第六章 伯努利过程和泊松过程

Probability and Statistics for Computer Scientists

第六章 伯努利过程和泊松过程

- 伯努利过程分析
- 泊松过程分析

§1 伯努利过程分析

与伯努利过程相关的随机变量及其性质

- **服从参数为 n 和 p 的二项分布.** 这是 n 次相继独立的试验成功的总次数 S 的分布. 它的分布列, 期望和方差是

$$p_S(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n,$$

$$\mathrm{E}[S] = np, \quad \mathrm{var}(S) = np(1-p).$$

§1 伯努利过程分析

- **服从参数为 p 的几何分布.** 相互独立重复的伯努利试验首次成功的总次数 T 的分布. 它的分布列, 期望和方差是

$$p_T(t) = p(1-p)^{t-1}, \quad t = 1, 2\cdots,$$

$$\mathrm{E}[T] = \frac{1}{p}, \quad \mathrm{var}(T) = \frac{1-p}{p^2}.$$

§1 伯努利过程分析

1.1 独立性和无记忆性

与伯努利过程相关的独立性质

- 对任意给定的时间 n, 随机变量序列 $X_{n+1}, X_{n+2}, \cdots$ (过程的将来) 也是伯努利过程, 而且与 $X_1, \cdots, X_n$ (过程的过去) 独立.
- 对任意给定的时间 n, 令 $\bar{T}$ 是时间 n 之后首次成功的时间, 则随机变量 $\bar{T} - n$ 服从参数为 p 的几何分布, 且与随机变量 $X_1, \cdots, X_n$ 独立.

§1 伯努利过程分析

例 6.2 计算机执行的任务分为两类：优先任务和非优先任务. 计算机将运行时间划分为互相连接的时间小区间，每个小区间称为"瞬间"(slot)，时间区间就实现了离散化. 计算机在每一个瞬间只有两个状态：忙碌或空闲. 这样计算机运行状态形成一个随机过程. 假定各个瞬间的忙闲是相互独立的. 又假定在每个瞬间的开始，优先任务以概率 p 到达，而且与其他瞬间是独立的. 当优先任务到达的时候，计算机执行优先任务，处于忙碌的状态. 非优先任务总是处于等待状态，只有在没有优先任务的前提下，才会执行. 当计算机执行非优先任务的时候，称计算机处于空闲的状态. 这样计算机在各瞬间的状态形成一个随机过程.

§1 伯努利过程分析

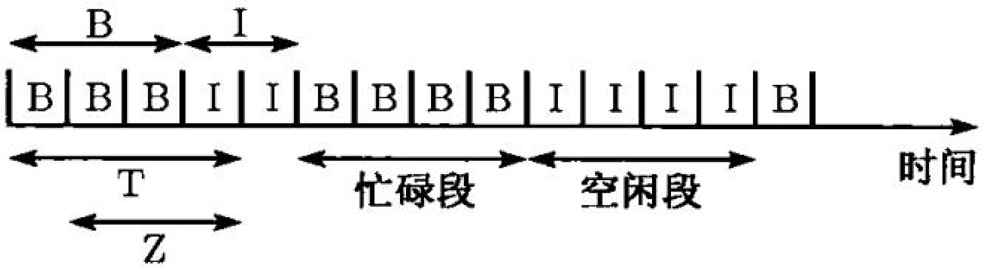
顺序相连的瞬间形成的时间区间称为段, 段的长度就是这个时间区间内的瞬间数.

