随机过程
Stochastic Processes

讲义8-9：泊松过程-1

目录

（课本第5章部分1）

8.1指数分布
(课本5.1-5.2)

8.1.1 指数分布

- 一个常见假定：某些随机变量服从指数分布
  - 非负的连续随机变量
  - 指数分布常用来描述事件间的时间间隔
  - 指数分布容易处理，且常是实际分布的一个良好近似
  - 指数分布的一个重要性质即无记忆性，不随时间改变
  - 指数分布是具备这个性质的唯一连续分布
  - 指数分布是泊松过程的基础

8.1.2 指数分布定义

- 概率密度函数：参数λ的指数分布X,概率密度函数为

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

  - ➢ 一般用于描述连续时间的长度

- 累积分布函数：

$$\mathrm{F}(x)=\int_{-\infty}^{x} f(y)dy=\begin{cases}1-e^{-\lambda x}, & x\ge 0\\ 0 & ,x<0\end{cases}$$

  - ➢ 注意到：$P\{X\le x\}=1-e^{-\lambda x}$
  - ➢ 注意到：$P\{X>x\}=1-P\{X\le x\}=e^{-\lambda x}$

8.1.2 指数分布定义

- 概率密度函数：参数λ的指数分布X,概率密度函数为

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

- 均值：

$$\mathrm{E}(X)=\int_{-\infty}^{\infty} xf(x)dx=\int_{0}^{\infty} x\lambda e^{-\lambda x}dx=\frac{1}{\lambda}$$

  - 参数λ常被称为速率参数
    - λ越大，间隔时间倾向于越短

- 矩母函数：

$$\phi(t)=\mathrm{E}(e^{tX})=\int_{0}^{\infty} e^{tX}\lambda e^{-\lambda x}dx=\frac{\lambda}{\lambda-t}$$

- 方差：

$$Var(X)=E[X^2]-E[X]^2=\frac{1}{\lambda^2}$$

  - $E[X^2]=\frac{d^2}{dt^2}\phi(t)|_{t=0}=\frac{2\lambda}{(\lambda-t)^3}|_{t=0}=\frac{2}{\lambda^2}$

8.1.2 指数分布定义-例子

- **例5.1**：考虑随机变化的报酬流$R(x)$，x指的是时刻。考虑一个折扣率$\alpha \geq 0$, 总折扣报酬R可表示为

$$R = \int_0^{\infty} e^{-\alpha x} R(x) dx$$

  - α常被称为连续复合利率
  - 总折扣报酬是无穷报酬流 $R(x)$ 的折现值（通过利率 α ）

- 可知，平均总折扣报酬$E[R]$可表示为

$$E[R] = \int_0^{\infty} e^{-\alpha x} E[R(x)] dx$$

8.1.2 指数分布定义–例子

●进一步，令T是以α为速率的指数随机变量，且独立于$R(x)$；请推导如下结论：

$$E\left[\int_0^T R(x)dx\right] = \int_0^\infty e^{-\alpha x} E[R(x)]dx$$

结论说明：

- 平均总折扣报酬(等式右侧) = 在以α为速率的指数随机时间里所得的期望总报酬（无折扣报酬，等式左侧）

推导：

- 要点：对所有$x \geq 0$，构建$I(x) = \begin{cases} 1, 若 x \leq T \\ 0, 若 x > T \end{cases}$
- 所以：$\int_0^T R(x)dx = \int_0^\infty R(x)I(x)dx$

8.1.2 指数分布定义-例子

- $E\left[\int_0^\infty R(x)I(x)dx\right] = \int_0^\infty E[R(x)I(x)]dx$

$$= \int_0^\infty E[R(x)]E[I(x)]dx$$

$$= \int_0^\infty E[R(x)]P\{T \geq x\}dx$$

$$= \int_0^\infty e^{-\alpha x}E[R(x)]dx$$

8.1.3 随机变量的无记忆性

- 无记忆性：$\forall s,t\geq 0$, 随机变量X均满足

$$P\{X>s+t|X>t\}=P\{X>s\}$$

说明，随机变量X具备无记忆性。

上述条件，等价于:

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

也等价于:

$$P\{X>s+t\}=P\{X>t\}P\{X>s\}$$

- 重要结论：指数分布有无记忆性，且是唯一具有这种性质的连续分布。

本课仅证明无记忆性：

➢ 回忆：$P\{X>s\}=e^{-\lambda s}$，由等价条件二可得

8.1.3 无记忆性的重要结论

- 基于无记忆性可推得：
  - $P\{X<s+t|X>t\}=P\{(X-t)<s|X>t\}$
    $=1-P\{X>s+t|X>t\}$
    $=P\{X<s\}$
  - 得到要点：条件于指数随机变量X大于某值，则超出该值的部分依然服从指数分布，且分布与原指数分布相同
  - 也即条件于$X>\mathrm{t}$, $(X-t)^{+}\sim Exp(\lambda)$

