高斯新分布扩展应用的研究

孔建新
（云南省会泽县老年科技工作者协会，云南 会泽 528403）

摘要：统计特征的多样性使随机变量频数分布还呈现双峰分布和多峰分布的形态。依据范剑青局部描述能降低统计误差的新理念和高斯分布的原理对其描述。本文以人口年龄频数分布的实例来进行拟合，由此对高斯新分布应用进行扩展性的研究。

关键词：高斯新分布；双峰分布；期望值；左期望差；右期望差

中图分类号：O211.3

Gaussian new-distribution the extended and application of study

Kong Jianxin
(Yunnan province Huize county elderly association of science and technology,YunNan HuiZe 528403)

Abstract: This diversity of statistical characteristics of random frequency distribution also showed a bimodal distribution and multi-peak distribution of the form.Partial description of the basis Jianqing Fan reduce statistical error of the new concepts and principles of its description of the Gaussian distribution.In this paper, the frequency distribution of the population age fitting examples.Gaussian new-distribution of this application scalability study.

Keywords: Gaussiannew-distribution;bimodaldistribution;expectedvalue;left-expecteddeviation;right-expected deviation

0 引言

自然和社会所显示随机变量的统计对象的不外乎包括单侧规范、双侧规范、多侧规范、自然规范四大类。前三类也称为单侧控制、双侧控制、多侧控制。自然规范是不同于前三类的一个新概念，指的是自然形成的随机变量的统计对象，一般难以进行人为的控制。如：气温的变化分布、降雨量分布等一些自然现象。以及像人口年龄频数分布等等。属于自然规范统计范畴随机变量的频数可能服从正态分布或偏斜分布或双峰分布或多峰分布或其它类型的分布，虽然其分布看似无规律可言的，但是针对某项自然规范的具体统计对象却是有规律可循的。

“自然界的事物基本上都很简单，所有的基础原理及主要问题都可以用数学方式表达，这是应用数学家的一个信仰。”这句话出自应用数学大师林家翘的一次演讲。他归国八年以来绝少面对媒体，但几乎每一次出现在镜头前他都会强调这一点——应用数学的意义在于揭示自然界和社会实际问题的规律。[1]本文依据范剑青局部描述能降低统计误差的新理念[2]和高斯分布的原理来推出双峰分布及多峰分布的数学模型。并以人口年龄频数分布为例展开讨论，从应用的角度出发揭示人口年龄频数的分布规律。目的是：为将进行的人口普查建立年龄频数分布的数学模型并由此说明高斯新分布扩展的现实意义和应用前景。

1 双峰分布和多峰分布存在的客观性

自然界和社会经济中凡是可以用数据表示的统计现象其随机变量的频数分布具有很大

作者简介：孔建新（1950-），男，质量工程师、高级统计师，研究方向：高斯分布. E-mail: kongfanjx@163.com

的随机性，是不可能只用某一种分布形态就可以概括来进行拟合的。正态分布、偏斜分布仅仅属于单峰分布的范畴，统称为：高斯新分布。而双峰分布和多峰分布的统计现象也是自然界客观存在的形态。所以也需要建立数学模型来表达拟合它。

双峰分布在某些领域较普遍。如：军事上枪械射击和炮火攻击目标的弹着点。体育上射击、射箭的中靶点，跳伞运动的着地点。还有，航天卫星回收的着陆点；地质勘探、石油钻井所出现的非垂直偏差分布。不仅如此，在人口普查中年龄的频数分布、一些产品在年内的销售量分布、某地年内每天（最高或最低)气温的变化分布都有可能出现两个或多个峰值。

