文章编号:1001-2486(2001)06-0109-05

小子样下命中概率的估计*

张金槐

(国防科技大学人文与管理学院,湖南 长沙 410073)

摘 要:研究了当试验次数较少时,命中圆域的概率估计。给出了自助估计法和融合估计方法,对于估计的精度进行了分析。方法易于实现,适用于工程应用。

关键词:命中概率;自助方法;Bayes 估计
中图分类号:O212.8 文献标识码:A

Estimating of Fitting Probability under the Circumstance of Small Sample

ZHANG Jin-huai
(College of Humanities and Management, National Univ.of Defense Technology, Changsha 410073, China)

Abstract: The estimation of fitting probability in a circle is studied. The bootstrap method and data-fusion estimation method are given. At the same time, the accuracy of the estimator is analysed. These methods can be applied in engineering practices.

Key words: fitting probability; bootstrap method; Bayes estimation

命中概率估计问题,已有较多论述。在文[1]中对于随机落点落入长方形域及圆域的概率估计进行了讨论。在现场试验数较大的场合,估计的性能较好。然而,在小子样场合,经典估计方法就值得商榷了。

人们早就熟悉 Bayes 小子样理论,它是现场试验次数较少,但具有较充分的验前信息可利用时的一种统计分析方法。还有一种近年来常被采用的自助方法,这个方法的特点是充分应用子样自身的信息,子样容量较小,而不考虑其他验前信息。因此,对于使用验前信息引起争议时,它是常被采用的方法。

1 自助估计方法

先从平面上随机落点问题说起。设(Y,Z)为落点坐标,且 $Y\sim N(m_Y,\sigma_Y^2)$,$Z\sim N(m_Z,\sigma_Z^2)$,Y、Z 独立,此时命中以 R 为半径的圆 C_R 内的概率为

$$P = P\{(Y,Z)\in C_R\} = \iint_{C_R}\frac{1}{2\pi\sigma_Y\sigma_Z}e^{-\frac{1}{2}\left[\frac{(Y-m_Y)^2}{\sigma_Y^2}+\frac{(Z-m_Z)^2}{\sigma_Z^2}\right]}dYdZ$$

如果落点具有圆散布(即 $\sigma_Y=\sigma_Z=\sigma$),且无系统性偏差(即 $m_Y=m_Z=0$),则命中 C_R 圆的概率为

$$P = 1-e^{-\frac{R^2}{2\sigma^2}} \tag{1}$$

我们将在上述前提下,讨论命中概率估计问题。

对于落点(Y,Z),记 $r=\sqrt{Y^2+Z^2}$,此时 r 属于瑞利分布,其分布密度为

$$P(r) = \frac{r}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}},r\geqslant 0 \tag{2}$$

设进行了 n 次试验,落点子样为(Y_i,Z_i),$i=1,\cdots,n$。记

* 收稿日期:2001-05-27
作者简介:张金槐(1930-),男,教授。

$$R_i = \sqrt{Y_i^2 + Z_i^2}, i = 1, \cdots, n$$

可以证明[1],命中 C_r 圆的概率 P 的无偏、最小方差估计(UMVU)为

$$\hat{P} = 1 - \left(1 - \frac{r^2}{t}\right)^{n-1} \tag{3}$$

其中,$t = \sum_{i=1}^{n} R_i^2$ 为 n 次试验中落点偏差(离原点的距离)$R_1, \cdots, R_n$ 的平方和。上述估计公式(3),在试验数 n 较大时,它是一种良好的估计,但当 n 比较小,则必须寻求其他方法。下面讨论在较小子样的情况下,如何充分地利用子样自身的信息,作出尽可能令人满意的结果。这就是自助方法。记

$$\begin{aligned} T_n &= \hat{P} - P = 1 - \left(1 - \frac{r^2}{\sum_1^n R_i^2}\right)^{n-1} - P \\ &= 1 - \left(1 - \frac{r^2}{\sum_1^n R_i^2}\right)^{n-1} - \left(1 - e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2}\right) \\ &= e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2} - \left(1 - \frac{r^2}{\sum_1^n R_i^2}\right)^{n-1} \end{aligned} \tag{4}$$

