DOI: 10.13203/j.whugis1998.03.014
第 23卷第 3期 武 汉 测 绘 科 技 大 学 学 报 Vol.23 No.3
1998年 9月 Journal of Wuhan Technical University of Surveying and Mapping Sept. 1998

p-范分布母体抽样分布的数字特征*

孙海燕 於宗俦

(青岛市勘察测绘研究院,青岛市人民路 2号,266033)

摘 要 给出了 p-范分布子样的 3个抽样分布—— i^p 分布、t_p 分布及 F_p 分布的分布函数及数学期望、方差与不确定度区间的计算公式,并对若干有关分布的应用问题进行讨论。

关键词 i^p 分布;t_p 分布;F_p 分布;熵;不确定度区间
分类号 P207.1

文献 [1]与文献 [2]分别定义了一元 p-范分布及 p-范母体的 3个抽样分布—— i^p 分布、t_p 分布和 F_p 分布。这样,对于一元 p-范母体的子样,就有了与一元正态母体子样平行的统计分析理论。为了便于这 3个抽样分布的应用,本文将给出它们分布函数及数学期望、方差、不确定度区间等数字特征的实用计算公式。

1 分布函数

1.1 i^p 分布的分布函数

由 i^p 分布的密度函数[2]及分布函数的定义知:

$$F(y)=\frac{\lambda^n}{\Gamma(n/p)}\int_0^y t^{n/p-1}e^{-\lambda^p t}\mathrm{d}t \tag{1}$$

式中,

$$\lambda=\left[\Gamma\left(\frac{3}{p}\right)\Big\backslash\Gamma\left(\frac{1}{p}\right)\right]^{1/2} \tag{2}$$

令 $x=\lambda^p t$,则 $t=\lambda^{-p}x$, $\mathrm{d}t=\lambda^{-p}\mathrm{d}x$,且当 $t=0$时,$x=0$;当 $t=y$时,$x=\lambda^p y$,所以,

$$F(y)=\frac{\lambda^n}{\Gamma(n/p)}\int_0^{\lambda^p y}(\lambda^{-p}x)^{n/p-1}(\lambda^{-p})\mathrm{d}x=$$

$$\frac{1}{\Gamma(n/p)}\int_0^{\lambda^p y}x^{n/p-1}e^{-x}\mathrm{d}x=\Gamma(n/p,\lambda^p y) \tag{3}$$

式中,

$$\Gamma(T,U)=\frac{1}{\Gamma(T)}\int_0^U x^{T-1}e^{-x}\mathrm{d}x \tag{4}$$

为不完全伽玛函数。

1.2 t_p 分布的分布函数

计算 t_p 分布的分布函数需要用到不完全贝塔函数:

$$B_x(a,b)=\frac{1}{B(a,b)}\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t \tag{5}$$

式中,$a>0$, $b>0$, $0\leqslant x\leqslant 1$.

下面推导 t_p 分布的分布函数 $F(x)$。由分布函数的性质知,当 $x>0$时,

$$F(x)=1-\int_x^{\infty}f(t)\mathrm{d}t=$$

$$1-k\int_x^{\infty}\left(1+\frac{t^p}{n}\right)^{-(n+1)/p}\mathrm{d}t$$

令 $y=[1+t^p/n]^{-1}$,则 $t=n^{1/p}[1/y-1]^{1/p}$, $\mathrm{d}t=(1/p)n^{1/p}(1/y-1)^{1/p-1}(-y^{-2})\mathrm{d}y$,且当 $t\to\infty$时,$y\to 0$;当 $t=x$时,$y=n/(n+x^p)$,所以,

$$F(x)=1+\frac{kn^{1/p}}{p}\int_{n/(n+x^p)}^0 y^{(n+1)/p}y^{1-1/p-2}\cdot$$

$$(1-y)^{1/p-1}\mathrm{d}y=1-\frac{kn^{1/p}}{p}\cdot$$

$$\int_0^{n/(n+x^p)}y^{n/p-1}(1-y)^{1/p-1}\mathrm{d}y=1-$$

$$\frac{kn^{1/p}}{p}B\left(\frac{n}{p},\frac{1}{p}\right)\frac{1}{B(n/p,1/p)}\cdot$$

$$\int_0^{n/(n+x^p)}y^{n/p-1}(1-y)^{1/p-1}\mathrm{d}y=1-$$

$$(1/2)B_{n/(n+x^p)}(n/p,1/p) \tag{6}$$

由于 t_p 分布对称,所以当 $x<0$时

$$F(x)=1-F(|x|)=\frac{1}{2}B_{n/(n+|x|^p)}\left(\frac{n}{p},\frac{1}{p}\right) \tag{7}$$

