

EVALUATION OF THE PRECISION OF STANDARDIZED PRECIPITATION INDEX (SPI) BASED ON YEARS 1954–1995 IN ŁÓDŹ

Edward Gąsiorek¹, Elżbieta Musiał¹

¹ Department of Mathematics, Wrocław University of Environmental and Life Sciences, Grunwaldzka St. 53, 50-357 Wrocław, Poland, e-mail: edward.gasiorek@up.wroc.pl; elzbieta.musial@up.wroc.pl

Received: 2015.07.14
Accepted: 2015.08.31
Published: 2015.10.01

ABSTRACT

This paper evaluates the accuracy of estimates of the Standardized Precipitation Index (SPI) using gamma, normal and log-normal distributions. In order to classify the above methods, the authors performed an analysis of the quality of theoretical distributions to empirical distribution, obtained on the basis of monthly precipitation sums during the vegetation season in a multi-year period 1954–1995 in Łódź.

Keywords: standardized precipitation index, gamma distribution, normal distribution, log-normal distribution.

INTRODUCTION

Standardized Precipitation Index (SPI) identifying different types of meteorological drought, can be determined by methods employing various distributions of probabilities. These methods are based on normal and gamma distributions [Krishnamoorthy et al. 2008, Thom 1958]. The authors [Gąsiorek et al. 2012] have shown that since methods based on normal and gamma distributions give similar SPI values, they may be used interchangeably. In their previous study, the authors [Gąsiorek et al. 2014] assessed the accuracy of estimates of the SPI, based on the data from Wrocław-Swojec observatory, with the use of gamma, normal and log-normal distributions.

The present study analyzes the quality of the appropriateness of gamma, normal and log-normal distributions to the empirical distribution obtained on the basis of monthly precipitation sums during the vegetation season in a multi-year period 1954–1995 in Łódź.

METHODS

The appropriateness of theoretical distribution do the empirical data can be presented graphically with the use of empirical distribution func-

tion and theoretical distribution function [Wibig 2012]. Kolmogorov-Smirnov statistic was applied as a measure of appropriateness of the theoretical distribution to the empirical distribution of monthly precipitation sums in a multi-year period [Kaczmarek 1970, Angelidis et al. 2012]. The value of Kolmogorov-Smirnov statistics is the maximum difference between the empirical cumulative distribution function and the theoretical cumulative distribution function. Thus, smaller values of the statistics indicate that the matched theoretical distribution deviates less from the empirical distribution.

Out of the transformations normalizing the empirical distribution, we have chosen the one for which the Kolmogorov-Smirnov statistic was the lowest, thus the match was best guaranteed. Having selected the transformation giving the best match, the verification of compliance with the normal distribution using the Shapiro-Wilk test was performed. The compliance with the gamma and log-normal distributions was checked with the χ^2 test. Kolmogorov-Smirnov statistic D_n is of the form:

$$D_n = \sup |F_n(x) - F_0(x)|$$

where: F_0 is the cumulative distribution function of the theoretical distribution and F_n denotes the empirical distribution function.

The value of D_n statistic is calculated following [Krysicki et al. 1997]:

- we sort the result of observations in the form of an increasing sequence $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, where $x_{(i)}$ denotes positional statistics (the i -th observation in the ordered sequence of data);
- we calculate $\frac{i}{n} - F_0(x_{(i)})$ where $i=1, \dots, n$;
- we calculate $d_n^+ = \max_i \left| \frac{i}{n} - F_0(x_{(i)}) \right|$;
- we calculate $F_0(x_{(i)}) - \frac{i-1}{n}$ where $i=1, \dots, n$;
- we calculate $d_n^- = \max_i \left| F_0(x_{(i)}) - \frac{i-1}{n} \right|$;
- we calculate $d_n = \max(d_n^+, d_n^-)$.

Gamma distribution density function is of the form:

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{dla } x > 0 \\ 0 & \text{dla } x \leq 0 \end{cases}$$

where:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

and distribution:

$$F(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x y^\alpha e^{-\frac{y}{\beta}} dy$$

In order to determine the parameters α and β , gamma distribution we used estimators proposed by Edwards and McKee [1977] calculated by Thom [1958] method of maximum likelihood:

$$\hat{\alpha} = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right), \hat{\beta} = \frac{\bar{x}}{\hat{\alpha}}$$

where: $A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n}$ and \bar{x} denotes the mean value.

