文章编号:1002-2082(2006)01-0073-06

目标探测分布律K-P法检验设计及应用

郝继平，李昕泽，孙宪全
(中国人民解放军92941部队94分队,辽宁 葫芦岛 125001)

摘 要： 对试验数据进行科学合理的误差分析是提高靶场科研鉴定试验质量的一项重要内容。若使用未经分布律检验的试验数据进行精度分析,很可能影响鉴定和定型结论的正确性,该文借助于Excel,Matlab和Origin等最新数据处理、绘图软件技术对大量数据实地研究,取得了如下结论:部分样本和总体标本分段后是符合正态分布的,进一步提出将柯尔莫戈洛夫(Kolmogorov)与皮尔逊(Pearson)χ^2两种方法作为靶场目标探测中大量数据分布律检验的统一规范,这样可大幅度提高试验的可信度。合理选择置信水平,可使试验航次数降到原试验航次数的10%~50%。该方法可推广到各类系统和单机试验中去。

关键词： 目标探测;靶场试验;分布律;K-P检验法;检验设计
中图分类号:O212.6 **文献标识码:**A

Design and application of K-P test for target detection distribution law

HAO Ji-ping, LI Xin-ze, SUN Xian-quan
(Unit of 92941, PLA, Liaoning 125001, China)

Abstract: Scientific error analysis for the test data is critical to improve the R&D product evaluation quality of the test range. Using testing data not verified by distribution law for accuracy analysis will mislead the evaluation and final conclusion. With the help of new data processing technique and drawing software (such as Excel, Matlab, Origin and so on), a lot of test data were studied, it is concluded the partial sample data and the whole sample data conform with normal distribution after they are separated into some parts. Kolmogorov and Pearson x^2 methods are introduced to function as a unified norm for verifying the distribution of large amount of data obtained from target detection in the test range. This norm can greatly increase the credibility of the test. If the confidence level is reasonably selected, the number of flights can be decreased to 10%~50%. The method has the potential to be used in almost all the system and device testing.

Key words: target detection; test range; distribution law; Kolmogorov-Pearson test method; test design

引言

在目标探测试验数据精度分析中,靶场通常将测量值和被测设备录取值的一次差数据作为一个总体来考虑,假定它为正态分布,然后对其均值或方差进行评价。因缺乏误差分析规范,经常出现多次处理之后,结果互不一致,加之不能实时处理,不但延长了测试周期,而且影响了鉴定和定型结论的正确性。

收稿日期:2005-11-15; 修回日期:2005-11-28

作者简介:郝继平(1956—),男,山西太原人,高级工程师,主要从事海上靶场舰载雷达、光电系统的鉴定,设计定型试验和测试等工作。

本文借助于Excel、Matlab和Origin等最新数据处理、绘图软件技术,在大量数据处理的基础上,提出了将柯尔莫戈洛夫(Kolmogorov)与皮尔逊(Pearson)x^2(简称K-P)2种方法作为靶场试验专用分布律检验设计方法,还分析了应用中出现的典型问题。

1 正态分布律检验设计

正态分布律检验的方法中,正态概率纸检验、W检验、D检验、符号检验及秩和检验都不适合靶场大数据量($n \geqslant 1\ 000$)定量分析的要求。斯米尔诺夫(Smirnov)检验借助于经验分布函数,只是比较了2个总体的真分布是否相同,它也不适用[1]。皮尔逊χ^2检验法使用范围广,可用于全样本,也可用于截尾样本。但由于它是分组处理样本的观察值,有时虽然原假设$F(x)=F_0(x)$不成立,但在另一种划分下,可有$F(a_i)-F(a_{i-1})=F_0(a_i)-F_0(a_{i-1})=p_i$,其中$i=1,2,\cdots,k$。此时虽然$F(x)\neq F_0(x)$,但这种划分并不影响统计量$\chi^2$的观察值。特别是对于截尾样本,分布的后一段可以完全不一样,而χ^2的数值却可以一样,因而很容易犯第2类错误,或称“取伪”错误,即把不真的假设H_0也接受下来。由于其精度要求比较低,数据量不限,虽存在“取伪”的可能,但相对来说还是一种较好的检验方法。

