In the name of GOD.

Stochastic Process

Fall 2024

Sharif University of Technology

Hamid R. Rabiee

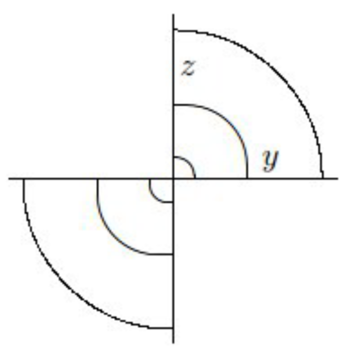
Homework 3 | Point Process, Poisson Process, Gaussian Process | Solution

1. Let X and Z be IID normalized Gaussian random variables. Let $Y = |Z|\text{Sgn}(X)$, where $\text{Sgn}(X)$ is 1 if $X \geq 0$ and -1 otherwise. Show that X and Y are each Gaussian, but are not jointly Gaussian. Sketch the contours of equal joint probability density.

   Solution:
   Note that Y has the magnitude of Z but the sign of X, so that X and Y are either both positive or both negative, i.e., their joint density is nonzero only in the first and third quadrant of the X, Y plane. Conditional on a given X, the conditional density of Y is twice the conditional density of Z since both Z and $-Z$ are mapped into the same Y. Thus

$$f_{XY}(x, y) = \left(\frac{1}{\pi}\right) \exp\left(-\frac{x^2 + y^2}{2}\right)$$

   for all x, y in the first or third quadrant.



2. A radioactive source emits particles according to a Poisson process of rate 2 particles per minute.

   (a) Compute the probability p_a that the first particle appears some time after 3 minutes and before 5 minutes.

   (b) Compute the probability p_b that exactly one particle is emitted in the time interval from 3 to 5 minutes.

   Solution:

   (a) Recall that the time intervals $T_1, T_2, \ldots$ for the jumps of the Poisson process are independent identically distributed exponential random variables of rate $\lambda = 2$. To say that the first particle appears some time after 3 minutes and before 5 minutes is the same as to say that $3 < T_1 < 5$. Hence

$$p_a = \mathbb{P}(3 < T_1 < 5) = \int_3^5 2e^{-2t}dt = -e^{-2t}\Big|_3^5 = e^{-6} - e^{-10}.$$

(b) For p_b, we ask also that there are no other particles arriving in the interval $[3, 5]$, i.e. that $T_1 + T_2 > 5$. Hence

$$p_b = \mathbb{P}(3 < T_1 < 5, T_1+T_2 > 5) = \int_3^5 f_{T_1}(t)\mathbb{P}(T_2 > 5-t)dt = \int_3^5 2e^{-2t}e^{-2(5-t)}dt = \int_3^5 2e^{-10}dt = 4e^{-10}.$$

3. Given a normal process $x(t)$ with $\eta_x = 0$ and $R_x(\tau) = 4e^{-2|\tau|}$, we form the random variables $z = x(t + 1)$, $w = x(t - 1)$:

   (a) Find $E[zw]$ and $E[(z + w)^2]$.

   (b) Find $f_z(z)$, $P\{z < 1\}$, and $f_{zw}(z, w)$.

Solution:
(a)

$$E\{zw\} = R_x(2) = 4e^{-4},$$
$$E\{z^2\} = E\{w^2\} = R_x(0) = 4,$$
$$E\{(z + w)^2\} = R_x(0) + R_x(0) + 2R_x(2) = 8(1 + e^{-4}).$$

(b)

- z is normally distributed: $z \sim \mathcal{N}(0, 4)$, so its probability density function is:

$$f_z(z) = \frac{1}{\sqrt{8\pi}} \exp\left(-\frac{z^2}{8}\right).$$

- The probability $P\{z < 1\}$ is:

$$P\{z < 1\} = \Phi\left(\frac{1}{2}\right) \approx 0.6915.$$

- The joint probability density function $f_{zw}(z, w)$ is bivariate normal:

$$f_{zw}(z, w) = \frac{1}{2\pi \cdot 2 \cdot 2\sqrt{1 - (e^{-4})^2}} \exp\left(-\frac{1}{2(1 - e^{-8})}\left[\frac{z^2}{4} + \frac{w^2}{4} - 2e^{-4}\frac{zw}{4}\right]\right),$$

  where $z \sim \mathcal{N}(0, 4)$, $w \sim \mathcal{N}(0, 4)$, and their correlation coefficient is e^{-4}.

