9. Important random processes

definitions, properties, and applications

- **Random walk:** genetic drifts, slowly-varying parameters, neuron firing
- **Gaussian:** popularly used by its tractability
- **Wiener/Brownian:** movement dynamics of particles
- **White noise:** widely used by its independence property
- **Markov:** population dynamics, market trends, page-rank algorithm
- **Poisson:** number of phone calls in a varying interval
- **ARMAX:** time series model in finance, engineering

Bernoulli random process

a (time) sequence of indepenent Bernoulli RV is an iid Bernoulli RP

example:

- $I[n]$ is an indicator function of the event at time n where $I[n] = 1$ when success and $I[n] = 0$ when fail
- let $D[n] = 2I[n] - 1$ and it is called **random step** process

$$D[n] = 1 \text{ or } -1$$

$D[n]$ can represent the *deviation* of a particle movement along a line

Sum process

the sum of a sequence of iid random variables, $X_1, X_2, \ldots$

$$S[n] = X_1 + X_2 + \cdots + X_n, \quad n = 1, 2, \ldots$$

where $S[0] = 0$, is called the **sum process**

- we can write $S[n] = S[n-1] + X_n$ (recursively)
- the sum process has *independent increments* in nonoverlapping intervals

  $$S[n]-S[n-1] = X_n, \quad S[n-1]-S[n-2] = X_{n-1}, \ldots, S[2]-S[1] = X_2$$

  (since X_k's are iid)
- the sum process has *stationary increments*

$$P(S[n] - S[k] = y) = P(S[n-k] = y), \quad n > k$$

autocovariance of the sum process:

- assume X_k's have mean m and variance σ^2
- $\mathbf{E}[S[n]] = nm$ (X_k's are iid)
- $\mathbf{var}[S[n]] = n\sigma^2$ (X_k's are iid)

we can show that

$$C(n,k) = \min(n,k)\sigma^2$$

the proof follows from letting $n \leq k$, and so $n = \min(n,k)$

$$\begin{aligned}
C(n,k) &= \mathbf{E}[(S[n]-nm)(S[k]-km)] \\
&= \mathbf{E}[(S[n]-nm)\{(S[n]-nm)+(S[k]-km)-(S[n]-nm)\}] \\
&= \mathbf{E}[(S[n]-nm)^2] + \mathbf{E}[(S[n]-nm)(S[k]-S[n]-(k-n)m)] \\
&= \mathbf{E}[(S[n]-nm)^2] + \mathbf{E}[(S[n]-nm)]\mathbf{E}[(S[k]-S[n]-(k-n)m)]
\end{aligned}$$

(apply that $S[n]$ has independent increments and $\mathbf{E}[S[n]-nm]=0$)

more properties of a sum process

- the joint pdf/pmf of $S(1), \ldots, S(n)$ is given by the product of pdf of $S(1)$ and the marginals of individual *increments*
  - X_k's are integer-valued
  - X_k's are continuous-valued
- the sum process is a Markov process (more on this)

Binomial counting process

let $I[n]$ be iid Bernoulli random process

the sum process $S[n]$ of $I[n]$ is then the **counting process**

- it gives the number successses in the first n Bernoulli trial
- the counting process is an increasing function
- $S[n]$ is binomial with parameter p (probability of success)

Random walk

let $D[n]$ be iid random step process where

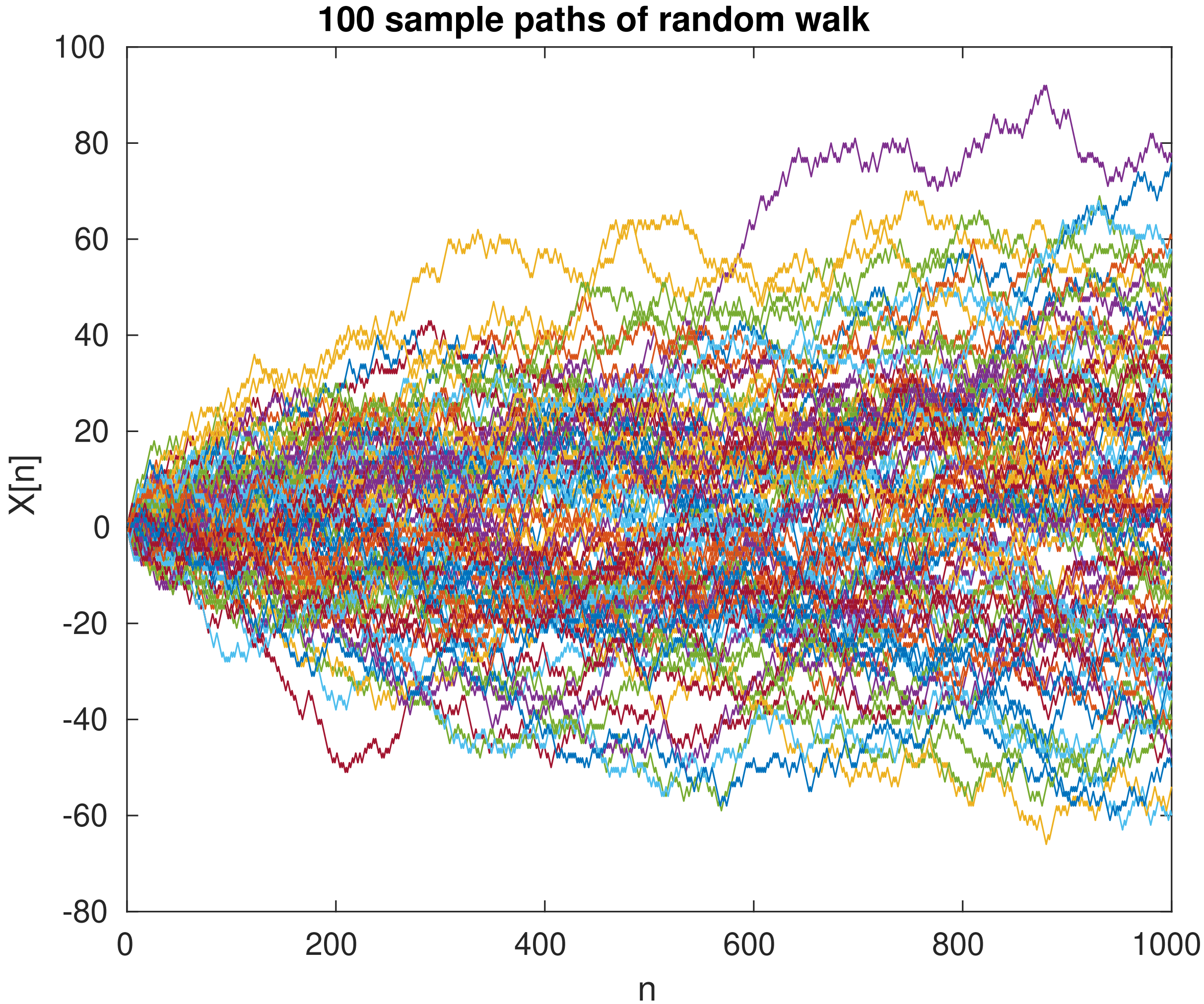
$$D[n] = \begin{cases} 1, & \text{with probability } p \\ -1, & \text{with probability } p \end{cases}$$

