

Aproximación de la Distribución Poisson Compuesta por medio de la Distribución Semi No Paramétrica

Diana Milena Velásquez Sierra

Director:

Norman Diego Giraldo

Ms. en Matemáticas

Profesor Asociado

Trabajo presentado como requisito para optar
al Título de Magíster en Estadística

Escuela de Estadística
Universidad Nacional de Colombia, Sede Medellín

Junio de 2009

Dedicatoria

A todas las personas que me acompañaron en este proceso.

Agradecimientos

Al profesor Norman Diego Giraldo por permitirme trabajar y aprender de él, por sus valiosos comentarios, aportes y sugerencias.

Resumen

En esta tesis se propone y desarrolla una aproximación numérica a las distribuciones compuestas, conocidas también como distribuciones de sumas aleatorias de variables aleatorias; este desarrollo es motivado por la dificultad conocida para evaluar de manera exacta las distribuciones compuestas, debido a que son mezclas infinitas. La aproximación propuesta es realizada a través de los estimadores de densidad Semi No Paramétrica (SNP), cuya distribución fue planteada y desarrollada por Gallant & Nychka [19]; dichos estimadores tiene propiedades que los hacen comparables con los estimadores Kernel y de interés para evaluar la aproximación de las distribuciones compuestas. La aproximación propuesta por los estimadores SNP, se compara con las aproximaciones a las distribuciones compuestas conocidas: la aproximación Normal Power, Gamma Trasladada y Gamma Incompleta, con el fin de evaluar la exactitud del ajuste; para esto se evalúan los casos en que la asimetría es menor y mayor al valor de 2, de manera similar al estudio realizado por Gendron & Crepeau [22] y Chaubey, *et al.* [6], y se calculan las medidas de distancia de los estadísticos utilizados para las pruebas de Kolmogorov Smirnov y Cramer-von Mises, con el fin de evaluar y comparar el ajuste entre las aproximaciones mencionadas.

La construcción de la aproximación SNP se realiza por medio de dos métodos: igualdad de momentos teóricos y estimación por máxima verosimilitud. Se construyeron funciones en R (R Development Core Team, 2008) para la estimación de la distribución SNP, la definición de la aproximación SNP y para las comparaciones entre las aproximaciones.

Finalmente, se obtiene que la aproximación SNP propuesta tiene buen desempeño en el ajuste a las distribuciones compuestas, superando a las demás aproximaciones evaluadas cuando el número de parámetros de la distribución SNP aumenta y cuando se emplea el método de estimación de máxima verosimilitud.

Palabras Clave: Distribuciones compuestas o de sumas aleatorias de variables aleatorias, Máxima verosimilitud, Distribución SNP, Método de igualdad de momentos, Aproximaciones a las Distribuciones Compuestas.

Índice General

1. Marco Teórico	3
1.1. Distribución Compuesta	3
1.1.1. Definición	3
1.1.2. Aproximaciones a Distribuciones Compuestas	7
1.1.3. Distribuciones de prueba para evaluar las aproximaciones	13
1.2. Distribución Semi No Paramétrica (SNP)	16
1.2.1. Definición del estimador de densidad SNP	16
1.2.2. Propiedades del estimador SNP	19
1.2.3. Aplicaciones del estimador de densidad SNP	20
1.3. Medidas de distancia para comparar las aproximaciones	21
2. Método de Aproximación SNP	23
2.1. Aproximación SNP a las distribuciones compuestas	24
2.1.1. Condiciones para emplear el estimador de densidad SNP	25
2.2. Aproximación SNP por el Método Máxima Verosimilitud	28
2.3. Aproximación SNP por el Método de Igualdad de Momentos	29
2.3.1. Aproximación SNP cuando $X \sim IG$	31

2.3.2.	Aproximación SNP cuando $X \sim \text{LogN}$	32
3.	Análisis de las Aproximaciones a la Distribución Compuesta	34
3.1.	Resultados de la aproximación SNP a la $PC(\lambda; IG(\mu, \sigma))$	36
3.1.1.	Caso de Asimetría $\gamma_S < 2$, cuando $S \sim PC(\lambda; IG(\mu, \sigma))$	36
3.1.2.	Caso de asimetría $\gamma_S > 2$, cuando $S \sim PC(\lambda; IG(\mu, \sigma))$	41
3.2.	Resultados de la aproximación SNP a la $PC(\lambda; LN(\mu, \sigma))$	46
3.2.1.	Caso de asimetría $\gamma_S < 2$, cuando $S \sim PC(\lambda; LN(\mu, \sigma))$	46
3.2.2.	Caso de asimetría $\gamma_S > 2$, cuando $S \sim PC(\lambda; LN(\mu, \sigma))$	51
4.	Aplicaciones	57
4.1.	Aplicación al Riesgo Operacional	57
4.2.	Modelos de Riesgo Colectivo	59
5.	Conclusiones	61
A.	Documentación y Código Funciones	63

Índice de Figuras

3.1. Gráficas de densidad: Distribución Inversa Gaussiana $IG(0.2, 2)$ y Distribución Compuesta $PC(20; IG(0.2, 2))$, con $\gamma_S < 2$. Ajuste Kernel y SNP_4	36
3.2. Comparación de las Aproximaciones a la Distribución Compuesta, cuando $S \sim PC(20; IG(0.2, 2))$ y $\gamma_S < 2$	37
3.3. Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(20; IG(0.2, 2))$, con tamaño de muestra $n = 2000$ y $\gamma_S < 2$	38
3.4. Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 2000$ y $\gamma_S < 2$	39
3.5. Gráficas de ajuste de los estimadores SNP con 4, 6, 8 y 10 parámetros de la distribución $PC(20; IG(0.2, 2))$, con tamaño de muestra $n = 6000$ y $\gamma_S < 2$	40
3.6. Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 6000$ y $\gamma_S < 2$	41
3.7. Gráficas de densidad: Distribución Inversa Gaussiana $IG(4, 0.5)$ y Distribución Compuesta $PC(10; IG(4.0.5))$, con $\gamma_S > 2$. Ajuste Kernel y SNP_4	42
3.8. Comparación de Aproximaciones a las Distribuciones Compuestas, cuando $S \sim PC(10; IG(4.0.5))$ y $\gamma_S > 2$	43
3.9. Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 2000$ y $\gamma_S > 2$	43
3.10. Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 2000$ y $\gamma_S > 2$	44
3.11. Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(10; IG(4, 0.5))$, con tamaño de muestra $n = 6000$ y $\gamma_S > 2$	45

3.12. Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 6000$ y $\gamma_S > 2$	46
3.13. Gráficas de densidad: Distribución Log-Normal $LN(1, 0.2)$ y Distribución Compuesta $PC(10; LN(1, 0.2))$, con $\gamma_S < 2$. Ajuste Kernel y SNP_4	47
3.14. Comparación de Aproximaciones a las Distribuciones Compuestas, cuando $S \sim PC(10; LN(1, 0.2))$ y $\gamma_S < 2$	47
3.15. Gráficas de densidad de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(10; LN(1, 0.2))$, con tamaño de muestra $n = 2000$ y $\gamma_S < 2$	49
3.16. Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 2000$ y $\gamma_S < 2$	49
3.17. Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(10; LN(1, 0.2))$, con tamaño de muestra $n = 6000$ y $\gamma_S < 2$	50
3.18. Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 6000$ y $\gamma_S < 2$	51
3.19. Gráficas de densidad: Distribución Log-Normal $LN(3, 1.3)$ y Distribución Compuesta $PC(15; LN(3, 1.3))$. Ajuste Kernel y SNP_4	52
3.20. Comparación de Aproximaciones a las Distribuciones Compuestas, cuando $S \sim PC(15; LN(3, 1.3))$ y $\gamma_S > 2$	53
3.21. Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(15; LN(3, 1.3))$, con tamaño de muestra $n = 2000$ y $\gamma_S > 2$	53
3.22. Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 2000$ y $\gamma_S > 2$	54
3.23. Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(15; LN(3, 1.3))$, con tamaño de muestra $n = 6000$ y $\gamma_S > 2$	55
3.24. Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 6000$ y $\gamma_S > 2$	55

Índice de Tablas

3.1. Medidas de ajuste de las Aproximaciones a la distribución $PC(20; IG(0.2, 2))$ cuando $\gamma_S < 2$	38
3.2. Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(20; IG(0.2, 2))$, cuando $\gamma_S < 2$	39
3.3. Medidas de ajuste de las Aproximaciones a la distribución $PC(20; IG(0.2, 2))$, cuando $\gamma_S < 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$	40
3.4. Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(20; IG(0.2, 2))$ cuando $\gamma_S < 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$	41
3.5. Medidas de ajuste de las Aproximaciones a la distribución $PC(10; IG(4.0.5))$ cuando $\gamma_S > 2$	42
3.6. Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(20; IG(0.2, 2))$, cuando $\gamma_S < 2$	44
3.7. Medida de ajuste de las Aproximaciones a la distribución $PC(10; IG(4, 0.5))$ cuando $\gamma_S > 2$, y los tamaños de muestra son: $n = 4000$ y $n = 6000$	45
3.8. Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros, a la distribución $PC(10; IG(4, 0.5))$ cuando $\gamma_S < 2$, y los tamaños de muestra son $n = 4000$ y $n = 6000$	46
3.9. Medidas de ajuste de las Aproximaciones a la distribución $PC(10; LN(1, 0.2))$ cuando $\gamma_S < 2$	48
3.10. Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(10; LN(1, 0.2))$, cuando $\gamma_S < 2$	48

3.11. Medidas de ajuste de las Aproximaciones a la distribución $PC(10; LN(1, 0.2))$, cuando $\gamma_S < 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$	50
3.12. Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(10; LN(1, 0.2))$ cuando $\gamma_S < 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$	51
3.13. Medidas de ajuste de las Aproximaciones a la distribución $PC(15; LN(3, 1.3))$ cuando $\gamma_S > 2$	52
3.14. Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(15; LN(3, 1.3))$. cuando $\gamma_S > 2$	54
3.15. Medidas de ajuste de las Aproximaciones a la distribución $PC(15; LN(3, 1.3))$, cuando $\gamma_S > 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$	54
3.16. Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(15; LN(3, 1.3))$ cuando $\gamma_S < 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$	56

Introducción

La distribución compuesta, conocida también como la suma aleatoria de variables aleatorias, se define como: $S = \sum_{j=1}^N X_j$, donde N y X_j son variables aleatorias independientes, ha motivado el desarrollo de múltiples aproximaciones, debido a que su función de distribución exacta presenta serios problemas numéricos y de convergencia.

