Bandas de confianza aproximadas para la función media del proceso de difusión lognormal homogéneo

Gutiérrez Jáimez, R., Román Román, P.
Rico Castro, N., Torres Ruiz, F.

rgjaimez@ugr.es, proman@ugr.es, nrico@ugr.es, fdeasis@ugr.es
Departamento de Estadística e I.O., Universidad de Granada

Resumen

En este artículo se obtienen, y comparan en términos de su error de cobertura y amplitud media, bandas de confianza aproximadas para la función media del proceso de difusión lognormal homogéneo.

Palabras Clave: Proceso de difusión lognormal, bandas de confianza.

1. Introducción

En la naturaleza existen multitud de fenómenos que pueden ser modelizados mediante una variable aleatoria positiva. Un modelo que ha sido extensamente utilizado es el modelo de la distribución lognormal, $X \rightarrow \Lambda(\mu, \sigma)$, siendo la inferencia sobre su media un problema ampliamente estudiado. Land [3] obtuvo un intervalo de confianza óptimo a partir del test insesgado uniformemente más potente para la media, el cual presenta el problema de ser computacionalmente tedioso e inestable para ciertos valores de los estadísticos $\bar{X}$ y S^2. En [5] se obtienen bandas de confianza, generalizando el método propuesto por Land, para funciones parámetricas generales que incluyen como caso particular la función media (y su versión condicionada) del proceso de difusión lognormal con factores exógenos, las cuales presentan también los problemas computacionales ya comentados.

De forma alternativa se han propuesto métodos de construcción de tests aproximados (más tratables desde el punto de vista computacional), entre los que destacamos el método de Cox [4], el método conservativo de Angus [1] y la aproximación bootstrap [2] de este último.

En este artículo se generalizan expresiones de intervalos de confianza aproximados para la media de la distribución lognormal al caso del proceso lognormal homogéneo, obteniéndose bandas de confianza para su función media y media condicionada. En concreto, se obtienen expresiones para bandas de confianza "naive", la adaptación del método de Cox, de la transformación de Patterson, del método conservativo de Angus y del método bootstrap que calcula los puntos críticos del método conservativo de Angus a partir de la generación de muestras bootstrap. Además, se presenta un estudio de simulación que permite comparar, en función del número de datos observados (n) en cada trayectoria simulada del proceso y del valor de σ^2 (probando previamente que no influye el valor de m), las bandas propuestas en términos de errores de cobertura y de amplitudes medias.

2. Planteamiento del problema y notación

Sea $\{X(t), t \geq t_0\}$ el proceso de difusión lognormal homogéneo, con momentos infinitesimales $A_1(t) = mx$ y $A_2(t) = \sigma^2 x^2$ con $\sigma > 0$ y $m \in R$. Tomando $a = m - \frac{\sigma^2}{2}$, la función media y su versión condicionada se pueden escribir en la forma $\exp\{\mu(t,s) + \lambda\sigma^2(t,s)\}$ con $\mu(t,s) = ln(x_0) + a(t - t_0)$, $\sigma^2(t,s) = (t - t_0)\sigma^2$ en el caso de la versión sin condicionar, $\mu(t,s) = ln(x_s) + a(t - s)$, $\sigma^2(t,s) = (t - s)\sigma^2$ para la versión condicionada y $\lambda = 1/2$, en ambos casos.

Sea $t_1, \ldots, t_n$ un conjunto de instantes de tiempo en los cuales se realiza un muestreo discreto obteniéndose las observaciones $x_1, \ldots, x_n$. Sean v_i los valores $(t_i - t_{i-1})^{-1/2}(\ln(x_i) - \ln(x_{i-1}))$ $i = 2, \ldots, n$. Para cada (t,s) fijado se pretende resolver el problema de contraste $H_0 : \mu(t,s) + \lambda\sigma^2(t,s) = \theta_0$ frente a $H_1 : \mu(t,s) + \lambda\sigma^2(t,s) \neq \theta_0$, o, equivalentemente

$$\begin{cases} H_0 : \dfrac{\mu(t,s) - \theta_0}{\sigma^2} = -\lambda \\ H_1 : \dfrac{\mu(t,s) - \theta_0}{\sigma^2} \neq -\lambda \end{cases}$$

A partir de él se construye un intervalo de confianza para $\mu(t,s)+\lambda\sigma^2(t,s)$. Para los valores apropiados de $\mu(t,s)$, λ y $\sigma^2(t,s)$, tomando exponenciales se obtiene en cada (t,s) un intervalo para la media del proceso lognormal homogéneo.