8.1.3 无记忆性的拓展

- 无记忆性：$\forall s, t \geq 0$, 随机变量X满足

$$P\{X > s + t | X > t\} = P\{X > s\}$$

  - 拓展1：t可以为随机变量，即T为非负随机变量，亦有

$$P\{X > s + T | X > T\} = P\{X > s\}$$

  - 拓展2：二元（多元）无记忆性，对于独立指数随机变量X, Y,有：

$$P\{X > s_1 + t_1, Y > s_2 + t_2 | X > t_1, Y > t_2\} \\ = P\{X > s_1\}P\{Y > s_2\}$$

8.1.3 无记忆性的例子

- 无记忆性：$\forall \mathrm{s}, \mathrm{t} \geq 0$, 随机变量$X$满足

$$P\{X > s + t | X > t\} = P\{X > s\}$$

  - 例1：设备寿命，给定已存活t小时，剩余存活时间仍服从原始寿命的指数分布

  - 例2：杰伦奶茶店主理人等顾客，已经等了3个小时，那么等到下一位顾客需要的额外时间，仍服从原始指数分布

8.1.3 无记忆性的例题

- 例5.2：假设顾客在银行时间服从均值10分钟的指数分布，求
（a）顾客用时超过15分钟的概率？
（b）假定顾客已用时10分钟，那么她用时超过15分钟的概率？

求解

- 定义X为顾客在银行的时间，有$X \sim Exp(\lambda) = Exp(0.1)$
- 问题（a）转化为概率公式：
$$P\{X > 15\}$$
- 问题（b）转化为概率公式：
$$P\{X > 10 + 5 | X > 10\} = P\{X > 5\}$$
  - 也即条件于$X > 10, (X - 10)^{+} \sim Exp(\lambda)$

8.1.3无记忆性的例题

- 例5.4：假设一次汽车事故损失金额数量是均值为1000的指数随机变量，假设保险公司仅赔付超出400的部分（亦称可扣除金额）。求保险公司赔付金额的期望和标准差。

求解：

- 令X是由一次事故导致的损失的金额数量，那么保险公司赔付金额为$(X-400)^+$
- 注意到条件于$X>400, Y=(X-400)^+ \sim Exp(1/1000)$
- 令$I=\begin{cases}1, 若X>400\\0, 若X\leq 400\end{cases}$， I为示性随机变量
  - $E[I]=P\{I=1\}=P\{X>400\}=e^{-0.4}$
- 希望得到：$E[Y|I]$，$Var[Y|I]$
  - $E[Y|I=1]=1000, E[Y|I=0]=0$
  - $Var[Y|I=1]=1000^2, Var[Y|I=0]=0$
- 所以：$E[Y|I]=1000I$; $Var[Y|I]=10^6 I$
- 可求$E[Y]$和 Var[Y] 请补齐

8.1.3 无记忆性的例题

- 例5.3：假设杰伦奶茶店有两个奶茶师傅，当你走入奶茶店时，发现俊杰和阿信正在接受两位师傅的服务，你被告知，只要俊杰或阿信任一人离开，你的服务会立刻开始。假设每位奶茶师傅服务一个顾客的时间服从均值为$\frac{1}{\lambda}$的指数分布。那么在你、俊杰、阿信三位顾客中，你是最后一位离开的概率？

解答：

- 考虑你首先发现一个柜员有空的时间，此时俊杰或者阿信刚刚离开，另外一个正在接受服务。

- 由指数分布的无记忆性推出，未走的这个人再花费在银行的额外时间仍旧服从均值为$1/\lambda$的指数分布。

- 你花费在银行的时间也服从均值为$1/\lambda$的指数分布。所以对称的，你在剩余的这位顾客之前完成服务的机会是1/2.

8.1.3 无记忆性还反映在失败率函数上

●失败/风险率函数$r(t)$：令正随机变量X具有分布函数F和密度f. 失败率函数 定义为：

$$r(t)=\frac{f(t)}{1-F(t)}$$

●理解：已存活t小时的部件X,考察它在附加时间dt内失效概率

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

$$=\frac{P\{X\in(t,t+dt)\}}{P\{X>t\}}\approx\frac{f(t)dt}{1-F(t)}=r(t)dt$$

- $r(t)$对应的是年龄t的部件（或人）在短时间内失败的速率
- 也称为风险率函数

8.1.3 无记忆性还反映在失败率函数上

- 指数随机变量的失败率函数为常数：

$$r(t)=\frac{f(t)}{1-F(t)}=\frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}=\lambda$$

  - 指数分布的无记忆性，必有这个结论
    - 由无记忆性可知，无论部件年龄多大，其剩余的寿命都服从相同的指数分布，所以任意时刻下，相应的短期内失败的速率应该是一样的。
  - 也说明了，λ称为速率，且对指数随机变量来说，为不同时间内都恒定的速率

