THOMAS SIMAR et JOSÉ PARIS, Louvain-la-Neuve

Le nombre de sinistres nécessaires pour en estimer valablement le coût moyen, dans le cas lognormal*

1. Dans une conversation avec Hans Schmitter, le second auteur avait appris que dans les milieux professionnels, on considérait que 25 000 sinistres étaient nécessaires pour en estimer valablement le coût moyen. Il n'est pas possible de vérifier une affirmation aussi générale. Aussi, a-t-il proposé au premier auteur, pour son travail de fin d'études, d'examiner la question lorsque le montant du sinistre suit une loi lognormale. Cet article présente les résultats obtenus.

2. Nous considérons d'abord le cas de la variable aléatoire (v. a.) lognormale à deux paramètres, $X \sim \Lambda(\mu, \sigma^2)$ ou ce qui est équivalent, $Y = \ln X$ est $N(\mu, \sigma^2)$. La fonction de fréquence de X est

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0 \tag{1}$$

et ses moments sont donnés par

$$\gamma_\tau = EX^\tau = e^{\tau\mu + \frac{1}{2}\tau^2\sigma^2} \tag{2}$$

En particulier, en introduisant les paramètres ϱ et φ, on a

$$\begin{aligned} \varrho = EX &= e^{\mu + \frac{1}{2}\sigma^2} \\ \operatorname{var} X &= e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) = \varrho^2\varphi^2 \end{aligned} \tag{3}$$

où

$$\varphi = \sqrt{e^{\sigma^2} - 1} \tag{4}$$

est le coefficient de variation de X égal à $\frac{\sqrt{\operatorname{var} X}}{EX}$.

* Ce travail a été soutenu par le contrat "Projet d'Actions de Recherche Concertées" (PARC No. 93/98-164) du Gouvernement Belge, auquel nous adressons notre grande gratitude.

A partir des valeurs de φ rencontrées en pratique, on peut déduire celles de σ^2 correspondantes. Le fait que φ ne dépend que de σ^2 nous permettra dans la suite un choix intéressant d'unités pour traiter le problème considéré.
La méthode usuelle pour estimer les paramètres μ et σ^2 est la méthode du maximum de vraisemblance. En passant aux v. a. Y, elle fournit les estimateurs

$$\begin{aligned} \widehat{\mu} &= \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \\ \widehat{\sigma}^2 &= V^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \end{aligned} \tag{5}$$

Ces deux estimateurs jouissent de propriétés bien connues. Rappelons que

1) $\widehat{\mu}$ est $N\left(\mu, \dfrac{\sigma^2}{n}\right)$

2) $\widehat{\mu}$ et $\widehat{\sigma}^2$ sont indépendants

3) $\dfrac{n\widehat{\sigma}^2}{\sigma^2}$ est $\chi^2(n-1)$

ou ce qui est équivalent

3′) $\widehat{\sigma}^2$ est $\Gamma\left(\dfrac{n}{2\sigma^2}, \dfrac{n-1}{2}\right)$

En plus la matrice d'information de Fisher pour un échantillon de taille n est

$$I_n(\mu, \sigma^2) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

Le problème considéré ne conduit pas à la recherche d'une estimation optimale du vecteur (μ, σ^2) mais se concentre plutôt sur la recherche d'un estimateur adéquat d'une fonction de ses composantes à savoir $\varrho = e^{\mu + \frac{1}{2}\sigma^2}$ ou ce qui est équivalent $\ln \varrho = \mu + \frac{1}{2}\sigma^2$. Ainsi, par la propriété d'invariance de l'estimateur du maximum de vraisemblance (théorème de Zehna) (Zacks (1971) p 223), l'estimateur du maximum de vraisemblance de ϱ est donné par

$$\widehat{\varrho} = e^{\widehat{\mu} + \frac{1}{2}\widehat{\sigma}^2} \tag{6}$$