La complejidad de la evaluación exacta de la distribución compuesta y el estudio y desarrollo de aproximaciones se evidencia en los siguientes artículos: Seal [40], Pentikainen [33], Pentikainen [35], Gendron & Crepeau [22], Panjer [32], entre otros; en los cuales se revisan y proponen algunas aproximaciones que involucran distribuciones conocidas como la Normal y la Gamma, e incluso métodos recursivos que parten de la discretización, como el método de Panjer.

En esta tesis se plantea una aproximación a la distribución compuesta por medio de los estimadores de densidad Semi No Paramétricos (SNP), cuya distribución, especificación y propiedades son expuestos en Gallant & Nychka [19]; los resultados al emplear los estimadores SNP son muy similares a los obtenidos por los estimadores kernel, esta conclusión fue obtenida por Fenton & Gallant [16] y se evidencia en los resultados de este trabajo.

La aproximación propuesta en esta tesis se desarrolla por los métodos de igualdad de momentos, este método ha sido empleado por varios autores para proponer aproximaciones a las distribuciones compuestas como Chaubey, *et al.* [6], y el método de estimación por máxima verosimilitud. La distribución desarrollada será llamada en adelante aproximación SNP.

La aproximación SNP propuesta se compara con algunas de las aproximaciones desarrolladas como: Normal Power, Gamma Traslada y Gamma Incompleta, entre otras, las cuales son revisadas por Gendron & Crepeau [22], Chaubey, *et al.* [6], entre otros.

Con el fin de analizar la calidad de la aproximación SNP propuesta a los datos generados por la distribución compuesta se emplean las medidas de distancia empleadas para la construcción de los estadísticos de prueba de Kolmogorov - Smirnov y de Cramer-von Mises.

En el Capítulo 1 se presentan los conceptos básicos de las distribuciones compuestas, algunas de las aproximaciones propuestas por otros autores, las distribuciones de N y X empleadas para evaluar la aproximación propuesta, y la distribución SNP y sus propiedades. En el Capítulo 2 se plantea la aproximación a la distribución compuesta por medio de la distribución SNP, bajo el supuesto de que $N \sim Poisson(\lambda)$, por lo tanto la distribución a la cual se detalla la aproximación es a la distribución Poisson Compuesta. En el Capítulo 3 se evalúa el ajuste de las aproximaciones SNP, Normal Power, Gamma Traslada y Gamma Incompleta, a la distribución Poisson Compuesta en diferentes escenarios establecidos por la asimetría de la distribución compuesta. En el Capítulo 4 se plantean posibles aplicaciones para la aproximación SNP. Al final del documento se encuentran los programas implementados en \mathbb{R} para el desarrollo de esta investigación.

Capítulo 1

Marco Teórico

1.1. Distribución Compuesta

1.1.1. Definición

Una variable aleatoria S se denomina una suma aleatoria de variables aleatorias si puede ser expresada de la siguiente manera:

$$S = X_1 + X_2 + \dots + X_N = \sum_{j=1}^N X_j \quad (1.1)$$

donde N es una variable aleatoria. Por lo tanto, S es la suma de variables aleatorias independientes e idénticamente distribuidas (i.i.d.), donde el número de términos de la suma está distribuido aleatoriamente y es independiente de los sumandos; de la ecuación (1.1) se puede concluir que $S = 0$ cuando $N = 0$.

Los siguientes supuestos se asumen en la definición de S :

1. N es una variable aleatoria discreta con valores en $\{0, 1, 2, \dots\}$.
2. X_1, X_2, \dots forman una sucesión de variables aleatorias i.i.d.. Para desarrollo de este trabajo se asumirá adicionalmente que son variables positivas continuas, es decir, la función de distribución acumulada es tal que $F_X(0) = 0$, y existe su derivada que se asume continua, $F'_X(x) = f_X(x)$.
3. Las variables N, X_1, X_2, \dots, X_N son mutuamente independientes.

La función de distribución acumulada de la distribución compuesta S está dada por:

$$\begin{aligned}
 F_S(x) &= P(S \leq x) \\
 &= \sum_{n=0}^{\infty} P(N = n)P(S = x|N = n) \\
 &= \sum_{n=0}^{\infty} p_n P(X_1 + X_2 + \dots + X_n \leq x|N = n) \\
 &= \sum_{n=0}^{\infty} p_n F^{*n}(x)
 \end{aligned} \tag{1.2}$$

donde $p_n = P(N = n)$ representa la probabilidad de ocurrencia de n eventos y $F^{*n}(x) = P(X_1 + X_2 + \dots + X_n \leq x)$ es la convolución n -ésima de la función F consigo misma. En general, la convolución de F está dada por las siguientes ecuaciones:

Si $n = 0$

$$F^{*n}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

Si $n = 1, 2, \dots$

$$F^{*n}(x) = \int_{-\infty}^{\infty} F^{*(n-1)}(x-y)dF(y)$$

Por lo tanto, la distribución compuesta es una mezcla infinita de distribuciones, con $F^{*0}(x) \equiv 1$, por lo que se convierte en la principal motivación para buscar aproximaciones a la distribución compuesta $F_S(x)$.

La función de distribución de probabilidad de S , definida como $f_S(x)$, se obtiene a través de la ecuación (1.2) y está dada por la siguiente ecuación:

$$f_S(x) = \sum_{n=0}^{\infty} p_n f^{*n}(x) \tag{1.3}$$

donde,

$$f^{*n}(x) = \begin{cases} \frac{d}{dx} F(x-y), & y \leq x \\ 0, & \text{en otros casos} \end{cases}$$

La función generadora de momentos (f.g.m.) para la distribución compuesta S , dada en Bowers [4], pag. 369, es:

$$M_S(t) = M_N [\log M_X(t)] \tag{1.4}$$

donde M_N y M_X son las f.g.m. de las variables aleatorias N y X , respectivamente. En adelante, se hará referencia a la sucesión de variables aleatorias: X_1, \dots, X_N a través de la variable aleatoria X .

Klugman, *et al.* [27], pag. 136, enumera las ventajas que existe al modelar la distribución de la variable aleatoria N por separado de la distribución de la variable aleatoria X , desde un punto de vista actuarial. Algunos autores mencionan ciertas distribuciones que son comunmente utilizadas para explicar las variables aleatorias N y X ; por ejemplo, Escalante [14] sugiere emplear las distribuciones Poisson, Binomial, Binomial Negativa y la Geométrica para explicar la variable aleatoria N ; Chaubey, *et al.* [6] emplea las distribuciones Inversa Gaussiana, Gamma, Pareto y Weibull para explicar el comportamiento la variable X . Rolski [39], pag. 103, menciona tres casos de interés de distribuciones compuestas en aplicaciones actuariales, generados cuando la variable aleatoria N es explicada por las distribuciones Poisson, Binomial Negativa y Geométrica, en cada uno de los casos se obtiene distribución Poisson Compuesta, Pascal o Binomial Negativa Compuesta y distribución Geométrica Compuesta, respectivamente.

Ejemplo 1. *Caso particular - Distribución Poisson Compuesta.* Considere que N se distribuye Poisson y suponga que la variable aleatoria X tiene una distribución continua, por lo tanto, se dice que la distribución de S es una distribución Poisson Compuesta. La función de distribución de probabilidad de la Poisson Compuesta se define a partir de la ecuación (1.3), donde p_n es la función de densidad de la distribución Poisson:

$$p_n = \frac{e^{-\lambda} \lambda^n}{n!}$$

Por lo tanto, la función de densidad de probabilidad de la Poisson Compuesta, $f_S(x)$, esta dada por:

$$f_S(x) = \sum_{n=0}^{\infty} \frac{e^{-\lambda} \lambda^n}{n!} f^{*n}(x) \quad (1.5)$$

La f.g.m. de la distribución Poisson Compuesta tiene la siguiente expresión (derivada de la ecuación (1.4)):

$$M_S(t) = \exp[\lambda(M_X(t) - 1)] \quad (1.6)$$

Por lo tanto, los primeros cuatro momentos centrales de la distribución Poisson Compuesta estan dados por el siguiente conjunto de ecuaciones:

$$\begin{aligned} \mu_{S1} &= 0 \\ \mu_{S2} &= \lambda E[X^2] \\ \mu_{S3} &= \lambda E[X^3] \end{aligned}$$

$$\mu_{S4} = \lambda E[X^4] + 3\lambda^2 (E[X^2])^2 \quad (1.7)$$

donde $E[X^k]$ es el k-ésimo momento de la variable aleatoria X , y la media y la varianza de la distribución Poisson Compuesta estan dadas por las siguientes ecuaciones:

$$\begin{aligned} E[S] &= \lambda E[X] \\ Var(S) &= \lambda E[X^2] \end{aligned} \quad (1.8)$$

Panjer & Willmot [31], pag. 171, describen la obtención de los momentos de la distribución Poisson Compuesta. Bowers [4] y Escalante & Arango [13], ofrecen más detalles de la distribución compuesta y de las convoluciones. ||

En muy pocos casos la distribución compuesta puede ser expresada de forma analítica, el siguiente ejemplo presenta uno de estos casos.

Ejemplo 2. (ver Rolski [39], pag. 104, example 1) Suponga que $N \sim Poisson(\lambda)$ y $X \sim i.i.d. Exp(\delta)$. Entonces la función de distribución de la variable compuesta S es de la forma:

$$F_S(x) = \sum_{n=0}^{\infty} e^{-\lambda} (\lambda^n / n!) F_X^{*n}(x),$$

donde $F_X^{*n}(x) = \delta^n \int_0^x e^{-\delta t} t^{n-1} dt / (n-1)!$, para $n = 1, 2, \dots$ La función de distribución $F_S(x)$ tiene probabilidad positiva en $x = 0$, correspondiente a la probabilidad $P(S = 0) = e^{-\lambda} = p_0$, y es derivable en $(0, \infty)$, con densidad dada por:

$$f_S^{(1)}(x) = e^{-(\lambda+x\delta)} \sum_{n=1}^{\infty} (x\lambda\delta)^n / (xn!(n-1)!), x > 0$$

Luego, es posible comprobar que esta última expresión se puede definir mediante una función Bessel modificada, $I_1(x)$, la cual está dada por:

$$I_1(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{2k+1}}{k! \Gamma(k+2)}$$

de tal forma que la densidad $f_S(x)$ se puede expresar como:

$$f_S(x) = p_0 \delta(x) + (1 - p_0) f_S^{(1)}(x) = p_0 \delta(x) + (1 - p_0) e^{-(\lambda+x\delta)} \sqrt{\lambda\delta/x} I_1(2\sqrt{\lambda\delta x}),$$

donde $\delta(x)$ es la función Delta de Dirac, que corresponde a la parte discreta de la densidad. La función $\delta(x)$ tiene un valor nulo en cualquier otro punto diferente de $x = 0$. ||

Este ejemplo muestra que en general se puede escribir la función de densidad de la variable S como una mezcla discreta - continua de la forma: $f_S(x) = p_0 \delta(x) + (1 - p_0) f_S^{(1)}(x)$, donde $f_S^{(1)}(x)$ es una función continua en $[0, \infty)$. Por tanto, el evento $(S = 0)$ no es relevante en la discusión de las

aproximaciones y en el resto de este trabajo se asumirá que se está trabajando con la parte continua de la distribución y las aproximaciones se referirán siempre a la aproximación de la distribución acumulada $F_S^{(1)}(x) = \int_0^x f_S^{(1)}(t)dt$ que se seguirá escribiendo, sin peligro de confusión, como $F_S(x)$.