8.1.3 无记忆性还反映在失败率函数上

- 结论：失败率函数唯一确定分布！

$$r(t)=\frac{f(t)}{1-F(t)}=\frac{\frac{d}{dt}F(t)}{1-F(t)}$$

基于失败率函数求解分布函数：

- 两边求积分，$\ln\big(1-F(t)\big)=-\int_0^t r(t)dt+C$
- $1-F(t)=e^C\exp\{-\int_0^t r(t)dt\}$
- 应用边缘条件$F(0)=0$，即，令$t=0$, 得到 $\mathrm{C}=0$
- 得到：$F(t)=1-\exp\{-\int_0^t r(t)dt\}$
  - 无记忆性分布必须是指数分布！
  - 因为若X无记忆，其失败率函数$r(t)$，必是常数
  - 若$r(t)$为常数，则$F(t)=1-e^{-ct}$，确定X为指数随机变量
- 这是后半学期的常用技能：常微分方程求解

8.1.3 失败率函数例子-超指数随机变量

- 例5.6.令$X_1,\dots,X_n$是速率为不同$\lambda_i, i=1,\dots,n$的独立指数随机变量。令T独立于这些随机变量，且假设

$$\sum_{j=1}^{n} P_j = 1, \text{其中} P_j = P\{T=j\}$$

- X_T称为超指数随机变量,有如下结论

$$f(t) = \sum_{i=1}^{n} \lambda_i P_i e^{-\lambda_i t}$$

$$F(t) = 1 - \sum_{i=1}^{n} P_i e^{-\lambda_i t}$$

$$r(t) = \frac{\sum_{j=1}^{n} \lambda_j P_j e^{-\lambda_j t}}{\sum_{i=1}^{n} P_i e^{-\lambda_i t}} = \sum_{j=1}^{n} \lambda_j P\{T=j | X>t\}$$

$$\lim_{t\to\infty} r(t) = \min \lambda_i$$

场景理解：

- 从n种电池中，随机抽取一个电池T，抽到类型i的概率为P_i，这个电池的寿命是X_T
- 如何保证抽到类型i的概率为P_i呢？
  - 罐子里类型i的电池比例为P_i，均匀随机抽即可

8.1.3 失败率函数例子-超指数随机变量

求解（详细作为课后任务请自行补充）：

先求 $F(t)$，$f(t)$，$r(t)$：

- 求 $F(t)$ 即可得到所有结果
- 条件于T，使用全概率公式！

  $$1-F(t)=P\{X>t\}=\sum_{i=1}^{n}P\{X>t|T=i\}P\{T=i\}$$

- 可得 $F(t)$，$f(t)$，$r(t)$

再得：$r(t)$的另一个表达式

- 注意到：

$$P\{T=j|X>t\}=\frac{P\{X>t|T=j\}P\{T=j\}}{P\{X>t\}}=\frac{P_je^{-\lambda_jt}}{\sum_{i=1}^{n}P_ie^{-\lambda_it}}$$

- 可得：$r(t)=\sum_{j=1}^{n}\lambda_jP\{T=j|X>t\}$

8.1.3 失败率函数例子-超指数随机变量

求解（详细作为课后任务请自行补充）：

最后表示$r(t)$的极限：

- 注意到，若$\forall i>1$有$\lambda_1<\lambda_i$，那么：

$$P\{T=1|X>t\}=\frac{P_1e^{-\lambda_1 t}}{P_1e^{-\lambda_1 t}+\sum_{i=2}^{n}P_ie^{-\lambda_i t}}$$
$$=\frac{P_1}{P_1+\sum_{i=2}^{n}P_ie^{-(\lambda_i-\lambda_1)t}}$$

- 所以：$\lim_{t\to\infty}P\{T=1|X>t\}=1$
- 类似可得：对$i\neq 1$，$\lim_{t\to\infty}P\{T=i|X>t\}=0$（请自行尝试一个$i=2$）
- 最终得到：

$$\lim_{t\to\infty}r(t)=\min\lambda_i$$

- 当t趋向于无穷，说明死亡需要很多时间，说明越可能是选中了那类最小失败率的电池！

8.1.4 指数分布的几个性质

- **结论1-i.i.d.**指数随机变量加和：令$X_1,\ldots,X_n$是均值$\frac{1}{\lambda}$的i.i.d. 指数随机变量，则

$$X_1+X_2+\cdots+X_n \text{服从参数为} n \text{和} \lambda \text{的伽马分布。}$$

  - 第一次作业题目之一

8.1.4 指数分布的几个性质

- **结论2-n个独立随机变量的最小值**

令$X_1, \dots, X_n$是速率$\lambda_i, i = 1, \dots, n$的独立指数随机变量，则

$$\min(X_1, \dots, X_n) \sim Exp(\sum_{i=1}^{n} \lambda_i)$$

推导：

- $$\begin{aligned} P\{\min(X_1, \dots, X_n) > x\} &= P\{X_1 > x, \dots, X_n > x\} \\ &= \prod_{i=1}^{n} P\{X_i > x\} \quad \text{独立性} \\ &= \exp\{-(\sum_{i=1}^{n} \lambda_i) x\} \end{aligned}$$

