SUR L'AJUSTEMENT DE LA LOI DE REPARTITION DE FISHER-TIPPETT DU TYPE I AU MOYEN DES ESTIMATEURS DE KIMBALL

R. SNEYERS et J. VAN ISACKER

Institut Royal Météorologique de Belgique

1. Introduction.

Il est bien connu que pour des variables aléatoires dont le domaine de définition s'étend de $-\infty$ à $+\infty$, sous des conditions bien déterminées, la théorie statistique des extrêmes conduit à la considération de la première asymptote de Fisher-Tippett [1], [2].

En particulier, si x est la variable étudiée et si y désigne la variable aléatoire de la fonction de répartition de la première asymptote de Fisher-Tippett donnée par la relation :

$$F(y) = \exp(-e^{-y}) \tag{1}$$

la loi des extrêmes de la variable x peut se définir à l'aide de la relation linéaire :

$$x = \mu + \sigma y \tag{2}$$

dans laquelle le paramètre de position μ et le paramètre d'échelle σ dépendent uniquement de la répartition initiale de x et de l'effectif des échantillons d'où sont issues les valeurs extrêmes.

En l'absence d'estimateurs conjointements exhaustifs pour μ et σ, divers estimateurs ont été successivement proposés par Gumbel [3], Lieblein [4], Kimball [5], Blom [6], Weiss [7] et Downton [8], et leur variété pose un problème de choix.

Les estimateurs de Gumbel obtenus par la méthode des moments ne semblent pas devoir être retenus à cause de l'efficacité médiocre et du biais mis en évidence par Lieblein.

Tous les autres estimateurs, qui sont des fonctions linéaires des observations ordonnées, ont, par contre, été obtenus par des méthodes optimales ou presque optimales. Parmi ceux-ci, les estimateurs de Downton présentent l'avantage que la matrice des covariances des estimateurs peut être calculée quel que soit l'effectif de l'échantillon. Comme, de plus, les estimateurs à coefficients quadratiques de cette dernière catégorie possèdent des efficacités asymptotiques proches du maximum, on peut considérer que ces derniers apportent une solution satisfaisante au problème.

Les estimateurs de Kimball ont cependant retenu notre attention en raison de la simplicité relative de leur calcul et pour autant que leur efficacité croisse suffisamment vite avec l'effectif de l'échantillon, ces estimateurs pourraient constituer également une solution acceptable.

Le but de la présente note a été de montrer qu'il en est réellement ainsi en même temps que de fournir une expression nouvelle du facteur de Kimball pour la correction du biais.

2. **Les estimateurs de Kimball.**

Si x_i désigne une série aléatoire et simple de valeurs extrêmes ordonnées $x_1 < x_2 < \ldots < x_n$, et si F_i est la valeur de $F(y)$ donnée par les relations 1.(1) et 1.(2) avec $x = x_i$, la combinaison des équations du maximum de vraisemblance conduit à l'équation :

$$\hat{\sigma} = \sum x_i \, (1 + \log F_i)/n \tag{1}$$

De plus, si l'on note que, pour la statistique d'ordre x_i, on a :

$$E(\log F_i) = \overline{\log F_i} = -\sum_{j=i}^{n} \frac{1}{j} \tag{2}$$

l'estimateur de Kimball, corrigé pour le biais, s'écrit :

$$\hat{\sigma} = \sum x_i \, (1 + \overline{\log F_i})/(n \, s_n) \tag{3}$$

où le facteur introduit pour le biais $1/s_n$ se définit par la récurrence :

$$s_m = s_{m-1} + [\sum_{1}^{m-1} (-1)^{i+1} \, \Delta^i \, \log 1]/m\,(m-1) \tag{4}$$

pour $m = 2, 3, \ldots, n$ et avec $s_1 = 0$.

En adjoignant à l'estimateur (3) l'estimateur (également sans biais) de μ :

$$\hat{\mu} = x - \gamma \hat{\sigma} \tag{5}$$

où $\overline{x} = \sum x_i/n$ et $\gamma = 0{,}5772\ldots$ est le nombre d'Euler, on obtient une solution complète du problème d'estimation.

