Revue de statistique appliquée

G. Hisleur

# Une estimation optimale des paramètres d'une loi normale

*Revue de statistique appliquée*, tome 15, n$^{o}$ 3 (1967), p. 43-50

<http://www.numdam.org/item?id=RSA_1967__15_3_43_0>

# UNE ESTIMATION OPTIMALE DES PARAMÈTRES D'UNE LOI NORMALE

G. HISLEUR

Laboratoire de Mathématiques Appliquées de Grenoble

## 1 - INTRODUCTION

On détermine le domaine d'acceptation de surface minimum pour les deux paramètres m et $\sigma$ d'un échantillon gaussien formé de n observations indépendantes. Le calcul pratique de ce domaine exige celui d'un seul coefficient qui est tabulé pour trois valeurs usuelles du niveau de signification ; le résultat est illustré par des graphiques qui mettent en évidence la différence entre le domaine ainsi obtenu et les intervalles de confiance usuels.

## 2 - DOMAINES DE SURFACE MINIMUM

2.1 - Théorème.- "*Soit* $f_n(z, t)$ *la fonction des deux réels* z *et* t *définie pour le paramètre* n , *entier, par :*

$$f_n(z, t) = \frac{n^{\frac{n}{2}}}{\sqrt{\pi} \cdot 2^{\frac{n}{2}-1} \Gamma\left(\frac{n-1}{2}\right)} \frac{1}{z^{n+1}} e^{-\frac{n}{2}\frac{1+t^2}{z^2}} \; ; \quad z \geqslant 0$$

*et soit* $F_n(\lambda)$ *la fonction du réel* $\lambda$ :

$$F_n(\lambda) = \iint\limits_{z \geqslant 0 \; ; \; f_n(z, t) \geqslant \lambda} f_n(z, t) \, dz \, dt$$

*alors si* $\lambda_{n,\alpha}$ *est la solution en* $\lambda$ *de l'équation :*

$$F_n(\lambda) = 1 - \alpha$$

*le domaine* $D_{n,\alpha}$ *de* $R^2$ *défini par :*

$$(z, t) \in D_{n,\alpha} \Longleftrightarrow f_n(z, t) \geqslant \lambda_{n,\alpha}$$

*est le domaine de surface minimum par rapport aux domaines de seuil* $\alpha$".

2.2 - Démonstration.

Soient $X_1, X_2, \ldots, X_n$ n variables aléatoires indépendantes, ayant pour loi commune la loi normale de moyenne m et de variance $\sigma^2$, et soit :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$S = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

d'après le théorème de Fisher, nous savons que les deux variables aléatoires :

$$U = \sqrt{n} \frac{\bar{X} - m}{\sigma}$$

$$V = \frac{n S}{\sigma^2}$$

sont indépendantes et suivent respectivement la loi normale centrée réduite et celle du $\chi^2$ à $n - 1$ degrés de liberté. Si nous exprimons $m$ et $\sigma$ en fonction de $U$ et $V$ il vient :

$$m = \bar{X} - \sqrt{S} \frac{U}{\sqrt{V}}$$

$$\sigma = S \sqrt{\frac{n}{V}}$$

Soient alors les nouvelles variables aléatoires :

$$Z = \sqrt{\frac{n}{V}}$$

$$T = -\frac{U}{\sqrt{V}} ;$$

si $C$ est un domaine de $R^2$ tel que :

$$\text{Prob } \{(Z , T) \in C\} = 1 - \alpha$$

on a :

S. Surface de C = Surface du domaine d'acceptation en $m$ et $\sigma$.

D'après le lemme de Neyman et Pearson, le domaine $C^*$ optimum sera défini par :

$$(z , t) \in C^* \Longleftrightarrow f_n(z , t) \geqslant \lambda_{n,\alpha}$$

où $f_n(z , t)$ est la densité de probabilité du couple $(Z , T)$. Or, un calcul élémentaire nous montre que :

$$f_n(z , t) = \frac{n^{\frac{n}{2}}}{\sqrt{\pi} \cdot 2^{\frac{n}{2}-1} \cdot \Gamma\left(\frac{n-1}{2}\right)} \frac{1}{z^{n+1}} e^{-\frac{n}{2} \cdot \frac{1+t^2}{z^2}} \quad ; \quad z \geqslant 0$$

et ainsi si on a calculé la correspondance :

$$(\alpha , n) \longrightarrow \lambda_{n,\alpha} : \iint f_n(z , t)\, dz\, dt = 1 - \alpha$$

$$z \geqslant 0 \; ; \; f_n(z , t) \geqslant \lambda_{n,\alpha}$$

on pourra tracer les domaines $C^*$ comme fonction de n et $\alpha$ sous $D_{n,\alpha}$.