8.1.4 指数分布的几个性质

- 结论3-指数随机变量小于另一个的概率

令X_1, X_2是均值$\frac{1}{\lambda_1}, \frac{1}{\lambda_2}$的独立指数随机变量，则

$$P\{X_1 < X_2\} = \frac{\lambda_1}{\lambda_1+\lambda_2}$$

推导：连续随机变量的全概率公式

- $P\{X_1 < X_2\}$

$$= \int_{-\infty}^{\infty} P\{X_1 < X_2 | X_1 = x\} \lambda_1 e^{-\lambda_1 x} dx$$

$$= \int_{-\infty}^{\infty} P\{x < X_2 | X_1 = x\} \lambda_1 e^{-\lambda_1 x} dx$$

$$= \int_{-\infty}^{\infty} P\{x < X_2\} \lambda_1 e^{-\lambda_1 x} dx \qquad \text{独立性}$$

$$= \frac{\lambda_1}{\lambda_1+\lambda_2}$$

8.1.4 指数分布的几个性质

- **结论4-** X_i是n个独立随机变量最小值的概率：令$X_1, \ldots, X_n$是速率$\lambda_i, i = 1, \ldots, n$的独立指数随机变量，$X_i$是其中最小的的概率为

$$\lambda_i / \sum_{j=1}^{n} \lambda_j$$

推导：

- $P\left\{X_i = \min_{j} X_j\right\} = P\left\{X_i < min_{j \neq i} X_j\right\}$

$$= \frac{\lambda_i}{\lambda_i + \sum_{j \neq i} \lambda_j}$$ 结论2+结论3

$$= \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$$

8.1.4 指数分布的几个性质

- **结论5**　　　$\min_i X_i$与X_i的大小次序独立

理解：

> $P\left\{X_{i_1} < \cdots < X_{i_n} \mid \min_i X_i > t\right\}$
> $= P\{X_{i_1} - t < \cdots < X_{i_n} - t \mid \min_i X_i > t\}$
> $= P\{X_{i_1} - t < \cdots < X_{i_n} - t \mid X_{i_1} > t, \ldots, X_{i_n} > t\}$
> $= P\{X_{i_1} < \cdots < X_{i_n}\}$

8.1.4 指数分布的几个性质

- 命题**5.1**：$X_1,\ldots,X_n$是速率$\lambda_1,\ldots,\lambda_n$的独立指数随机变量，则

$$\min(X_1,\ldots,X_n) \sim Exp(\textstyle\sum_{i=1}^{n}\lambda_i)$$

且

$$\min_i X_i \text{与} X_i \text{的大小次序独立}$$

➢ 结论2+结论5

8.1.4 指数分布的几个性质

- **例5.8.** 你到达银行时，银行两位独立的办事员都在服务一位用户，没人在排队。只要哪个办事员有空，你就会得到服务。若办事员i的服务时间是速率为$\lambda_i (i=1,2)$的指数分布. 令T为你待在银行的时间，求$E[T]$.

8.1.4 指数分布的几个性质

- **例5.8.** 你到达银行时，银行两位独立的办事员都在服务一位用户，没人在排队。只要哪个办事员有空，你就会得到服务。若办事员i的服务时间是速率为$\lambda_i(i=1,2)$的指数分布. 令T为你待在银行的时间，求$E[T]$.

求解：

要点：$T=\min(R_1,\ R_2)+S$

- 令R_1,R_2为两位用户在你进入银行后的剩余用时
  - 由无记忆性，可知$R_1\sim Exp(\lambda_1),R_2\sim Exp(\lambda_2)$,且独立
- 令S为你的服务时间

- $E(S)=E[S|R_1<R_2]P\{R_1<R_2\}+E[S|R_1\geq R_2]P\{R_1\geq R_2\}$
- $P\{R_1<R_2\}=\frac{\lambda_1}{\lambda_1+\lambda_2}$ $\qquad$ $P\{R_1\geq R_2\}=\frac{\lambda_2}{\lambda_1+\lambda_2}$
- $E[S|R_1<R_2]=\frac{1}{\lambda_1}$ $\qquad$ $E[S|R_2<R_1]=\frac{1}{\lambda_2}$
- $E[T]=\frac{3}{\lambda_1+\lambda_2}$

8.1.4 指数分布的几个性质

- **例5.7. (分配问题的贪婪算法)**
- n份工作分配给n个人，每人一份
- 费用是 $C(i,j),(i,j=1,\ldots,n)$, i指人，j指工作
- 假定$C(i,j)$构成一组n^2个独立指数随机变量，每个速率1.
- 标准：总费用最小
- 本例目标：比较以下两种算法。

