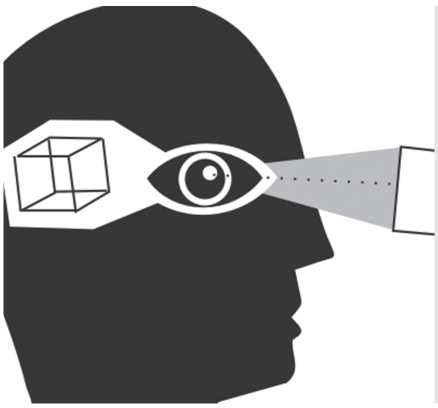
REVISTA Universidad EAFIT
Vol. 42. No. 144. 2006. pp. 81-98

Comparación por intervalos
entre diferentes métodos de estimación de la media de la distribución Poisson



Juan Carlos Correa M.
Doctor en Estadística. Profesor de la Escuela de Estadística de la Universidad Nacional de Colombia- Sede Medellín.
jccorrea@unal.edu.co

Francisco Castrillón M.
Profesor de la Escuela de Estadística de la Universidad Nacional de Colombia - Sede Medellín.
fjcastri@unal.edu.co

Recepción: 27 de mayo de 2006 I Aceptación: 25 de junio de 2006

Resumen

La construcción de intervalos de confianza para la estimación del parámetro λ de la distribución Poisson es un problema importante en el trabajo estadístico aplicado. Se revisan diferentes procedimientos para su construcción y se realiza un estudio de simulación para mostrar que para muestras grandes los procedimientos basados en el criterio de la raíz y el de razón de verosimilitudes tienen un comportamiento similar y más eficiente que el de los otros procedimientos usualmente utilizados.

Palabras Clave

Distribución Poisson
Estimación
Intervalo de confianza

Estimate by interval of the Poisson distribution measure

Abstract

The construction of confidence intervals for the estimate of the λ parameter of the Poisson distribution is an important problem in applied statistic work. We reviewed different procedures for its construction and we carried out a simulation study to analyze the performance of the real confidence levels when comparing them to the theoretical levels.

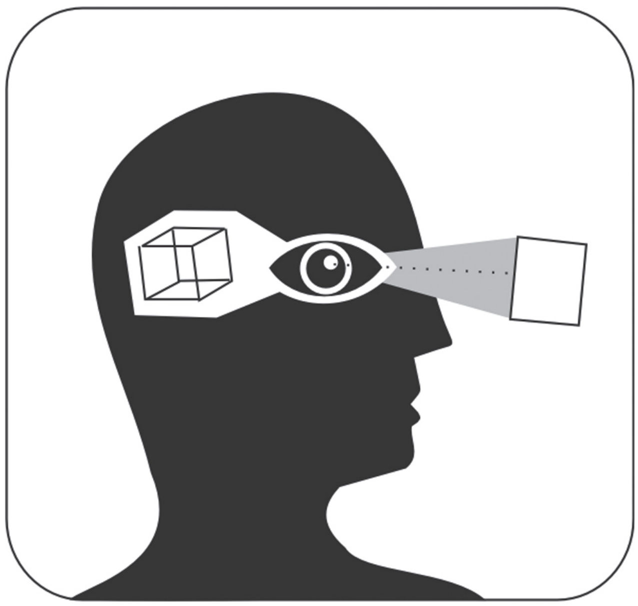
Key words

Poisson distribution
Estimation
Confidence interval

Introducción

La *distribución Poisson*[1] es importante para describir ocurrencias *aletorias* siendo estas ocurrencias o bien objetos en un espacio determinado o eventos en el tiempo. Una distribución aleatoria de objetos en el espacio es aquella en la cual cada porción del espacio tiene la misma probabilidad de contener un objeto y la ocurrencia de un objeto en una porción del espacio de ninguna manera influencia la ocurrencia de cualquier otro de los objetos en cualquier porción del espacio. Una distribución de eventos en el tiempo es aquella en la cual cada periodo de tiempo tiene igual oportunidad de abarcar un evento, y la ocurrencia de un evento cualquiera es independiente de la ocurrencia de cualquier otro evento.

La distribución Poisson encuentra aplicaciones en campos tales como la salud, economía y finanzas e ingeniería, entre otros, para modelar conteos de ocurrencias de eventos de interés. Ejemplos típicos son el numero de accidentes de transito por día, lo cual permite desarrollar cartas de control; el numero de casos de cierta enfermedad con lo cual los epidemiólogos pueden alertar sobre la posible aparición de una epidemia; en entidades financieras una variable de interés es el numero de clientes que incumplen una obligación.



Debido al importante trabajo aplicado en la modelación de problemas de conteo, se han considerado diversos métodos para estimar la media de esta distribución (Pearson and Hartley, 1966; Hahn, Gerald J. and Meeker, 1991; Wardell, D.G. 1997). En casos prácticos, el problema para el analista es la carencia de reglas claras sobre cuál fórmula escoger.