本文通过《中国统计年鉴》2006 年全国人口变动情况抽样调查样本数据[3]表 1 资料为例说明多峰分布存在的客观必然性。

表 1 2006 年全国人口变动情况抽样调查样本数据

Tab. 1 Changes in the 2006 national population sample survey data

项目	年龄		人口数	比重	人口数（人）		比重（ % ）		性比（%）
序号	组距	组中值	人	%	男	女	男	女	男 / 女#
4#	307#	5#	93889#	813::7##	66454#	5:768#	51:::4##	516336##	4531:6##
5#	80<#	:#	:38;;#	81<4;8##	6;<75#	64979#	615984##	519867##	456139##
6#	4304?#	45#	;<469#	:?7:6:##	7;38:#	743:<#	7135<7##	617776##	4491<<##
7#	4804<#	4:#	438356#	;1;38:##	887;4#	7<875#	71984;##	71486<##	4441<<##
8#	5305?#	55#	:9493#	916;8:##	6:5:4#	6;;;<#	614583##	61593:##	<81;7##
9#	5805<#	5:#	:7443#	91546;##	68;9;#	6;575#	6133:7##	615397##	<61:<##
:#	6306?#	65#	<66<;#	:1;643##	78;4<#	7:8;3#	61;74:##	61<;<7##	<9163##
;#	6806<#	6:#	446<85#	<18877##	89366#	8:<53#	719<;4##	71;896##	<91:7##
<#	7307?#	75#	448:;4#	<1:3::##	8:5:9#	8;838#	71;357##	71<387##	<:1<3##
43#	7807<#	7:#	:97<9#	91746<##	6;57;#	6;57:#	61539<##	61539;##	433133##
44#	8308?#	85#	<393:#	:18<:3##	78976#	77<97#	61;5:3##	61::33##	434184##
45#	8808<#	8:#	9;5::#	81:57:##	6786<#	66:6:#	51;<8<##	51;5;:##	43516;##
46#	9309?#	95#	7;;;9#	713<;<##	57<7:#	56<6<#	513<4:##	5133:5##	437154##
47#	9809<#	9:#	6<<<9#	616868##	53793#	4<869#	41:488##	4196;3##	4371:6##
48#	:30:7#	:5#	659<5#	51:744##	49475#	49883#	416867##	416;:9##	<:186##
49#	:80:<#	::#	53965#	41:5<<##	<;83#	43:;4#	31;58<##	31<36<##	<4169##
4:#	;30;7#	;5#	43;58#	31<3:9##	7937#	9554#	316;93##	318549##	:7134##
4;#	;80;<#	;:#	7475#	3167:6##	4894#	58;4#	31463<##	315497##	9317;##
4<#	<30<7#	<5#	4463#	313<7:##	699#	:97#	31363:##	313974##	7:1<4##
53#	<8≤#	<:#	5:<#	313567##	:8#	537#	313396##	3134:4##	691:9##
Ù #	总计	Ù #	44<5999#	433#	937636#	8;;695#	83199;6##	7<1664:##	4351:4##