Los estimadores de máxima verosimilitud de los parámetros son

$$\hat{a} = \frac{1}{t_n - t_1} \ln\left(\frac{x_n}{x_1}\right)$$

$$\hat{\sigma}^2 = \frac{1}{n-1}(\mathbf{v} - \hat{a}\mathbf{u})'(\mathbf{v} - \hat{a}\mathbf{u})$$

donde $\mathbf{v}$=$(v_2, \ldots, v_n)'$ y $\mathbf{u}$=$(\sqrt{t_2 - t_1}, \ldots, \sqrt{t_n - t_{n-1}})'$. La distribución de estos estimadores es

$$\hat{a} \to N(a, \frac{\sigma^2}{t_n - t_1})$$

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \to \chi^2_{n-2}$$

Tomando la notación

$$
\begin{array}{ll}
\mu(t,s) = \ln(x_s) + (t-s)a & \mu(t) = \ln(x_1) + (t-t_1)a \\
B(t,s) = \ln(x_s) + (t-s)\hat{a} & B(t) = \ln(x_1) + (t-t_1)\hat{a} \\
\sigma^2(t,s) = (t-s)\sigma^2 & \sigma^2(t) = (t-t_1)\sigma^2 \\
S^2(t,s) = (t-t_1)\frac{(n-1)\hat{\sigma}^2}{n-2} & S^2(t) = (t-t_0)\frac{(n-1)\hat{\sigma}^2}{n-2}
\end{array}
$$

para la función media y para la función media condicionada, se tiene que $(B(t,s), S^2(t,s))$ es suficiente y completo para $(\mu(t,s), \sigma^2(t,s))$ y que $(B(t), S^2(t))$ es suficiente y completo para $(\mu(t), \sigma^2(t))$. Estos estadísticos se utilizarán a la hora de resolver el problema de contraste planteado.

3. Métodos aproximados de transformación y directos

Los métodos de transformación obtienen un límite de confianza para $E[g(X(t))]$ donde g es una transformación cualquiera. Este límite se transforma mediante alguna función apropiada para dar un límite aproximado para $E[X(t)]$. El método naive (referencia) es un método de transformación donde se construye, para cada (t,s) fijo, un intervalo de confianza para $E[\ln(X(t))] = \mu(t,s)$, que se puede notar como μ_α, y se transforman los intervalos obtenidos tomando exponenciales $\xi_\alpha = \exp\{\mu_\alpha\}$. Este método es sesgado salvo cuando el valor de σ^2 es extremadamente pequeño. Para eliminar el sesgo del método naive, Patterson propone distintas transformaciones de los intervalos. Una de ellas es $\xi_\alpha = \exp\left\{\mu_\alpha + \lambda S^2(t,s)\right\}$.

Los métodos directos se basan en distribuciones aproximadas de estimadores de $E[X(t)]$ o de alguna función directa de ella, de forma que se evitan fallos debidos a la dependencia de $E[X(t)]$ en $\mu(t,s)$ y σ^2. Como inconveniente se tiene que si se supone la normalidad de los estimadores de una función paramétrica sin rango restringido puede dar lugar a intervalos de confianza inadmisibles. El método de Cox está basado en el estadístico $B(t,s)+\lambda S^2(t,s)$ y en la suposición de normalidad del mismo. El método de Angus se basa en un estadístico pivotal $\dfrac{B(t,s)+\lambda S^2(t,s) - \big(\mu(t,s)+\lambda\sigma^2(t,s)\big)}{\sqrt{S^2(t,s)C(t,s) + 2\lambda^2 S^4(t,s)/(n-2)}}$ donde $C(t,s) = \dfrac{t-t_0}{t_n - t_0}$. La idea de utilizar este estadístico surge del hecho de que $B(t,s)+\lambda S^2(t,s)$ es el UMVUE de $\mu(t,s)+\lambda\sigma^2(t,s)$ con varianza asintótica $C(t,s)\sigma^2(t,s)\dfrac{2\lambda^2\sigma^4}{n-2}$. Este método también surge de la construcción de intervalos de confianza bilaterales basados en el criterio de razón de verosimilitudes.