T_n 表示估计的误差。由子样 $R_1, \cdots, R_n$ 可以作出抽样分布。事实上,由于 $R_i \sim$ 瑞利分布,因此只需对分布的未知分布参数 σ 作估计就可以了。R_i 的概率密度函数为

$$P(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, r \geqslant 0$$

注意到瑞利分布的随机变量 R 具有方差 $\left(2 - \frac{\pi}{2}\right)\sigma^2$,即

$$D[R] = \left(2 - \frac{\pi}{2}\right)\sigma^2$$

因此,当获得子样$(R_1, \cdots, R_n)$后,可作出 $D[R]$的估计,它是

$$\hat{D}[R] = \frac{1}{n-1}\sum_{i=1}^{n}(R_i - \bar{R})^2$$

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i$$

于是 σ^2 的估计为

$$\hat{\sigma}^2 = \frac{1}{\left(2 - \frac{\pi}{2}\right)} \cdot \frac{1}{n-1}\sum_{i=1}^{n}(R_i - \bar{R})^2$$

由此,R 的抽样分布密度为

$$\hat{f}^*(r) = \frac{r}{\hat{\sigma}^2} e^{-\frac{r^2}{2\hat{\sigma}^2}}, r \geqslant 0$$

由抽样分布产生新的子样(再生子样)$R_1^*, \cdots, R_n^*$,由此,又可作出 P 的估计,记作 $\hat{P}(\hat{f}^*)$,它为

$$\hat{P}(\hat{f}^*) = 1 - \left(1 - \frac{r^2}{\sum_1^n R_i^{*2}}\right)^{n-1}$$

作

$$\begin{aligned} S_n^* &\triangleq \hat{P}(\hat{f}^*) - \hat{P} \\ &= \left(1 - \frac{r^2}{\sum_1^n R_i^2}\right)^{n-1} - \left(1 - \frac{r^2}{\sum_1^n R_i^{*2}}\right)^{n-1} \end{aligned} \tag{5}$$

S_n^* 为 T_n 的自助统计量。以 S_n^* 的分布去迫近 T_n 的分布，这是自助法的核心思想。由于再生子样可以重复生成，记

$$R^{*(j)} = (R_1^{*(j)}, \cdots, R_n^{*(j)}), j = 1, \cdots, N$$

它表示第 j 次产生的再生子样，对每个 $R^{*(j)}$ 均可作出 S_n^*，记为 $S_n^{*(j)}$：

$$S_n^{*(j)} = \hat{P}(\hat{f}^{*(j)}) - \hat{P}, j = 1, \cdots, N$$

于是对每个 $S_n^{*(j)}$，可以算出 P 的近似取值，记它为 $P^{(j)}$，即

$$\begin{aligned} P^{(j)} &= \hat{P} - T_n \\ &\simeq \hat{P} - S_n^{*(j)}, j = 1, \cdots, N \end{aligned}$$

这样得到了 P 的 N 个可能取值，于是有 P 的自助估计

$$\hat{P}_B = \frac{1}{N}\sum_{j=1}^{N}[\hat{P} - S_n^{*(j)}] \tag{6}$$

如果需要，还可以去估计出 $\mathrm{Var}(\hat{P})$。事实上，重复上面得到 $\hat{P}$ 的方法，可以得到多个 $\hat{P}$。例如，作出 M 个 P 的估计，如 $\hat{P}^{(S)}, S = 1, \cdots, M$，于是 $\mathrm{Var}(\hat{P}_B)$ 的估计为

$$\begin{aligned} \hat{\mathrm{Var}}(\hat{P}_B) &= \frac{1}{M-1}\sum_{S=1}^{M}(\hat{P}^{(S)} - \bar{P})^2 \\ \bar{P} &= \frac{1}{M}\sum_{S=1}^{M}\hat{P}^{(S)} \end{aligned} \tag{7}$$

还可以用随机加权自助方法，获取 P 的估计。事实上，运用随机加权法时，只需将 $\hat{P}(\hat{f}^*)$ 改作

$$\hat{P}(V) = 1 - \left(1 - \frac{r^2}{\sum_{i=1}^{n} V_i R_i^2}\right)^{n-1}$$

其中 $(V_1, \cdots, V_n)$ 为 $D(1,1,\cdots,1)$ 随机向量，即 $\mathrm{Dirichlet}(1,\cdots,1)$ 随机向量。此时，将(5)式中的 S_n^* 改作

$$S_n(V) = \hat{P}(V) - \hat{P}$$

于是，用产生多组 $D(1,1,\cdots,1)$ 随机向量的方法代替重复产生再生子样 $R^{*(j)}$，如产生 N 组 $D(1,1,\cdots,1)$ 随机向量，记作