En la perspectiva de ayudarle en su decisión, se ha realizado un estudio de simulación que permite analizar -por distintos métodos- el comportamiento de los niveles de confianza reales al ser comparados con los niveles nominales. Generalmente se han utilizado métodos tradicionales tales como el basado en el Teorema Central del Límite y el de máxima verosimilitud, y también métodos basados en transformaciones como la logarítmica y la raíz cuadrada. Aquí, además de estos, se recurrió también al estudio del *método exacto,* al de *razón de verosimilitud relativa,* al *método bootstrap* y al método basado en el *modelo lineal generalizado.*

[1] Descrita en 1837 por Simeon-Denis Poisson (1781-1840), matemático y físico frances; aunque aparentemente fue descrita previamente por Abraham de Moivre (1667-1754). También fue descrita independientemente por otros incluyendo a “Student” (W.S. Goset, 1876-1937) en 1909 (Haight, 1967)

Si se asume que $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una distribución Poisson con función de masa dada por: (Meyer, 1986)

$$p_x(X = x) = \frac{\lambda^{\lambda x}\exp(-\lambda)}{X!}$$

para $\lambda > 0$ y $x > 0$ La media de la distribución es $\mu = \lambda$ y su varianza $\sigma^2 = \lambda$. El estimador de máxima verosimilitud para λ es $\bar{X}$ (Canavos, 1988).

1. Estimación por intervalos de confianza

Los límites de confianza para el parámetro de la distribución Poisson - el cual es a su vez la media y la varianza poblacional- dependen del método de estimación que se va a utilizar, bien sea el método exacto, por transformación o mediante el uso del Teorema Central del Límite, entre otros. Como se podrá ver, no todos son iguales, algunos son robustos, uniforme y asintóticamente mejores que otros y, por supuesto están los poco eficientes y por tanto poco recomendables para su uso.

1.1 Método basado en el Teorema Central del Límite (T. C. L.)

El Teorema Central del Límite establece que si $S = \sum_{i=1}^{n} X_i$ es una muestra de variables aleatorias independientes e idénticamente distribuidas con una distribución de probabilidad no especificada y que tienen una media μ y una varianza σ^2 finita, el promedio muestral $\bar{X} = (1/)n \sum_{i=1}^{n} X_i$ tiene una distribución con media μ y varianza σ^2/n que tiende hacia una distribución normal conforme n tiende a ∞ (Canavos, 1988). De este modo, si el tamaño muestral es lo suficientemente grande, podemos aplicar este teorema para hallar un intervalo de confianza para λ:

$$\left(\bar{X} - Z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{X} + Z_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$

donde $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$. Este es el intervalo propuesto en la mayoría de textos de estadística básica (Wonnacott y Wonnacott, 1979; Roussas, 1973; Walpole, 1992; Meyer, 1986; Mood *et al.*, 1974).

1.2 Método basado en la Máxima Verosimilitud de Fisher

La función de verosimilitud es de la forma:

$$L(\theta; x_1, \ldots x_n) = L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta)$$

El estimador máximo verosímil es el $\hat{\theta}$ que maximiza esta función. Se sabe que si $\hat{\theta}$ es el estimador máximo verosímil para θ (puede ser un vector), $\hat{\theta}$ es asintóticamente normal θ (Serfling, 1980); es decir, $\hat{\theta} \sim AN\ (\theta, I^{-1}(\theta))$, siendo $I(\hat{\theta})$ la matriz de información de Fisher o matriz Hessiana, I, en la que la (i, j) ésima entrada está dada por:

$$I_{ij} = E\left[-\frac{\partial^2 \log L(\theta; X)}{\partial\theta_i \partial\theta_j}\right]$$

Entonces, en el caso de una distribución Poisson, $\hat{\lambda} = \bar{X} \sim AN\left(\lambda, \frac{\lambda}{n}\right)$. Así, el intervalo de confianza para λ está dado por:

$$\left(\bar{X} - Z_{\alpha/2}\sqrt{\frac{\bar{X}}{n}}, \bar{X} + Z_{\alpha/2}\sqrt{\frac{\bar{X}}{n}}\right)$$

Nótese la similitud con el intervalo basado en el Teorema Central del Límite; aquí en vez de s está $\bar{X}$.

1.3 Método Exacto

Se sabe que $S = \sum_{i=1}^{n} X_i$ se distribuye Poisson con parámetro $n\lambda$ (Kalbfleish, 1985). El intervalo de confianza para la proporción de ocurrencia media de una distribución Poisson (Hahn y Meeker, 1991), siendo $\hat{\lambda} = \frac{s}{n}$ y s el número de ocurrencias, está dado por:

$$\left[\frac{0.5\chi^2_{(\alpha/2;2s)}}{n}, \frac{0.5\chi^2_{(1-\alpha/2;2s+2)}}{n}\right]$$

donde $\chi^2_{(\alpha/2;2s)}$, $\chi^2_{(1-\alpha/2;2s+2)}$, son los valores críticos inferior y superior de la variable aleatoria chi-cuadrada con 2*s* y 2*s*+2 grados de libertad respectivamente, que corresponden al porcentaje $\alpha/2$ de la distribución; esto es $P(X^2 > \chi^2_{(\alpha/2;2s)}) = \alpha/2$ y $P(X^2 > \chi^2_{(1-\alpha/2;2s+2)}) = 1 - \alpha/2$.

1.4 Intervalo basado en transformaciones

Si *X* es una variable aleatoria con función de densidad de probabilidad f_X y $y = h(x)$ es una transformación bajo ciertas condiciones de continuidad y medibilidad, de modo que exista la transformación inversa $x = h^{-1}(y)$, entonces, $Y = h(X)$ es una variable aleatoria con función de densidad (Roussas, 1973):

$$f_Y(y) = f_X[h^{-1}(y)]\left|\frac{d}{dy}h^{-1}(y)\right|$$

A continuación se usarán dos transformaciones que cumplen las condiciones exigidas, una es la transformación logaritmo y otra es la transformación raíz cuadrada. Ambas están basadas en la propiedad de normalidad asintótica.