the random walk process $X[n]$ is defined by

$$X[0] = 0, \quad X[n] = \sum_{k=1}^{n} D[k], \quad k \geq 1$$

- the random walk is a sum process
- we can show that $\mathbf{E}[X[n]] = n(2p-1)$
- the random walk has a tendency to either grow if $p > 1/2$ or to decrease if $p < 1/2$

a random walk example as the sum of Bernoulli sequences with $p = 1/2$



$\mathbf{E}[X(n)] = 0$ and $\mathbf{var}[X(n)] = n$ (variance grows over time)

properties:

- $X[n]$ has *independent stationary increments* in nonoverlapping time intervals
$$P[X[m] - X[n] = y] = P[X[m-n] = y]$$
(increments in intervals of the same length have the same distribution)

- a random walk is related to an **autoregressive process** since
$$X[n+1] = X[n] + D[n+1]$$
(widely used to model financial time series, biological signals, etc)
$$\text{stock price:} \quad \log X[n+1] = \log X[n] + \beta D[n+1]$$

- extension: if $D[n]$ is a Gaussian process, we say $X[n]$ is a **Gaussian random walk**

Gaussian process

an RP $X(t)$ is a **Gaussian** process if the samples

$$X_1 = X(t_1), X_2 = X(t_2), \quad X_k = X(t_k)$$

are jointly Gaussian RV for all k and all choices of $t_1, \ldots, t_k$

that is the joint pdf of samples from time instants is given by

$$f_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-(1/2)(x-m)^T \Sigma^{-1}(x-m)}$$

$$m = \begin{bmatrix} m(t_1) \\ m(t_2) \\ \vdots \\ m(t_k) \end{bmatrix}, \Sigma = \begin{bmatrix} C(t_1,t_1) & C(t_1,t_2) & \cdots & C(t_1,t_k) \\ C(t_2,t_1) & C(t_2,t_2) & \cdots & C(t_2,t_k) \\ \vdots & \vdots & & \vdots \\ C(t_k,t_1) & \cdots & & C(t_k,t_k) \end{bmatrix}$$

properties:

- Gaussian RPs are specified completely by the mean and covariance functions
- Gaussian RPs can be both continuous-time and discrete-time
- linear operations on Gaussian RPs preserve Gaussian properties

example: let $X(t)$ be a zero-mean Gaussian RP with

$$C(t_1, t_2) = 4e^{-3|t_1 - t_2|}$$

find the joint pdf of $X(t)$ and $X(t+s)$
we see that

$$C(t, t+s) = 4e^{-3s}, \quad \mathbf{var}[X(t)] = C(t,t) = 4$$

therefore, the joint of pdf of $X(t)$ and $X(t+s)$ is the Gaussian distribution parametrized by

$$f_{X(t),X(t+s)}(x_1, x_2) = \frac{1}{(2\pi)|\Sigma|^{1/2}} e^{-(1/2)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}$$

where

$$\Sigma = \begin{bmatrix} 4 & 4e^{-3s} \\ 4e^{-3s} & 4 \end{bmatrix}$$

example: let $X(t)$ be a Gaussian RP and let $Y(t) = X(t+d) - X(t)$

- mean of $Y(t)$ is

$$m_y(t) = \mathbf{E}[Y(t)] = m_x(t+d) - m_x(t)$$

- the autocorrelation of $Y(t)$ is

$$\begin{aligned} R_y(t_1,t_2) &= \mathbf{E}[(X(t_1+d) - X(t_1))(X(t_2+d) - X(t_2))] \\ &= R_x(t_1+d, t_2+d) - R_x(t_1+d, t_2) - R_x(t_1, t_2+d) + R_x(t_1, t_2) \end{aligned}$$

- the autocovariance of $Y(t)$ is then

$$\begin{aligned} C_y(t_1,t_2) &= \mathbf{E}[(X(t_1+d) - X(t_1) - m_y(t_1))(X(t_2+d) - X(t_2) - m_y(t_2))] \\ &= C_x(t_1+d, t_2+d) - C_x(t_1+d, t_2) - C_x(t_1, t_2+d) + C_x(t_1, t_2) \end{aligned}$$

since $Y(t)$ is the sum of two Gaussians then $Y(t)$ must be Gaussian

- any k-time samples of $Y(t)$

$$Y(t_1), Y(t_2), \ldots, Y(t_k)$$

is linear transformation of jointly Gaussians, so $Y(t_1), \ldots, Y(t_k)$ have jointly Gaussian pdf