## 3 - COMPARAISON AVEC LA METHODE USUELLE

En pratique on se contente du domaine non optimal suivant : la région critique du test associé à notre problème d'estimation est le complémentaire du rectangle :

$$N^{-1}\left(\frac{\varepsilon}{2}\right) \leqslant u \leqslant -N^{-1}\left(\frac{\varepsilon}{2}\right)$$

$$Q_{n-1}^{-1}\left(\frac{\delta}{2}\right) \leqslant v \leqslant Q_{n-1}^{-1}\left(1-\frac{\delta}{2}\right)$$

où les deux nombres $\varepsilon$ et $\delta$ sont tels que :

$$(1-\varepsilon)(1-\delta) = 1-\alpha\ ,$$

et en ce qui concerne le domaine d'acceptation en m et $\sigma$ cela donne le trapèze :

$$\sqrt{s}\ \sqrt{\frac{n}{Q_{n-1}^{-1}\left(1-\frac{\delta}{2}\right)}} \leqslant \sigma \leqslant \sqrt{s}\ \sqrt{\frac{n}{Q_{n-1}^{-1}\left(\frac{\delta}{2}\right)}}$$

$$\bar{x} + \frac{\sigma}{\sqrt{n}} N^{-1}\left(\frac{\varepsilon}{2}\right) \leqslant m \leqslant \bar{x} - \frac{\sigma}{\sqrt{n}} N^{-1}\left(\frac{\varepsilon}{2}\right)$$

s et $\bar{x}$ représentant la moyenne et la variance empiriques des observations.

On retrouvera sur le graphique (a) pour $n = 15$ et $\alpha = 1\ \%$ le domaine limité par la courbe $f_{20}(z\ ,\ t) \geqslant \lambda_{20,0.05}$ et à titre comparatif le trapèze obtenu par la méthode usuelle.

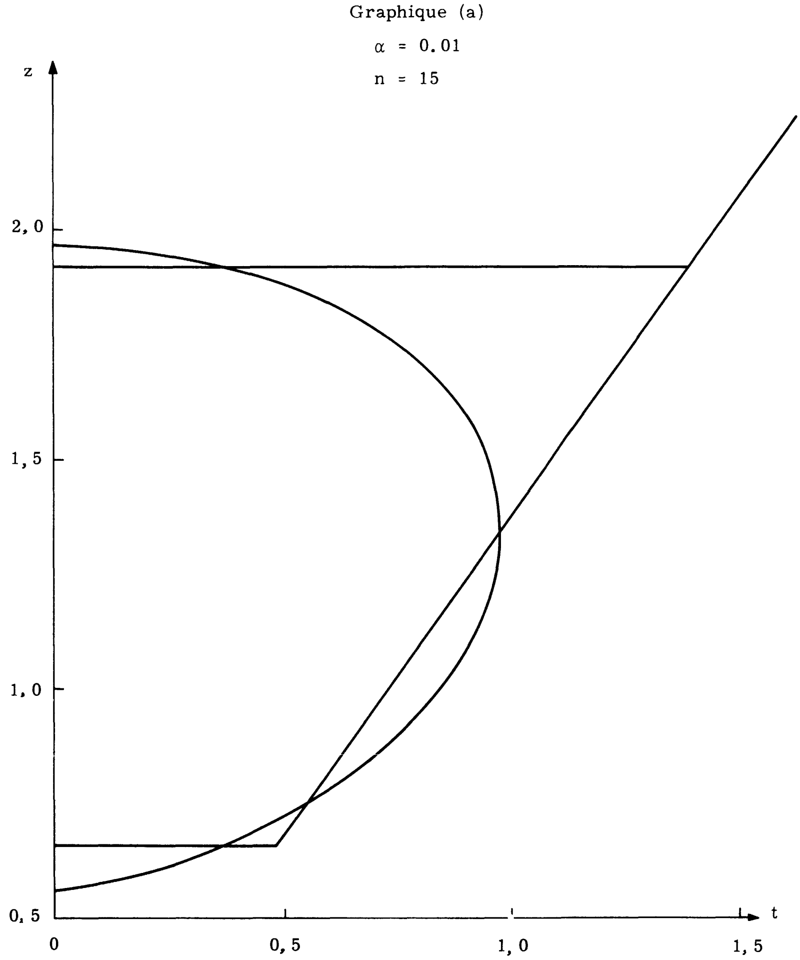
## 4 - COMPARAISON AVEC LES TABLES DE NEYMAN ET PEARSON