根据表 1 人口数作人口年龄频数直方图。见图 1 （频数单位：千人）。

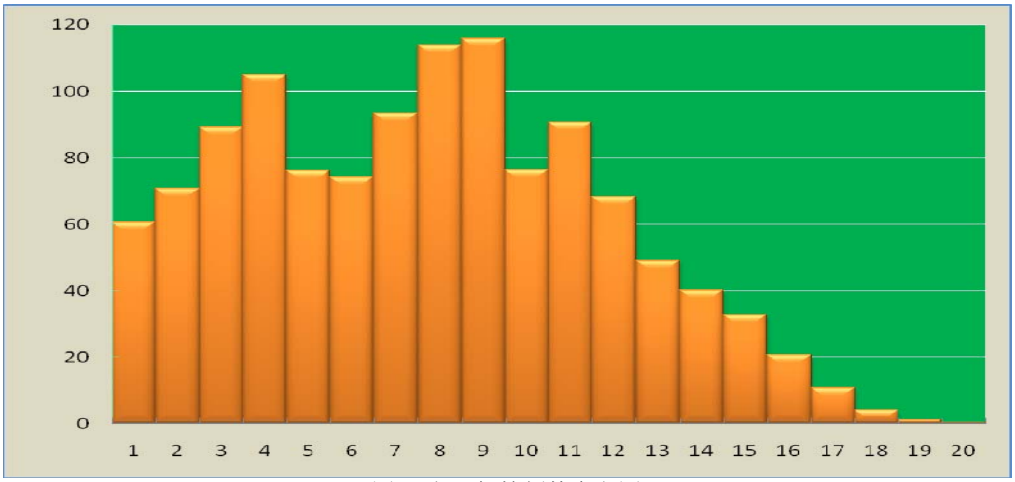

图 1 人口年龄频数直方图

Fig.1 population age frequency histogram

根据表 1 作男性人口年龄频数直方图。见图 2（频数单位：千人）。

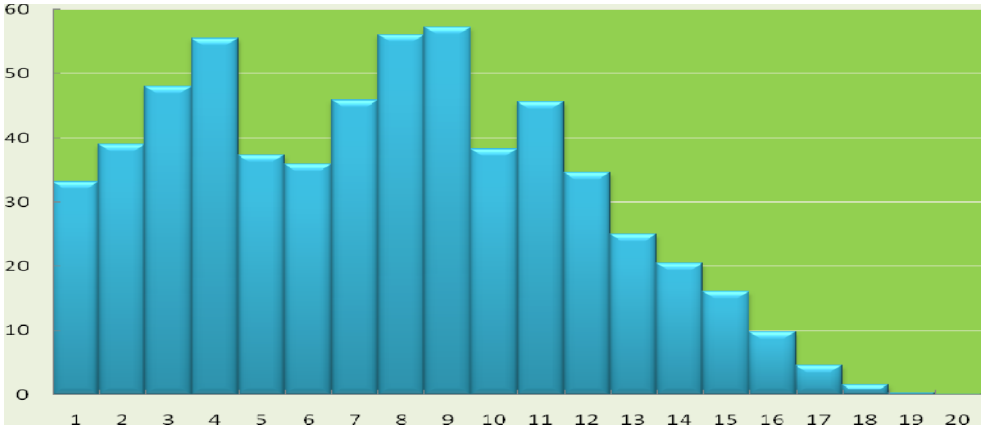

图 2 男性人口年龄频数直方图

Fig.2 Male population age frequency histogram

根据表 1 作女性人口年龄频数直方图。见图 2（频数单位：千人）。

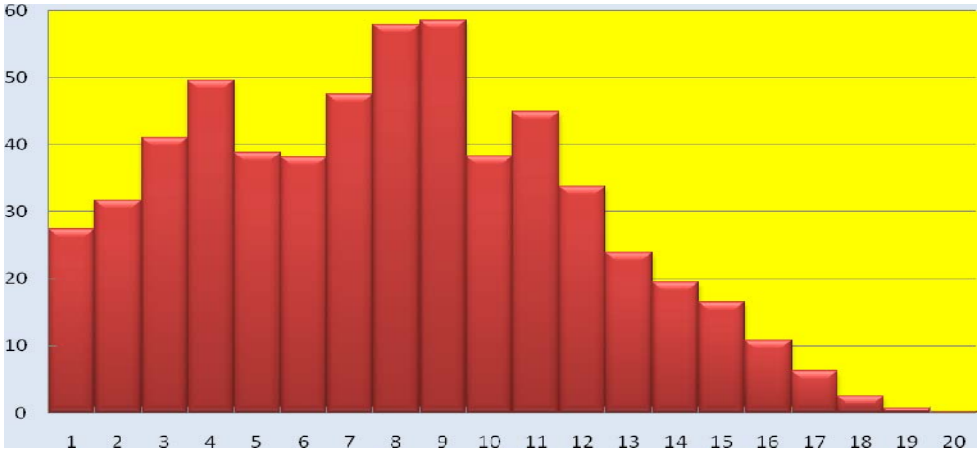

图 3 女性人口年龄频数直方图

Fig.3 Female population age frequency histogram

根据以上表 1 和图中说明：多峰分布是客观存在的统计现象。而多峰分布则是双峰分布的推广，所以有必要建立一个能够拟合描述双峰分布曲线的数学模型，依据的是高斯分布的原理。

高斯是一个伟大的数学家，重要的贡献不胜枚举。其中 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}exp[-\frac{(x-\mu)^2}{2\sigma^2}]$ 模型“由于具有分布密度形式的误差分布对后来的数理统计及其它相关学科的发展影响极大，故此命名为高斯分布或正态分布。德国 10 马克纸币上有一高斯头像且在头像后面印有高斯分布钟型曲线，这表明在高斯的一切科学贡献中,对人类文明影响最大的莫过于这一项。到了 19 世纪可以说成了高斯分布统治的年代，而到了 20 世纪小样本理论建立后，高斯正态分布又显示了它的强大的优越性。[4]”就此特别强调：偏斜分布、高斯新分布、双峰分布、多峰分布的数学模型完全是建立在高斯分布的原理之上，仅仅是对高斯分布在应用中进行的扩展。由此验证了崔恒建教授所述“高斯正态分布又显示了它的强大的优越性。”

2 双峰分布数学模型建立的理论基础及推论

“正态分布是德国数学家高斯（C.F.Gauss，1777～1855）在研究误差理论时最早使用这一分布，所以正态分布又称高斯分布。

若－∞＜μ＜∞，σ ＞0 为两个实数，则由下列密度函数

$$f(x)=\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],\quad -\infty<x<\infty \qquad (1)$$

确定的随机变量 X 的分布称为正态分布，记为 $N(\mu,\sigma^2)$。[5] ”