Pour le problème analysé, l'usage d'intervalles de confiance est plus approprié que des raffinements de l'estimation ponctuelle. Toutes les méthodes basées sur l'estimateur du maximum de vraisemblance ne présentent pas la même facilité d'application, ni la même efficacité. Ainsi, par exemple, Finney (1941) a obtenu l'estimateur sans biais de variance minimale de ϱ sous la forme

$$\widehat{\varrho}_1 = e^{\overline{Y}} \, {}_0F_1\left(\frac{n-1}{2}, \frac{(n-1)^2}{4n} S_Y^2\right) \tag{7}$$

dans laquelle

$$S_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \tag{8}$$

est l'estimateur sans biais de σ^2 et ${}_0F_1$ est une fonction hypergéométrique généralisée définie par

$${}_0F_1(\alpha, z) = \sum_{j=0}^{\infty} \frac{z^j}{(\alpha)_j j!} \tag{9}$$

Ce résultat a permis à Hoyle (1968) d'obtenir pour ϱ un intervalle de confiance du type

$$\widehat{\varrho}_1 \pm t_{n-1,1-\alpha/2} \sqrt{\widehat{\operatorname{var} \widehat{\varrho}_1}} \tag{10}$$

dans lequel

$$\widehat{\operatorname{var} \widehat{\varrho}_1} = \widehat{\varrho}_1{}^2 - e^{2\overline{Y}} \, {}_0F_1\left(\frac{n-1}{2}, \frac{(n-1)(n-2)}{2n} S_Y^2\right) \tag{11}$$

En plus de sa difficulté de mise en oeuvre, cette méthode présente des inconvénients importants relevés par Land: possibilité de borne inférieure négative et convergence, en fonction de n, du niveau de confiance réel, soit la valeur de la probabilité

$$P\left[\widehat{\varrho}_1 \pm t_{n-1,1-\alpha/2}\sqrt{\widehat{\operatorname{var}\widehat{\varrho}_1}}\,\right]$$

vers le niveau de confiance nominal α, extrêmement lente. Cette procédure ne semble donc pas adéquate pour l'objectif poursuivi. Ce qui nous intéresse particulièrement, c'est de trouver le plus facilement possible un intervalle de confiance de coefficient de confiance donné et de longueur satisfaisante pour les besoins du praticien. Pour ce faire, nous serons amenés à analyser d'une part la méthode exacte et deux méthodes approchées pour comparer les résultats de celles-ci à ceux de la méthode exacte.

3 Méthode approchée 1

Contrairement à la méthode de Finney, l'estimateur du maximum de vraisemblance $\widehat{\varrho}$ ne donne pas un estimateur sans biais mais fournit immédiatement un intervalle de confiance asymptotique pour ϱ, de niveau $1-\alpha$, sous la forme

$$e^{\overline{Y}+\frac{1}{2}V^2}\left(1 \pm z_{1-\frac{\alpha}{2}} V\sqrt{\frac{1+\frac{V^2}{2}}{n}}\right) \tag{12}$$

où $z_{1-\frac{\alpha}{2}}$ est le quantile de la v. a. normale réduite. Ce résultat, non exploité par Land, se déduit facilement du fait que $\widehat{\varrho}$ est asymptotiquement $N\big(\varrho, \frac{1}{n}\varrho^2\sigma^2(1+\frac{\sigma^2}{2})\big)$. Il se démontre par les techniques habituelles. L'une de celles-ci est développée dans Zacks (1971) p 248. La longueur de cet intervalle est aléatoire et égale à

$$L = 2e^{\overline{Y}+\frac{1}{2}V^2} z_{1-\frac{\alpha}{2}} V\sqrt{\frac{1+\frac{V^2}{2}}{n}} \tag{13}$$

Par le théorème de Zehna,

$$\hat{L}_n = \frac{\sqrt{n}L}{2z_{1-\frac{\alpha}{2}}} \tag{14}$$

est l'estimateur du maximum de vraisemblance de la quantité

$$l = h(\mu, \sigma^2) = e^{\mu + \frac{1}{2}\sigma^2} \sigma \sqrt{1 + \frac{\sigma^2}{2}} \tag{15}$$

