# Stochastic Processes

# 随机过程

Chao Qian （钱超）

Associate Professor, Nanjing University, China

Email: qianc@nju.edu.cn
Homepage: http://www.lamda.nju.edu.cn/qianc/

# Stochastic process

A stochastic process is a collection $\{X(t) \mid t \in T\}$ of random variables

- $X(t)$ is a random variable
- $t$ is often interpreted as time, and $X(t)$ is called the state of the process at time $t$

➢ Discrete-time stochastic process:

The index set $T$ is a countable set

➢ Continuous-time stochastic process:

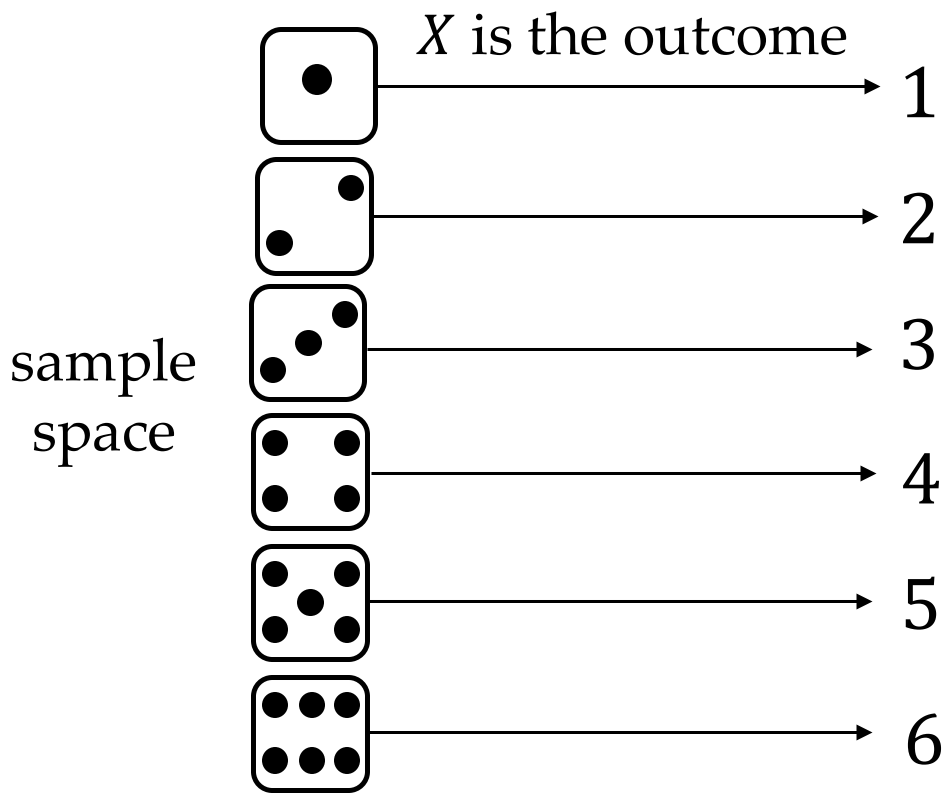
The index set $T$ is a continuum

# Random variable

A random variable $X: S \rightarrow R$ is a function that assigns a real value to each outcome in the sample space $S$

For example,

random experiment

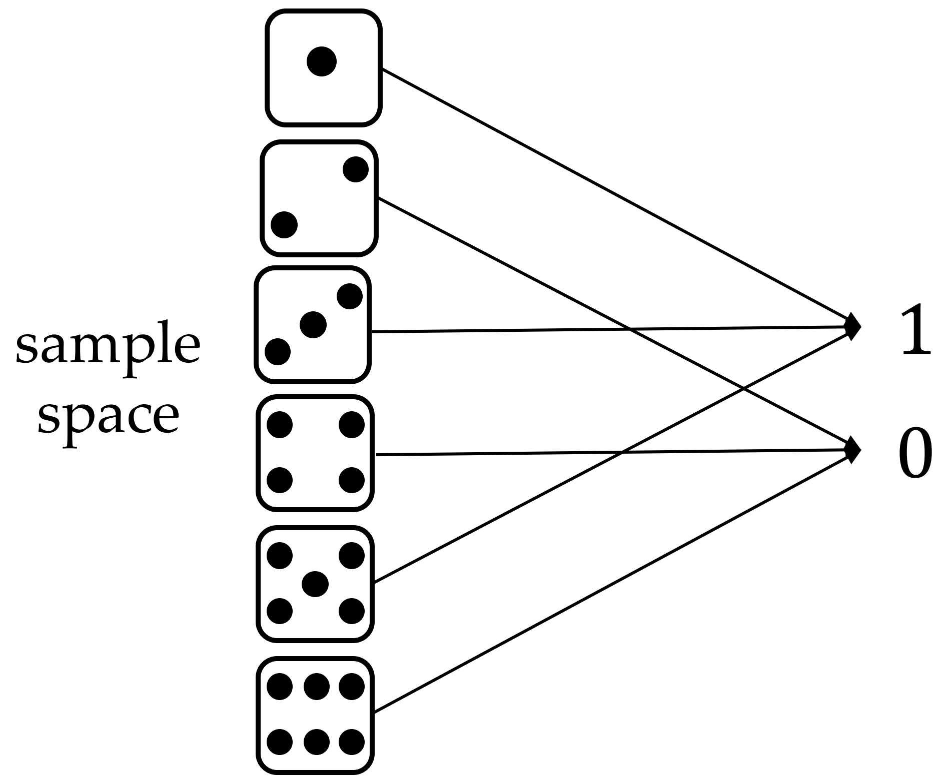


# Random variable

For example,

$X$ indicates the oddness

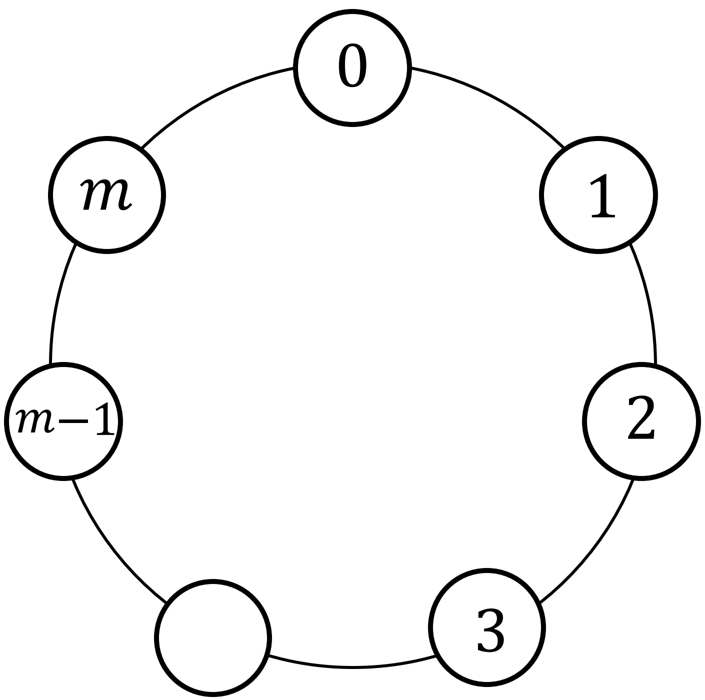
random experiment




$$P(X^{-1}(A)) \quad = \quad P(X \in A)$$

# Example of stochastic process

**Example:** Consider a particle that moves along a set of $m+1$ nodes, labelled $0, 1, \ldots, m$, that are arranged around a circle. At each step the particle is equally likely to move one position in either the clockwise or counterclockwise direction.