$$V^{(j)} = (V_1^{(j)}, \cdots, V_n^{(j)}), j = 1, \cdots, N$$

对每个 $V^{(j)}$，计算出 $S_n^{(j)}(V), j = 1, \cdots, N$。于是可得 P 的随机加权自助估计(相应于公式(6)中的 $\hat{P}_B$)为

$$\hat{P}_V = \frac{1}{N}\sum_{j=1}^{N}(\hat{P} - S_n^{(j)}(V)) \tag{7'}$$

这里顺便指出产生 $D(1,1,\cdots,1)$ 随机向量 $(V_1, \cdots, V_n)$ 的方法。首先产生 $[0,1]$ 上均匀分布的一组随机数 $v_1, \cdots, v_{n-1}$，将它们由小到大重新排序，得次序统计量

$$v_{(1)} \leqslant v_{(2)} \leqslant \cdots \leqslant v_{(n-1)}$$

然后，取 $v_{(0)} = 0, v_{(n)} = 1$，记

$$v_i = v_{(i)} - v_{(i-1)}, i = 1, \cdots, n$$

则 $(V_1, \cdots, V_n)$ 即为 $D(1,1,\cdots,1)$ 随机向量。

一般地说，随机加权自助方法较之重复产生再生子样的自助方法有较高的估计精度。

2 多源信息下命中概率的融合估计

2.1 加权融合估计

如果在现场试验之前，具有多种关于落点的信息，例如数学仿真获取的落点信息、半实物仿真获取的落点信息、其他类似的试验转换而来的信息等。这里讨论具有某种验前的落点信息(数据)和现场试验这两方面信息时的命中概率的融合估计方法，多源信息的情况可仿此进行。

设某种验前落点子样为 $R_1^{(0)},\cdots,R_{n_0}^{(0)}$,则运用前面的方法,可以估算出命中 C_r 圆的概率估计及估计的方差,记作 $\hat{P}^{(0)},\mathrm{Var}(\hat{P}^{(0)})$;由现场试验所获得的命中概率估计及其方差记为 $\hat{P},\mathrm{Var}(\hat{P})$,则最终 P 的融合估计为

$$\hat{P}_{融}=\mathrm{Var}(\hat{P}_{融})[(\mathrm{Var}(\hat{P}^{(0)}))^{-1}\hat{P}^{(0)}+(\mathrm{Var}(\hat{P}))^{-1}\hat{P}] \tag{8}$$

$$[\mathrm{Var}(\hat{P}_{融})]^{-1}=[\mathrm{Var}(\hat{P}^{(0)})]^{-1}+[\mathrm{Var}(\hat{P})]^{-1} \tag{9}$$

这种融合估计方法,其实质是将融合估计看作验前估计和现场估计的线性组合。此时,如果 $\hat{P}^{(0)},\hat{P}$ 均为 UMVU 估计,那么 $\hat{P}_{融}$ 为命中 C_r 圆概率的 UMUV 估计,且

$$\mathrm{Var}(\hat{P}_{融})=\frac{1}{[\mathrm{Var}(\hat{P}^{(0)})]^{-1}+[\mathrm{Var}(\hat{P})]^{-1}}<\frac{1}{[\mathrm{Var}(\hat{P}^{(0)})]^{-1}}=\mathrm{Var}(\hat{P}^{(0)})$$

同样,$\mathrm{Var}(\hat{P}_{融})<\mathrm{Var}(\hat{P})$,这说明融合估计较之 $\hat{P}^{(0)}$ 或 $\hat{P}$ 具有较高的精度。

这里需要指出的是上述融合估计是在验前信息与现场信息为相容的情况下得出的。

如果有多种验前信息,则可按(8)、(9)式分层逐次融合,直至获得最后的多源信息融合估计值。

2.2 准 Bayes 融合估计方法

设现场试验之前,具有落点的信息,它表示为验前子样 $R_1^{(0)},\cdots,R_{n_0}^{(0)}$,而现场试验的落点子样为

$$R_1,\cdots,R_n$$

在上述情况下,要去确定命中 C_r 圆的概率 $P=1-\mathrm{e}^{-\frac{r^2}{2\sigma^2}}$ 的估计。为此,先作出 $\sigma^2\triangleq D$ 的 Bayes 估计。

$D=\sigma^2$ 为瑞利分布的参数,为了对 D 进行 Bayes 估计,只需计算出 D 的验密度 $P(D\mid R_1,\cdots,R_n)$。先注意下列引理[3]