La distribución compuesta dada en el Ejemplo 2 y en el caso en el que la distribución de N es una Binomial Negativa y la distribución de X una Exponencial, (ver Ejemplo 2 en [39], pag. 104), son los únicos ejemplos conocidos en los cuales es posible expresar la distribución compuesta mediante una fórmula analítica. Para los casos en los que se asume que X tiene distribución absolutamente continua es muy difícil o imposible encontrar tal distribución. Esto explica el porqué los métodos de aproximación, incluyendo simulación, han sido un área activa de investigación desde la década de 1960. En la siguiente sección se detallan algunos desarrollos encaminados a la obtención de aproximaciones.

1.1.2. Métodos de Aproximación a la Distribución Compuesta

Como se mencionó anteriormente, en muchos casos la distribución compuesta presenta dificultades en el momento de ser estimada y evaluada a través de una solución analítica, lo que ha motivado el desarrollo de métodos numéricos, algoritmos recursivos, técnicas asintóticas y métodos de aproximación numérica.

Las aproximaciones para las distribuciones compuestas han sido estudiadas por muchos años y analizadas por varios autores, algunos de ellos son: Bowers [3], Seal [40], Pentikainen [33], Pentikainen [34], Pentikainen [35], Gendron & Crepeau [22], Chaubey, *et al* [6], Barbour & Chryssaphinou [2], Panjer [32], entre otros que revisan y proponen aproximaciones para las distribuciones compuestas.

Varios métodos han sido desarrollados para lograr una aproximación lo más exacta posible a las distribuciones compuestas, es decir a $F_S(x)$, algunos de ellos son: la aproximación Normal, la aproximación Normal Power (NP), que es una variante de la aproximación Normal, la aproximación Gamma Incompleta y Gamma Traslada, el método recursivo de Panjer, entre otras evaluaciones numéricas. Algunas de estas aproximaciones fueron construidas mediante la igualdad de momentos, método que será evaluado más adelante y por medio del cual se hace una propuesta de estimación a la distribución compuesta a través de la distribución Semi No Paramétrica. Los métodos de aproximación que se emplean en esta tesis para comparar la calidad del ajuste con la aproximación que se sustenta a través de la distribución Semi No Paramétrica, son descritos a continuación. Al final se realiza una breve revisión de otros métodos que se han empleado para construir aproximaciones a las distribuciones compuestas.

Aproximación Normal Power

La aproximación Normal emplea la función de distribución Normal a través de la estandarización de la variable aleatoria X :

$$F_S(x) \approx \Phi(Z) = \Phi\left(\frac{s - \mu_S}{\sigma_S}\right)$$

donde μ_S y σ_S son la media y desviación estándar estimadas de S , respectivamente, y que pueden ser derivados de la ecuación (1.4). En el caso en el que F_S sigue una distribución Poisson Compuesta, μ_S y σ_S pueden ser estimadas a través de las expresiones dadas en (1.9), según Chaubey, *et al.* [6], pag. 217.

La aproximación Normal Power (NP) modifica la definición de la aproximación Normal por medio de un ajuste a la variable estandarizada Z :

$$F_S(x) \approx \Phi(z + \Delta z) = \Phi\left[\sqrt{\frac{9}{\gamma_S^2} + 1} + \frac{6z}{\gamma_S} - \frac{3}{\gamma_S}\right] \quad (1.9)$$

donde γ_S es la asimetría de la distribución compuesta S y $z = (s - \mu_S)/\sigma_S$; en el caso en el que S tiene una distribución Poisson Compuesta, ver Ejemplo 1., la asimetría está dada por:

$$\gamma_S = \frac{\lambda E[X^3]}{(\lambda E[X^2])^{3/2}} \quad (1.10)$$

Pentikainen [33], pag. 281, aclara que si Δz es pequeño se puede utilizar la siguiente expresión para la aproximación NP:

$$F_S(x) \approx \Phi\left(z - \frac{\gamma_S}{6}(z^2 - 1)\right)$$

Según Gendron & Crepeau [22], pag. 254, ésta aproximación es buena mientras la asimetría de la distribución compuesta no exceda en valor a 2. Para que la aproximación NP funcione adecuadamente se sugiere que la distribución de la variable aleatoria X no sea muy asimétrica.

Pentikainen [33] defiende el uso y la implementación de esta aproximación y demuestra su buen desempeño numéricamente, mientras Seal [40] critica su uso y sustenta que la aproximación Gamma Incompleta arroja mejores resultados y adicionalmente la defiende ya que la precisión no depende del valor de la asimetría como en la aproximación NP. Por otro lado, Ramsay [38], comenta en su artículo que la aproximación NP es bastante precisa cuando $z > 1$ y la exactitud o precisión de esta aproximación disminuye a medida que γ_S aumenta.

Aproximación Gamma Incompleta

Seal [40], pag. 214, describe la aproximación Gamma Incompleta a través de la siguiente expresión:

$$F_S(x) \approx \frac{1}{\Gamma(\alpha)} \int_0^{\alpha+z\sqrt{\alpha}} e^{-y} y^{\alpha-1} dy \equiv P(\alpha, \alpha + z\sqrt{\alpha})$$

donde Γ es la función gamma, $\alpha = \frac{4}{\gamma^2}$ y z es definida en la descripción de la aproximación anterior.

Seal [40], muestra la superioridad en ajuste de esta aproximación sobre la aproximación NP. Seal [41], comenta de nuevo acerca de la calidad en el ajuste de dicha aproximación y la emplea para construir una aproximación a la distribución de probabilidad de ruina. Gendron & Crepeau [22], concluyen que la aproximación Gamma Incompleta produce mejores estimaciones a la $F_S(x)$ que las demás aproximaciones analizadas, incluso cuando el valor de la asimetría es relativamente alto.

La función de distribución Gamma Incompleta ha sido ampliamente empleada para la construcción de aproximaciones dada su flexibilidad. Bowers [3] emplea la función de la Gamma Incompleta para construir una aproximación a la distribución compuesta por medio de sumas de funciones ortogonales. La aproximación se describe en la siguiente fórmula, la cual emplea los primeros cinco momentos de la distribución compuesta:

$$\begin{aligned} F_S(x) = & \Gamma(x, \alpha) - A \left[\frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)} - \frac{2x^{\alpha+1} e^{-x}}{\Gamma(\alpha+2)} + \frac{x^{\alpha+2} e^{-x}}{\Gamma(\alpha+3)} \right] \\ & + B \left[\frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)} - \frac{3x^{\alpha+1} e^{-x}}{\Gamma(\alpha+2)} + \frac{3x^{\alpha+2} e^{-x}}{\Gamma(\alpha+3)} - \frac{x^{\alpha+3} e^{-x}}{\Gamma(\alpha+4)} \right] \\ & - C \left[\frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)} - \frac{4x^{\alpha+1} e^{-x}}{\Gamma(\alpha+2)} + \frac{6x^{\alpha+2} e^{-x}}{\Gamma(\alpha+3)} - \frac{4x^{\alpha+3} e^{-x}}{\Gamma(\alpha+4)} + \frac{x^{\alpha+4} e^{-x}}{\Gamma(\alpha+5)} \right] \end{aligned}$$

donde $\Gamma(x, \alpha)$ es la función Gamma Incompleta y está dada por:

$$\Gamma(x, \alpha) = \frac{1}{\Gamma(\alpha)} \int_0^x z^{\alpha-1} e^{-z} dz$$

y

$$\begin{aligned} A &= \frac{\mu_3 - 2\alpha}{3!} \\ B &= \frac{\mu_4 - 12\mu_3 - 3\alpha^2 + 18\alpha}{4!} \\ C &= \frac{\mu_5 - 20\mu_4 - (10\alpha - 120)\mu_3 + 60\alpha^2 - 144\alpha}{5!} \end{aligned}$$

donde, μ_k hace referencia el momento k-ésimo de la distribución compuesta.

Adicionalmente existe una restricción, la cual es que el primer y el segundo momento de la distribución compuesta deben de ser iguales, para esto, Bowers [3], pag. 127, sugiere construir una

variable aleatoria que permita escalar la variable original, con el fin de obtener los mismos valores en los dos primeros momentos.

Aproximación Gamma Traslada

Esta aproximación está basada en la distribución Gamma con tres parámetros, la cual consiste en aproximar la distribución compuesta S por medio de $k+Y$, donde k es el parámetro de traslación y Y una variable aleatoria con distribución $Gamma(\alpha, \beta)$, con función de distribución de probabilidad dada por:

$$f(y) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}$$

La aproximación es construida por el método de igualdad de momentos, que consiste en la solución del sistema de ecuaciones conformado por los primeros tres momentos centrales de la Gamma Traslada y los momentos centrales de la distribución compuesta S .

Con el fin de ilustrar el desarrollo de esta aproximación, se establece el sistema de ecuaciones con la distribución Gamma Traslada y la distribución Poisson Compuesta, considerando que la variable aleatoria N sigue una distribución Poisson, para tener más detalle de esta distribución compuesta ver Ejemplo 1. A través del sistema de ecuaciones se obtiene la estimación de los parámetros de la Gamma Traslada (k, α, β) en función de los parámetros de la distribución Poisson Compuesta S . La aproximación Gamma Traslada está descrita en McNeil *et al.* [29], pag. 477, y Bowers, *et al.* [4], pag. 388.

El sistema de ecuaciones resultante al igualar los primeros tres momentos de ambas distribuciones es:

$$\begin{aligned} k + \frac{\alpha}{\beta} &= \lambda E[X] \\ \frac{\alpha}{\beta^2} &= \lambda E[X^2] \\ \frac{2}{\sqrt{\alpha}} &= \frac{E[X^3]}{\sqrt{\lambda E[X^2]^3}} \end{aligned}$$

donde $E(X^k)$ corresponde al k-ésimo momento de la variable aleatoria X .

Resolviendo el sistema planteado, se obtienen los parámetros de la distribución Gamma Traslada en función de los parámetros de la distribución Poisson Compuesta:

$$k = \lambda E[X] - \frac{2\lambda(E[X^2])^2}{E[X^3]}$$

$$\alpha = \frac{4\lambda(E[X^2])^3}{(E[X^3])^2}$$

$$\beta = \frac{2E[X^2]}{E[X^3]}$$

La aproximación Gamma Traslada no se encuentra implementada en el R, sin embargo, su implementación se puede realizar de manera sencilla.

