Chia-Ping Chen

Professor
Department of Computer Science and Engineering
National Sun Yat-sen University

Probability
A random process is a kind of random experiment. It evolves in time and generates a set of random values as outcome.

Every $\omega \in \Omega$ corresponds to a sequence or function $X(\omega)$, which is called a sample path or sample function.

Random processes are often used as mathematical models to analyze sequential data.
A **discrete-time** random process has a discrete index set. A **continuous-time** random process has a continuous index set.

A random variable in a discrete-time random process $X$ is denoted by $X_n$.

A random variable in a continuous-time random process $X$ is denoted by $X_t$. 
Bernoulli Processes 伯努利過程
A random process that generates binary values of 0 or 1 is called an **arrival process**. Value 1 means arrival, and value 0 means no arrival.

A random variable in an arrival process is a Bernoulli random variable.

An arrival process can be used to model
- speech activity
- winning
- renewal
A Bernoulli process is a discrete-time arrival process with i.i.d. Bernoulli random variables. It is the simplest arrival process.

A Bernoulli process $X$ with $X_n \sim \text{Bernoulli}(p)$ is denoted by $X \sim \text{BerPro}(p)$.
Example 6.1 獨立性

Let $X$ be a Bernoulli process.

- Random variable $U$ is the number of arrivals from time 1 to 5, and random variable $V$ is the number of arrivals from time 6 to 10. Then $U \perp \perp V$.
- Random variable $W$ is the first odd-index time with an arrival, and random variable $Y$ is the first even-index time with an arrival. Then $W \perp \perp Y$. 
Suppose

\[ X \sim \text{BerPro}(p) \]

The **number of arrivals** from time 1 to time \( n \) is binomial

\[ S_n \sim \text{binomial}(n, p) \]

The **first arrival time** is geometric

\[ T \sim \text{geometric}(p) \]
Consider

\[ X \sim \text{BerPro}(p) \]

Suppose the time index is reset at a time \( l > 0 \), so the restarted process \( X' \) is related to the original process \( X \) by

\[ X' : X'_{n} = X_{l+n}, \quad n = 1, 2, \ldots \]

Then

\[ X' \sim \text{BerPro}(p) \]
Consider

$$X \sim \text{BerPro}(p)$$

Suppose the time index is reset at a *random* time $L > 0$, so the restarted process $X'$ is related to the original process $X$ by

$$X' : X'_n = X_{L+n}, \quad n = 1, 2, \ldots$$

If $L$ depends only on $\{X_n, n \leq L\}$, i.e. $L$ is independent of $X_{L+1} X_{L+2} \ldots$, then

$$X' \sim \text{BerPro}(p)$$
Example 6.2 電腦很忙

A computer executes 2 types of jobs, **priority** and **non-priority**. A slot is **busy** if the computer executes a priority job, **idle** otherwise. We call a string of idle slots, flanked by busy slots, an **idle period**. We call a string of busy slots, flanked by idle slots, a **busy period**.

A priority job occurs with probability $p$ at the beginning of each slot, independent of other slots, and requires one full slot to execute. A non-priority job is always available and is executed at a given slot if no priority job is available.
Let $T$ be the time of the first idle slot. Then

$$T \sim \text{geometric}(1 - p)$$

Let $B$ be the length of the first busy period. Then

$$B \sim \text{geometric}(1 - p)$$

Let $I$ be the length of the first idle period. Then

$$I \sim \text{geometric}(p)$$

Let $Z$ be the number of slots after the first slot of the first busy period, including the first subsequent idle slot. Then

$$Z = B \Rightarrow Z \sim \text{geometric}(1 - p)$$
Suppose \( X \sim \text{BerPro}(p) \). Let \( N \) be the first time that an arrival immediately follows an arrival. What is the probability

\[
P(X_{N+1} = 0 \cap X_{N+2} = 0)
\]

that there are no arrivals in the next two time slots?

