

O PORAZDELITVI LETNIH EKSTREMOV NALIVOV V SLOVENIJI
ABOUT THE DISTRIBUTION OF YEARLY EXTREMES OF SHOWERS IN
SLOVENIA

551.577.37:519.2

JOŽE ROŠKAR
Meteorološki zavod SRS, Ljubljana

SUMMARY

The negative sign of constant k which gives the density function of yearly extremes of showers was tested. Jenkinson /1/ supposed that constant k , which is found in the general solution of Fisher-Tippet's equation

$$x = a(1 - e^{-ky}),$$

has the negative sign for showers. The prediction of extreme intensities of showers for a defined period depends to a great extent on the sign of constant k . In this paper it is shown that Jenkinson's supposition is doubtful if applied on yearly extremes of showers in Slovenia. According to Jenkinson, the probability that the constant k has a positive sign if applied on any set of data for showers, is very close to zero. In our example this probability is 0.64. This value shows that constant k for showers in Slovenia has a positive sign as a rule. The consequence of this statement is that another basis must be taken for prediction of extremes showers. These can be found in a more general solution of Fisher-Tippet's equation.

POVZETEK

Testirali smo negativnost koeficienta k za gostoto verjetnosti letnih ekstremov nalivov. A. F. Jenkinson /1/ je domneval, da je konstanta k , ki nastopa pri splošni rešitvi Fisher-Tippetove enačbe, negativna. Fisher-Tippetova funkcijska enačba je

$$P^S(x) = P(a_s + b_s),$$

za katero je Jenkinson našel naslednjo rešitev

$$x = a(1 - e^{-ky}).$$

Od znaka konstante k v gornji rešitvi je močno odvisna ocena napovedi ekstrema za neko določeno obdobje. V tem prispevku smo pokazali, da je Jenkinsonova

domneva dvomljiva vsaj za letne ekstreme nalivov v Sloveniji. Po Jenkinsonu bi morala biti verjetnost, da je konstanta k poljubnega niza podatkov o nalivih pozitivna, zelo blizu 0. V našem primeru pa je ta verjetnost ocenjena na 0.64, kar kaže, da so konstante k za nalive v Sloveniji praviloma pozitivne. Posledica te ugotovitve je vprašanje, kako sedaj računati ocene za napoved ekstremnih nalivov. Verjetno bo treba najti splošnejšo rešitev Fisher-Tippetove funkcijske enačbe.

UVOD

Sistematična meteorološka opazovanja obsegajo relativno kratko časovno dobo. V praksi pa nas določeni problemi nemalokrat postavijo pred nalogo, napovedati ekstreme za neki parameter vremena za daljšo dobo, za recimo 50, 100 ali več let. Seveda lahko nalogo rešimo le na osnovi obstoječih opazovanj. V Sloveniji so nizi opazovanj dokaj kratki. Večina sistematičnih opazovanj se je začela po drugi svetovni vojni; redki so kraji, ki se ponašajo z daljšimi nizi podatkov (Ljubljana - 125 let). Kljub temu pa se le da nekaj narediti. Fisher in Tippet sta že leta 1928 postavila funkcijsko enačbo, s katero se lahko dobi porazdelitvena funkcija za ekstreme. V statistični teoriji so zelo redki avtorji, ki se ukvarjajo z anomalijami dane populacije. Eden od njih je A. F. Jenkinson. Ob rešitvi Fisher-Tippetove enačbe je razvil nekaj splošnih prijemov za praktično aplikacijo v meteorologiji.

Njegovo metodo smo preiskusili za nalive v Sloveniji. Podatki o nalivih so posebno pomembni pri nizkih gradnjah, kot so ceste, melioracije, kanalizacije itd. Rezultati, ki smo jih dobili, pa se ne ujemajo z predpostavkami avtorja metode. Naloga tega prispevka je, da prikažemo razhajanja, ki smo jih ugotovili, in nakažemo možnosti, da se odstopom od domnev teorije izognemo.