1：按人$i=1,\ldots,n$顺序，分配未分配工作中费用最低的工作

2：全局算法，按min 未分配$C(i,j)$，逐对 安排（人,工作）

8.1.4 指数分布的几个性质

算法1：

- C_1（第一个人,对应n份工作最小值）$\sim \text{Exp}(n)$
- C_2（第二个人,对应的$n-1$份工作）$\sim \text{Exp}(n-1)$
  - C_1的情况，没有提供任意关于C_2的额外信息！
- C_3（第三个人,对应的$n-2$份工作）$\sim \text{Exp}(n-2)$

…

- 总期望费用E_1[总费用] $= E[C_1 + \cdots C_n] = \sum_{i=1}^{n} 1/i$

8.1.4 指数分布的几个性质

- **算法2：**
  - C_1（n^2份工作最小值）$\sim \mathrm{Exp}(n^2)$

  - C_2 (共考虑$(n-1)^2$个大于C_1的指数随机变量的最小值) $= C_1$+随机变量超出C_1的部分的最小值
    - C_2等于C_1加上$(n-1)^2$个速率1的独立指数变量的最小值
    - **核心思路：** C_1给出的信息，只是剩余$(n-1)^2$个指数随机变量都比C_1大
      - 而大于C_1的部分，仍然是独立的指数随机变量

  - $\mathrm{E}[C_2] = E[C_1] + \frac{1}{(n-1)^2} = \frac{1}{n^2} + \frac{1}{(n-1)^2}$

8.1.4 指数分布的几个性质

进一步，同理可推：

- C_3 (共考虑$(n-2)^2$个大于C_2的指数随机变量) $= C_2$+随机变量超出C_2的部分的最小值
  - $E[C_3] = E[C_2] + \frac{1}{(n-2)^2} = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \frac{1}{(n-2)^2}$

- C_4 (共考虑$(n-3)^2$个大于C_3的指数随机变量) $= C_3$+随机变量超出C_3的部分的最小值
  - $E[C_4] = E[C_3] + \frac{1}{(n-3)^2} = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \frac{1}{(n-2)^2} + \frac{1}{(n-3)^2}$

…

- 可知期望费用E_2[总费用] $= E[C_1 + \cdots C_n] = \frac{n}{n^2} + \frac{n-1}{(n-1)^2} + \frac{n-2}{(n-2)^2} + \cdots + 1 = \sum_{i=1}^{n} \frac{1}{i}$
  - 与算法1相等

8.1.5 指数随机变量的卷积

•亚指数随机变量：$X_1, \ldots, X_n$是速率不同$(\lambda_1, \ldots, \lambda_n)$的独立指数随机变量.

•$S = \sum_{i=1}^{n} X_i$称为亚指数随机变量，概率密度函数为：

$$f_S(t) = f_{X_1+\cdots+X_n}(t) = \sum_{i=1}^{n} C_{i,n}\lambda_i e^{-\lambda_i t}\text{，其中}C_{i,n} = \prod_{j\neq i}\frac{\lambda_j}{\lambda_j-\lambda_i}$$

- ➢ 卷积的逐步推导+归纳法（课后任务）

•S的尾分布函数：

$$P\{S > t\} = \int_t^{\infty}\sum_{i=1}^{n} C_{i,n}\lambda_i e^{-\lambda_i x}\,dx = \sum_{i=1}^{n} C_{i,n} e^{-\lambda_i t}$$

- ➢ 后续的连续时间马氏链会用到这个结论

8.1.5 指数随机变量的卷积

- S的失败率函数：

$$r_S(t) = \frac{\sum_{i=1}^{n} C_{i,n}\lambda_i e^{-\lambda_i t}}{\sum_{i=1}^{n} C_{i,n} e^{-\lambda_i t}}$$

- 当$t \to \infty$的情况:

$$\lim_{t\to\infty} r_S(t) = \min(\lambda_1, \dots, \lambda_n) = \lambda^*$$

  - 分子分母同乘$e^{\lambda^* t}$即可：

$$\frac{C_{*,n}\lambda^* + \sum_{i\neq *} C_{i,n}\lambda_i e^{-(\lambda_i-\lambda^*)t}}{C_{*,n} + \sum_{i\neq *} C_{i,n} e^{-(\lambda_i-\lambda^*)t}}$$

  - 这也说明，当t很大时，一个存活到年龄t的亚指数分布的部件的剩余寿命，近似于一个指数随机变量描述的时长，且该随机变量速率为求和项中的指数随机变量对应的最小速率！

8.1.5 指数随机变量的卷积

- 考克斯随机变量（例**5.11.**）：$X_1, \dots, X_m$是速率不同$(\lambda_1, \dots, \lambda_m)$的独立指数随机变量.
- 令N独立于这些随机变量，可取值$1, \dots, m$，假设$P_n = P\{N = n\}$,且$\sum_{n=1}^{m} P_n = 1$
- 则$Y = \sum_{i=1}^{N} X_i$称为考克斯随机变量，其概率密度函数为:

$$f_Y(t) = \sum_{n=1}^{m} P_n \sum_{i=1}^{n} C_{i,n} \lambda_i e^{-\lambda_i t} ,$$