柯尔莫戈洛夫检验可以克服这一缺点,其正态分布律检验是单个样本的分布拟合问题的检验方法,它比皮尔逊χ^2检验更为精细。虽然该方法精度要求高,较难找到满足条件的大量数据,但它科学严谨,目前是公认的最好方法。

1.1 设计原理

1.1.1 K法设计原理

假设统计量H_0: $F(x)=F_0(x)$

其中,H_0为基本假设,又称零假设或原假设;$F(x)$是总体分布函数;$F_0(x)$是完全已知的连续型分布函数。

我们知道,当样本容量n充分大时,经验分布函数$F_n(x)$与总体分布函数$F(x)$相当接近。所以,当H_0成立且n较大时,$F_n(x)$与$F_0(x)$的差距不太大。鉴于此,俄国数学家柯尔莫戈洛夫提出用统计量

$$D_n=\sup_{-\infty<x<+\infty}|F_n(x)-F_0(x)|$$

作为H_0的检验统计量,并导出了D_n的精确分布和$\sqrt{n}D_n$的极限分布$Q(z)$。D_n的精确分布为[1]

$$P\left(D_n<Z+\frac{1}{2n}\right)=\begin{cases}0 & Z<0\\ \int_{\frac{1}{2n}-Z}^{\frac{1}{2n}+Z}\int_{\frac{3}{2n}-Z}^{\frac{3}{2n}+Z}\cdots\int_{\frac{2n-1}{2n}-Z}^{\frac{2n-1}{2n}+Z}f(y_1,\cdots,y_n)(\mathrm{d}y_1,\cdots,\mathrm{d}y_n) & 0\leqslant Z<\dfrac{2n-1}{2n}\\ 1 & Z\geqslant\dfrac{2n-1}{2n}\end{cases}$$

其中

$$f(y_1,\cdots,y_n)=\begin{cases}n! & 0<y_1<\cdots<1\\ 0 & \text{其它}\end{cases}$$

$\sqrt{n}D_n$的极限分布为

$$Q(z)=\lim_{n\to\infty}P(\sqrt{n}D_n<Z)=\sum_{k=-\infty}^{\infty}(-1)^k e^{-2Z^2k^2}=1-2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2Z^2k^2}$$

由于$F_n(x)$和$F_0(x)$都是x的单调非降函数,所以偏差$|F_n(x)-F_0(x)|$的上限可在n点$X_{(i)}$处找。因而,柯尔莫戈洛夫检验是在样本的每个顺序统计量$X_{(i)}$上求样本经验分布函数和假设的分布函数之间的偏差中较大的,即

$$d_i=\max\left\{\left|F_0(X_{(I)})-\frac{i-1}{n}\right|,\left|\frac{i}{n}-F_0(X_{(i)})\right|\right\}$$

其中,$i=1,2,\cdots,n$; 最大的一个d_i就是柯尔莫戈洛夫检验统计量D_n的取值,即$D_n=\max\{d_1,d_2,\cdots,d_n\}$。

若$F_n(x)$与$F_0(x)$拟合很好,则D_n的值比较小;反之,当D_n较大时,$F_n(x)$和$F_0(x)$拟合得不好。

柯尔莫戈洛夫检验的规则是:对给定的显著性水平$\alpha(0<2<1)$,若$D_n>D_{n,\alpha}$,则拒绝H_0,否则接受H_0。当$n<100$时,可从文献[1]的附表6查得临界值$D_{n,\alpha}$,使$P(D_n\geqslant D_{n,\alpha})=\alpha$。当$n\geqslant 100$时,可从柯尔莫戈洛夫检验统计量$D_n$的极限分布表(见文献[1]中附表7)查得$\lambda_{1-\alpha}$值,使$Q(\lambda_{1-\alpha})=1-\alpha$,即$P(D_n\geqslant\lambda_{1-\alpha}/\sqrt{n})\approx 1-Q(\lambda_{1-\alpha})=\alpha$。此时,$D_{n,\alpha}\approx\lambda_{1-\alpha}/\sqrt{n}$。

