

Poisson para tasa de defectos en construcción de software**Poisson for defect rate in software construction**

LÓPEZ-PADILLA, Gilberto†*, SORIA-GARCÍA, Myriam y SANTOS-POMPA, Diego Marcelo

*Universidad Tecnológica de León, México.*ID 1^{er} Autor: *Gilberto, López-Padilla* / ORC ID: 0000-0002-9431-1656, Researcher ID Thomson: S-4615-2018ID 1^{er} Coautor: *Myriam, Soria-García* / ORC ID: 0000-0001-5718-9949, Researcher ID Thomson: V-9022-2019ID 2^{do} Coautor: *Diego Marcelo, Santos-Pompa* / ORC ID: 0000-0002-2174-2877

DOI: 10.35429/JCS.2020.11.4.23.26

Recibido: 15 de Enero, 2020; Aceptado 30 de Junio, 2020

Resumen

Este trabajo documenta el proceso del modelado estadístico de una tasa de defectos probables, utilizando la distribución Poisson, en la construcción de software. Esta tasa es muy importante ya que se usa para la estimación de los defectos en productos de la fábrica de software de la Universidad Tecnológica de León, UTL. La fábrica de software de la UTL ha implementado el modelo CMMI para el desarrollo en su nivel 4 y los resultados preliminares obtenidos fueron parte del trabajo alcanzar el nivel; y pueden ser de gran ayuda para organismos y empresas interesados en la implementación de dicho modelo.

Software, Defectos, Estimación, CMMI**Abstract**

This paper documents the process of statistical modeling of a probable defect rate, using the Poisson distribution, in software construction. This rate is very important since it is used to estimate defects in products of the Technological University of León's, UTL, software factory. The UTL's software factory has implemented the CMMI model for development at level 4 and the preliminary results obtained were part of the work reaching the level; and they can be of great help for organizations and companies interested in the implementation of this model.

Software, Defects, Estimation, CMMI

Citación: LÓPEZ-PADILLA, Gilberto, SORIA-GARCÍA, Myriam y SANTOS-POMPA, Diego Marcelo. Poisson para tasa de defectos en construcción de software. Revista de Simulación Computacional. 2020. 4-11:23-26.

* Correspondencia al autor (Correo electrónico: gpadilla@utleon.edu.mx)

† Investigador contribuyendo como primer autor

Introducción

La UTL debe producir estimaciones de defectos que estén apegadas a modelos de calidad, para validar sus buenas prácticas cuando se buscó el nivel 4 de CMMI.

En los procesos de la UTL se considera que un producto está “libre” de defectos, si los defectos que pueden presentarse en un producto entregado no son más del promedio de defectos observados históricamente más 3 desviaciones estándar, esto es, seis sigmas; lo que significa un límite empírico de 99.97% de los defectos probables por debajo del cual un producto no cumpliría con la especificación de calidad.

Los productos de software de la UTL se ponderan en puntos de función sin ajustar, PFSA, que tienen una equivalencia en horas, y se ha demostrado en trabajos anteriores que el esfuerzo de construcción se puede modelar usando una distribución estadística para una tasa que relaciona el esfuerzo con los PFSA.

Para el caso de los defectos, si tenemos n defectos observados históricamente para un producto ponderado en u PFSA, se propone la siguiente tasa, δ , como una densidad de defectos por PFSA:

$$\delta = n/u \quad (1)$$

Metodología

Usando datos históricos, se usarán herramientas estadísticas para observar la distribución que siguen, esperando, como dictan muchos autores de textos de estadística; entre ellos, Luis Rincón, que sea la distribución de Poisson.

Si esta tasa puede describirse como una distribución Poisson, entonces se puede usar δ como la λ de Poisson. La fórmula (1) podría reescribirse de la siguiente forma:

$$\lambda = n/u \quad (2)$$

Teniendo λ , se conocerían la media y la varianza de la distribución Poisson y por consiguiente la desviación estándar, pues la media y la varianza para esta distribución son iguales: $\mu = \lambda$ y $\sigma^2 = \lambda$ (Humberto Gutiérrez, 2013).