4. In one of the ancient cities, there was a traditional restaurant that served as the main gathering place for the city's residents. Every morning, all the people lined up in front of the restaurant and entered one by one. Once inside, they chose a table and stayed there until the end of the day. One of the favorite pastimes of the residents of this city is choosing their table at random. Now, suppose Nth person in line wants to enter while $N - 1$ people are already seated . This person

has two options: can sit at one of the tables that already have people, with the probability of choosing table k (which currently has n_k people seated) given by $\frac{n_k}{N-1+\alpha}$, or alternatively, can choose a new table with a probability of $\frac{\alpha}{N-1+\alpha}$, where n_k is the number of people already seated at table k, and α is a fixed constant. Considering this seating process, if the total population of the city is M, and α is equal to 1, determine the average number of tables occupied in one day. Express the answer in terms of H_n ($H_n = \sum_{i=1}^{n} \frac{1}{i}$).

Solution:

n_t number of tables at time t, $Z_t = n_t - n_{t-1}$

$$P(Z_t = k) = \begin{cases} \frac{\alpha}{t-1+\alpha} & \text{if } k = 1 \\ \frac{t-1}{t-1+\alpha} & \text{if } k = 0 \end{cases}$$

$$n_t = 1 + \sum_{\tau=2}^{t} Z_\tau \quad \Rightarrow \quad \mathbb{E}[n_t] = 1 + \sum_{\tau=2}^{t} \mathbb{E}[Z_\tau]$$

$$\mathbb{E}[n_t] = 1 + \sum_{\tau=2}^{t} \frac{\alpha}{\tau - 1 + \alpha} = H_t$$

5. (a) Let $X_1 \sim N(0, \sigma_1^2)$ and $X_2 \sim N(0, \sigma_2^2)$ be independent random variables. Show that $X_1 + X_2$ follows the distribution $N(0, \sigma_1^2 + \sigma_2^2)$.
   (b) Let W_1, W_2 be i.i.d. normalized Gaussian random variables. Show that $a_1 W_1 + a_2 W_2$ is Gaussian, $\mathcal{N}(0, a_1^2 + a_2^2)$.
   (c) Using the result from part (b), to show that all linear combinations of i.i.d. normalized Gaussian random variables are Gaussian.

Solution:

(a) Sum of Independent Normal Variables

Let $Z = X_1 + X_2$. Since X_1 and X_2 are independent, the density of Z is the convolution of the X_1 and X_2 densities. For simplicity, assume $\sigma_{X_1}^2 = \sigma_{X_2}^2 = 1$.

$$\begin{aligned} f_Z(z) = f_{X_1}(z) * f_{X_2}(z) &= \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z - x) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-(z-x)^2/2} dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(x^2+z^2-2zx)/2} dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(x-z/2)^2 - z^2/4} dx \\ &= \frac{1}{\sqrt{\pi}} e^{-z^2/4} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-(x-z/2)^2} dx \\ &= \frac{1}{\sqrt{\pi}} e^{-z^2/4}, \end{aligned}$$

since the last integral evaluates to 1 (as it represents the integral of a Gaussian pdf with mean $z/2$ and variance $1/2$). Thus, Z is Gaussian with zero mean and variance 2.