\( N \) depends on \( X_{\leq N} \), so

\[
X' : X_{N+1}, X_{N+2}, \ldots
\]

is a Bernoulli process \( X' \sim \text{BerPro}(p) \), and

\[
P(X_{N+1} = 0 \cap X_{N+2} = 0) = P(X'_1 = 0 \cap X'_2 = 0) = (1 - p)^2
\]
Let $X$ be a Bernoulli process. Let $Y_k$ be the **arrival time** of the $k$th arrival of $X$, and $T_k$ be the **inter-arrival time** between arrival $(k-1)$ and arrival $k$ of $X$. Then

$$Y_k = T_1 + \cdots + T_k$$

and

$$T_k = Y_k - Y_{k-1}$$
The inter-arrival times of $X \sim \text{BerPro}(p)$ are i.i.d. geometric random variables with parameter $p$.

The first arrival time $T_1$ is geometric, with

$$T_1 \sim \text{geometric}(p)$$

$X'$ is $X$ reset at $T_1$, so $X' \sim \text{BerPro}(p)$ by the fresh-start property. The inter-arrival time $T_2$ of $X$ is the first arrival time $Y_1'$ of $X'$, so

$$T_2 = Y_1' \sim \text{geometric}(p)$$
Example 6.4 雨天

It has been observed that after a rainy day, the number of days until it rains again is geometrically distributed with parameter $p$, independent of the past. Find the probability that it rains on both the 5th and the 8th day of the month.

\[ P(X_5 = 1 \cap X_8 = 1) = p^2 \]
Suppose \( X \sim \text{BerPro}(p) \). What is the PMF of \( Y_k \), the \( k \)th arrival time?

Event \( Y_k = n \) is equivalent to \( k - 1 \) arrivals from time 1 to time \( n - 1 \) and an arrival at time \( n \). It also requires \( n \geq k \). Thus

\[
p_{Y_k}(n) = \binom{n - 1}{k - 1} p^{k-1} (1 - p)^{n-1-(k-1)} \times \binom{n - 1}{k - 1} p^k (1 - p)^{n-k}, \quad n \geq k
\]

This is also known as the Pascal PMF.
Example 6.5 篮球

In each playing minute, Lin commits a foul with probability $p$ and no foul with probability $1 - p$. Committed fouls are independent. Lin plays 30 minutes if he does not foul out. What is the PMF of Lin’s playing time $Z$?

Lin’s fouls can be seen as arrivals of $X \sim \text{BerPro}(p)$. Let $Y_k$ be the $k$th arrival time of $X$. Then

$$Z = \min(Y_6, 30)$$

We have equivalent events

$$(Z = 30) = (Y_6 \geq 30), \quad (Z = n) = (Y_6 = n), \quad n = 6, \ldots, 29$$

so

$$p_Z(n) = \begin{cases} 
  p_{Y_6}(n), & 6 \leq n \leq 29 \\
  1 - \sum_{n' = 6}^{29} p_{Y_6}(n'), & n = 30
\end{cases}$$
Let the arrivals of an arrival process $X \sim \text{BerPro}(p)$ be split stochastically. Specifically, an arrival is accepted with probability $q$, and not accepted with probability $1 - q$. Then the accepted arrivals of $X$ is an arrival process $Y \sim \text{BerPro}(pq)$.

The unaccepted arrivals of $X$ is another arrival process $Z \sim \text{BerPro}(p(1 - q))$. 
Let $W$ be the merged process of $X \sim \text{BerPro}(p)$ and $Y \sim \text{BerPro}(q)$. Specifically, $W$ has an arrival if either $X$ or $Y$ has an arrival. Then $W$ is an arrival process with

$$W \sim \text{BerPro}(p + q - pq)$$
Poisson Processes

泊松過程
In a continuous-time arrival process, an arrival can occur at any time.