Nous pouvons également prendre comme test associé celui qui consiste à vérifier si un échantillon de taille n a été tiré d'une population normale de moyenne m et de variance $\sigma^2$. La méthode utilisée par Neyman et Pearson [1] consiste alors à choisir des valeurs représentatives des deux paramètres, soient $\bar{x}$ et s, et à tester l'hypothèse simple $(m\ ,\ \sigma^2)$ contre l'hypothèse simple $(\bar{x}\ ,\ s)$. On est ainsi ramené à la résolution en $\mu$ de l'équation :

$$\iint f_n(z\ ,\ t)\ dz\ dt = 1-\alpha$$

$$z \geqslant 0\ ;\ \frac{1}{z^n}\ e^{-\frac{n}{2}\frac{1+t^2}{z^2}} \geqslant \mu$$

C'est l'intégrale que Neyman et Perason ont tabulée en fonction du paramètre $\mu$. Sur le graphique (b) nous avons représenté pour $n = 20$ et




$\alpha$ = 5 % le domaine d'acceptation correspondant, ainsi que le domaine de surface minimum.

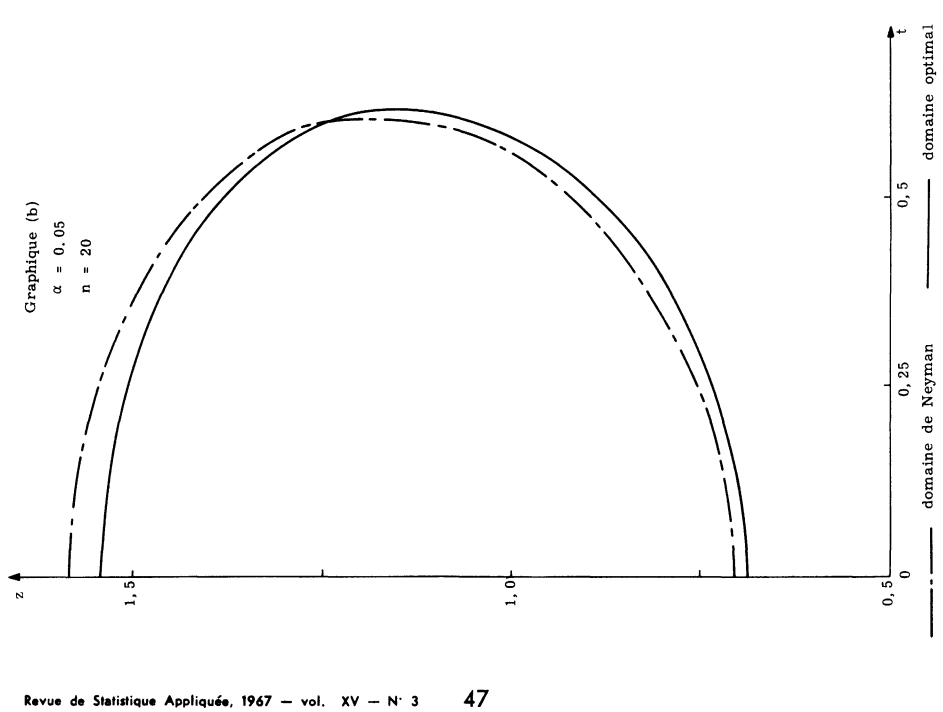
## 5 - UTILISATION POUR L'ESTIMATION DE $\sigma$