Puisque L est une fonction de estimateurs $\overline{Y}$ et V^2, la méthode delta (Bishop, Fienberg, Holland (1975), p 486) montre que

$$\sqrt{n}(\hat{L}_n - l) \xrightarrow{\text{loi}} (0, I_h^{-1}(\mu, \sigma^2)) \tag{16}$$

où

$$I_h^{-1}(\mu, \sigma^2) = H(\mu, \sigma^2) I^{-1}(\mu, \sigma^2) H(\mu, \sigma^2)' \tag{17}$$

dans laquelle

$$I^{-1}(\mu, \sigma^2) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \tag{18}$$

$$\begin{aligned} H(\mu, \sigma^2) &= \left(\frac{\partial}{\partial \mu} h(\mu, \sigma^2) \quad \frac{\partial}{\partial \sigma^2} h(\mu, \sigma^2) \right) \\ &= \left(h(\mu, \sigma^2) \quad \frac{1}{2} h(\mu, \sigma^2) + \varrho \frac{1 + \sigma^2}{2\sqrt{\sigma^2 + \frac{\sigma^4}{2}}} \right) \end{aligned} \tag{19}$$

On en déduit que

$$I_h^{-1}(\mu, \sigma^2) = \varrho^2 \left[\sigma^4 \left(1 + \frac{\sigma^2}{2} \right)^2 + \sigma^4 (1 + \sigma^2) + \frac{\sigma^2}{2} \frac{(1 + \sigma^2)^2}{(1 + \frac{\sigma^2}{2})} \right] \tag{20}$$

Pour que la longueur L de l'intervalle de confiance soit inférieure à une quantité δ avec une probabilité $1 - \varepsilon$, on impose la restriction

$$P[L \leq \delta] = 1 - \varepsilon$$

Mais pour que cette restriction fournisse des résultats utilisables en pratique, il faut faire un bon choix des unités en lesquelles on va exprimer δ. Celui-ci résulte du fait que

$$P[L \leq \delta] = P\left[\hat{L}_n \leq \frac{\sqrt{n}\delta}{2 z_{1-\alpha/2}} \right] = 1 - \varepsilon \tag{21}$$

et puisque $\hat{L}_n$ est asymptotiquement normale, on trouve

$$\delta = \left(\frac{2z_{1-\alpha/2}}{n}\right)\left(l\sqrt{n} + z_{1-\varepsilon}\sqrt{I_h^{-1}(\mu, \sigma^2}\right) \tag{22}$$

Il est capital de noter que le terme ϱ peut être mis en évidence dans le second membre de cette dernière relation et dès lors $\frac{\delta}{\varrho}$ représente le quantile de niveau $1 - \varepsilon$ de la v. a. longueur de l'intervalle de confiance quand celle-ci est exprimée en des unités "coût moyen de sinistres". Il en résulte que, de ce fait, ce quantile ne dépend plus que de σ^2 ou encore du coefficient de variation φ de la v. a. X mesurée.
L'expression

$$\frac{\delta}{\varrho} = \frac{2z_{1-\alpha/2}}{\sqrt{n}}\sigma\sqrt{1 + \frac{\sigma^2}{2}}\left(1 + \frac{z_{1-\varepsilon}}{\sqrt{n}} \quad \frac{\sqrt{\frac{1}{2} + 3\sigma^2 + \frac{7}{2}\sigma^4 + \frac{9}{8}\sigma^6 + \frac{1}{8}\sigma^8}}{1 + \frac{\sigma^2}{2}}\right) \tag{23}$$

permet de répondre à la question: lorsque le coefficient de variation est φ, combien faut-il faire d'observations pour obtenir un intervalle de confiance de niveau $(1 - \alpha)$ sur le coût moyen du sinistre quand on veut que la longueur de cet intervalle exprimée en unités ϱ soit dans $100(1 - \varepsilon)\,\%$ des cas inférieure à $\frac{\delta}{\varrho}$. Dans le cas $1 - \alpha = 0.95$, $1 - \varepsilon = 0.99$ la réponse peut être trouvée à l'aide de l'annexe 1 lorsque le coefficient de variation est compris entre 3.5 et 12 (valeurs usuelles) et le rapport $\frac{\delta}{\varrho}$ entre 0.04 et 0.10. A titre d'exemple, notons que si une compagnie d'assurances souhaite obtenir un intervalle de confiance de niveau 0.95 dont la longueur soit dans 99 % des cas au plus égale à 5 % du coût moyen, c'est-à-dire $\frac{\delta}{\varrho} = 0.05$, alors elle doit disposer de 76 320 sinistres, si $\sigma^2 = 3.9$. Le fait de ne pas connaître la valeur exacte de σ^2 n'est pas un problème majeur car, en pratique, la valeur de n est toujours grande.