4. Generalización de los intervalos de confianza aproximados

4.1. Métodos de transformación: Banda naive y transformación

de Patterson

El método naive propone construir los intervalos de confianza para cada (t,s) fijos a partir del intervalo de confianza para $E[\ln(X(t))]$, tomando como transformación ξ_α la exponencial del intervalo obtenido, de forma que la banda se calcula como

$$\exp\left\{B(t,s) \pm t_{n-2;\alpha/2}\sqrt{C(t,s)}S(t,s)\right\}$$

La transformación que propone Patterson es $\xi_\alpha = \exp\left\{\mu_{\alpha+\lambda S^2(t,s)}\right\}$ de forma que la expresión de la banda es

$$\exp\left\{B(t,s) \pm t_{n-2;\alpha/2}\sqrt{C(t,s)}S(t,s) + \lambda S^2(t,s)\right\}$$

4.2. Métodos directos: Método de Cox, método de Angus y método bootstrap

El método de Cox está basado en el UMVUE de $\mu(t,s) + \lambda\sigma^2(t,s)$, que es $B(t,s) + \lambda S^2(t,s)$ y en la suposición de normalidad de éste,

$$B(t,s) + \lambda S^2(t,s) \to N\left(\mu(t,s) + \lambda\sigma^2(t,s), S^2(t,s)C(t,s) + \frac{2\lambda^2}{n}S^4(t,s)\right)$$

de forma que se obtiene la expresión de la banda de confianza

$$\exp\left\{B(t,s) + \lambda S^2(t,s) \pm z_{\alpha/2}\sqrt{C(t,s)S^2(t,s) + 2\lambda^2\frac{S^4(t,s)}{n}}\right\}$$

Este método de construcción aparece en [4] y es tratado en [6] tomando una distribución del estadístico diferente, la varianza del estadístico no se calcula a partir del UMVUE de la varianza, sino a partir la sustitución de σ^2 por su estimación S^2 en la expresión de la varianza del estimador. El estadístico de partida y su distribución en este caso son

$$B(t,s) + \lambda S^2(t,s) \to N\left(\mu(t,s) + \lambda\sigma^2(t,s), S^2(t,s)C(t,s) + \frac{2\lambda^2}{n-2}S^4(t,s)\right)$$

llegando a la expresión para el cálculo de la banda de confianza

$$\exp\left\{B(t,s) + \lambda S^2(t,s) \pm z_{\alpha/2}\sqrt{C(t,s)S^2(t,s) + 2\lambda^2\frac{S^4(t,s)}{n-2}}\right\}$$

El método de Angus propone la construcción de bandas de confianza a partir de un estadístico pivotal que se distribuye como

$$\frac{N + \frac{\lambda}{C(t,s)}\sigma(t,s)(W-1)}{\sqrt{W\left(1 + \frac{2\lambda^2 W}{C(t,s)(n-2)}\sigma^2(t,s)\right)}}$$

donde W y N son independientes, $N \to N(0,1)$ y $W \to \frac{\chi^2_{n-2}}{n-2}$. Para cada (t,s) fijo la función de distribución es monótona creciente en $\sigma(t,s)$. Tomando límites

$$\lim_{\sigma(t,s)\to 0} \frac{N + \frac{\lambda}{C(t,s)}\sigma(t,s)(W-1)}{\sqrt{W\left(1 + \frac{2\lambda^2 W}{C(t,s)(n-2)}\sigma^2(t,s)\right)}} = \frac{N}{\sqrt{W}} \to t_{n-2}$$

$$\lim_{\sigma(t,s)\to \infty} \frac{N + \frac{\lambda}{C(t,s)}\sigma(t,s)(W-1)}{\sqrt{W\left(1 + \frac{2\lambda^2 W}{C(t,s)(n-2)}\sigma^2(t,s)\right)}} = \sqrt{\frac{n-2}{2}}\left(1 - \frac{1}{W}\right)$$

La banda de confianza se calcula como

$$\left(\exp\left\{B(t,s) + \lambda S^2(t,s) - t_{n-2;\alpha/2}\sqrt{C(t,s)S^2(t,s) + 2\lambda^2\frac{S^4(t,s)}{n-2}}\right\},\right.$$

$$\left.\exp\left\{B(t,s) + \lambda S^2(t,s) + \sqrt{\frac{n-2}{2}}\left(\frac{n-2}{\chi^2_{n-2;1-\alpha/2}} - 1\right)\sqrt{C(t,s)S^2(t,s) + 2\lambda^2\frac{S^4(t,s)}{n-2}}\right\}\right)$$

El método conservativo de Angus es similar al método directo de Cox. La diferencia en este caso es que los intervalos de confianza se ajustan de modo que la probabilidad de cobertura mínima sea igual a su valor nominal. En el método directo de Cox el tamaño del test no está controlado.