The simplest continuous-time arrival process is a Poisson process, with uniform and independent arrivals.
Let $N(\delta)$ be the number of arrivals of a Poisson process in a period of length $\delta$. By assumption, $N(\delta)$ is approximately Bernoulli

$$P(N(\delta) = k) = \begin{cases} 
1 - \lambda \delta + o(\delta), & k = 0 \\
\lambda \delta + o(\delta), & k = 1 \\
o(\delta), & k > 1 
\end{cases}$$

where $\lambda > 0$ is a parameter.
The parameter $\lambda$ of a Poisson process is the expected number of arrivals per unit time.

$$E[N(\delta)] = \lambda \delta + o(\delta)$$

$$\lim_{\delta \to 0^+} \frac{E[N(\delta)]}{\delta} = \lambda$$

Thus, the parameter $\lambda$ is the **arrival rate**. A Poisson process with rate $\lambda$ is denoted by

$$X \sim \text{PoiPro}(\lambda)$$
The number of arrivals of in a Poisson process in any finite period is a Poisson random variable.

Consider a Poisson process with rate $\lambda$. For very small $\delta$, the number of arrivals $N(\delta)$ is Bernoulli

$$N(\delta) \sim \text{Bernoulli}(\lambda \delta)$$

A period of length $\tau$ can be partitioned into $n$ disjoint periods, each of length $\delta = \frac{\tau}{n}$. $N(\tau)$ is the sum of $n$ Bernoulli random variables with parameter $\lambda \delta$. Therefore

$$N(\tau) \sim \text{binomial}(n, \lambda \delta) \approx \text{Poisson}(n \lambda \delta) = \text{Poisson}(\lambda \tau)$$

The approximation is exact as $n \to \infty$. 
Number of periods: $n = \frac{\tau}{\delta}$

Probability of success per period: $p = \lambda \delta$

Expected number of arrivals: $np = \lambda \tau$
Example 6.8 電子邮件

Bill gets e-mails according to a Poisson process at a rate of \( \lambda = 0.2 \) messages per hour. He checks email every hour. What is the probability of no new message? 1 new message?

\[ N(1) \sim \text{Poisson}(0.2 \times 1) \]
Arrivals of customers at a local supermarket are modeled by a Poisson process with a rate of $\lambda = 10$ customers per minute. $M$ is the number of customers arriving between 9:00 and 9:10, and $N$ is the number of customers arriving between 9:30 and 9:35. What is the PMF of $M + N$?

\[(M + N) \sim (\text{Poisson}(10 \times 10) + \text{Poisson}(10 \times 5)) = \text{Poisson}(150)\]
The first arrival time of a Poisson process is an **exponential random variable**.

Consider a Poisson process with rate $\lambda$. We have equivalent events

$$(Y_1 > t) = \{\text{no arrivals in } (0, t)\} = (N(t) = 0)$$

So

$$F_{Y_1}(t) = P(Y_1 \leq t)$$

$$= 1 - P(Y_1 > t)$$

$$= 1 - P(N(t) = 0)$$

$$= 1 - e^{-\lambda t}$$

That is

$$Y_1 \sim \text{exponential}(\lambda)$$
Consider a Poisson process with rate $\lambda$.

Suppose the time index is reset at $u > 0$, so the restarted process is

$$X' : X'_t = X_{u+t}$$

Then $X' \sim \text{PoiPro}(\lambda)$.

Suppose the time index is reset at random $U > 0$, where $U$ depends only on $\{X_t, t \leq U\}$, so the restarted process is

$$X' : X'_t = X_{U+t}$$

Then $X' \sim \text{PoiPro}(\lambda)$. 
You and your partner go to a badminton court, and have to wait until the players in the court finish playing. Assume that their playing time has an exponential PDF. Then the PDF of your waiting time has the same exponential PDF, regardless of when they started playing.
When you enter the bank, you find all 3 tellers are busy serving customers, and there are no other customers in queue. After you, no more customers are allowed to enter. Assume that the service times for you and for each of the customers being served are i.i.d. exponential random variables. What is the probability that you will be the last to leave?