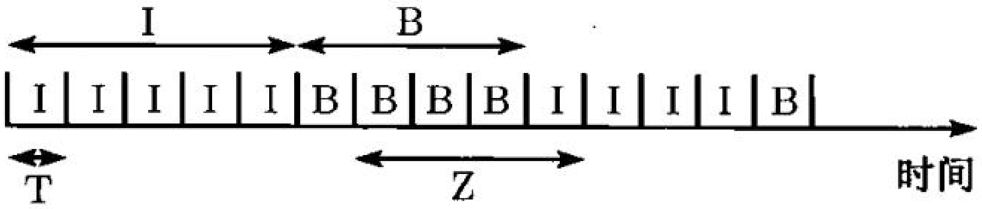
(a) $T=$ 首个空闲瞬间的时间指标;

(b) $B=$ 首个忙碌段的时间长度 (即忙碌段中含有的忙碌瞬间的个数);

(c) $I=$ 首个空闲段的时间长度;

(d) $Z=$ 第一个忙碌瞬间之后直到出现首个空闲瞬间的瞬间数 (含这个空闲瞬间, 但不含第一个忙碌瞬间).





§1 伯努利过程分析

T 是服从参数为 $1-p$ 的几何分布随机变量, 其分布列是

$$p_T(k) = p^{k-1}(1-p), \qquad k = 1, 2, \cdots,$$

均值和方差是

$$\mathrm{E}[T] = \frac{1}{1-p}, \quad \mathrm{var}(T) = \frac{p}{(1-p)^2}.$$

我们考虑第一个忙碌时间段. 起始于第一个忙碌瞬间, 称之为瞬间 L

直到出现下一个空闲瞬间 (包括这个瞬间) 的瞬间数 Z

$Z = B$, 所以 B 与 T 一样, 具有相同的分布列.

§1 伯努利过程分析

如果我们将空闲瞬间和忙碌瞬间的位置对换, 把 p 换成 $1-p$, 则第一个空闲段的长度 I 与第一个忙碌段的长度具有一样的分布列, 所以

$$p_I(k) = (1-p)^{k-1}p, \quad k = 1, 2, \cdots, \quad \mathrm{E}[I] = \frac{1}{p}, \quad \mathrm{var}(I) = \frac{1-p}{p^2}.$$

最后注意到上述结论对第 2, 3, 4 等忙碌 (或空闲) 段, 都是成立的. 所以计算得出的分布列也可以应用在任何第 i 个忙碌 (或空闲) 段. □

§1 伯努利过程分析

例 6.3 (随机时间的重新新开始) 设 N 是第一次遇到连续两次成功的时刻 (即, N 是满足 $X_i = X_{i-1} = 1$ 的第一个 i). 现求概率 $\mathrm{P}(X_{N+1} = X_{N+2} = 0)$, 即紧接着两次实验都失败的概率.

一旦条件 $X_{N-1} = X_N = 1$ 满足的话, 从那时开始, 未来的过程由独立的伯努利实验组成.

所以 $\mathrm{P}(X_{N+1} = X_{N+2} = 0) = (1-p)^2$.

§1 伯努利过程分析

现在对上述结论进行严格的证明

$$\begin{aligned}\mathrm{P}(X_{N+1}=X_{N+2}=0) &= \sum_{n=1}^{\infty}\mathrm{P}(N=n)\mathrm{P}(X_{N+1}=X_{N+2}=0|N=n)\\ &= \sum_{n=1}^{\infty}\mathrm{P}(N=n)\mathrm{P}(X_{n+1}=X_{n+2}=0|N=n).\end{aligned}$$

因为 N 确定后, 事件 $\{N=n\}$ 发生, 当且仅当 $X_1,\cdots,X_n$ 满足某个特定的条件, 而这些随机变量与 X_{n+1},X_{n+2} 是独立的, 所以

$$\mathrm{P}(X_{n+1}=X_{n+2}=0|N=n)=\mathrm{P}(X_{n+1}=X_{n+2}=0)=(1-p)^2.$$

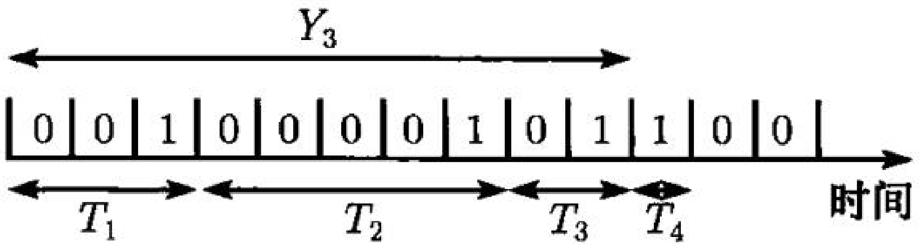
故

$$\mathrm{P}(X_{N+1}=X_{N+2}=0)=\sum_{n=1}^{\infty}\mathrm{P}(N=n)(1-p)^2=(1-p)^2. \qquad \square$$