1.4.1 Transformación logaritmo

Si $\bar{X} \sim N(\lambda, /n)$, entonces $log(\bar{X}) \sim AN\left(log(\lambda), \frac{1}{n\lambda}\right)$, asumiendo que $\lambda>0$ (Serfling, 1980). El intervalo de confianza para λ está dado por

$$\left[\bar{X}\exp\left(-z_{\alpha/2}/\sqrt{n\bar{X}}\right), \bar{X}\exp\left(z_{\alpha/2}/\sqrt{n\bar{X}}\right)\right]$$

1.4.2 Transformación raíz cuadrada

La transformación raíz cuadrada tiene varianza constante. En este caso, $\sqrt{\bar{X}} \sim AN\left(\sqrt{\lambda}, \frac{1}{4n}\right)$ (How & Craig, 1978).

Por tanto el intervalo de confianza para λ viene dado por:

$$\left[\left(\sqrt{\bar{X}} - Z_{\alpha/2}\frac{1}{2\sqrt{n}}\right), \left(\sqrt{\bar{X}} - Z_{\alpha/2}\frac{1}{2\sqrt{n}}\right)^2\right]$$

1.5 Intervalos basados en la Razón de Verosimilitud Relativa

Kalbfleish (1985) presenta la metodología para construir intervalos de verosimilitud. Si $L(\lambda)$ es la función de verosimilitud, la *función de verosimilitud relativa* se define como

$$R(\lambda) = \frac{L(\lambda)}{L(\hat{\lambda})}$$

En el caso de la distribución de Poisson esta razón viene dada por:

$$R(\lambda) = \left(\frac{\lambda}{\bar{X}}\right)^{nX} e^{-n(\lambda-\bar{X})}$$

El conjunto de valores de λ para los cuales $R(\lambda) \geq$ p, es llamado el *intervalo de 100% p de verosimilitud* para λ. El intervalo del 14.7% de verosimilitud corresponde al intervalo de confianza del 95%.

El procedimiento consiste entonces, en hallar las raíces de $R(\lambda) \geq 0.147$ o de

$$r(\lambda) = n\bar{X}\log\left(\frac{\lambda}{\bar{X}}\right) - n(\lambda - \bar{X}) \geq \log(0.147)$$

Esta ecuación se resuelve numéricamente.

1.6 Bootstrap

El método bootstrap proporciona una manera directa y sencilla para hallar intervalos simultáneos para el parámetro de la distribución Poisson. Para hallarlos se procede así:

1. Se estima el parámetro por máxima verosimilitud a partir de la muestra:

$$\hat{\lambda} = \frac{1}{n}\sum_{=1}^{n} X_i$$

2. Genere *M* muestras de tamaño *n* de una distribución Poisson con parámetro $\hat{\lambda}$. Para cada muestra estime el parámetro λ, digamos que para la muestra *j* el estimador es $\hat{\lambda}_j$.
3. Para los $\left\{\hat{\lambda}_j\right\}_{j=1}^{M}$ se construye un histograma y se calculan los percentiles 0.025 y 0.975, denotándolos por $\hat{\lambda}^{0.025}$ y $\hat{\lambda}^{0.975}$. El intervalo del 95% para el parámetro de la distribución Poisson es entonces, $(\hat{\lambda}^{0.025}, \hat{\lambda}^{0.975})$.

1.7 Intervalos de confianza utilizando el MLG

Bajo el modelo lineal generalizado, MLG (McCulloch & Searle, 2001), podemos construir un intervalo de confianza para λ utilizando un modelo sólo con intercepto de la regresión Poisson. El modelo que se debe utilizar es $\ln(\lambda) = \beta_0$, ya que no hay covariables y la funcion de enlace es la logarítmica. Se estima, entonces, β_0 y se halla el error estándar del estimador, s.e.($\hat{\beta}_0$). Así, el intervalo de confianza para λ está dado por:

$$\left[\exp\left(\hat{\beta}_0 - z_{\alpha/2} s.e.(\hat{\beta}_0)\right), \exp\left(\hat{\beta}_0 + z_{\alpha/2} s.e(\hat{\beta}_0)\right)\right]$$

2. Resultados de la simulación

Para comparar los distintos métodos de estimación se creó un programa en R, Versión 2.1.0. Se generaron muestras de una distribución Poisson para cada combinación de n y λ. De cada combinación se hicieron 5000 simulaciones y se halló un intervalo de confianza del 95% para la tasa nominal por los distintos métodos anotados. Se calculó la longitud promedio de los 5000 intervalos y se determinó la proporción de intervalos que contienen el valor de la tasa; a esta proporción se le dio el nombre de *Nivel de Confianza Real.*

Se presentan tablas para cada combinación de *n=10, 20, 50, 100* y *200* y λ*=0.1, 0.25, 0.50, 1, 2 y 5.* Se muestran los percentiles $P_{0.05}$ y $P_{0.95}$, así como la mediana y la media de las longitudes de los intervalos, junto con los niveles de confianza reales para cada método de estimación: TLC, Verosimilitud, Exacto, Transformación Raíz Cuadrada, Transformación Logarítmica, Bootstrap, Razón de Verosimilitud y MLG. En la última columna aparece el *Nivel de Confianza Real. k* es el porcentaje de muestras con *cero* éxitos.