Revisión de otros Métodos de Aproximación

A continuación se presentan algunas de las revisiones realizadas sobre las aproximaciones a las distribuciones compuestas:

- Seal [40], justifica las ventajas del ajuste de la aproximación Gamma Incompleta a la distribución compuesta $F_S(x)$, sobre la aproximación NP. En el mismo año Pentikainen [33], hace referencia al artículo de Seal [40] y mediante una evaluación concluye que no se encuentran diferencias significativas en la exactitud entre los métodos NP y Gamma Incompleta, además concluye que la aproximación Normal es aceptable sólo cuando la distribución de X no es muy asimétrica.
- Diez años más adelante, Pentikainen [35] realiza una comparación de posibles aproximaciones a la función de distribución compuesta $F_S(x)$ y comenta acerca de los métodos utilizados para la construcción de las aproximaciones; el método de igualdad de momentos ha sido empleado para definir algunas de las aproximaciones analizadas, Pentikainen [35] también comenta acerca de algunas dificultades de éste método en casos específicos. La aproximación que analiza es una modificación de la aproximación de Haldane y las compara con la aproximación NP y la fórmula de Wilson - Hilferty, que es un método exacto, ambas aproximaciones son descritas en su artículo.
- Gendron & Crepeau [22], evalúan varias aproximaciones a la distribución compuesta cuando $N \sim Poisson(\lambda)$ y $X \sim IG(\mu, \sigma)$. El ajuste de las siguientes aproximaciones fue analizado: la aproximación Normal, NP, de Edgeworth, de Esscher y la Gamma Incompleta; en éste estudio se compara la calidad de estas aproximaciones según la asimetría de la distribución Poisson Compuesta y concluyen que la aproximación Gamma Incompleta es la que mejor ajusta a los casos evaluados, además que el desempeño de esta aproximación no es tan sensible al valor de la asimetría como el de las demás aproximaciones.
- Chaubey, *et al.* [6] proponen realizar un ajuste a la distribución compuesta $F_S(x)$ por medio de la distribución Inversa Gaussiana, motivados por la capacidad de esta distribución para modelar datos asimétricos; y comparan sus resultados con otras aproximaciones ya existentes.

Además, proponen otra aproximación que emplea una mezcla entre las distribuciones Gamma e Inversa Gaussiana (IG). En este artículo se consideran las distribuciones Poisson, Binomial y Binomial Negativa para la distribución de p_n y las distribuciones IG, Gamma, Pareto y Weibull para la distribución $F_X(x)$. Las aproximaciones propuestas en este artículo son construidas mediante la igualdad de los momentos de la distribución compuesta y los momentos de la distribución IG o bien, los correspondientes a la mezcla entre las distribuciones Gamma e IG. Mediante cálculos numéricos concluyen que la aproximación IG propuesta arroja resultados muy similares a los de la aproximación Gamma Incompleta, sin embargo, los resultados de la aproximación mezclada (Gamma e IG) proporciona un mejor ajuste.

- Klugman, *et al.* [27], comentan acerca de algunos métodos de aproximación que han sido empleados para las distribuciones compuestas. El método de los momentos es conocido y muy empleado en las aproximaciones, además es simple y fácil de aplicar, sin embargo, es criticado en este texto ya que, dependiendo de la distribución que se emplee para la construcción de la aproximación, depende el buen ajuste del método, sobretodo en la cola derecha de la distribución. Otro método analizado es el método del cálculo directo, como se comentó anteriormente, se menciona que éste medio presenta dificultades en la evaluación de la n -ésima convolución de la distribución de la variable aleatoria X y que no en todos los casos es viable. Otro método revisado por ellos es el método recursivo, el cuál aclaran que es más efectivo en un 99.9% en la reducción del tiempo de la estimación.
- Panjer [32], pag. 174, describe detalladamente el método recursivo de Panjer, el cual es un algoritmo que permite la estimación de la $F_S(x)$ mediante la discretización de la función de distribución de $F(x)$. Existen básicamente dos requisitos para calcular la aproximación a la distribución compuesta: el primer requisito es que la distribución de N sea discreta (Poisson, Binomial Negativa, Geométrica, entre otras), y el segundo requisito es que la distribución de las X_j pueda ser discretizada. Escalante [14] obtiene la distribución de probabilidad de la F_S a través del método de Panjer. Además, Sundt [43] hace una revisión amplia y detallada del método recursivo de Panjer.

Como se mencionó anteriormente en las referencias presentadas se pueden hallar otras aproximaciones propuestas para las distribuciones colectivas, como la aproximación de Edgeworth que está basada en la función de distribución acumulada de la Normal Estándar; la aproximación Esscher, que según Gendron & Crepeau [22], arroja mejores resultados en el ajuste de la cola que la aproximación de Edgeworth; la aproximación Normal, que fue parcialmente descrita en la definición de la aproximación NP, entre otras. Otros métodos con base en la Transformada Rápida de Fourier y con base en el método de Stein son revisados por Panjer [32] y Barbour & Chryssaphinou [2]. Es posible realizar la simulación de la distribución compuesta de manera sencilla, adicionalmente en la librería actuar de R hay implementadas algunas de las aproximaciones definidas anteriormente.

1.1.3. Distribuciones de prueba para evaluar las aproximaciones

Hasta el momento es claro que la precisión de algunas de las aproximaciones dependen de las distribuciones que modelan el comportamiento de las variables aleatorias N y X . Por tal motivo en esta sección se exponen las distribuciones que se emplean para el desarrollo de la aproximación propuesta en esta tesis.

Las distribuciones más comunes para modelar la variable aleatoria N fueron mencionadas anteriormente cuando se hacía referencia al artículo de Escalante [14]. Las propiedades de emplear cada una de las distribuciones discretas sugeridas para modelar la variable aleatoria N (Binomial, Poisson, entre otras) son mencionadas por varios autores como Panjer & Willmot [31], Bowers [4] y Panjer [32]. El empleo de la distribución Poisson es muy frecuente y las siguientes propiedades lo sustentan:

- La suma de n variables aleatorias independientes Poisson se distribuye Poisson con valor de parámetro igual a la suma de los n parámetros de las variables iniciales. Esta propiedad se define a través de un teorema por Panjer [32], pag. 109.
- Bajo el supuesto que N hace referencia al número de eventos y se distribuye Poisson con parámetro λ , suponga que cada evento puede ser clasificado en una de m categorías, cada una con probabilidades p_1, \dots, p_m independientes de los otros eventos. Entonces el número de eventos N_1, \dots, N_m , correspondiente a las categorías $1, \dots, m$, son variables aleatorias Poisson mutuamente independientes con medias $\lambda p_1, \dots, \lambda p_m$

Por lo tanto, se empleará la distribución Poisson para modelar la variable aleatoria N , ya que se ha considerado como una de las distribuciones más sencillas y flexibles de emplear. De esta manera no se descarta el empleo de otras distribuciones discretas para el desarrollo de aproximaciones a las distribuciones compuestas, simplemente se deja para trabajos futuros.

Algunas de las distribuciones empleadas en la literatura para determinar el comportamiento de la variable aleatoria X son mencionadas a continuación:

- Gendron & Crepeau [22], utilizan la distribución Poisson para modelar la variable aleatoria N y las distribuciones Inversa Gaussiana para explicar la distribución de la variable aleatoria X .
- Chaubey *et al.* [6], emplean las distribuciones Gamma, Pareto, Weibull e Inversa Gaussiana para modelar la variable aleatoria X . Además, mencionan que una distribución ampliamente empleada en actuaría y riesgos es la distribución Log-Normal.

- Otros autores como Consul & Jain [9], Pentikainen [35], Bowers [4] y Embrechts & Puccetti [12], justifican el uso de la distribución Log-Normal para modelar la variable aleatoria X . Estos autores emplean la distribución compuesta en sus artículos para casos prácticos relacionados con temas de seguros, riesgos y actuaría.

A continuación se describen las distribuciones continuas empleadas para la variable aleatoria X en el desarrollo de esta tesis:

- **Distribución Inversa Gaussiana**

Folks & Chhikara [17], hacen una breve reseña acerca de la distribución Inversa Gaussiana (IG) donde se remontan al año 1915 cuando Schrödinger da el primer paso en la construcción de esta distribución y mencionan la evolución que tuvo ésta en los años 1945 y 1947, en los que Tweedie y Wald, respectivamente, aportaron avances para llegar a lo que hoy conocemos como la distribución Inversa Gaussiana.

La forma de la distribución esta dada por:

$$f(x; \mu, \sigma) = \left(\frac{\sigma}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left\{\frac{-\sigma(x - \mu)^2}{2\mu^2 x}\right\}, x > 0 \quad (1.11)$$

donde μ y σ son constantes positivas.

Si la variable aleatoria X se distribuye Inversa Gaussiana, entonces se puede expresar que $X \sim IG(\mu, \sigma)$.

Esta función de densidad pertenece a la familia exponencial y tiene las características de ser unimodal y sesgada positivamente, además de la particularidad que cuando $\mu = 1$ la distribución obtenida es llamada la distribución de Wald. Como en la distribución Gamma, la suma de variables que se distribuyen Inversas Gaussianas es Inversa Gaussiana.

La media y la varianza de la distribución Inversa Gaussiana están dados por:

$$\begin{aligned} E[X] &= \mu \\ Var[X] &= \frac{\mu^3}{\sigma} \end{aligned}$$

Los parámetros μ y σ son interpretados como los parámetros de localización y escala, respectivamente.

Seshadri [42], presenta características particulares de la distribución Inversa Gaussiana y algunas propiedades que entran más al detalle de esta distribución.

La f.g.m. de la distribución IG está definida como:

$$M_x(t) = \exp[\sigma\mu^{-1}\{1 - (1 - 2t\mu^2\sigma^{-1})^{1/2}\}] \quad (1.12)$$

Mediante la f.g.m. se obtienen las siguientes expresiones para los cuatro primeros momentos positivos de la distribución Inversa Gaussiana:

$$\begin{aligned} E[X] &= \mu \\ E[X^2] &= \mu^2 + \mu^3\sigma^{-1} \\ E[X^3] &= \mu^3 + 3\mu^4\sigma^{-1} + 3\mu^5\sigma^{-2} \\ E[X^4] &= \mu^4 + 6\mu^5\sigma^{-1} + 15\mu^6\sigma^{-2} + 15\mu^7\sigma^{-3} \end{aligned} \quad (1.13)$$

Gendron & Crepeau [22], evalúan varias aproximaciones a $F_S(x)$, cuando la variable aleatoria X sigue una distribución IG. El buen ajuste de las aproximaciones se evalúa a través del parámetro de asimetría de la distribución Poisson Compuesta.

