寿命估计

Estimation of Lifetime

郑祖康 著

GUANGXI NORMAL UNIVERSITY PRESS
广西师范大学出版社
·桂林·

寿命估计

Estimation of Lifetime

YINGYONG TONGJI YU XINXI CONGSHU

应用统计与信息丛书

郑祖康 著

GUANGXI NORMAL UNIVERSITY PRESS
广西师范大学出版社
·桂林·

图书在版编目（CIP）数据

寿命估计／郑祖康著. —桂林：广西师范大学出版社，2002. 3

（应用统计与信息丛书）

ISBN 7-5633-3473-4

Ⅰ. 寿… Ⅱ. 郑… Ⅲ. 疲劳寿命估算

Ⅳ. 0346.2

中国版本图书馆 CIP 数据核字（2000）第 010373 号

广西师范大学出版社出版发行

（桂林市中华路 36 号　邮政编码：541001

网址：http://www.bbtpress.com.cn）

出版人：萧启明

全国新华书店经销

广西师范大学出版社印刷厂印刷

（广西桂林市临桂县金山路 168 号　邮政编码：541100）

开本：890 mm×1 240 mm　1/32

印张：4.875　字数：140 千字

2002 年 3 月第 1 版　2002 年 3 月第 1 次印刷

印数：0 001～1 000　定价：12.00 元

前　言

如何进行寿命试验，如何进行寿命估计，是工程技术、生物医学领域的重要课题，从产品的耐用性研究到人类各种疾病治疗的研究，都有广泛的应用. 本书就这一专题进行了论述，介绍了当前使用的一些理论和方法，特别介绍了我国学者在这方面的工作.

本书除了准备知识之外，有三方面的内容：第一部分是定时截断与定数截断下的寿命估计，包括无失效数据分析、定时定数试验的改进与选择. 第二部分是随机截断下的寿命估计，内容有 K 类估计、和型估计等. 由于干扰随机变量的分布可以自行设计，因此也有较好的结果. 第三部分是有关污染数据的寿命估计，阐述了参数方法、非参数方法、小样本方法等. 各章的后面都附有参考文献，可供读者参考.

与寿命试验、寿命估计相关的研究还有区间截断数据分析(interval censored data analysis)和数据挖掘(data mining)等，从理论上讲，这些方面尚未成熟，所以本书没有收入. 应该指出，加速寿命试验是一个重要的方面，考虑到目前已有这方面的专著，本书也不再涉及了.

感谢广西师范大学出版社为本书出版作出的努力，主编余鑫晖教授亲自做了大量的工作，才使这本学术专著问世. 由于本人的水平有限，书中难免有错误之处，敬请读者批评指正.

作　者

2000 年 10 月于上海复旦大学

目　　录

第一章　准备知识

§1　关于寿命的一些基本概念

寿命估计，包括它的分布、均值、方差、分位数等的估计在工程和生物医学中被人们频繁地使用，各种各样的与寿命有关的数据统计分析已发展成一个重要的专题，从产品的耐用性研究到人类各种疾病的研究，都有广泛的应用. 从数学意义上讲，我们可以把寿命理解为非负随机变量，记为 X，它的特征可用下列函数或数字特征来刻画.

1.1　基本函数

1. 生存函数 $S(t)$

$$S(t)=P(X>t)=1-P(X\leqslant t)=1-F(t), \tag{1.1}$$

其中 $F(t)$ 为 X 分布函数，$S(t)$ 又称为可靠度或可靠度函数. 显然，$S(0^-)=1$，$S(\infty)=0$. 当 X 有分布密度函数 $f(t)$ 时，有

$$S(t)=\int_t^\infty f(u)\mathrm{d}u.$$

我们常记 X 的数学期望为 μ，方差为 σ^2，中位数为 m.

