

第 4 讲 年龄与剩余寿命, Poisson 过程的参数估计

1 年龄与剩余寿命

2 泊松过程的检验与参数估计

- 泊松过程的检验
- 用 $N(t)$ 估计 λ
- 用 S_n 估计 λ

引理 1.1

对于强度为 λ 的泊松过程 $\{N(t)\}$, 用 $N(t-)$ 表示区间 $[0, t)$ 内发生的事件数, 则

- (1) $N(t) - N(t-) = 0$ a.s.;
- (2) $N[s, t] = N(t) - N(s-)$ 是闭区间 $[s, t]$ 内发生的事件数;
- (3) $N[s, t] = N(s, t]$ a.s..

一个使用寿命服从指数分布 $\mathcal{E}(\lambda)$ 的部件一旦失效后马上被换上同型号的备用部件继续工作. 用 $N(t)$ 表示 $(0, t]$ 中更换的部件数, 则 $\{N(t)\}$ 是强度为 λ 的泊松过程. 通常称更换部件的时刻为呼叫时刻. 用 $\{S_n\}$ 表示相应的泊松流.

当 $N(t) = n$ 时, S_n 是时间 t 之前最后一次呼叫时刻, S_{n+1} 是时间 t 之后第一次呼叫时刻. 于是我们知道, $S_{N(t)}$ 是时间 t 前的最后一次呼叫时刻, $S_{N(t)+1}$ 是时间 t 之后的第一个呼叫时刻. 于是

$$S_{N(t)} \leq t < S_{N(t)+1}.$$

定义

$$A(t) = t - S_{N(t)}, \quad R(t) = S_{N(t)+1} - t.$$

$A(t)$ 是 t 时服役的部件的使用年龄, $R(t)$ 是 t 时服役的部件的剩余寿命. 以后将 $A(t)$ 简称为 年龄, 将 $R(t)$ 简称为 剩余寿命.

定理 1.2

在上面的定义下, 有如下的结果:

- (1) $R(t) \sim \mathcal{E}(\lambda)$;
- (2)
$$P(A(t) \leq u) = \begin{cases} 1 - e^{-\lambda u}, & u \in [0, t), \\ 1, & u \geq t; \end{cases}$$
- (3) $A(t)$ 与 $R(t)$ 独立.

证明.

- (1) 对 $v \geq 0$ 有 $\{R(t) > v\} = \{N(t, t+v) = 0\}$. 于是 $P(R(t) > v) = e^{-\lambda v}$, 即 $R(t) \sim \mathcal{E}(\lambda)$.
- (2) 由于 $A(t) \leq t$, 所以对 $u \geq t$, $P(A(t) \leq u) = 1$. 对于 $u < t$, 由 $\{A(t) > u\} = \{N[t-u, t] = 0\}$ 和引理 1.1 (1), 我们有

$$P(A(t) > u) = P(N(t - u, t] = 0) = P(N[t - u, t] = 0) = e^{-\lambda u},$$

于是 (2) 成立.

- (3) 对于 $u, v \geq 0$, 从 $N[t-u, t]$ 和 $N(t, t+v]$ 独立得到 $A(t)$ 和 $R(t)$ 独立.

定理 1.2 (1) 说明时间 t 时服役的部件剩余寿命 $R(t)$ 也服从指数分布 $\mathcal{E}(\lambda)$. 该性质是由指数分布的无记忆性所决定的.
利用上面结论和关系式

$$A(t) = t - S_{N(t)}, \quad R(t) = S_{N(t)+1} - t.$$

我们有

推论 1.3

$S_{N(t)}$ 和 $S_{N(t)+1}$ 的分布函数分别为

$$P(S_{N(t)} \leq s) = \begin{cases} e^{-\lambda(t-s)} & 0 \leq s \leq t, \\ 1, & s > t, \end{cases}$$

$$P(S_{N(t)+1} \leq s) = \begin{cases} 1 - e^{-\lambda(s-t)} & 0 \leq s < t, \\ 0, & s \geq t. \end{cases}$$

指数分布的数学期望是 λ^{-1} . 如果用 $X(t)$ 表示时刻 t 时服役的部件的使用寿命, 则有

$$X(t) = A(t) + R(t), \quad E[X(t)] = E[A(t)] + E[R(t)] > \lambda^{-t}.$$

于是, t 时刻服役的部件的平均寿命比同型号的备用部件的平均使用寿命 $E[X_1] = \lambda^{-1}$ 要长.