■ Distribución Log-Normal

La distribución Log-Normal es utilizada en aplicaciones muy diversas, como: la biología, la confiabilidad, seguros y riesgos, entre otros campos de aplicación, por su sencillez y fácil implementación. Una de las ventajas más mencionadas acerca de esta distribución es que se ajusta muy bien a las distribuciones empíricas y es muy flexible y fácil de estimar. Estas características motivan el empleo de ésta distribución para modelar la variable aleatoria X .

La función de distribución de probabilidad de X , cuando $X \sim \text{LogN}(\mu, \sigma)$ es:

$$f(x; \mu, \sigma) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left[-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right] \quad (1.14)$$

Los momentos de la distribución Log-Normal existen, sin embargo, no hay una expresión definida para explicar la f.g.m. En su lugar, existe una función recursiva que permite calcular los momentos y que se define como:

$$E[X^k] = \exp[k\mu + k^2\sigma^2/2] \quad (1.15)$$

Los primeros cuatro momentos hallados a través de la anterior ecuación, están dados por:

$$\begin{aligned} E[X] &= \exp[\mu + \sigma^2/2] \\ E[X^2] &= \exp[2\mu + 2\sigma^2] \\ E[X^3] &= \exp[3\mu + 9\sigma^2/2] \\ E[X^4] &= \exp[4\mu + 8\sigma^2] \end{aligned} \quad (1.16)$$

1.2. Distribución Semi No Paramétrica (SNP)

La distribución Semi No Paramétrica (SNP) fue definida por Gallant & Nychka [19], quienes motivados por dar solución a la falta de especificación en algunas aplicaciones de los modelos econométricos y a las implicaciones de emplear el método de máxima verosimilitud en ellos, plantean a través de la distribución SNP un estimador cuyo objetivo principal es aproximar una función de densidad desconocida a través de las series de Hermite.

Fenton & Gallant [16] resaltan esta distribución como un método conveniente para estimar simultáneamente los parámetros de un modelo no lineal y la densidad de un proceso latente por máxima verosimilitud, además obtienen como conclusión relevante para este trabajo que los estimadores SNP funcionan asintóticamente similar a los estimadores kernel.

Durante la época de los 90 se estudió con más profundidad las propiedades no paramétricas de los estimadores SNP y se generaron desarrollos importantes que conllevaron a que estos estimadores fueran implementados en áreas como la economía, las finanzas y hasta la ciencias de la salud, según Fenton & Gallant [15].

La definición de este estimador de densidad es dada a continuación, junto con la descripción de algunas propiedades y aplicaciones que motivan su implementación en el desarrollo de esta tesis.

1.2.1. Definición del estimador de densidad SNP

Gallant & Nychka [19], proponen un método de estimación Semi No Paramétrico, cuyo objetivo principal es estimar una función de densidad desconocida a través de las series de Hermite. Un requisito para que el método sea consistente es que el número de términos de la serie de Hermite aumenten con el tamaño muestral, por lo tanto, la aproximación del método mejora a medida que el tamaño de la muestra se amplía.

Gallant & Nychka [19], definen el estimador SNP considerando el resultado de Phillips [36], quien muestra que la forma ERA (Extended Rational Approximant)

$$h(u) = \frac{P^2(u)}{Q^2(u)} \phi^2(u|\tau, \Sigma)$$

puede aproximar cualquier densidad sujeta a ciertas restricciones. En la ecuación anterior, $P(u)$ y $Q(u)$ son polinomios y $\phi(u|\tau, \Sigma)$ es la función de densidad de la Normal multivariada. Para que $h(u)$ sea una función de densidad, es necesario restringir a los polinomios para que la función integre a 1.

Gallant & Nychka [19], restringen la función $h(u)$ a un espacio H , que está definido por un conjunto de funciones de densidades aceptables, específicamente al espacio H_k cuyas densidades son de la forma Hermite:

$$h(u) = P_K^2(u - \tau)\phi^2(u|\tau, \Sigma)$$

donde $P_K(u - \tau)$ es un polinomio de grado K . Para obtener más detalles de los espacios de las densidades y de sus especificaciones diríjase a Gallant & Nychka [19], pag 367.

Fenton & Gallant [16], pag. 82, mencionan que «Gallant & Nychka (1987) nombran el indicador semi no paramétrico para sugerir que el método se encuentra en medio de los métodos paramétricos y no paramétricos: los algoritmos estándar paramétricos son usados aunque el método tiene propiedades no paramétricas».

La ventaja de este método es que no limita con respecto a un modelo particular, Klaauw & Koning [26], pag. 2, mencionan que la densidad Normal es usada como base y agregan que cualquier función de densidad con función generadora de momentos podría ser empleada en lugar de la distribución Normal. Por lo tanto, una alternativa del método propuesto por Gallant & Nychka [19], es optar por la densidad de la distribución Normal en forma univariada. El caso de la distribución SNP definida con la distribución Normal univariada es definido por Fenton & Gallant [16].

Polinomios de Hermite

Los polinomios de Hermite son polinomio ortogonales definidos como

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} \quad (1.17)$$

y cumple con las siguiente propiedades (para más detalles diríjase a Fenton & Gallant [16], pag. 84):

- Ecuación diferenciable
- Posee una relación de recurrencia

$$H_n(x) = xH_{n-1}(x) - nH_{n-2}(x)$$

- Posee una relación diferenciable

$$\frac{d}{dx} H_n(x) = nH_{n-1}(x)$$

A través del polinomio Hermite normalizado, definido como

$$\bar{H}_n(x) = (\sqrt{2\pi n!})^{1/2} H_n(x)$$

se puede obtener la relación recurrente para el polinomio de Hermite normalizado:

$$\bar{H}_n(x) = (x\bar{H}_{n-1}(x) - \sqrt{n-1}\bar{H}_{n-2}(x))/\sqrt{n} \quad (1.18)$$

Definición paramétrica de la distribución SNP

León, Mencía & Sentana [28], exponen las propiedades paramétricas de la distribución SNP; estos autores son los primeros en analizar la distribución SNP presentada por Gallant & Nychka [19] desde el punto de vista paramétrico. La distribución SNP es definida por la siguiente ecuación:

$$f_{SNP}(x; d) = \frac{\phi(x)}{d'd} \left(\sum_{j=0}^k d_j H_j(x) \right)^2 \quad (1.19)$$

donde $x \in R$, $d = (d_0, d_1, \dots, d_k)' \in R^k$, $k = 1, 2, \dots$ es un vector de parámetros, $H_j(x)$ es el polinomio de Hermite normalizado definido en la ecuación (1.18) y $\phi(x)$ es la función de la normal estándar. Existe una restricción para que se cumpla que f_{SNP} sea una función de probabilidad: $d_0 = 1$.

La ecuación (1.19) es la definición de la de la Distribución SNP que se empleará para realizar una aproximación a las distribuciones compuestas en esta tesis.

Una expresión alternativa de la densidad f_{SNP} puede ser obtenida al resolver la expresión que se encuentra al cuadrado:

$$f_{SNP}(x; d) = \phi(x) \sum_{m=0}^{2k} \gamma_m(d) H_m(x)$$

donde $\gamma_m(d) = \frac{d' A_m d}{d'd}$, donde A_m es una matriz simétrica cuadrada cuyos elementos son definidos a través de la siguiente expresión:

$$a_{ij,m} = \frac{(i!j!m!)^{1/2}}{\left(\frac{i+j-m}{2}\right)! \left(\frac{i+m-j}{2}\right)! \left(\frac{m+j-i}{2}\right)!}$$

si $m \in \Gamma$ y en otro caso es cero, y

$$\Gamma = \left\{ m \in \mathbb{N} : |i-j| \leq m \leq i+j; \frac{i-j+m}{2} \in \mathbb{N} \right\}$$

A partir de las expresiones anteriores es posible definir la función generadora de momentos de la

densidad SNP:

$$E[e^{tx}] = \exp\left(\frac{t^2}{2}\Lambda(d, t)\right) \quad (1.20)$$

donde

$$\Lambda(d, t) = \sum_m^{2k} \gamma_m(d) \frac{t^m}{\sqrt{m!}}$$

Mediante la función de densidad de probabilidad o la f.g.m. pueden ser hallados los momentos de la distribución SNP. Para el caso de estudio inicial de esta tesis se emplea, la densidad SNP definida con cuatro parámetros, es decir $k = 4$, por razones de simplicidad y tratabilidad, se deja abierta la propuesta para emplear esta densidad con mayor número de parámetros. Los primeros cuatro momentos de la función de densidad definida con cuatro parámetros, $f_{SNP}(d_1, d_2, d_3, d_4)$ están dados por:

$$\begin{aligned} \mu_{SNP1} &= \frac{2d_1 + 2\sqrt{2}d_1d_2 + 2\sqrt{3}d_2d_3 + 4d_3d_4}{1 + d_1^2 + d_2^2 + d_3^2 + d_4^2} \\ \mu_{SNP2} &= \frac{1 + 4\sqrt{3}d_2d_4 + 9d_4^2 + 2\sqrt{6}d_1d_3 + 7d_3^2 + 5d_2^2 + 2\sqrt{2}d_2 + 3d_1^2}{1 + d_1^2 + d_2^2 + d_3^2 + d_4^2} \\ \mu_{SNP3} &= \frac{6d_1 + 2\sqrt{6}d_3 + 12\sqrt{2}d_1d_2 + 18\sqrt{3}d_2d_3 + 4\sqrt{6}d_1d_4 + 48d_3d_4}{1 + d_1^2 + d_2^2 + d_3^2 + d_4^2} \\ \mu_{SNP4} &= \frac{3 + 12\sqrt{2}d_2 + 15d_1^2 + 39d_2^2 + 75d_3^2 + 123d_4^2 + 20\sqrt{6}d_1d_3 + 56\sqrt{3}d_2d_4 + 4\sqrt{6}d_4}{1 + d_1^2 + d_2^2 + d_3^2 + d_4^2} \end{aligned} \quad (1.21)$$

1.2.2. Propiedades del estimador SNP

Los estimadores SNP poseen propiedades prácticas al momento de emplearlos para estimar funciones de densidad, éstas son mencionadas a continuación:

- Fenton & Gallant [16] concluyen que el comportamiento del estimador de densidades SNP es asintóticamente equivalente al estimador kernel.
- El estimador de densidad SNP pertenece a la clase de estimadores sieve. La explicación de estos estimadores puede ser encontrada en Chen [8], quien expone que los modelos semi no paramétricos son más flexibles y robustos que los modelos paramétricos, pero a su vez tiene complicaciones en la definición de espacios parametrales. El método de sieve proporciona una manera de tratar estas complejidades a través de la optimización de un criterio sobre el espacio de parámetros original. Para asegurar la consistencia del método se requiere que la

complejidad de los sieve aumente a medida que el tamaño de muestral aumenta. Chen [8], pag. 5574, menciona como un ejemplo de un estimador sieve al estimador de densidad SNP cuyo espacio parametral, H_K , es usado para aproximar una densidad desconocida.