2. 危险率函数 $\lambda(t)$

$$\lambda(t)=\lim_{\Delta t\to 0^+}\frac{1}{\Delta t}P(X\leqslant t+\Delta t\mid X>t), \tag{1.2}$$

这里假定极限存在. $\lambda(t)$ 又称为损坏函数、失效率函数等. 当 X 的密度函数 $f(t)$ 存在时，又有

$$\lambda(t)=\frac{f(t)}{1-F(t)}=\frac{f(t)}{S(t)}=-\frac{S'(t)}{S(t)}, \tag{1.3}$$

或者

$$S(t)=\exp\left[-\int_0^t\lambda(u)\mathrm{d}u\right], \tag{1.4}$$

而且满足

$$\int_0^{\infty} \lambda(u)\mathrm{d}u=\infty.$$

有时候，寿命变量 X 须当做一个离散变量来处理，比如寿命被分组. 设 X 取值为 $t_1,t_2,\cdots(0\leqslant t_1<t_2<\cdots)$，记

$$p(t_j)=P(X=t_j),\ j=1,2,\cdots,$$

则其生存函数为

$$S(t)=P(X>t)=\sum_{j:t_j>t} p(t_j), \tag{1.5}$$

危险率函数为

$$\lambda(t_j)=P(X=t_{j+1}\mid X>t_j)=\frac{p(t_{j+1})}{S(t_j)},\ j=1,2,\cdots. \tag{1.6}$$

注意到 $p(t_{j+1})=S(t_j)-S(t_{j+1})$，则

$$\lambda(t_j)=1-\frac{S(t_{j+1})}{S(t_j)}.$$

3. 累积危险率函数 $\Lambda(t)$

$$\Lambda(t)=\int_0^t \lambda(u)\mathrm{d}u. \tag{1.7}$$

$\Lambda(t)$ 又称为累积损坏函数、累积失效率函数等. 类似地，我们有

$$\Lambda(+\infty)=\infty,$$

$$S(t)=\mathrm{e}^{-\Lambda(t)}. \tag{1.8}$$

同时密度函数 $f(t)$ 也可以用 $\lambda(t)$，$\Lambda(t)$ 来表示，就是

$$f(t)=\lambda(t)S(t)=\lambda(t)\exp\left[-\int_0^t \lambda(u)\mathrm{d}u\right]=\lambda(t)\mathrm{e}^{-\Lambda(t)}. \tag{1.9}$$

图 1-1 给出了三个连续分布的危险率函数. 这三个危险率函数本质上有很大的不同：(a) 单调上升，我们称 X 具有递增损坏速度 (Increasing Failure Rate)，简记为 IFR；(b) 单调下降，我们称 X 具有递减损坏速度

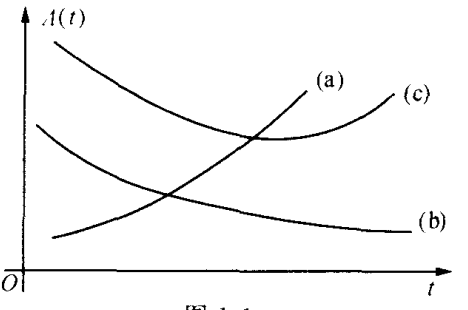

图 1-1

(Decreasing Failure Rate)，简记为 DFR；(c)称为“浴盆状”或 U 形危险率函数. 在实际中，这三种形状的危险率函数的模型是很有用的，尤以“浴盆状”的危险率函数为最. 例如，对一总体中的每一个体从出生到死亡进行跟踪，就会发现“浴盆状”的危险率函数常常是适合的. 如人的寿命，由于先天性缺陷和婴儿疾病，在婴儿阶段死亡率比较高，而后死亡率随年龄的增长而下降，在 20～50 岁段表现出低而稳定，随后又随年龄的增长死亡率升高.

有时我们也考虑所谓平均剩余寿命 $m(t)$ 和平均危险率函数 $\bar{\lambda}(t)$，其定义如下：

$$m(t)=E(X-t|X>t)=\frac{\int_t^{\infty}S(u)\mathrm{d}u}{S(t)}, \tag{1.10}$$

$$\bar{\lambda}(t)=\frac{1}{t}\int_0^t\lambda(u)\mathrm{d}u, \tag{1.11}$$

其中 $m(t)$ 表示个体活过 t 时后尚能存活的期望时间，$\bar{\lambda}(t)$ 则表示 $[0,t]$ 段平均的 $\lambda(t)$.

1.2 参数模型

在分析寿命数据以及在涉及对老化或失效过程模型化的问题中，常用许多参数模型. 参数模型占有重要地位，如果被研究总体能纳入参数模型，常会事半功倍. 下面列出几个常见的参数模型，它们都已被人们较为深刻地研究过，掌握它们不仅便于实际工作，而且利于对非参数模型的理解.