3. **Calcul du facteur de Kimball pour le biais.**

De la récurrence 2.(4), on tire :

$$s_n = \sum_{1}^{n} (-1)^{i+1} \left(\frac{1}{i} - \frac{1}{n}\right) \Delta^i \log 1 \tag{1}$$

ce qui peut s'écrire :

$$s_n = A_n - B_n/n \tag{2}$$

si l'on pose d'une manière récurrente :

$$A_n = A_{n-1} - \frac{(-1)^n}{n} \Delta^n \log 1, \qquad \text{avec } A_1 = \log 2$$

$$B_n = B_{n-1} - (-1)^n \Delta^n \log 1, \qquad \text{avec } B_1 = \log 2 \tag{3}$$

Par ailleurs, pour éviter la complexité du calcul de $\Delta^n \log 1$ lorsqu'on se base sur la définition :

$$\Delta^n \log x = \Delta^{n-1} \log (x+1) - \Delta^{n-1} \log (x) \tag{4}$$

du fait qu'il doit s'accomplir en multiple précision, c'est-à-dire en utilisant plusieurs dizaines de chiffres significatifs, on a préféré tenir compte de la transformation de Laplace [9] :

$$\Delta \log x = \log (x+1) - \log x = \int_0^\infty \frac{-(e^{-s}-1)\ e^{-xs}}{s} ds \tag{5}$$

d'où l'on tire :

$$\Delta^n \log x = \int_0^\infty \frac{-(e^{-s}-1)^n\ e^{-xs}}{s} ds \tag{6}$$

et, par conséquent :

$$\Delta^n \log 1 = \int_0^\infty \frac{-(e^{-s}-1)^n\ e^{-s}}{s} ds \tag{7}$$

En posant ensuite :

$$t = 1 - e^{-s} \tag{8}$$

il vient finalement :

$$\Delta^n \log 1 = (-1)^n \int_0^1 \frac{t^n\, dt}{\log (1-t)} \tag{9}$$

ce qui ramène le calcul à celui d'une intégrale entre des limites finies.

Les valeurs de $b_n = 1/s_n$ obtenues de cette manière sont indiquées au tableau 1 ; elles sont très probablement correctes avec cinq décimales. De plus, si l'on procède à une représentation graphique de $\log (b_n - 1)$

en fonction de $\log(n-1)$, on constate un comportement asymptotique linéaire très rapide.

En réalité, pour $n > 11$, la relation :

$$\log(b_n - 1) = -0{,}97562 \log(n-1) + 1{,}043532 - 1{,}950309/\log(n-1) + 3{,}574231/\log^2(n-1) - 2{,}383483/\log^3(n-1) \quad (10)$$

fournit des valeurs approchées de b_n avec une erreur moindre que 10^{-5}.

4. Efficacités empiriques des estimateurs de Kimball.

Pour chaque effectif n = 2, 3, 4, 5, 6, 10, 20, 30, 50, 100, 200, 300, 400 et 500, on a constitué cinq mille échantillons aléatoires de valeurs de F et les estimations des paramètres μ et σ ont été calculées pour chaque échantillon à partir des valeurs $x = -\log(-\log F)$ au moyen des estimateurs 2.(3) et 2.(5).

Sachant que dans ce cas $E(\mu) = \gamma$ et $E(\sigma) = 1$, la matrice des covariances de ces estimateurs a été évaluée au moyen des relations :

$$\text{var}\ \hat{\mu} = \sum (\hat{\mu} - \gamma)^2/5000, \qquad \text{var}\ \hat{\sigma} = \sum (\hat{\sigma} - 1)^2/5000$$

$$\text{cov}\ (\hat{\mu}, \hat{\sigma}) = \sum (\hat{\mu} - \gamma)(\hat{\sigma} - 1)/5000 \quad (1)$$

L'efficacité des estimateurs considérés a ensuite été déterminée au moyen des rapports :

$$e(\hat{\mu}) = \frac{\text{var}_0\ \mu}{\text{var}\ \hat{\mu}}, \qquad e(\hat{\sigma}) = \frac{\text{var}_0\ \sigma}{\text{var}\ \hat{\sigma}}, \qquad e(\hat{\mu}, \hat{\sigma}) = \frac{\text{cov}_0\ (\mu, \sigma)}{\text{cov}\ (\hat{\mu}, \hat{\sigma})} \quad (2)$$

sachant qu'avec $\sigma = 1$ les limites inférieures de Cramer-Rao donnent :

$$\text{var}_0\ \mu = [1 + 6(1-\gamma)^2/\pi^2]/n = 1{,}10866/n$$

$$\text{var}_0\ \sigma = (6/\pi^2)/n = 0{,}60793/n$$

$$\text{cov}_0\ (\mu, \sigma) = [6(1-\gamma)/\pi^2]/n = 0{,}25702/n \quad (3)$$

De plus, en ce qui concerne la précision des efficacités ainsi obtenues, on notera qu'en vertu du théorème central limite les estimateurs $\hat{\mu}$ et $\hat{\sigma}$ possèdent une distribution conjointe approximativement normale. On en déduit que l'erreur type associée aux efficacités précédentes est environ de $\sqrt{2/5000}$, soit 0,02, pour $e(\hat{\mu})$ et pour $e(\hat{\sigma})$.