Resultados

Tabla 1. $\lambda = 0.1$, $n = 10$, k = 0.3692

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.39200	0.39200	0.47891	0.83667	0.99588
Vero.	0.39200	0.39200	0.48103	0.67896	0.99841
Exacto	0.55463	0.55463	0.63461	0.81486	0.96290
Raíz	0.39200	0.39200	0.48103	0.67896	0.96290
Log	0.69585	0.69585	0.73346	0.83344	0.96290
Boot.	0.30000	0.30000	0.41633	0.70000	0.96544
R.V.	0.43393	0.43393	0.51660	0.70340	0.96290
MLG	0.69585	0.69585	0.73346	0.83344	0.96290

Tabla 2. $\lambda = 0.1$, $n = 20$, k = 0.1220

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.19600	0.26979	0.28678	0.45856	0.99727
Vero.	0.19600	0.27719	0.28850	0.43827	0.99863
Exacto	0.27732	0.34912	0.36131	0.50224	0.98246
Raíz	0.19600	0.27719	0.28850	0.43827	0.98246
Log	0.34792	0.37484	0.39275	0.49659	0.93530
Boot.	0.15000	0.25000	0.26572	0.45000	0.98292
R.V.	0.21737	0.29149	0.30378	0.44695	0.98246
MLG	0.34792	0.37484	0.39275	0.49654	0.93530

Tabla 3. $\lambda = 0.1, n = 50,$ k = 0.0092

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.10974	0.16800	0.17158	0.24255	0.87707
Vero.	0.11087	0.17531	0.17236	0.23520	0.87767
Exacto	0.13965	0.20090	0.19842	0.25939	0.98728
Raíz	0.11087	0.17531	0.17236	0.23520	0.95155
Log	0.14994	0.19863	0.19824	0.25229	0.96810
Boot.	0.10000	0.18000	0.16761	0.22050	0.86758
R.V.	0.11696	0.17848	0.17621	0.23798	0.93238
MLG	0.14994	0.19863	0.23798	0.93238	0.96810

Tabla 4. $\lambda = 0.1, n = 100,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.08586	0.12327	0.12235	0.15813	0.92320
Vero.	0.08765	0.12396	0.12239	0.15680	0.92120
Exacto	0.10045	0.13595	0.13446	0.16838	0.97260
Raíz	0.08765	0.12396	0.12239	0.15680	0.95440
Log	0.09931	0.13205	0.13088	0.16315	0.95820
Boot.	0.09000	0.12000	0.12081	0.15000	0.90520
R.V.	0.08908	0.12553	0.12364	0.15712	0.94000
MLG	0.09931	0.13205	0.13088	0.16315	0.95820

Tabla 5. $\lambda = 0.1, n = 200,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.06917	0.08689	0.08686	0.10512	0.92980
Vero.	0.07066	0.08765	0.08698	0.10371	0.94680
Exacto	0.07654	0.09336	0.09270	0.10931	0.95420
Raíz	0.07066	0.08765	0.08698	0.10371	0.93760
Log	0.07420	0.09048	0.08988	0.10610	0.94100
Boot.	0.07000	0.08500	0.08612	0.10500	0.92380
R.V.	0.07133	0.08708	0.08731	0.10340	0.95420
MLG	0.07420	0.09048	0.08988	0.10610	0.94100

Tabla 6. $\lambda = 0.1, n = 200,$ k = 0.0802

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.39200	0.59879	0.61918	0.87654	0.98870
Vero.	0.39200	0.67896	0.62630	0.87654	0.99870
Exacto	0.55463	0.81486	0.76845	1.0045	0.98434
Raíz	0.39200	0.67896	0.62630	0.87654	0.98434
Log	0.69585	0.83344	0.81614	0.99317	0.95477
Boot.	0.30000	0.60000	0.58635	0.90000	0.99174
R.V.	0.43393	0.70240	0.65162	0.89390	0.98434
MLG	0.69585	0.83339	0.81613	0.99317	0.95477

Tabla 7. $\lambda = 0.25, n = 20,$ k = 0.008

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.26979	0.41212	0.42395	0.64224	0.87117
Vero.	0.27719	0.43827	0.42759	0.58800	0.86935
Exacto	0.34912	0.50224	0.49287	0.64847	0.98850
Raíz	0.27719	0.43827	0.49287	0.64847	0.98850
Log	0.37484	0.49659	0.49288	0.63073	0.96532
Boot.	0.25000	0.45000	0.41595	0.55125	0.86109
R.V.	0.29149	0.44695	0.43761	0.59530	0.93024
MLG	0.37484	0.49654	0.49287	0.63072	0.96532

Tabla 8. $\lambda = 0.25, n = 50,$ k = 0.000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.20191	0.26996	0.27348	0.35418	0.93160
Vero.	0.20743	0.27159	0.27481	0.34174	0.92220
Exacto	0.23217	0.29522	0.29849	0.36463	0.96600
Raíz	0.20743	0.27159	0.27481	0.34174	0.94560
Log	0.22693	0.28631	0.28974	0.35337	0.95080
Boot.	0.20000	0.26050	0.27145	0.34000	0.93320
R.V.	0.21059	0.27355	0.27710	0.34390	0.94560
MLG	0.22693	0.28631	0.28974	0.35335	0.95080