1. 指数分布

指数分布的应用很广，从研究产品的寿命到病患者的存活时间都经常用指数分布作为模型，其主要特征可用危险率函数为常数来刻画，即

$$\lambda(t)=\lambda\ (t\geqslant 0). \tag{1.12}$$

可以进一步算出

$$\Lambda(t)=\lambda t,$$
$$S(t)=e^{-\lambda t},$$
$$f(t)=\lambda e^{-\lambda t}.$$

从历史上看，指数分布是首先得到广泛应用的寿命分布模型，其原因是形式简单又适合用来描述许多对象的寿命. 当然危险率函数的常数假定是比较粗糙的. 可以算出指数分布的均值为 λ^{-1}，方差为 λ^{-2}. 当 $\lambda=1$ 时，我们称指数分布为标准指数分布.

2. 威布尔分布

威布尔分布也许是使用最为广泛的寿命分布，在涉及寿命问题时，包括生物、工业产品的寿命，都广泛提倡用威布尔分布. 它有两个参数 $\lambda(>0)$，$\beta(>0)$，容易满足各种情况. 可以算出

$$S(t)=e^{-(\lambda t)^{\beta}}(t\geqslant 0),$$
$$f(t)=\lambda\beta(\lambda t)^{\beta-1}e^{-(\lambda t)^{\beta}},$$
$$\lambda(t)=\lambda\beta(\lambda t)^{\beta-1},$$
$$\Lambda(t)=(\lambda t)^{\beta}.$$

由此可得威布尔分布的均值和方差分别为

$$\mu=\frac{1}{\lambda}\Gamma(1+\beta^{-1}),$$

$$\sigma^2=\frac{1}{\lambda^2}[\Gamma(1+2\beta^{-1})-\Gamma^2(1+\beta^{-1})],$$

其中 Γ 是 Gamma 函数. 对于不同的 β 值，威布尔分布的密度函数呈不同的形状，因此 β 也称为形状参数. 图 1-2 和图 1-3 分别画出了 $\lambda=1$

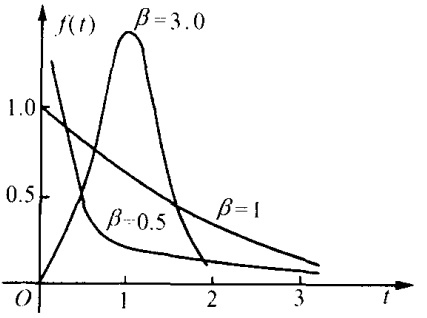

图 1-2

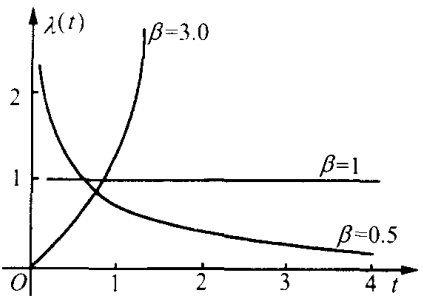

图 1-3

时，$\beta=0.5$，$\beta=1$，$\beta=3.0$ 的威布尔分布的密度函数和危险率函数的图象. 威布尔分布的危险率函数在 $\beta>1$ 时是单调递增的，在 $\beta<1$ 时是单调递减的，在 $\beta=1$ 时恒为常数. 从这两个图也可以看出威布尔分布是相当灵活的，它对许多类型的寿命数据都能给出很好的描述. 注意 λ 只是一个尺度参数，不同的 λ 值只会改变横轴的刻度，不会改变曲线的基本形状. 当 $\beta=1$ 时，威布尔分布退化为指数分布. 正因为这种多变性，威布尔分布能适合许多实际情况.