以高斯分布的数学模型为基础，考虑不对称因素存在的客观性，有必要将参数 σ 分解为左右 σ_-，σ_+，将 σ^2 分解为左右 σ_-^2、σ_+^2，将分解的不对称参数的元素溶入模型所形成的新分布称为：高斯新分布（Gauss new-distribution），定义如下：

若－∞＜μ＜∞，σ_- ＞0，σ_+ ＞0 为三个实数，则由下列密度函数：

$$f(x)=\begin{cases}(\sqrt{2\pi}\,\sigma_-)^{-1}exp\left[-\dfrac{(x-\mu)^2}{2\sigma_-^2}\right] & x\leqslant\mu\\[2ex](\sqrt{2\pi}\,\sigma_+)^{-1}exp\left[-\dfrac{(x-\mu)^2}{2\sigma_+^2}\right] & x\geqslant\mu\end{cases} \qquad (2)$$

由以上密度函数确定的随机变量 X 的分布称为高斯新分布。记为 $N(\mu,\sigma_-,\sigma_+)$。

当 $\sigma_-=\sigma_+$ 时呈正态分布；当 $\sigma_-\neq\sigma_+$ 时呈偏斜分布。

从以上高斯新分布的数学模型清楚表明，当 $\sigma_-=\sigma_+$ 时（2）式就还原为（1）式。由此导出高斯新分布完全是高斯分布参数扩展的结果。

随机变量的频数分布在单峰的条件下由于不对称偏斜分布的客观存在必然使平均值偏移峰值。从高斯分布的性质得出：期望值 μ 一定是在分布峰值点上的取值，从而导出在非对称的情况下期望值在峰值点上的取值使之不一定等于平均值。为此需要对期望值（expected value）重新定义如下：随机变量的频数分布在单峰的条件下是分布曲线最大值点上的取值。满足随机变量 $x=\mu$ 时，$f(\mu)=(\sqrt{2\pi}\,\sigma_-)^{-1}$ 或 $f(\mu)=(\sqrt{2\pi}\,\sigma_+)^{-1}$ 为 $f(x)$ 最大值的条件。

由于期望值两边不一定对称，所以期望值两边随机变量与期望值离散程度的变异指标就不相等，期望差、左期望差、右期望差由此引出。

期望差（expected deviation）：随机变量与期望值频数比率离差平方之和的平方根。也称整体期望差。符号记为：σ。由于期望值两边不一定对称，就存在左期望差和右期望差。其定义的计算方法[6]在以下计算表 2 中充分体现。

左期望差（left-expected deviation）：小于等于期望值的随机变量与期望值频数比率离差平方之和的平方根。符号记为：σ_-。

右期望差（right-expected deviation）：大于等于期望值的随机变量与期望值频数比率离差平方之和的平方根。符号记为：σ_+。

从以上高斯新分布及给出相关术语的新定义进行必要扩展，双峰分布相应的有两个期望值及对应的两组左右期望差。

从高斯新分布的数学模型进行参数的再扩展可以得到描述双峰分布的数学模型。

期望值 1 为 μ_1、期望值 2 为 μ_2。对应的期望差就分别有左期望差 1 记为：σ_{1-}、右期望差 1 记为：σ_{1+}；左期望差 2 记为：σ_{2-}、右期望差 2 记为：σ_{2+}。

上述确定的双峰分布两个期望值和对应的两组左右期望差的参数，还需要设定两峰间的谷值。若统计的对象属于多侧规范的条件，则两峰间的谷值必然是被确定的标准值的中心点零值；若统计的对象属于自然规范的条件，则两峰间的谷值必然是不能被确定点的值；符号记为：v 。得出以下不同规范的两种双峰分布的数学表达式。

在多侧规范的条件下双峰分布的数学表达式：

$$f(x)=\begin{cases}(\sqrt{2\pi}\,\sigma_{1-})^{-1}\ exp[-\dfrac{(x-\mu_1)^2}{2\sigma_{1-}^2}] & x\leqslant\mu_1\\ (\sqrt{2\pi}\,\sigma_{1+})^{-1}\ exp[-\dfrac{(x-\mu_1)^2}{2\sigma_{1+}^2}] & \mu_1\leqslant x\leqslant 0\\ (\sqrt{2\pi}\,\sigma_{2-})^{-1}\ exp[-\dfrac{(x-\mu_2)^2}{2\sigma_{2-}^2}] & 0\leqslant x\leqslant\mu_2\\ (\sqrt{2\pi}\,\sigma_{2+})^{-1}\ exp[-\dfrac{(x-\mu_2)^2}{2\sigma_{2+}^2}] & \mu_2\leqslant x\end{cases} \qquad (3)$$