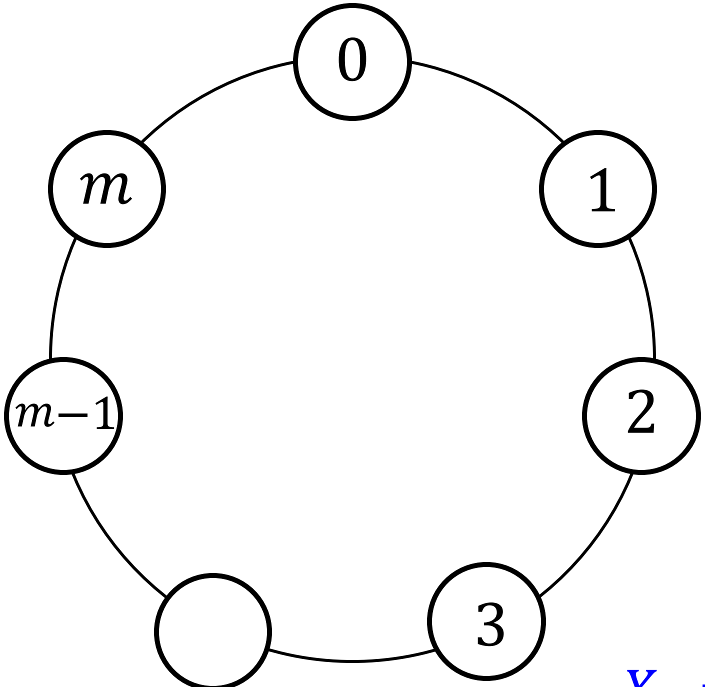



$X_n$: position of the particle after $n$ steps

$$P(X_{n+1} = i + 1 \mid X_n = i) \\ = P(X_{n+1} = i - 1 \mid X_n = i) \\ = 1/2$$

$\{X_n \mid n = 0,1,2, \ldots\}$ is a stochastic process

# Example of stochastic process




$X_n$: position of the particle after $n$ steps

$$P(X_{n+1} = i + 1 \mid X_n = i) = P(X_{n+1} = i - 1 \mid X_n = i) = 1/2$$

$\{X_n \mid n = 0,1,2, \ldots\}$ is a stochastic process

$X_0 = 0$

Suppose that the particle starts at 0 and continues to move around according to the above rules until all the nodes have been visited.

What is the probability that node $i$ is the last one visited?

# Example of stochastic process

**Solution:**

**Target event $E_i$:** $i$ is the last visited node until visiting all nodes

**Random variable $T_k$:** the first time that the particle visits $k$

$$P(E_i) = P(E_i \mid T_{i-1} < T_{i+1})P(T_{i-1} < T_{i+1}) + P(E_i \mid T_{i-1} > T_{i+1})P(T_{i-1} > T_{i+1})$$
$$= P(W_m)P(T_{i-1} < T_{i+1}) + P(W_m)P(T_{i-1} > T_{i+1})$$
$$= P(W_m)$$

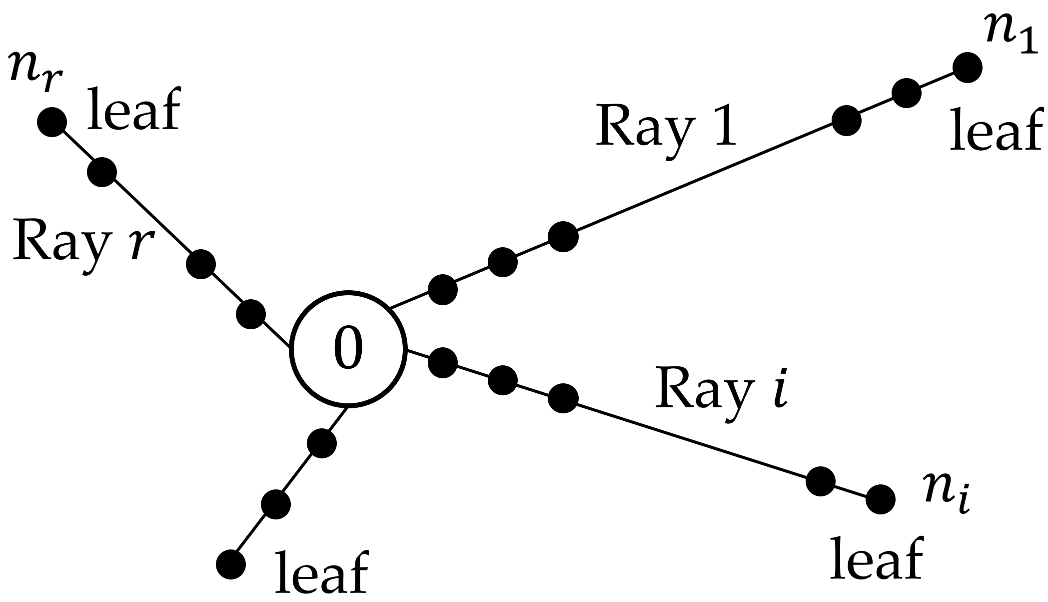
before node $i$ is visited, node $i+1$ is visidted

$$\sum_{i=1}^{m} P(E_i) = 1 \Longrightarrow P(E_i) = \frac{1}{m}$$

**Event $W_m$:** a gambler who starts with 1 unit, and wins 1 when a fair coin turns up heads and loses 1 when it turns up tails, will have his fortune go up by $m-1$ before he goes broke

# Example of stochastic process

**Example:** Consider a particle moves along the vertices of the graph so that it is equally likely to move from whichever vertex it is presently at to any of the neighbors of that vertex




a star graph with $r$ rays

$X_n$: position of the particle after $n$ steps

$\{X_n \mid n = 0,1,2,\ldots\}$ is a stochastic process

Starting at 0, what is the probability that the first visited leaf is on ray $i$?

# Example of stochastic process

**Solution:**

**Target event $\boldsymbol{E_i}$:** the first visited leaf is on ray $i$

**Event $\boldsymbol{C_k}$:** the first visited ray is $k$

$$P(E_i) = \sum_{k=1}^{r} P\left(E_i | C_k\right) \frac{1}{r}$$

$$P(E_i \mid C_i) = P\left(W_{n_i}\right) + \left(1 - P\left(W_{n_i}\right)\right) P(E_i) = \frac{1}{n_i} + \left(1 - \frac{1}{n_i}\right) P(E_i)$$

$$\forall j \neq i: P\left(E_i \mid C_j\right) = 0 + \left(1 - P\left(W_{n_j}\right)\right) P(E_i) = \left(1 - \frac{1}{n_j}\right) P(E_i)$$

$$P(E_i) r = \frac{1}{n_i} + \sum_{k=1}^{r} \left(1 - \frac{1}{n_k}\right) P(E_i) \quad \Longrightarrow \quad P(E_i) = \frac{\frac{1}{n_i}}{\sum_{k=1}^{r} \frac{1}{n_k}}$$

**Event $\boldsymbol{W_m}$:** a gambler who starts with 1 unit, and wins 1 when a fair coin turns up heads and loses 1 when it turns up tails, will have his fortune go up by $m - 1$ before he goes broke