3. 极值分布

极值分布为 Gumbel 等人所研究，它也有两个参数：$b>0$，$-\infty<u<+\infty$. 它与威布尔分布有密切的关系. 如果 X 具有威布尔分布，那么 $X^*=\ln X$ 就具有极值分布，相应的参数 $b=\beta^{-1}$，$u=-\ln\lambda$. 不难看出极值分布的生存函数为

$$S(t)=\exp\left[-\exp\left(\frac{t-u}{b}\right)\right]\quad(-\infty<t<+\infty),$$

密度函数为

$$f(t)=b^{-1}\left[\left(\frac{t-u}{b}\right)-\exp\left(\frac{t-u}{b}\right)\right]\quad(-\infty<t<+\infty).$$

在数据分析时使用对数寿命时间常常是方便的，因此当寿命时间服从威布尔分布时，取对数后它就服从极值分布了. 可以算出，极值分布的均值为 $u-\gamma b$（γ 为欧拉常数），方差为 $\frac{\pi^2}{6}b^2$. 特别地，当 $u=0$，$b=1$ 时，称为标准极值分布.

4. 对数正态分布

对数正态分布也是一个广泛使用的寿命分布模型，当 $Y=\ln X$ 服从正态分布时，就称 X 服从对数正态分布. 它也有两个参数 $\mu^*>0$，$\sigma^*>0$. 它的生存函数为

$$S(t)=1-\Phi\left(\frac{\ln t-\mu^*}{\sigma^*}\right)\quad(t>0),$$

其中 $\Phi(\cdot)$ 为标准正态分布函数，密度函数为

$$f(t)=\frac{1}{\sqrt{2\pi}\,\sigma^* t}\exp\left[-\frac{1}{2}\left(\frac{\ln t-\mu^*}{\sigma^*}\right)^2\right]\quad(t>0).$$

可以验证，对数正态分布的均值和方差分别为 $\exp\left(\mu^*+\frac{\sigma^{*2}}{2}\right)$ 和 $[\exp(\sigma^{*2})-1][\exp(2\mu^*+\sigma^{*2})]$.

5. 广义 Gamma 分布

广义 Gamma 分布是一个有三个参数 $\alpha>0,\beta>0,\lambda>0$ 的分布，其密度函数为

$$f(t)=\frac{\lambda\beta}{\Gamma(\alpha)}(\lambda t)^{\alpha\beta-1}\mathrm{e}^{-(\lambda t)^{\beta}}(t>0).$$

当 $\alpha=\beta=1$ 时为指数分布，当 $\alpha=1$ 时为威布尔分布，当 $\beta=1$ 时为 Gamma 分布. 可以证明，对数正态分布是 $\alpha\to\infty$ 时广义 Gamma 分布的极限分布. 由此可见，广义 Gamma 分布是一个相当灵活的三个参数分布族，有助于模型的选择.

6. Rayleigh 分布

Rayleigh 分布是一个两个参数($\lambda_0>0,\lambda_1>0$)的分布，它的特点是危险率函数呈线性状态，即

$$\lambda(t)=\lambda_0+\lambda_1 t.$$

由此可以推出

$$\Lambda(t)=\lambda_0 t+\frac{1}{2}\lambda_1 t^2,$$

$$S(t)=\exp\left(-\lambda_0 t-\frac{1}{2}\lambda_1 t^2\right),$$

$$f(t)=(\lambda_0+\lambda_1 t)\exp\left(-\lambda_0 t-\frac{1}{2}\lambda_1 t^2\right).$$

更一般的情况是危险率函数呈多项式状态，即

$$\lambda(t)=\sum_{i=0}^{n}\lambda_i t^i.$$

§2 不完全数据与寿命试验

如何进行寿命试验，如何利用数据估计寿命变量的均值、方差以至整个分布是一个重要的问题. 在解决这些问题之前，先来考察一下我们获得的数据，不难发现我们的数据会出现这样或那样的问题. 比如说，