理解:

- 考克斯随机变量Y可用来描述，病人接受治疗的总时长
- 假定治疗包含m个有顺序的步骤，各步骤时长为$X_1, \dots, X_m$
- 条件于病人已经到达了步骤n, 病人在步骤n结束后跳出治疗过程的概率为：

$$r(n) = P\{N = n | N \geq n\} = \frac{P_n}{\sum_{i=n}^{m} P_i}$$

  - 注意到，这里是离散时间版本的失败率函数

8.1.5 指数随机变量的卷积

- 考克斯随机变量（例**5.11.**）：$X_1,\dots,X_m$是速率不同$(\lambda_1,\dots,\lambda_m)$的独立指数随机变量.
- 令N独立于这些随机变量，可取值$1,\dots,m$，假设$P_n=P\{N=n\}$,且$\sum_{n=1}^{m}P_n=1$
- 则$Y=\sum_{i=1}^{N}X_i$称为考克斯随机变量，其概率密度函数为:

$$f_Y(t)=\sum_{n=1}^{m}P_n\sum_{i=1}^{n}C_{i,n}\lambda_i e^{-\lambda_i t}\ ,$$

推导$f_Y(t)$:

$$\begin{aligned}f_Y(t)&=\sum_{n=1}^{m}f_Y(t|N=n)\,P_n\\&=\sum_{n=1}^{m}f_{X_1+\cdots X_n}(t|N=n)\,P_n\\&=\sum_{n=1}^{m}f_{X_1+\cdots X_n}(t)\,P_n\\&=\sum_{n=1}^{m}P_n\sum_{i=1}^{n}C_{i,n}\lambda_i e^{-\lambda_i t}\end{aligned}$$

8.2计数过程
(课本5.3.1)

8.2.1. 计数过程定义

- 计数过程$\{N(t), t \geq 0\}$：$N(t)$表示时刻t为止发生事件的总数
  - 仅是在时刻t是不行的！
  - 例1：$N(t)$等于在或早于时刻t进入一个特定商店的人数
  - 例2：一个小孩诞生，计做一个事件，在或早于时刻t诞生的总人数（即使人已经去世，也要计算，因为属于发生的事件数）
  - 例3：足球队员进一球，计做一个事件，在或早于时刻t进球总数

8.2.1. 计数过程定义

- 计数过程$\{N(t), t \geq 0\}$：正式定义，必须满足下列条件
  1. $N(t) \geq 0$
  2. $N(t)$取整数值
  3. 若$s < t$, 则$N(s) \leq N(t)$
  4. 对$s < t$, 则$N(t) - N(s)$表示在区间$(s, t]$中发生的事件个数

- 计数过程是一个很广泛的过程
  - 泊松过程是一种最重要的计数过程

8.2.2. 独立增量

- 独立增量：发生在不相交时间区间中的事件的个数彼此独立，这类计数过程称为具有独立增量
  - 例：$N(10)$ 与 $N(15) - N(10)$ 独立
  - 计数过程不一定具有独立增量

- 数学表达：对任何正整数n和$0 \leq t_1 < t_2 < \cdots < t_n$
  - $N(0), N(t_1) - N(0), N(t_2) - N(t_1), \dots, N(t_n) - N(t_{n-1})$ 相互独立
  - 一般用$N(s,t] = N(t) - N(s)$表示区间$(t,s]$发生的事件数

8.2.3. 平稳增量

- 平稳增量：如果任何时间区间中发生的事件个数的分布仅依赖于时间区间的长度，这类计数过程称为具有平稳增量
  - 区间$(s, s+t)$中发生的事件个数的分布对一切s都相同
  - 计数过程不一定具有平稳增量

- 数学表达：对任何$s > 0$和$0 \leq t_1 < t_2$,
  - $N(t_1, t_2], N(t_1 + s, t_2 + s]$同分布

8.3 泊松过程定义
（课本5.3.2）

8. 3. 1. $o(h)$函数的定义

- $o(h)$的定义：函数$f(\cdot)$称为$o(h)$，若 $\lim\limits_{h\to 0}\frac{f(h)}{h}=0$
  - $f(h)$很小，$f(h)$比h更快接近0

- 例5.12：下列函数是$o(h)$吗？
  - 例1：$f(x)=x^2$？ 是，因为$\lim\limits_{h\to 0}\frac{f(h)}{h}=\lim\limits_{h\to 0}h=0$
  - 例2：$f(x)=x$？ 否，因为$\lim\limits_{h\to 0}\frac{f(h)}{h}=\lim\limits_{h\to 0}1=1$
  - 例3：$f(\cdot)$和$g(\cdot)$是$o(h)$, $f(\cdot)+g(\cdot)$？ 是
  - 例4：$f(\cdot)$是$o(h)$, c$f(\cdot)$？ 是，因为c $\lim\limits_{h\to 0}\frac{f(h)}{h}=c*0$
  - 例5：多个$o(h)$的线性组合？ 是，由例3&4