El método bootstrap propone utilizar valores bootstrap del estadístico pivote usado en el método de Angus, generando valores con distribución normal estándar y valores chi-cuadrado de forma independiente. Con los valores obtenidos se calculan los percentiles para la construcción del intervalo de confianza. El proceso será generar k valores $N_1^*, \ldots, N_k^*$ con distribución $N(0,1)$ y k valores $W_1^*, \ldots, W_k^*$ con distribución $\chi^2/(n-2)$ de forma independiente. Con estos valores se calcula T_i^*, $i = 1, \ldots, k$ como

$$T_i^* = \frac{N_i^* + \frac{\lambda}{\sqrt{C(t,s)}}S(t,s)(W_i^*-1)}{\sqrt{W_i^*\left(1 + \frac{2\lambda^2}{C(t,s)(n-2)}W_i^* S^2(t,s)\right)}}$$

debido a la distribución del estadístico pivotal que se está utilizando. Los valores T_i^* se ordenan en $T_{(1)}^* < \ldots < T_{(k)}^*$ y se calcula $t_{1-\alpha/2}^{boot} = T[(1-\alpha/2)k]$ y $t_{\alpha/2}^{boot} = T[(\alpha/2)k]$ donde $[a]$ denota el entero menor o igual a a. Con estos valores, se construye la banda de confianza bootstrap

$$\left(\exp\left\{B(t,s)+\lambda S^2(t,s)-t_{1-\alpha/2}^{boot}\sqrt{C(t,s)S^2(t,s)+2\lambda^2\frac{S^4(t,s)}{n-2}}\right\},\right.$$

$$\left.\exp\left\{B(t,s)+\lambda S^2(t,s)-t_{\alpha/2}^{boot}\sqrt{C(t,s)S^2(t,s)+2\lambda^2\frac{S^4(t,s)}{n-2}}\right\}\right)$$

4.3. Método propuesto

Tomando como punto de partida los métodos de transformación, se puede observar que en la construcción de las bandas se tiene en cuenta solamente la variabilidad de $\mu(t,s)$ en ambos métodos. El método que se propone a continuación considera la variabilidad de $\sigma^2(t,s)$ calculando un intervalo de confianza para σ^2. En la construcción de la banda se tienen en cuenta el intervalo de confianza construido para cada uno de los parámetros y se conjugan los extremos en la siguiente expresión:

$$\left(\exp\left\{B(t,s)-t_{n-2;\alpha/2}S(t,s)\sqrt{C(t,s)}+\lambda\frac{n-2}{\chi^2_{n-2;\alpha/2}}S^2(t,s)\right\}\right.$$

$$\left.\exp\left\{B(t,s)+t_{n-2;\alpha/2}S(t,s)\sqrt{C(t,s)}+\lambda\frac{n-2}{\chi^2_{n-2;1-\alpha/2}}S^2(t,s)\right\}\right)$$

5. Estudio de simulación. Comparación de los métodos

Para comparar la bondad de las bandas de confianza se realizan simulaciones del proceso, según el procedimiento descrito por Rao, con $m = 0{,}5$ y σ^2 variando entre los valores 0,001, 0,01, 0,1 y 1. El número de datos simulados en cada trayectoria varía, tomando los valores 11, 101 y 1001, de forma que para cada combinación de valores de n y de σ^2 se hace una simulación de 100 trayectorias. A partir de cada una de las trayectorias simuladas se realiza la estimación de los parámetros y a partir de ellos se calculan las bandas de confianza propuestas anteriormente. Las medidas que se utilizan para comparar las bandas de confianza serán la probabilidad de cobertura de la banda, la amplitud media, el error de cobertura...

6. Agradecimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Ciencia y Tecnología, Proyecto de Investigación BFM 2002-03633.

7. Bibliografía

[1] Angus, J. E. (1988). Inferences on the lognormal mean for complete samples. *Communications in Statistics -Simulation and Computation*, 17, 1307-1331.

[2] Angus, J. E. (1994). Bootstrap one-sided confidence intervals for lognormal mean. *Statistician*, 43, 395-401.

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

[4] Land, C. E. (1972). An evaluation of approximate confidence interval estimation of a population mean using tranformed sample data. *Technometrics*, 8, 535-537.

[5] Gutiérrez Jáimez, R., Román Román, P., Romero Molina, D. y Torres Ruiz, F. (2003). Forecasting for the univariate lognormal diffusion process with exogenous factors. *Cybernetics and Systems: An International Journal*, 34, 709-724.

[6] Zhou, X. & Gao, S. (1997). Confidence intervals for the log-normal mean. *Statistics in Medicine*, 16, 783-790.