- for example, find joint pdf of $Y(t)$ and $Y(t+s)$: need only mean and covariance
  - $m_y(t)$ and $m_y(t+s)$
  - covariance is given by

$$\Sigma = \begin{bmatrix} C_y(t,t) & C_y(t,t+s) \\ C_y(t,t+s) & C_y(t+s,t+s) \end{bmatrix}$$

Wiener process

consider the random step on page 9-2

symmetric walk ($p = 1/2$), magnitude step of M, time step of h seconds

let $X_h(t)$ be the accumulated sum of random step up to time t

- $X_h(t) = M(D[1] + D[2] + \cdots + D[n]) = MS[n]$ where $n = [t/h]$
- $\mathbf{E}[X_h(t)] = 0$
- $\mathbf{var}[X_h(t)] = M^2 n$

Wiener process $X(t)$**:** obtained from $X_h(t)$ by shrinking the magnitude and time step to zero in a *precise way*

$$h \to 0, \quad M \to 0, \quad \text{with } M = \sqrt{\alpha h} \text{ where } \alpha > 0 \text{ is constant}$$

(meaning; if $v = M/h$ represents a particle speed then $v \to \infty$ as displacement M goes to 0) properties of Wiener (also called Wiener-Levy) process:

- $\mathbf{E}[X(t)] = 0$ (zero mean of all time)
- $\mathbf{var}[X(t)] = (\sqrt{\alpha h})^2 \cdot (t/h) = \alpha t$ (stays finite and nonzero)
- $X(t) = \lim_{h \to 0} M(D[1] + \cdots + D[n]) = \lim_{n \to \infty} \sqrt{\alpha t} \frac{S[n]}{\sqrt{n}}$ approaching the sum of an *infinite* number of RV
- by CLT, pdf $X(t)$ approaches Gaussian with mean zero and variance αt

$$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-\frac{x^2}{2\alpha t}}$$

- $X(t)$ has independent stationary increments (from random walk form)
- Wiener process is a Gaussian random process ($X(t_k)$ is obtained as linear transformation of increments)
- Wiener process is used to model **Brownian motion** (movement of particles in fluid)

- the covariance function of Wiener process is

$$C(t_1, t_2) = \alpha \min(t_1, t_2), \quad \alpha > 0$$

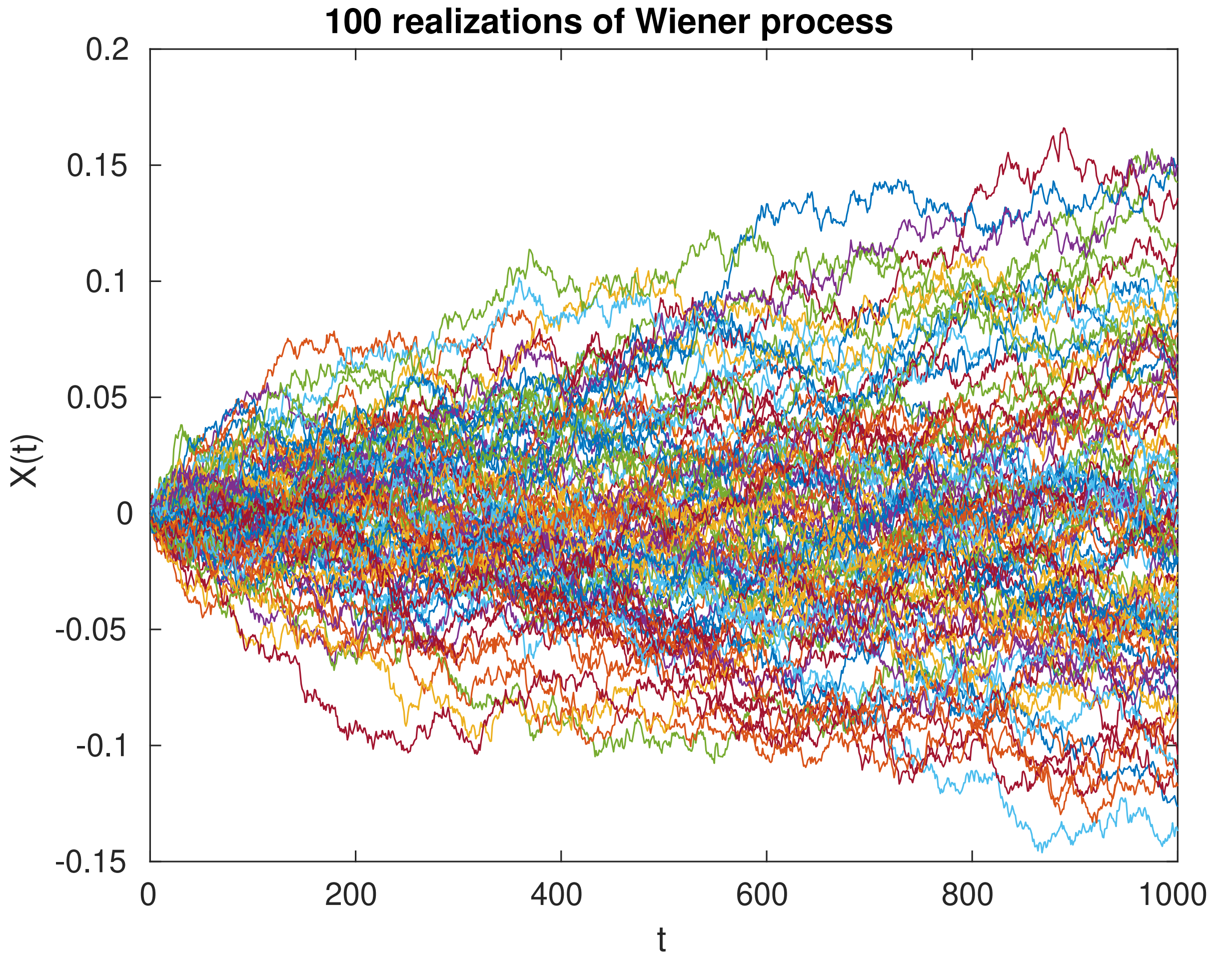
to show this, let $t_1 \geq t_2$,

$$\begin{aligned} C(t_1, t_2) &= \mathbf{E}[X(t_1)X(t_2)] = \mathbf{E}[(X(t_1) - X(t_2) + X(t_2))X(t_2)] \\ &= \mathbf{E}[(X(t_1) - X(t_2)X(t_2)] + \mathbf{var}[X(t_2)] \\ &= 0 + \alpha t_2 \end{aligned}$$