JENKINSONOVA REŠITEV FISHER-TIPPETOVE FUNKCIJSKE ENAČBE

Fisher in Tippet sta že leta 1928 pokazala, da lahko dobimo porazdelitveno funkcijo za ekstreme po funkcijski enačbi

$$P^S(x) = P(a_s + b_s), \quad (1)$$

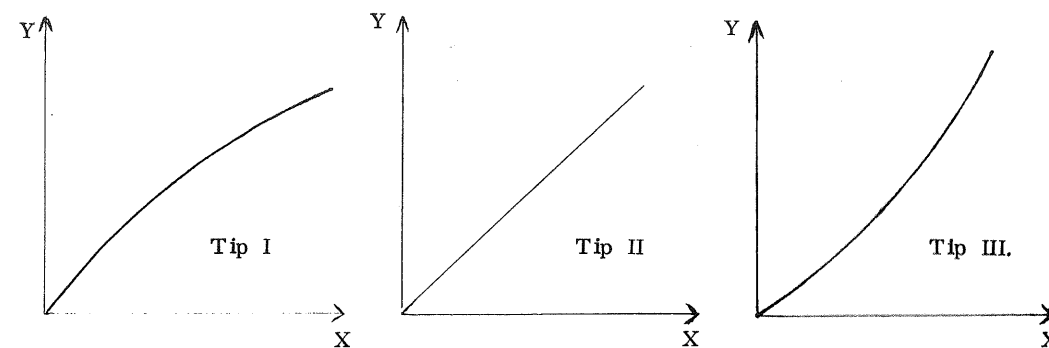
kjer sta a_s in b_s funkciji od s . Jenkinson je v /1/ pokazal, da je krivulja

$$x = a(1 - e^{-ky}), \quad (2)$$

kjer je $ak \geq 0$, splošna rešitev enačbe (1). V naravi najdemo po Jenkinsonu tri tipične oblike krivulje (2):

- | | | |
|------------|---|----------|
| 1. $k < 0$ | - | tip I |
| 2. $k = 0$ | - | tip II |
| 3. $k > 0$ | - | tip III. |

Te oblike prikazuje slika 1.



Slika 1 Tipične oblike krivulj, ki jih najdemo v naravi.

Po avtorju /1/ ustreza oblika po tipu I nalivom in maksimalnim padavinam, oblika po tipu II pritisku, oblika po tipu III pa temperaturam in pretokom. Nas zanima torej tip I, kadar je $k < 0$. Poglejmo, kakšne so mejne vrednosti v tem primeru:

$$\begin{aligned} y &= -\infty, & x &= a, & dy/dx &= \infty; \\ y &= 0, & x &= 0, & dy/dx &= 1/ak; \\ y &= \infty, & x &= \infty, & dy/dx &= 0. \end{aligned}$$

Zaradi narave problema je razumljivo, da lahko izmerimo posamezne nalive, ki so ekstremno visoki, kar se sklada z mejnimi vrednostmi.

Ob rešitvi (2) si oglejmo gostoto porazdelitve za ekstreme. Po /1/ lahko gostoto porazdelitve maksimalnih vrednosti za obdobje T časovnih enot zapišemo kot

$$dP = d(e^{-Te^{-y}}). \quad (3)$$

Pričakovana vrednost v T časovnih enotah je potem

$$E(X_T) = \int_{-\infty}^{\infty} x d(e^{-Te^{-y}}) = \int_{-\infty}^{\infty} a(1 - e^{-ky}) d(e^{-Te^{-y}}). \quad (4)$$

Vidimo, da je pričakovana vrednost $E(X_T)$ bistveno odvisna od predznaka konstante k . Pri negativni konstanti dobimo mnogo večje vrednosti kot pri pozitivni. Ta ugotovitev se sklada tudi z naravo nalivov, namreč da lahko izmerimo zelo ekstremne vrednosti. Pri računanju $E(X_T)$ za nekatere kraje v Sloveniji smo ugotovili, da je konstanta k pozitivna, kljub temu da smo obravnavali nalive. To nas je privedlo do natančnejšega testiranja negativnosti ali pozitivnosti konstante k za čimveč krajev v Sloveniji, za katere razpolagamo z relativno daljšimi nizi opazovanj. Test naj potrdi ali zavrže domnevo, da je konstanta k za nalive negativna.

