

一种用于旱涝分析的 降水概率指标

—Gamma分布概率指标

黄嘉佑

(北京大学地球物理系)

提 要

本文介绍以Gamma分布为基础的降水概率指标以及相应的旱涝级别。对北京地区7月、夏季(6—8月)及年的降水概率指标及早涝级别序列分析表明,它在描述气候状态上较传统的旱涝级别序列更为合理,分布更接近于正态。分析还表明, Gamma分布对降水量的描述有较高的稳定性。

一、前 言

近年来由于气候灾害频发,使气象学者对气候变化及其影响的研究甚为关注。例如华北地区,特别是北京地区,由于连年的干旱以及工农业的发展,水资源的匮乏愈来愈严重。但是如何衡量一个地区的旱涝程度呢?气象、水文及农业工作者都有自己的标准来划分旱、涝。目前在气象上常用的旱涝指标是用降水量距平、距平百分率或涝等级来划分。这是以平均值为基点的旱涝指标。这种指标的缺点是没有考虑到降水量变化幅度的特点,对于一些内陆地区,由于降水量变化幅度大,而使距平值很大,很难与其它地区进行比较。故有同时考虑地区降水量的平均值和标准差的旱涝指标。这类指标有标准化降水量,即把降水量的距平值与多年样本的标准差值相比,以消除站点降水量变化幅度差异的影响。还有用这种标准化降水量的某些特殊值作为划分旱涝程度的依据。例如把降水量标准化值的临界值 -1.17 , -0.33 , 0.33 及 1.17 分别定义为五个级别的分界线,把降水量划分为5、4、3、2、1等级别,

对应为旱,偏旱,正常,偏涝,涝等气候状况[1]。

无论是以降水量的平均值,或者以平均值和标准差为基点划分的旱涝指标,均是以降水量遵从正态分布为前提条件的,即以平均值作为旱或涝气候状态的对称点。考虑划分的旱涝级别也是以正态分布为出发点的,上面提到的降水量标准化值的临界值分别对应于“降水量大于(相应)临界值”事件出现的概率为 0.121 , 0.371 , 0.629 , 0.879 。把降水量看成遵从以平均值(或期望)为中心点两侧是对称的正态分布,与事实有很大的出入。即使是年降水量,除在我国南方一些站点还可看成近似正态分布,而在北方大部分测站则不能看成为正态分布[2]。至于月和季降水量的分布与正态分布差别就更大[3]。因此把并不遵从正态分布的降水量按正态分布来划分旱涝级别或制作旱涝指标,对真实气候状态的描述显然会带来很大偏差,使用这种偏离实际气候状态的指标来分析气候变化规律,有可能并不反映真正的规律性。而且对这类并不遵从正态分布的统计量作统计检验,也不符合常用统计理论的基本假定。

因此,目前国外有关气候的分析总结中的旱涝级别划分,采用一种 Gamma分布概率指标来描述降水量所遵从的分布[4]。相应的五级旱涝级别对应的临界概率值分别定义为0.10, 0.30, 0.70, 0.90[5]。显然,这种划分较之我国的传统划分,扩大了“正常”级别出现的概率,而缩小了极端的“旱”、“涝”级别出现的概率。

本文着重介绍以降水量遵从 Gamma分布为出发点制作降水概率指标的过程。重点分析北京地区7月及夏季(6—8月)降水量的概率指标,并与传统的旱涝指标作比较。

二、方法

我们把某一地区月或季降水量看成为随机变量X,它遵从Gamma分布。Gamma分布的概率密度函数表示为

$$f(x) = \frac{1}{\beta^\gamma \Gamma(\gamma)} x^{\gamma-1} e^{-x/\beta} \quad (x > 0) \quad (1)$$