在灯泡寿命的试验中，我们发现灯泡 A 工作了1 000 h 后仍没有坏，由于试验条件的限制，我们不得不停止试验，我们并不知道灯泡 A 的寿命，但知道它的寿命长于1 000 h. 某医院研究一种特效药对患某种疾病的人生命的影响，这些病人可以在不同的时间进入该药物的治疗研究，一旦进入研究之后，也有种种可能中途离开. 这主要表现在：(i)病人由于搬家等原因不再到此医院看病；(ii)由于疗效不甚理想，病人觉得没有必要继续进行该药物的治疗，或者由于病人的体质原因有较强烈的副作用而不得不中止治疗等. 于是研究人员对此病人“不知所终”，但却记录下最后就诊的日子. 对股票市场中某一股票的买入成本分析，它的分布与正态分布或对数正态分布有显著的差异，细致的研究表明：它的分布函数似乎是两个不同的均值与方差的正态分布函数的混合物. 直观上，这两个不同均值与方差的正态分布可能分别反映了两个集团——机构大户和中小散户的买入成本. 反映人体肥胖程度的 BMI 系数的分布也有类似的现象. 生产过程中我们须定时进行抽样检查，以保证产品的质量，这种定时检查可能每隔 30 min 或 60 min 或更多的时间，应视具体的产品而定. 由于生产过程是连续的，当某次检查发现产品出了问题，我们只能确定出问题的时刻介于两个检查点之间，而并不知道确切的时间. 某地搞了一次家庭消费支出调查，被抽到的家庭被要求填写调查表中的 25 项内容，调查表回收后仍发现有漏填的项目，还发现一些明显的错误，如在家庭人口中填写了“3”，却在粮食消费上填写了“每月 250 kg”. 诸如此类的问题，在各种报表上也时有发现. 我们把上述这些有毛病的数据统称为“不完全数据”. 它们的形式虽然各不相同，但其来源大致有两种可能：一是“自然形成”的，即由于不可控制或不可预知的干扰和污染，只能采集到一些不完全数据；二是“人为造成”的，如前面所述灯泡寿命试验的问题，我们不能因为某一个灯泡“坚持工作”而无限制地延长试验时间. 从寿命试验的角度看，既要处理“自然形成”的受干扰、污染的数据，也要处理因省钱省力省时而“人为造成”的数据. 我们的任务是从这些不完全数据中提取我们所需的信息. 下面我们简述传统的寿命试验方法.

1. 定时截断法(第一类截断)

仍考虑我们上面提到的灯泡寿命试验. 图 2-1 表示 20 个灯泡寿命

试验的结果. 由于试验条件的限制(人力或设备等),这个试验的持续时间不能超过1 000 h. 图 2-1 中,"·"表示灯泡毁坏,"×"表示虽然到达1 000 h(试验必须终止),但灯泡仍未毁坏. 从图中可见,第 3、6、7、12、17、18、20 号灯泡的寿命均超过了1 000 h,但各自的寿命究竟有多长呢,却无法得知. 它们被1 000 h 这个界线截断,这种截断方式称为第一类截断(Type I Censoring).

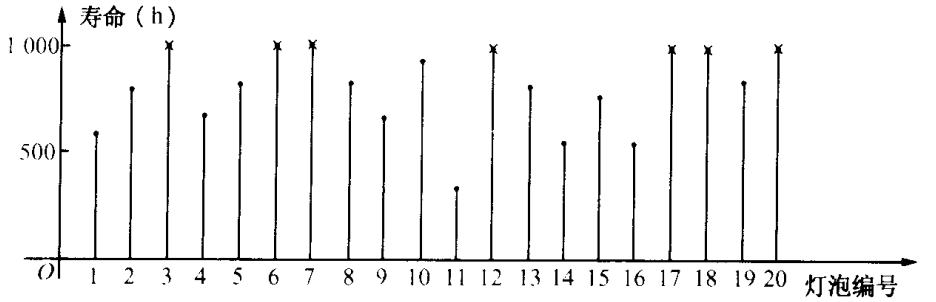

图 2-1

现在要问:这些灯泡的平均寿命究竟是多少呢? 显然,毛病出在那些带"×"的数据上,我们称之为截断数据. 这些数据不能轻易地抛弃掉,否则估得的寿命将偏小,同时白白地耗费了人力钱财. 那么,如何来利用这些数据呢? 从假设检验的角度看,我们要问:$H_0:\mu=750$(μ 表示均值)能否成立? 当"×"的比例增加时,我们的困难也增加了.