在上述推导过程中 k为常数,且

$$k=p\Gamma\left(\frac{n+1}{p}\right)\Big\backslash\left(2n^{1/p}\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{n}{p}\right)\right) \tag{8}$$

1.3 F_p 分布的分布函数

令

$$A=\left(\frac{m}{p}\right)^{m/p}/B\left(\frac{m}{p},\frac{n}{p}\right) \tag{9}$$

收稿日期:1997-07-25.孙海燕,男,38岁,高级工程师,现从事 GPS研究。
* 国家自然科学基金资助项目,编号 49574202

对 F_p 分布的密度函数[2]积分得:

$$F(x)=\int_0^x A z^{m/p-1}\left(1+\frac{m}{n}z\right)^{-\frac{m+n}{p}}dz=$$

$$1-A\int_x^\infty z^{m/p-1}\left(1+\frac{m}{n}z\right)^{-(m+n)/p}dz$$

令 $y=[1+(m/n)z]^{-1}$,则 $z=(n/m)[(1-y)/y]$, $dz=-(n/m)y^{-2}dy$;当 $z\to\infty$ 时, $y\to 0$,当 $z=x$ 时, $y=n/(n+mx)$,于是

$$F(x)=1-A\int_{n/(n+mx)}^0\left(\frac{n}{m}\frac{1-y}{y}\right)^{m/p-1}\cdot$$

$$y^{(m+n)/p}\left(-\frac{n}{m}y^{-2}\right)dy=$$

$$1-\frac{(m/n)^{m/p}(m/n)^{-m/p}}{B(m/p,n/p)}\int_0^{n/(n+mx)}y^{n/p-1}\cdot$$

$$(1-y)^{m/p-1}dy=1-B_{n/(n+mx)}(n/p,m/p) \tag{10}$$

2 数学期望与方差

2.1 χ^p 分布的数学期望与方差

由数学期望及方差的定义知:

$$E(y)=\int_0^\infty y\chi^p(y)dy=\frac{\lambda^n}{\Gamma(n/p)}\cdot$$

$$\int_0^\infty y^{n/p}e^{-\lambda^p y}dy=\frac{\lambda^{-p}}{\Gamma(n/p)}\int_0^\infty u^{n/p}e^{-u}du=$$

$$\frac{\Gamma(n/p+1)}{\lambda^p\Gamma(n/p)}=\frac{n}{p\lambda^p} \tag{11}$$

$$D(y)=\int_0^\infty[y-E(y)]^2\chi^p(y)dy=$$

$$\int_0^\infty\left[y^2-\frac{2ny}{p\lambda^p}+\left(\frac{n}{p\lambda^p}\right)^2\right]\chi^p(y)dy=$$

$$[\Gamma(n/p+2)]/[\lambda^{2p}\Gamma(n/p)-$$

$$2n^2/(p^2\lambda^{2p})=n/(p\lambda^{2p}) \tag{12}$$

令 $p=2$,则得 χ^2 分布的数学期望与方差:

$$E(y)=n/(2\lambda^2)=n$$

$$D(y)=n/(2\lambda^4)=2n$$

2.2 t_p 分布的数学期望与方差

由于 $f(t)=f(-t)$,故

$$E(t)=0 \tag{13}$$

由方差的定义知:

$$D(t)=\int_{-\infty}^\infty t^2f(t)dt=2\int_0^\infty t^2f(t)dt=$$

$$\frac{p\Gamma[(n+1)/p]}{n^{1/p}\Gamma(1/p)\Gamma(n/p)}\int_0^\infty t^2\left(1+\frac{t^p}{n}\right)^{-(n+1)/p}dt \tag{14}$$

由公式

$$\int_0^\infty\frac{x^{m-1}}{(1+bx^a)^{m+n}}dx=$$

$$a^{-1}b^{-m/a}B(m/a,m+n-m/a) \tag{15}$$

得:

$$I=\int_0^\infty t^2\left(1+\frac{t^p}{n}\right)^{-(n+1)/p}dt=\frac{n^{3/p}}{p}\cdot$$

$$B[3/p,(n-2)/p]=n^{3/p}\Gamma(3/p)\cdot$$

$$\Gamma[(n-2)/p]/\{p\Gamma[(n+1)/p]\} \tag{16}$$

把(16)式代入(14)式得:

$$D(t)=\frac{n^{2/p}\Gamma(3/p)\Gamma[(n-2)/p]}{\Gamma(1/p)\Gamma(n/p)} \tag{17}$$