当样本$X_1,X_2,\cdots,X_n$中没有重复数据时,可按前式计算D_n的值。当样本$X_1,X_2,\cdots,X_n$中有重复数据时,可按如下步骤要求取D_n值:将样本$X_1,X_2,\cdots,X_n$从小到大排列(重复数据合并为一个),即$X_{(1)}<X_{(2)}<\cdots<X_{(m)}$,其中,$1\leqslant m\leqslant n$。设$n_i$为对应$X_{(i)}$在样本中出现的频数,则有

$n_i \geqslant 1, n_1+n_2+\cdots+n_m=n$;

$F_n(X_{(i)})=\dfrac{n_1+n_2+\cdots+n_{i-1}}{n}$

令 $d_i=\max\{|F_0(X_{(i)})-F_n(X_{(i)})|, |F_n(X_{(i+1)})-F_0(X_{(i)})|\}, i=1,2,\cdots,m$。

则 $F_n(X_{(m+1)})=1, D_n=\max\{d_1,d_2,\cdots,d_m\}$

1) 正态分布律的K法检验

通过样本 $X_1,X_2,\cdots,X_n$ 来检验总体的分布是否服从正态 $N(\mu,\sigma^2)$。由于 $N(\mu,\sigma^2)$ 中有2个未知参数 μ 和 σ^2,不能直接用柯尔莫戈洛夫统计量 D_n 来检验正态分布的假设。对于未知参数 μ 和 σ^2,如果用其无偏估计量 $\hat{\mu}$ 和 $\hat{\sigma}^2$,即

$$\hat{\mu}=\overline{X}$$

$$\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

代替,则要检验的假设实际上是

$$H_0: F(x)=F_0(x,\overline{X},\hat{\sigma}^2)=\frac{1}{\sqrt{2\pi}\hat{\sigma}}\int_{-\infty}^{x}\exp\left\{-\frac{(t-\overline{X})^2}{2\hat{\sigma}^2}\right\}\mathrm{d}t$$

最初的柯尔莫戈洛夫正态分布律检验中,取检验统计量为

$$\hat{D}_n=\sup_{-\infty<x<\infty}|F_0(x,\overline{X},\hat{\sigma}^2)-F_n(x)|$$

对于给定的显著性水平 α,检验规则为:若 $\hat{D}_n \geqslant \hat{D}_{n\alpha}$,则拒绝 H_0;否则,接受 H_0。为求 $\hat{D}_{n\alpha}$,需要知道 $\hat{D}_n$ 的分布函数,显然 $\hat{D}_n$ 的分布不同于 D_n 的分布。对于 $\alpha=0.01, 0.05, 0.10, 0.15, 0.20$, Lilliefors 计算了 $\hat{D}_n$ 的临界值 $\hat{D}_{n\alpha}$(见文献[1]中附表8)。$\hat{D}_n$ 值的计算方法与 D_n 值的计算方法相同。

2) 改进后的正态分布律K法检验

1971年Finklestein和Schafter提出了改进的正态分布律检验的统计量。改进的统计量 H_0 为

$$S_n^*=\sum_{i=1}^{n}d_i$$

式中,S_n^* 为提高检验功效。当 $\alpha=0.01, 0.05, 0.10, 0.15, 0.20$ 时,求出 S_n^* 的临界值 $S_{n\alpha}^*(n\leqslant 30)$,见文献[1]中附表9。

显然,$d_i(i=1,2,\cdots,n)$ 值大,S_n^* 的值也大,此时分布函数 $F_0(x,\hat{\theta})$ 的曲线与经验分布函数 $F_n(x)$ 拟合不好,应拒绝 H_0。对显著性水平 α,检验的规则为:若 $S_n^*\geqslant S_{n\alpha}^*$ 时,拒绝 H_0,否则就接受 H_0。因表[1]中 $n\leqslant 30$,所以它对于大数据量的情况不适用。