在自然规范的条件下双峰分布的数学表达式：

$$f(x)=\begin{cases}(\sqrt{2\pi}\,\sigma_{1-})^{-1}\ exp[-\dfrac{(x-\mu_1)^2}{2\sigma_{1-}^2}] & x\leqslant\mu_1\\ (\sqrt{2\pi}\,\sigma_{1+})^{-1}\ exp[-\dfrac{(x-\mu_1)^2}{2\sigma_{1+}^2}] & \mu_1\leqslant x\leqslant v\\ (\sqrt{2\pi}\,\sigma_{2-})^{-1}\ exp[-\dfrac{(x-\mu_2)^2}{2\sigma_{2-}^2}] & v<x\leqslant\mu_2\\ (\sqrt{2\pi}\,\sigma_{2+})^{-1}\ exp[-\dfrac{(x-\mu_2)^2}{2\sigma_{2+}^2}] & \mu_2\leqslant x\end{cases} \qquad (4)$$

以上双峰分布数学表达式推而广之就可以扩展描述实例表 1 资料直方图展示的三峰分布。

3 三峰分布参数的设定及数学模型的建立

根据表 1 资料得出的三个直方图显示，人口年龄频数分布呈三峰形态。依据范剑青教授“局部建模的优点在于可以大大降低误差。[2]”的新理念，以表 1 资料所对应的三个图形的分布来看，由于性别比不是相差太大，所以分布图形也比较相似，选择图 3 为例展开说明。拟合分布曲线需要根据资料或图形进行各峰区的划分及对应参数的设定。

3.1 分布峰区的划分和对应参数的设定及数学模型的建立

以表 1 女性年龄频数分布资料所对应图 3 的直方图来说明。

3.1.1 峰区的划分及左偏分布、右偏分布的定义

从图 3 可以将频数的分组进行划分：

1～5 组划为第一峰区；

6～10 组划为第二峰区；

11～20 组划为第三峰区。

第一峰区与第二峰区以峰值为中心存在两组左右期望差。

第三峰区的分布是一个单减函数，左边的期望差等于零，称为：右偏分布或正偏分布。若右边的期望差等于零，则为单增函数，称为：左偏分布或负偏分布。

由于右偏分布的左期望差等于零，左偏分布的右期望差等于零，则与高斯新分布的定义中“若$-\infty<\mu<\infty$，$\sigma_{-}>0$，$\sigma_{+}>0$为三个实数”及数学模型（2）式矛盾。所以需要对左偏分布和右偏分布的新概念分别定义如下：

左偏分布定义：若$-\infty<\mu<\infty$，$\sigma_{-}>0$且$\sigma_{+}=0$为三个实数，则由下列密度函数：

$$f(x) = (\sqrt{2\pi}\,\sigma_{-})^{-1}\ exp[-\frac{(x-\mu)^2}{2\sigma_{-}^2}],\ x \leqslant \mu \qquad (5)$$

由以上密度函数确定的随机变量X的分布称为左偏分布或负偏分布。记为$N(\mu, \sigma_{-}, 0)$。

由左偏分布定义（5）式确定$\sigma_{+}=0$，则$(\sqrt{2\pi}\,\sigma_{+})^{-1}exp[-\frac{(x-\mu)^2}{2\sigma_{+}^2}]$不存在。

右偏分布定义：若$-\infty<\mu<\infty$，$\sigma_{+}>0$且$\sigma_{-}=0$为三个实数，则由下列密度函数：

$$f(x) = (\sqrt{2\pi}\,\sigma_{+})^{-1}\ exp[-\frac{(x-\mu)^2}{2\sigma_{+}^2}],\ x \geqslant \mu \qquad (6)$$

由以上密度函数确定的随机变量X的分布称为右偏分布或正偏分布。记为$N(\mu, 0, \sigma_{+})$。

由右偏分布定义（6）式确定$\sigma_{-}=0$，则$(\sqrt{2\pi}\,\sigma_{-})^{-1}exp[-\frac{(x-\mu)^2}{2\sigma_{-}^2}]$不存在。