# Properties of stochastic process

A stochastic process $\{X(t) \mid t \in T\}$ is said to have independent increments if $\forall t_0 < t_1 < t_2 < \cdots < t_n$, the random variables

$$X(t_1) - X(t_0), X(t_2) - \mathrm{X}(t_1), \dots, X(t_n) - X(t_{n-1})$$

are independent

A stochastic process $\{X(t) \mid t \in T\}$ is said to have stationary increments if $\forall s > 0$,

$$X(t+s) - X(t)$$

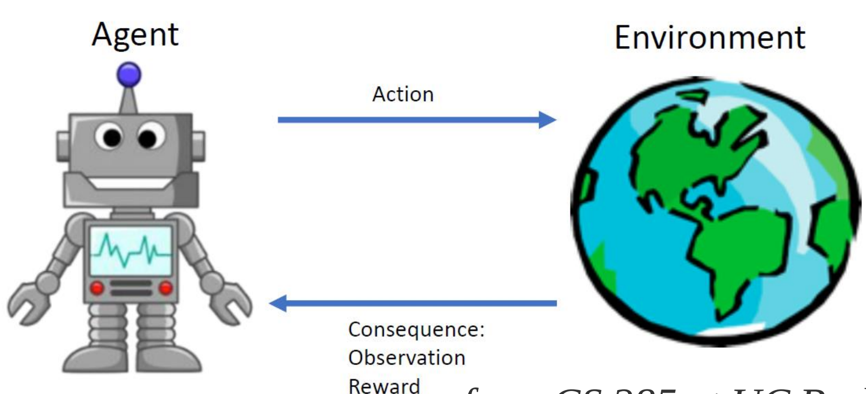
has the same distribution for all $t$

# Stochastic process in AI

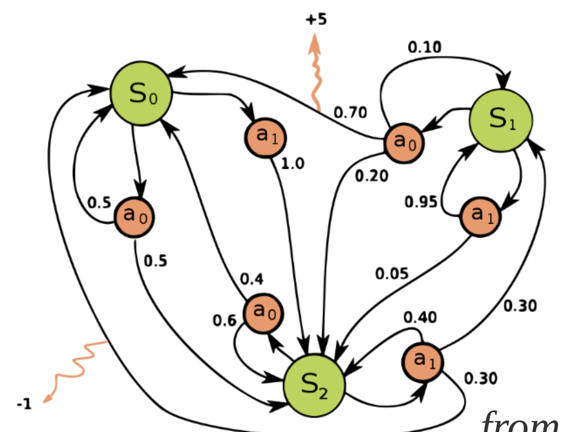
**Markov decision process** in reinforcement learning:

- State space $S$
- Action space $A$
- Transition function $P(s' \mid s, a)$ : the probability of transitioning into state $s'$ upon taking action $a$ in state $s$
- Reward function $R(s, a)$: the immediate reward associated with taking action $a$ in state $s$

Stochastic Process



*from CS 285 at UC Berkeley*



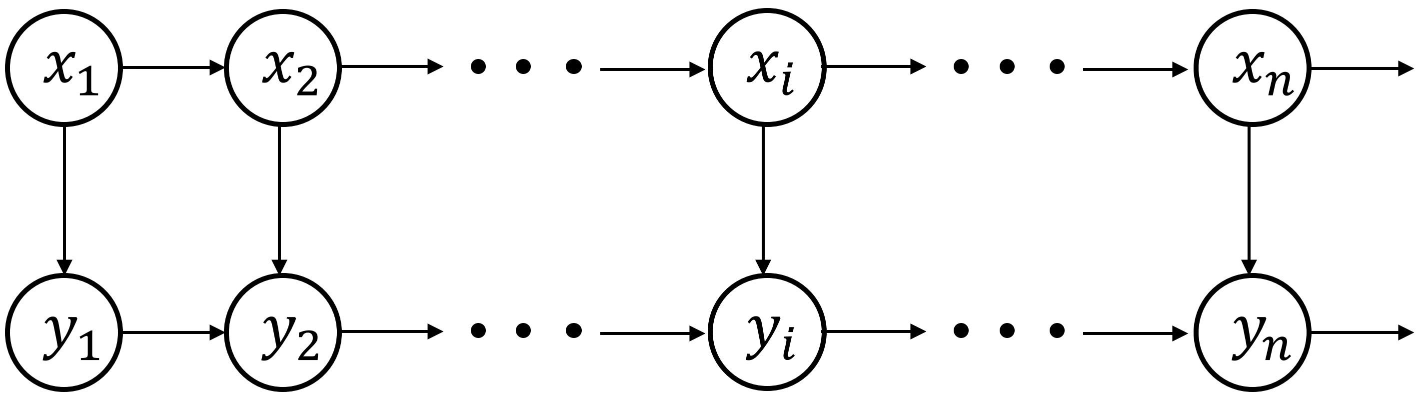
*from Wikipedia*

# Stochastic process in AI

**Hidden Markov model:** →

- Machine Learning
- Pattern Recognition
- Nature Language Processing

hidden variables




observed variables

$$P(x_1, y_1, \dots, x_n, y_n) = P(x_1)P(y_1|x_1)\prod_{i=2}^{n} P(x_i|x_{i-1})P(y_i|x_i)$$

$\{x_n | \, n \geq 1\}$ and $\{y_n | \, n \geq 1\}$ are two stochastic processes

# Stochastic process in AI

**Gaussian process** (GP) is a collection of random variables, such that every finite collection of those random variables has a multivariate normal distribution

Stochastic Process

## Bayesian optimization

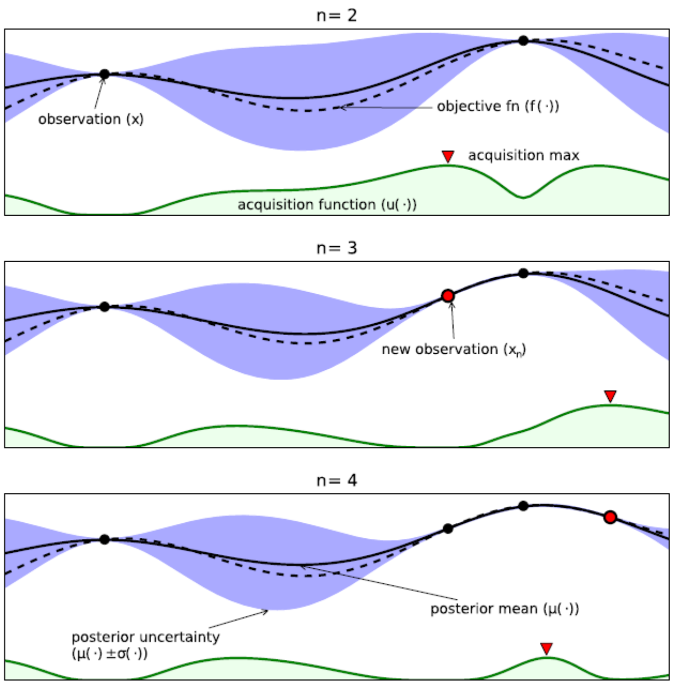
**Algorithm 1** BO Framework

**Input**: iteration budget $T$

**Process**:

let $D_0 = \emptyset$;
**for** $t = 1 : T$ **do**
  $\boldsymbol{x}_t = \arg\max_{\boldsymbol{x} \in \mathcal{X}} acq(\boldsymbol{x})$;
  evaluate $f$ at $\boldsymbol{x}_t$ to obtain $y_t$;
  augment the data $D_t = D_{t-1} \cup \{(\boldsymbol{x}_t, y_t)\}$ and update the GP model
**end for**

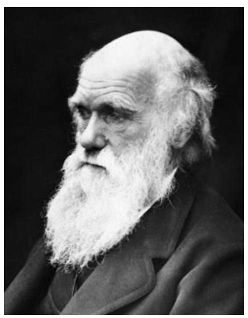
$\arg\max_{\boldsymbol{s}} f(\boldsymbol{s})$




regards the $f$ value at each data point as a random variable, and assumes satisfying a joint Gaussian distribution

