)

Let \( Y_i = X_i - X_{i-1} \)

If \( \epsilon > 0 \) and \( |Y_i| \leq \epsilon \) \( \forall i \), then

\[ P \left( \frac{X_n - X_0}{\epsilon} \geq \frac{\epsilon}{\sqrt{n}} \right) \leq \exp \left( -\frac{\epsilon^2}{2 \sum \mathbb{E}[Y_i^2]} \right) \]

Analogue of Chernoff bound, but without the independence assumption

Simple Jensen's lemma

Let \( Y \in [-1, 1] \) be a RV s.t. \( E[Y] = 0 \). For any \( t > 0 \),

\[ E[e^{tY}] \leq e^{t^2/2} \]

Proof

\[ e^{tx} \leq \frac{1}{2} (1+x) e^t + \frac{1}{2} (1-x) e^{-t} \]

\[ -(ax_1 + (1-a)x_2) \leq a f(x_1) + (1-a) f(x_2) \]

Take \( x_1 = -1, x_2 = 1 \)

\[ a = \frac{1}{2} (1+x) \quad x_1 = 1 \]

\[ 1-a = \frac{1}{2} (1-x) \quad x_2 = -1 \]

\[ a - 4 + a = 2a - 1 = 2 + y - 1 = y \]
Proof

\[ E[e^{tY}] \leq E\left[ \frac{1}{2} (s^t Y) e^t + \frac{1}{2} (s^t Y) e^{-t} \right] \]

\[ = \frac{1}{2} e^t + \frac{1}{2} e^{-t} = \frac{1}{2} \left[ 1 + \sum_{k=1}^{\infty} \frac{t^k}{k!} + \ldots + 1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} \ldots \right] \]

\[ = 1 + \frac{t^2}{2} + \frac{t^4}{4!} \ldots = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!} \leq \sum_{n=0}^{\infty} \frac{(t/2)^n}{n!} = e^{t^2/2} \]

**Azuma's inequality**

For any \( t > 0 \),

\[ \Pr\{ X_n - X_0 > \lambda t \} \leq \Pr\{ e^{t(X_n - X_0)} > e^{\lambda t} \} \]

Applying Markov's inequality,

\[ \Pr\{ e^{t(X_n - X_0)} > e^{\lambda t} \} \leq E[e^{t(X_n - X_0)}] e^{-\lambda t} \]

\[ = e^{-\lambda t} E[e^{t(Y_n + X_{n-1} - X_0)}] = e^{-\lambda t} E[E[e^{t(Y_n + X_{n-1} - X_0)|S_{n-1}]]] \]

\[ = E[E[e^{t(Y_n + X_{n-1} - X_0)|S_{n-1}]]] = E[e^{t(X_{n-1} - X_0)} E[e^{tY_n|S_{n-1}}]] \]

Apply lemma above to \( \frac{Y_n}{c_n} \in [-1, 1] \)

\[ \leq e^{t(X_{n-1} - X_0)} e^{t^2c_n^2/2} \]

\[ E[e^{t(Y_n/c_n)}] \leq e^{t^2/2} \]

Replace \( t \) with \( c_n t \)

Hence,

\[ \Pr\{ X_n - X_0 > \lambda \} \leq e^{-\lambda t} e^{t^2c_n^2/2} E[e^{t(X_{n-1} - X_0)}] \]

Apply procedure inductively to get

\[ \Pr\{ X_n - X_0 > \lambda \} \leq e^{-\lambda t} e^{(t^2/2) \Sigma \alpha^2} \]

and optimize with \( t = \frac{\lambda}{\Sigma \alpha^2} \)

\( \square \)
Brownian motion (Wiener process)

\[ \sigma^2 > 0 \]
\[ W = (W_t : t \geq 0) \]

P1. \[ P(W_0 = 0) = 1 \]

P2. \( W \) has independent increments

P3. \( W_t - W_s \) has a \( \mathcal{N}(0, \sigma^2(t-s)) \) distribution for \( t \geq s \)

P4. \( P(W_t \text{ is continuous function of } t) = 1 \)

Clearly, Brownian motion is a martingale process.