左偏分布描述了一个单增函数，右偏分布描述了一个单减函数。

左偏分布和右偏分布属于偏斜分布的统计范畴，必然也是属于高斯新分布的统计范畴，它们分别是偏斜分布的特例，存在条件的限制。借此对以上高斯新分布的定义进行补充。

高斯新分布（Gauss new-distribution）定义补充如下：

若$-\infty<\mu<\infty$，$\sigma_{-}>0$，$\sigma_{+}>0$为三个实数，则由下列密度函数：

$$f(x) = \begin{cases} (\sqrt{2\pi}\,\sigma_{-})^{-1}exp[-\frac{(x-\mu)^2}{2\sigma_{-}^2}] & x \leqslant \mu \\ (\sqrt{2\pi}\,\sigma_{+})^{-1}exp[-\frac{(x-\mu)^2}{2\sigma_{+}^2}] & x \geqslant \mu \end{cases} \qquad (同\ 2)$$

由以上密度函数确定的随机变量X的分布称为高斯新分布。记为$N(\mu, \sigma_{-}, \sigma_{+})$。

当$\sigma_{-} = \sigma_{+}$时，呈正态分布；当$\sigma_{-} \neq \sigma_{+}$时，呈偏斜分布。

若$\sigma_{-} = 0$时则$(\sqrt{2\pi}\,\sigma_{-})^{-1}exp[-\frac{(x-\mu)^2}{2\sigma_{-}^2}]$不存在，呈右偏分布；

若$\sigma_{+} = 0$时则$(\sqrt{2\pi}\,\sigma_{+})^{-1}exp[-\frac{(x-\mu)^2}{2\sigma_{+}^2}]$不存在，呈左偏分布。

左偏分布和右偏分布新概念提出的理由之一是：其存在具有客观性。之二是为建立多峰分布的数学模型提供理论依据。就是说凡是客观存在的分布形态都需要有数学模型来表达，完全符合应用数学大师林家翘教授“应用数学的意义在于揭示自然界和社会实际问题的规律。”的数学应用的思想。

3.1.2 对应峰区参数的设定

根据以上峰区的划分设定对应的参数。

从左至右三峰区的期望值分别为：μ_1 、μ_2、μ_3 。

第一峰区与第二峰区间的谷值为：v 。

第二峰区与第三峰区间的谷值显然是小于第三峰区峰的期望值 μ_3 。

第一峰区的左右期望差分别为： σ_{1-} 、 σ_{1+} 。

第二峰区的左右期望差分别为： σ_{2-} 、 σ_{2+} 。

第三峰区的左右期望差分别为： σ_{3-}、 σ_{3+} 。

由于 $\sigma_{3-}=0$ 则 $(\sqrt{2\pi}\,\sigma_{3-})^{-1}\ exp[-\frac{(x-\mu)^2}{2\sigma_{3-}^2}]$不存在。

根据以上设定的参数，建立资料表 1 直方图呈现三峰分布曲线的数学模型如下：

$$f(x)=\begin{cases}(\sqrt{2\pi}\,\sigma_{1-})^{-1}\ exp[-\frac{(x-\mu_1)^2}{2\sigma_{1-}^2}] & x\leqslant\mu_1\\ (\sqrt{2\pi}\,\sigma_{1+})^{-1}\ exp[-\frac{(x-\mu_1)^2}{2\sigma_{1+}^2}] & \mu_1\leqslant x\leqslant v\\ (\sqrt{2\pi}\,\sigma_{2-})^{-1}\ exp[-\frac{(x-\mu_2)^2}{2\sigma_{2-}^2}] & v<x\leqslant\mu_2\\ (\sqrt{2\pi}\,\sigma_{2+})^{-1}\ exp[-\frac{(x-\mu_2)^2}{2\sigma_{2+}^2}] & \mu_2\leqslant x<\mu_3\\ (\sqrt{2\pi}\,\sigma_{3+})^{-1}\ exp[-\frac{(x-\mu_3)^2}{2\sigma_{3+}^2}] & \mu_3\leqslant x\end{cases} \qquad (7)$$

以上（7）式拟合描述了表 1 资料对应的图 1、图 2、图 3 的分布曲线，并达到拟合的优度。

4 三峰分布参数的计算及对应的数学表达式

从表 1 资料中已经得出三峰的期望值 μ_1 、μ_2、μ_3 分别为：17、42、52 。离散参数左右期望方差和左右期望差的计算公式和方法在前期的研究中已经作了多方面的详细介绍，在这里不再赘述。选择表 1 图 3 女性人口资料，应用统计表进行计算见表 2 ：