Dado que el estimador SNP es un estimador sieve, su representación puede ser en forma paramétrica, lo que permite que metodologías de inferencia, como pruebas de hipótesis y la estimación de medidas como el AIC o BIC, puedan ser calculadas.

1.2.3. Aplicaciones del estimador de densidad SNP

En esta sección se mencionan algunas de las aplicaciones de los estimadores de densidad SNP, además de mencionar algunas referencias claves en el estudio de este tema.

- La distribución SNP tiene una relación directa con los desarrollos en serie de polinomios ortogonales de una función de densidad. Detalles sobre aplicaciones en Estadística de estos desarrollos, y detalles de la teoría básica se pueden consultar en el capítulo 8 del texto de Khuri [25].
- Gallant & Tauchen [21] emplean la densidad SNP para explicar el comportamiento de los retornos de las acciones.
- Gabler, Laisney & Lechner [18], usan la densidad SNP para modelar la distribución de error bivariada de un modelo de selección binaria.
- Por la flexibilidad y capacidad de los estimadores de densidad SNP, estos estimadores han sido usados en el análisis de series de tiempo no lineales, algunos de los artículos conocidos en esta aplicación son: Brunner [5] y Gallant, *et al.* [20].
- Klaauw & Koning [26], emplean la función de densidad SNP para estimar las distribuciones de las alteraciones o de los errores. Este artículo detalla de manera clara la distribución SNP y sus propiedades, según la definición dada por Gallant & Nychka [19]. Una de las propiedades mencionadas es que a medida que se incrementa el número de términos en el polinomio, los estimadores de densidad presentan mejor desempeño.
- Coppejans & Gallant [10] emplean el estimador de densidad SNP como un atributo de la estimación estructural del método eficiente de los momentos.
- León, Mencía & Sentana [28], plantean las propiedades paramétricas de la densidad SNP y además, la emplean para modelar la valoración de opciones.

1.3. Medidas de distancia para comparar las aproximaciones

Las medidas de distancia son elementos fundamentales en el análisis de datos y han sido empleadas en múltiples áreas con el objetivo de comparar dos poblaciones, muestras o distribuciones. Adicionalmente las medidas de distancia son herramientas para la construcción de pruebas de hipótesis, para el estudio de propiedades asintóticas de los estimadores, la construcción de estimadores, etc. Por lo tanto, las medidas de distancia son un elemento teórico intuitivo que facilita el análisis y la comparación entre dos distribuciones o muestras.

En el caso de estudio, las medidas de distancia son empleadas para comparar el ajuste de las aproximaciones a las distribuciones compuestas. En esta tesis se emplearan las medidas de distancia de Kolmogorov Smirnov y la medida de Cramer-von Mises ponderada. A continuación, se presenta un detalle de cada una de las estas medidas de distancia.

Medida de Distancia de Kolmogorov - Smirnov

El estadístico de Kolmogorov - Smirnov esta construido por la siguiente medida de distancia:

$$KS = \sup_{x \in \mathbb{R}} |F(\hat{x}) - F(x)| \quad (1.22)$$

donde $F(x)$ corresponde a la función de distribución y $F(\hat{x})$ es la función de distribución aproximada. Por lo tanto, el valor de KS es la máxima distancia que exista entre ambas funciones de distribución acumulada.

El estadístico de Kolmogorov es frecuentemente empleado para la construcción de pruebas de hipótesis que verifican la igualdad de dos muestras, distribuciones o de una muestra frente a una distribución de referencia.

Medida de Distancia de Cramer-von Mises

La medida de distancia que soporta el estadístico de prueba de Cramer-von Mises está dada por:

$$CVM = \int_{-\infty}^{\infty} (F(x) - F(\hat{x}))^2 dF(x)$$

Hogg & Klugman [23], definen la medida de distancia de Cramer-von Mises de la siguiente manera:

$$\frac{1}{n} \sum_{i=0}^n (F(x) - F(\hat{x}))^2$$

Adicionalmente a esta definición, Hogg & Klugman [23] mencionan que es posible incluir en esta ecuación una ponderación $w(x)$ para darle mayor peso a ciertos puntos de la distribución. Por lo tanto, la siguiente expresión se define como el estimador ponderado de la distancia mínima de Cramer-von Mises:

$$CVM = \sum_{i=0}^n w(x_i) (F(x_i) - F(\hat{x}_i))^2 \quad (1.23)$$

donde la función $w(x)$ puede ser definida por la siguiente expresión, la cual le asigna mayor peso a los valores finales de la distribución, que es precisamente donde se desea ser más estrictos con respecto al ajuste:

$$w(x) = \frac{n}{F(x)(1 - F(x))}$$

Una definición más amplia es dada en Hogg & Klugman [23], pag. 83 y 135. Esta función de distancia fue empleada por Hurlimann [24], el cual menciona que esta medida es especialmente útil para realizar pruebas empíricas de las colas de la distribución.

Capítulo 2

Método de Aproximación SNP

El objetivo principal de esta tesis es proponer y desarrollar una aproximación a las distribuciones compuestas por medio de la distribución Semi No Paramétrica (SNP), con el fin de proporcionar un método de aproximación efectivo para el análisis de distribuciones compuestas en las que se torna complejo un desarrollo exacto.

Asumiendo que la distribución de X , denotada por $F_X(x)$, depende de m parámetros, $(\theta_1, \dots, \theta_m)$, y asumiendo N tiene una distribución denotada por p_n , que depende de n parámetros $\lambda_1, \dots, \lambda_n$, se puede obtener una distribución compuesta denotada por $S \sim F_S(\lambda_1, \dots, \lambda_n, F(x; \theta_1, \dots, \theta_m))$. El propósito es aproximar la distribución de S , $F_S(x)$, mediante la distribución SNP, $F_{SNP}(x; d_1, \dots, d_k)$, tal que

$$\sup_{x \geq 0} \{|F_S(x) - F_{SNP}(x)|\} \quad (2.1)$$

sea lo más pequeña posible. Como se mencionó anteriormente, el número de parámetros de la distribución SNP seleccionado en esta tesis es 4, es decir: se empleará la función $F_{SNP}(x; d_1, \dots, d_4)$, lo cual garantiza la existencia de los cuatro primeros momentos y el manejo teórico de la distribución. La aproximación propuesta a través de la distribución SNP para las distribuciones compuestas será analizada para el caso específico de la Poisson Compuesta, es decir, cuando $N \sim Poisson(\lambda)$. Se deja abierto el tema y la revisión de las aproximaciones para cuando la distribución compuesta está conformada por otras distribuciones que modelan el comportamiento de la variable aleatoria N . Por lo tanto, en el estudio de esta tesis asumiremos que $S \sim PC(\lambda, F(x; \theta_1, \dots, \theta_m))$, la especificación de esta función se presenta en el Ejemplo 1., ecuación (1.5).

Para el desarrollo de la aproximación es necesario estimar los parámetros d_1, \dots, d_4 de la distribución SNP, en función de $\lambda, \theta_1, \dots, \theta_m$, los parámetros de la Poisson Compuesta; para la estimación de estos parámetros se emplean los siguientes métodos:

- Método de los Momentos (conocido como *moment matching methodology*): Igualdad de los primeros cuatro momentos teóricos de la distribución Poisson Compuesta y la distribución SNP.
- El método de estimación por máxima verosimilitud.

Ambos métodos de aproximación son evaluados para los casos considerados de análisis. Específicamente el método de aproximación por la igualdad de los momentos es utilizado sólo cuando se garantiza que los momentos de la distribución Poisson Compuesta pueden ser definidos.

Una vez se obtengan los parámetros estimados de la distribución SNP, la aproximación SNP será la evaluación de la función de distribución SNP en estos parámetros. El desempeño de la aproximación SNP es analizado cuando la variable aleatoria X es modelada por diferentes distribuciones, las consideradas en este trabajo son: Inversa Gaussiana y Log-Normal. Adicionalmente, la aproximación SNP se compara con las aproximaciones descritas anteriormente: Normal, la Normal Power (NP), Gamma Incompleta y Gamma Traslada.

En las siguientes secciones se justifica el empleo de la aproximación SNP para las distribuciones compuestas y especifican los procedimientos para la construcción de la aproximación SNP con ambos métodos de estimación.

2.1. Aproximación SNP a las distribuciones compuestas

La propuesta de esta tesis es aproximar la función de distribución acumulada (f.d.a.) de la distribución compuesta $F_S(x)$ mediante la correspondiente distribución SNP(d), indicada por $F_{SNP}(x)$, en esta sección se analizan las propiedades que favorecen esta aproximación y algunas de las condiciones para su desarrollo.

Una de las primeras condiciones para el desarrollo de la aproximación es que la definición de ésta debe ser absolutamente continua, por lo que se excluyen el caso $F_S(0) > 0$; por lo tanto, lo que se busca es aproximar F_S por una $F_{SNP}(x)$ que cumple $F_{SNP}(0) = 0$.

En la literatura se encuentran dos procedimientos para plantear este tipo de aproximaciones:

1. Aproximar la variable S por una $SNP(d)$, es decir, $F_S(x) \approx F_{SNP}(x), x > 0$.
2. Aproximar la variable S estandarizada con una $SNP(d)$, es decir:

$$P(S \leq x) = P((S - \mu_S)/\sigma_S \leq (x - \mu_S)/\sigma_S) \approx F_{SNP}((x - \mu_S)/\sigma_S)$$

El primer procedimiento se aplica en el método de la Gamma Traducida y en el método propuesto por Chaubey *et al.* [6], quien propone una aproximación a las distribuciones compuestas a través de la distribución Inversa Gaussiana por el método de igualdad de momentos. El segundo procedimiento se aplica en los métodos Normal, Normal Power, Edgeworth, y Gamma (según lo presenta Bowers [3]). En el desarrollo de esta tesis se empleará el segundo procedimiento para plantear la aproximación SNP, guardando consistencia con la distribución base empleada para la distribución SNP, la normal estándar.

Por lo tanto, la variable aleatoria S será estandarizada para aplicarle tanto el método de igualdad de momentos, como el método de máxima verosimilitud. En la siguiente sección se justificará el empleo de las distribuciones SNP para aproximar las distribuciones compuestas. En las siguientes secciones se especificará al detalle el desarrollo de la aproximación con ambos métodos de estimación.