The trick used here is called completing the square. For the term $x^2 + \alpha z x + \beta z^2$, add and subtract $\frac{\alpha^2 z^2}{4}$, resulting in:

$$x^2 + \alpha z x + \frac{\alpha^2 z^2}{4} = \left(x + \frac{\alpha z}{2}\right)^2.$$

This transformation allows us to integrate in Gaussian form.

Repeating the same steps for arbitrary $\sigma_{X_1}^2$ and $\sigma_{X_2}^2$, we get the Gaussian density with mean 0 and variance $\sigma_{X_1}^2 + \sigma_{X_2}^2$.

(b) Linear Combination of Independent Gaussians

You could repeat all the steps from (a), but an insightful approach is to let:

$$X_i = a_i W_i \quad \text{for } i = 1, 2.$$

Since W_1 and W_2 are i.i.d. standard Gaussian random variables, the variance of X_i is:

$$\text{Var}(X_i) = a_i^2 \quad \text{for } i = 1, 2.$$

Hence, the sum $a_1 W_1 + a_2 W_2$ is Gaussian with variance:

$$\text{Var}(a_1 W_1 + a_2 W_2) = a_1^2 + a_2^2.$$

Thus, $a_1 W_1 + a_2 W_2 \sim \mathcal{N}(0, a_1^2 + a_2^2)$.

(c) Induction for General Linear Combination of Gaussians

We now prove by induction that any linear combination of i.i.d. Gaussian random variables is Gaussian.

The inductive hypothesis is that for a sequence $\{W_i; i \geq 1\}$ of i.i.d. normal random variables, and a sequence of constants $\{\alpha_i; i \geq 1\}$, the sum:

$$\sum_{i=1}^{n} \alpha_i W_i \sim \mathcal{N}\left(0, \sum_{i=1}^{n} \alpha_i^2\right)$$

for some $n \geq 1$.

Base Case: For $n = 2$, from part (b), we have:

$$a_1 W_1 + a_2 W_2 \sim \mathcal{N}(0, a_1^2 + a_2^2).$$

Inductive Step: Assume the hypothesis is true for $n = k$, i.e.:

$$\sum_{i=1}^{k} \alpha_i W_i \sim \mathcal{N}\left(0, \sum_{i=1}^{k} \alpha_i^2\right).$$

Now consider the sum for $n = k + 1$:

$$X = \sum_{i=1}^{k} \alpha_i W_i.$$

By the inductive hypothesis, $X \sim \mathcal{N}(0, \sum_{i=1}^{k} \alpha_i^2)$. Since W_{k+1} is independent of X, and $W_{k+1} \sim \mathcal{N}(0, 1)$, we have:

$$X + \alpha_{k+1} W_{k+1} \sim \mathcal{N}\left(0, \sum_{i=1}^{k+1} \alpha_i^2\right).$$

Thus, by the principle of induction, for all $n \geq 1$, any linear combination of i.i.d. Gaussian random variables is Gaussian.

6. Earthquakes occur in a given region in accordance with a Poisson process with rate 5 per year.

   (a) What is the probability that there will be at least two earthquakes in the first half of 2020?

   (b) Assuming that the event in part (a) occurs, what is the probability that there will be no earthquakes during the first 9 months of 2021?

   (c) Assuming that the event in part (a) occurs, what is the probability that there will be at least four earthquakes over the first 9 months of the year 2020?

Solution:

(a) The rate of earthquakes per year is given as 5. For the first half of 2020 (i.e., 6 months), the rate is:

$$\lambda_{6 \text{ months}} = \frac{5}{2} = 2.5 \text{ earthquakes per 6 months.}$$

The number of earthquakes, X, follows a Poisson distribution with parameter $\lambda = 2.5$:

$$X \sim \text{Poisson}(2.5).$$

We are interested in finding the probability that there are at least 2 earthquakes:

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1).$$

$$P(X = 0) = \frac{e^{-2.5}(2.5)^0}{0!} = e^{-2.5},$$

$$P(X = 1) = \frac{e^{-2.5}(2.5)^1}{1!} = 2.5e^{-2.5}.$$

Now:

$$P(X \geq 2) = 1 - e^{-2.5} - 2.5e^{-2.5}.$$

$$P(X \geq 2) \approx 1 - 0.0821 - 0.2052 = 0.7127.$$

Thus, the probability that there will be at least 2 earthquakes in the first half of 2020 is approximately:

$$\boxed{0.7127}.$$

(b) The earthquakes in 2021 will be independent of those of 2020.