\[
\frac{1}{4} \ ? \ \frac{1}{3} \\
\]

Let $Y_k$ be the $k$th arrival time of a Poisson process with rate $\lambda$. Then

$$f_{Y_k}(t) = \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t}, \quad t \geq 0$$

$$P(t < Y_k < t + \delta) = P(N(t) = k - 1) \times P(N(\delta) = 1)$$

$$= \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} \times (\lambda \delta + o(\delta))$$

Then

$$f_{Y_k}(t) = \lim_{\delta \to 0^+} \frac{F_{Y_k}(t + \delta) - F_{Y_k}(t)}{\delta} = \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t}$$
Example 6.12 国税局热线

You call the IRS hotline. You are the 56th person waiting to be served. Callers depart according to a Poisson process with a rate of \( \lambda = 2 \) per minute. How long will you have to wait on average until your service starts, and what is the probability that you will have to wait for more than 30 minutes?

Your wait until everyone in front of you have departed, so your waiting time is

\[
W = Y_{56}
\]

\[
E[W] = E[Y_{56}] = \frac{56}{2} = 28
\]

\[
P(W > 30) = P(Y_{56} > 30) = \int_{30}^{\infty} \frac{2^{56} t^{55}}{(55)!} e^{-2t} dt
\]

Exercise: Estimate the probability by CLT.
By the fresh-start property, the inter-arrival times of a Poisson process are \textbf{i.i.d.} exponential variables.

Consider a Poisson process with rate $\lambda$. A sample path of $X$ can be generated as follows.

- Sample inter-arrival times
  \[ T_i \sim \text{exponential}(\lambda) \]

- Set $X_t$ to 1 at
  \[ t = T_1, T_1 + T_2 \ldots \]

  Set $X_t$ to 0 otherwise.
Let the arrivals of an arrival process $X \sim \text{PoiPro}(\lambda)$ be split stochastically: an arrival is **accepted** with probability $q$, and **not accepted** with probability $1 - q$. The accepted arrivals define an arrival process

$$Y \sim \text{PoiPro}(\lambda q)$$

The unaccepted arrivals define another arrival process

$$Z \sim \text{PoiPro}(\lambda(1 - q))$$

\[
\begin{align*}
P(N(\delta) = 1) &= [\lambda \delta + o(\delta)]q = (\lambda q)\delta + o(\delta) \\
P(N(\delta) = 0) &= [1 - \lambda \delta + o(\delta)] \cdot 1 + [\lambda \delta + o(\delta)] \cdot (1 - q) \\
&= 1 - (\lambda q)\delta + o(\delta) \\
P(N(\delta) = n) &= o(\delta), \quad n \geq 2
\end{align*}
\]
Let $W$ be the merged process of $X \sim \text{PoiPro}(\lambda_1)$ and $Y \sim \text{PoiPro}(\lambda_2)$. Then

$$W \sim \text{PoiPro}(\lambda_1 + \lambda_2)$$

For the number of arrivals of $W$ in a small period

$$\mathbb{P}(N(\delta) = 0) = (1 - \lambda_1 \delta + o(\delta))(1 - \lambda_2 \delta + o(\delta))$$
$$= 1 - (\lambda_1 + \lambda_2)\delta + o(\delta)$$

$$\mathbb{P}(N(\delta) = 1) = \lambda_1 \delta(1 - \lambda_2 \delta) + (1 - \lambda_1 \delta)(\lambda_2 \delta) + o(\delta)$$
$$= (\lambda_1 + \lambda_2)\delta + o(\delta)$$

So

$$W \sim \text{PoiPro}(\lambda_1 + \lambda_2)$$
The arrivals of packets at a node of a data network define a Poisson process with rate $\lambda$. Each packet is either a local packet (with probability $p$) or a transit packet (with probability $1 - p$), independent of other arrivals and independent of the arrival times.