Pour $e(\hat{\mu}, \hat{\sigma})$, on note que, pour N échantillons, on a :

$$\frac{\text{var cov}(\hat{\mu}, \hat{\sigma})}{\text{cov}_0{}^2(\hat{\mu}, \hat{\sigma})} = \frac{1}{N}\left(3 + \frac{1 - 2\rho^2}{\rho^2}\right) \tag{4}$$

d'où il résulte qu'avec $\rho = 0{,}31307$ l'erreur-type est environ de $\sqrt{11{,}2/5000}$, soit 0,047.

Les efficacités obtenues dans ces conditions ont été indiquées au tableau 2. En outre, pour permettre la comparaison, on a fait figurer également les efficacités exactes calculées par Downton.

Il apparaît ainsi que, pour les effectifs n considérés, les efficacités $e(\hat{\mu})$ et $e(\hat{\sigma})$ croissent régulièrement avec n. De plus, si pour $n = 10$ on constate que l'efficacité de l'estimateur $\hat{\mu}$ est pratiquement optimale, les valeurs obtenues pour le rapport $e(\hat{\sigma})$ montrent que l'efficacité de l'estimateur $\hat{\sigma}$ est très comparable à celle de l'estimateur à coefficients quadratiques de Downton, dont l'efficacité asymptotique est 93,6 %.

Enfin, on notera encore que le rapport $e(\hat{\mu}, \hat{\sigma})$ converge aussi rapidement vers 100 %.

5. **Conclusion.**

Le calcul des efficacités empiriques des estimateurs de Kimball révèle que ces estimateurs constituent une bonne solution du problème de l'estimation des paramètres de position et d'échelle de la première asymptote de Fisher-Tippett. En outre, les efficacités élevées obtenues permettent de considérer cette solution comme une solution de rechange à celle proposée par Downton.

Du point de vue pratique, il en résulte que lors de l'évaluation de l'erreur d'estimation associée à la valeur du fractile d'ordre P donnée par la relation :

$$\hat{x}_P = \hat{\mu} + \hat{\sigma}\, y_P \tag{1}$$

où y_P est le fractile correspondant de la loi réduite 1.(1), et dont la variance s'obtient au moyen de la relation :

$$\text{var}\,\hat{x}_P = \text{var}\,\hat{\mu} + 2\,y_P\,\text{cov}(\hat{\mu}, \hat{\sigma}) + y^2_P\,\text{var}\,\hat{\sigma} \tag{2}$$

on pourra adopter pour var $\hat{\mu}$, cov $(\hat{\mu}, \hat{\sigma})$ et var $\hat{\sigma}$ les approximations fournies par les limites inférieures de Cramer-Rao 4.(3), c'est-à-dire adopter :

$$\text{var}\,\hat{x}_P = [\text{var}_0\,\mu + 2\,y_P\,\text{cov}_0(\mu, \sigma) + y^2_P\,\text{var}_0\,\sigma]\,\hat{\sigma}^2 \tag{3}$$

TABLEAU 1. — *Valeurs du facteur de Kimball* : $b_n = 1/s_n$.

