Journal of Computer Applications ISSN 1001-9081 2014-12-15
计算机应用,2014,34(S2):83-84,90 CODEN JYIIDU http://www.joca.cn

文章编号:1001-9081(2014)S2-0083-02

累积正态分布函数的逼近函数综述

王晶晶,杨正瓴*
(天津大学 电气与自动化工程学院,天津 300072)
(* 通信作者电子邮箱 zlyang@tju.edu.cn)

摘 要: Fisher z 变换是一个显式的初等函数,用来逼近累积正态分布函数(标准正态分布的累积分布函数)。介绍了累积正态分布函数逼近函数的评价标准,对有代表性的逼近函数表达式及相应的最大距离误差值进行归纳总结。

关键词: 正态分布;累积分布函数;逼近函数;误差;Fisher z 变换
中图分类号: TP391.8 **文献标志码:** A

Review of approximating functions to cumulative normal distribution function

WANG Jingjing, YANG Zhengling*
(*School of Electrical Engineering and Automation, Tianjin University, Tianjin* 300072, *China*)

Abstract: Fisher z transformation is an explicit elementary function, which can be used to approximate the cumulative normal distribution function (cumulative distribution function of the standard normal distribution). The evaluation criteria of approximating the cumulative normal distribution function were introduced. The expressions of some typical approximating functions and the absolute maximal distance error between cumulative normal distribution function and the approximating functions were summarized.

Key words: normal distribution; cumulative distribution function; approximating function; error; Fisher z transformation

0 引言

标准正态分布的累积分布函数(又称为累积正态分布函数)不是显式的初等函数,其值可以通过函数逼近等方法近似得到。数理统计学的专著与教材中通常以表格的形式给出累积分布概率的数值。同时,在当前大数据研究中,虽然数值计算结果能够相当精确地找到复杂系统中不同变量间的因果关系,但是人们还是需要通过显式的函数关系去理解复杂的系统,即增强数值结果的可解释性。为此,兼顾精确性和简洁性的显式初等函数,一直是逼近标准正态分布的累积分布函数的研究方向之一。本文将对 1969 年 Cody[1] 提出的有理 Chebyshev 逼近之后的相关研究进行扼要回顾。

在概率论和数理统计学[2] 中,标准正态分布的概率密度函数的表达式如下:

$$f(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2) \tag{1}$$

其中实数 $x \in (-\infty, +\infty)$。

标准正态分布的累积分布函数如下:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp(-t^2/2)\,\mathrm{d}t = \frac{1}{2}[1 + erf(x/\sqrt{2})] \tag{2}$$

其中 $x, t \in (-\infty, +\infty)$ 都是实数,$\Phi(x) \in [0,1]$。其中 erf 是 Gauss 误差函数:

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_{0}^{x}\exp(-t^2)\,\mathrm{d}t \tag{3}$$

实数 $erf(x) \in [-1, +1]$。

1 评价标准

记标准正态分布的累积分布函数的逼近函数为 $Q(x)$。评价该函数的标准有两个:第一个标准是 $\max|Q(x) - \Phi(x)|$,即某逼近函数与标准正态分布的累积分布函数两者之间最大距离误差绝对值;误差值越小,逼近效果就越好。另一个标准是逼近函数的简洁性,越是简洁的逼近函数,其数值计算的速度越快,这个指标在实时系统中是十分重要的。

2 研究现状

标准正态分布是概率论和数理统计学中重要的分布,中心极限定理表明它是几乎所有独立同分布随机变量的大样本容量时的极限分布。研究人员自 20 世纪初就开始寻找初等函数去逼近标准正态分布的累积分布函数。经过不断发展和完善,他们已经找到了很多逼近效果比较好的初等逼近函数,下面介绍一些有代表性的研究者及其所取得的成果。

1915 年,Fisher 提出了一个显式初等函数——Fisher z 变换[3],如下:

$$z_r = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = \operatorname{arctanh}(r) \tag{4}$$

其中 $z_r \in (-\infty, +\infty)$,$r \in [-1, +1]$ 都是实数。式(4)经过反变换之后得到如下等式,可以被用来近似式(2),即:

$$Q(x) = \frac{1}{2}\left[1 + \frac{\exp(2x) - 1}{\exp(2x) + 1}\right] \tag{5}$$

当 $x \approx \pm 0.731\,693\,636\,946$ 时,Fisher z 变换逼近标准正态分布累积分布函数 $\Phi(x)$ 的绝对误差值

收稿日期:2014-01-20。 基金项目:天津市应用基础及前沿技术研究计划项目(09JCYBJC07700)。
作者简介:王晶晶(1988-),女,河南驻马店人,硕士研究生,主要研究方向:智能交通; 杨正瓴(1964-),男,河北灵寿人,副教授,博士,主要研究方向:复杂时间序列预测、智能交通。