Remarques:

1) A priori, il est difficile de se faire une idée de la qualité de cette méthode puisque deux approximations y sont faites.

2) La possibilité pour la limite inférieure de la relation (12) d'être négative n'est théoriquement pas exclue. Néanmoins, elle est très peu probable si n est assez grand par rapport à V^2, ce qui est toujours le cas en assurance.

4 Méthode approchée 2

Le même développement peut être effectué à partir de la méthode de Cox qui consiste à estimer $\beta = \ln \varrho$ plutôt que ϱ. Par le théorème de Lehmann-Scheffé, (Zacks (1971) p 104).

$$\widehat{\beta} = \overline{Y} + \frac{1}{2} S_Y^2 \tag{24}$$

est l'estimateur sans biais de variance minimale. Par les propriétés des estimateurs du maximum de vraisemblance dans le cas normal, rappelées plus haut,

$$\frac{(n-1)S_Y^2}{\sigma^2} \quad \text{est} \quad \chi^2(n-1)$$

ou ce qui est équivalent

$$S_Y^2 \quad \text{est} \quad \Gamma\left(\frac{n-1}{2\sigma^2}, \frac{n-1}{2}\right)$$

Il en résulte que

$$\begin{aligned} ES_Y^2 &= \sigma^2 \\ \operatorname{var} S_Y^2 &= \frac{2\sigma^4}{n-1} \\ ES_Y^4 = E(S_Y^2)^2 &= \frac{n+1}{n-1}\sigma^4 \end{aligned}$$

Ainsi, la variance de $\widehat{\beta}$ est donnée par

$$\operatorname{var} \widehat{\beta} = \frac{\sigma^2}{n} + \frac{\sigma^4}{2(n-1)} \tag{25}$$

Elle peut être estimée par

$$\widehat{\operatorname{var}\beta} = \frac{S_Y^2}{n} + \frac{S_Y^4}{2(n+1)} \tag{26}$$

dont on vérifie facilement qu'il est un estimateur sans biais. Dès lors, si n est assez grand,

$$\frac{\widehat{\beta} - \beta}{\sqrt{\widehat{\operatorname{var}\beta}}} \sim N(0,1) \tag{27}$$

On en déduit donc les extrémités de l'intervalle de confiance pour ϱ, soit

$$e^{\overline{Y} + \frac{1}{2}S_Y^2 \pm z_{1-\alpha/2}\sqrt{\frac{S_Y^2}{n} + \frac{S_Y^4}{2(n+1)}}} \tag{28}$$

Notons qu'ici les extrémités de l'intervalle de confiance pour ϱ sont toujours positives. On peut reprendre la même recherche à propos de la valeur de n que celle faite au point précédent. Cette méthode demande cependant plus de temps calcul que la première pour fournir des résultats comparables. Une partie des résultats obtenus à l'aide du logiciel Matlab est reprise dans l'annexe 2. Elle confirme ceux de l'annexe 1.

5 La méthode exacte

Pour valider les méthodes approchées, il était nécessaire de comparer les résultats obtenus à ceux fournis par la méthode exacte. Celle-ci proposée par Land consiste à rechercher directement un intervalle de confiance exact pour $\mu + \frac{1}{2}\sigma^2$ en utilisant d'une part la relation bien connue entre test d'hypothèses et intervalle de confiance et, d'autre part, les propriétés fondamentales des estimateurs de μ et σ^2 dans les populations normales. En exploitant les logiciels Matlab et Mathematica, nous avons vérifié les résultats de Land qui fournissent pour ϱ des intervalles de confiance du type

$$\left(e^{\overline{Y} + m^*(0, S^2, 1-\frac{\alpha}{2})}, e^{\overline{Y} + m^*(0, S^2, \frac{\alpha}{2})}\right) \tag{29}$$