注意 $A(t)$ 的分布函数在 t 处有一个跳跃高度 $e^{-\lambda t}$, 用定理 1.2 (2) 得到, 当 $t \rightarrow \infty$ 时,

$$\begin{aligned} E[A(t)] &= \int_0^\infty u dP(A(t) \leq u) \\ &= \int_0^t \lambda u e^{-\lambda u} du + t e^{-\lambda t} \\ &\rightarrow \int_0^\infty \lambda u e^{-\lambda u} du = \frac{1}{\lambda}, \end{aligned}$$

所以有

$$\lim_{t \rightarrow \infty} E[X(t)] = \lim_{t \rightarrow \infty} (E[A(t)] + E[R(t)]) = 2\lambda^{-1} = 2E[X_1].$$

上式说明了, 如果泊松过程在无穷远之前就开始了, 则 t 时服役的部件的平均使用寿命是同型号备用部件的平均寿命的两倍.

注记 1.4

发生上述现象并不足为怪. 因为实际使用寿命长的部件在 t 时被遇到的概率比实际寿命短的部件被遇到的概率要大.

例 1.5

北京前门的公交车站，每 6 分钟发出一辆开往颐和园的公交车。由于随机因素的干扰，汽车到达圆明园站时，两车之间的间隔时间成为独立同分布，服从指数分布的随机变量。设某乘客等可能地到达车站候车，计算

- (1) 他在前门候车时的平均候车时间；
- (2) 他在圆明园站候车时的平均候车时间。

解.

- (1) 用 T 表示甲到达前门站的时间. 对于任意长度为 6 分钟的发车间隔 $(0, 6]$, 已知 $T \in (0, 6]$ 时, T 在 $(0, 6]$ 中均匀分布. 所以平均候车时间是 3 分钟.
- (2) 根据题意, 公交车按照强度为 λ 的泊松过程 $\{N(t)\}$ 到达圆明园站. 由于这路公交车都要经过圆明园站, 所以平均每 6 分钟停靠一辆, 即有 $E[N(6)] = 6\lambda = 1$. 于是 $\lambda = 1/6$. 按照剩余寿命的定义, 在 t 时到达的乘客的候车时间为 $R(t)$. 由 $R(t) \sim \mathcal{E}(\lambda)$ 知道平均候车时间为 $E[R(t)] = 6(\text{分钟})$. □

注记 1.6

上例的结果告诉我们, 市内公交车的始发站和终点站距离太远时, 会延长乘客的平均候车时间和降低公交车的平均利用率.

设 $X_1, X_2, \dots, X_n$ 是来自总体 X 的样本, θ 是 X 的未知参数, 但是已知 $\theta \in \Theta_0 + \Theta_1$, 这里 $\Theta_0 + \Theta_1$ 是互不相交的参数集合. 为判断 $\theta \in \Theta_0$ 是否成立, 作假设检验, 其中

原假设 $H_0: \theta \in \Theta_0$, 备择假设 $H_1: \theta \in \Theta_1$.

设 W 是一个事件, 当 W 发生就否定 H_0 , 称 W 为检验的拒绝域. 设 α 是 $(0, 1)$ 中的常数. 如果对一切的 $\theta \in \Theta_0$, 有

$$P_{\theta}(W) \leq \alpha,$$

则称拒绝域 W 的检验水平或显著性水平是 α .

按照泊松过程的性质, 要检验计数过程 $\{N(t), t \geq 0\}$ 是否是泊松过程, 可转化为下面的检验问题之一:

1. 检验 $\{X_n, n \geq 1\}$ 是否独立同指数分布;
2. 对任意 $t > 0$, 检验在 $N(t) = 1$ 下 $S_1 = X_1$ 是否是 $[0, t]$ 上均匀分布;
3. 给定 $T > 0$, 检验在 $N(T) = n$ 下 $S_1, S_2, \dots, S_n$ 的条件分布是否与 $[0, T]$ 上 n 个独立均匀分布随机变量的顺序统计量有相同的分布.

这里我们仅具体讨论最后一种检验方法.