$\max|Q(x)-\Phi(x)|$是0.044 227 990 450 3。Fisher z变换被当前国内外教材普遍使用[4-7]。

1969年，Cody[1]提出了有理Chebyshev逼近，这是一个相当复杂的逼近函数。该逼近函数与标准正态分布的累积分布函数的最大距离误差$\max|Q(x)-\Phi(x)|$小于6×10^{-19}。该函数被Matlab用作计算$\Phi(x)$的数值方法。

1977年，Page[8]提出的初等显式逼近函数形式为：

$$Q(x)=1-\frac{1}{1+\exp(\alpha_1x^3+\alpha_2x)} \tag{6}$$

其中$\alpha_1=0.070\,565\,992$，$\alpha_2=1.597\,6$。该逼近函数的绝对误差$\max|Q(x)-\Phi(x)|$小于1.4×10^{-4}。

1978年，Hamaker[9]提出了如下形式的逼近函数：

$$Q(x)=\frac{1}{2}\left[1+\{1-\exp[-(0.806x(1-0.018x))^2]\}^{1/2}\right] \tag{7}$$

当$0\leqslant x\leqslant4$时，其绝对误差$\max|Q(x)-\Phi(x)|$小于0.005。

1990年，Lin[10]提出逼近函数：

$$Q(x)=1-\frac{1}{1+\exp(4.2\pi x/(9-x))} \tag{8}$$

当$0\leqslant x<9$时，其绝对误差$\max|Q(x)-\Phi(x)|$小于6.8×10^{-3}。

1996年，Waissi和Rossin[11]提出了逼近函数：

$$Q(x)=\frac{1}{1+\exp(-\sqrt{\pi}(\beta_1x^5+\beta_2x^3+\beta_3x))} \tag{9}$$

其中：$\beta_1=-0.000\,440\,6$，$\beta_2=0.041\,819\,8$，$\beta_3=0.900\,000\,0$。当$-8<x<+8$时，其绝对误差$\max|Q(x)-\Phi(x)|$小于4.31×10^{-5}。

2002年，Bryc[12]提出了一致逼近函数：

$$Q(x)=1-\frac{x+3.333}{\sqrt{2\pi}x^2+7.32x+2\times3.333}\exp(-x^2/2);\quad x\geqslant0 \tag{10}$$

其绝对误差$\max|Q(x)-\Phi(x)|$小于7.1×10^{-4}。

进一步，一个更精确的逼近函数如下：

$$Q(x)=1-(x^2+5.575\,192\,695x+12.774\,363\,24)\exp(-x^2/2)/(\sqrt{2\pi}x^3+14.387\,181\,47x^2+31.535\,319\,77x+2\times12.774\,363\,24) \tag{11}$$

当$x>0$时，其绝对误差$\max|Q(x)-\Phi(x)|$小于1.9×10^{-5}。

2008年，Aludaat和Alodat[13]找到了如下形式的逼近函数：

$$Q(x)=\frac{1}{2}\left[1+\{1-\exp(-\sqrt{\pi/8}x^2)\}^{1/2}\right];\quad x\geqslant0 \tag{12}$$

其绝对误差$\max|Q(x)-\Phi(x)|$小于0.001 973 23。

2009年，Yun[14]提出了基于双曲正切函数的逼近累积正态分布函数的函数形式：

$$Q(x)=\frac{1}{2}\left\{1+\tanh\left[\frac{r}{2j}\left(\frac{1}{(1-x/a)^j}-\frac{1}{(1+x/a)^j}\right)\right]\right\};\quad 0\leqslant x\leqslant a \tag{13}$$

其中$a=\sqrt{\pi/2}r$。当$j=1$，$r=4.04$，$a=5.07\,59$时，其绝对误差$\max|Q(x)-\Phi(x)|$小于等于1.8×10^{-3}；当$j=2(4,6,8,10)$时，其绝对误差$\max|Q(x)-\Phi(x)|$小于等于8.9×10^{-4}；其中当$j=10$，$r=18.2$，$a=22.810$时，Yun的该逼近函数的绝对误差$\max|Q(x)-\Phi(x)|$最小。

2013年，杨正瓴等[15]首先提出了二次根式函数形式的逼近函数如下：

$$Q(x)=\frac{1}{2}\left(1+\frac{x}{\sqrt{k+x^2}}\right) \tag{14}$$

并把上述函数和Fisher z变换结合起来如下：

$$\Phi(x)\approx\begin{cases}\dfrac{1}{2}\left(1+\dfrac{x}{\sqrt{k+x^2}}\right), & -1.519<x<1.519\\[2ex] \dfrac{1}{2}\left(1+\dfrac{\exp(2x)-1}{\exp(2x)+1}\right), & \text{其他}\end{cases} \tag{15}$$