8.3.1. $o(h)$函数的定义

- $o(h)$的作用：使得命题更加简洁
  - 若X是密度为f的连续随机变量，且失败率函数为$\lambda(t)$
  - 近似命题

$$P\{t < X < t + h\} \approx f(t)h;$$

$$P\{t < X < t + h | X > t\} \approx \lambda(t)h$$

  - 可准确表示为

$$P\{t < X < t + h\} = f(t)h + o(h);$$

$$P\{t < X < t + h | X > t\} = \lambda(t)h + o(h)$$

8.3.2. 泊松过程的定义

●泊松过程定义：计数过程$\{N(t),t\geq 0\}$称为具有速率$\lambda(\lambda>0)$的泊松过程，若下列四个条件满足

1. $N(0)=0$
2. $\{N(t),t\geq 0\}$过程有平稳增量和独立增量
3. $P\{N(t+h)-N(t)=1\}=\lambda h+o(h)$
4. $P\{N(t+h)-N(t)\geq 2\}=o(h)$

- 要点：时间区间很短时，至多发生一件事！
- $P\{N(t+h)-N(t)=0\}=1-\lambda h+o(h)$

- 重要性质1：泊松过程也具备无记忆性
- 重要性质2：在任意长度为t的区间中的事件个数服从均值λt的泊松分布

8.3.2. 泊松过程的无记忆性

- 泊松过程的无记忆性：过了一段时间，从新的时间点开始，从新计数，该泊松过程与原泊松过程性质完全一致！
  - 新泊松过程与原泊松过程完全一致
    - 验证四条性质

8.3.3. 泊松过程性质-时间区间内事件数分布

- **定理5.1**：若$\{N(t), t \geq 0\}$是速率为$\lambda(\lambda > 0)$的泊松过程，则对一切$s > 0, t > 0, N(s+t) - N(s)$是均值为$\lambda t$的泊松随机变量。也就是，在任意长度为$t$的区间中的事件个数是均值为$\lambda t$的泊松随机变量。

思路：

- 由平稳增量性可知，只要说明$N(t) \sim Pois(\lambda t)$即可说明对一切$s > 0, t > 0, N(s+t) - N(s)$是均值为$\lambda t$的泊松随机变量

思路1：

- 随机变量的拉普拉斯变换唯一确定分布，验证$N(t)$的拉普拉斯变换与$Pois(\lambda t)$相等即可

8.3.3. 泊松过程性质-时间区间内事件数分布

- **定理5.1**：若$\{N(t), t \geq 0\}$是速率为$\lambda(\lambda > 0)$的泊松过程，则对一切$s > 0, t > 0, N(s+t) - N(s)$是均值为$\lambda t$的泊松随机变量。也就是，在任意长度为$t$的区间中的事件个数是均值为$\lambda t$的泊松随机变量。

证明（求$N(t)$的拉普拉斯变换）：

- 固定$u > 0$，并且令 $g(t) = E[\exp\{-uN(t)\}]$
- 我们推导$g(t)$的一个微分方程如下，对$h \to 0$：

$$g(t+h) = E[\exp\{-uN(t+h)\}]$$

$$= E\left[\exp\{-uN(t)\}\exp\left\{-u\big(N(t+h) - N(t)\big)\right\}\right]$$

$$= E[\exp\{-uN(t)\}]\,\mathrm{E}\left[\exp\left\{-u\big(N(t+h) - N(t)\big)\right\}\right]$$

$$= g(t)\mathrm{E}\left[\exp\left\{-u\big(N(t+h) - N(t)\big)\right\}\right]$$

- 泊松过程的定理3,4可知

$$P\{N(t+h) - N(h) = 0\} = 1 - \lambda h + o(h)$$

$$P\{N(t+h) - N(h) = 1\} = \lambda h + o(h)$$

$$P\{N(t+h) - N(h) \geq 2\} = o(h)$$

8.3.3. 泊松过程性质-时间区间内事件数分布

- 注意到$N(t+h)-N(t) \sim N(h)$,取条件$=0,=1,\geq 2$:

$$E\left[\exp\left\{-u\left(N(t+h)-N(t)\right)\right\}\right]$$
$$=1-\lambda h+o(h)+e^{-u}\left(\lambda h+o(h)\right)+o(h)$$
$$=1-\lambda h+e^{-u}\lambda h+o(h)$$

- 所以，$g(t+h)=g(t)(1+\lambda h(e^{-u}-1))+o(h)$
- 可知，$\frac{g(t+h)-g(t)}{h}=g(t)\lambda(e^{-u}-1)+\frac{o(h)}{h}$
- 令$h\to 0$，$g'(t)=g(t)\lambda(e^{-u}-1)$