Given: $\lambda = 5$ per year, assuming occurrence to be homogeneous, we can assume $\lambda = 3.75$ for 9 months.

$$P(X=0) = \frac{e^{-3.75}3.75^0}{0!} = \boxed{0.0235}$$

(c) In the first half of 2020 since the event in part (a) occurs, 2, 3, or more than 4 earthquakes will occur. Hence, we have 3 cases:

Case I: 2 earthquakes in the first half of 2020 Hence, at least 2 earthquakes need to occur in the next 3 months which are independent of the previous 6 months. Hence,

$$P(\text{case I}) = \frac{P(X=2 \text{ in the first half}) \times P(X \geq 2 \text{ in the next quarter})}{P(X \geq 2 \text{ in the first half})}$$

$$= \frac{\frac{e^{-2.5}2.5^2}{2!} \times \left(1 - \frac{e^{-1.25}1.25^0}{0!} - \frac{e^{-1.25}1.25^1}{1!}\right)}{0.7127} = 0.128$$

Case II: 3 earthquakes in the first half of 2020 Hence, at least 1 earthquake needs to occur in the next 3 months which are independent of the previous 6 months. Hence,

$$P(\text{case II}) = \frac{P(X=3 \text{ in the first half}) \times P(X \geq 1 \text{ in the next quarter})}{P(X \geq 2 \text{ in the first half})}$$

$$= \frac{\frac{e^{-2.5}2.5^3}{3!} \times \left(1 - \frac{e^{-1.25}1.25^0}{0!}\right)}{0.7127} = 0.214$$

Case III: 4 or more earthquakes in the first half of 2020 This is pretty straightforward.

$$P(\text{case III}) = \frac{P(X \geq 4 \text{ in the first half})}{P(X \geq 2 \text{ in the first half})}$$

$$= \frac{1 - \frac{e^{-2.5}2.5^0}{0!} - \frac{e^{-2.5}2.5^1}{1!} - \frac{e^{-2.5}2.5^2}{2!} - \frac{e^{-2.5}2.5^3}{3!}}{0.7127} = 0.340$$

Hence, the total required probability will be:

$$P(X \geq 4 \text{ in the first 9 months} \mid X \geq 2 \text{ in the first 6 months}) = \boxed{0.682}$$

7. A stochastic process $\{X(t), t \geq 0\}$ is said to be stationary if $X(t_1), \ldots, X(t_n)$ has the same joint distribution as $X(t_1 + a), \ldots, X(t_n + a)$ for all $n, a, t_1, \ldots, t_n$.

(a) Prove that a necessary and sufficient condition for a Gaussian process to be stationary is that $\text{Cov}(X(s), X(t))$ depends only on $t - s$, $s \leq t$, and $\mathbb{E}[X(t)] = c$.

(b) Let $\{X(t), t \geq 0\}$ be Brownian motion and define

$$V(t) = e^{-\alpha t/2} X(a e^{\alpha t}).$$

Show that $\{V(t), t \geq 0\}$ is a stationary Gaussian process. It is called the Ornstein-Uhlenbeck process. (**Hint:** Brownian motion is a Gaussian process W_t such that $\mathbb{E}[W_t] = 0$ and $\text{cov}(W_t, W_s) = \min(t, s)$.)