2.1.1. Condiciones para emplear el estimador de densidad SNP

En esta sección se presentan algunas conclusiones que sustentan la propuesta de aproximar distribuciones compuestas con la distribución SNP. Fenton & Gallant [15], presentan las condiciones suficientes para que una función de densidad pueda ser aproximada a través de la distribución SNP, es decir, para que el estimador de densidad SNP pueda ser adecuadamente empleado; estas condiciones son analizadas para el caso en el que la función de densidad es una distribución compuesta, con el fin de justificar la propuesta de aproximación que en esta tesis se plantea.

Fenton & Gallant [15], con el objetivo de comparar el comportamiento de los estimadores SNP con otros estimadores de densidad, como el kernel, establecen condiciones suficientes para que la aproximación de una función de densidad mediante la distribución SNP sea válida. Suponga que la densidad a ser aproximada es denotada por $f(x)$, es decir, la distribución de S , la cual es definida continua en \mathbb{R} . El supuesto principal para que la aproximación sea válida en Fenton & Gallant [15], pag. 720, es que la función $f(x)$ puede ser escrita de la siguiente forma:

$$f(x) = g^2(x) \exp(-x^2/2). \quad (2.2)$$

Adicionalmente, las derivadas $g_j(x) = (d^j/dx^j)g(x)$ deben existir hasta cierto orden $k \geq 1$, y deben satisfacer la condición

$$\int_{-\infty}^{\infty} g_j^2(x) \exp(-x^2/2) dx < \infty, \quad j = 0, 1, \dots, k. \quad (2.3)$$

Y además, se exige que para toda pareja de números $a_0 \in \mathbb{R}, a_1 > 0$ existan $k_0, k_1 > 0$ tales que

$$P(S^2 \geq a_0 + a_1 t) \leq k_0 e^{-k_1 \sqrt{t}}, \quad (2.4)$$

para S variable aleatoria con densidad $f(x)$. Esta condición se cumple si las colas de la distribución de S decrecen de manera exponencial a cero. Si se cumplen las condiciones (2.2), (2.3), (2.4) el estimador de densidad SNP ajusta de manera adecuada la función $f(x)$.

Para construir la aproximación se debe realizar el procedimiento que se describe a continuación. Se asume una muestra aleatoria de la variable S de tamaño n , denotada por $x_t, t = 1, \dots, n$, y se estandariza la variable x_t , se denota $z_t = (x_t - \bar{x})/s$ como la variable estandarizada con s la desviación estándar y \bar{x} la media de la muestra aleatoria. Luego, se estima un modelo SNP por máxima verosimilitud o por el método de igualdad de momentos con base en la muestra anterior, por lo tanto, se deben estimar p_k coeficientes $d_j, j = 1, \dots, p_k$; a medida que n el valor de k también, por lo tanto, se asume que el número de coeficientes de la aproximación SNP aumenta como función del tamaño de la muestra de la variable cuya densidad se quiere aproximar. La densidad SNP con base en estos p_k parámetros estimados se indica por $\hat{f}_k(x)$. Entonces, el Teorema 1 de Fenton & Gallant [15], pag. 720, concluye que si $p_k = n^\alpha$, donde $0 < \alpha < 1$, entonces:

$$\int_{-\infty}^{\infty} |f(x) - \hat{f}_k(x)| dx = A(\alpha, k)$$

donde $A(\alpha, k)$ converge a cero casi en todas partes, más rápidamente que $n^{\delta-1/2+\alpha/2}$ para $\delta > 0$ pequeño.

Fenton y Gallant [15] señalan que a partir de este resultado se puede observar que los estimadores SNP son muy similares a los estimadores no paramétricos de densidades tipo kernel. Por ejemplo, cuando $n = 1000$ la tasa de convergencia más rápida se obtiene con $\alpha = 1/5$. En este caso $p_k \approx 4$.

Para el desarrollo de esta tesis se plantea considerar la variable S modelada por una distribución compuesta, por lo tanto, se debe comprobar que las condiciones (2.2), (2.3), (2.4), se cumplan. Las primeras dos condiciones se cumplen, ya que S es una distribución compuesta y por lo tanto, $P(S \geq 0) = 1$. La condición (2.4) implica más detalle y estudio.

Willmot & Lin [44], expresan que la función de $f(x)$ puede ser escrita como: $f(x) = \sum_{j=0}^{\infty} p_j g^{*j}(x)$, donde $g^{*j}(x)$ representa las convoluciones de $g(x)$. Por lo tanto, para cumplir con la condición (2.4) se requiere tener un resultado de la forma:

$$P(S > t) \leq k_0 e^{-k_1 t} \tag{2.5}$$

Como anteriormente se ha mencionado, la variable aleatoria $S > 0$, por lo que S positiva si $t > 0$; por lo tanto, si se cumple (2.5) entonces

$$P(S^2 > t) = P(S > \sqrt{t}) \leq k_0 e^{-k_1 \sqrt{t}}$$

que es equivalente a (2.4). Luego, la condición que debe cumplirse para la aproximación SNP sea válida para las distribuciones compuestas, cuando estas tienen distribuciones asociadas $g(x)$ suficientemente suaves, es (2.5).

En la literatura existen resultados que garantizan estas condiciones cuando S tiene una distribución compuesta. Por ejemplo, Willmot & Lin [44], demuestran que la distribución compuesta puede ser escrita de la forma requerida por Fenton & Gallant [15] para cumplir con las condiciones necesarias para obtener un buen ajuste por el estimador SNP; el siguiente resultado lo demuestra: Suponga que la distribución compuesta puede ser escrita de la forma $f(x) = \sum_{j=0}^{\infty} p_j g^{*j}(x)$ y asuma que existen: una constante $\phi_2 \in (0, 1)$ tal que $a_n \leq \phi_2 a_{n-1}$ para $a_n = \sum_{j=n+1}^{\infty} p_j = P(N \geq n)$ y otra constante $k_2 > 0$ tal que $M_X(k_2) = 1/\phi_2$ donde $M_X(t)$ es la f.g.m de la variable X con densidad $g(x)$, la cual se asume existe en una vecindad del origen. Entonces se cumple que

$$P(S > t) \leq \frac{1 - p_0}{\phi_2} e^{-k_2 t} \quad (2.6)$$

La condición para que exista ϕ_2 en el caso $N \sim Poisson(\lambda)$ está en Willmot & Lin [44]. La condición para que exista k_2 depende del modelo asumido para X ; por ejemplo, en el caso de asumir una distribución Inversa Gaussiana, $X \sim IG(\mu, \beta)$, se tiene que la f.g.m. está dada por la ecuación (1.12), por lo tanto, la solución de $M_x(k_2) = 1/\phi_2$ se puede encontrar y es igual a

$$k_2 = \log(1/\phi_2) \frac{2\beta - \mu \log(1/\phi_2)}{2\mu\beta} \quad (2.7)$$

Nótese que ϕ_2 debe satisfacer en este caso $e^{-\beta/\mu} < \phi_2 < 1$ para que exista la solución k_2 . Como resumen de la discusión anterior podemos enunciar el siguiente resultado.

Teorema 1. Suponga una variable S distribuída Poisson Compuesta, $S \sim PC(\lambda, F(x))$, con $X \sim F(x)$ distribuída Inversa Gaussiana, $IG(\mu, \beta)$. Entonces, asumiendo que existen constantes ϕ_2 tal que $P(N \geq n + 1 | N \geq n) \leq \phi_2, \forall n = 1, 2, \dots$, y k_2 definida por (2.7), entonces se cumplen las condiciones (2.2), (2.3) y (2.4) y por tanto, $\int_{-\infty}^{\infty} |f_S(x) - \hat{f}_n(x)| dx \rightarrow 0, n \rightarrow \infty$ con probabilidad uno.

Otros casos para la distribución de la variable X requieren análisis similares. Por ejemplo, para el caso en el cual X se distribuye Log-Normal(μ, σ), no existen, hasta donde nuestro conocimiento llega, resultados similares a los obtenidos por Willmot & Lin [44]. Una explicación es que no es posible garantizar que puede encontrarse la solución k_2 tal que $M_X(k_2) = 1/\phi_2$, debido a que la f.g.m. de la distribución Log-Normal no tiene expresión analítica, no conocemos cómo podría asegurarse la existencia de la solución. Sin embargo, existen resultados similares, como el siguiente, en Panjer y Willmot [31], pag. 348.

$$P(S > x) \sim E(N) \sigma^3 (2\pi)^{-1/2} \frac{x}{\log(x) - \mu)^2} e^{-(\log(x) - \mu)^2 / (2\sigma^2)}, \quad x \rightarrow \infty$$

Lo cual garantiza que la distribución Log-Normal es de la clase subexponencial.

2.2. Aproximación SNP por el Método Máxima Verosimilitud

El método de máxima verosimilitud es uno de los principales métodos de estimación paramétrica conocido por las buenas propiedades de sus estimadores, ya que estos cumplen con las siguientes propiedades: son asintóticamente insesgados, invariantes y eficientes.

Para aplicar el método de estimación por máxima verosimilitud se parte del supuesto de que existe una muestra aleatoria, denotada por: z_1, z_2, \dots, z_n , con función de distribución de probabilidad $f(z, \omega)$ donde $\omega = (\omega_1, \dots, \omega_k)$ son los k parámetros, $f(z|\omega_1, \dots, \omega_k)$, cuya función de máxima verosimilitud existe y es definida como:

$$L(\omega|z) = \prod_{i=1}^n f(z_i|\omega_1, \dots, \omega_k)$$

Los estimadores máximo verosímiles son aquellos valores de $\hat{\omega}_i$ que maximizan la función de verosimilitud.

Para la construcción de la aproximación SNP es necesario conocer la función de verosimilitud de dicha distribución; por lo tanto, si la función de distribución de probabilidad de la SNP está dada por $f_{SNP}(x; d_1, d_2, d_3, d_4)$, la función de verosimilitud se define de la siguiente manera:

$$\begin{aligned} L(x; d_1, \dots, d_4) &= \prod_{i=1}^n f_{SNP}(x_i; d_1, d_2, d_3, d_4) \\ &= \prod_{i=1}^n \frac{e^{-x_i^2/2} (1 + d_1 h_1(x_i) + d_2 h_2(x_i) + d_3 h_3(x_i) + d_4 h_4(x_i))^2}{\sqrt{2\pi}(1 + d_1^2 + d_2^2 + d_3^2 + d_4^2)} \\ &= \left(\frac{1}{\sqrt{2\pi}(1 + d_1^2 + d_2^2 + d_3^2 + d_4^2)} \right)^n e^{-\sum_{i=1}^n x_i^2/2} \\ &\quad \prod_{i=1}^n (1 + d_1 h_1(x_i) + d_2 h_2(x_i) + d_3 h_3(x_i) + d_4 h_4(x_i))^2 \end{aligned} \quad (2.8)$$

Por lo tanto, el logaritmo de la función de máxima verosimilitud tiene la siguiente expresión:

$$\begin{aligned} \log(L(x; d_1, d_2, d_3, d_4)) &= -\log(\sqrt{2\pi}(1 + d_1^2 + d_2^2 + d_3^2 + d_4^2)) - \sum_{i=1}^n x_i^2/2 \\ &\quad + \sum_{i=1}^n \log((1 + d_1 h_1(x_i) + d_2 h_2(x_i) + d_3 h_3(x_i) + d_4 h_4(x_i))^2) \end{aligned} \quad (2.9)$$

donde \log corresponde al logaritmo natural.