Mean function

\[ \mu_W(t) = E[W_t] = E[W_t - W_0] = 0 \]

Correlation function: \( s \leq t \)

\[ R_W(s, t) = E[W_s W_t] = E[(W_s - W_0)(W_t - W_0)] \]
\[ = E[(W_s - W_0)(W_t - W_s + W_s - W_0)] \]
\[ = E[(W_s - W_0)(W_t - W_s)] + E[(W_s - W_0)(W_s - W_0)] = \sigma^2 s \]

or more generally,

\[ R_W(s, t) = \sigma^2 \min(s, t) \]

Brownian motion is a Gaussian process

P4 - P3 \( \Rightarrow \) \( W \) is Gaussian \( \mu_W = 0, R_W(s, t) = \sigma^2 \min(s, t) \)

Converse is also true

P4 is special
Counting processes

Counting function $\mathcal{f}$

$\mathcal{f}(0) = 0$

$\mathcal{f}$ nondecreasing

$\mathcal{f}$ right continuous

$\mathcal{f}$ is integer valued

\[ \mathcal{f}(b) - \mathcal{f}(a) = \text{# of times something happened in time interval } (a, b], \ b > a \]

Can also describe above via count times $\{t_i : i \geq 1\}$

or waiting times $\mathcal{w}_1 = t_0$, $\mathcal{w}_i = t_i - t_{i-1}$, $i \geq 2$

(intercount)

\[ \mathcal{f}(t) = \sum_{n=1}^{\infty} I(t \geq t_n) \]

\[ t_n = \min \{ t : \mathcal{f}(t) \geq n \} \]

\[ t_n = w_1 + \ldots + w_n \]

Poisson process

For $\lambda > 0$, we say that $N = (N_t : t \geq 0)$ is a Poisson RP with parameter $\lambda$ if

(P1). $N$ is a counting process

(P2). $N$ has independent increments

(P3). $N(t) - N(s)$ has a Poisson ($\lambda(t-s)$) distribution, $t > s$. 

Counting process:

a process s.t. w.p. 1 its sample path is a counting process
In the proof we used differences of two "martingale RVs". We say that \( X_1, X_2, X_3, \ldots \) is a martingale difference sequence if \( S_n = X_1 + \cdots + X_n \) \((S_0 = 0)\) is a martingale. Bennett's and Bernstein's inequalities for martingale difference sequences are proved similarly to Azuma's inequality. (Check it!)

**Brownian motion**

Random motion of a particle in a fluid or gas resulting from collisions with other fast moving particles in the substance. Einstein, 1905 (Nobel prize in Physics, 1921)

"Diffusion equation" has the solution

\[
P(x,t) = \frac{N}{\sqrt{4\piDt}} e^{-\frac{x^2}{4Dt}}
\]

- number of particles per unit volume, time \( t \)
- Normal, \( \mu = 0 \), var \( \sigma^2 = 2Dt \)

\( W = (W_t : t \geq 0) \)

1. \( P(W_0 = 0) = 1 \) (start from resting position)
2. \( W \) has independent increments
3. \( W_t - W_s \) has a \( \mathcal{N}(0, \sigma^2(t-s)) \) distribution, \( t > s \)
4. \( \mathbb{P}(W_t \text{ is continuous in } t^0) = 1 \)

\( W \) is a martingale!

