

我国日平均气温概率分布的初步研究[†]

鞠笑生

江志红 戴廷仁

(国家气候中心,北京 100081) (南京气象学院应用气象学系,南京 210044)

摘要 利用分布在华北、华东、华中、华南的5个代表站(北京、太原、上海、武汉、广州)1978~1987年的温度资料,运用气候统计特征量研究、分析了我国日平均气温的概率分布。结果表明:多数站点气温的概率分布呈正态分布,少数站点个别月份为负偏态分布。

关键词 气候统计特征量,正态分布,偏态分布,平均气温

分类号 P423.36

根据长期气象记录,利用气候学统计原理和方法,估计出能综合反映气象序列分布规律的度量称之为气候统计特征量或气候指标。气候统计特征量主要用于对气象要素的记录序列进行分析和比较,以描述气候变化特征,分析气候演变的客观规律,从而预测未来气候变化的基本趋势。

在生产实践和科学研究中,应用最广泛的气候要素是温度,所以本文主要应用气候统计特征量来研究温度的概率分布规律。有关温度统计指标的形式很多,如平均温度(日、候、旬、月、年)、极端温度(平均最高、平均最低、极端最高、极端最低)等,气象工作者已对这些指标的时空分布作了大量的研究。但对一个随机变量来讲,概率分布是其最完全的概率特征,对温度的各种指标的研究必须包括对其概率分布的研究。

铃木荣一(1968年)^[1]曾系统地总结过日本各地气温的概率分布,指出日、旬、月平均气温以及最高气温多数属于正态分布,但候平均气温、日最高气温、月最低气温和年最低气温有的属于 Pearson I 型分布,并且发现年平均温度有双峰型趋势。

本文利用均匀分布在华北、华东、华中、华南的5个代表站点(北京、太原、上海、武汉、广州)近10年的日平均温度(1、7月)及1月日最低温度、7月日最高温度资料,初步分析了上述地区气温的概率分布特征。

1 气候特征量

随机变量概率分布的性质可以通过特征量来描述,因此在具体研究概率分布之前,首先要计算统计特征量。通常使用的统计特征量有下面几种:

- (1)平均值 \bar{X} 这是经常使用的气候特征量,表征频率分布的中心位置。
- (2)均方差 σ 它是刻划观测值频率分布离散程度的统计特征量。

[†] 国家气象局短平快项目资助课题

收稿日期:1995-03-18;改回日期:1995-10-16

(3) 偏度系数 A 用于衡量频率分布的不对称程度。

$$A = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n\sigma^3} \quad (1)$$

当 $A > 0$ 时, 为正偏; $A < 0$ 时, 为负偏; $A = 0$ 时, 为对称分布。计算表明, $|A| \leq 0.25$ 时, 偏度较小; $0.25 < |A| \leq 0.5$ 时, 偏度中等; $|A| > 0.5$ 时, 偏度较大。

(4) 峰度系数 E 它是频率分布陡度的度量。

$$E = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{n\sigma^4} - 3 \quad (2)$$

峰度系数数值范围从 -2 至 ∞ , 当 $E < 0$ 时, 为低峰态; 当 $E > 0$ 时, 为高峰态; $E = 0$ 时, 表示正态。通常情况下, 峰度大小可以用峰度系数绝对值判断, 当 $|E| \leq 0.5$ 时, 峰度比较小; 当 $0.5 < |E| \leq 1.0$ 时, 峰度中等; 而当 $|E| > 1.0$ 时, 峰度比较大。

2 统计原理和方法

概率分布是一个随机变量最完全的概率特征, 任何变量的概率估计和计算都必须建立在一定概率分布模式的基础上。在未知温度指标服从何种分布模式之前, 先介绍两种概率分布模式。

(1) 正态分布 正态分布是连续随机变量分布中最常见的一种, 在概率统计理论中具有特别重要地位, 其概率密度函数为

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (3)$$

其中 m 和 σ 为参数, 可用下式进行估计

$$\left. \begin{aligned} m &\doteq \hat{m} = \frac{1}{n} \sum_{i=1}^n x_i \\ \sigma &\doteq \hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{m})^2} \end{aligned} \right\} \quad (4)$$

由此可见, 通过对样本计算可得到 m 和 σ 的估计值, 然后可以进行概率分布拟合。

(2) 耿贝尔分布 对于极端温度值, 常用耿贝尔分布拟合。其分布函数为

$$F(x) = e^{-e^{-\alpha(x-u)}} \quad (\alpha > 0, \quad -\infty < u < +\infty) \quad (5)$$

式中 α 和 u 为两个参数。

由耿贝尔法, 可得到 α 与 u 的估计值

$$\left. \begin{aligned} \alpha &\doteq \hat{\alpha} = \frac{S_y}{S_{x_m}} \\ u &\doteq \hat{u} = \bar{x}_m - \frac{S_{x_m} \bar{y}}{S_y} \end{aligned} \right\} \quad (6)$$