using $X(t_1) - X(t_2)$ and $X(t_2)$ are independent (when $t_1 \geq t_2$)

if $t_2 < t_1$, we do the same and obtain $C(t_1, t_2) = \alpha t_1$

sample paths of Wiener process when $\alpha = 2$



$\mathbf{E}[X(t)] = 0$ and $\mathbf{var}[X(t)] = \alpha t$ (variance grows over time)

White noise process

a random process $X(t)$ is white noise if

- $\mathbf{E}[X(t)] = 0$ (zero mean for all t)
- $\mathbf{E}[X(t)X(s)] = 0$ for $t \neq s$ (uncorrelated with another time sample)

in another word,

- the correlation function of a white noise is the *impulse function*

$$R(t_1, t_2) = \alpha\delta(t_1 - t_2), \quad \alpha > 0$$

- power spectral density is flat (more on this): $S(\omega) = \alpha, \quad \forall\omega$
- $X(t)$ has infinite power, varies extremely rapidly in time, and is most unpredictable

those two properties of white noise are derived from the definition that

white Gaussian noise process is the *time derivative* of Wiener process

recall the correlation of wiener process is $R_{\mathrm{wiener}}(t_1, t_2) = \alpha \min(t_1, t_2)$

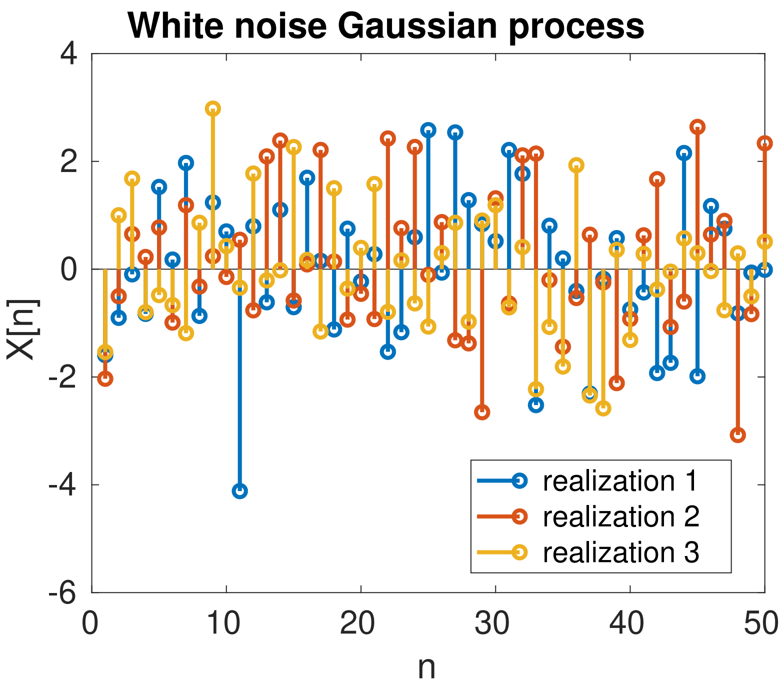
$$
\begin{aligned}
R(t_1, t_2) = \mathbf{E}[X(t_1)X(t_2)] &= \mathbf{E}\left[\frac{\partial}{\partial t_1} X_{\mathrm{wiener}}(t_1) \cdot \frac{\partial}{\partial t_2} X_{\mathrm{wiener}}(t_2)\right] \\
&= \frac{\partial}{\partial t_1}\frac{\partial}{\partial t_2} R_{\mathrm{wiener}}(t_1, t_2) = \frac{\partial}{\partial t_1}\frac{\partial}{\partial t_2} \begin{cases} \alpha t_2, & t_2 < t_1 \\ \alpha t_1, & t_2 \geq t_1 \end{cases} \\
&= \frac{\partial}{\partial t_1} \alpha u(t_1 - t_2), \quad u \text{ is the step function}
\end{aligned}
$$

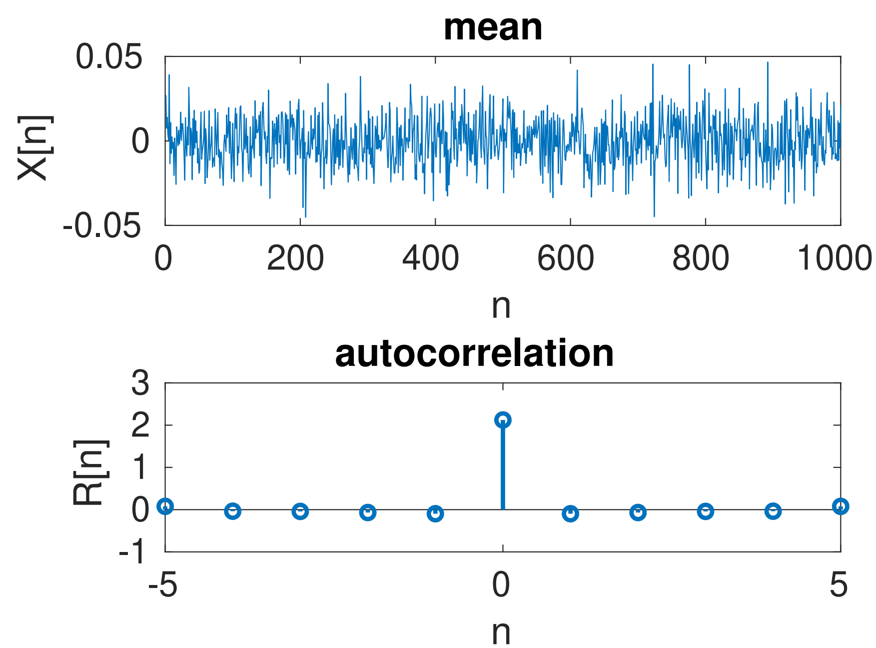
but u is not differentiable at $t_1 = t_2$, so the second derivative does not exist

instead, we generalize this notion using delta function

$$R(t_1, t_2) = \alpha \delta(t_1 - t_2)$$

example of white noite Gaussian process with variance 2





- mean function (averaged over 10,000 realizations) is close to zero
- sample autocorrelation is close to a delta function where $R(0) \approx 2$