根据以下表 2 计算得出如下结果：

第一峰区

期望值 $\mu_1=17$ ；谷值 $v=24$（根据表 1 第 5 组组距 22～24 的上限确定为 24）。

左期望方差： $\sigma_{1-}^2=82.96$ ，左期望差： $\sigma_{1-}=9.11$ 。

右期望方差： $\sigma_{1+}^2=15.27$，右期望差： $\sigma_{1+}=3.91$ 。

第二峰区

期望值 $\mu_2=42$ 。

左期望方差： $\sigma_{2-}^2=85.61$ ，左期望差： $\sigma_{2-}=9.25$ 。

右期望方差： $\sigma_{2+}^2=14.17$ ，右期望差： $\sigma_{2+}=3.76$ 。

第三峰区

期望值 μ_3=52 。

左期望方差： σ_{3-}^2 =0 ， 左期望差： σ_{3-} =0 。

右期望方差： σ_{3+}^2 =229.45 ； 右期望差： σ_{3+} =15.15 。

表 2 2006 年全国人口女性年龄分布样本资料计算表

Tab. 2 Female population in 2006 sample data computation of the age distribution

分组序号	分布区间	年龄分组组中值	人口频数人	人口频数比率	与期望值离差	离差平方	左比率离差平方	右比率离差平方
4#		5#	5:768#	3154<9##	048##	558#	7<174##	Ù #
5#	第	:#	64979#	315866##	043##	433#	58166##	Ù #
6#		45#	743:<#	3165;;##	08##	58#	;155##	Ù #
7#	一	4:#	57::4#	314<;6##	3##	3#	3133##	Ù #
#		小计	457<64#	413333##	Ù #	Ù #	Ù #	Ù #
7#	峰	4:#	57::4#	316;<4##	3##	3#	Ù #	3133##
8#		55#	6;;;<#	31943<##	8##	58#	Ù #	4815:##
#	区	小计	96993#	413333##	期望方差（合计）		;51k9##	4815:##
#		合计	4;;8<4##	Ù #	期望差		<144##	61k4##
9#		5:#	6;575#	315544##	048##	558##	7<1:7##	Ù #
:#	第	65#	7:8:<#	315:83##	043##	433##	5:183##	Ù #
;#		6:#	8:<4<#	31667;##	08##	58##	;16:##	Ù #
<#	二	75#	5<5851 8#	3149<4##	3##	3##	3133##	Ù #
#		小计	4:5<<518#	413333##	Ù #	Ù #	Ù #	Ù #
<#	峰	75#	5<5851 8#	317667##	3##	3##	Ù #	3133##
43#		7:#	6;57;#	318999##	8##	58##	Ù #	4714:##
#	区	小计	9:8331 8#	413333##	期望方差（合计）		;8194##	4714:##
#		合计	5737<6##	Ù #	期望差		<158##	61:9##
44#		85#	557;5#	413333##	3##	3##	3133#	Ù #
#		小计	557;5#	413333##	Ù #	Ù #	Ù #	Ù #
44#	第	85#	557;5#	314976##	3##	3##	Ù #	3133##
45#		8:#	66:6;#	315799##	8##	58##	Ù #	914:##
46#		95#	56<6<#	314:83##	43##	433##	Ù #	4:183##
47#	三	9:#	4<869#	31475;##	48##	558##	Ù #	65146##
48#		:5#	49883#	314543##	53##	733##	Ù #	7;16<##
49#		::#	43:;5#	313:;;##	58##	958##	Ù #	7<159##
4:#	峰	;5#	9554#	313788##	63##	<33##	Ù #	731k6##
4;#		;:#	58;4#	3134;<##	68##	4558##	Ù #	56144##
4<#		<5#	:97#	313389##	73##	4933##	Ù #	;1k7##
53#	区	<:#	537#	313348##	78##	5358##	Ù #	6135##
Ù #		小计	469:<:#	413333##	期望方差（合计）		3133##	55<178##
#		合计	48<5:<#	Ù #	期望差		3133##	48148##