The arrival of local packets is a Poisson process with rate $\lambda p$, i.e.

$$\text{PoiPro}(\lambda p)$$
At a post office, the arrivals of people with letters to mail define a Poisson process with rate $\lambda_1$, and the arrivals of people with packages to mail define a Poisson process with rate $\lambda_2$. The arrivals are independent.

The arrivals of people with letters or packages to mail define a Poisson process

$$W \sim \text{PoiPro}(\lambda_1 + \lambda_2)$$

A person arrives at the post office. He wants to mail a letter with probability

$$\frac{\lambda_1}{\lambda_1 + \lambda_2}$$
Example 6.15 二燈泡

Two light bulbs have lifetimes $T_a$ and $T_b$ which are independent exponential random variables with parameters $\lambda_a$ and $\lambda_b$. What is the distribution for the first time a bulb burns out?

$$Y_1 = \min(T_a, T_b) \Rightarrow (Y_1 > t) = (T_a > t \cap T_b > t)$$

$$\Rightarrow P(Y_1 > t) = P(T_a > t) P(T_b > t)$$

$$\Rightarrow 1 - P(Y_1 < t) = e^{-\lambda_a t} e^{-\lambda_b t}$$

$$\Rightarrow P(Y_1 < t) = 1 - e^{-(\lambda_a + \lambda_b)t} = F_{Y_1}(t)$$

So

$$f_{Y_1}(t) = (\lambda_a + \lambda_b) e^{-(\lambda_a + \lambda_b)t}$$

That is

$$Y_1 \sim \text{exponential}(\lambda_a + \lambda_b)$$
Example 6.16 三燈泡

Three light bulbs have lifetimes which are independent exponential random variables with a common parameter $\lambda$. What is the expected value of the time until the last bulb burns out?

\[ Y_1^{(3)} \sim \text{exponential}(3\lambda) \]
\[ Y_1^{(2)} \sim \text{exponential}(2\lambda) \]
\[ Y_1^{(1)} \sim \text{exponential}(\lambda) \]

\[
\mathbb{E} \left[ Y_1^{(3)} + Y_1^{(2)} + Y_1^{(1)} \right] = \mathbb{E} \left[ Y_1^{(3)} \right] + \mathbb{E} \left[ Y_1^{(2)} \right] + \mathbb{E} \left[ Y_1^{(1)} \right] = \frac{1}{3\lambda} + \frac{1}{2\lambda} + \frac{1}{\lambda}
\]
Random Sum 隨機和
Definition

Recall that a random sum is the sum of i.i.d. random variables

\[ Y = X_1 + \cdots + X_N \]

where the number of terms in the sum, \( N \), is another random variable, and \( N \) is independent of the i.i.d. random variables \( X_1 \ldots \).
A geometric sum of geometric random variables is a geometric random variable.

Specifically, if

\[ N \sim \text{geometric}(q) \]

and

\[ X_i \sim \text{geometric}(p) \]

then

\[ (X_1 + \cdots + X_N) \sim \text{geometric}(pq) \]
Split $Z \sim \text{BerPro}(p)$ with probability $q$ to $Z' \sim \text{BerPro}(pq)$. Let $X_i$ be the $i$th inter-arrival time of $Z$, $N$ be the number of arrivals in $Z$ until an arrival is accepted by $Z'$, and $Y_1'$ be the first arrival time of $Z'$.

$$X_i \sim \text{geometric}(p), \quad N \sim \text{geometric}(q), \quad Y_1' \sim \text{geometric}(pq)$$

Since

$$Y_1' = X_1 + \cdots + X_N$$

we have

$$(X_1 + \cdots + X_N) \sim \text{geometric}(pq)$$
A geometric sum of exponential random variables is an exponential random variable.