其中, $y_m = -\ln[-\ln(1 - \frac{m}{N+1})]$, $m = 1, 2, \dots, N$ 。

3 拟合优度检验

在进行气象要素概率分布拟合时, 所选理论分布模型是否适宜, 通常要进行分布的适度测验, 也就是分布统计假设检验。

本文无论对正态分布还是对耿贝尔分布进行拟合优度检验,都是使用 χ^2 检验。原理如下:若假设样本满足某个理论分布(如正态分布或耿贝尔分布),则将样本分成 K 组,分别计算各组的理论频数和实测频数,则不论样本的原始分布为何种分布,其统计量

$$\chi^2 = \sum_{i=1}^K \frac{(f_K - P_K)^2}{P_K} \quad (7)$$

是服从自由度为 $K-N-1$ 的 χ^2 分布的随机变量,其中 f_K 为第 K 组的实测频数; P_K 为第 K 组的理论频数; N 为相应理论分布需估计的参数个数。计算出统计量 χ^2 后,便可以进行拟合优度假设检验。具体步骤如下。

(1)先处理原始资料,分别计算样本的均值和均方差,根据(1)和(2)式计算样本概率分布的偏度系数和峰度系数。

(2)根据拟合优度检验原理,将 n 个样本分成 K 组,计算出每组的实测频数,分成的组数 K 满足在 $2.5\lg n \sim 5\lg n$ 之间。

(3)根据计算出的样本值和均方差值,再按照(4)或(6)式计算出假设理论分布参数的估计值,这样就可以根据假设的理论分布模型的公式(见(3)和(5)式)和参数的估计值计算出相应的理论频数。

(4)根据实测频数和理论频数值,绘制出分布的概率直方图。

(5)假设检验,对于实测频数小于5的组应与相邻组合并,直到大于、等于5为止。根据(7)式计算统计量 χ^2 。

对于一定的信度 $\alpha=0.05$ 或 $\alpha=0.01$,可查表求出自由度为 $K-3$ 的 χ^2 分布临界值。若 $\chi^2 > \chi_{\alpha}^2$,样本不满足该理论分布,若 $\chi^2 < \chi_{\alpha}^2$,样本满足该理论分布。

4 结果分析

上述统计量中,偏度系数和峰度系数是描述概率分布型最重要的统计特征量,根据偏度系数和峰度系数的计算结果可以决定温度指标所服从的概率分布模型。计算和模拟结果表明:

(1)无论是日平均温度还是日极值温度,大多数台站概率分布的偏度系数和峰度系数的值都比较小,且这两个指标概率分布的偏度系数的绝对值多数都小于0.5,峰度系数的绝对值多数都小于1。由此可见,温度指标的概率分布大都接近对称分布,因此,可用正态分布模型进行拟合。

(2)北京、太原、上海、武汉、广州5个站的1月日平均温度和日最低温度的概率分布都能通过信度 $\alpha=0.05$ 的 χ^2 检验,呈正态分布。而7月日平均温度的概率分布只有北京和太原能通过信度 $\alpha=0.05$ 的 χ^2 检验,呈正态分布,上海、武汉、广州均不能通过信度 $\alpha=0.01$ 的 χ^2 检验,呈偏态分布。对7月日最高温度,上海、武汉、广州均不能通过信度 $\alpha=0.01$ 的 χ^2 检验,呈偏态分布。只有北京能通过信度 $\alpha=0.05$ 的 χ^2 检验,太原能通过信度 $\alpha=0.01$ 的 χ^2 检验,呈正态分布。

(3)从广州7月日平均温度及日最高温度的概率直方图(图1、2)可见,这两个温度指标的概率分布均呈负偏分布。该指标在上海和武汉的概率分布也是如此(图略)。可见7月日平均温度和日最高温度的概率分布因地区不同而有明显变化,北方两站呈正态分布,南方三站呈负偏分布。

(4)从上海、武汉、广州7月日平均温度及日最高温度的偏度系数和峰度系数(表1)可以看出,无论是日平均温度还是日最高温度,其概率分布的偏度系数均为负值,呈负偏分布,而且

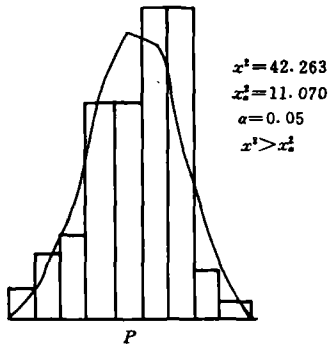


图 1 广州 7 月日平均温度概率直方图
(折线为正态分布)

Fig. 1 Vertical histogram of Guanzhou daily mean temperature probability in July
(The broken line denotes the normal distribution)

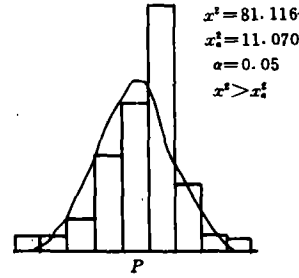


图 2 广州 7 月日最高温度概率直方图
(折线为正态分布)