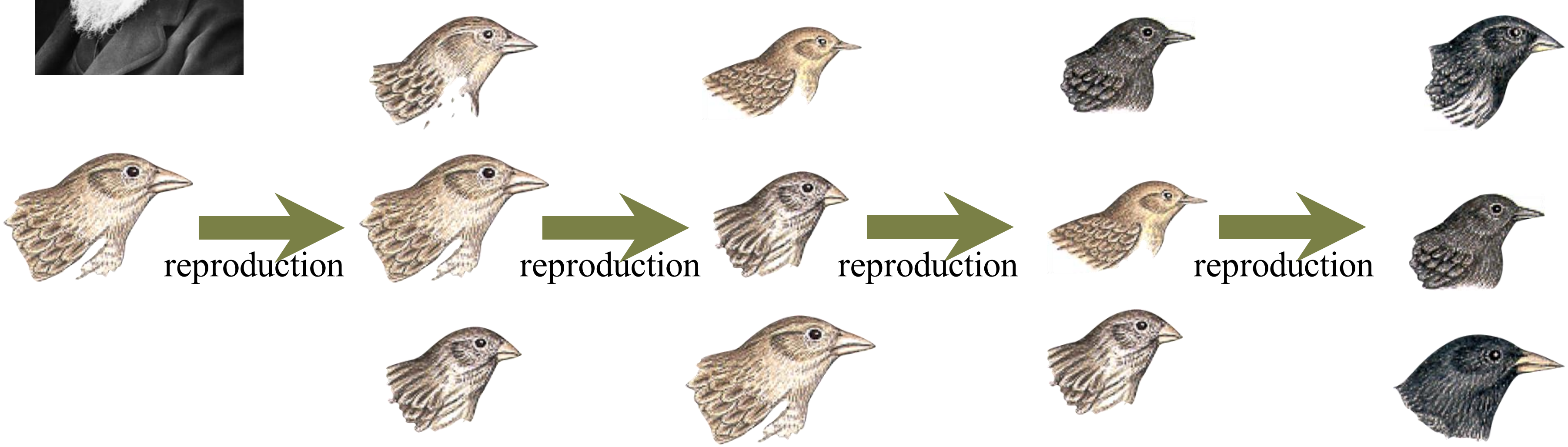
# Stochastic process in AI

**Evolutionary algorithms (EAs)** are a kind of randomized heuristic optimization algorithms, inspired by nature evolution (reproduction with variation + nature selection)



Charles Darwin
1809-1882

Nature Evolution

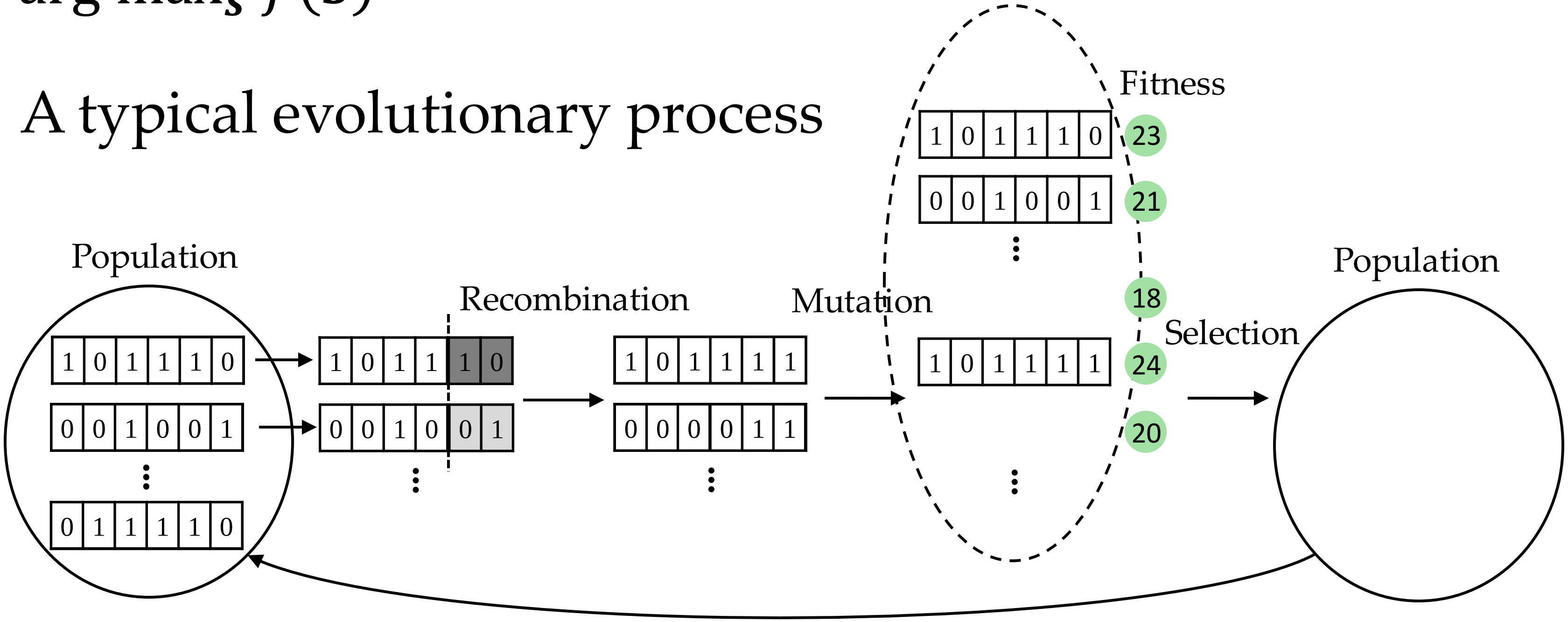


# Stochastic process in AI

**Evolutionary algorithms (EAs)** are a kind of randomized heuristic optimization algorithms, inspired by nature evolution (reproduction with variation + nature selection)

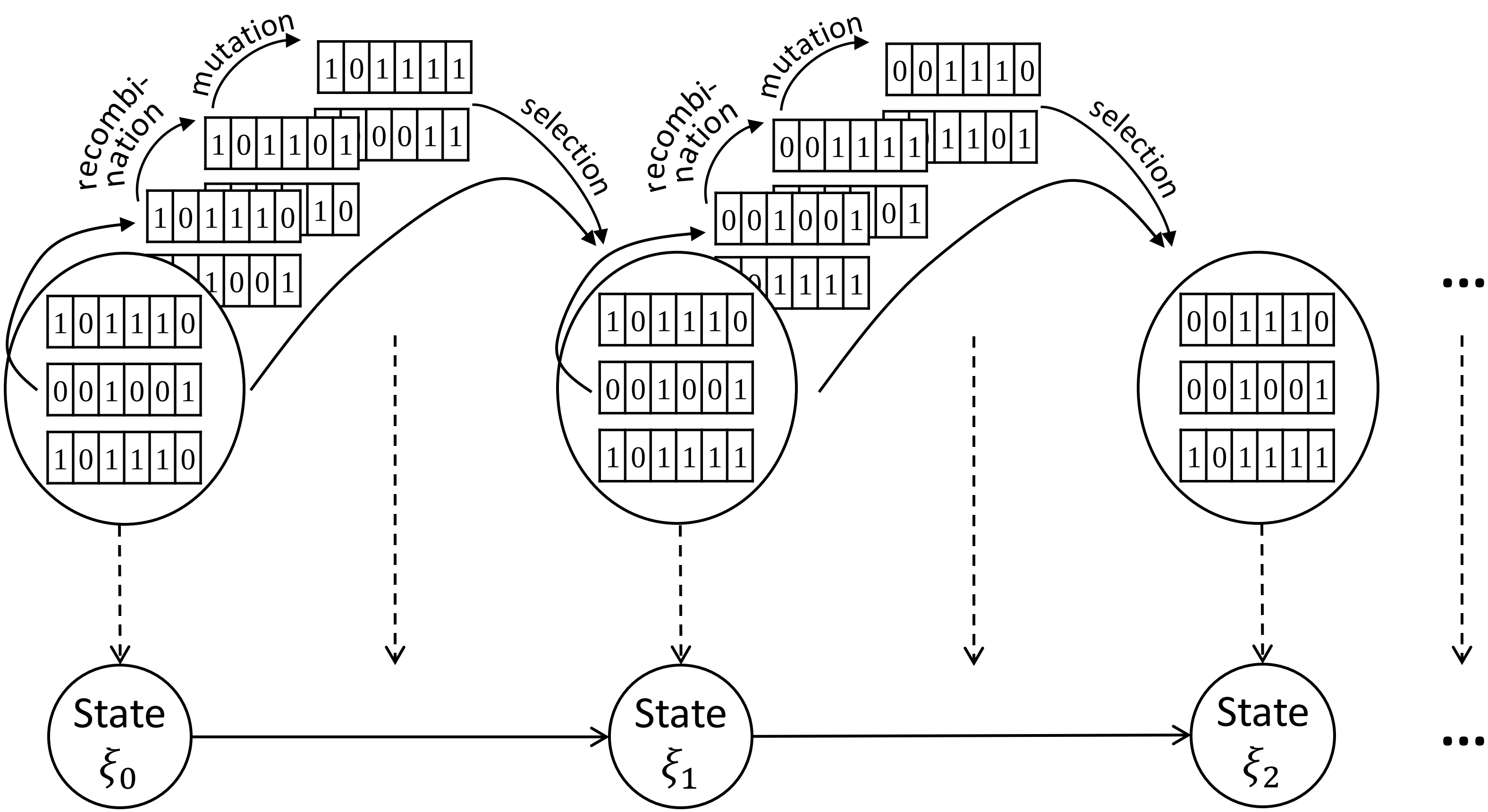
$\arg\max_{\boldsymbol{s}} f(\boldsymbol{s})$