Density function of the log-normal distribution is as follows:

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & \text{dla } x > 0 \\ 0 & \text{dla } x \leq 0 \end{cases}$$

while the distribution function has the form:

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

where Φ denotes the cumulative distribution function of standard normal distribution.

Since the gamma function is undefined for $x=0$ and precipitation distribution may contain zeros, the cumulative probability $H(x)$, is calculated by the following formula

$$H(x) = (1 - q)F(x)$$

where q is the probability of zero.

Standardized precipitation index is determined by the formula:

$$SPI = \Phi^{-1}(F(x))$$

In order to determine the standardized precipitation index approximation method can be applied quantile of the standard normal distribution [Abramowitz et al. 1965].

For an approximation of the empirical distribution by the normal distribution is used as a transformation $\sqrt[3]{x}$, $\sqrt[3]{x+10}$ i $\ln x$, and the SPI index was obtained after standardization of the transformed data.

Estimators for the parameters μ and σ is determined using the method of maximum likelihood:

$$\hat{\mu} = \frac{\sum \ln x}{n}$$

$$\hat{\sigma}^2 = \frac{\sum (\ln x - \mu)^2}{n}$$

In order to verify the compliance of probability distribution with the empirical distribution of precipitation sums during the vegetation season, the χ^2 test was used.

The depiction of the gamma distribution fit by graphic comparison of the empirical distribution function with the theoretical one, was presented on the example of July in a multi-year period 1954–1995 in Łódź (Figure 1). The lower the value of d_n statistics, the better fit of the theoretical distribution to the empirical one.

RESULTS

Table 1 contains the values of applied statistics D_n for the transformations normalizing monthly precipitation sums in years 1954–1995, with the use of Kolmogorov-Smirnov statistics.

Data on the verification of hypotheses on compliance with normal distribution by the Shapiro-Wilk test, and on compliance with the log-normal and gamma distributions by the χ^2 test, were presented in Table 2.

Then, in order to compare the fit results for the individual distributions to the empirical data, the values of D_n statistics were set for gamma, log-normal and normal distributions (Table 3). The determined values of Kolmogorov-Smirnov statistics for normal, gamma and log-normal distributions, listed in Table 3, are presented in a graphical form (Figure 2).

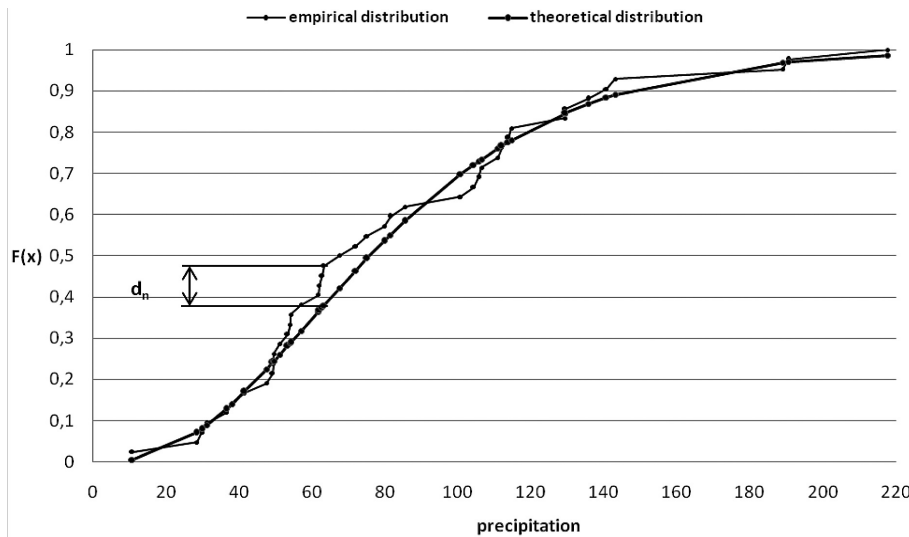


Figure 1. Empirical distribution function and fitted gamma distribution function for July in years 1954–1995 in Łódź with a marked value of d_n statistics