Solution:

(a) If the Gaussian process is stationary, then for $t > s$:

$$\begin{pmatrix} X(t) \\ X(s) \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} X(t-s) \\ X(0) \end{pmatrix}$$

Thus $\mathbb{E}[X(s)] = \mathbb{E}[X(0)]$ for all s and $\text{Cov}(X(t), X(s)) = \text{Cov}(X(t-s), X(0))$ for all $t < s$.

Now, assume $\mathbb{E}[X(t)] = c$ and $\text{Cov}(X(t), X(s)) = h(t-s)$. For any $T = (t_1, \ldots, t_k)$, define vector $X_T = (X(t_1), \ldots, X(t_k))^\top$. Let $\tilde{T} = (t_1 - a, \ldots, t_k - a)$. If $\{X(t)\}$ is a Gaussian process, then both X_T and $X_{\tilde{T}}$ are multivariate normal and it suffices to show that they have the same mean and covariance. This follows directly from the fact that they have the same element-wise mean c and the equal pairwise covariances, $\text{Cov}(X(t_i - a), X(t_j - a)) = h(t_i - t_j) = \text{Cov}(X(t_i), X(t_j))$.

(b) Since all finite dimensional distributions of $\{V(t)\}$ are normal, it is a Gaussian process. Thus from part (a), it suffices to show the following:

$$\mathbb{E}[V(t)] = e^{-\alpha t/2}\mathbb{E}[X(ae^{\alpha t})] = 0.$$

Thus, $\mathbb{E}[V(t)]$ is constant.

For $s \leq t$,

$$\text{Cov}(V(s), V(t)) = e^{-\alpha(t+s)/2}\text{Cov}(X(ae^{\alpha s}), X(ae^{\alpha t})) = e^{-\alpha(t+s)/2}ae^{\alpha s} = ae^{-\alpha(t-s)/2},$$

which depends only on $t - s$.

8. Let X_t and Y_t represent two independent Poisson processes with arrival rates λ_1 and λ_2, respectively, where these rates indicate the hourly arrival rate of customers at stores 1 and 2.

   (a) What is the probability that a customer arrives at store 1 before any customers arrive at store 2?

   (b) What is the probability that, during the first hour, the combined total number of customers arriving at both stores is exactly four?

   (c) Given that exactly four customers arrived across the two stores, what is the probability that all four arrived exclusively at store 1?

   (d) Let T denote the arrival time of the first customer at store 2. Then, X_T represents the count of customers at store 1 by the time the first customer arrives at store 2. Determine the probability distribution of X_T.

Solution:

   (a) Probability that a customer arrives in store 1 before any customers arrive in store 2

   Let T_{1i} and T_{2i} be the times of the i-th customer arriving at store 1 and store 2, respectively. Let $T = \min(T_1, T_2)$, then we want to compute the probability $P(T_1 = T)$. This can be computed as:

$$P(T_1 = T) = \int_0^\infty P(T_2 > t)dP(T_1 = t) = \int_0^\infty e^{-\lambda_2 t}\lambda_1 e^{-\lambda_1 t}dt = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

   Thus, the probability that a customer arrives in store 1 before any customers arrive in store 2 is given by $\frac{\lambda_1}{\lambda_1+\lambda_2}$.

(b) The total number of customers arriving at the two stores by time $t = 1$ is $N(1) = N_1(1) + N_2(1)$, where $N_1(1) \sim \text{Poisson}(\lambda_1)$ and $N_2(1) \sim \text{Poisson}(\lambda_2)$. Thus, $N(1) \sim \text{Poisson}(\lambda_1 + \lambda_2)$, and we have:

$$P(N(1) = 4) = \frac{(\lambda_1 + \lambda_2)^4 e^{-(\lambda_1+\lambda_2)}}{4!}.$$

(c) Given $N_1(1)+N_2(1) = 4$, the number of customers in store 1 follows a binomial distribution:

$$P(N_1(1) = 4 \mid N_1(1) + N_2(1) = 4) = \binom{4}{4} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^4 = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^4.$$

(d) Let T denote the time of the first arrival in store 2. Then X_T is the number of customers in store 1 by time T. Find the distribution of X_T.