Tabla 9. $\lambda = 0.25, n = 50,$ k = 0.000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.15759	0.19365	0.19416	0.23437	0.93220
Vero.	0.16163	0.19600	0.19488	0.22857	0.94700
Exacto	0.17316	0.20726	0.20616	0.23966	0.95680
Raíz	0.16163	0.19600	0.19488	0.22857	0.93840
Log	0.16778	0.20106	0.20003	0.23290	0.94600
Boot.	0.16000	0.19000	0.19334	.23000	0.92800
R.V.	0.16200	0.19595	0.19540	0.22957	o.95680
MLG	0.16778	0.20106	0.20003	0.23289	0.94600

Tabla 10. $\lambda = 0.25, n = 50,$ k = 0.000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.11887	0.13754	0.13816	0.15864	0.94080
Vero.	0.12240	0.13859	0.13829	0.15433	0.94820
Exacto	0.12791	0.14404	0.14374	0.15973	0.95080
Raíz	0.12240	0.13859	0.13829	0.15433	0.94120
Log	0.12442	0.14037	0.14009	0.15593	0.94200
Boot.	0.12000	0.13513	0.13739	0.15500	0.93720
R.V.	0.12170	0.13839	0.13843	0.15422	0.95000
MLG	0.12442	0.14037	0.14009	0.15593	0.94200

Tabla 11. $\lambda = 0.5, n = 10,$ k = 0.082

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.52267	0.86674	0.83918	1.31320	0.87840
Vero.	0.55437	0.87654	0.85572	1.17600	0.87497
Exacto	0.69825	1.00450	0.98618	1.29690	0.98467
Raíz	0.55437	0.87654	0.85572	1.17600	0.95301
Log	0.74969	0.99317	0.98566	1.26150	0.96532
Boot.	0.50000	0.90000	0.83264	1.10000	0.86490
R.V.	0.58398	0.82032	0.86307	1.18880	0.93366
MLG	0.74969	0.99314	0.98565	1.26150	0.96532

Tabla 12. $\lambda = 0.5, n = 10,$ k = 0.082

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.41212	0.59653	0.60194	0.82912	0.91998
Vero.	0.43827	0.61981	0.60944	0.75910	0.92278
Exacto	0.50224	0.67975	0.66985	0.81724	0.97700
Raíz	0.43827	0.61981	0.60944	0.75910	0.95560
Log	0.49659	0.66026	0.65203	0.79192	0.96420
Boot.	0.45000	0.60000	0.60074	0.75000	0.91378
R.V.	0.44695	0.62613	0.61592	0.76426	0.94279
MLG	0.49654	0.66024	0.65202	0.79192	0.96420

Tabla 13. $\lambda = 0.5, n = 50,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.30136	0.38943	0.38802	0.48316	0.93720
Vero.	0.32325	0.39200	0.39086	0.45715	0.94540
Exacto	0.34631	0.41452	0.41341	0.47931	0.95400
Raíz	0.32325	0.39200	0.39086	0.45715	0.93780
Log	0.33556	0.40212	0.40112	0.46580	0.94200
Boot.	0.32000	0.38050	0.38752	0.46000	0.93220
R.V.	0.32492	0.39339	0.39224	0.45898	0.95400
MLG	0.33556	0.40211	0.40211	0.46579	0.94200

Tabla 14. $\lambda = 0.5, n = 100,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.23278	0.27553	0.27597	0.32189	0.93880
Vero.	0.24465	0.27719	0.27677	0.30866	0.94360
Exacto	0.25566	0.28808	0.28767	0.31946	0.94840
Raíz	0.24465	0.27719	0.27677	0.30866	0.93680
Log	0.24860	0.28075	0.28036	0.31186	0.94180
Boot.	0.24000	0.28000	0.27502	0.31000	0.93460
R.V.	0.24338	0.27878	0.27712	0.30827	0.94840
MLG	0.24867	0.28074	0.28035	0.31185	0.94180

Tabla 15. $\lambda = 0.5, n = 200,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.17318	0.19452	0.19571	0.21912	0.94100
Vero.	0.17964	0.19600	0.19585	0.21291	0.93920
Exacto	0.18498	0.20131	0.20117	0.21820	0.94800
Raíz	0.17964	0.19600	0.19585	0.21291	0.94200
Log	0.18101	0.19726	0.19711	0.21407	0.94160
Boot.	0.17500	0.19500	0.19473	0.21500	0.93780
R.V	0.17922	0.19670	0.19585	0.21245	0.94220
MLG	0.18100	0.19725	0.19711	0.21406	0.94160

Tabla 16. $\lambda = 1, n = 10,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.70366	1.16870	1.19020	1.77730	0.90440
Vero.	0.87654	1.23960	1.22090	1.51820	0.92160
Exacto	1.00450	1.35950	1.34160	1.63450	0.97300
Raíz	0.87654	1.23960	1.22090	1.51820	0.95340
Log	0.99317	1.32050	1.30590	1.58380	0.96000
Boot.	0.90000	1.20000	1.20510	1.50000	0.90900
R.V.	0.82032	1.25130	1.22530	1.52850	0.91400
MLG	0.99314	1.32050	1.30590	1.58380	0.96000

Tabla 17. $\lambda = 0.5, n = 20,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.60328	0.84840	0.86406	1.15430	0.93300
Vero.	0.70669	0.87654	0.87234	1.03710	0.95080
Exacto	0.76542	0.93360	0.92952	1.09310	0.96000
Raíz	0.70669	0.87654	0.87234	1.03710	0.94200
Log	0.74201	0.90487	0.90120	1.06100	0.94620
Boot.	0.70000	0.85000	0.86536	1.05000	0.93220
R.V.	0.71241	0.88088	0.87624	1.03980	0.95960
MLG	0.74200	0.90487	0.90120	1.06100	0.94620