式中 $\beta(>0)$ 及 $\gamma(>0)$ 为参数,分别称为尺度和形状参数, Γ 表示通常的Gamma函数,即

$$\Gamma(\gamma) = \int_0^\infty x^{\gamma-1} e^{-x} dx \quad (2)$$

用容量为n的降水资料样本 $x_i (i=1, 2, \dots, n)$,用最大或然法可得出(1)式概率密度函数中两个参数的估计值,表示为[6]

$$\begin{cases} \hat{\gamma} = \frac{1 + \sqrt{1 + 4A/3}}{4A} \\ \hat{\beta} = \bar{x} / \hat{\gamma} \end{cases} \quad (3)$$

其中

$$A = \lg \bar{x} - \frac{1}{n} \sum_{i=1}^n \lg x_i$$

式中 \bar{x} 为样本平均值。在近似估计形状参数的过程中需要作订正, Thom[6]用列表形式给出不同的 $\hat{\gamma}$ 计算值的订正值 $\Delta \hat{\gamma}$ 。但是,从订正值随 $\hat{\gamma}$ 的变化来看,可以发现它们是非线性关系,可用多项式回归方程来描述。经过试验,用下面的求订正值 $\Delta \hat{\gamma}$ 的方程来描

述这种关系,可以达到令人满意的效果,即表示为

$$\Delta \hat{\gamma} = 0.0285 - 0.1800 \hat{\gamma} + 0.0024 \hat{\gamma}^2 \quad (4)$$

当计算得到的 $\hat{\gamma}$ 值后,减去由(4)式求出的订正值 $\Delta \hat{\gamma}$,即可得到较精确的形状参数 $\hat{\gamma}$ 的估计值,然后再使用(3)式可得到尺度参数 $\hat{\beta}$ 的最大或然估计值。

确定概率密度函数中的参数后,对于某一年的降水量 x_0 ,可求出“随机变量X小于 x_0 ”事件的概率为

$$P(X < x_0) = \int_0^{x_0} f(x) dx \quad (5)$$

利用数值积分可以计算用(1)式密度函数代入(5)式后的事件概率近似估计值。由此,逐年的降水量可以转换为遵从Gamma分布的概率估计值,得到相应的降水概率指标序列。再按分级概率临界值的定义,可划分出相应的旱涝级别序列。

三、北京地区7月降水概率指标

为了解释求取降水概率指标的优点,我们以北京7月降水量为例,取1951—1980年期间作为求概率密度函数中参数的基本样本。这一期间的样本已被许多国家采用为气候分析基本样本(见文献[4])。表1给出这一期间北京7月降水频数分布。从表中可见,

表1 北京7月降水频数分布(1951—1980)

区间 (mm)	组中值 (mm)	频数
0—60	30	0
60—120	90	4
120—180	150	6
180—240	210	8
240—300	270	7
300—360	330	4
360—420	390	0
420—480	450	0
480—540	510	0
540—600	570	1

降水量小于240 mm出现的次数达18次之多, 占总数的60%。而该月30年平均降水量为192.5mm, 这意味按小于平均值为偏旱来衡量气候旱涝标准的话, 那么北京地区出现偏旱情况为57%。这显然属于频率分布中的偏态情况。为了说明这一点, 我们分别计算样本的偏度系数及峰度系数。即用下式

$$g_1 = m_3/m_2^{3/2} \quad \text{及} \quad g_2 = \frac{m_4}{m_2^2} - 3$$

[其中 $m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$ ($k=2, 3, 4$)]

算得偏度(g_1)为0.946, 峰度(g_2)为1.980。若它们遵从正态分布, 这两参数95%的置信区间应分别为 ± 0.794 及 ± 0.251 。显然, 北京7月降水量不能看成是遵从正态分布的。