où m^* est fourni par la table de Land que nous avons vérifiée. Malheureusement la forme analytique de m^* n'est pas connue et la méthode delta ne pourra pas être appliquée à la longueur de l'intervalle de confiance obtenu. Des considérations similaires à celles faites plus haut nous ont permis, à partir de simulations, de déduire la valeur de n pour répondre à la question posée. Cet effort nous permet de conclure que la méthode approchée 1 fournit des résultats très satisfaisants pour les besoins du praticien. Comme son utilisation est beaucoup plus simple que celle des deux autres méthodes, il pourra donc utiliser l'annexe 1 pour trouver les valeurs de n qui l'intéresse.

6 La lognormale à trois paramètres

Cette v. a. X est caractérisée par sa densité

$$f(x) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x-\gamma)-\mu)^2}{2\sigma^4}} \qquad x > \gamma$$

$$\varrho = EX = \gamma + e^{\mu + \frac{1}{2}\sigma^2}$$

$$\operatorname{var} X = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

$$\text{coefficients de variation} = \sqrt{e^{\sigma^2} - 1}$$

Le cas où γ est connu se ramène à ce qui précède en supposant que $X - \gamma$ est lognormale. Si γ est inconnu, la recherche des estimateurs du maximum de vraisemblance des trois paramètres et les difficultés de calcul correspondantes ont suscité beaucoup de travaux. Ainsi, Hill (1963) a montré que les maxima globaux de la vraisemblance conduisent à des estimateurs inadmissibles. Il est dès lors indispensable d'utiliser des techniques alternatives. Celles-ci soulèvent aussi beaucoup de questions. Pour traiter notre problème, nous avons considéré uniquement la méthode du maximum de vraisemblance local car la recherche des estimateurs correspondants peut finalement se réduire à une maximisation unidimensionnelle et de plus, ces estimateurs jouissent de bonnes propriétés asymptotiques. A l'aide du logiciel Matlab, nous avons réalisé des simulations qui ont montré leur convergence pratiquement certaine. A partir de ces estimateurs, nous avons trouvé un intervalle de confiance pour ϱ en supposant que son estimateur $\widehat{\varrho}$ suivait asymptotiquement une loi normale de moyenne ϱ et de variance

$$h(\mu, \sigma^2, \gamma) \quad I_n^{-1} \quad h(\mu, \sigma^2, \gamma)' \tag{30}$$

dans laquelle $h(\mu, \sigma^2, \gamma)$ est le vecteur des dérivées partielles de l'espérance $\gamma + e^{\mu + \frac{1}{2}\sigma^2}$ et I_n^{-1} est l'inverse de la matrice d'information de Fisher. Puisque celle-ci requiert le calcul d'espérances qui dans le cas présent sont compliquées, on utilisera plutôt la matrice J_n^{-1} où J_n est la matrice des dérivées secondes de la vraisemblance dans laquelle on remplace les paramètres par leur estimateur du maximum de vraisemblance.
L'intervalle de confiance approché correspondant pour la moyenne est alors de la forme

$$\widehat{\varrho} \pm z_{1-\alpha/2}\sqrt{hJ_n^{-1}h'} \tag{31}$$

Par simulation, nous avons déterminé le niveau de confiance effectivement atteint par cette procédure lorsque $n = 100$ et $\alpha = 0.05$.
Ces bons résultats s'expliquent par une bonne approximation de $\operatorname{var}\widehat{\varrho}$ par $hJ_n^{-1}h'$. L'intervalle de confiance obtenu est déjà satisfaisant lorsque $n = 100$. Lorsque $n = 500$, $\gamma = 5\,000$, $\varrho = 40\,000$ alors le niveau de confiance atteint est 0.94833 si $\varphi = 7$ et 0.94667 si $\varphi = 9$.
La longueur de l'intervalle de confiance (31) est

$$2z_{1-\alpha/2}\sqrt{hJ_n^{-1}h'} \tag{32}$$

Elle permet comme plus haut de déterminer la valeur de n nécessaire pour atteindre un objectif fixé dans l'estimation de ϱ. Des considérations identiques à propos du choix des unités peuvent être faites et finalement les valeurs appropriées de n peuvent encore être déduites de l'annexe 1. A titre d'exemple, examinons deux situations:

a) γ connu ($\gamma = 5\,000$, $\varphi = 3.5$, $\gamma + \varrho = 40\,000$), alors la table 1 montre que $n = 10\,5000$ si on veut avoir une longueur de l'intervalle de confiance de niveau 0.95 qui soit inférieure à 3 500 dans 99 % des cas.
b) si γ est inconnu mais si son estimation laisse supposer qu'une valeur de 5 000 est acceptable, alors à l'aide de 300 simulations, nous avons

trouvé que le quantile correspondant à la situation du point a) passait de 3 500 à 3 570.7. Ce résultat est logique puisqu'on a utilisé une estimation au lieu de la valeur exacte.

Nous avons effectué d'autres comparaisons similaires pour arriver à la même conclusion.

Remerciements: Nous voulons remercier le rapporteur dont les remarques ont sensiblement amélioré la lisibilité du texte.

References

[1] Aitchison, J. and Brown, J. A. C. (1957), *The Lognormal Distribution,*Cambridge University Press, Cambridge.

[2] Bishop, Y. M., Fienberg, S. E., Holland P. W. (1975), *Discrete Multivariate Analysis:* Theory and Practice, MIT Press.

[3] Crow, E. L. and Shimizu, K. (eds) (1988), *Lognormal Distributions: Theory and Applications,* Marcel Dekker, New York.

[4] Finney, D. J. (1941), *On the distribution of a variate whose logarithm is normally distributed,* J. R. Stat. Soc. Suppl, 7, 155–161.

[5] Griffiths, David A. (1980), *Interval Estimation for the Three-parameter Lognormal Distribution via the Likelihood Function,* Applied Statistics, 29, 58–68.

[6) Hill, B. M. (1963), *The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic,* J. Amer. Stat. Assoc, 58, 72–84.

[7] Hoyle, M. H. (1968), *The estimation of variances after using a Gaussianatinq transformation,* Ann. Math. Statist. 39, 1125–1143.

[8] Land, C. E. (1971), *Confidence intervals for linear functions of the normal mean and variance,* Annals of Mathematical Statistics, 42, 1187–1205.

[9] Land, C. E. (1972), *An evaluation of approximate confidence intervals methods for lognormal means,* Technometrics, 14, 145–158.

[10] Land, C. E. (1973), *Standard Confidence Limits for Linear Functions of the Normal Mean and Variance,* Journal of the American Statistical Association, 68, 960–963.

[11] Simar, Th. (1996), *Détermination du nombre d'observations nécessaires pour obtenir une estimation valable de la moyenne d'une lognormale,* Mémoire présenté en vue du grade d'Ingénieur Civil en Mathématiques Appliquées, UCL, Louvain-la-Neuve.

[12] Zacks, S. (1971), *The theory of statistical inference,* Wiley, New York.

Thomas Simar et José Paris
Institut de Statistique et CORE
Voie du Roman Pays 20
Louvain-la-Neuve
Belgique