Table 1. Best normalizing transformations of monthly precipitation sums in a multi-year period 1954–1995 with the use of Kolmogorov-Smirnov statistics

Month	Kolmogorov-Smirnov statistics			Chosen transformation
	$\sqrt[3]{P}$	$\sqrt[3]{P+10}$	$\ln P$	
April	0.0635	0.0748	0.0742	$\sqrt[3]{P}$
May	0.0870	0.0977	0.0894	$\sqrt[3]{P}$
June	0.0860	0.0904	0.0814	$\ln P$
July	0.0969	0.1069	0.0946	$\ln P$
August	0.0638	0.0731	0.0862	$\sqrt[3]{P}$
September	0.1044	0.1020	0.1492	$\sqrt[3]{P+10}$

P – monthly precipitation sums in the analyzed month in years 1954–1995.

Table 2. Consistency of gamma, normal and log-normal distributions of monthly precipitations sums with empirical distribution in the years 1954–1995 in Łódź

Month	Theoretical distribution							p-value
	Log-normal			Gamma			Normal transformation	
	p-value	μ	σ^2	p-value	α -shape	β -scale		
April	0.59	3.50	0.34	0.68	3.54	10.87	$Y = \sqrt[3]{P}$	0.86
May	0.59	3.88	0.24	0.58	4.61	11.79	$Y = \sqrt[3]{P}$	0.93
June	0.86	4.12	0.27	0.81	4.02	17.44	$Y = \ln P$	0.71
July	0.13	4.27	0.36	0.11	3.30	25.53	$Y = \ln P$	0.25
August	0.96	4.08	0.28	0.85	4.00	16.91	$Y = \sqrt[3]{P}$	0.99
September	0.03	3.67	0.47	0.05	2.82	16.83	$Y = \sqrt[3]{P+10}$	0.83

Table 3. The values of D_n statistics calculated for subsequent months for gamma, log-normal and normal distributions

Month	Value of D_n statistics		
	Gamma distribution	Log-normal distribution	Normal distribution
April	0.062	0.074	0.064
May	0.087	0.089	0.087
June	0.086	0.081	0.081
July	0.096	0.095	0.095
August	0.063	0.086	0.064
September	0.107	0.149	0.102

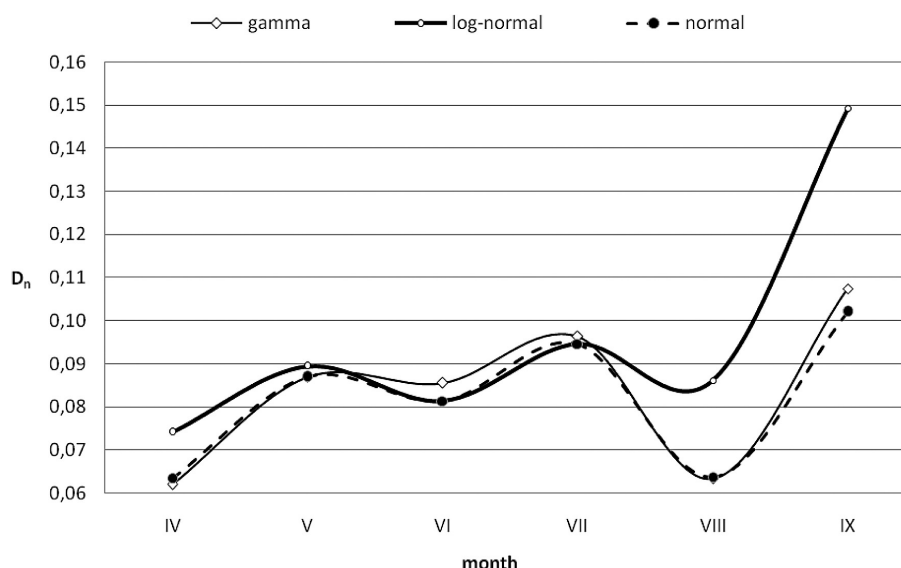


Figure 2. The values of D_n statistics calculated for subsequent months for gamma, log-normal and normal distributions