对(5)式用数值积分方法可求出这一期间“小于各年降水量”事件的概率估计值。根据分级概率临界值0.10, 0.30, 0.70及0.90可划分相应的级别, 结果发现划分的旱涝级别与我国传统划分方法所得的级别有所不同。表2给出级别不同的年份降水量标准化值、Gamma分布概率估计值以及根据各自的标准所划分的级别。从比较中可见, 原先认为是偏旱状态的年份(如1953, 1956, 1968, 1971, 1976), 按Gamma分布的标准仅看成为正常年份。其原因是北京地区降水少的年份多。北京有“十年九旱”的气候谚语描述这种状态。事实上, 在30年中, 低于平均值(192.5mm)的年份达17年之多, 占总年数的57%。按正态分布的分布规律划分自然是偏旱的情况较多。而事实上, 由于北京降水偏少常发生。那么把某些“旱”年看为“正常”年份是应该的。至于原先定为“涝”年的两年, 例如在1979年, 降水量为228.1 mm, 是属于常见降水量的范围内(见表1), 划为“偏涝”是不太合适的, 划为“正常”更合理一些; 1969年, 降水量虽然较

表 2 旱涝级别不同的年份比较

年份	降水量(mm)	标准化值	概率值	原级别	新级别
1953	140.8	-0.54	0.35	4	3
1956	131.4	-0.64	0.31	4	3
1968	138.8	-0.53	0.34	4	3
1969	310.6	1.23	0.88	1	2
1971	154.0	-0.40	0.41	4	3
1976	159.7	-0.34	0.44	4	3
1979	228.1	0.37	0.69	2	3

(注) 原级别表示根据正态分布划分的级别, 新级别表示根据Gamma分布划分的级别。

多, 但比起1959年511.1 mm的降水量来说, 不能说是很多, 故划为“涝”年也是不合适的, 只能算是“偏涝”。显然, 使用Gamma分布划分旱涝状态比较合理地反映真正的气候状态。当然, 这种状态是针对气象特点而言的, 至于水资源需求迫切的工农业来说, 北京近年来干旱严重的情况也是符合实际的。

另外, 从旱涝级别频数分布比较(见表3)也可以看到, 在30年期间用正态分布划

表 3 旱涝级别频数分布比较

级别	1	2	3	4	5
正态	2	9	7	9	3
Gamma	1	9	13	4	3

分的级别呈“M”型分布, “正常”年份反而比“偏涝”或“偏旱”的年份少。用正态分布划分的级别本来是希望反映降水量的正态分布, 但是结果反而不是正态分布。而用Gamma分布划分的结果则接近于正态分布, 划分的旱涝状态又比较符合实际气候状态。例如, 划分“涝”的1年是1959年, 降水量为511.1mm, 划为“旱”的3年(1951, 1965, 1980)其降水量都小于70mm。

四、北京夏季及年的降水概率指标

用1951—1980年夏季(6—8)月及年的北京降水量资料进行类似分析。计算它们的样本偏度及峰度系数, 季降水量为1.235及2.094, 年降水量为1.078及1.376。显著性

检验表明,它们均不能视为正态分布。比较而言,夏季降水量偏态情况尤为严重。降水量500mm以下的年份达18年之多,占总样本容量约60%,而降水量500—1200 mm的多雨年份仅出现12年。用 χ^2 检验还发现,夏季及年降水量均可遵从以尺度参数82.663(夏季)及84.491(年)以及形状参数5.839(夏季)及7.625(年)的Gamma分布。

表4给出分别用正态分布标准及用Gamma分布标准所作的旱涝级别划分,并显示不同级别的分布状况。从表中可见,无论是夏季和年的旱涝级别,用正态分布标准划分后,均显示有偏(偏旱)的分布,而用Gamma分布标准划分,则大致遵从正态分布。以年降水量为例,1952年降水总量为

表4 夏季及年旱 级别频数分布

级 别		1	2	3	4	5
夏 季	正态	3	5	9	10	3'
	Gamma	3	4	16	2	5
年	正态	4	4	9	12	1
	Gamma	3	5	13	7	2

547.3mm,按正态分布的标准,其标准化值为-0.40,划为“4”级,而实际上该年降水量仅比30年的平均值(664.2mm)少116.9mm。划为正常状态较好。另外,1962年降水量为366.9mm,差不多为平均值的一半,却因为标准化值(-1.14)按习惯划为偏旱,而Gamma分布的标准可划为特旱。由此可见,用Gamma分布对夏季及年降水量作旱涝级别划分也是比较合理的。