1.1.2　P法设计原理

设总体 X 的值为 $1,2,\cdots,r$ 个离散型随机变量,其取值的概率为 p_i,即

$$P\{X=i\}=p_i$$

式中,$i=1,2,\cdots,r$;$\sum_{i=1}^{r}p_i=1$。设 $X=(X_1,X_2,\cdots,X_n)$ 是从总体样本中抽取的简单随机子样,$(x_1,x_2,\cdots,x_n)$ 是子样观察值,n_i 表示 $(x_1,x_2,\cdots,x_n)$ 中取值为 i 的个数,即子样中出现 $(X=i)$ 的频数,则得到 $n_1,n_2,\cdots,n_r$ 中每个 n_i 都是子样的函数,所以它们也是随机变量,并且 $\sum_{i=1}^{r}n_i=n$ 和 $(n_1,n_2,\cdots,n_r)$ 服从多项分布。它的概率分布是

$$\frac{n!}{n_1!\ n_2!\ \cdots n_n!}p_1^{n_1}\times p_2^{n_2}\times\cdots\times p_r^{n_r}$$

我们知道,频率是概率的反映,如果总体的概率分布是 $(p_1,p_2,\cdots,p_r)$,那么当观察个数 n 愈来愈大时,频率 n_i/n 与 p_i 之间的差异会愈来愈小。因此,频率 n_i/n 和 p_i 之间的差异程度可以反映出 $(p_1,p_2,\cdots,p_r)$ 是否为总体的真实概率分布。

Pearson首先提出运用下面的统计量来衡量它们的差异程度。

$$\chi^2=\sum_{i=1}^{r}\frac{(n_i-np_i)^2}{np_i}$$

这个统计量又称为Pearson统计量。如果 $(p_1,p_2,\cdots,p_r)$ 是总体服从其真实概率分布,则统计量 χ^2 要小些;若它不服从其真实概率分布,则 χ^2 就有偏大的趋势。因此可以用它来作为多项分布的检验统计量。

1.2　检验步骤

1.2.1　K法检验步骤[2]

目标探测试验数据分布律检验步骤如下[1]:

1) 提出假设确定分布及参数。

假设 H_0: $F(x)=F_0(x,\overline{X},\hat{\sigma}^2)=\Phi(\dfrac{x-\mu}{\sigma})$

H_1: $F(x)\neq F_0(x,\overline{X},\hat{\sigma}^2)=\Phi(\dfrac{x-\mu}{\sigma})$

式中,$F_0(x)$ 是已知的分布;$F(x)$ 转换为标准正态分布。设2者有相同分布。S 是未知参数的个数,正态分布为2个。然后计算未知参数无偏估计 $\hat{\mu}$ 和 $\hat{\sigma}$,其中自由度为 $r-s-1$。

2) 确定统计量及参数

统计量 $\hat{D}_n=\sup\limits_{-\infty<x<\infty}|F_0(x,\overline{X},\hat{\sigma}^2)-F_n(x)|$。

参数为 n_i, $\Phi(ui)$, $F_N(X_{(I)})$, $F_N(X_{(I+1)})$, $|\Phi(ui)-F_N(X_{(I)})|$, $|F_n(X_{(i+1)})-\Phi(ui)|$, d_i。$d_i=\max\{|F_0(X_{(i)})-F_n(X_{(i)})|, |F_n(X_{(i+1)})-F_0(X_{(i)})|\}$;$S_n^*$。

3) 选取置信度

根据否定域 $\hat{D}_n \geqslant \hat{D}_{n\alpha}$,查出置信限 $\hat{D}_{n\alpha}$;根据否定域 $S_n^* \geqslant S_{n\alpha}^*$,查出置信限 $S_{n\alpha}^*$。通常选取 $\alpha=5\%$,查 $\hat{D}_n$ 的临界值($\hat{D}_{n\alpha}$)表,得 $\hat{D}_{n\alpha}$。查 S_n^* 的临界值($S_{n\alpha}^*$)表,得 $S_{n\alpha}^*$。