Figure 2 shows that the values of D_n statistics for normal and gamma distributions, characterizing the accuracy of the fit of the theoretical distribution to the empirical distribution, are similar. For months of the vegetation season, the values of D_n statistics for normal distribution and gamma distribution vary slightly (April – 0.002, May – identical, June – 0.005, July – 0.001, August – 0.001, September – 0.005), which shows the high accuracy of estimates by means of both distributions. The D_n statistics values for the log-normal distribution, characterizing the accuracy of appropriateness of the theoretical distribution to the empirical distribution, differ significantly for April – 0.01 August – 0.23 and September – 0.042, while for the remaining months of the vegetation season, these differences are minor (May – 0.002, June – 0.005, July – 0.001).

The results indicate that the evaluation of meteorological drought with the standardized precipitation index gives nearly identical classifications when using gamma and normal distributions, differing significantly from those obtained by the log-normal distribution use.

CONCLUSIONS

Evaluation of quality of fit of theoretical distributions to empirical distribution, obtained on the basis of monthly precipitation sums in the multi-year period (1954–1995) in Łódź with the use of Kolmogorov-Smirnov statistics, indicates that:

- the best fit is obtained using the approximation of the empirical distribution to the gamma or normal distribution, with the chosen data transposition,
- the weakest fit to the empirical distribution is obtained by the use of log-normal distribution,
- methods based on normal and gamma distributions give similar SPI values, thus they may be used interchangeably.

REFERENCES

1. Abramowitz M., Stegun I.A. 1965. Handbook of mathematical formulas, graphs and mathematical tables. Dover Publications, New York.
2. Angelidis P., Maris F., Kotsovinos N., Hrisanthou V. 2012. Computation of drought index SPI with alternative distribution function. *Water Resource Manage.*, 26, 2453–2473,
3. Edwards DC., McKee TB. 1997. Characteristics of 20th century drought in the United States at multiple scales. *Climatology Report 97-2*, Department of Atmospheric Science, Colorado State University, Fort Collins.
4. Gąsiorek E., Grządziel M., Musiał E., Rojek M. 2012. The comparison of standardized precipitation index (SPI) evaluated by gamma and normal distribution for monthly precipitation sums. *Infrastruktura i Ekologia Terenów Wiejskich PAN 3/III 2012*, 197–208.
5. Gąsiorek E., Musiał E., Rojek M. 2014. Assessment of accuracy of precipitation index (SPI) determined by different probability distributions. *Inżynieria Ekologiczna*, 40, 92–99.

6. Kaczmarek Z. 1970. Metody statystyczne w hydrologii i meteorologii. Warszawa, WKiŁ.
7. Krishnamoorthy K., Mathew T., Mukherjee S., 2008. Normal-based methods for a Gamma Distribution: prediction and tolerance intervals and stress-strength reliability. *Technometrics*, February 2008, 50(1), 69–78.
8. Kryszicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M. 1997. Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach. Wydawnictwo Naukowe PWN,
9. Łabędzki L. 2006a. Susze rolnicze, Zarys problematyki oraz metody monitorowania i klasyfikacji. Woda Środowisko Obszary Wiejskie, Rozprawy naukowe i monografie, No. 17.
10. Łabędzki L, 2006b. O dwóch metodach wyznaczania wskaźnika suszy meteorologicznej SPI. Materiały XXXVI Seminarium Zastosowań Matematyki, 41–47.
11. McKee T.B., Doesken N.J., Kleist J. 1993. The relationship of drought frequency and duration to time scales. Proc, 8th Conf, Applied Climatology, 17-22 January 1993, Anaheim, California, 179–184.
12. McKee T.B., Doesken N.J., Kleist J. 1995. Drought monitoring with multiple time scales. Preprints 9th Conf, Applied Climatology, 15-20 January 1995, Dallas, Texas, 233–236.
13. Thom H.C.S. 1958. A note on the gamma distribution. *Monthly Weather Review*, 86, 117–122.
14. Wibig J. 2012. Warunki wilgotnościowe w Polsce w świetle wskaźnika standaryzowanego klimatycznego bilansu wodnego. Woda-Środowisko- Obszary Wiejskie, t. 12, z. 2(38), 329–340.