§1 伯努利过程分析

1.2 相邻到达间隔时间

$$T_1 = Y_1, \quad T_k = Y_k - Y_{k-1}, \quad k = 2, 3, \cdots, \ Y_k = T_1 + \cdots + T_k.$$



伯努利过程另一种描述

(1) 开始于一串相互独立的，参数为 p 的几何分布随机变量序列 $T_1, T_2, \cdots$，它们是相邻到达时间间隔.

(2) 观测成功 (或到达) 的时间为 $T_1, T_1 + T_2, T_1 + T_2 + T_3$, 等等.

§1 伯努利过程分析

1.3 第k次到达的时间

第 k 次到达的时间的性质

- 第 k 次到达的时间等于前 k 个相邻到达时间之和

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

而且 $T_1, \cdots, T_k$ 独立同分布, 服从参数为 p 的几何分布.

- Y_k 的期望, 方差分别为

$$\mathrm{E}[Y_k] = \mathrm{E}[T_1] + \cdots + \mathrm{E}[T_k] = \frac{k}{p},$$

$$\mathrm{var}[Y_k] = \mathrm{var}[T_1] + \cdots + \mathrm{var}[T_k] = \frac{k(1-p)}{p^2}.$$

§1 伯努利过程分析

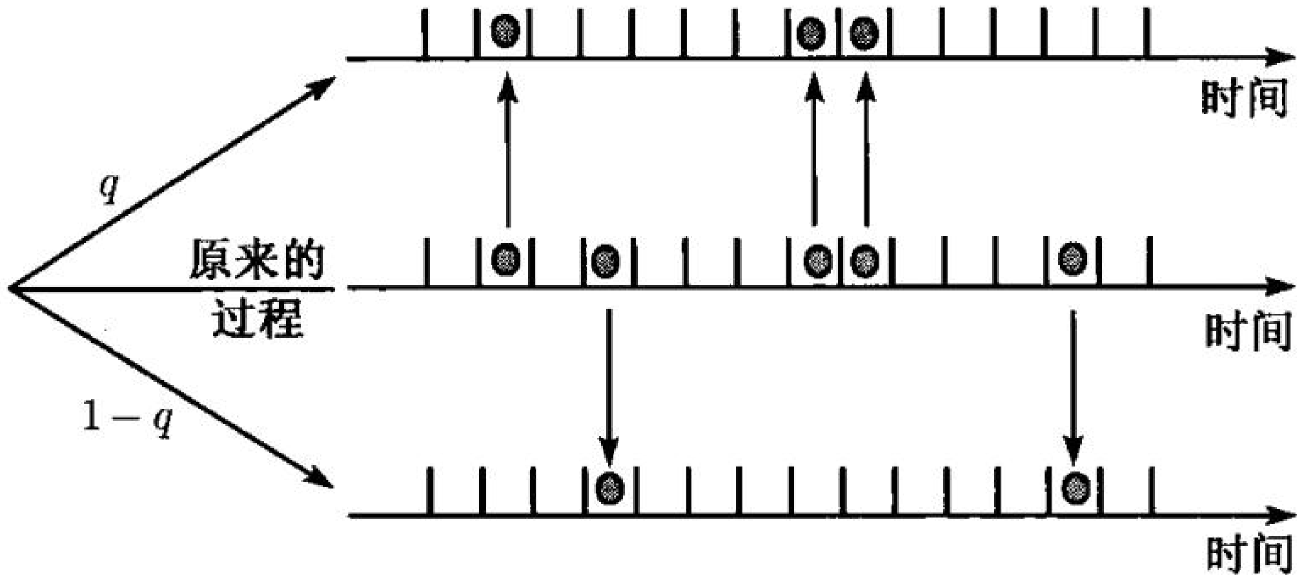
- Y_k 的分布列是

$$p_{Y_k}(t)=\binom{t-1}{k-1}p^k(1-p)^{t-k},\quad t=k,k+1,\cdots,$$

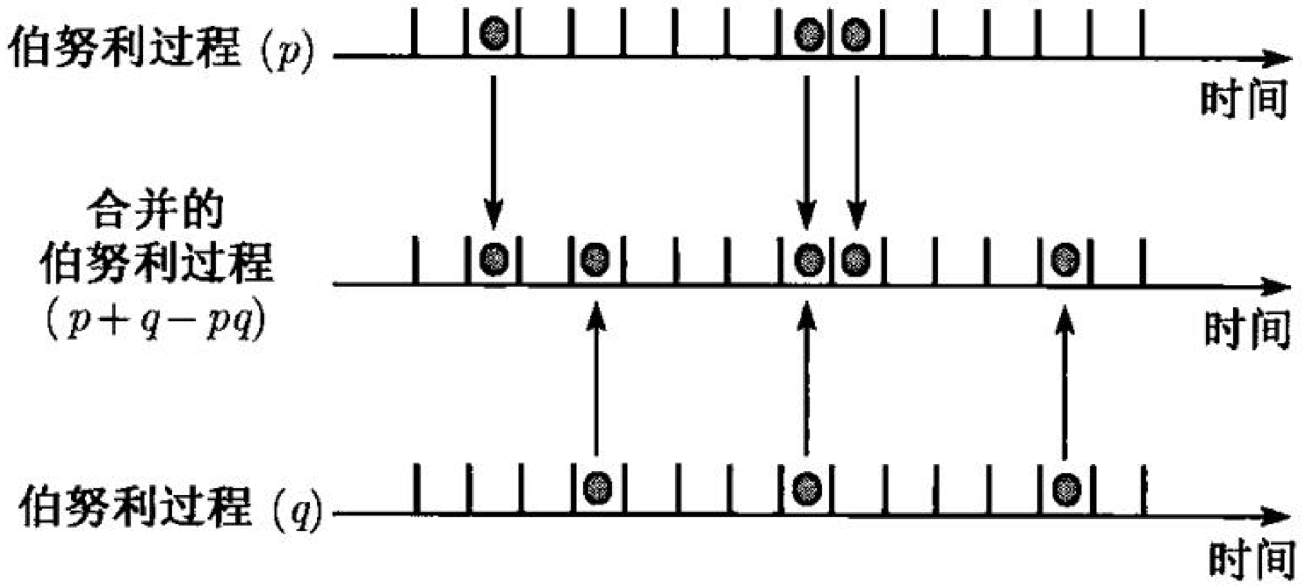
这就是有名的阶数为 k 的帕斯卡分布.

§1 伯努利过程分析

1.4 伯努利过程的分裂与合并



§1 伯努利过程分析



§1 伯努利过程分析

1.5 二项分布的泊松近似

二项分布的泊松近似

- 参数为 λ 的泊松分布的随机变量 Z 取非负整数值，其分布列如下

$$p_Z(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \cdots$$

均值和方差是

$$E[Z] = \lambda, \quad \mathrm{var}(Z) = \lambda.$$

- 当 $n \to \infty, p = \lambda/n$ 时，二项分布的概率

$$p_S(k) = \frac{n!}{(n-k)!k!} \cdot p^k(1-p)^{n-k}$$

收敛到 $p_Z(k)$，其中 λ 是常数，k 是任意的非负整数.