五、参数稳定性的比较

为了进一步说明用Gamma分布制作的降水概率指标的优越性,还可以从参数稳定性的角度来说明。因为通常使用的旱涝级别是基于降水量遵从正态分布来划分的,而正态分布主要依赖于2个参数,即平均值与标准差。对于基于Gamma分布的概率值来划分旱涝级别,同样也取决于分布函数中的2个

参数,即尺度与形状参数。比较这4个参数在不同时段的样本中变化情况是有必要的。

选取本世纪以来(1901—1980年期间)6个30年的滑动样本(以10年为间隔滑动),分别计算不同时段样本中4个参数。表5给出这4个参数在7月、夏季及年的降水量中总平均值及标准差。从表中可见,无论在北京的各种尺度降水量中,Gamma分布中的参数

表5 参数变化比较

	平均值				标准差			
	v	β	\bar{x}	s	v	β	\bar{x}	s
7月	3.26	63.46	221.0	121.8	0.31	8.68	19.04	15.59
夏季	6.24	76.49	471.9	197.3	1.40	6.04	10.69	12.86
年	3.66	73.79	632.7	210.3	1.79	8.20	22.92	21.54

(注) v, β 表示Gamma分布的形状和尺度参数;
 \bar{x} , s表示正态分布的平均值和标准差。

变化都不大,比较而言,正态分布的参数变化幅度平均比Gamma分布的参数要大(见表5中标准差一项)。比较还发现,1951—1980年所计算的Gamma分布在7月、夏季及年的参数在不同时段中变化不大,由于Gamma分布参数较为稳定,因此也可以将这一时段的分布参数用于其它时段的旱涝级别划分中[4]。

六、结论与讨论

本文介绍以降水量遵从Gamma分布制成的降水概率指标,以及由降水概率转换成旱涝级别序列。

对北京地区7月,夏季(6—8月)及年降水概率指标及相应的旱涝级别序列分析表明,利用Gamma分布概率划分的旱涝级别较之传统的旱涝级别在描述真正的气候状态方面更为合理,在级别序列的样本中,不同级别的分布较传统的级别序列更接近于正态分布。对滑动样本中不同分布的参数变化性分析表明,Gamma分布中的参数较之正态分布的参数具有更高的稳定性。因此基于1951—1980年样本所产生的Gamma分布参数是

具有代表性的, 可以适用于其它时段中。

当然, 对于我国南方一些台站, 年降水量能较好的遵从正态分布, 对这些地区, 旱涝级别划分仍沿用传统方法也是可以的。不过, 由于Gamma分布也可以描述正态分布的特殊情况, 且前已被许多国家采用作为分析旱涝气候状态的主要工具。

参 考 文 献

(1) 中央气象局等, 中国近五百年旱涝分布图集, 地图出版社, pp.332, 1980。

(2) 徐尔灏, 论年雨量之常态性, 气象学报, 21, 17—34, 1950。

(3) 黄嘉佑, 气象统计预报讲义, 北京大学地球物理系, 1979。

(4) Arkin, P.A., The global climate for December 1988—February 1989.

(5) Ropelewski, C.F., J.E.Janowiak and M.S. Halpert, The analysis and display of real time surface climate data. Mon. Wea. Rev., 113, 1101—1106, 1985.

(6) Thom, H. C. S., A note on the Gamma distribution, Mon. Wea. Rev., 86, 117—121, 1958.

An index of precipitation probability for dryness/wetness analysis

Huang Jiayou

(Department of Geophysics, Peking University)

Abstract

The index of precipitation probability and the grade of dryness/wetness based on Gamma distribution are presented in this paper. Sequential analysis on probability index of precipitation (in July, summer and in the year) and of the grade of dryness/wetness in Beijing area suggest that the series of the index and the grade of dryness/wetness based on Gamma distribution are better than the ones based on Normal distribution in describing the climatic states, The distribution of the former is closer to Normal than the latter. The results also show that Gamma distribution is more stable with time in presenting the precipitation.