El área OPP del nivel 4 de CMMI para el desarrollo, tiene el propósito de establecer y mantener una comprensión cuantitativa del desempeño de procesos seleccionado [y la practica especifica 1.5 es] seleccionar y mantener modelos de desempeños de procesos (CMMI Product Team, 2010), para esto, la UTL utiliza simuladores que usan el método Montecarlo que requieren una media y una desviación estándar de una distribución que modele la medida algún proceso, y con base en ello producir estimaciones. En este caso se usará λ para producir la simulación y se despejará n para obtener una estimación de defectos para un producto ponderado en PFSA.

$$n = \lambda u \quad (3)$$

Los datos usados para la determinación de la distribución de δ , fueron datos de dos proyectos de la fábrica de software gestionados con CMMI nivel 3. Los datos incluían la siguiente información: orden cronológico, nombre de proyecto, nombre del requerimiento, PFSA, y defectos encontrados en las pruebas: unitarias, por pares, de integración y regresión. Es un conjunto de 27 requerimientos ponderados en PFSA y se analizaron, en conjunto para contar con el mayor número de registros posible, con la herramienta estadística Minitab haciendo pruebas de bondad de ajuste para Poisson. Dado que todas las pruebas buscan esencialmente lo mismo, no se hizo distinción entre las pruebas, de haberse hecho, no se hubiera contado con suficientes registros para el análisis.

La tabla 1 muestra un resumen de los datos a analizar. El nombre del requerimiento en esta tabla solo refleja el orden en que se codificó, importante para realizar gráficos de control, pero su generación y uso no se aborda en este documento.

Proyecto	Requerimiento	PFSA	Defectos
P1	RF01	0.60	7
P1	RF02	0.60	5
P1	RF03	0.30	6
P1	RF04	0.40	2
P1	RF05	0.40	7
P1	RF06	0.30	6
P1	RF07	0.30	2
P1	RF08	0.40	3
P1	RF09	0.50	4
P1	RF10	0.30	5
P1	RF11	0.70	4
P1	RF12	1.00	2
P2	RF01	0.60	1
P2	RF02	0.40	1
P2	RF03	0.60	0
P2	RF04	0.30	0
P2	RF05	0.30	4
P2	RF06	0.30	1
P2	RF07	0.30	4
P2	RF08	0.30	0
P2	RF09	0.30	0
P2	RF10	0.30	1
P2	RF11	0.30	2
P2	RF12	0.40	1
P2	RF13	0.30	5
P2	RF14	0.30	3
P2	RF15	0.40	1

Tabla 1 Datos históricos de defectos de dos proyectos de la fábrica de software de la UTL

Resultados

Asumiendo que $\lambda = \delta$, un primer intento de análisis se hizo para el conjunto completo de requerimientos, usando la formula número 2 para obtener una tasa de defectos para un solo PFSA. Sin embargo, la tasa produce números con parte fraccionaria; y un análisis de bondad de ajuste para Poisson no es posible pues no existen defectos con parte fraccionaria, además de que Poisson es una distribución Discreta.

La segunda aproximación a la solución del problema fue agrupar los requerimientos por tamaño, encontrado que mayoría de ellos tenían .30 PFSA. En este caso, la hipótesis nula, “los datos siguen una distribución de Poisson”, se hubiera descartado si el valor p hubiera sido menor o igual que un nivel de significancia de .05; pero para este análisis p fue de 0.202. Para confirmar la hipótesis nula, el valor crítico de la distribución Chi-cuadrada de 3 grados de libertad (GL) y una significancia de .05 es de 7.81473 y el estadístico Chi-cuadrada obtenido de los datos analizados es 4.62162, que es menor al valor crítico.

De esta manera, se obtuvo una λ de 2.78571 defectos por cada .30 PFSA. Esta λ , es la media y la varianza de la distribución de defectos que sirvieron de establecer línea base de defectos en la construcción de software en la fábrica de software de la UTL. La figura 1 muestra un resumen del análisis.

Obtenida λ de la distribución de Poisson, se realizó una escalación de λ para el intervalo de un PFSA. De acuerdo con Gallager (2011) “Para un proceso de Poisson de tasa λ , y para cualquier $t > 0$, la FMP [función de masa de probabilidad] para $N(t)$ (por ejemplo, el número de llegadas en $(0,t]$) viene dado por la [ecuación (4) de la] FMP de Poisson”

$$P_{N(t)}(n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \tag{4}$$

Entonces, existe un intervalo $(a, b]$ con una longitud de un PFSA que está dividido en m instantes; y se asume que se ha obtenido una $\lambda t = 2.78571$ para el instante $t_0 = .3$. Si se despeja λ , sabríamos que el factor de escala es 3.33 y λ es 9.2857 para nuestra distribución de Poisson.