Fig. 2 Vertical histogram of Guanzhou daily maximum temperature probability in July
(The broken line denotes the normal distribution)

随着纬度降低,偏度系数的绝对值越大。峰度系数只有广州 7 月日最高温度为正值,呈高峰态分布,其他站均为低峰态分布。

表 1 上海、武汉和广州 7 月日平均温度及日最高温度的偏度系数和峰度系数

Table 1 Coefficients of skewness and kurtosis of daily mean and maximum temperatures in July for Shanghai, Wuhan and Guangzhou

	日平均温度			日最高温度		
	上海	武汉	广州	上海	武汉	广州
偏度系数	-0.116	-0.494	-0.738	-0.348	-0.756	-0.980
峰度系数	-0.316	-0.520	-0.204	-0.489	-0.251	1.344

(5)用耿贝尔分布对上海、武汉、广州 3 站的 7 月日最高温度进行拟合,得到实测频数和理论频数值(见表 2)。由表 2 可见,实际的频数分布与耿贝尔理论分布差别很大,拟合优度检验结果表明,上述 3 站 7 月日最高温度也不满足耿贝尔分布($\alpha = 0.01$)。

5 结 论

(1)大部分台站的日平均温度和日极值温度的偏度系数值和峰度系数值都比较小,其概率分布大都接近正态分布。

(2)各站 1 月日平均温度和日最低温度均为正态分布。7 月日平均温度和日最高温度的概率分布因分布地区不同而有明显变化,北方两站(北京、太原)呈正态分布,南方 3 站(上海、武汉、广州)呈负偏分布。

(3)南方 3 站 7 月日平均温度和日最高温度随着纬度降低,偏度系数的绝对值增大。峰度系数只有广州 7 月日最高温度为正值,呈高峰态分布,其他两站 7 月最高温度以及 3 个站 7 月日平均温度均为低峰态分布。

表 2 上海、武汉、广州 7 月日最高温度频数
Table 2 Frequency of daily maximum temperature in July
for Shanghai, Wuhan and Guangzhou

上海				武汉				广州			
组中值	实测 频数	正态分布 理论频数	耿贝尔分布 理论频数	组中值	实测 频数	正态分布 理论频数	耿贝尔分布 理论频数	组中值	实测 频数	正态分布 理论频数	耿贝尔分布 理论频数
22.7	0	1.03	0	21.7	0	0.18	0	23.6	0	0.94	0
24.3	6	4.00	0.15	23.4	1	1.03	0	25.1	1	5.46	0
25.9	19	11.89	5.56	25.1	11	4.42	0.18	26.6	7	20.40	0
27.5	34	26.96	32.64	26.8	20	13.91	7.31	28.1	10	49.29	1.15
29.1	21	46.69	65.55	28.5	26	32.17	40.73	29.6	18	76.98	28.25
30.7	49	61.76	71.09	30.2	40	54.64	74.10	31.1	44	77.72	85.63
32.3	80	62.38	54.87	31.9	38	68.19	72.45	32.6	71	50.72	89.22
33.9	52	48.13	35.24	33.6	86	62.50	50.98	34.1	123	21.40	55.83
35.5	36	28.36	20.54	35.3	75	42.09	30.27	35.6	28	5.83	22.84
37.1	12	12.76	11.37	37.0	13	20.82	16.51	37.1	7	1.03	12.59
38.7	1	4.39	12.99	38.7	0	7.56	17.47	38.6	1	0.12	9.49

参 考 文 献

- 1 铃木荣一[日]. 气象统计学. 北京: 地人印书馆, 1968
- 2 屠其璜, 丁裕国, 王俊德, 等. 气象应用概率统计学, 北京: 气象出版社, 1984
- 3 么枕生, 丁裕国. 气候统计. 北京: 气象出版社, 1990
- 4 Ковышева Н В, Годьберг М А. Мемогтмескпе укэжэнтпе но смэптэмтескпе олдрэомке мемэоролэптэсктэ иэгощ, лэмпирэг ттпнэомэопжтэм, 1990

INITIAL INVESTIGATION OF PROBABILITY DISTRIBUTION OF DAILY MEAN TEMPERATURE OVER CHINA

Ju Xiaosheng¹⁾ Jiang Zhihong²⁾ Dai Tingren²⁾

(1) National Climate Center of China, Beijing 100081, 2) Department of Applied Meteorology, NIM, Nanjing 210044)

Abstract Using 1978 ~ 1987 temperature of Beijing, Taiyuan, Shanghai, Wuhan and Guangzhou located in N, E, S, and Central China and with the aid of climatologically characteristic statistics, investigation is carried out of the probability distribution of daily mean temperatures over the country. Results show that the distribution is of normal character with a limited number of stations showing negative skew nature.

Keywords climatologically characteristic statistic, normal distribution, skew distribution, mean temperature