Since $T \sim \text{Exp}(\lambda_2)$, the number of customers in store 1 by time T follows a Poisson distribution:

$$P(X_T = k) = \int_0^\infty P(X_T = k \mid T = t) f_T(t)\, dt = \int_0^\infty \frac{(\lambda_1 t)^k e^{-\lambda_1 t}}{k!} \lambda_2 e^{-\lambda_2 t}\, dt.$$

This simplifies to:

$$P(X_T = k) = \frac{\lambda_2}{k!} \int_0^\infty (\lambda_1 t)^k e^{-(\lambda_1+\lambda_2)t}\, dt = \frac{\lambda_2}{k!} \frac{k!}{(\lambda_1 + \lambda_2)^{k+1}} (\lambda_1)^k = \frac{\lambda_2}{\lambda_1 + \lambda_2} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k.$$

Thus, $X_T \sim \text{Geom}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)$:

$$P(X_T = k) = \frac{\lambda_2}{\lambda_1 + \lambda_2} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k.$$

9. Suppose $X(t)$ is a Gaussian process, with $X(0) = 0$ with probability 1. Suppose that $X_t + X_s \sim \mathcal{N}\left(0, \sqrt{|t-s|}\right)$.

(a) Calculate the auto-covariance function.

(b) Calculate the distribution of $(X(t_1), X(t_2), \ldots, X(t_n))$.

(c) Prove that such a process doesn't exist.

Solution:

(a) Auto-Covariance Function

We know:

$$X_t + X_s \sim \mathcal{N}\left(0, \sqrt{|t-s|}\right),$$

so:

$$\text{Var}(X_t + X_s) = \text{Var}(X_t) + \text{Var}(X_s) + 2\,\text{Cov}(X_t, X_s).$$

Since $X(t) = 0$ with probability 1, we have $\text{Var}(X_t) = 0$ and $\text{Var}(X_s) = 0$. Therefore:

$$2\,\text{Cov}(X_t, X_s) = \sqrt{|t-s|},$$

which gives:

$$\text{Cov}(X(t), X(s)) = R_X(t, s) = \frac{\sqrt{|t-s|}}{2}.$$

(b) Joint Distribution

Since $X(t)$ is a Gaussian process, the joint distribution $(X(t_1), X(t_2), \ldots, X(t_n))$ is multivariate normal with:

$$\mathbb{E}[X(t_1), X(t_2), \ldots, X(t_n)] = (0, 0, \ldots, 0),$$

and the covariance matrix Σ is given by:

$$\Sigma_{ij} = \frac{\sqrt{|t_i - t_j|}}{2}.$$

Thus, the joint distribution is:

$$(X(t_1), X(t_2), \ldots, X(t_n)) \sim \mathcal{N}(0, \Sigma),$$

where $\Sigma_{ij} = \frac{\sqrt{|t_i - t_j|}}{2}$.

(c) Non-existence

We calculate the variance of $X_t - X_s$:

$$\mathrm{Var}(X_t - X_s) = \mathrm{Var}(X_t) + \mathrm{Var}(X_s) - 2\mathrm{Cov}(X_t, X_s),$$

where $\mathrm{Var}(X_t) = 0$, $\mathrm{Var}(X_s) = 0$, and:

$$\mathrm{Cov}(X_t, X_s) = \frac{\sqrt{|t - s|}}{2}.$$

Thus, the equation becomes:

$$\mathrm{Var}(X_t - X_s) = 0 + 0 - 2 \times \frac{\sqrt{|t - s|}}{2} = -\sqrt{|t - s|} < 0 \quad \text{for } t \neq s.$$

Since variance cannot be negative, this leads to a contradiction. Therefore, such a process does not exist.