4) 计算统计量 $\hat{D}_n$ 和 S_n^*

$$\hat{D}_n = \max\{d_1, d_2, \cdots, d_m\}$$

$$S_n^* = \sum_{i=1}^{n} d_i$$

5) 统计推断

当 $\hat{D}_n < \hat{D}_{n\alpha}$ 时,接受 H_0,当 $\hat{D}_n \geqslant \hat{D}_{n\alpha}$ 时,拒绝 H_0;或 $S_n^* < S_{n\alpha}^*$ 时,接受 H_0, $S_n^* > S_{n\alpha}^*$ 时,拒绝 H_0。2 者接受 H_0 时结果无显著差异。

1.2.2 P 法检验步骤[2]

目标探测试验数据分布律 P 法检验步骤如下:

1) 提出基本假设,确定分布及参数

假设 H_0: $F(x) = F_0(x)$,确定分布参数。$F_0(x)$ 是已知的分布(一般为正态分布),设 2 者分布相同,计算分布参数 $\hat{\mu}$、$\hat{\sigma}$。

2) 确定统计量及参数

统计量中自由度为 $r-s-1$,$n\hat{p}_i \geqslant 5$, $n_i \geqslant 5$。设统计量分为 r 个区间,列表计算 n_i,$n\hat{p}_i$,$n_i - n\hat{p}_i$,$(n_i - n\hat{p}_i)^2$,$(n_i - n\hat{p}_i)^2/n\hat{p}_i$; S 是未知参数的个数,正态分布为 μ 和 σ。统计量可表示成:

$$\hat{\chi}^2 = \sum_{i=1}^{r} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

3) 选取置信度

$\alpha = 5\%$。

4) 查出置信限,确定否定域

查 χ^2 检验表,当自由度为 $r-s-1$,$\alpha=5\%$ 时,得到 χ_α^2 的值。当确定否定域 $\hat{\chi}^2 \geqslant \chi_\alpha^2$ 时,否定 H_0,拒绝接受。

1.3 检验结果

1.3.1 K 法检验

在显著性水平下 $\alpha = 0.1$ 时,检验假设为

$$H_0: F(x) = \Phi(\frac{x-\mu}{\sigma})$$

$$H_1: F(x) \neq \Phi(\frac{x-\mu}{\sigma})$$

其中 $F(x)$ 为分布函数; $\Phi(x)$ 为标准正态分布函数。解 $\Phi(\frac{x-\mu}{\sigma})$ 中含有的 2 个未知参数 μ 和 σ^2,其无偏估计为

$$\hat{\mu} = \overline{X} = 1.404$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = 0.0478^2$$

将某靶场试验数据中100 个采样值[4]按从小到大的次序排列(重复数据合并为1 个),如表1 中的第1 列。表1 给出了各相关项的计算值。其中,d_i 中的最大者就是统计量$\hat{D}_n$ 的值。