等价的，$\frac{g'(t)}{g(t)}=\lambda(e^{-u}-1)$

- $\frac{g'(t)}{g(t)}=\lambda(e^{-u}-1)$这种形式的微分方程求解是要会的！

8.3.3. 泊松过程性质-时间区间内事件数分布

常微分方程求解：

- 注意到$\frac{g'(t)}{g(t)}$是$\ln g(t)$的导数，积分推出

$$\ln(g(t)) = \lambda(e^{-u} - 1)t + C$$

- 由边界条件：$g(0) = E\left[e^{-uN(0)}\right] = 1$
  - 得到：$C = 0$
- 即，$N(t)$的Laplace变换是$E\left[e^{-uN(t)}\right] = g(t) = e^{\lambda t(e^{-u}-1)}$
- 与均值λt的泊松随机变量同形式
- 说明$N(t)$~$\text{Poisson}(\lambda t)$

\# 结论得证

- 课后任务：验证$Poisson(\lambda t)$的拉普拉斯变换形式

8.3.3. 二项分布理解时间区间内事件数分布

思路2：

- 将时间区间均等分成尽量小份，每一小份至多发生一个事件，则每一小份是否发生事件可看做一个伯努利试验。
- 时间区间内发生的总事件数，等于这些伯努利试验相加，因而构成二项分布。
- 由讲义1中讲过的，在n极大，p极小，$np=$常数的情况下，二项分布趋向于泊松随机变量；可知时间区间内发生的总事件数服从泊松分布。
- 所以定理5.1也被称为一般性的Law of rare events.
  - 特殊的Law of rare events，就是讲义1的结论：在n极大，p极小，$np=$常数的情况下，二项分布趋向于泊松随机变量。

8.3.3. 二项分布理解时间区间内事件数分布

- 思路2-泊松过程中的二项分布：
  - 将$[0,t]$均分成k份($k \to \infty$)
  - 注意到：

$$N(t) = \text{发生事件的时间区间数}$$

  - 由定义中条件4：单一份内2个或以上事件的概率趋0
  - 由定义中条件3：事件发生的概率为$\frac{\lambda t}{k} + o(\frac{t}{k})$
  - 所以$N(t) \sim Binomial(k, \frac{\lambda t}{k} + o(\frac{t}{k}))$
  - 令$k \to \infty$，有$\frac{\lambda t}{k} + o(\frac{t}{k}) \to 0$，且有$k * \left(\frac{\lambda t}{k} + o\left(\frac{t}{k}\right)\right) = \lambda t + \lim_{k\to\infty} \frac{t o\left(\frac{t}{k}\right)}{\frac{t}{k}} = \lambda t$
  - 由特殊的Law of rare events：

$$\text{所以} N(t) \sim Binomial\left(k, \frac{\lambda t}{k} + o\left(\frac{t}{k}\right)\right) \sim Poisson(\lambda t)$$

8.3.5.时间区间内事件数分布定义泊松过程

- **泊松过程等价定义1(**泊松过程第**2**种定义**):** 计数过程$\{N(t), t \geq 0\}$称具有速率$\lambda(\lambda > 0)$的泊松过程，若下列条件满足

  1. $N(0) = 0$
  2. $\{N(t), t \geq 0\}$过程有独立增量
  3. 在任意长度为t的时间区间内的事件数是均值为λt的泊松随机变量。即$N(s, s+t] \overset{D}{=} Poisson(\lambda t), \forall s, t \ \geq 0.$

理解等价性：

- 新条件3中已经蕴含了平稳增量
- 新条件3能否推出定义中的条件3与条件4?

8.3.5. 时间区间内事件数分布定义泊松过程

- **泊松过程等价定义1(泊松过程第2种定义):** 计数过程$\{N(t), t \geq 0\}$称具有速率$\lambda(\lambda > 0)$的泊松过程，若下列条件满足
  1. $N(0) = 0$
  2. $\{N(t), t \geq 0\}$过程有独立增量
  3. 在任意长度为t的时间区间内的事件数是均值为λt的泊松随机变量。即$N(s, s+t] \overset{D}{=} Poisson(\lambda t), \forall s, t \geq 0.$

等价性验证：

- $P(N(s, s+t] = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \ n = 0,1 \ldots$

得到定义的条件3&4:

- $P\{N(t+h) - N(t) = 1\} = P\{N(t, t+h] = 1\} = e^{-\lambda h}\lambda h$
  $= \lambda h(1 - \lambda h + o(h)) = \lambda h + o(h)$
- $P\{N(t, t+h] \geq 2\} = 1 - e^{-\lambda h}\lambda h - e^{-\lambda h}$
  $= 1 - [\lambda h + o(h)] - [1 - \lambda h + o(h)] = o(h)$
  - 由泰勒展开$e^{-\lambda h}$可得