当$k=1.010\,019\,038\,949\,07$时，其绝对误差$\max|Q(x)-\Phi(x)|$小于0.031 303 884 628 9，是著名的Fisher z变换的误差的70.7%。

同年，杨正瓴等[16]又发现了Sigmoid-like形式的逼近函数，如下：

$$Q(x)=\frac{1}{1+\exp(-kx)} \tag{16}$$

当$k=1.701\,744\,541\,09$时，其绝对误差$\max|Q(x)-\Phi(x)|$小于0.009 457 228 328 68，是著名的Fisher z变换的误差的21.4%。

3 结语

上述标准正态分布累积分布函数$\Phi(x)$的逼近函数具有如下普遍特征：一类以1969年Cody的有理Chebyshev逼近为代表，逼近误差很小，但函数形式比较复杂，它们的计算速度较慢；另一类以2013年杨正瓴等的二次根式函数为代表，在满足一定精度的条件下，逼近函数形式简单，计算速度很快。二次根式逼近函数的计算速度是Cody有理Chebyshev逼近的20余倍。

能够在整个实数范围内直接使用的逼近函数，有Fisher z变换、1977年的Page逼近，以及2013年杨正瓴等提出的2个逼近函数。有些逼近函数只在特定的自变量区间范围内能够达到某一精度范围，比如Hamaker和Lin等提出的函数，这些逼近函数在实际使用中不太方便。

逼近函数的简洁性(对应快的数值计算速度)和逼近误差(精确性)是相互矛盾的，而在实际的实时系统如控制系统[17]中，计算速度往往比计算结果的精确性更加重要。由于2009年Yun的逼近函数，以及2013年杨正瓴等提出的2个逼近函数是兼顾简洁性和精确性的较好逼近函数，因此它们都是替换目前国内外教材中使用的Fisher z变换的主要候选逼近函数。

参考文献：

[1] CODY W J. Rational Chebyshev approximations for the error function [J]. Mathematics of Computation, 1969, 23: 631-637.

[2] 史道济，张玉环. 应用数理统计[M]. 天津：天津大学出版社，2008: 242-257.

[3] FISHER R A. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population [J]. Biometrika, 1915, 10(4): 507-521.

[4] CHATTERJEE S K. Statistical thought: a perspective and history [M]. New York: Oxford University Press, 2003.

[5] RILEY K F, HOBSON M P, BENCE S J. Mathematical methods for physics and engineering: a comprehensive guide [M]. Cambridg: Cambridge University Press, 2002.

(下转第90页)

对于发送方,响应方和入侵者模型,在每次消息接收发送时,也要对 KDC 的消息收发方式的改变作出调整;且由于多了一个域,在入侵者进程中,入侵者拦截伪装消息的方式也要更改。

5 结语

本文介绍了通过 Spin 进行安全协议形式化验证的一般方法,具体提出了一种改进的双方密钥分配中心(KDC)通信协议 Promela 语义模型,然后用 Spin 检测工具检测,发现原协议不满足用 LTL 公式表示的安全性,得到了协议中的一种可被入侵者利用的攻击序列。针对该攻击漏洞,给出了一种可行的改进方案,有效修补该漏洞。并且给出了新协议的 Promela 语义模型。比起原方法,验证效率有了较大提高。这种简化入侵者攻击进程模型的方法思路,对于其他一些安全协议的基于 Spin 的形式化验证也有一定的参考作用。

基于 Spin 的形式化验证方法,需要自己熟悉使用 Spin 模型检测工具,有很大的操作难度,下一步的工作可以研发一种操作更简单、界面更直观的软件系统[16]。

参考文献:

[1] 范红,冯登国. 安全协议理论与方法[M]. 北京: 科学出版社, 2003.

[2] 卿斯汉. 安全协议[M]. 北京: 清华大学出版社, 2005.

[3] NEEDHAM R, SCHROEDER M. Using encryption for authentication in large networks of computers[J]. Communications of the ACM, 1978,21(12):993 - 999.

[4] LOWE G. An attack on the Needham-Schroeder public-key authentication protocol [J]. Information Processing Letters, 1995, 56:131 - 133.

[5] BURROWS M, ABADI M, NEEDHAM R M. A logic of authentication[J]. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 1989, 426(1871): 233 - 271.