一般地,设寿命随机变量 $X_1,X_2,\cdots,X_n$ 独立同分布且具有未知分布函数 $F(x)$[或未知密度函数 $f(x)$]. 在试验中,这些产品有相同的起始点,即同时开始试验. 所谓定时截断,是指有一族预先定下的常数 C_i $(i=1,\cdots,n)$,分别与 X_i 相截,仅观察到

$$\begin{cases} Z_i=\min(X_i,C_i), \\ \delta_i=I_{(X_i\leqslant C_i)}, \end{cases} \tag{2.1}$$

这里 δ_i 是指示函数,即 $\delta_i=1$(若 $X_i\leqslant C_i$),或 $\delta_i=0$(若 $X_i>C_i$). 这些 C_i 也可以相同,图 2-1 中,$C_1=C_2=\cdots=C_{20}=1\ 000$ h.

在定时截断中,信息损失较多,特别是尾部的信息,寿命变量大于 $\max C_i$ 的那一段信息将完全失去. 在某些场合,由于截断力度过强(C_i 值太小),很少出现真实寿命数据,甚至没有寿命数据出现,即出现无失

效数据的情况，这给统计推断带来极大的困难，我们将在第二章详细论述.

2. 定数截断法(第二类截断)

我们也可以把上面灯泡试验的定时截断法作一些改变，把规则改为如果 8 个灯泡毁坏了，那么我们就停止试验. 图 2-2 表示了试验的结果.

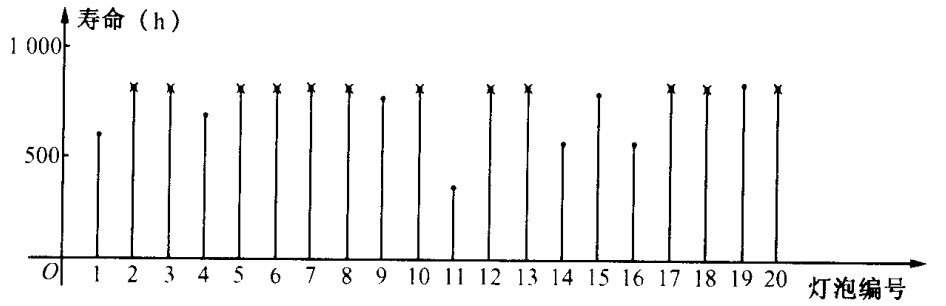

图 2-2

在这 8 个灯泡中，19 号灯泡是最后毁坏的，此时试验停了下来，其余 12 个灯泡就带上了"×". 这种形式的截断称之为定数截断或第二类截断. 一般地说，用 n 个产品做试验，试验在有 r 个失效时就停止了，其中 r 是预先指定的($1\leqslant r\leqslant n$). 也就是说，n 个被观察的产品中有 r 个最小的才被观察到，因为在某些情况要等到 n 个产品都失效要花很长的时间，这样的截断是省时省钱的. 设 $X_1, X_2, \cdots, X_n$ 为独立同分布的寿命变量，$X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 为其顺序统计量，那么(2.1)式就变成

$$Z_{(i)}=\begin{cases}X_{(i)}, & i=1,2,\cdots,r,\\ X_{(r)}, & i=r+1,\cdots,n,\end{cases} \tag{2.2}$$

$$\delta_{(i)}=\begin{cases}1, & i=1,2,\cdots,r,\\ 0, & i=r+1,\cdots,n,\end{cases} \tag{2.3}$$

这里 $Z_{(i)}$ 是 Z_i 的顺序统计量，并假定这些 X_i 都不相同. 定数截断法的统计推断有一定的方便性，设 X_i 的分布函数为 $F(x)$，密度函数为 $f(x)$，则 $X_{(1)}, X_{(2)}, \cdots, X_{(r)}$ 的联合分布密度函数为

$$\frac{n!}{(n-r)!}f(x_{(1)})\cdot\cdots\cdot f(x_{(r)})[1-F(x_{(r)})]^{n-r}. \tag{2.4}$$

统计推断便可建立在(2.4)之上. 特别地，当 $X_1, X_2, \cdots, X_n$ 独立同分布时，$F(x)=1-e^{-\lambda x}$，$f(x)=\lambda e^{-\lambda x}$，即 X_i 服从指数分布时，$X_{(1)}, X_{(2)}, \cdots, X_{(r)}$ 的联合分布密度函数为

$$\begin{aligned}&\frac{n!}{(n-r)!}\left[\prod_{i=1}^{r}\lambda e^{-\lambda X_{(i)}}\right]\left[e^{-\lambda X_{(r)}}\right]^{n-r}\\&=\frac{n!}{(n-r)!}\lambda^r\exp\left\{-\lambda\left[\sum_{i=1}^{r}X_{(i)}+(n-r)X_{(r)}\right]\right\}\\&=\frac{n!}{(n-r)!}\lambda^r e^{-\lambda T},\end{aligned}$$