将以上计算的具体参数结果代入（7）式得出 2006 年全国人口女性年龄分布样本资料的数学模型（8）式如下：

$$f(x)=\begin{cases}(\sqrt{2\pi}\ 9.11)^{-1}\ exp[-\dfrac{(x-17)^2}{2\times 82.96}] & x\leqslant 17\\ (\sqrt{2\pi}\ 3.91)^{-1}\ exp[-\dfrac{(x-17)^2}{2\times 15.27}] & 17\leqslant x\leqslant 24\\ (\sqrt{2\pi}\ 9.25)^{-1}\ exp[-\dfrac{(x-42)^2}{2\times 85.61}] & 24<x\leqslant 42\\ (\sqrt{2\pi}\ 3.76)^{-1}\ exp[-\dfrac{(x-42)^2}{2\times 14.17}] & 42\leqslant x<52\\ (\sqrt{2\pi}\ 15.15)^{-1}\ exp[-\dfrac{(x-52)^2}{2\times 229.45}] & 52\leqslant x\end{cases} \qquad (8)$$

以上（8）式是根据表 2 计算结果代入（7）式的数学表达式。但是在实践中应该按分组组距的具体数字，需要对峰区划分进行调整。从表 1 第 2 峰区与第 3 峰区的谷值是第 10 组，组距为 45～49，第 3 峰值（期望值 μ_3）所在 11 组的组距为 50～54。所以（8）式的区间与第 3 峰区相连的区间划分 $42 \leqslant x < 52$ 应该调整为 $42 \leqslant x < 50; 52 \leqslant x$ 整为 $50 \leqslant x$ 。见以下（9）式：

$$f(x) = \begin{cases} (\sqrt{2\pi}\ 9.11)^{-1}\ exp[-\dfrac{(x-17)^2}{2\times 82.96}] & x \leqslant 17 \\ (\sqrt{2\pi}\ 3.91)^{-1}\ exp[-\dfrac{(x-17)^2}{2\times 15.27}] & 17 \leqslant x \leqslant 24 \\ (\sqrt{2\pi}\ 9.25)^{-1}\ exp[-\dfrac{(x-42)^2}{2\times 85.61}] & 24 < x \leqslant 42 \\ (\sqrt{2\pi}\ 3.76)^{-1}\ exp[-\dfrac{(x-42)^2}{2\times 14.17}] & 42 \leqslant x < 50 \\ (\sqrt{2\pi}\ 15.15)^{-1}\ exp[-\dfrac{(x-52)^2}{2\times 229.45}] & 50 \leqslant x \end{cases} \qquad (9)$$

从调整后的（9）式可以得出结论：多峰分布在频数分组的情况下，连接谷值区间的划分是以所在组距的上下限来确定的。如果谷值连接的是下一峰区的峰值，则此期望值被连接的数据是期望值所在组的下限，而不是组中值，在实际应用中须注意深刻理解。

5 遗留期待探析的问题

在前期多项研究成果的基础上，根据“左右方差计算公式的推导与应用[6]”解决了本文数学模型离散参数的计算问题。有关按统计特征分为单侧规范、双侧规范、多侧规范、自然规范四大类的问题将涉及随机变量的频数在统计实践中呈现单峰分布、双峰分布、多峰分布的客观形态。本文建立双峰分布、多峰分布的数学模型都是依据高斯分布的原理。是否可以归属于高斯新分布的统计范畴，以及多峰分布任意区间概率的计算问题将有待下一课题来探析。

6 结论

依据范剑青教授局部描述能提高统计精确度的新理念建立的双峰分布和多峰分布的数学模型仅仅是对高斯新分布应用的推广。而高斯新分布则是建立在高斯分布的理论基础上。从多峰分布的数学模型（7）式的 5 个数学表达式中清楚显示：它们各是正态分布的一半，由此就可以应用高斯分布来分部描述。这将验证一个重要的结论：应用正态分布来分部描述多峰分布显示了高斯分布又一强大的优越性。

建立多峰分布数学模型的目的是：为进行的第六次人口普查结果的相关资料来揭示人口年龄频数的分布规律。从而促进对高斯新分布的广泛应用，进而推动对高斯新分布的深刻理解。

[参考文献] (References)

[1] 刘文嘉.林家翘:大师之忧[N].光明日报 2010 年 5 月 7 日.第 1 版.
[2] lixing.范剑青:把数学作为解决社会问题的工具 [OL] 经济学论坛-中国经济学教育科研网 2008-2-13
[3] 中国统计年鉴 [OL] http://www.sei.gov.cn/try/hgjj/yearbook/2007/indexCh.htm
[4] 崔恒建,陈秋华.高斯分布的启示 [J] 数学通报 2000.第 4 期 P42 页
[5] 茆诗松.统计手册 [M].北京:科学出版社 2003.1
[6] 孔建新. 左右方差计算公式的推导与应用[OL].中国科技论文在线, 2010.8.24.