令 $p=2$,代入(17)式,则得 t 分布的方差:

$$D(t)=[n\cdot(1/2)\Gamma(1/2)\Gamma(n/2-1)]/$$

$$[\Gamma(1/2)(n/2-1)\Gamma(n/2-1)]=n/(n-2)$$

2.3 F_p 分布的数学期望与方差

以 $f(z)$ 表示 F_p 分布的密度函数[2],则

$$E(z)=\int_0^\infty zf(z)dz=\frac{(m/n)^{m/p}}{B(m/p,n/p)}\cdot$$

$$\int_0^\infty z^{m/p}(1+mz/n)^{-(m+n)/p}dz=\frac{(m/n)^{m/p}}{B(m/p,n/p)}\cdot$$

$$\left(\frac{m}{n}\right)^{-m/p-1}B\left(\frac{m}{p}+1,\frac{n}{p}-1\right)=\frac{n}{n-p} \tag{18}$$

$$D(z)=\int_0^\infty[z-E(z)]^2f(z)dz=$$

$$\frac{(m/n)^{m/p}}{B(m/p,n/p)}\int_0^\infty\left(z-\frac{n}{n-p}\right)^2\cdot$$

$$z^{m/p-1}\left(1+\frac{m}{n}z\right)^{-(m+n)/p}dz=\frac{(m/n)^{m/p}}{B(m/p,n/p)}\cdot$$

$$\int_0^\infty z^{m/p+1}\left(1+\frac{m}{n}z\right)^{-(m+n)/p}dz-\frac{2n^2}{(n-p)^2}+$$

$$\frac{n^2}{(n-p)^2}=\frac{n^2}{m}\frac{m+p}{(n-p)(n-2p)}-$$

$$\frac{n^2}{(n-p)^2}=\frac{n^2p(m+n-p)}{m(n-p)^2(n-2p)} \tag{19}$$

由式(18)、(19)知,只有当 $n>p$ 及 $n>2p$ 时, F_p 分布才存在数学期望和方差。

3 熵与不确定度区间

根据文献[3],随机变量的不确定度区间是该随机变量出现的基本区间。这一区间既可作为其分布的离散性的度量,也可作为对服从该分布的统计量进行假设检验的置信区间。不确定度区间的定义为:

$$d=e^{H(y)} \tag{20}$$

式中, $H(y)$为分布的熵,由下式计算:

$$H(y)=-\int_{-\infty}^\infty f(y)\ln f(y)dy \tag{21}$$

下面我们首先计算 χ^p 分布的熵。由(21)式知:

$$H(y)=-\int_0^\infty\chi^p(y)\ln\chi^p(y)dy=$$

$$\ln\left[\frac{\Gamma(n/p)}{\lambda^n}\right]-\left(\frac{n}{p}-1\right)\int_0^{\infty}\chi^p(y)\ln y\,\mathrm{d}y+\frac{n}{p} \tag{22}$$

令

$$I=\int_0^{\infty}\chi^p(y)\ln y\,\mathrm{d}y \tag{23}$$

并考虑到

$$\ln y=\lim_{h\to\infty}h\left(y^{\frac{1}{h}}-1\right) \tag{24}$$

可得:

$$I=\frac{\lambda^n}{\Gamma(n/p)}\int_0^{\infty}y^{n/p-1}\mathrm{e}^{-\lambda^p y}\ln y\,\mathrm{d}y=$$
$$\frac{\lambda^n}{\Gamma(n/p)}\lim_{h\to\infty}h\int_0^{\infty}y^{n/p-1}\mathrm{e}^{-\lambda^p y}\left(y^{1/h}-1\right)\mathrm{d}y=$$
$$\frac{\lambda^n}{\Gamma(n/p)}\lim_{h\to\infty}h\left[(\lambda^{-p})^{n/p+1/h}\int_0^{\infty}x^{n/p+1/h-1}\mathrm{e}^{-x}\mathrm{d}x-(\lambda^{-p})^{n/p}\int_0^{\infty}x^{n/p-1}\mathrm{e}^{-x}\mathrm{d}x\right]=\frac{1}{\Gamma(n/p)}\cdot$$
$$\lim_{h\to\infty}h\left[\lambda^{-p/h}\Gamma(n/p+1/h)-\Gamma(n/p)\right]=$$
$$\frac{1}{\Gamma(n/p)}\lim_{h\to\infty}\left\{\frac{(\lambda^{-p})^{1/h}\Gamma\left(\frac{n}{p}+\frac{1}{h}\right)-\Gamma\left(\frac{n}{p}\right)}{\frac{1}{h}}+h\left[(\lambda^{-p})^{1/h}-1\right]\Gamma\left(\frac{n}{p}\right)\right\}=$$
$$\frac{1}{\Gamma(n/p)}\left[\Gamma'\left(\frac{n}{p}\right)+\Gamma\left(\frac{n}{p}\right)\ln(\lambda^{-p})\right]=$$
$$\psi(n/p)-p\ln\lambda \tag{25}$$

式中,$\psi(x)=\Gamma'(x)/\Gamma(x)$ 为普西函数。将 (25) 式代入 (22) 式可得:

$$H(y)=\ln\left[\frac{\Gamma(n/p)}{\lambda^n}\right]+\frac{n}{p}-\left(\frac{n}{p}-1\right)\cdot\left[\psi\left(\frac{n}{p}\right)-p\ln\lambda\right]=\ln\left[\frac{\Gamma(n/p)}{\lambda^n}\cdot\mathrm{e}^{n/p-(n/p-1)[\psi(n/p)-p\ln\lambda]}\right] \tag{26}$$

由 (20) 式知 χ^p 分布的不确定度区间为:

$$d=\Gamma(n/p)\lambda^n\mathrm{e}^{\{n/p-(n/p-1)[\psi(n/p)-p\ln\lambda]\}} \tag{27}$$

限于篇幅,下面我们直接给出 t_p、F_p 变量的熵与不确定度区间的公式。其实在这些公式的推导过程中,除了运用 (24)式将被积函数化为一个收敛的函数列,待积分完成后再用

$$\lim_{h\to\infty}\frac{f(x+1/h)-f(x)}{1/h}=f'(x) \tag{28}$$

将求极限符号去掉这样两个技巧之外,并无什么特别之处。

t_p 变量的熵与不确定度区间为:

$$H(t)=-\ln k-\frac{n+1}{p}\left[\psi\left(\frac{n}{p}\right)-\psi\left(\frac{n+1}{p}\right)\right]=\ln\left[(1/k)\mathrm{e}^{[(n+1)/p]\{\psi[(n+1)/p]-\psi(n/p)\}}\right] \tag{29}$$

$$d=\mathrm{e}^{H(t)}=\frac{1}{k}\mathrm{e}^{\frac{n+1}{p}\left[\psi\left(\frac{n+1}{p}\right)-\psi\left(\frac{n}{p}\right)\right]} \tag{30}$$

F_p 变量的熵与不确定度区间为:

$$H(z)=\ln\left[\left(\frac{n}{m}\right)^{m/p}B\left(\frac{m}{p},\frac{n}{p}\right)\mathrm{e}^{g(m,n,p)}\right] \tag{31}$$

$$d=\left(\frac{n}{m}\right)^{m/p}B\left(\frac{m}{p},\frac{n}{p}\right)\mathrm{e}^{g(m,n,p)} \tag{32}$$

在 (31),(32) 两式中

$$g(m,n,p)=-\left(\frac{m}{p}-1\right)\ln\left(\frac{n}{m}\right)-\left(\frac{n}{p}+1\right)\psi\left(\frac{n}{p}\right)-\left(\frac{m}{p}-1\right)\psi\left(\frac{m}{p}\right)+[(m+n)/p]\psi[(m+n)/p] \tag{33}$$

4　应用中的若干问题

如何用 χ^p、t_p 及 F_p 这 3个抽样分布对 p-范母体的子样进行统计分析,在方法上与用 χ^2、t、F 分布分析正态母体的子样是完全相同的。关于随机变量的熵及不确定度的区间的含义及应用,在文献 [3]中已作过比较详细的介绍,对于这些问题本文就不再讨论。在这些抽样分布的应用中,主要的困难在于分布函数及不确定度区间的计算,也就是几个特殊函数的计算。本文将给出它们的实用计算公式。用这些公式,可以很方便地编程计算这些分布的分布函数值及不确定度区间。