§1 伯努利过程分析

验证泊松定律. $\lambda = np$.

$$P_s(k) = \frac{n!}{(n-k)!\cdot k!}\cdot p^k\cdot(1-p)^{n-k}.$$

$$= \frac{n(n-1)\cdots(n-k+1)}{k!}\cdot\frac{\lambda^k}{n^k}\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n}{n}\cdot\frac{n-1}{n}\cdots\frac{n-k+1}{n}\cdot\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k}.$$

$$\xrightarrow{n\to\infty}: \left(1-\frac{\lambda}{n}\right)^{-k}\to 1 \qquad \left(1-\frac{\lambda}{n}\right)^n\to e^{-\lambda}.$$

$$\therefore\ P_s(k)\xrightarrow{n\to\infty} e^{-\lambda}\cdot\frac{\lambda^k}{k!}$$

§1 伯努利过程分析

例 6.7 有 n 个字符连成一串组成一个信息包, 在一个有噪声的通道中传输. 每个字符有 $p = 0.000\,1$ 的概率在传输中传错, 而且不同字符的传输过程是独立的. 问为保证在传输中发生错误的概率不超过 0.01, 这时 n 应该为多少?

每个字符的传输可视为一个独立的伯努利试验. 所以整个信息包发生错误传输的概率为

$$1 - \mathrm{P}(S = 0) = 1 - (1 - p)^n,$$

$$1 - (1 - 0.000\,1)^n < 0.01, \text{ 即 } n < \frac{\ln 0.999}{\ln 0.999\,9} = 10.004\,5.$$

§1 伯努利过程分析

同样我们也可使用泊松近似的方法来计算 $\mathrm{P}(S=0)$, 即 $\mathrm{P}(S=0)=\mathrm{e}^{-\lambda}$, 这里 $\lambda = np = 0.000\,1 \cdot n$. 由条件 $1-\mathrm{e}^{-0.000\,1\cdot n} < 0.001$, 可以得到

$$n < -\frac{\ln 0.999}{0.000\,1} = 10.005.$$

n 是一个整数, 两种方法都得出相同的结果: n 最多是 10. □

§2 泊松过程分析

$$P(k,\tau) = \mathrm{P}(\text{在时间段长度为 } \tau \text{ 的时间内, 有 } k \text{ 个到达}).$$

泊松过程的定义

一个到达过程, 被称为强度为 λ 的泊松过程, 如果该过程具有如下性质:

(a) (**时间同质性**) k 次到达的概率 $P(k,\tau)$ 在相同长度 τ 的时间内都是一样的.

(b) (**独立性**) 一个特定时间段里到达的数目与其他时间段里到达的历史是独立的.

§2 泊松过程分析

(c) (**小区间概率**) 概率 $\mathrm{P}(k,\tau)$ 满足如下关系

$$P(0,\tau)=1-\lambda\tau+o(\tau),$$

$$P(1,\tau)=\lambda\tau+o_1(\tau),$$

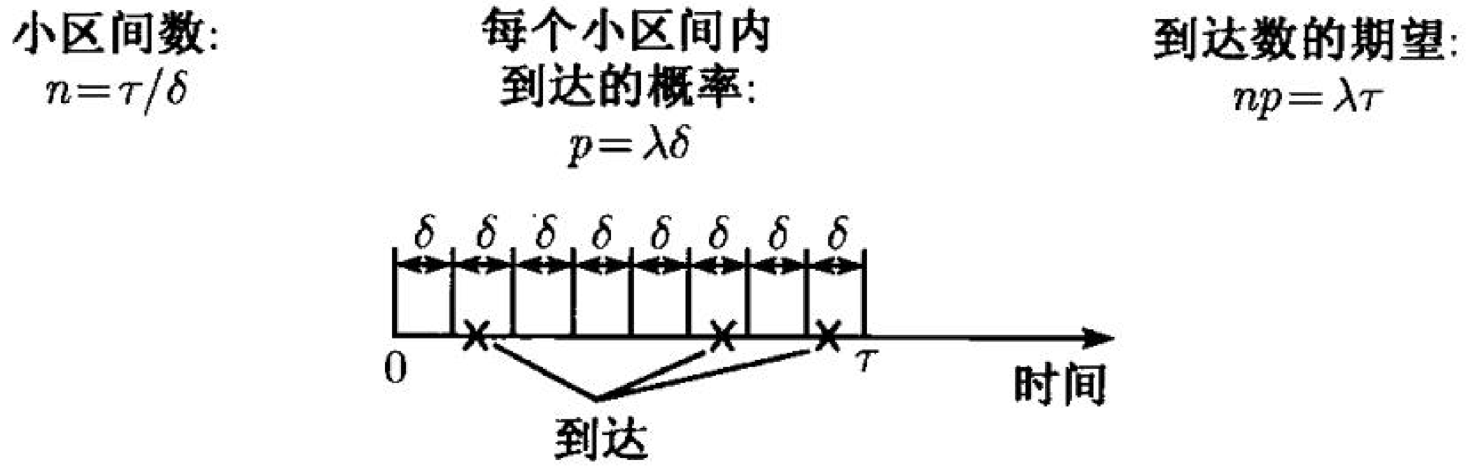
$$P(k,\tau)=o_k(\tau),\quad k=2,3,\cdots$$