表1 统计量计算表

Table 1 Calculation for statistics

$X_{(i)}$	n_i	$ui=\frac{X_{(i)}-\hat{\mu}}{\hat{\sigma}}$	$\Phi(ui)$	$F_n(X_{(i)})$	$F_n(X_{(i+1)})$	$\lvert\Phi(ui)-F_n(X_{(i)})\rvert$	$\lvert F_n(X_{(i+1)}-\Phi(ui))\rvert$	d_i
1.27	1	−2.803	0.002 555	0	0.01	0.002 56	0.007 445	0.007 445
1.30	1	−2.176	0.014 63	0.01	0.02	0.004 63	0.005 37	0.005 37
1.31	1	−1.967	0.024 42	0.02	0.03	0.004 42	0.005 58	0.005 58
1.32	2	−1.757	0.039 2	0.03	0.05	0.009 2	0.010 8	0.010 8
1.34	4	−1.339	0.090 12	0.05	0.09	0.040 12	0.000 12	0.040 12
1.35	3	−1.13	0.129 2	0.09	0.12	0.039 2	0.009 2	0.039 2
1.36	7	−0.921	0.178 8	0.12	0.19	0.058 8	0.011 2	0.058 8
1.37	11	−0.711	0.238 9	0.19	0.30	0.048 9	0.061 1	0.061 1
1.38	4	−0.502	0.308 5	0.3	0.34	0.008 5	0.031 5	0.031 5
1.39	9	−0.293	0.385 9	0.34	0.43	0.045 9	0.044 1	0.045 9
1.40	7	−0.084	0.468 1	0.43	0.50	0.038 1	0.031 9	0.038 1
1.41	7	0.125 5	0.551 7	0.50	0.57	0.051 7	0.018 3	0.05 17
1.42	15	0.334 7	0.629 3	0.57	0.72	0.059 3	0.090 7	0.090 7
1.43	5	0.543 9	0.705 4	0.72	0.77	0.014C6	0.064 6	0.064 6
1.44	5	0.753 1	0.773 4	0.77	0.82	0.003 4	0.046 6	0.046 6
1.45	6	0.962 3	0.831 5	0.82	0.88	0.011 5	0.048 5	0.048 5
1.46	2	1.171 5	0.879	0.88	0.90	0.001	0.021	0.021
1.47	2	1.380 8	0.916 21	0.90	0.92	0.016 21	0.003 79	0.016 21

(续表1)

$X_{(i)}$	n_i	$ui=\frac{X_{(i)}-\hat{\mu}}{\hat{\sigma}}$	$\Phi(ui)$	$F_n(X_{(i)})$	$F_n(X_{(i+1)})$	$\lvert\Phi(ui)-F_n(X_{(i)})\rvert$	$\lvert F_n(X_{(i+1)}-\Phi(ui))\rvert$	d_i
1.48	4	1.59	0.944 08	0.92	0.96	0.024 08	0.015 92	0.024 08
1.49	1	1.799 2	0.964 07	0.96	0.97	0.004 07	0.005 93	0.005 93
1.50	1	2.00 84	0.977 25	0.97	0.98	0.007 25	0.002 75	0.007 25
1.53	1	2.636	0.995 855	0.98	0.99	0.014 586	0.005 855	0.015 855
1.55	1	3.054 4	0.998 856	0.99	1	0.008 86	0.001 144	0.008 856

经计算[6]得出$\hat{D}_n=0.090\ 7$。从文献[1]中附表8查得$\hat{D}_{n\alpha}=\hat{D}_{100,0.01}=\frac{1.031}{\sqrt{100}}=0.103\ 1$。因$0.090\ 7<0.103\ 1$,故在水平$\alpha=0.1$时接受$H_0$,即认为服从正态分布$N(1.404,(0.047\ 8)^2)$。

若取$\alpha=0.05$,则$\hat{D}_{n\alpha}=\hat{D}_{100,0.05}=\frac{0.866}{\sqrt{100}}=0.086\ 6$。因为$0.090\ 7>0.086\ 6$,故$\alpha=0.05$时拒绝$H_0$。不服从正态分布。

1.3.2　P法检验

若用Pearsonχ^2检验,显著水平$\alpha=0.05$时接受H_0。

根据所给数据[5]得:$n=100$,$\hat{\mu}=\overline{x}=1.404\ 2$,$\hat{\sigma}=0.047\ 76$。在$\chi^2$检验中,要求分组时每组中的观察个数不少于5个。根据数据,按0.4分段等间隔地分5个区,如表2所示。

表2中自由度是$5-2-1=2$。对于$\alpha=0.05$,得$\chi^2_{0.05}(2)=5.991$,从表2可知$\hat{\chi}^2=7.679>5.991=\hat{\chi}^2_{0.05}$。因此,我们认为试验结果与正态分布这一假设有显著差异。对于$\alpha=0.02$,得$\chi^2_{0.02}(2)=7.824$,$\hat{\chi}^2=7.679<\chi^2_{0.02}(2)=7.824$,有显著差异。