A typical evolutionary process



# Stochastic process in AI

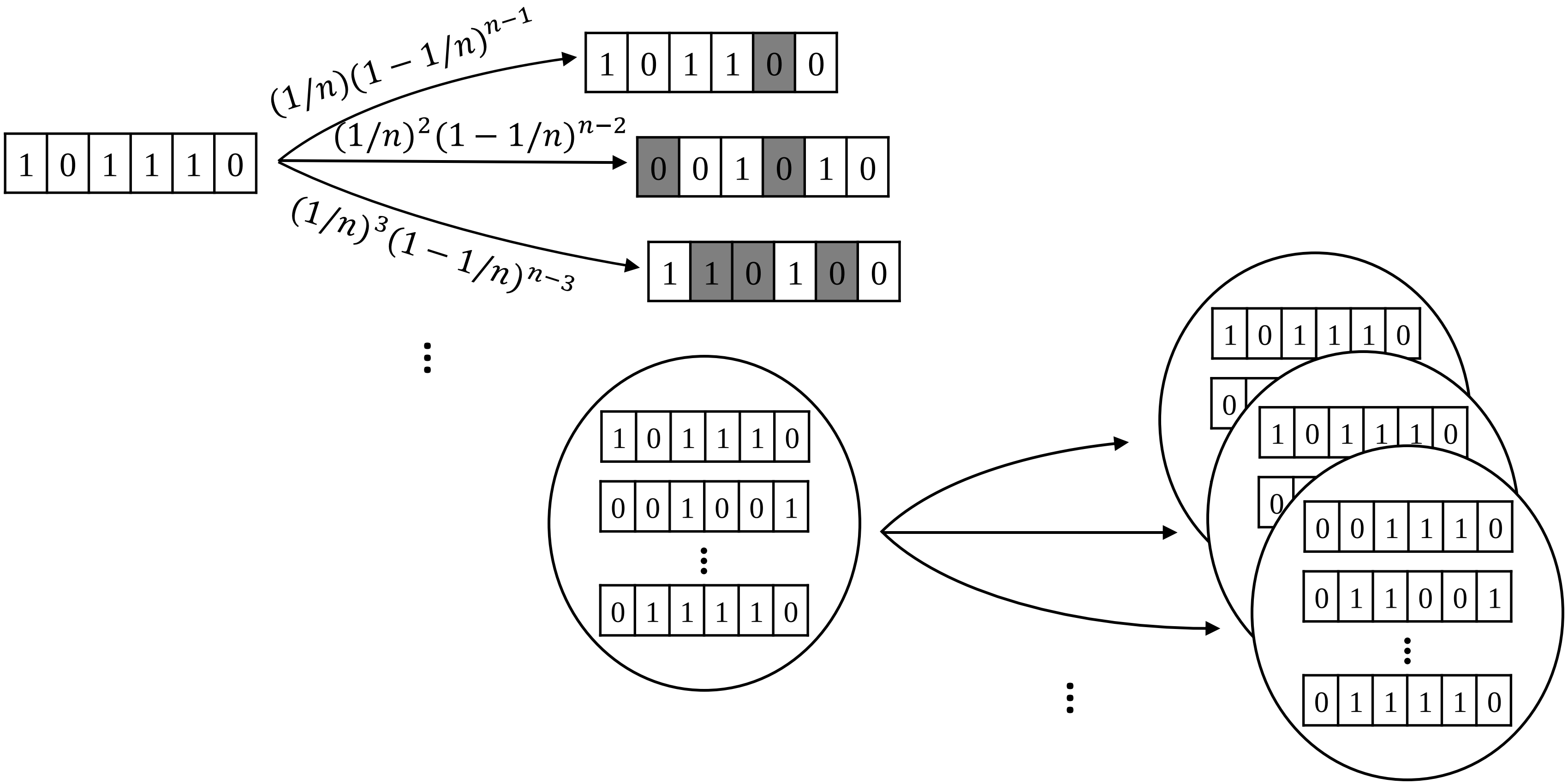
## Evolutionary algorithms



# Stochastic process in AI

The process of generating new population is randomized

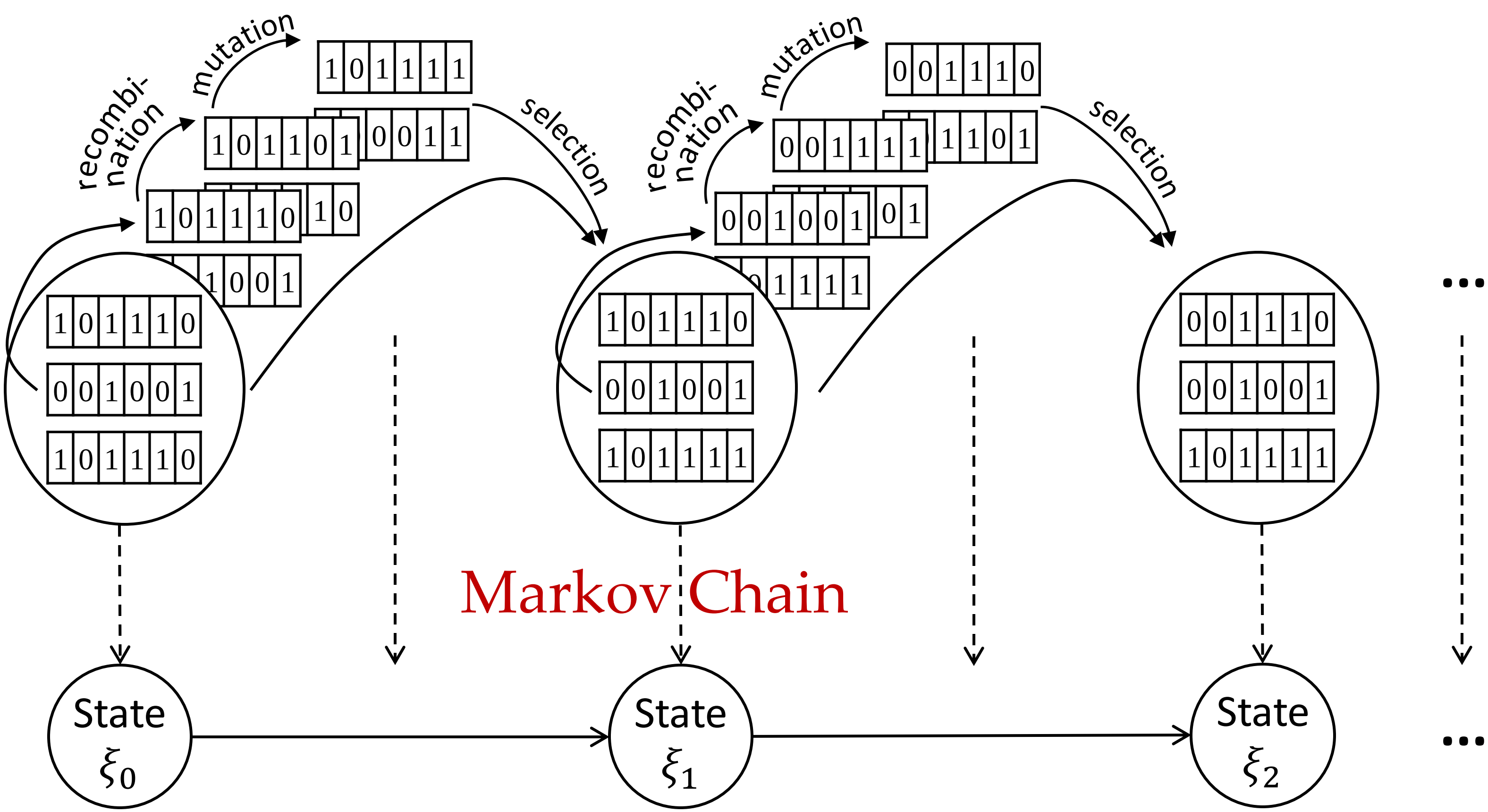
e.g., bit-wise mutation: flips each bit with prob. $1/n$