n	b_n	n	b_n	n	b_n	n	b_n
2	2.88539	39	1.05954	75	1.03188	111	1.02196
3	1.96061	40	1.05811	76	1.03148	112	1.02177
4	1.65033	41	1.05675	77	1.03109	113	1.02159
5	1.49408	42	1.05545	78	1.03071	114	1.02141
6	1.39966	43	1.05421	79	1.03034	115	1.02123
7	1.33628	44	1.05303	80	1.02998	116	1.02106
8	1.29072	45	1.05190	81	1.02963	117	1.02089
9	1.25634	46	1.05082	82	1.02928	118	1.02072
10	1.22944	47	1.04979	83	1.02895	119	1.02056
11	1.20781	48	1.04879	84	1.02862	120	1.02039
12	1.19002	49	1.04784	85	1.02830	121	1.02023
13	1.17513	50	1.04693	86	1.02799	122	1.02007
14	1.16247	51	1.04605	87	1.02768	123	1.01992
15	1.15157	52	1.04520	88	1.02738	124	1.01977
16	1.14208	53	1.04439	89	1.02709	125	1.01962
17	1.13375	54	1.04360	90	1.02680	126	1.01947
18	1.12637	55	1.04284	91	1.02652	127	1.01932
19	1.11979	56	1.04211	92	1.02625	128	1.01918
20	1.11388	57	1.04141	93	1.02598	129	1.01904
21	1.10854	58	1.04072	94	1.02572	130	1.01890
22	1.10370	59	1.04007	95	1.02546	131	1.01876
23	1.09928	60	1.03943	96	1.02521	132	1.01863
24	1.09524	61	1.03881	97	1.02496	133	1.01849
25	1.09152	62	1.03822	98	1.02472	134	1.01836
26	1.08809	63	1.03764	99	1.02448	135	1.01823
27	1.08492	64	1.03708	100	1.02425	136	1.01811
28	1.08197	65	1.03653	101	1.02402	137	1.01798
29	1.07923	66	1.03601	102	1.02380	138	1.01786
30	1.07667	67	1.03549	103	1.02358	139	1.01774
31	1.07428	68	1.03500	104	1.02336	140	1.01762
32	1.07204	69	1.03451	105	1.02315	141	1.01750
33	1.06993	70	1.03404	106	1.02294	142	1.01738
34	1.06794	71	1.03359	107	1.02274	143	1.01726
35	1.06607	72	1.03314	108	1.02254	144	1.01715
36	1.06430	73	1.03271	109	1.02234	145	1.01704
37	1.06263	74	1.03229	110	1.02215	146	1.01693
38	1.06105						

n	b_n	n	b_n	n	b_n	n	b_n
147	1.01682	195	1.01286	290	1.00882	400	1.00649
148	1.01671	200	1.01255	300	1.00854	410	1.00634
149	1.01660	205	1.01226	310	1.00827	420	1.00620
150	1.01650	210	1.01198	320	1.00803	430	1.00606
155	1.01599	215	1.01172	330	1.00780	440	1.00593
160	1.01552	220	1.01146	340	1.00758	450	1.00580
165	1.01507	230	1.01099	350	1.00737	460	1.00568
170	1.01465	240	1.01055	360	1.00718	470	1.00557
175	1.01425	250	1.01015	370	1.00699	480	1.00546
180	1.01387	260	1.00978	380	1.00682	490	1.00535
185	1.01352	270	1.00944	390	1.00665	500	1.00525
190	1.01318	280	1.00912				

TABLEAU 2.

Efficacités empiriques conjointes des estimateurs de Kimball (en pourcent)

(Entre parenthèses les efficacités exactes calculées par Downton).

n	$e(\hat{\mu})$	$e(\hat{\sigma})$	$e(\hat{\mu}, \hat{\sigma})$
2	81,4 (84,1)	40,8 (42,7)	— 272,3
3	93,4 (91,7)	57,0 (57,3)	275,6
4	95,1 (94,5)	66,9 (65,0)	183,9
5	94,9 (95,8)	72,6 (69,9)	171,6
6	93,9 (96,6)	73,1 (73,3)	126,8
10	101,6	82,6	113,7
20	98,9	87,1	99,8
30	102,5	89,5	107,3
50	103,5	89,8	95,5
100	97,3	89,3	95,8
200	98,8	92,1	91,3
300	100,6	92,9	98,7
400	100,1	94,1	98,5
500	99,9	95,9	103,3

BIBLIOGRAPHIE

[1] FISHER R.A. et TIPPETT L.H.C. : Limiting form of the frequency distributions of the largest or smallest number of a sample. Proc. Cambridge Phil. Soc., Vol. 24, 1928, pp. 80-190.

[2] GUMBEL E.J. : Statistics of Extremes. Columbia University Press, New York, 1958.

[3] GUMBEL E.J. : Statistical Theory of Extreme Values and Some Practical Applications, National Bureau of Standards, Applied Mathematics Series, 33, 1954.

[4] LIEBLEIN J. : A New Method of Analyzing Extreme Value Data, Washington D.C. National Advisory Committee for Aeronautics, Tech. Note 3053, 1954.

[5] KIMBALL B.F. : The bias in certain estimates of the parameters of the extreme-value distribution, Ann. Math. Stat., Vol. 27, 1956, p. 758.

[6] BLOM G. : Statistical Estimates and Transformed Beta-variables, Wiley, New York, 1958.

[7] WEISS L. : On the estimation of scale parameters, Naval Res. Logist. Quart., Vol. 8, 1961, pp.245-256.

[8] DOWTON F. : Linear estimates of parameters in the extreme value distribution, Technometrics, Vol. 8, n° 1, 1966, pp. 3-17.

[9] Handbook of Mathematical Functions, National Bureau of Standards, Applied Mathematics Series, 55, 1964.