这里 τ 的函数 $o(\tau)$ 和 $o_k(\tau)$ 满足

$$\lim_{\tau\to 0}\frac{o(\tau)}{\tau}=0,\quad \lim_{\tau\to 0}\frac{o_k(\tau)}{\tau}=0.$$

§2 泊松过程分析

2.1 区间内到达的次数



在时间 τ 到达 k 次的概率 $P(k,\tau)$ 近似地等于以每次实验成功概率为 $p=\lambda\delta$, 进行 $n=\tau/\delta$ 次独立伯努利试验, 而成功 k 次的 (二项) 概率.

§ 2 泊松过程分析

$$P(k,\tau)=\mathrm{e}^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!},\qquad k=0,1,\cdots.$$

注意, 由 $\mathrm{e}^{-\lambda\tau}$ 的泰勒展开, 可以得到

$$P(0,\tau)=\mathrm{e}^{-\lambda\tau}=1-\lambda\tau+0(\tau),$$

$$P(1,\tau)=\lambda\tau\mathrm{e}^{-\lambda\tau}=\lambda\tau-\lambda^2\tau^2+O(\tau^3)=\lambda\tau+o_1(\tau),$$

跟性质 (c) 相符.

利用泊松分布的均值和方差的公式, 可以得到

$$\mathrm{E}[N_\tau]=\lambda\tau,\quad \mathrm{var}(N_\tau)=\lambda\tau,$$

其中 N_τ 表示在时间长度为 τ 的时间段中到达的次数.

首次到达的时间 T $F_T(t)=\mathrm{P}(T\leqslant t)=1-\mathrm{P}(T>t)=1-P(0,t)=1-\mathrm{e}^{-\lambda t},\quad t>0.$

$$f_T(t)=\lambda\mathrm{e}^{-\lambda t},\qquad t>0.$$

§2 泊松过程分析

泊松过程相关的随机变量及其性质

- **服从参数为 $\lambda\tau$ 的泊松分布.** 这是泊松过程的强度为 λ, 在时间长度为 τ 的区间内到达的总次数 N_τ 的分布. 它的分布列, 期望和方差分别是

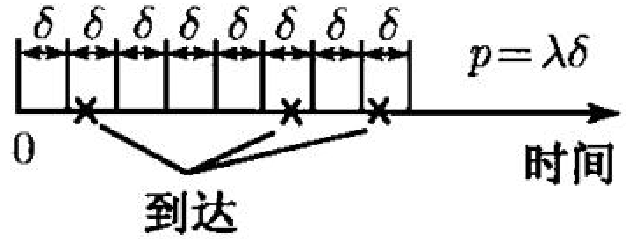
$$p_{N_\tau}(k) = P(k,\tau) = \mathrm{e}^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}, \quad k = 0, 1, \cdots,$$

$$\mathrm{E}[N_\tau] = \lambda\tau, \quad \mathrm{var}(N_\tau) = \lambda\tau.$$

- **服从参数为 λ 的指数分布.** 这是首次到达的时间 T 的分布. 它的分布列, 期望和方差是

$$f_T(t) = \lambda\mathrm{e}^{-\lambda t}, \quad t \geqslant 0, \quad \mathrm{E}[T] = \frac{1}{\lambda}, \quad \mathrm{var}(T) = \frac{1}{\lambda^2}.$$