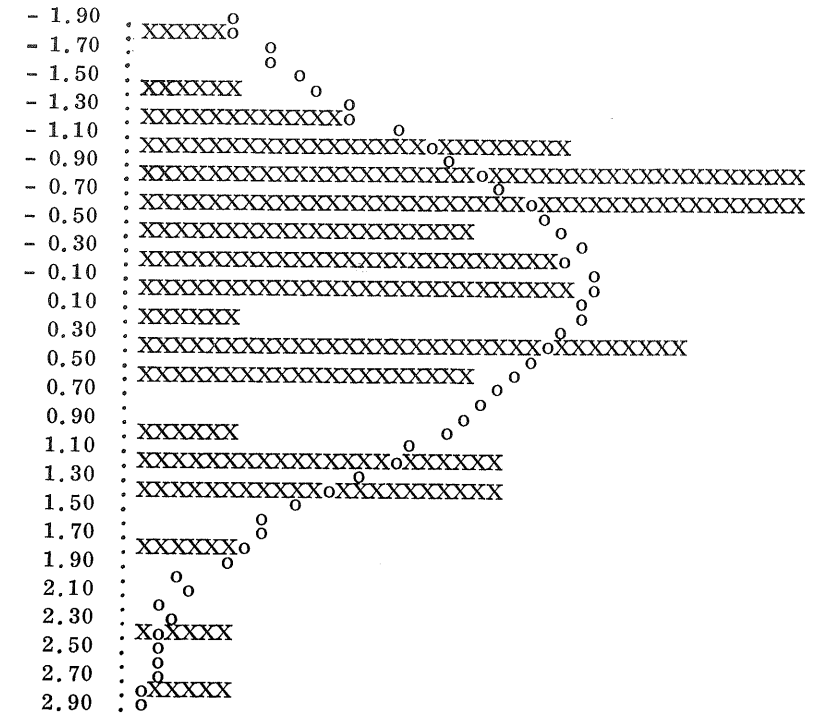
TESTIRANJE NEGATIVNOSTI KONSTANTE k

V tabeli 1 in 1a so izračunane vrednosti konstante k za izbrane kraje v Sloveniji. Za vsak kraj smo izračunali konstanto k za 10 minutni, 1 urni in 6 urni naliv. Vse skupaj smo vzeli kot enotno populacijo, saj nas zanima samo negativnost ali pozitivnost. Prav tako je število let opazovanj v posameznih krajih različno, to pa je spet brez pomena glede na problem, ki smo si ga zastavili. Imamo torej 51 različnih konstant k ; naloga je ugotoviti, ali so praviloma pozitivne ali negativne. Najprej ugotovimo, kakšna je porazdelitev za konstanto k iz tabel 1 in 1a. Poprečna vrednost je $\bar{k} = 0.0053$, vzorčna varianca ali standardna deviacija pa $s^2 = 0.0466$. Slika 2 nam predstavlja to porazdelitev. S χ^2 - kvadrat testom ugotovimo, da je normalna porazdelitev s parametroma \bar{k} in s^2 različna od empirične iz tabel 1 in 1a le z zelo velikim tveganjem. Dobimo $\chi^2 = 9.5$ pri 10 prostostnih stopnjah, kar nam da tveganje približno 0.5.

Iz slike 2 je tudi razvidno, da je empirična porazdelitev asimetrična in pomaknjena v levo. Z ustrezno transformacijo jo je mogoče pomakniti v desno. Izvedimo tako transformacijo:

$$y + a = e^x, \text{ ali} \\ \ln(y + a) = x. \quad (5)$$

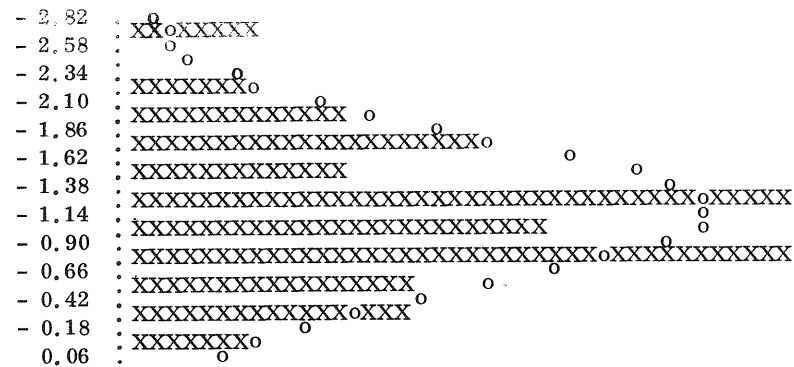
Osnovno porazdelitev pomaknimo za a v desno, da lahko definiramo logaritem. Za a vzemimo minimum osnovne populacije, torej $a = -0.3770$.