Poisson process

let $N(t)$ be the number of event occurrences in the interval $[0, t]$

properties:

- non-decreasing function (of time t)
- integer-valued and continuous-time RP

assumptions:

- events occur at an average rate of λ events per seconds
- the interval $[0, t]$ is divided into n subintervals and let $h = t/n$
- the probability of *more than one event* occurrences in a subinterval is negligible compared to the probability of observing one or zero events
- whether or not an event occurs in a subinterval is independent of the outcomes in other subintervals

meaning:

- the outcome in each subinterval can be viewed as a Bernoulli trial
- these Bernoulli trials are independent
- $N(t)$ can be *approximated* by the binomial counting process

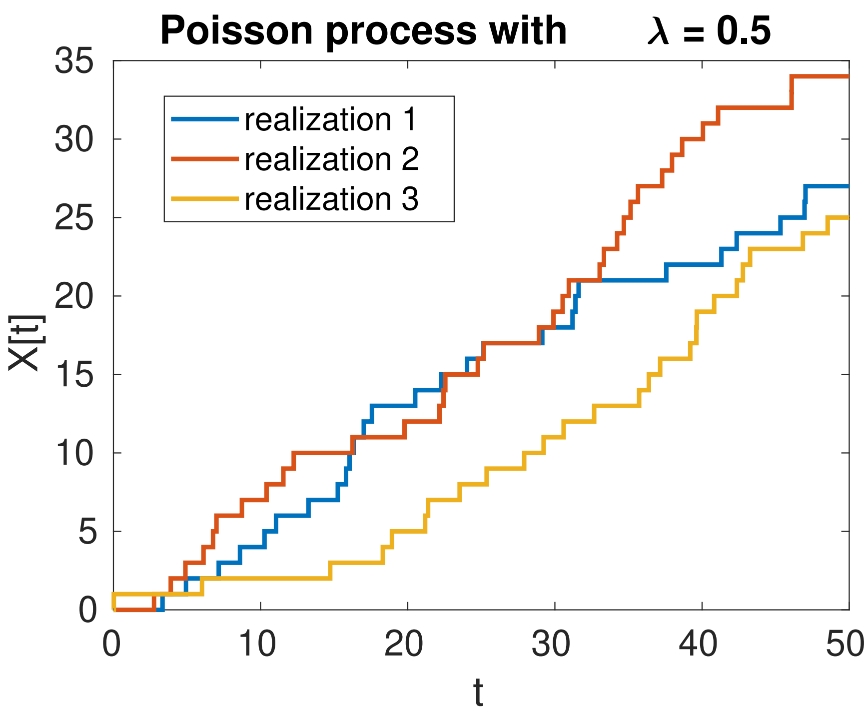
Binomial counting process:

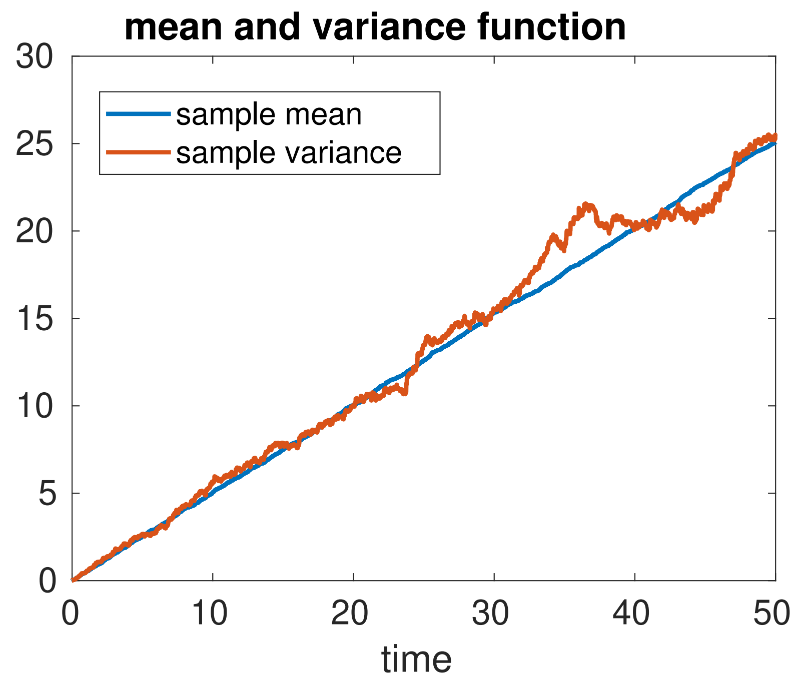
- let the probability of an event occurrence in subinterval is p
- average number of events in $[0, t]$ is $\lambda t = np$
- let $n \to \infty$ ($h = t/n \to 0$) and $p \to 0$ while $np = \lambda t$ is fixed
- from the following approximation when n is large

$$P(N(t) = k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k} \approx \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \ldots$$

$N(t)$ has a Poisson distribution and is called a **Poisson process** example of Poisson process $\lambda = 0.5$

- generated by taking cumulative sum of n-sequence Bernoulli with and $p = \lambda T/n$ where $n = 1000$ and $T = 50$