引理 若 D 的验前密度为逆 Gamma 分布 $g(D;\alpha_0,\beta_0)$,

即
$$g(D;\alpha_0,\beta_0)=\frac{\alpha_0^{\beta_0}}{\Gamma(\beta_0)}D^{-(\beta_0+1)}\mathrm{e}^{-\frac{\alpha_0}{D}},D>0$$

则当获得子样 $R_1,\cdots,R_n$ 后,D 的验后密度仍为逆 Gamma 分布,它为

$$P(D\mid R_1,\cdots,R_n)=g(D;\alpha_1,\beta_1) \tag{10}$$

其中,分布参数 α_1,β_1 为

$$\begin{cases}\alpha_1=\alpha_0+\dfrac{S^2}{2}\\ \beta_1=\beta_0+n\end{cases} \tag{10'}$$

式中,$S^2=\sum_{i=1}^{n}R_i^2$。α_0,β_0 是验前分布的参数,它由验前信息确定。如可取

$$\alpha_0=\frac{1}{2}\sum_{i=1}^{n_0}(R_i^{(0)})^2,\beta_0=n_0$$

在平方损失函数之下,D 的 Bayes 估计为 $\mathrm{E}[D\mid R_1,\cdots,R_n]$,它为逆 Gamma 分布的期望值,即

$$\mathrm{E}[D\mid R_1,\cdots,R_n]=\frac{\alpha_1}{\beta_1-1}=\frac{1}{n_0+n}\left[\frac{1}{2}\sum_{i=1}^{n_0}(R_i^{(0)})^2+\frac{1}{2}\sum_{i=1}^{n}R_i^2\right] \tag{11}$$

记此估计为 $\hat{D}_B\triangleq\hat{\sigma}_B^2$,则取命中 C_r 圆的概率的

$$\hat{P}'=1-\mathrm{e}^{-\frac{r^2}{2\hat{\sigma}_B^2}} \tag{12}$$

称 $\hat{P}'$ 为准 Bayes 估计或伪(pseudo)Bayes 估计。这种估计,工程上易于实现。且知 D 的验后方差为

$$\mathrm{Var}[D\mid R_1,\cdots,R_n]=\frac{\alpha_1^2}{(\beta_1-1)^2(\beta_1-2)} \tag{13}$$

式中 α_1,β_1,由(10′)表示。

下面作出 $\hat{P}'$的置信分布,记

$$\hat{\sigma}_B^2 = a + bt$$

式中,
$$\alpha = \frac{1}{2(n_0+n)}\sum_{i=1}^{n_0}(R_i^{(0)})^2, b = \frac{1}{2(n_0+n)}, t = S_n^2$$

在 $\sigma=1$ 的情况下,$t\sim\chi^2(2n)$(如果 $\sigma\neq1$,则 $t/\sigma^2\sim\chi^2(2n)$),注意到

$$\hat{P}' = 1 - e^{-\frac{r^2}{2(a+bt)}}$$

且 $d\hat{P}'/dt<0$。因此 $\hat{P}'$为 t 的单调递减函数,利用 $t\sim\chi^2(2n)$,可建立如下关系:

$$P\{t > x_\alpha^2\} = 1-\alpha$$

式中 α 与 x_α^2 的关系可查 χ^2—分布表获得。当 $t>x_\alpha^2$ 时,它等价于 $\hat{P}'<J_\alpha$,此处

$$J_\alpha = 1 - e^{-\frac{r^2}{2(a+bx_\alpha^2)}}$$

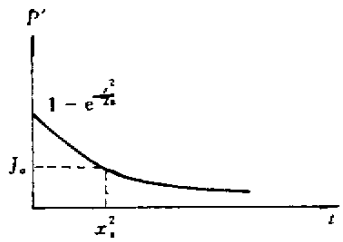

图 1 $\hat{P}'\sim t$ 关系曲线图

Fig.1 Relation of $\hat{P}'\sim t$

于是,$P\{0<\hat{P}'<J_\alpha\}=1-\alpha$,即在置信概率 $1-\alpha$ 之下,$\hat{P}'$的置信区间为$(0,J_\alpha)$。见图 1。

参考文献:

[1] 张金槐 . 命中概率估计[J]. 飞行器测控学报,2001.

[2] 张金槐,唐雪梅 .Bayes 方法[M]. 长沙 :国防科技大学出版社,第二版,1995.

[3] 张金槐 . 再入飞行器落点精度散布鉴定中验前概率的确定[J]. 飞行器测控技术,1991(2) :1-7.