Slika 2 Empirična in teoretična porazdelitev osnovne populacije konstante k ; empirično predstavljajo X - i (vsak $X = 0.30\%$), teoretično pa krogci. Podatki so normirani.

Sedaj predpostavljamo, da je porazdelitev, ki smo jo tako dobili, normalna s parametri μ in σ^2 , torej $N(\mu, \sigma^2)$. Naša osnovna porazdelitev bi naj bila torej logaritemsko normalna. To moramo seveda preveriti. Pri tem naletimo na prvo težavo pri izbiri števila celic. Ker pa to ni problem te naloge, navedemo samo tabelo, v kateri so za različno število celic izračunani χ^2 in ustrezajoča tveganja (Tabela 2). Očitno je, da so razlike med tako dobljeno teoretično in empirično porazdelitvijo manjše, kot v primeru brez transformacije. Za nadaljno obravnavo je vseeno, koliko celic vzamemo, saj smo v vseh primerih, razen pri koraku $\Delta x = 0.3$, dobili iste ocene za μ in σ^2 . Izberimo si tako $\Delta x = 0.24$; imamo 11 celic. Porazdelitev v tem primeru predstavlja slika 3. Vidimo, da res izgubimo precej asimetrije, ki smo jo imeli v primeru na sliki 2. Oцени za μ in σ^2 sta v tem primeru: $\hat{\mu} = -1.1572$ in $\hat{\sigma}^2 = 0.3816$. Izračunajmo maksimalno likelihood oceno za ta primer. Iz $/2/$ vzemo, da je ta ocena za logaritemsko normalno porazdelitev naslednja

$$\hat{\mu} = \hat{E}(z) = e^{\bar{x} + s^2/2},$$



Slika 3 Empirična in teoretična porazdelitev logaritemske populacije konstante k; empirično predstavljajo X - i (vsak X = 0.45%), teoretično pa krogi.

kjer je \bar{x} ocena za pričakovano vrednost logaritemsko normalne porazdelitve, s^2 pa ocena za varianco. V našem primeru lahko torej napišemo

$$\hat{V} = \widehat{E(y + a)} = e^{\hat{\mu} + \hat{\sigma}^2/2}. \quad (6)$$

Upoštevač transformacijo (5) dobimo maksimalno likelihood oceno za osnovno porazdelitev

$$\hat{V}_0 = \hat{V} - a. \quad (7)$$

Kakšne vrednosti lahko zavzame tako dobljena ocena? Po [2] je razlika $n \sqrt{V_0} (\hat{V}_0 - V_0)$ porazdeljena asimptotično normalno s porazdelitvijo $N(0, i^{-1})$, kjer je i Fisherjeva mera informacije, definirana na sledeč način:

$$i(V) = -E(\partial^2 \log f(x; V) / \partial V^2). \quad (8)$$

Imamo torej

$$\sqrt{i n} (\hat{V}_0 - V_0) \overset{a}{\sim} N(0, 1). \quad (9)$$

Za naš primer smo izračunali Fisherjevo mero informacije v dodatku. Upoštevač (9) bo interval zaupanja s tveganjem naslednji

$$\hat{V}_0 - z_{\alpha} / \sqrt{i n} \leq \hat{V}_0 \leq \hat{V}_0 + z_{\alpha} / \sqrt{i n}. \quad (10)$$