- the rate of Poisson process grows as λt for $t \in [0, T]$
- the mean and variance functions (approximate over 100 realizations) have linear trend over time

joint pmf: for $t_1 < t_2$,

$$\begin{aligned} P[N(t_1) = i, N(t_2) = j] &= P[N(t_1) = i]P[N(t_2) - N(t_1) = j - i] \\ &= P[N(t_1) = i]P[N(t_2 - t_1) = j - i] \\ &= \frac{(\lambda t_1)^i e^{-\lambda t_1}}{i!} \frac{(\lambda(t_2 - t_1))^j e^{-\lambda(t_2 - t_1)}}{(j - i)!} \end{aligned}$$

autocovariance: $C(t_1, t_2) = \lambda \min(t_1, t_2)$

for $t_1 \leq t_2$,

$$\begin{aligned} C(t_1, t_2) &= \mathbf{E}[(N(t_1) - \lambda t_1)(N(t_2) - \lambda t_2)] \\ &= \mathbf{E}[(N(t_1) - \lambda t_1)\{N(t_2) - N(t_1) - \lambda t_2 + \lambda t_1 + (N(t_1) - \lambda t_1)\}] \\ &= \mathbf{E}[(N(t_1) - \lambda t_1)]\mathbf{E}[(N(t_2) - N(t_1) - \lambda(t_2 - t_1)] + \mathbf{var}[N(t_1)] \\ &= \lambda t_1 \end{aligned}$$

we have used *independent and stationary increments* property examples:

- random telegraph signal
- the number of car accidents at a site or in an area
- the requests for individuals documents on a web server
- the number of customers arriving at a store

Time between events in Poisson process

let T be the time between event occurrences in a Poisson process

- the probability involving T follows

$$
\begin{aligned}
P[T > t] &= P[\text{no events in } t \text{ seconds}] = (1-p)^n \\
&= \left(1 - \frac{\lambda t}{n}\right)^n \to e^{-\lambda t}, \quad \text{as } n \to \infty
\end{aligned}
$$

  T is an exponential RV with parameter λ

- the interarrival time in the underlying binomial proces are independent geometric RV
- the sequence of interarrival times $T[n]$ in a Poisson process form an iid sequence of *exponential* RVs with mean $1/\lambda$
- the sum $S[n] = T[1] + \cdots + T[n]$ has Erland distribution

Markov process

for any time instants, $t_1 < t_2 < \cdots < t_k < t_{k+1}$, if

discrete-valued

$$P[X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k, \ldots, X(t_1) = x_1] = P[X(t_{k+1} = x_{k+1} \mid X(t_k) = x_k]$$

continuous-valued

$$f(x_{k+1} \mid X(t_k) = x_k, \ldots, X(t_1) = x_1) = f(x_{k+1} \mid X(t_k) = x_k)$$

then we say $X(t)$ is a **Markov** process

joint pdf conditioned on several time instants reduce to pdf conditioned on the *most recent* time instant

properties:

- pmf and pdf of Markov processes are conditioned on several time instants can reduce to pmf/pdf that is only conditioned on the *most recent* time instant
- an integer-valued Markov process is called a **Markov chain** (more details on this)
- the sum of iid sequence where $S[0] = 0$ is a Markov process
- a Poisson process is a continuous-time Markov process
- a Wiener process is a continous-valued Markov process
- in fact, any *independent-increment* process is also Markov

to see this, for a discrete-valued RP,

$$\begin{aligned}
&P[X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k, \ldots, X(t_1) = x_1] \\
&\quad = P[X(t_{k+1}) - X(t_k) = x_{k+1} - x_k \mid X(t_k) = x_k, \ldots, X(t_1) = x_1] \\
&\quad = P[X(t_{k+1}) - X(t_k) = x_{k+1} - x_k \mid X(t_k) = x_k] \quad \text{by independent increments} \\
&\quad = P[X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k]
\end{aligned}$$

more examples of Markov process

- birth-death Markov chains: transitions only between adjacent states are allowed

$$p(t+1) = Pp(t), \quad P \text{ is tri-diagonal}$$

- M/M/1 queue (a queuing model): continuous-time Markov chain

$$\dot{p}(t) = Qp(t)$$

Discrete-time Markov chain

a Markov chain is a random sequence that has n possible states:

$$X(t) \in \{1, 2, \dots, n\}$$

with the property that

$$\mathbf{prob}(\ X(t+1) = i \ \mid \ X(t) = j\ ) = p_{ij}$$

where $P = [p_{ij}] \in \mathbf{R}^{n \times n}$

- p_{ij} is the **transition probability** from state j to state i
- P is called the **transition matrix** of the Markov chain
- the state $X(t)$ still cannot be determined with *certainty*
- $\{1, 2, \dots, n\}$ is called *label* (simply mapped to integers)

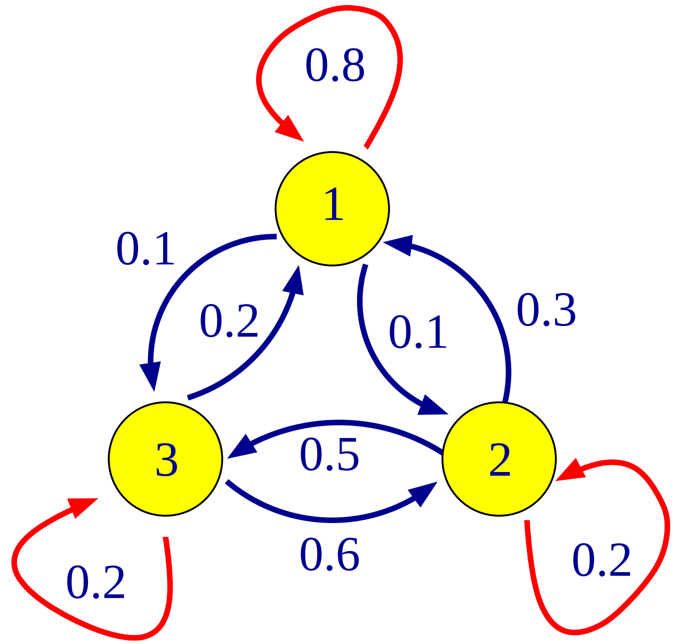
example:

a customer may rent a car from any of three locations and return to any of the three locations