Tabla 18. $\lambda = 1, n = 50,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.44293	0.54823	0.54940	0.66600	0.93760
Vero.	0.48961	0.55437	0.55301	0.61732	0.94900
Exacto	0.51163	0.57616	0.57481	0.63892	0.95140
Raíz	0.48961	0.55437	0.55301	0.61732	0.94100
Log	0.49769	0.56150	0.56019	0.62372	0.94260
Boot.	0.48000	0.54050	0.54955	0.62000	0.93580
R.V.	0.49015	0.55455	0.55358	0.61798	0.95140
MLG	0.49768	0.56150	0.56019	0.62372	0.94260

Tabla 19. $\lambda = 1, n = 100,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.33696	0.38890	0.39009	0.44732	0.94460
Vero.	0.35927	0.39200	0.39141	0.42401	0.94320
Exacto	0.36996	0.40263	0.40204	0.43459	0.95460
Raíz	0.35927	0.39200	0.39141	0.42401	0.94800
Log	0.36202	0.39451	0.39394	0.42634	0.94760
Boot.	0.35000	0.39000	0.38941	0.43000	0.93940
R.V.	0.35848	0.39139	0.39143	0.42314	0.94800
MLG	0.36201	0.39451	0.39394	0.42634	0.94760

Tabla 20. $\lambda = 1, n = 200,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.24922	0.27576	0.27627	0.30440	0.94620
Vero.	0.26076	0.27719	0.27698	0.29335	0.94660
Exacto	0.26600	0.28241	0.28220	0.29855	0.95180
Raíz	0.26076	0.027719	0.27698	0.29335	0.94560
Log	0.26171	0.27807	0.27787	0.29419	0.94820
Boot.	0.25500	0.27500	0.27543	0.29513	0.94340
R.V.	0.26049	0.27631	0.27685	0.29257	0.94820
MLG	0.26170	0.27807	0.27787	0.29418	0.94820

Tabla 21. $\lambda = 2, n = 10$, k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	1.02050	1.65800	1.70410	2.49300	0.90740
Vero.	1.41340	1.75310	1.74550	2.07430	0.94480
Exacto	1.53080	1.86720	1.85980	2.18620	0.95220
Raíz	1.41340	1.75310	1.74550	2.07430	0.93460
Log	1.48400	1.80970	1.80320	2.12200	0.93880
Boot.	1.40000	1.70000	1.72900	2.10000	0.92240
R.V.	1.37020	1.75980	1.74700	2.08090	0.93820
MLG	1.48400	1.80970	1.80320	2.12200	0.93880

Tabla 22. $\lambda = 2, n = 20$, k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.87538	1.20660	1.21960	1.59550	0.93240
Vero.	1.07350	1.23960	1.23150	1.38590	0.94140
Exacto	1.12930	1.29460	1.28650	1.44040	0.96040
Raíz	1.07350	1.23960	1.23150	1.38590	0.94980
Log	1.09660	1.25960	1.25170	1.40370	0.95260
Boot.	1.05000	1.20120	1.22330	1.40000	0.93740
R.V.	1.07660	1.24120	1.23350	1.38890	0.94200
MLG	1.09660	1.25950	1.25170	1.40370	0.95260

Tabla 23. $\lambda = 2, n = 50$, k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.64297	0.78047	0.78132	0.93007	0.94000
Vero.	0.71855	0.78400	0.78444	0.84803	0.94320
Exacto	0.73992	0.80526	0.80570	0.86918	0.95340
Raíz	0.71855	0.784000	0.78444	0.84803	0.94660
Log	0.72404	0.78903	0.78948	0.85267	0.94700
Boot.	0.70050	0.78000	0.78018	0.86000	0.94140
R.V.	0.72016	0.78378	0.78427	0.84883	0.94660
MLG	0.72404	0.78903	0.78948	0.85267	0.94700

Tabla 24. $\lambda = 2, n = 100,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.48412	0.55297	0.55385	0.62666	0.94440
Vero.	0.52145	0.55437	0.55431	0.58669	0.94340
Exacto	0.53192	0.56481	0.56475	0.59710	0.95040
Raíz	0.52145	0.55437	0.55431	0.58669	0.94640
Log	0.52334	0.55615	0.55609	0.58837	0.94520
Boot.	0.51000	0.55025	0.55163	0.59025	0.94300
R.V.	0.52116	0.55381	0.55410	0.55410	0.94820

Tabla 25. $\lambda = 2, n = 200,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.35533	0.39058	0.39096	0.42638	0.94500
Vero.	0.37548	0.39200	0.39191	0.40832	0.94880
Exacto	0.38064	0.39715	0.39706	0.41347	0.95300
Raíz	0.37548	0.39200	0.39191	0.40832	0.95020
Log	0.95020	0.39263	0.39254	0.40892	0.94960
Boot.	0.36512	0.39012	0.39010	0.39010	0.94360
R.V.	0.37671	0.39125	0.39164	0.40831	0.94680
MLG	0.37614	0.39263	0.39254	0.40892	0.94960

Tabla 26. $\lambda = 5, n = 10,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	1.67330	2.67790	2.70330	3.85190	0.91360
Vero.	2.44800	2.77190	2.76060	3.08660	0.94780
Exacto	2.55810	2.88080	2.86960	3.19460	0.94160
Raíz	2.44800	2.77190	2.76060	3.08660	0.94160
Log	2.48840	2.80750	2.79660	3.11860	0.94760
Boot.	2.40000	2.70250	2.74290	3.10000	0.93760
R.V.	2.42300	2.77480	2.75900	3.08710	0.95060
MLG	2.48840	2.80750	2.79660	3.11860	0.94760