Izberimo si 5% tveganje in izračunajmo interval zaupanja za naš primer. Dobimo

$$z_{0.05} / \sqrt{i n} = 0.0125$$

in je interval zaupanja

$$-0.0091 \leq \hat{V}_0 \leq 0.0159. \quad (11)$$

Sedaj se vprašajmo, kolikšna je verjetnost, da je \hat{V}_0 negativen. Odgovor na to je tudi odgovor na naš problem. Imamo dva dogodka:

- A_1 je dogodek, da se unija danih podintervalov, ki pokrivajo \hat{V}_0 , ujema s negativnim poltrakom (da mi lahko trdimo, da je prava vrednost \hat{V}_0 negativna);
- A_2 je dogodek, da se unija danih podintervalov, ki pokrivajo \hat{V}_0 , ujema s pozitivnim poltrakom (da mi lahko trdimo, da je prava vrednost \hat{V}_0 nenegativna).

Če upoštevamo interval zaupanja (11), dobimo

$$\begin{aligned} P(A_1) &= 0.36 \\ P(A_2) &= 0.64. \end{aligned} \quad (12)$$

Gre za residualni verjetnosti, ki ju lahko vzamemo za oceno verjetnosti dogodkov A_1 in A_2 . Torej je verjetnost, da je poljuben k iz našega osnovnega vzorca pozitiven, precej večja od verjetnosti, da je negativen.

ZAKLJUČEK

Tako smo vsaj za slovenski prostor ugotovili, da je vprašljiva domneva, da je konstanta k iz enačbe (2) za nalive negativna. To dejstvo seveda precej spremeni uporabo Jenkinsonove teorije. Iz enačbe (4) je jasno, da je pričakovana vrednost $E(X_T)$ močno odvisna od predznaka konstante k . Vsekakor pa metode, opisane v [1] ne moremo uporabiti v primeru, ko je konstanta k pozitivna, pa čeprav gre v bistvu samo za ocene. Postavlja se vprašanje, če nimamo morda prekratkih nizov podatkov. Toda tudi pritrđen odgovor nas ne odvezuje razširitev rešitve (2), saj smo postavljeni velikokrat pred nalogo, napovedati ocene absolutnih ekstremov za daljšo dobo, čeprav imamo na razpolago kratek niz.

ZAHVALA

Zahvaljujem se tov. dipl.ing.mat. Veljku Boletu za izčrpne nasvete pri obravnavi tega problema.

LITERATURA

- /1/ A. F. Jenkinson: The frequency distribution of the annual maximum (or minimum) values of meteorological elements; Quaterly journal of the royal meteorological society, No. 348, London, 1955.
- /2/ Alexander M. Mood and Franklin A. Graybill: Introduction to the theory of statistics; McGraw-Hill book Company, New York, 1963.
- /3/ Vladimir Vranić: Vjerojatnost i statistika; Tehnička knjiga, Zagreb, 1965.
- /4/ Arhiv Meteorološkega zavoda SRS, Ljubljana.

Kraj	Naliv	Število let opazovanj	k	ln (k + 0.3770)
MARIBOR	10 ^m	27	0.084	-0.774
	1 ^h	27	0.111	-0.717
	6 ^h	27	0.251	-0.465
LJUBLJANA	10 ^m	25	-0.059	-1.146
	1 ^h	25	-0.184	-1.645
	6 ^h	25	-0.120	-1.359
ŠMARTNO PRI SLOVENJGRADCU	10 ^m	20	0.000	-0.976
	1 ^h	20	-0.304	-2.617
	6 ^h	20	0.299	-0.392
KOPER	10 ^m	16	0.239	-0.485
	1 ^h	16	0.098	-0.744
	6 ^h	16	0.379	-0.280
BOVEC	10 ^m	14	0.604	-0.019
	1 ^h	14	0.084	-0.774
	6 ^h	14	-0.269	-2.226
ČEPOVAN	10 ^m	15	0.214	-0.526
	1 ^h	15	-0.168	-1.565
	6 ^h	15	-0.234	-1.945
SAVICA	10 ^m	25	0.111	-0.717
	1 ^h	25	0.516	-0.113
	6 ^h	25	-0.136	-1.423
RAKITNA	10 ^m	16	0.070	-0.805
	1 ^h	16	-0.136	-1.423
	6 ^h	16	0.070	-0.805
NOVO MESTO	10 ^m	21	0.111	-0.717
	1 ^h	21	-0.184	-1.645
	6 ^h	21	-0.120	-1.359
MURSKA SOBOTA	10 ^m	16	-0.044	-1.100
	1 ^h	16	-0.014	-1.013
	6 ^h	16	0.043	-0.868
BRNIK	10 ^m	10	-0.089	-1.245
	1 ^h	10	-0.105	-1.302
	6 ^h	10	-0.044	-1.100
ČRNOMELJ	10 ^m	10	-0.201	-1.737
	1 ^h	10	-0.234	-1.945
	6 ^h	10	0.275	-0.428