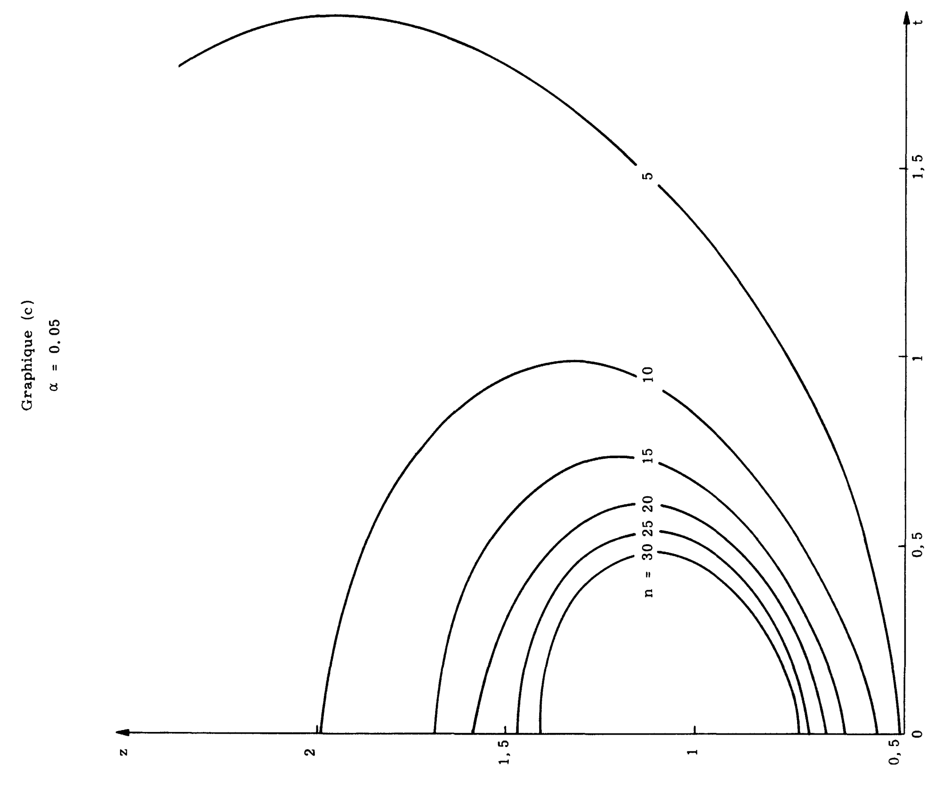
Nous pouvons remarquer que les domaines $D_{n,\alpha}$ sont convexes, symétriques par rapport à l'axe des z, et sont de plus limités par la courbe :

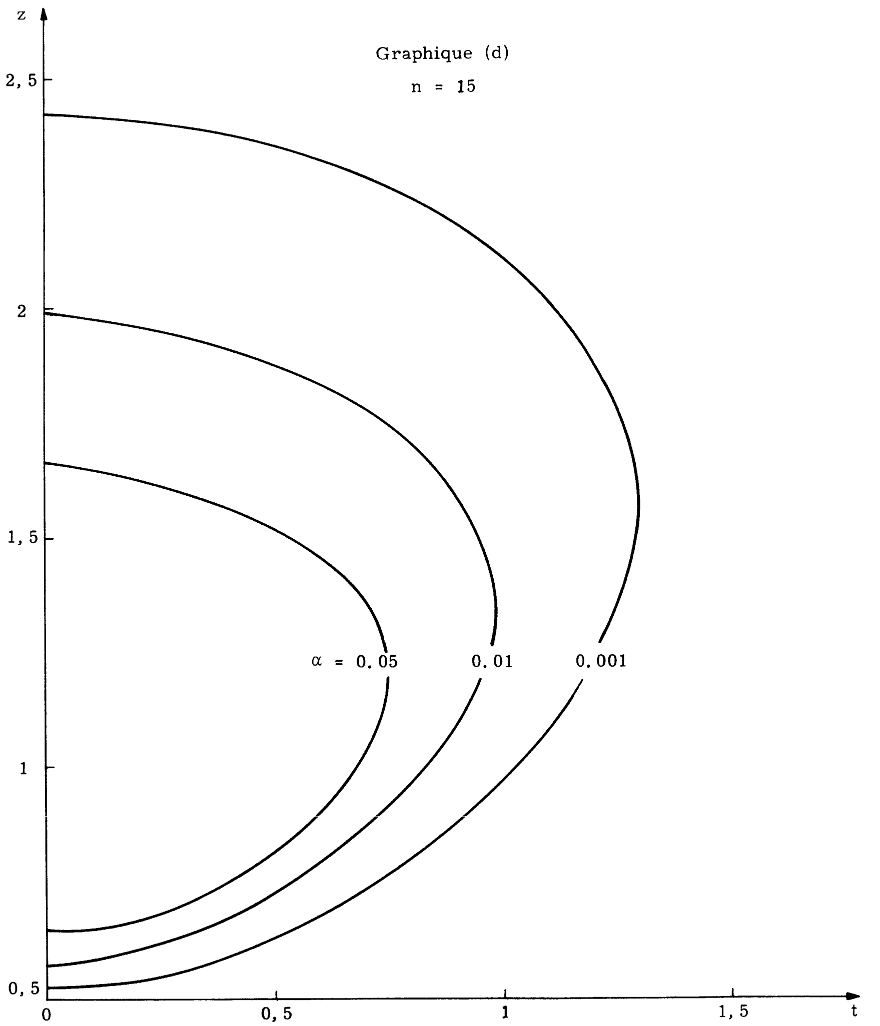
$$1 + t^2 = 2\,\frac{n+1}{n}\,z^2\,[A_{n,\alpha} - \mathrm{Log}\ z] \quad ; \quad z \geqslant 0$$