# Stochastic process in AI

**Evolutionary algorithms**

Stochastic Process




random variable random variable random variable

# Outline of this course

- **Lecture 1: Preliminaries**
- **Lecture 2: Poisson process**
- **Lecture 3: Renewal process**
- **Lecture 4: Markov chain**
- **Lecture 5: Martingale**
- **Lecture 6: Random walk**
- **Lecture 7: Brownian motion**

# 相关教材

- An Introduction to Stochastic Modeling, 4th edition, 2010
  *by Mark Pinsky and Samuel Karlin*

- Basic Stochastic Processes, 1999
  *by Zdzislaw Brzezniak and Tomasz Zastawniak*

- Stochastic Processes, 2nd edition, 1995
  *by Sheldon M. Ross*

- Introduction to Stochastic Processes, 2013
  *by Erhan Cinlar*

- Essentials of Stochastic Processes, 2nd edition, 2012
  *by Richard Durrett*

# 课程相关信息

课程时间：周五下午14:00-16:00

课程主页：

http://www.lamda.nju.edu.cn/SP22/

课程讨论QQ群：569309099

随机过程

不随机过

助教：卞超、王雨桐

答疑时间：周五下午16:00-17:30、逸A-502

成绩计算：4次平时作业（15%）、期末考试（40%）

# Stochastic Processes

## Lecture 1: Preliminaries

Chao Qian （钱超）

Associate Professor, Nanjing University, China

Email: qianc@nju.edu.cn
Homepage: http://www.lamda.nju.edu.cn/qianc/

# Independence of random variables

Two random variables $X$ and $Y$ are said to be **independent** if

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

for all $x$ and $y$

Random variables $X_1, X_2, \dots, X_n$ are said to be **mutually independent** if for any subset $I \subseteq \{1,2, \dots, n\}$ and any $x_i$,

$$P\left(\bigwedge_{i \in I} (X_i = x_i)\right) = \prod_{i \in I} P(X_i = x_i)$$

# Expectation of random variables

The **expectation** of a discrete random variable $X$ is

$$E[X] = \sum_{x} x \cdot P(X = x)$$

where the sum is over all $x$ in the range of $X$

Two common ways of calculating $E[X]$:

- Let $X = X_1 + X_2 + \cdots + X_n$, then $E[X] = \sum_{i=1}^{n} E[X_i]$
- $E[X] = E[E[X \mid Y]]$

# How to calculate the expectation

Let $X = X_1 + X_2 + \cdots + X_n$, then $\quad E[X] = \sum_{i=1}^{n} E[X_i]$

**Proof:**

$$E[X+Y] = \sum_{x}\sum_{y}(x+y)P(X=x, Y=y)$$

$$= \sum_{x} x \sum_{y} P\,(X=x, Y=y) + \sum_{y} y \sum_{x} P\,(x=x, Y=y)$$

$$= \sum_{x} x\, P(X=x) + \sum_{y} y\, P(Y=y)$$

$$= E[X] + E[Y]$$

$$\Downarrow$$

$$E[X_1 + \cdots + X_n] = E[X_1 + \cdots + X_{n-1}] + E[X_n] = \cdots = \sum_{i=1}^{n} E\,[X_i]$$

# How to calculate the expectation

$$E[X] = E[E[X \mid Y]]$$

**Proof:**

$$\begin{aligned}
E[E[X \mid Y]] &= \sum_y E\,[X \mid Y = y\,]P(Y = y) \\
&= \sum_y \sum_x x\,P\,(\,X = x \mid Y = y\,)P(Y = y) \\
&= \sum_x x \sum_y P\,(X = x, Y = y) \\
&= \sum_x x\,P(X = x) \\
&= E[X]
\end{aligned}$$

# Example

**Example:** There are $n$ keys with the same shape, where only one can unlock the door. Each key is selected randomly without replacement. Let $X$ denote the number of selected keys until unlocking the door. Calculate $E[X]$.

**Solution1:** By the definition of expectation

$$E[X] = \sum_{k=1}^{n} k \cdot P(X = k)$$

$$= \sum_{k=1}^{n} k \cdot \frac{n-1}{n} \times \frac{n-2}{n-1} \times \cdots \times \frac{n-(k-1)}{n-(k-2)} \times \frac{1}{n-(k-1)}$$

$$= \sum_{k=1}^{n} k \cdot \frac{1}{n} = \frac{n+1}{2}$$

# Example

Let $X = X_1 + X_2 + \cdots + X_n$, then $\quad E[X] = \sum_{i=1}^{n} E[X_i]$

**Solution2:**

Let $X_i = \begin{cases} 1 & \text{the first } i-1 \text{ tries all fail} \\ 0 & \text{otherwise} \end{cases}$

Then, $X_1 = 1$

$$\forall i \geq 2, E[X_i] = P(X_i = 1) = \frac{n-1}{n} \times \frac{n-2}{n-1} \times \cdots \times \frac{n-(i-1)}{n-(i-2)}$$

$$= \frac{n-(i-1)}{n}$$

Thus, $E[X] = 1 + \sum_{i=2}^{n} \frac{n-(i-1)}{n} = \frac{n+1}{2}$

# Example

$$E[X] = E[E[X \mid Y]]$$

**Solution3:**

Let $Y = \begin{cases} 1 & \text{the first try succeeds} \\ 0 & \text{otherwise} \end{cases}$,

and $X_n$ denote the random variable $X$ corresponding to $n$ keys

Then, $E[X_n] = E[E[X_n \mid Y]]$

$$= \frac{1}{n}E[X_n \mid Y = 1] + (1 - \frac{1}{n})E[X_n \mid Y = 0]$$

$$= \frac{1}{n} + (1 - \frac{1}{n})(1 + E[X_{n-1}]) = 1 + (1 - \frac{1}{n})E[X_{n-1}]$$

$\Longrightarrow$ $$E[X_n] = \frac{n+1}{2}$$

# Poisson and binomial distribution

A discrete random variable $X$ is said to have a **Poisson distribution with parameter $\boldsymbol{\lambda}$, $\boldsymbol{\lambda > 0}$**, if

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for } k = 0,1,2,\ldots$$

Expectation: $E[X] = \lambda$ Variance: $Var[X] = \lambda$

A discrete random variable $X$ is said to have a **Binomial distribution with parameters $\boldsymbol{n}$ and $\boldsymbol{p} \in [\mathbf{0}, \mathbf{1}]$**, if

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0,1,2,\ldots,n$$

Expectation: $E[X] = np$ Variance: $Var[X] = np(1-p)$

# Poisson and binomial distribution

Poisson distribution with parameter $\lambda = np$ can be used as an approximation to binomial distribution with parameters $n$ and $p$ if $n$ is sufficiently large and $p$ is sufficiently small.