**Mean function**

\[
\mu(W_t) = E[W_t] = E[W_t - W_0] = 0
\]

\( s \leq t \)

\[
\begin{align*}
R_W(s,t) &= E[W_s W_t] = E[(W_s - W_0)(W_s - W_t + W_t - W_s)] \\
&= E[(W_s - W_0)^2] = 2\sigma^2 s
\end{align*}
\]

or

\[
R_W(s,t) = 2\sigma^2 \min(s,t)
\]
Brownian motion is Gaussian
\[(W_{t_1}, \ldots, W_{t_n}) = \mathcal{L}(W_{t_1}, W_{t_2}, W_{t_3}, \ldots, W_{t_n})\]

Counting processes

Poisson Process

Counting function \( f: \) integer-valued \((\geq 0)\), nondecreasing, \( RC \)
\[0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq \ldots\]
\[\uparrow \quad \text{count times}\]
\[U_i = t_i - t_{i-1}\]
\[\uparrow \quad \text{intercount times}\]
\[f(t) = \sum_{i=1}^{n} \mathbb{I}(t \geq t_i)\]
\[t_n = \min \{ t: f(t) > n \}\]
\[t_n = U_1 + U_2 + \ldots + U_n\]

Counting process: \( RP \) for which sample path is a counting function \( w_p 1\)

Poisson process will rule a
\[N = (N_t: t \geq 0) \text{ s.t.}\]
\[P_1. \text{ } N \text{ is a counting process}\]
\[P_2. \text{ } N \text{ has independent increments}\]
\[P_3. \text{ } N(t) - N(s) \sim \text{ Pois}(\lambda(t-s)), \quad t > s\]

Claims

\( N \) Poisson iff the intercount times are iid \( \sim \text{Exp}(\lambda) \)

\[ \forall \epsilon > 0, \text{ } N_t \text{ is } \text{Pois}(\lambda \epsilon) \]

and given \( N_t = n \) have pdf of \( t_1, t_2, t_3, \ldots, t_n \)
\[f(t_1, t_2, \ldots, t_n \mid N_t = n) = \frac{n!}{t_1 ! \cdot t_2 ! \cdot \ldots \cdot t_n !} \cdot e^{-\lambda \epsilon}, \quad t_1 < t_2 < \ldots < t_n \]
\[0, \text{ otherwise}\]
Start with disjoint intervals
\((t_1-e, t_1], (t_2-e, t_2], \ldots, (t_n-e, t_n]\)

\((e\) is chosen small enough as needed to make the intervals disjoint\)

\[ P \left( \bigcap_{i=1}^{n} (t_i-e, t_i] \right) = \prod_{i=1}^{n} P(t_i-e, t_i] \]

\[ = \prod_{i=1}^{n} N(t_i-e) = 0,\ N(t_i-e) = 1 - e^{-\lambda(t_i-e)}, \text{for } i = 1, \ldots, n \]

\[ = e^{-\lambda(t_2-t_1)} e^{-\lambda(t_3-t_2)} e^{-\lambda(t_n-t_{n-1})} \]

\[ = (e\lambda)^n e^{-\lambda t_n} \quad (\text{since we used } e^{n-x} = e^{n-x}) \]

Hence

\[ \sum_{i=t_1}^{t_n} T_i(t_1, \ldots, t_n) = \begin{cases} \sum_{i=t_1}^{t_n} e^{-\lambda t_i}, & 0 < t_1 < \cdots < t_n \\ 0, & \text{otherwise} \end{cases} \]

Put \( U_i = t_i, \ U_i = t_{i+1} - t_i \)

Hence, we have the Jacobian

\[ \frac{\partial u}{\partial t} = \begin{bmatrix} \frac{\partial u}{\partial t_1} \\ \vdots \\ \frac{\partial u}{\partial t_n} \end{bmatrix} = \begin{bmatrix} e^{-\lambda t_1} \\ \vdots \\ e^{-\lambda t_n} \end{bmatrix} \]

\[ = \begin{bmatrix} \sum_{i=t_1}^{t_n} U_i(t_1, \ldots, t_n) = \begin{cases} \sum_{i=t_1}^{t_n} e^{-\lambda t_i}, & 0 < t_1 < \cdots < t_n \\ 0, & \text{otherwise} \end{cases} \end{bmatrix}, \ U_i \in \mathbb{R}^+\]

\( U_i \)'s are iid, \( \sim \text{Exp}(\lambda) \)