[6] GONG L, NEEDHAM R, YAHALOM R. Reasoning about belief in cryptographic protocols[C]// Proceedings of the 1990 IEEE Computer Society Symposium on Research in Security and Privacy. Piscataway: IEEE, 1990: 234 - 248.

[7] MAO W, BOYD C. Towards formal analysis of security protocols [C]// Proceedings of the 1993 IEEE Computer Security Foundations Workshop VI. Piscataway: IEEE, 1993: 147 - 158.

[8] ABADI M, TUTTLE M R. A semantics for a logic of authentication [C]// Proceedings of the Tenth Annual ACM Symposium on Principles of Distributed Computing. New York: ACM, 1991: 201 - 216.

[9] van OORSCHOT P. Extending cryptographic logics of belief to key agreement protocols[C]// Proceedings of the 1st ACM Conference on Computer and Communications Security. New York: ACM, 1993: 232 - 243.

[10] SYVERSON P F, van OORSCHOT P C. On unifying some cryptographic protocol logics[C]// Proceedings of the 1994 IEEE Computer Society Symposium on Research in Security and Privacy. Piscataway: IEEE, 1994: 14 - 28.

[11] CHEN H, CLARK J A, JOCOB J L. A search-based approach to the automated design of security protocols, YCS 376[R]. New York: University of York, Department of Computer Science, 2004.

[12] LOWE G. Towards a completeness result for model checking of security protocols[J]. Journal of Computer Security, 1999, 7(2): 89 - 146.

[13] ROSCOE A W, GOLDSMITH M H. The perfect 'spy' for model-checking cryptoprotocols[C]// Proceedings of DIMACS Workshop on the Design and Formal Verification of Cryptographic Protocols. Piscataway: IEEE, 1997: 574.

[14] SCHNEIDER S. Security properties and CSP[C]// Proceedings of the 1996 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 1996: 174 - 187.

[15] LOWE G. Breaking and fixing the Needham-Schroeder public-key protocol using FDR[C]// Tools and Algorithms for the Construction and Analysis of Systems. Heidelberg: Springer Berlin, 1996: 147 - 166.

[16] YANG H, WU J. Formal verification of RGPS-S[C]// Proceedings of the 2011 International Conference on Business Computing and Global Informatization. Washington, DC: IEEE Computer Society, 2011: 599 - 602.

[17] GNESI S, LENZINI G, LATELLA D, *et al*. An automatic SPIN validation of a safety critical railway control system[C]// DSN 2000: Proceedings of the 2000 IEEE International Conference on Dependable Systems and Networks. Piscataway: IEEE, 2000: 119 - 124.

[18] 王巧丽. SPIN 模型检测的研究与应用[D]. 贵阳: 贵州大学, 2006.

(上接第84页)

[6] MORRISON D F. Multivariate statistical methods [M]. Australia: Thomson Brooks/Cole, 2005.

[7] OLVER F W J, LOZIER D M, BOISVERT R F, *et al*. NIST handbook of mathematical functions [M]. Cambridge: Cambridge University Press, 2010.

[8] PAGE E. Approximation to the cumulative normal function and its inverse for use on a pocket calculator [J]. Journal of Applied Statistics, 1977, 26: 75 - 76.

[9] HAMAKER H C. Approximating the cumulative normal distribution and its inverse [J]. Journal of Applied Statistics, 1978, 27: 76 - 77.

[10] LIN J T. A simpler logistic approximation to the normal tail probability and its inverse [J]. Journal of the Royal Statistical Society: Series C: Applied Statistics, 1990, 39(2): 255 - 257.

[11] WAISSI G R, ROSSIN D F. A sigmoid approximation to the standard normal integral [J]. Applied Mathematics and Computation, 1996, 77: 91 - 95.

[12] BRYC W. A uniform approximation to the right normal tail integral [J]. Applied Mathematics and Computation, 2002, 127: 365 - 374.

[13] ALUDAAT K M, ALODAT M T. A note on approximating the normal distribution function [J]. Applied Mathematical Sciences, 2008, 2(9): 425 - 429.

[14] YUN B I. Approximation to the cumulative normal distribution using hyperbolic tangent based functions [J]. Journal of the Korean Mathematical Society, 2009, 46(6): 1267 - 1276.

[15] YANG Z, DUAN Z, WANG J, *et al*. Quadratic radical function better than Fisher z transformation [J]. Transactions of Tianjin University, 2013, 19(5): 381 - 384.

[16] 宋延文. 基于稳健统计的公路短时交通流组合预测[D]. 天津: 天津大学, 2012.

[17] 高志强, 胡晓勤. 基于抗体浓度的实时网络风险控制系统的设计与实现[J]. 计算机应用, 2013, 33(10): 2842 - 2845.