表2　频数计算表

Table 2　Calculation for frequency data

频数n_i	概率$\hat{p}_i$	理论频数$n\hat{p}_i$	$n_i-n\hat{p}_i$	$(n_i-n\hat{p}_i)^2$	$(n_i-n\hat{p}_i)^2/n\hat{p}_i$
5	0.090	9.021	−4.012	16.096	1.786
29	0.215	21.488	7.512	56.430	2.626
38	0.324	32.43	5.570	31.025	0.957
18	0.250	24.97	−6.970	48.581	1.946
10	0.121	12.1	−2.100	4.410	0.364
100					7.679

1.3.3　K-P法检验结果

以上分析说明检验结果基本相同。K法在$\alpha=0.1$水平下接受H_0,P法在$\alpha=0.05$水平下接受,但检验的精度略有区别。

2　检验结果讨论

2.1　检验A组结果

检验某试验数据,A组为总体的不分距离段的数据,看它是否服从正态分布。

根据误差数据样本得$n=1\ 385$,取值范围在$[-1.33,0.28]$。$\hat{\mu}=\overline{x}=-0.446\ 6$,$\hat{\sigma}=0.171$,$\hat{\sigma}^2=0.029\ 24$。

K法检验结果是:通常取$\alpha=0.05$,统计推断结果为,当$\hat{D}_n<\hat{D}_{n\alpha}$时,接受$H_0$;当$\hat{D}_n\geqslant\hat{D}_{n\alpha}$时,否定$H_0$。$\hat{D}_n=0.20\geqslant\hat{D}_{n\alpha}=0.024$,否定$H_0$。

P法检验结果是:选取$\alpha=0.05$。当$\hat{\chi}^2<\chi^2_\alpha$时,接受$H_0$;当$\hat{\chi}^2\geqslant\chi^2_\alpha$时,否定$H_0$。经计算$\hat{\chi}^2=255.9>\hat{\chi}^2_{0.05}=1.41$,拒绝接受。因此,我们认为试验结果与这一假设正态分布有显著差异。

由以上分析可以看出,K-P法2种检验结果一致,即总体不分距离段数据,且假设为正态分布,不能接受H_0。这说明,对总体数据进行正态分布的检验是十分必要的,其结果往往不符合正态分布的规律,不作区别地全部按正态分布律进行评价是不合理的。

2.2　检验D组结果

D组为数个误差序列的组合。

根据D组误差数据样本可得:

$n=281$,$\hat{\mu}=\overline{x}=-0.417\ 4$,$\hat{\sigma}=0.098\ 69$,$\hat{\sigma}^2=0.009\ 74$。

K法检验结果是:选取$\alpha=0.05$,并按0.02等间隔地分了27个区。经计算得$\hat{D}_n=0.078\ 6\geqslant\hat{D}_{n\alpha}=(0.886/\sqrt{n})=0.053$,因此否定$H_0$。尽管结果是否定,但2数值0.078 6和0.053比较接近。$S_n^*=$

14.57，$S^*_{281,\alpha} \geqslant S^*_{30,\alpha}=2.59$,所以$S^*_n$ 的算法不适用。选取$\alpha=0.1$则可以接受。

P法检验结果是:选取$\alpha=0.05$,那么$\hat{\chi}^2=8.33<\hat{\chi}^2_{1-0.05}=16.92$,因此检验结果与这一假设正态分布没有显著差异。

K-P法2种检验表明,由若干个误差数据组成的D组误差数据,其样本较符合正态分布律。

2.3　检验B组结果

B组为单一误差序列误差数据。

根据误差数据样本可得$n=151$,$\hat{\mu}=\overline{x}=-0.432\ 52$,$\hat{\sigma}=0.108\ 04$,$\hat{\sigma}^2=0.011\ 7$。