Rented from location

1	2	3	
0.8	0.3	0.2	**1**
0.1	0.2	0.6	**2**
0.1	0.5	0.2	**3**

Returned to location



Properties of transition matrix

let P be the transition matrix of a Markov chain

- all entries of P are real *nonnegative* numbers
- the entries in any column are summed to 1 or $\mathbf{1}^T P = \mathbf{1}^T$:

$$p_{1j} + p_{2j} + \cdots + p_{nj} = 1$$

  (a property of a **stochastic matrix**)

- 1 is an eigenvalue of P
- if q is an eigenvector of P corresponding to eigenvalue 1, then

$$P^k q = q, \quad \text{for any } k = 0, 1, 2, \ldots$$

Probability vector

we can represent probability distribution of $x(t)$ as n-vector

$$p(t) = \begin{bmatrix} \mathbf{prob}(\ x(t) = 1\ ) \\ \vdots \\ \mathbf{prob}(\ x(t) = n\ ) \end{bmatrix}$$

- $p(t)$ is called a **state probability vector** at time t
- $\sum_{i=1}^{n} p_i(t) = 1$ or $\mathbf{1}^T p(t) = 1$
- the state probability propagates like a linear system:

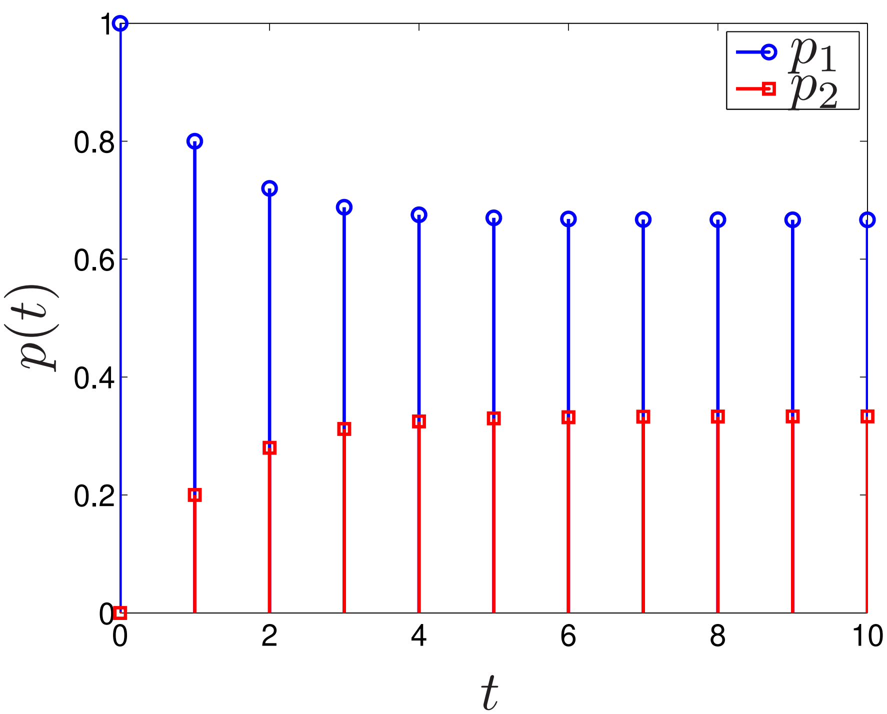
$$p(t+1) = Pp(t)$$

- the state PMF at time t is obtained by multiplying the initial PMF by P^t

$$p(t) = P^t p(0), \quad \text{for } t = 0, 1, \ldots$$

example: a Markov model for packet speech

- two states of packet speech: contain 'silent activity' or 'speech activity'
- the transition matrix is $P = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix}$
- the initial state probability is $p(0) = (1, 0)$
- the packet in the first state is 'silent' with certainty



- eigenvalues of P are 1 and 0.4
- calculate P^t by using 'diagonalization' or 'Cayley-Hamilton theorem'

$$P^t = \begin{bmatrix} (5/3)(0.4 + 0.2 \cdot 0.4^t) & (2/3)(1 - 0.4^t) \\ (1/3)(1 - 0.4^t) & (5/3)(0.2 + 0.4^{t+1}) \end{bmatrix}$$

- $P^t \rightarrow \begin{bmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{bmatrix}$ as $t \rightarrow \infty$ (all columns are the same in limit!)
- $\lim_{t\rightarrow\infty} p(t) = \begin{bmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} p_1(0) \\ 1 - p_1(0) \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$

$p(t)$ does not depend on the *initial state probability* as $t \rightarrow \infty$

what if $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$?

- we can see that

$$P^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad P^3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \dots$$

- P^t does not converge but oscillates between two values

under what condition $p(t)$ converges to a constant vector as $t \to \infty$?