Specifically, if

\[ N \sim \text{geometric}(q) \]

and

\[ X_i \sim \text{exponential}(\lambda) \]

then

\[ (X_1 + \cdots + X_N) \sim \text{exponential}(\lambda q) \]
Split $\mathbf{W} \sim \text{PoiPro}(\lambda)$ with probability $q$ to $\mathbf{W}' \sim \text{PoiPro}(\lambda q)$. Let $X_i$ be the $i$th inter-arrival time of $\mathbf{W}$, $N$ be the number of arrivals in $\mathbf{W}$ until an arrival is accepted by $\mathbf{W}'$, and $Y_1'$ be the first arrival time of $\mathbf{W}'$.

$X_i \sim \text{exponential}(\lambda), \ N \sim \text{geometric}(q), \ Y_1' \sim \text{exponential}(\lambda q)$

Since

$$Y_1' = X_1 + \cdots + X_N$$

we have

$$(X_1 + \cdots + X_N) \sim \text{exponential}(\lambda q)$$
A binomial sum of Bernoulli random variables is a binomial random variable. Specifically, if

\[ N \sim \text{binomial}(m, p) \]

and

\[ X_i \sim \text{Bernoulli}(q) \]

then

\[ (X_1 + \cdots + X_N) \sim \text{binomial}(m, pq) \]
Split $Y \sim \text{BerPro}(p)$ with probability $q$ to $Y' \sim \text{BerPro}(pq)$. Let $X_i$ be a Bernoulli random variable such that $X_i = 1$ if the $i$th arrival of $Y$ is accepted by $Y'$, and $X_i = 0$ otherwise. Let $N$ be the number of arrivals in $Y$ during the first $m$ slots. Let $N_a$ be the number of arrivals in $Y'$ during the first $m$ slots.

$$X_i \sim \text{Bernoulli}(q), \ N \sim \text{binomial}(m, p), \ N_a \sim \text{binomial}(m, pq)$$

Since

$$N_a = X_1 + \cdots + X_N$$

we have

$$(X_1 + \cdots + X_N) \sim \text{binomial}(m, pq)$$
A Poisson sum of Bernoulli random variables is a Poisson random variable.

Specifically, if

\[ N \sim \text{Poisson}(\lambda) \]

and

\[ X_i \sim \text{Bernoulli}(q) \]

then

\[ (X_1 + \cdots + X_N) \sim \text{Poisson}(\lambda q) \]
Split $W \sim \text{PoiPro}(\lambda)$ with probability $q$ to $W' \sim \text{PoiPro}(\lambda q)$.

Let $X_i$ be a Bernoulli random variable that indicates whether the $i$th arrival of $W$ is accepted by $W'$. Let $N$ be the number of arrivals in $W$ during $[0, 1]$. Let $N_a$ be the number of arrivals in $W'$ during $[0, 1]$.

\[ X_i \sim \text{Bernoulli}(q), \ N \sim \text{Poisson}(\lambda), \ N_a \sim \text{Poisson}(\lambda q) \]

Since

\[ N_a = X_1 + \cdots + X_N \]

we have

\[ (X_1 + \cdots + X_N) \sim \text{Poisson}(\lambda q) \]
Summary: Bernoulli Process

Bernoulli process

\[ X \sim \text{BerPro}(p), \; X_n \sim \text{Bernoulli}(p) \]

# of arrivals

\[ S_n \sim \text{binomial}(n, p) \]

Inter-arrival time

\[ T_n \sim \text{geometric}(p) \]

Splitting and merging

\[ Y \sim \text{BerPro}(pq), \; W \sim \text{BerPro}(p + q - pq) \]
### Summary: Poisson Process

- Poisson process with rate $\lambda$
  
  \[ X \sim \text{PoiPro}(\lambda), \quad N(\delta) \sim \text{Bernoulli}(\lambda\delta) \]

- \# of arrivals
  
  \[ N(\tau) \sim \text{Poisson}(\lambda\tau) \]

- Inter-arrival times
  
  \[ T_n \sim \text{exponential}(\lambda) \]

- Splitting and merging
  
  \[ Y \sim \text{PoiPro}(\lambda p), \quad W \sim \text{PoiPro}(\lambda_1 + \lambda_2) \]