Tabla 27. $\lambda = 5, n = 20,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	1.42480	1.91780	1.93380	2.49350	0.93120
Vero.	1.79640	1.96000	1.96060	2.12010	0.94820
Exacto	1.84980	2.01310	2.01380	2.17300	0.95320
Raíz	1.79640	1.96000	1.96060	2.12010	0.94940
Log	1.81010	1.97260	1.97320	2.13170	0.94620
Boot.	1.75000	1.95000	1.95140	2.15000	0.94400
R.V.	1.79880	1.96100	1.96090	2.12130	0.94940
MLG	1.81010	1.97260	1.97320	2.13170	0.94620

Tabla 28. $\lambda = 5, n = 50,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	1.02530	1.23100	1.23280	1.44910	0.94760
Vero.	1.17340	1.23960	1.23960	1.30250	0.95400
Exacto	1.19420	1.26040	1.26040	1.32320	0.95580
Raíz	1.17340	1.23960	1.23960	1.30250	0.95320
Log	1.17670	1.24280	1.24280	1.30550	0.95580
Boot.	1.14050	1.24000	1.23320	1.32050	0.94940
R.V.	1.17400	1.23970	1.23890	1.30150	0.95320
MLG	1.17670	1.24280	1.24280	1.30550	0.95580

Tabla 29. $\lambda = 5, n = 100,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.76673	0.87099	0.87178	0.97770	0.94520
Vero.	0.84348	0.87566	0.87577	0.90840	0.94740
Exacto	0.85376	0.88593	0.88604	0.88604	0.95160
Raíz	0.84348	0.87566	0.87577	0.90840	0.94780
Log	0.84465	0.87679	0.87690	0.90948	0.95000
Boot.	0.82000	0.87025	0.87171	0.93000	0.94680
R.V.	0.84318	0.87568	0.87513	0.90605	0.95060
MLG	0.84464	0.87678	0.87689	0.90947	0.95000

Tabla 30. $\lambda = 5, n = 200,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	0.5666	0.61881	0.61916	0.67267	0.94600
Vero.	0.60316	0.61981	0.61965	0.63572	0.94760
Exacto	0.60825	0.62490	0.62474	0.64080	0.95060
Raíz	0.60316	0.61981	0.61965	0.63572	0.94860
Log	0.60357	0.62020	0.62005	0.63610	0.94720
Boot.	0.58500	0.61513	0.61656	0.65025	0.94660
R.V.	0.60096	0.61998	0.61920	0.63477	0.94860
MLG	0.60356	0.62020	0.62005	0.63610	0.94720

Tabla 31. $\lambda = 10, n = 10,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	2.35200	3.78930	3.82350	5.39390	0.91740
Vero.	3.59270	3.92000	3.91790	4.24010	0.94120
Exacto	3.69960	4.02630	4.02420	4.34590	0.95280
Raíz	3.59270	3.92000	3.91790	4.24010	0.94520
Log	3.62020	3.94510	0.94510	4.26340	0.94640
Boot.	3.50000	3.90000	3.89440	4.30000	0.93920
R.V.	3.57650	3.91890	3.91370	4.24040	0.94440
Glm	3.62002	3.94510	3.94320	4.26340	0.94640

Tabla 32. $\lambda = 10, n = 20,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	2.02280	2.72630	2.74700	3.53920	0.93380
Vero.	2.60760	2.77190	2.76920	2.92690	0.94660
Exacto	2.66000	2.82400	2.82140	2.97900	0.95520
Raíz	2.60760	2.77190	2.76920	2.92690	0.95060
Log	2.61710	2.78070	2.77810	2.93530	0.95220
Boot.	2.55000	2.75000	2.75420	2.95120	0.94680
R.V.	2.60700	2.77080	2.76830	2.92540	0.95520
Glm	2.61710	2.78070	2.77810	2.93530	0.95220

Tabla 33. $\lambda = 10\ n = 50,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	1.45060	1.73550	1.73810	2.02610	0.94300
Vero.	1.68700	1.75310	1.75240	1.81510	0.94960
Exacto	1.70750	1.77360	1.77300	1.83560	0.95120
Raíz	1.68700	1.75310	1.75240	1.81510	0.94860
Log	1.68930	1.75530	1.75470	1.81730	0.94900
Boot.	1.64000	1.74050	1.74450	1.86000	0.94660
R.V.	1.68520	1.75180	1.75130	1.81350	0.95120
Glm	1.68930	1.75530	1.75470	1.81730	0.94900

Tabla 34. $\lambda = 10, n = 100,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TCL	1.09050	1.23200	1.23420	1.37780	0.94180
Vero.	1.20570	1.23900	1.23910	1.27140	0.94420
Exacto	1.21590	1.24920	1.24930	1.28160	0.94700
Raíz	1.20570	1.23900	1.23910	1.27140	0.94460
Log	1.20650	1.23980	1.23990	1.27220	0.94340
Boot.	1.17000	1.23050	1.23330	1.30050	0.94140
R.V.	1.20260	1.23940	1.23810	1.27040	0.94420
Glm	1.20650	1.23980	1.23990	1.27220	0.94340