4.1　不完全伽玛函数的计算

不完全伽玛函数的定义为:

$$\Gamma(\alpha,\beta)=\frac{1}{\Gamma(\alpha)}\int_0^{\beta}x^{(\alpha-1)}\mathrm{e}^{-x}\mathrm{d}x$$

令

$$P(\alpha,\beta)=\int_0^{\beta}x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x \tag{34}$$

$$Q(\alpha,\beta)=\int_{\beta}^{\infty}x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x \tag{35}$$

则

$$\Gamma(\alpha,\beta)=\frac{P(\alpha,\beta)}{\Gamma(\alpha)}=1-\frac{Q(\alpha,\beta)}{\Gamma(\alpha)} \tag{36}$$

当 $\beta<\alpha+1$ 时,可用级数

$$P(\alpha,\beta)=\mathrm{e}^{-\beta}\beta^{\alpha}\sum_{k=0}^{\infty}\frac{\Gamma(\alpha)}{\Gamma(\alpha+1+k)}\beta^k \tag{37}$$

的前几项计算 $P(\alpha,\beta)$。而当 $\beta\geq\alpha+1$ 时,可用下面的连分式来进行计算:

$$Q(\alpha,\beta)=\mathrm{e}^{-\beta}\beta^{\alpha}\cfrac{1}{x+\cfrac{1-\alpha}{1+\cfrac{1}{x+\cfrac{2-\alpha}{1+\cfrac{2}{x+\cdots+\cfrac{n-\alpha}{1+\cfrac{n}{x+\cdots}}}}}}} \tag{38}$$

4.2 不完全贝塔函数的计算

由式(5)定义的不完全贝塔函数具有如下的对称关系与极限值

$$\begin{aligned} &B_x(a,b) = 1 - B_{1-x}(b,a) \\ &B_0(a,b) = 0,\quad B_1(a,b) = 1 \end{aligned} \tag{39}$$

而且不完全贝塔函数可以用下列连分式表示:

$$\begin{aligned} &B_x(a,b) = \frac{x^a(1-x)^b}{a,B(a,b)}\mathrm{h}(x) \\ &\mathrm{h}(x) = \cfrac{1}{1+\cfrac{d_1}{1+\cfrac{d_2}{1+\cdots}}} \end{aligned} \tag{40}$$

其中,

$$\begin{cases} d_{2k-1} = -\dfrac{(a+k)(a+b+k)x}{(a+2k)(a+2k+1)} \\ d_{2k} = \dfrac{k(b-k)x}{(a+2k-1)(a+2k)} \end{cases}$$

$$k = 1, 2, \cdots$$

当 $x < (a+1)/(a+b+2)$ 时,连分式(40)式收敛很快。而当 $x > (a+1)/(a+b+2)$ 时,可以用对称关系(39)式来计算。

4.3 普西函数的计算

普西函数 $\mathrm{j}(x) = \Gamma'(x)/\Gamma(x)$有如下渐近表达式:

$$\mathrm{j}(x) = \ln(x-1) + 1/[2(x-1)] - 1/[12(x-1)^2] + 1/[120(x-1)^4] - \cdots \tag{41}$$

当 x 较大时,(41)式的前 4项就可给出足够的精度。当 x 较小时,可选适当的正整数 k,由(41)式计算 $\mathrm{j}(x+k)$,然后由下式计算 $\mathrm{j}(x)$:

$$\mathrm{j}(x) = \mathrm{j}(x+k) - \sum_{i=0}^{k-1}\frac{1}{x+i} \tag{42}$$

参考文献

1 孙海燕,张方仁. p-范分布与 p-范最小平差. 武汉测绘科技大学学报, 1991, 16(4): 38～55

2 孙海燕,於宗俦. p-范分布母体的抽样分布. 武汉测绘科技大学学报, 1998, 23(2): 118～120

3 孙海燕. 熵与不确定度区间. 武汉测绘科技大学学报, 1994, 19(1): 67～70

The Digital Characters of the p-norm sample Distribution

Sun Haiyan　Yu Zongchou

(Qingdao Geotechnical Investigating and Surveying Research Institute, 2 Renmin Road, Qingdao, China, 266033)

Abstract　This paper gives the calculating formulas of the distribution function and some digital characters of the p-norm sample distributions— the i^p、t_p and F_p distribution.

Key words　i^p distribution; t_p distribution; F_p distribution; entropy; uncertainty interval