Tabela 1 Podatki za konstanto k in ln(k + 0.3770) po posameznih postajah.

Kraj	Naliv	Število let opazovanj	k	ln (k + 0.3770)
GOMANCE	10 ^m	13	-0.059	-1.146
	1 ^h	13	-0.014	-1.013
	6 ^h	13	-0.120	-1.359
JAVORJE	10 ^m	11	0.310	-0.375
	1 ^h	11	-0.184	-1.645
	6 ^h	11	-0.168	-1.565
POSTOJNA	10 ^m	11	-0.074	-1.194
	1 ^h	11	-0.120	-1.359
	6 ^h	11	-0.014	-1.013
TEMNICA	10 ^m	13	-0.184	-1.645
	1 ^h	13	-0.269	-2.226
	6 ^h	13	-0.3770	./.
VEDRIJAN	10 ^m	11	-0.234	-1.945
	1 ^h	11	0.322	-0.358
	6 ^h	11	-0.304	-2.617

Tabela 1a Nadaljevanje tabele 1.

n	Δx	P	χ ²
15	0.18	0.97	6.6
13	0.20	0.99	1.5
12	0.22	0.85	7.4
11	0.24	0.93	5.1
10	0.26	0.80	6.2
10	0.28	0.91	4.2
9	0.30	0.96	2.8
8	0.32	0.92	3.0
8	0.34	0.85	4.3
7	0.36	0.97	1.7
7	0.38	0.99	1.1

Tabela 2 χ² in ustrezne verjetnosti za različno število celic logaritemske populacije konstante k.

DODATEK

Po definiciji Fisherjeve mere informacije potrebujemo gostoto verjetnosti za naš primer. Ta je

$$f(y; \hat{\nu}) = e^{-(\log(y+a) - \mu)^2 / 2 \sigma^2} / (y+a)$$

oziroma, ako postavimo substitucijo $u = y + a$, dobimo

$$f(u; \hat{\nu}) = (1/u) e^{-(\log u - \mu)^2 / 2 \sigma^2}$$

Po logaritmiranju dobimo

$$\log f(u; \hat{\nu}) = -\log u - (1/2 \sigma^2) (\log u - \mu)^2$$

Če upoštevamo maksimalno likelihood oceno (6), dobimo

$$\log f(u; \hat{\nu}) = -\log u - (\log u - \mu)^2 / 4 (\log \hat{\nu} - \mu)$$

Po dvakratnem odvajanju na $\hat{\nu}$ dobimo

$$\partial^2 \log f(x; \hat{\nu}) / \partial \hat{\nu}^2 = -M (\log u - \mu)^2$$

kjer je

$$M = -(4 + \sigma^2) / \hat{\nu}^2 \sigma^6$$

Tako imamo

$$i(\hat{\nu}) = -M \int_0^{\infty} (1/u) (\log u - \mu)^2 e^{-(\log u - \mu)^2 / 2 \sigma^2} \sigma^2 du$$

Po substituciji $\log u = v$ dobimo

$$i(\hat{\nu}) = -M \int_{-\infty}^{\infty} (v - \mu)^2 e^{-(v - \mu)^2 / 2 \sigma^2} dv = -M \sigma^2$$

Tako je torej

$$i(\hat{\nu}) = (4 + \sigma^2) / \hat{\nu}^2 \sigma^4$$

V našem primeru, ko je $\hat{\mu} = -1.1572$ in $\hat{\sigma}^2 = 0.3816$, je

$$i(\hat{\nu}_0) = 209.0$$