其中

$$T=\sum_{i=1}^{r}X_{(i)}+(n-r)X_{(r)}. \tag{2.5}$$

直观上 T 是试验的“总”时间，是一个很有用的统计量. 再令

$$\begin{aligned}&W_1=nX_{(1)},\\&W_2=(n-1)(X_{(2)}-X_{(1)}),\\&\cdots\cdots\\&W_i=(n-i+1)(X_{(i)}-X_{(i-1)}),\\&\cdots\cdots\\&W_r=(n-r+1)(X_{(r)}-X_{(r-1)}).\end{aligned}$$

显然 $T=\sum_{i=1}^{r}W_i$ 以及 $\frac{\partial(W_1,W_2,\cdots,W_r)}{\partial(X_{(1)},X_{(2)},\cdots,X_{(r)})}=\frac{n!}{(n-r)!}$，从而 $W_1, W_2, \cdots, W_r$ 的联合分布密度函数为

$$\lambda^r\exp\left(-\lambda\sum_{i=1}^{r}w_i\right)\quad(w_i>0), \tag{2.6}$$

即 $2T\lambda\sim\chi^2_{2r}$. 这一性质容易用来构造置信区间或进行假设检验. 还可以进一步算出

$$E(X_{(i)})=\frac{1}{\lambda}\sum_{j=1}^{i}\frac{1}{n-j+1}.$$

3. 随机截断法

在定时截断法中，C_i 是事先固定的常数，但事实上却常常是随机的，我们改用 Y_i 来记. 所谓随机截断，是指独立同分布的寿命随机变量 $X_1, X_2, \cdots, X_n$ 分别受到随机变量 $Y_1, Y_2, \cdots, Y_n$ 的干扰，我们只能观察

到其中小的一个以及判断出是 X_i 还是 Y_i，即

$$\begin{cases} Z_i=\min(X_i,Y_i), \\ \delta_i=I_{(X_i\leqslant Y_i)}. \end{cases} \tag{2.7}$$

一般地，我们还假定这些 Y_i 也是独立同分布的，且$\{X_i\}$序列与$\{Y_i\}$序列是独立的. 若进一步假定 X_i 的分布函数为 $F(x)$，密度函数为 $f(x)$，Y_i的分布函数为 $G(y)$，密度函数为 $g(y)$，则我们可以写出$\{Z_i\}$的似然函数

$$\begin{aligned} L &= \prod_{i=1}^{n} \{f(Z_i)[1-G(Z_i)]\}^{\delta_i}\{g(Z_i)[1-F(Z_i)]\}^{1-\delta_i} \\ &= \prod_{u} f(Z_i)\prod_{c}[1-F(Z_i)]\prod_{c} g(Z_i)\prod_{u}[1-G(Z_i)], \end{aligned} \tag{2.8}$$

其中 $\prod_{u}$ 表示通过所有 $\delta_i=1$ 的数据(即非截断数据)，$\prod_{c}$ 表示通过所有 $\delta_i=0$ 的数据(即截断数据). 注意，(2.8)式后两个乘积不包含关于 X_i 分布的信息.

如果未知函数 $F(x)(f(x))$是参数形式的，(2.8)则提供了求最大似然估计的途径. 设 $\underset{\sim}{\theta}=(\theta_1,\theta_2,\cdots,\theta_p)'$是关于 X_i 分布的 p 维参数，求它的最大似然估计，就是要寻找$\hat{\underset{\sim}{\theta}}$，使

$$L^*=\Big[\prod_{u} f(Z_i)\Big]\Big\{\prod_{c}[1-F(Z_i)]\Big\}$$

最大. 当 F 的性质较好时(我们所需的各阶导数均存在且连续)，可以用通常求最大似然估计的办法微分 $\ln L^*$，并使之为零来算出 $\hat{\underset{\sim}{\theta}}$，即令

$$\begin{aligned} \frac{\partial}{\partial\theta_j}\ln L^* &= \sum_{u}\frac{\partial}{\partial\theta_j}\ln f(Z_i)+\sum_{c}\frac{\partial}{\partial\theta_j}\ln[1-F(Z_i)] \\ &= \sum_{i=1}^{n}\frac{\partial}{\partial\theta_j}\ln L_i^* \\ &= 0\ (j=1,2,\cdots,p), \end{aligned} \tag{2.9}$$