**Brun's sieve:** Let $X$ be a bounded nonnegative integer-valued random variable. If, for all $i \geq 0$,

$$E\left[\binom{X}{i}\right] \approx \lambda^i / i!$$

To show that a binomially distributed random variable $X$ satisfies this equation

then

$$P(X = j) \approx \frac{\lambda^j e^{-\lambda}}{j!} \qquad \text{for } j = 0,1,2,\ldots$$

Poisson distribution

# Poisson and binomial distribution

**Brun's sieve:** Let $X$ be a bounded nonnegative integer-valued random variable. If, for all $i \geq 0$,

$$E\left[\binom{X}{i}\right] \approx \lambda^i / i! \quad \Rightarrow \quad P(X = j) \approx \frac{\lambda^j e^{-\lambda}}{j!} \qquad \text{for } j = 0,1,2,\ldots$$

**Proof:** Let $I_j = \begin{cases} 1 & \text{if } X = j \\ 0 & \text{otherwise} \end{cases} \qquad j \geq 0$

Then, $I_j = \binom{X}{j}(1-1)^{X-j} = \binom{X}{j}\sum_{k=0}^{X-j}\binom{X-j}{k}(-1)^k$

$$= \sum_{k=0}^{\infty}\binom{X}{j}\binom{X-j}{k}(-1)^k = \sum_{k=0}^{\infty}\binom{X}{j+k}\binom{j+k}{k}(-1)^k$$

Thus, $P(X=j) = E[I_j] = \sum_{k=0}^{\infty} E\left[\binom{X}{j+k}\right]\binom{j+k}{k}(-1)^k$

$$\approx \sum_{k=0}^{\infty}\frac{\lambda^{j+k}}{(j+k)!}\binom{j+k}{k}(-1)^k = \frac{\lambda^j}{j!}\sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!} \approx \frac{\lambda^j e^{-\lambda}}{j!}$$

by Taylor series

# Poisson and binomial distribution

**To prove that** a binomial random variable $X$ with parameters $n$ and $p$ satisfies

$$E\left[\binom{X}{i}\right] \approx \lambda^i / i!$$

**Proof:**

- $X$ can be viewed as the number of successes in $n$ independent trials where each is a success with probability $p$
- For each of the $\binom{n}{i}$ sets of $i$ trials, define

$$\forall j \in \left\{1,2,\ldots,\binom{n}{i}\right\}:\ X_j = \begin{cases} 1 & \text{if all the } i \text{ trails are successes} \\ 0 & \text{otherwise} \end{cases}$$

# Poisson and binomial distribution

**To prove that** a binomial random variable $X$ with parameters $n$ and $p$ satisfies

$$E\left[\binom{X}{i}\right] \approx \lambda^i / i!$$

**Proof:** $\binom{X}{i} = \sum_{j=1}^{C_n^i} X_j$

$$E\left[\binom{X}{i}\right] = E\left[\sum_{j=1}^{C_n^i} X_j\right] = \sum_{j=1}^{C_n^i} E\left[X_j\right] = C_n^i \cdot p^i = \frac{n(n-1)\cdots(n-i+1)}{i!} \cdot p^i$$

Case 1: $i$ is small enough relative to $n$, e.g. $i \in o(n)$

$$E\left[\binom{X}{i}\right] \approx \frac{n^i}{i!} \cdot p^i = \lambda^i / i!$$

Case 2: $i$ is not small relative to $n$, e.g. $i \in \Theta(n)$

$$E\left[\binom{X}{i}\right] \leq \frac{n^i}{i!} \cdot p^i \to 0$$

holds when $n$ is large enough, $p$ is small enough and $np$ is not large

# Poisson and binomial distribution

**How to bound the approximation error for general $\boldsymbol{n}$ and $\boldsymbol{p}$?**

Let $X = \sum_{i=1}^{n} X_i$, where the $X_i$ are Bernoulli random variables with respective means $p_i$, $i = 1, \dots, n$. Set $\lambda = \sum_{i=1}^{n} p_i$ and let $A$ denote a set of nonnegative integers. To bound

$$\left| P(X \in A) - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!} \right|$$

**Remark:**

$X_i$ are not necessarily independent, and $p_i$ can be different

More general than binomial distribution

# Poisson and binomial distribution

To bound $$\left| P(X \in A) - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!} \right|$$

**Proof:** Define a function $g$ for which

$$P(X \in A) - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!} = E[\lambda g(X+1) - X g(X)]$$

Let $g(0) = 0$, $\forall j \geq 0$: $g(j+1) = \frac{1}{\lambda}\left[ I(j \in A) - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!} + j g(j) \right]$

Then, $\lambda g(j+1) - j g(j) = I(j \in A) - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!}$

$$\Longrightarrow E[\lambda g(X+1) - X g(X)] = E[I(X \in A)] - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!}$$

$$= P(X \in A) - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!}$$

# Poisson and binomial distribution

To bound $\left| P(X \in A) - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!} \right|$

**Proof:**

- Analyze $E[\lambda g(X+1) - Xg(X)]$

$$E[\lambda g(X+1)] = E[\sum_{i=1}^{n} p_i \, g(X+1)] = \sum_{i=1}^{n} p_i \, E[g(X+1)]$$

$$E[Xg(X)] = \sum_{i=1}^{n} p_i \, E[g(X) \mid X_i = 1] = \sum_{i=1}^{n} p_i \, E[g(V_i + 1)]$$

following the lemma next page

$P(V_i = k) = P(\sum_{j \neq i} X_j = k \mid X_i = 1)$

$$|E[\lambda g(X+1)] - E[Xg(X)]| = \left| \sum_{i=1}^{n} p_i \left( E[g(X+1)] - E[g(V_i + 1)] \right) \right|$$
$$\leq \sum_{i=1}^{n} p_i \, E[|g(X+1) - g(V_i + 1)|]$$

# Poisson and binomial distribution

**Lemma:** For any random variable $R$,

$$E[XR] = \sum_{i=1}^{n} p_i E[R \mid X_i = 1]$$

**Proof:**

$$E[XR] = E\left[\sum_{i=1}^{n} R\, X_i\right] = \sum_{i=1}^{n} E\,[RX_i]$$

$$= \sum_{i=1}^{n} E\,[E[RX_i \mid X_i]]$$

$$= \sum_{i=1}^{n} p_i\, E[R \mid X_i = 1]$$

# Poisson and binomial distribution

$$\left|P(X \in A) - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!}\right| \leq \sum_{i=1}^{n} p_i E[|g(X+1) - g(V_i+1)|]$$

**Lemma:** For any $\lambda$ and $A$,

$$|g(j) - g(j-1)| \leq \min\{1, 1/\lambda\}$$

$$\begin{aligned} |g(X+1) - g(V_i+1)| &= |g(X+1) - g(X) + \cdots + g(V_i+2) - g(V_i+1)| \\ &\leq |g(X+1) - g(X)| + \cdots + |g(V_i+2) - g(V_i+1)| \\ &= |X - V_i| \cdot \min\{1,1/\lambda\} \end{aligned}$$

$$\begin{aligned} \left|P(X \in A) - \textstyle\sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!}\right| &\leq \textstyle\sum_{i=1}^{n} p_i \, E[|g(X+1) - g(V_i+1)|] \\ &\leq \textstyle\sum_{i=1}^{n} p_i \cdot E[|X - V_i|] \cdot \min\{1,1/\lambda\} \end{aligned}$$

# Poisson and binomial distribution

**How to bound the approximation error for general $n$ and $p$?**

Let $X = \sum_{i=1}^{n} X_i$, where the $X_i$ are Bernoulli random variables with respective means $p_i$, $i = 1, \ldots, n$. Set $\lambda = \sum_{i=1}^{n} p_i$ and let $A$ denote a set of nonnegative integers.