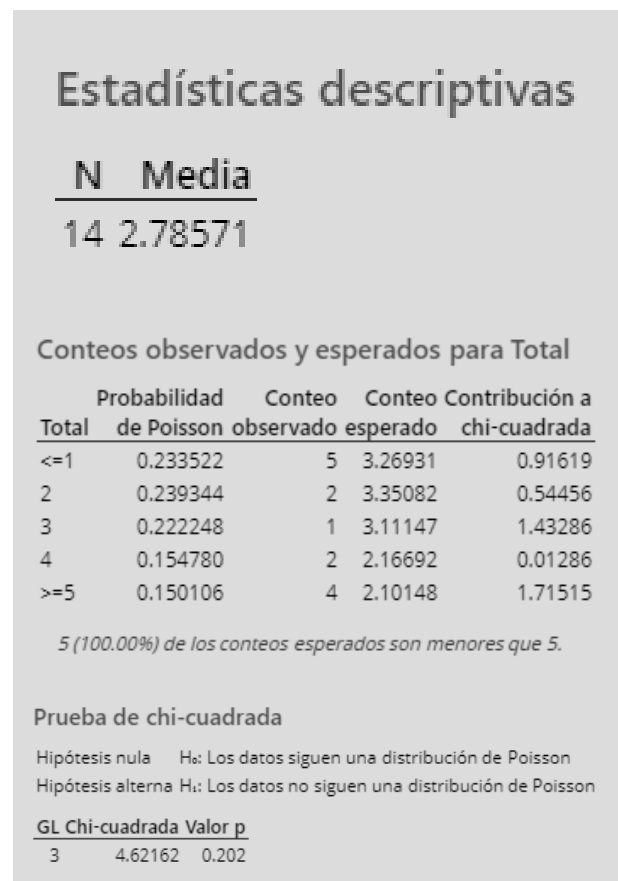


Figura 1 Estadísticas descriptivas de la distribución de defectos en productos de la fábrica de software de la UTL

Conclusiones

Buscar si la distribución de λ (en nuestro caso δ) sigue la distribución Poisson, es un error. La teoría dice que la distribución de Poisson modela el número de eventos que ocurren por unidad (Humberto Gutiérrez, 2013), pero no que la media de la distribución, λ , también va a seguir a distribución Poisson. Se menciona esto porque existe la tendencia a manejar la densidad de defectos de la misma forma en que se maneja la densidad de esfuerzo en construcción de software, debido a que la distribución de la tasa de esfuerzo por unidad se puede modelar usando la distribución estándar, pero la distribución de la tasa de defectos por unidad no se pudo modelar por Poisson, y en todo caso se puede asumir que la distribución muestral de la tasa δ , que son medias de Poisson, será normal.

Aunque la tasa de defectos se pudo modelar como una distribución Poisson, se encontraron varias situaciones que pueden hacer poco práctica esta solución:

- a) Se tuvieron que agrupar los requerimientos de acuerdo con su tamaño en PFSA, esto produjo que, de 27 registros de varios tamaños, solo se usaran 14 con un tamaño de .3 PFSA
- b) Justificar el escalamiento de λ , en el SCAMPI (Standard CMMI Appraisal Method for Process Improvement) para obtener el nivel 4 de CMMI, produjo controversias entre los evaluadores y el equipo CMMI de la UTL, pues no se contaba con justificación matemática presentada en este documento.
- c) Si bien los evaluadores del SCAMPI aceptaron la λt propuesta, no aceptaron que la λ escalada fuera usada como parámetro en simulaciones de Montecarlo para Poisson, ya que no se había obtenido directamente de la herramienta estadística, propuesta por el equipo CMMI de la UTL, Minitab.
- d) Como no se pudo usar la λ escalada como línea base, usar λt implicó hacer la transformación de la tasa δ para cada uno de los requerimientos en particular lo cual se vuelve muy laborioso a medida que el número de requerimientos de un proyecto crece.

Como trabajo a futuro, se buscará normalidad en la densidad de defectos para producir un conjunto de parámetros de distribución estadística que sean más fáciles de manejar en las simulaciones de Montecarlo y fáciles de justificar en un SCAMPI de CMMI.

Referencias

- CMMI Product Team. (2010). *CMMI® for Development, Version 1.3*. Pittsburgh, PA.: Carnegie Mellon University Press.
- Gallager, R. G. (2011, 01 31). *DISCRETE STOCHASTIC PROCESSES*. Retrieved from Chapter 2: Poisson processes: https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-262-discrete-stochastic-processes-spring-2011/course-notes/MIT6_262S11_chap02.pdf
- Humberto Gutiérrez, R. d. (2013). *Control estadístico de la calidad y Seis Sigma*. México: McGRAW-HILL/INTERAMERICANA EDITORES, S.A. DE C.V.
- Minitab, LLC. (2020, 11 24). *Minitab-Soporte*. Retrieved from Interpretar todos los estadísticos y gráficas para Prueba de bondad de ajuste para Poisson: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/goodness-of-fit-test-for-poisson/interpret-the-results/all-statistics-and-graphs/>
- Rincón, L. (2007). *Curso elemental de PROBABILIDAD Y ESTADÍSTICA*. México: Facultad de Ciencias UNAM.