où le coefficient $A_{n,\alpha}$ est simplement déduit de $\lambda_{n,\alpha}$ et n. Z est compris entre les deux valeurs $Z_{min}$ et $Z_{max}$.

On peut penser à utiliser $Z_{min}$ et $Z_{max}$ comme intervalle de confiance pour $\sigma^2$.


Graphique (b)
α = 0.05
n = 20
z
1,5
1,0
0,5
0
0,25
0,5
t
domaine de Neyman
domaine optimal


Graphique (c)
α = 0.05
z
2
1,5
1
0,5
0
0,5
1
1,5
t
5
10
15
n = 30 25 20




Il ne sera pas, bien entendu de surface minimum mais les résultats seront meilleurs que ceux obtenus en prenant :

$$\frac{n.s}{Q_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right)} \leqslant \sigma^2 \leqslant \frac{n.s}{Q_{n-1}^{-1}\left(\frac{\alpha}{2}\right)}$$

Ainsi pour $n = 7$ et $\alpha = 0.013$ nous avons obtenu pour :

a) l'intervalle de confiance de surface minimum

$$(0.27659) . s \leqslant \sigma^2 \leqslant (5.59552) . s$$

b) l'intervalle de confiance établi à l'aide de $Z_{min}$ et $Z_{max}$ :

$$(0.27766) . s \leqslant \sigma^2 \leqslant (5.59559) . s$$

c) l'intervalle de confiance approximatif :

$$(0.38212) . s \leqslant \sigma^2 \leqslant (10.01723) . s$$

Tables.

Nous avons tabulé le coefficient $A_{n,\alpha}$ pour $n = 5(1)\ 30$ et $\alpha = 0.1\ \%$, $1\ \%$, $5\ \%$. De plus, pour chaque couple $(n, \alpha)$ nous avons indiqué les deux valeurs $Z_{min}$ et $Z_{max}$.

On trouvera représenté sur le graphique (c) pour $\alpha = 5\ \%$ les courbes correspondant aux valeurs $n = 5(5)\ 30$ et sur le graphique (d) les courbes correspondant aux valeurs $\alpha = 0.1\ \%$, $1\ \%$, $5\ \%$, $n$ étant égal à 15.

En ce qui concerne la table les calculs ont été réalisés sur calculateur et les valeurs numériques ne sont fournies qu'à titre d'illustration de la procédure établie, qui permet, en fait, de calculer $A_{n,\alpha}$ pour tout couple $(n, \alpha)$.

## BIBLIOGRAPHIE

[1] NEYMAN J. and PEARSON E.S. 1928 a - Biometrika 20 A, 175-240. Cet ouvrage étant trop ancien nous n'avons pas pu le trouver. Ses références ont été fournies par : Greenwood, J.A. and Hartley, H.O. 1962, *Guide to Tables in Mathematical Statistics*, Princeton University Press.