§2 泊松过程分析



	泊松	伯努利
到达时间	连续	离散
到达次数的分布	泊松	二项
相邻到达时间的分布	指数	几何
到达率	λ/单位时间	p/每次试验

§2 泊松过程分析

2.2 独立性和无记忆性

泊松过程的独立性质

- 对任意给定的时间 $t>0$, 时间 t 之后的过程也是泊松过程, 而且与时间 t 之前 (包括时间 t) 的历史过程相互独立.
- 对任意给定的时间 t, 令 $\bar{T}$ 是时间 t 之后首次到达的时间, 则随机变量 $\bar{T}-t$ 服从参数为 λ 的指数分布, 且与时间 t 之前 (包括时间 t) 的历史过程相互独立.

§2 泊松过程分析

2.3 相邻到达时间

泊松过程另一种描述

(1) 开始于一串相互独立并且公共参数为 λ 的指数随机变量序列 $T_1, T_2, \cdots$，它们是相邻到达时间.

(2) 过程的到达的时间为 $T_1, T_1+T_2, T_1+T_2+T_3$, 等等. 这样形成的随机过程就是泊松过程.

§2 泊松过程分析

2.4 第k次到达的时间

第 k 次到达的时间的性质

- 第 k 次到达的时间等于前 k 个相邻到达时间之和

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

  而且 $T_1, \cdots, T_k$ 独立同分布，服从参数为 λ 的指数分布.

- Y_k 的期望、方差为

$$\mathrm{E}[Y_k] = \mathrm{E}[T_1] + \cdots + \mathrm{E}[T_k] = \frac{k}{\lambda},$$

$$\mathrm{var}(Y_k) = \mathrm{var}(T_1) + \cdots + \mathrm{var}(T_k) = \frac{k}{\lambda^2}.$$

§2 泊松过程分析

- Y_k 的分布密度是

$$f_{Y_k}(y)=\frac{\lambda^k y^{k-1}\mathrm{e}^{-\lambda y}}{(k-1)!},\qquad y\geqslant 0,$$

这就是有名的阶数为 k 的埃尔朗分布②.

§2 泊松过程分析

证明 Y_k 的分布密度公式

当下面两个事件同时发生.

(a) 事件 A: 在时间段 $[y, y+\delta]$ 到达了一次;

(b) 事件 B: 在时间 y 之前恰好发生了 $k-1$ 次.

这两个事件发生的概率分别是

$$\mathrm{P}(A) \approx \lambda\delta,$$

$$\mathrm{P}(B) = P(k-1, y) = \frac{\lambda^{k-1} y^{k-1} \mathrm{e}^{-\lambda y}}{(k-1)!}.$$

事件 A 与 B 是相互独立的, 所以

$$\delta f_{Y_k}(y) \approx \mathrm{P}(y \leqslant Y_k \leqslant y+\delta) \approx \mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) \approx \lambda\delta \frac{\lambda^{k-1} y^{k-1} \mathrm{e}^{-\lambda y}}{(k-1)!},$$

所以

$$f_{Y_k}(y) = \frac{\lambda^{k} y^{k-1} \mathrm{e}^{-\lambda y}}{(k-1)!}, \quad y \geqslant 0.$$

§2 泊松过程分析

2.5 泊松过程的分裂与合并

例 6.15 (竞争指数) 两个灯泡[①]具有独立的寿命 T_a 和 T_b, 它们分别服从参数为 λ_a 和 λ_b 的指数分布. 问两个灯泡首次烧坏的时间 $Z=\min\{T_a,T_b\}$ 的分布是什么?