$$\left| P(X \in A) - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!} \right| \le \sum_{i=1}^{n} p_i \cdot \min\{1, 1/\lambda\} \cdot E[|X - V_i|]$$

where $P(V_i = k) = P(\sum_{j \neq i} X_j = k \mid X_i = 1)$

**Remark:**

$X_i$ are not necessarily independent, and $p_i$ can be different

# Poisson and binomial distribution

**How to bound the approximation error for general $n$ and $p$?**

For binomial random variable $X$,

$X_i$ are independent, and $p_i$ are the same, denoted as $p$

$$\left| P(X \in A) - \sum_{i \in A} \frac{\lambda^i e^{-\lambda}}{i!} \right| \le \sum_{i=1}^{n} p_i \cdot \min\{1, 1/\lambda\} \cdot E[|X - V_i|]$$

$$= \min\{np^2, p\}$$

by independence

$E[|X_i|] = p_i$

# Exponential distribution

A continuous random variable $X$ is said to have an **exponential distribution with parameter $\boldsymbol{\lambda}$, $\boldsymbol{\lambda > 0}$**, if its *probability density function* is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

or, equivalently, if its *cumulative distribution function* is

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Expectation: $E[X] = 1/\lambda$ Variance: $Var[X] = 1/\lambda^2$

# Exponential distribution

An exponentially distributed random variable $X$ has the **memoryless** property:

$$\forall s, t \geq 0, P(X > s + t \mid X > t) = P(X > s)$$

**Proof:**

$$P(X > s + t \mid X > t) = \frac{P(X > s + t, X > t)}{P(X > t)} = \frac{P(X > s + t)}{P(X > t)}$$

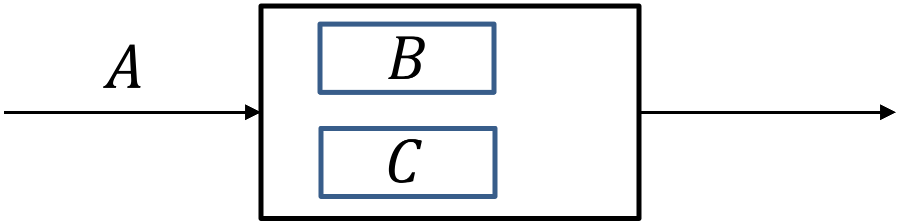
$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s)$$

# Exponential distribution

An exponentially distributed random variable $X$ has the **memoryless** property:

$$\forall s, t \geq 0, P(X > s + t \mid X > t) = P(X > s)$$

**Application:** Consider a post office having two clerks, and suppose that when $A$ enters the system, $B$ and $C$ are being served by these two clerks, respectively. Suppose also that $A$ will begin to be served once either $B$ or $C$ leaves. If the amount of time a clerk spends with a customer is exponentially distributed with parameter $\lambda$, what is the probability that $A$ is the last to leave the post office?




$1/2$

# Failure (Hazard) rate function

Consider a non-negative, continuous random variable $X$ having distribution $F$ and density $f$, let $\bar{F}(x) = P(X > x) = 1 - F(x)$.

The **failure (or hazard) rate function $\boldsymbol{\lambda(t)}$** is defined by

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)}$$

**Intuitive explanation:** Let $X$ denote the lifetime of some item

$$P(X \in (t, t + dt) \mid X > t) = \frac{P(x \in (t, t + dt), X > t)}{P(X > t)}$$

$$\approx \frac{f(t) \cdot dt}{\bar{F}(t)} = \lambda(t) \cdot dt$$

# Failure (Hazard) rate function

Failure rate function $\lambda(t)$ for the exponential distribution:

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

The failure rate function $\lambda(t)$ uniquely determines the cumulative distribution $F$

**Proof:**

By definition of $\lambda(t)$, $\lambda(t) = \frac{f(t)}{\bar{F}(t)} = \frac{-\frac{d\bar{F}(t)}{dt}}{\bar{F}(t)}$

Then, $\int_0^t -\lambda(t)dt = \log \bar{F}(t)\Big|_0^t = \log \bar{F}(t) - \log \boxed{\bar{F}(0)}$ $\quad = P(X > 0) = 1$

Thus, $\bar{F}(t) = e^{-\int_0^t \lambda(t)dt}$

# Probability inequalities

**Markov inequality:** If $X$ is a nonnegative random variable, then for any $a > 0$,

$$P(X \geq a) \leq \frac{E[X]}{a}$$

**Proof:**

Let $Y = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$, then $Y \leq \frac{X}{a}$

Thus, $E[Y] \leq \frac{E[X]}{a}$

# Probability inequalities

**Chernoff bound:** If $X$ is random variable, then for $a > 0$,

$$P(X \geq a) \leq e^{-ta}E[e^{tX}] \text{ for all } t > 0$$

$$P(X \leq a) \leq e^{-ta}E[e^{tX}] \text{ for all } t < 0$$

**Proof:**

For all $t > 0$: $P(X \geq a) = P(tX \geq ta) = P(e^{tX} \geq e^{ta})$

$$\leq \frac{E[e^{tX}]}{e^{ta}}$$

by Markov inequality

For all $t < 0$: $P(X \leq a) = P(tX \geq ta) = P(e^{tX} \geq e^{ta})$

$$\leq \frac{E[e^{tX}]}{e^{ta}}$$

# Probability inequalities

**Chernoff bound:** If $X$ is random variable, then for $a > 0$,

$$P(X \geq a) \leq e^{-ta} E[e^{tX}] \text{ for all } t > 0$$

$$P(X \leq a) \leq e^{-ta} E[e^{tX}] \text{ for all } t < 0$$

**Application:** If $X$ is Poisson with mean $\lambda$, i.e.,

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \text{ for } k = 0,1,2, \ldots$$

$$P(X \geq j) \leq e^{-tj} E[e^{t\alpha}] = \sum_{k=0}^{\infty} e^{tk} \cdot \frac{X^k e^{-\lambda}}{k!}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} \cdot e^{\lambda e^t} = \boxed{e^{\lambda(e^t - 1) - tj}}$$

$(\lambda e^t - j) \cdot e^{\lambda(e^t - 1) - tj}$ derivative

Let $t = \ln \frac{j}{\lambda}$ (assume $j > \lambda$), then $P(X \geq j) \leq e^{j-\lambda} \cdot \left(\frac{j}{\lambda}\right)^{-j} = \frac{e^{-\lambda}(e\lambda)^j}{j^j}$

# Probability inequalities

**Jensen's inequality:** If $f$ is a convex function, then

$$E[f(X)] \geq f(E[X])$$

provided the expectations exist

**Proof:**

Let $\mu = E[X]$

$\geq 0$ because $f$ is convex

Then, $$f(X) = f(\mu) + f'(\mu)(X-\mu) + \frac{\boxed{f''(\xi)}}{2}(X-\xi)^2$$
$$\geq f(\mu) + f'(\mu)(X-u)$$

Thus, $$E(f(X)] \geq f(\mu) + f'(\mu)(E[X]-\mu) = f(\mu) = f(E[X])$$

# Limit theorems

**Strong Law of Large Numbers:** If $X_1, X_2, \ldots$ are independent and identically distributed with mean $\mu$, then

$$P\left(\lim_{n\to\infty}(X_1 + \cdots + X_n)/n = \mu\right) = 1$$

**Central Limit Theorem:** If $X_1, X_2, \ldots$ are independent and identically distributed with mean $\mu$ and variance $\sigma^2$, then

$$\lim_{n\to\infty} P\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

# Summary

- What is stochastic process
- Stochastic process in AI
- Preliminaries

**References:** Chapter 1 & 10, Stochastic Processes, 2nd edition, 1995, *by Sheldon M. Ross*