Tabla 35. $\lambda = 10, n = 200,$ k = 0.0000

Método	P0.05	Mediana	Media	P0.95	NCR
TLC	0.80055	0.87495	0.87463	0.94909	0.94920
Vero.	0.86084	0.87676	0.87680	0.89260	0.95120
Exacto	0.86590	0.88181	0.88185	0.89766	0.95200
Raíz	0.86084	0.87676	0.87680	0.89260	0.95060
Log	0.86113	0.87704	0.87708	0.89288	0.95060
Boot.	0.83000	0.87050	0.87297	0.92012	0.94940
R.V.	0.85958	0.87432	0.87607	0.89312	0.94940
Glm	0.86112	0.87704	0.87708	0.89288	0.95060

3. Ejemplo

Las siguientes *Y*, por mm, son fallas de un alambre delgado de cobre de longitud *L*(Montgomery & Runger, 1996):

2 2 3 2 2 2 3 2 1 3 2 3 1 4 1 3 3 2 3 4 3 3 2 3 0 4 0 8 2 1 4 3 1 3 2 2

El estimador de máxima verosimilitud para la media de fallas por mm. es $\hat{\lambda} = \hat{Y} = 2.4722$. La siguiente tabla muestra los intervalos de confianza correspondientes a los metodos estudiados.

Tabla 36. Intervalos de confianza

Método	Límite inferior	Límite superior	Longitud intervalo
TCL	2.01365	2.93079	0.91714
Vero.	1.95859	2.98585	1.02726
Exacto	1.98539	3.04228	1.05689
Raiz	1.98527	3.01252	1.02725
Log.	2.00843	3.04304	1.03461
R.V.	0.32402	2.56317	2.23915
Boostrap	2.00000	2.97222	0.97222
MLG	2.00843	3.04304	1.03461

Se puede observar, entonces, que el intervalo más grande es el de la razon de verosimilitud relativa, mientras el mas pequeño es el generado por el Teorema Central del Limite que, junto con el boostrap son los mas precisos. Coinciden los de la transformacion logaritmo y el modelo lineal generalizado.

Conclusiones

- Para tasas cercanas a cero y tamaños de muestra pequeños, hay mayor probabilidad de encontrar muestras que no contienen al verdadero valor de la tasa.
- Si se tienen en cuenta las longitudes medias de los intervalos, así como los niveles reales de pertenencia de la tasa a estos intervalos, de los ocho métodos de estimación empleados, los uniformemente mejores fueron el criterio de la raíz y el de razón de verosimilitud. El método exacto es también muy bueno comparado con estos para tamaños muestrales grandes; sin embargo, son mejores los dos primeros. Entre el criterio de la raíz y el de razón de verosimilitud no se puede escoger uno; algunas veces es mejor el de la raíz y otras, el segundo.
- Contrariamente a lo que se podría pensar, el método más deficiente de todos para estimar la media de la distribución Poisson, aún para tamaños muestrales grandes, es el basado en el Teorema Central del Límite. Y, de nuevo contra lo que se puede creer, da la apariencia de ser el mejor método para tasas cercanas a cero ($\lambda \leq 0.25$) y tamaños muestrales pequeños ($n \leq 10$).
- El método de máxima verosimilitud presenta un comportamiento similar al del TCL, aunque con niveles reales mas altos uniformemente

De lo expuesto anteriormente, se deduce que no se puede recomendar el uso del TCL para estimar por intervalos el parámetro de la distribución Poisson aún para tamaños de muestras grandes; en cambio para este propósito es mejor utilizar el criterio de la raíz o el de razón de verosimilitudes. El comportamiento irregular de los otros métodos probados vía simulación, también los hace poco recomendables haciéndolos unas veces muy eficientes y otras no; es decir, no tienen un patrón de comportamiento determinado.

Bibliografía

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: John Wiley and Sons.

McCulloch, C.E. y Searle, S.R. (2001). *Generalizad and Mixed Models.* New York: John Wiley and Sons,

Haight, F.A. (1967). *Handbook of the Poisson Distribution*. New York: John Wiley and Sons. pp. 117.

Canavos, G. (1988). *Probabilidad y Estadística: Aplicaciones y Métodos*. McGraw Hill: Madrid

Kalbfleish, J.G. (1985). *Probability and Statistical Inference. Vol. 2*. (2 ed). New York: Springer-Verlag.

Meyer, P.L. (1986). *Probabilidad y aplicaciones estadísticas*. (2 ed.). México: Addison Wesley Iberoamericana.

Mood , A.M, Graybill, F.A. y Boes, D.C. (1974). *Introduction to the Theory of Statistics*. (3 ed.). Tokio: McGraw-Hil.

Roussas, G.G. (1973). *A First Course in Mathematical Statistics*. Massachusetts: Addison-Wesley Publishing.

Pearson, E. S., and H. O. Hartley. (1966). *Biometrika Tables for Statisticicians.* Vol. 1. (3 ed.). Cambridge: University Press. 264 p.

Walpole, R.E. y Myers, R.H. (1992). *Probabilidad y Estadística*. (4 ed.). México: McGraw Hill.

Wardell, D.G. (1997) "Small-Sample Interval Estimation of Bernoulli and Poisson Parameters". En: *The American Statistician*. Vol. 51, No. 4, pp. 321-32

Wonnacott, T.H. y Wonnacott, R.J. (1979). *Fundamentos de Estadística para Administración y Economía*. México: Limusa.

Hahn, Gerald J. and Meeker, William Q. (1991). Statistical Intervals A Guide for Practitioners. New York: John Wiley and Sons. 123. p.