提出检验假设 $H_0 : \{N(t), t \geq 0\}$ 是泊松过程. 令 $\sigma_n = \sum_{k=1}^n S_k$. 当 H_0 成立时, 我们有

$$E[\sigma_n | N(T) = n] = E\left[\sum_{i=1}^n U_i\right] = \frac{nT}{2},$$

$$\text{Var}[\sigma_n | N(T) = n] = \text{Var}\left(\sum_{i=1}^n U_i\right) = \frac{nT^2}{12}.$$

这里 $\{U_i, 1 \leq i \leq n\} \sim U[0, t]$ 独立同分布. 由中心极限定理我们有

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(\frac{\sigma_n - \frac{n}{2}T}{T\sqrt{\frac{n}{12}}} | N(T) = n\right) &= \lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n Y_i - \frac{n}{2}T}{T\sqrt{\frac{n}{12}}} \leq x\right) \\ &= \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du. \end{aligned}$$

于是对于充分大的 n , 有

$$P\left(\frac{\sigma_n}{t} \leq \frac{1}{2}\left[n + x\left(\frac{n}{3}\right)^{\frac{1}{2}}\right] \mid N(T) = n\right) \approx \Phi(x).$$

如果给定置信水平 $\alpha = 0.05$, 则当

$$\frac{\sigma_n}{T} \in \frac{1}{2}\left[n \pm 1.96\left(\frac{n}{3}\right)^{\frac{1}{2}}\right]$$

时, 接受 H_0 , 否则拒绝 H_0 .

设 $N(t)$ 是强度为 λ 的泊松过程. 由 $N(t)$ 的概率分布

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 0, 1, \dots$$

我们知基于观测 $N(t) = n$ 的似然函数是

$$L(\lambda) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

对数似然函数是

$$l(\lambda) = \ln L(\lambda) = n \ln \lambda - \lambda t + c_0, \quad c_0 \text{ 是常数.}$$

由 $l'(\lambda) = 0$ 得到 λ 的最大似然估计 n/t .

一般情况下, $[0, t]$ 内发生了 $N(t)$ 个事件, λ 的最大似然估计是

$$\hat{\lambda} = \frac{N(t)}{t}.$$

由于 $EN(t) = \lambda t$, 所以 $E\hat{\lambda} = \lambda$. 这说明 $\hat{\lambda}$ 是无偏估计.

对于每个 $t > 0$, 有正整数 n 使得 $n - 1 < t \leq n$ 成立. 这时

$$\frac{N(n-1)}{n} \leq \frac{N(t)}{t} \leq \frac{N(n)}{n-1}.$$

利用独立增量性和平稳增量性知 $N(j-1, j]$ 是独立同分布的随机变量. 注意到

$$N(n) = \sum_{j=1}^n N(j-1, j],$$

利用强大数律我们有

$$\lim_{n \rightarrow \infty} \frac{N(n-1)}{n} = \lim_{n \rightarrow \infty} \frac{N(n-1)}{n-1} \frac{n-1}{n} = E[N(1)] = \lambda \text{ a.s.},$$

$$\lim_{n \rightarrow \infty} \frac{N(n)}{n-1} = \lim_{n \rightarrow \infty} \frac{N(n)}{n} \frac{n}{n-1} = E[N(1)] = \lambda \text{ a.s..}$$

于是

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \lambda \text{ a.s.},$$

这说明 $\hat{\lambda}$ 是 λ 的强相合估计.

再利用独立增量性和平稳增量性我们知

$$Y_j = N(jt/n) - N((j-1)t/n), \quad j = 1, 2, \dots, n$$

是独立同分布的随机变量，其部分和

$$N(t) = \sum_{j=1}^n Y_j$$

的数学期望和方差都是 λt . 从中心极限定理我们知对于较大的 λt ,

$$\xi = \frac{N(t) - \lambda t}{\sqrt{\lambda t}} \sim N(0, 1)$$

近似成立.

给定置信水平 $1 - \alpha$, 用 $z_{\alpha/2}$ 表示标准正态分布的 $\alpha/2$ 上分位数:

$\Phi(z_{\alpha/2}) = 1 - \alpha/2$. 利用 $P(|\xi| \leq z_{\alpha/2}) \approx 1 - \alpha$, 对于较大的 λt , 可以得到 λ 的置信水平为 $1 - \alpha$ 的近似置信区间

$$J_\lambda = \left[\frac{b - \sqrt{b^2 - 4ac}}{2a}, \frac{b + \sqrt{b^2 - 4ac}}{2a} \right],$$

其中 $a = t^2$, $b = 2tN(t) + tz_{\alpha/2}^2$, $c = N^2(t)$.

实际上, $\xi \leq z_{\alpha/2}$ 当且仅当

$$|N(t) - \lambda t|^2 \leq \lambda t z_{\alpha/2}^2,$$

即 $g(\lambda) = a\lambda^2 - b\lambda + c \leq 0$. 解该不等式即得 λ 的置信水平为 $1 - \alpha$ 的近似置信区间.