对任意的 $z\geqslant 0$, 有

$$\begin{aligned}
F_Z(z) &= \mathrm{P}(\min\{T_a,T_b\}\leqslant z)\\
&= 1-\mathrm{P}(\min\{T_a,T_b\}>z)\\
&= 1-\mathrm{P}(T_a>z,T_b>z)\\
&= 1-\mathrm{P}(T_a>z)\mathrm{P}(T_b>z)\\
&= 1-\mathrm{e}^{-\lambda_a z}e^{-\lambda_b z}\\
&= 1-\mathrm{e}^{-(\lambda_a+\lambda_b)z}.
\end{aligned}$$

假设 T_a 和 T_b 分别是强度为 λ_a 和 λ_b 的泊松过程 首次到达的时间是 $\min\{T_a,T_b\}$.

§2 泊松过程分析

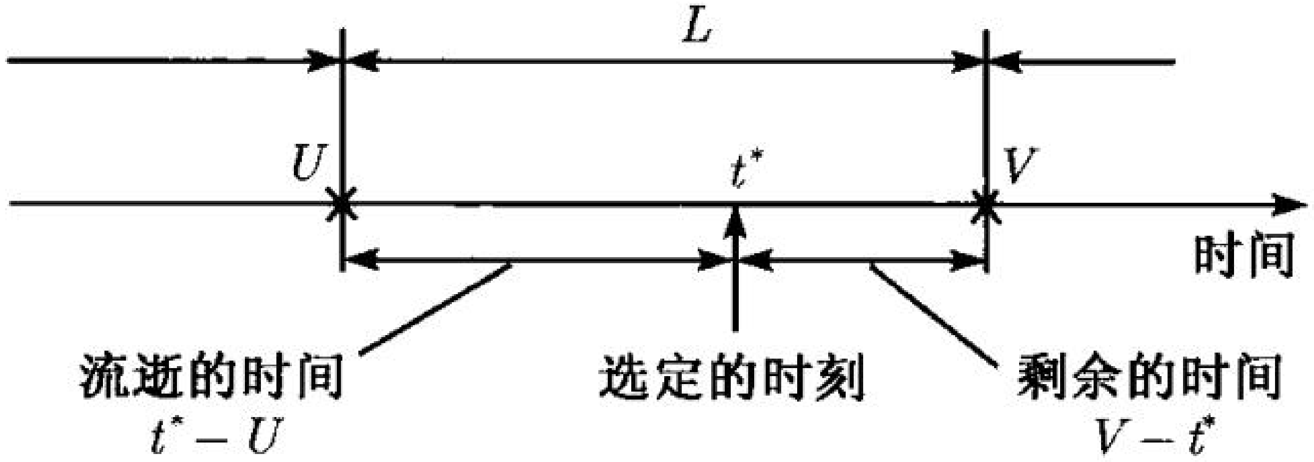
随机数个独立随机变量和的性质

设 $N, X_1, \cdots, X_n$ 是独立随机变量, 其中 N 取非负整数. 当 $N>0$ 时, 定义 $Y=X_1+\cdots+X_N$, 当 $N=0$ 时, 定义 $Y=0$.

- 如果 X_i 的分布是参数为 p 的伯努利分布, N 的分布是参数为 m 和 q 的二项分布, 则 Y 的分布是参数为 m 和 pq 的二项分布.
- 如果 X_i 的分布是参数为 p 的伯努利分布, N 的分布是参数为 λ 的泊松分布, 则 Y 的分布是参数为 λp 的泊松分布.
- 如果 X_i 的分布是参数为 p 的几何分布, N 的分布是参数为 q 的几何分布, 则 Y 的分布是参数为 pq 的几何分布.
- 如果 X_i 的分布是参数为 λ 的指数分布, N 的分布是参数为 q 的几何分布, 则 Y 的分布是参数为 λq 的指数分布.

§2 泊松过程分析

2.6 随机插入的悖论



作业

8. 从早上 8 点到 9 点这段繁忙时间里, 交通事故的发生数服从一个强度为每小时 5 次的泊松分布, 在早上 9 点到 11 点之间, 交通事故的发生数服从一个独立的频率为每小时 3 次的泊松分布. 试求：早上 8 点到 11 点之间发生事故总次数的分布函数.