K法检验结果是:选取$\alpha=0.05$,经计算得$\hat{D}_{151}=0.19 \geqslant \hat{D}_{151,0.05}=0.072$,否定$H_0$。尽管是否定的结果,但2数值0.19、0.072也比较接近,接近符合正态分布律。若降低要求,选取$\alpha=0.1$可以接受。

P法检验结果是:选取$\alpha=0.05$。统计推断结果为$\hat{\chi}^2<\chi^2_2$时,接受H_0;当$\hat{\chi} \geqslant \chi^2_\alpha$时,否定$H_0$,显然$\hat{\chi}^2=1.94<\hat{\chi}^2_{0.05}=5.991$。因此我们认为试验结果与这一假设正态分布没有显著差异。

K-P法2种检验表明,单一误差序列误差数据组成的B组,其样本较符合正态分布律。

2.4　分段后总体分布结果

检验C组试验数据,取$n=1\ 200$,$\hat{\mu}=\overline{x}=-0.404\ 89$,$\hat{\sigma}=0.105\ 62$,$\hat{\sigma}^2=0.011\ 156$。

K法检验结果是:选取$\alpha=0.05$。经计算得,$\hat{D}_n=0.18 \geqslant \hat{D}_{n\alpha}=0.886/\sqrt{n}=0.025\ 6$,否定$H_0$。尽管是否定的结果,但2数值0.18、0.03相差不算大。若降低要求,选取$\alpha=0.1$,可以接受。

P法检验结果是:选取$\alpha=0.05$。按0.1等间隔地分了6个区。经计算得$\hat{\chi}^2=4.13<\hat{\chi}^2_{0.05}=7.82$,因此检验试验结果与这一假设正态分布没有显著差异。

K-P法2种检验表明,总体误差数据分段后,其样本较符合正态分布律。

2.5　综合误差的检验方法

综上所述,总体不分距离段的结果与正态分布存在很大差异。在这种情况下,应该采用单航次或距离分段方式进行检验和评价。

一台观测设备中包含着许多误差源,要评定该设备的精度,就需要把各种因素的误差进行综合。当只用1个指标来表示一台测量设备的精度时,需把随机误差与系统误差2者进行综合,这个指标称为准确度。它表征所有测量结果与真值的符合程度。

对系统误差的合成及随机误差与系统误差的合成还没有统一的、普遍接受的方法。这个问题不单纯是数学问题,特别是当把各系统误差分量看成随机量进行合成时,要确定它们的实际分布将会非常困难,而且选择不同的分布形式将导致不同的置信因子。因此可采用系数合成法、广义方和根法和绝对值和法。一方面要改进试验数据的测量录取质量,另一方面应考虑采用综合误差的处理方法。

3　结束语

随着靶场试验理论、试验方法研究的不断深入,目标探测数据的有效处理对试验评定可以产生重大影响。本文提出的将Kolmogorov法与Pearsonχ^2法用于靶场测量的数据分布律的检验切实可行。目前已用于靶场数据分析及处理系统中,可直接服务于武器系统的科研及生产。这2种方法为靶场目标探测多组大量数据分布律检验提供了统一规范,可大幅度提高可信度。如果合理选择置信水平,试验航次数可以降到原航次数的10%～50%。该方法可推广到各类系统和单机的实际试验结果分析中。

参考文献:

[1]　吴翊,李永乐,胡庆军.应用数理统计[M].长沙:国防科技大学出版社,1995:98-130.

[2]　周概容.概率论与数理统计[M].北京:高等教育出版社,1984:493-507.

[3]　刘海燕.概率论与数理统计(下)[M].北京:国防工业出版社,2001:69-79.

[4]　石博强,滕贵法,李海鹏,等.MATLAB数学计算范例教程[M].北京:中国铁道出版社,2004:133-178.

[5]　孙志刚,杨聪.Excel在经济与数理统计中的应用[M].北京:中国电力出版社,2004:109-113.

[6]　郝红伟,施光凯.数据分析和绘图软件——Origin实例教程[M].北京:中国电力出版社,2003:104-129.