Second pair of equivalence follows from observing that

\[ \{ N(t_1) = n \} \equiv \left\{ (t_1, t_2, \ldots, t_n) \in A_{n, 2} \right\} \]

\[ A_{n, 2} = \left\{ t \in \mathbb{R}^{n+1} : 0 < t_1 < \cdots < t_n < t \leq 2 \right\} \]

Integrating over \( t \leq 2 \)

\[ \frac{\lambda^n e^{-\lambda t}}{\lambda^n} = \begin{cases} \lambda^n e^{-\lambda t}, & 0 < t_1 < \cdots < t_n < 2 \\ 0, & \text{else} \end{cases} \]

\[ \frac{\lambda^n e^{-\lambda t}}{\lambda^n} \text{ is a const.} \]

\[ \frac{\lambda^n e^{-\lambda t}}{\lambda^n} = \frac{1}{\text{vol}(\mathbb{R}^n)} \quad \text{Vol}(\mathbb{R}^n) = \frac{2^n}{n!} \]
\[ \frac{n^ne^{-n^2}}{n!} = e^{-1} \]

and since \[ P(N_n = ny) = e^{-ny} \]

the second pair of equivalences follows.

Poisson process is not a martingale.

Has same mean and covariance function as Brownian motion with \[ \sigma^2 = \lambda \]

**Stationary RPs** (recall we only deal with second order processes)

\[ X = (X_t; t \in \mathbb{T}) \]

is said to be stationary if for any \( n \), any \( t_1, t_2, \ldots, t_n \), and any \( s \in \mathbb{T} \)

\[ (X_{t_1}, \ldots, X_{t_n}) \text{ and } (X_{t_1+s}, \ldots, X_{t_n+s}) \]

have the same distribution

(Shift invariance of statistics)

**Consequences:**

For \( n=1 \), all \( X_t \)'s have the same distribution and \( \mu_X(t) \), \( E[X_t^2] \) do not depend on \( t \)

\( n=2 \) \( E[X_{t+s}, R_{X}(t_1+s, t_2+s) \) do not depend on \( s \)

**Wide-sense stationarity**

\[ \mu_X(t) = \mu_X(t+s) \forall s \]

\[ R_X(t_1, t_2) = R_X(t_1+s, t_2+s) \forall s \]

or

\[ E[X_{t+s}] = \mu_X, \ E[X_{t+s}X_{t+s}] = R_X(t_1-t_2) \]

If Gaussian process is WSS, then it is stationary.
Examples (Discrete time)

Let \( X_n = Z_{n-1} + Z_n \quad \forall n \geq 1 \) (Moving average process)

\( Z_0, Z_1, \ldots, Z_n, \ldots \) are i.i.d. \( \sim N(0,1) \)

Find the mean and autocorrelation function

Is the process stationary? WSS?

\[
\begin{align*}
E[X_n] &= \frac{1}{2} E[Z_{n-1}] + E[Z_n] = 0 \\
R_x(m, n) &= E[X_m X_n] \\
&= E[(Z_{m-1} + Z_m)(Z_{n-1} + Z_n)] \\
&= E[Z_{m-1} Z_n + \text{terms}] \\
&= \begin{cases} 
1, & |n-m| = 1 \\
2, & n = n \\
0, & \text{unless } n = m \text{ or } |n-m| = 1
\end{cases}
\end{align*}
\]

Hence, process is WSS. But \( Z_i \) are \( \sim N(0,1) \)

hence process is Gaussian, and stationary

Repeat for autoregressive process

\( X_n = \frac{1}{2} X_{n-1} + Z_n, \quad X_0 = 0, \quad \forall n \geq 1 \)

\( Z \) i.i.d. \( \sim N(0,1) \)

Result:

\[
\begin{align*}
R_x(m, n) &= 2^{-|m-n|} E[X_m] \\
E[X_m^2] &= \frac{4}{3} \left( 1 - \frac{1}{4^n} \right)
\end{align*}
\]