Annexe 1

	delta/rho												
c.v.	0.04	0.045	0.05	0.055	0.06	0.065	0.07	0 075	0.08	0.085	0.09	0.095	0.1
3.5	0 6046	0.4812	0.3926	0.3267	0.2765	0.2372	0.2060	0.1806	0.1598	0.1425	0.1280	0.1156	0.1050
4	0.6984	0.5558	0.4534	0.3773	0.3193	0.2739	0.2378	0.2085	0.1845	0.1645	0.1477	0.1334	0.1212
4.5	0.7876	0.6267	0.5112	0.4255	0.3600	0.3088	0.2681	0.2351	0.2080	0.1854	0.1665	0.1504	0.1366
5	0.8726	0.6944	0.5664	0.4713	0.3987	0.3420	0.2969	0.2603	0.2303	0.2054	0.1844	0.1665	0.1512
5.5	0.9537	0.7588	0.6189	0.5150	0.4357	0.3737	0.3244	0.2844	0.2516	0.2243	0.2014	0.1819	0.1652
6	1.0312	0.8204	0.6691	0.5567	0.4710	0.4040	0.3506	0.3074	0.2720	0.2425	0.2176	0.1966	0.1785
6.5	1.1053	0.8793	0.7171	0.5967	0.5047	0.4329	0.3758	0.3295	0.2914	0.2598	0.2332	0.2106	0.1913
7	1.1763	0.9358	0.7632	0.6350	0.5371	0.4607	0.3998	0.3506	0.3101	0.2764	0.2481	0.2241	0.2035
7.5	1.2446	0.9901	0.8074	0.6718	0.5682	0.4874	0.4230	0.3708	0.3280	0.2924	0.2624	0.2370	0.2152
8	1.3103	1.0423	0.8500	0.7072	0.5981	0.5130	0.4452	0.3903	0.3452	0.3078	0.2762	0.2495	0.2265
8.5	1.3735	1.0926	0.8910	0.7413	0.6270	0.5377	0.4666	0.4091	0.3619	0.3226	0.2895	0.2614	0.2374
9	1.4346	1.1412	0.9306	0.7742	0.6548	0.5616	0.4873	0.4272	0.3779	0.3368	0.3023	0.2730	0.2479
9.5	1.4937	1.1882	0.9688	0.8060	0.6817	0.5846	0.5073	0.4448	0.3934	0.3506	0.3147	0.2842	0.2580
10	1.5508	1.2336	1.0059	0.8368	0.7077	0.6069	0.5267	0.4617	0.4084	0.3640	0.3267	0.2950	0.2679
10.5	1.6062	1.2776	1.0417	0.8666	0.7329	0.6286	0.5454	0.4781	0.4229	0.3769	0.3383	0.3055	0.2774
11	1.6599	1.3203	1.0765	0.8955	0.7574	0.6495	0.5636	0.4941	0.4370	0.3895	0.3495	0.3156	0.2866
11.5	1.7121	1.3618	1.1103	0.9236	0.7812	0.6699	0.5813	0.5096	0.4507	0.4017	0.3605	0.3255	0.2955
12	1.7628	1.4021	1.1432	0.9510	0.8043	0.6897	0.5985	0.5246	0.4640	0.4135	0.3711	0.3351	0.3042

Annexe 2

	delta/rho			
	0.04	0.06	0.08	0.1
c.v.				
4	0.7161	0.3310	0.1932	0.1281
5	0.8993	0.4162	0.2433	0.1615
6	1.0668	0.4944	0.2893	0.1923
7	1.2209	0.5664	0.3318	0.2207
8	1.3636	0.6332	0.3712	0.2470
9	1.4965	0.6955	0.4080	0.2717

Résumé

Dans ce travail, nous montrons comment trouver le nombre d'observations nécesssaires pour estimer la moyenne d'une variable aléatoire lognormale, avec une précision souhaitée quand on a fait un bon choix d'unités pour exprimer la longueur de l'intervalle de confiance correspondant. Nous comparons deux méthodes approchées avec la méthode exacte et nous montrons que la plus simple fournit d'excellents résultats. Nous étendons les résultats au cas de la variable aléatoire lognormale à trois paramètres. Des résultats numériques sont fournis.

Summary

The present paper indicates how to find the sample size necessary to estimate the mean of a lognormal random variable with an adapted precision related to a good choice of units to express the length of the confidence interval. We compare two approximate methods to the exact one and we observe that the simplest one gives excellent results. We extend the solution to the case of a lognormal distribution with three parameters. Pratical numerical results are given.

Zusammenfassung

In dieser Arbeit zeigen wir, wie die Anzahl Beobachtungen bestimmt werden kann, die notwendig sind, um den Mittelwert einer lognormal verteilten Zufallsvariable mit einer gewünschten Genauigkeit zu schätzen, falls die Breite des Konfidenzintervalls in gut gewählten Masseinheiten ausgedrückt ist. Wir vergleichen zwei Näherungsmethoden mit der exakten und zeigen, dass die einfachste Methode ausgezeichnete Resultate ergibt. Wir erweitern die Ergebnisse zum Fall der dreiparametrig lognormal verteilten Zufallsvariable. Numerische Resultate sind gegeben.