定理 2.1

设 $\{N(t)\}$ 是强度为 λ 的泊松过程, 则 $\hat{\lambda} = \frac{N(t)}{t}$ 是 λ 的强相合无偏估计. 对于较大的 λt , λ 的置信水平为 $1 - \alpha$ 的近似置信区间是

$$J_\lambda = \left[\frac{b - \sqrt{b^2 - 4ac}}{2a}, \frac{b + \sqrt{b^2 - 4ac}}{2a} \right], \quad (2.1)$$

例 2.2

上海证券交易所开盘后, 在 10:00 至 10:30 之间一共有 36 万次成交. 假定在这个时间段内股票的成交构成一个泊松流, 以秒为时间单位.

- (1) 估计泊松流的强度 λ ;
- (2) 在置信水平 0.95 下, 计算强度 λ 的置信区间;
- (3) 当 $\lambda = 200$ 时, 计算 5 秒内的成交次数大于 1050 次的概率.

解.

(1) 因为在 $t = 30 \times 60$ 秒内有 $N(t) = 36$ 万次成交, 所以 λ 的最大似然估计是

$$\hat{\lambda} = \frac{36 \times 10^4}{30 \times 60} = 200.$$

(2) 对于置信水平 $1 - \alpha = 0.95$, 我们有 $z_{\alpha/2} = 1.96$. 容易计算出 (2.1) 中的

$$\begin{aligned} N(t) &= 36 \times 10^4, \\ a &= t^2 = 1800^2 = 3240000, \\ b &= (2tN(t) + tz_{\alpha/2}^2) = 1.296 \times 10^9, \\ c &= N^2(t) = 1.296 \times 10^11, \end{aligned}$$

将 a, b, c 代入 (2.1) 得到置信区间 $[199.9846, 200.0154]$.

解.

(3) 因为每秒钟平均有 200 次交易, 所以

$$\xi = \frac{N(5) - s\lambda}{\sqrt{5\lambda}} \sim N(0, 1)$$

近似成立. 于是

$$\begin{aligned} P(N(5) > 1050) &= P\left(\frac{N(5) - 5\lambda}{\sqrt{5\lambda}} > \frac{1050 - 5\lambda}{\sqrt{5\lambda}}\right) \\ &\approx P\left(\xi > \frac{5}{\sqrt{10}}\right) \\ &= 1 - \Phi(1.5811) \approx 0.057. \end{aligned}$$



在实际问题中如果可以观测到第 n 次到达的发生时刻 $S_n = s_n$, 我们知 λ 的似然函数

$$L(\lambda) = \frac{\lambda^n}{\Gamma(n)} s_n^{n-1} e^{-\lambda s_n}.$$

对数似然函数是

$$l(\lambda) = \ln L(\lambda) = n \ln \lambda - \lambda s_n + c_0, \quad c_0 = \text{常数}.$$

由 $l'(\lambda) = n/\lambda - s_n = 0$, 得到 λ 的最大似然估计 $\tilde{\lambda} = n/s_n$. 实际问题中观测到 S_n 的最大似然估计是 $\tilde{\lambda} = n/S_n$.

注意到 $E[\tilde{\lambda}] = \frac{n}{n-1}\lambda > \lambda$, 所以 $\tilde{\lambda}$ 不是 λ 的无偏估计. 但是他是 λ 的强相合估计, 这是因为由强大数律我们有当 $n \rightarrow \infty$ 时,

$$\tilde{\lambda}^{-1} = \frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n} \rightarrow E[X_1] = 1/\lambda \text{ a.s.},$$

于是有 $\tilde{\lambda} \rightarrow \lambda$ a.s..

下面再计算 λ 的置信区间. 首先我们有 $\eta = 2\lambda S$ 的密度函数

$$f(t) = \frac{1}{2^n \Gamma(n)} t^{n-1} e^{-t/2}, \quad t > 0.$$

这正是 $2n$ 个自由度的 χ^2 分布的密度函数, 于是 $\eta = 2\lambda S_n \sim \chi^2(2n)$. 用 $\chi_{\alpha/2}^2$ 表示 $\chi^2(2n)$ 分布的上 $\alpha/2$ 分位数, 利用

$$P(\chi_{1-\alpha/2}^2 \leq 2\lambda S_n \leq \chi_{\alpha/2}^2) = 1 - \alpha,$$

得到 λ 的置信水平为 $1 - \alpha$ 的置信区间

$$\left[\frac{\chi_{1-\alpha/2}^2}{2S_n}, \frac{\chi_{\alpha/2}^2}{2S_n} \right].$$