Definition: a transition matrix is **regular** if some integer power of it has all *positive* entries

Fact: if P is regular and let w be *any* probability vector, then

$$\lim_{t \to \infty} P^t w = q$$

where q is a **fixed** probability vector, independent of t

Steady state probabilities

we are interested in the **steady state probability vector**

$$q = \lim_{t\to\infty} p(t) \qquad \text{(if converges)}$$

- the steady-state vector q of a regular transition matrix P satisfies

$$\lim_{t\to\infty} p(t+1) = P \lim_{t\to\infty} p(t) \qquad \Longrightarrow \qquad Pq = q$$

  (in other words, q is an eigenvector of P corresponding to eigenvalue 1)

- if we start with $p(0) = q$ then

$$p(t) = P^t p(0) = 1^t q = q, \quad \textbf{for all } t$$

  q is also called the **stationary state PMF** of the Markov chain

example: weather model ('rainy' or 'sunny')

probabilities of weather conditions given the weather on the preceding day:

$$P = \begin{bmatrix} 0.4 & 0.2 \\ 0.6 & 0.8 \end{bmatrix}$$

(probability that it will rain tomorrow given today is sunny, is 0.2)

given today is sunny with probability 1, calculate the probability of a rainy day in long term

Gauss-Markov process

let $W[n]$ be a white Gaussian noise process with $W[1] \sim \mathcal{N}(0, \sigma^2)$

definition: a Gauss-Markov process is a first-order autoregressive process

$$X[1] = W[1], \quad X[n] = aX[n-1] + W[n], \quad n \geq 1, \quad |a| < 1$$

- clearly, $X[n]$ is Markov since the state $X[n]$ only depends on $X[n-1]$
- $X[n]$ is Gaussian because if we let

$$X_k = X[k], \quad W_k = W[k], \quad k = 1, 2, \ldots, n \quad \text{(time instants)}$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n-1} \\ X_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ a & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a^{n-2} & a^{n-3} & \vdots & 1 & 0 \\ a^{n-1} & a^{n-2} & \cdots & a & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_{n-1} \\ W_n \end{bmatrix}$$

  pdf of $(X_1, \ldots, X_n)$ is Gaussian for all n

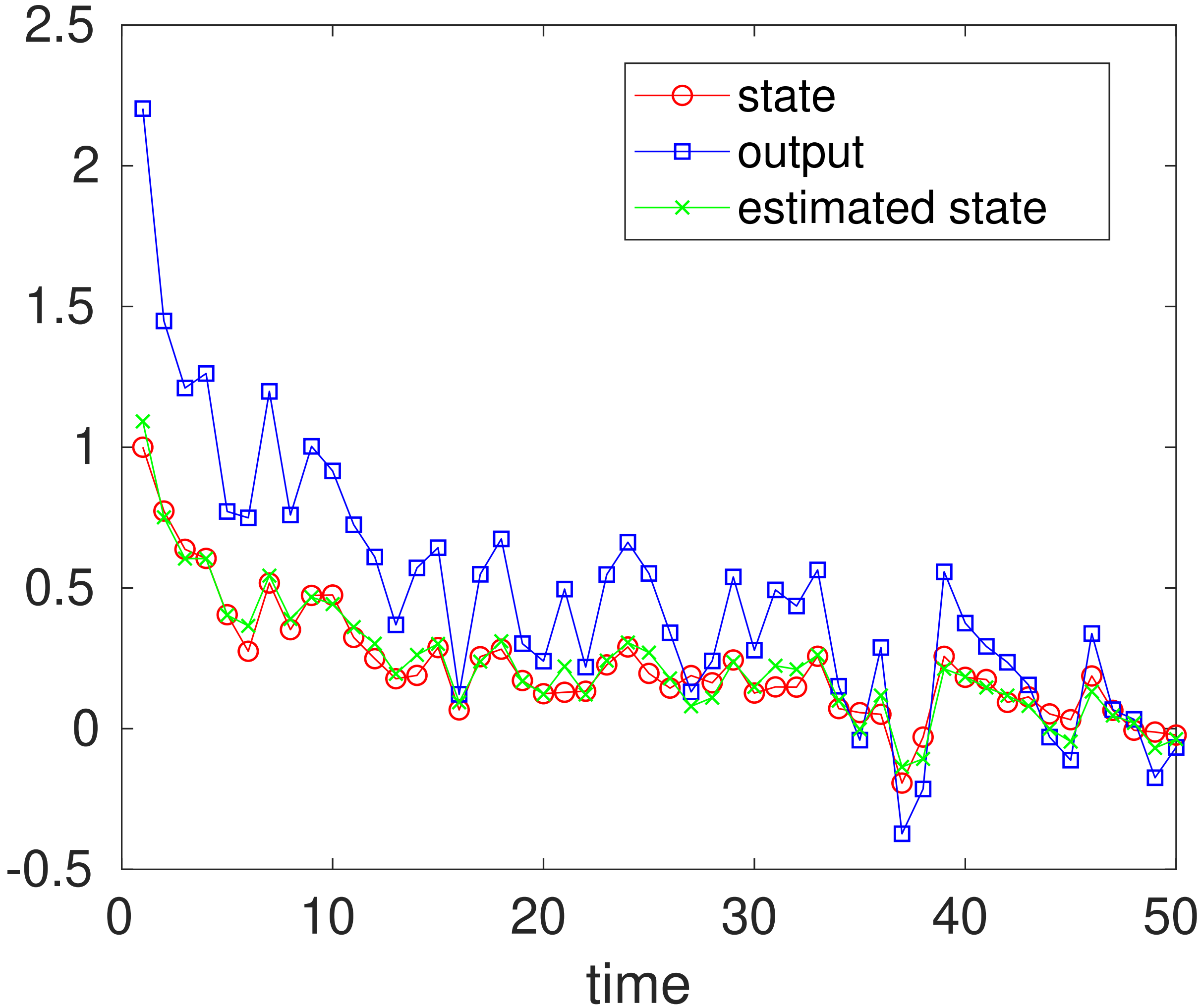
questions involving a Gauss-Markov process

setting:

- we can observe $Y[n] = X[n] + V[n]$ where V represents a sensor noise
- only Y can be observed, but we do not know X

question: can we estimate $X[n]$ from information of $Y[n]$ and statistical properties of W and V?

solution: yes we can. one choice is to apply a Kalman filter example: $a = 0.8$, $Y[k] = 2X[k] + V[k]$



$X[k]$ is estimated by Kalman filter

References

Chapter 9 in
A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd edition, Pearson Prentice Hall, 2009

Chapter 9 in
H. Stark and J. W. Woods, *Probability, Statistics, and Random Processes for Engineers*, 4th edition, Pearson, 2012