[2] BARRA J.R. 1964 - *Cours de Statistique Mathématique*, Faculté de des Sciences de Grenoble.

| N | Alpha = 0.001 | | | Alpha = 0.01 | | | Alpha = 0.05 | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | ZMIN | ZMAX | A | ZMIN | ZMAX | A | ZMIN | ZMAX |
| 4 | | | | 1.9391 | 0.3692 | 6.8942 | 1.4012 | 0.4199 | 3.9576 |
| 5 | | | | 1.5724 | 0.4116 | 4.7292 | 1.1686 | 0.4639 | 3.0790 |
| 6 | 1.8100 | 0.3958 | 6.0390 | 1.3541 | 0.4451 | 3.7574 | 1.0305 | 0.4981 | 2.6345 |
| 7 | 1.5966 | 0.4218 | 4.8450 | 1.2090 | 0.4727 | 3.2109 | 0.9392 | 0.5260 | 2.3655 |
| 8 | 1.4346 | 0.4451 | 4.0875 | 1.1057 | 0.4960 | 2.8617 | 0.8744 | 0.5492 | 2.1841 |
| 9 | 1.3176 | 0.4647 | 3.6077 | 1.0289 | 0.5160 | 2.6205 | 0.8262 | 0.5690 | 2.0533 |
| 10 | 1.2254 | 0.4823 | 3.2631 | 0.9690 | 0.5334 | 2.4418 | 0.7888 | 0.5862 | 1.9536 |
| 11 | 1.1537 | 0.4975 | 3.0138 | 0.9214 | 0.5489 | 2.3051 | 0.7596 | 0.6011 | 1.8765 |
| 12 | 1.0921 | 0.5119 | 2.8114 | 0.8824 | 0.5628 | 2.1961 | 0.7534 | 0.6145 | 1.8130 |
| 13 | 1.0427 | 0.5245 | 2.6561 | 0.8501 | 0.5753 | 2.1075 | 0.7153 | 0.6265 | 1.7601 |
| 14 | 1.0005 | 0.5361 | 2.5281 | 0.8227 | 0.5867 | 2.0336 | 0.6980 | 0.6376 | 1.7148 |
| 15 | 0.9642 | 0.5468 | 2.4211 | 0.7992 | 0.5971 | 1.9709 | 0.6835 | 0.6475 | 1.6765 |
| 16 | 0.9332 | 0.5566 | 2.3318 | 0.7791 | 0.6066 | 1.9177 | 0.6710 | 0.6566 | 1.6433 |
| 17 | 0.9062 | 0.5657 | 2.2555 | 0.7614 | 0.6155 | 1.8710 | 0.6601 | 0.6649 | 1.6142 |
| 18 | 0.8816 | 0.5743 | 2.1870 | 0.7458 | 0.6237 | 1.8301 | 0.6505 | 0.6726 | 1.5884 |
| 19 | 0.8603 | 0.5823 | 2.1286 | 0.7320 | 0.6313 | 0.7938 | 0.6420 | 0.6798 | 1.5654 |
| 20 | 0.8413 | 0.5897 | 2.0767 | 0.7196 | 0.6385 | 1.7613 | 0.6344 | 0.6865 | 1.5446 |
| 21 | 0.8241 | 0.5968 | 2.0306 | 0.7085 | 0.6452 | 1.7323 | 0.6275 | 0.6928 | 1.5257 |
| 22 | 0.8084 | 0.6034 | 1.9886 | 0.6984 | 0.6515 | 1.7058 | 0.6213 | 0.6986 | 1.5085 |
| 23 | 0.7942 | 0.6098 | 1.9508 | 0.6892 | 0.6575 | 1.6816 | 0.6157 | 0.7042 | 1.4928 |
| 24 | 0.7817 | 0.6156 | 1.9178 | 0.6810 | 0.6630 | 1.6599 | 1.6599 | 0.6106 | 1.4784 |
| 25 | 0.7697 | 0.6213 | 1.8862 | 0.6734 | 0.6684 | 1.6398 | 0.6059 | 0.7143 | 1.4651 |
| 26 | 0.7587 | 0.6267 | 1.8574 | 0.6663 | 0.6735 | 0.6210 | 0.6016 | 0.7189 | 1.4527 |
| 27 | 0.7490 | 0.6317 | 1.8318 | 0.6599 | 0.6783 | 1.6040 | 0.5977 | 0.7233 | 1.4414 |
| 28 | 0.7395 | 0.6367 | 1.8069 | 0.6539 | 0.6829 | 1.5879 | 0.5940 | 0.7276 | 1.4306 |
| 29 | 0.7311 | 0.6413 | 1.7849 | 0.6484 | 0.6872 | 1.5731 | 0.5906 | 0.7316 | 1.4206 |
| 30 | 0.7229 | 0.6459 | 1.7634 | 0.6432 | 0.6914 | 1.5990 | 0.5874 | 0.7354 | 1.4111 |