其中

$$L_i^*=\begin{cases} f(Z_i), & \text{如果 } \delta_i=1, \\ 1-F(Z_i), & \text{如果 } \delta_i=0. \end{cases}$$

$\sum_{u}$ 表示通过所有 $\delta_i=1$ 的数据，$\sum_{c}$ 表示通过所有 $\delta_i=0$ 的数据. 记

$$\frac{\partial}{\partial \underset{\sim}{\theta}}\ln L^{*}=\left(\frac{\partial}{\partial\theta_1}\ln L^{*},\frac{\partial}{\partial\theta_2}\ln L^{*},\cdots,\frac{\partial}{\partial\theta_p}\ln L^{*}\right)',$$

$$\frac{\partial^2}{\partial \underset{\sim}{\theta}^2}\ln L^{*}=\begin{pmatrix}\frac{\partial^2}{\partial\theta_1\partial\theta_1}\ln L^{*}\cdots\frac{\partial^2}{\partial\theta_1\partial\theta_p}\ln L^{*}\\ \vdots\qquad\qquad\vdots\\ \frac{\partial^2}{\partial\theta_p\partial\theta_1}\ln L^{*}\cdots\frac{\partial^2}{\partial\theta_p\partial\theta_p}\ln L^{*}\end{pmatrix},$$

以及

$$I(\underset{\sim}{\theta})=-\frac{\partial^2}{\partial \underset{\sim}{\theta}^2}\ln L^{*},$$

$$I^{*}(\underset{\sim}{\theta})=EI(\underset{\sim}{\theta}).$$

这样，最大似然估计可由(2.9)式求得(通常要用 Newton-Raphson 方法)，并且在正则条件下近似地有

$$\hat{\underset{\sim}{\theta}}\sim N(\underset{\sim}{\theta},I^{*-1}(\underset{\sim}{\theta})). \tag{2.10}$$

这一事实常被用来作假设检验，在原假设 $\underset{\sim}{\theta}=\underset{\sim}{\theta}^{\circ}$ 下，

$$(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}^{\circ})I^{*}(\underset{\sim}{\theta}^{\circ})(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}^{\circ})\sim\chi_p^2\quad(\text{Wald}), \tag{2.11}$$

$$-2\ln\frac{L^{*}(\underset{\sim}{\theta}^{\circ})}{L^{*}\hat{\underset{\sim}{\theta}}}\sim\chi_p^2\quad(\text{Neyman-Pearson}), \tag{2.12}$$

以及

$$\left[\frac{\partial}{\partial\underset{\sim}{\theta}}\ln L^{*}(\underset{\sim}{\theta}^{\circ})\right]'I^{*-1}(\underset{\sim}{\theta}^{\circ})\left[\frac{\partial}{\partial\underset{\sim}{\theta}}\ln L^{*}(\underset{\sim}{\theta}^{\circ})\right]\sim\chi_p^2\quad(\text{Rao}). \tag{2.13}$$

我们用下面两个例子来说明(2.9)的用途.

例 2.1 在随机截断下指数分布参数 λ 的最大似然估计.

记 $n_u=\sum_{i=1}^{n}\delta_i$ 为非截断数据的个数，则

$$L^{*}=\exp\Big(-\lambda\sum_{u}Z_i\Big)\exp\Big(-\lambda\sum_{c}Z_i\Big)=\exp\Big(-\lambda\sum_{i=1}^{n}Z_i\Big)\lambda^{n_u},$$

$$\ln L^{*}=-\lambda\sum_{i=1}^{n}Z_i+n_u\ln\lambda,$$

$$\frac{\partial\ln L^{*}}{\partial\lambda}=-\sum_{i=1}^{n}Z_i+\frac{n_u}{\lambda},$$