El método de máxima verosimilitud consiste en estimar los valores de los parámetros de la distribución SNP, d_1, d_2, d_3, d_4 , que maximicen la probabilidad de obtener los valores que se dieron en la muestra, dicha muestra se genera a partir de simulación de la distribución Poisson Compuesta. Los estimadores máximo verosímiles son calculados a partir de la derivada de la función de verosimilitud con respecto a cada parámetro d_i ; normalmente se toma el logaritmo natural de la función de máxima verosimilitud para facilitar la obtención de los estimadores. Como anteriormente se mencionó, este método produce estimadores que son funciones de estadísticos suficientes si el estimador máximo verosímil es único, proporciona estimadores eficientes, es decir, de varianza mínima, y los estimadores obtenidos son asintóticamente insesgados.

Para estimar los parámetros d_1, d_2, d_3, d_4 de la distribución SNP a través del método de máxima verosimilitud una muestra aleatoria de la distribución Poisson Compuesta es generada. El procedimiento para hallar los estimadores de los parámetros comienza con la simulación de los datos de la distribución Poisson Compuesta ($PC(\lambda; F_X(x))$) para cada caso de distribución $F_X(x)$ considerado, dichos datos son estandarizados con la media y la desviación estándar de la distribución compuesta dadas por la ecuación (1.9). La estimación de los parámetros de la distribución SNP es hallado a través de una función implementada en R, la función del logaritmo de máxima verosimilitud, la cual se optimiza por la función `optim`.

Una ventaja de la implementación de este método de estimación es que podrá ser empleado en el caso particular en el que el sistema de ecuaciones de momentos no pueda ser definido debido a que la distribución de las X_j no posee al menos los cuatro primeros momentos. Adicionalmente, es implementado para objetivos de comparación en la estimación de las distribuciones compuestas.

2.3. Aproximación SNP por el Método de Igualdad de Momentos

El método de los momentos es uno de los más antiguos procedimientos que se conocen para obtener estimadores puntuales, con la virtud de ser uno de los métodos más simples de estimación. Este método consiste en igualar los momentos teóricos de una distribución $D1$ con los momentos muestrales de una distribución $D2$, con el fin de estimar los parámetros de la distribución $D1$ en función de los valores muestrales de la distribución $D2$. Para lograrlo, se plantean tantas ecuaciones como parámetros a ser estimados; al final, los estimadores de los parámetros son aquellos valores que dan solución al sistema de ecuaciones conformado.

El método, llamado en esta tesis, igualdad de momentos es una modificación del método de momentos anteriormente definido, ya que, en lugar de igualar los momentos teóricos de una distribución $D1$ a los momentos muestrales de la distribución $D2$, el sistema se conforma por los momentos teóricos de ambas distribuciones; de esta manera se conforma un sistema de ecuaciones el cual es

resuelto cuando se halla la definición de los parámetros de la distribución $D1$ en función de los parámetros de la distribución $D2$.

En la revisión realizada sobre las aproximaciones a las distribuciones compuestas, se encontró que muchas de las aproximaciones son definidas empleando el método de igualdad de momentos, por lo tanto, para construir la aproximación SNP este método de estimación es evaluado. Una evidencia de esto es la construcción de la aproximación Gamma Traslada, presentada en el capítulo anterior. Panjer [32], pág. 181, también aplica éste método. Chaubey, *et al.* [6], pág. 223, proponen la aproximación Inversa Gaussiana a $F_S(x)$ a través de este método.

Para definir la aproximación SNP es necesario hallar los parámetros de la distribución SNP, es decir d_1, d_2, d_3, d_4 , en función de los parámetros de la distribución Poisson Compuesta, que son $\lambda, \theta_1, \dots, \theta_m$, donde $\theta_1, \dots, \theta_m$ son los m parámetros de la distribución de la variable aleatoria X . Para obtener la aproximación SNP se emplea el método de igualdad de los momentos, creando un sistema no lineal de cuatro ecuaciones con cuatro incógnitas, al igualar los cuatro primeros momentos de la distribución SNP con los momentos estandarizados de la distribución Poisson Compuesta:

$$\mu_{SNPj} = E[Z^j], j = 1, \dots, 4 \quad (2.10)$$

con incógnitas d_1, \dots, d_4 . Donde μ_{SNPj} corresponden al momento j -ésimo de la distribución SNP, los primeros cuatro momentos de la distribución SNP están dados en las ecuaciones (1.21); $E[Z^j]$ corresponde al j -ésimo momento de Z , la cual, bajo el supuesto de que $S \sim PC(\lambda; F_X(x; \theta_1, \dots, \theta_m))$, es:

$$Z = \frac{S - \mu_S}{\sigma_S}$$

la variable estandarizada de la variable aleatoria S , donde μ_S y σ_S son la media y la desviación estándar de la distribución de S , dadas en la ecuación (1.8).

El sistema de ecuaciones conformado para hallar la aproximación SNP de la distribución Poisson Compuesta, en forma general, debe ser planteado de la siguiente forma:

$$\begin{aligned} \mu_{SNP1} &= 0 \\ \mu_{SNP2} &= 1 \\ \mu_{SNP3} &= \frac{\lambda E[X^3]}{(\lambda E[X^2])^{\frac{3}{2}}} \\ \mu_{SNP4} &= \frac{\lambda E[X^4] + 3\lambda^2 (E[X^2])^2}{(\lambda E[X^2])^2} \end{aligned} \quad (2.11)$$

Por lo tanto, se conforma un sistema de ecuaciones no lineales que debe ser resuelto por métodos numéricos. Este sistema se puede resolver siempre y cuando los momentos de la variable aleatoria

X_j existan. En caso contrario, la estimación de los parámetros puede ser resuelta por el método de estimación máxima verosimilitud.

La solución del sistema de ecuaciones no lineales se resuelve en \mathbf{R} a través de la función `optim`, que emplea varios métodos de optimización, entre ellos el método quasi-Newton. El método quasi-Newton es un método de optimización, también conocido como método o algoritmo de métrica variable, y es empleado para hallar máximos y mínimos locales. Este método está basado en el método de Newton para encontrar el punto estacionario de una función, donde el gradiente es 0.

La aproximación SNP a la distribución Poisson Compuesta es evaluada para cuando la variable aleatoria X_j es modelada por las distribuciones continuas: IG y Log-Normal. Un planteamiento similar es realizado por Chaubey, *et al.* [6], pag. 217, ya que en este artículo se evalúan las aproximaciones cuando $F_X(x)$ adopta las siguientes distribuciones: IG, Gamma, Pareto y Weibull.

A continuación se plantean los sistemas de ecuaciones no lineales a ser resueltos para cada una de estas distribuciones de $F_X(x)$ consideradas. Los resultados de la Aproximación SNP son presentados en el siguiente capítulo.

2.3.1. Aproximación SNP cuando $X \sim IG$

La distribución Inversa Gaussiana, como fue comentado anteriormente, posee la propiedad reproductiva, es decir, la convolución de n distribuciones IG independientes es nuevamente IG, lo cual resulta útil para la evaluación de la distribución de la variable aleatoria S . Es por esta razón, que esta distribución ha sido atractiva para explicar las severidades y evaluar el desempeño de las distribuciones compuestas, tal como lo hacen Gendron & Crepeau [22].

Para desarrollar la aproximación SNP es necesario construir el sistema de ecuaciones conformado por los momentos de la distribución SNP y los momentos estandarizados de la distribución Poisson Compuesta cuando $X \sim IG(\mu, \sigma)$, $S \sim PC(\lambda; IG(\mu, \sigma))$.

La definición teórica de los momentos de la distribución Poisson Compuesta cuando $X \sim IG(\mu, \sigma)$, se obtienen a partir de la ecuación (1.7) y los momentos de la distribución IG son obtenidos a través de las ecuaciones dadas en (1.13). A continuación se presentan los momentos de la variable estandarizada Z , cuando $S \sim PC(\lambda; IG(\mu, \sigma))$:

$$\begin{aligned} \mu_{Z_{PC_1(\lambda; IG)}} &= 0 \\ \mu_{Z_{PC_2(\lambda; IG)}} &= 1 \\ \mu_{Z_{PC_3(\lambda; IG)}} &= \frac{\lambda(\mu^3 + 3\mu^4\sigma^{-1} + 3\mu^5\sigma^{-2})}{(\lambda(\mu^2 + \mu^3\sigma^{-1}))^{\frac{3}{2}}} \end{aligned}$$

$$\mu_{Z_{PC_4}(\lambda; IG)} = \frac{\lambda(\mu^4 + 6\mu^5\sigma^{-1} + 15\mu^6\sigma^{-2} + 15\mu^7\sigma^{-3}) + 3\lambda^2(\mu^2 + \mu^3\sigma^{-1})^2}{(\lambda(\mu^2 + \mu^3\sigma^{-1}))^2} \quad (2.12)$$

El objetivo es estimar los parámetros (d_1, d_2, d_3, d_4) en función de los parámetros de la Poisson Compuesta (λ, μ, σ) mediante el método de igualdad de los momentos, tal como lo demuestra el siguiente diagrama:

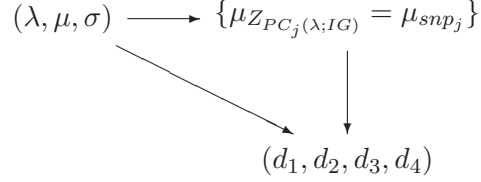


Diagrama 1. Composición del sistema de momentos

Por lo tanto, el sistema de ecuaciones no lineales a resolver es el presentado en la ecuación (2.11) resultante de igualar las ecuaciones (1.21) y (2.12) es:

$$\begin{aligned} \mu_{SNP1} &= \mu_{Z_{PC_1}(\lambda; IG)} \\ \mu_{SNP2} &= \mu_{Z_{PC_2}(\lambda; IG)} \\ \mu_{SNP3} &= \mu_{Z_{PC_3}(\lambda; IG)} \\ \mu_{SNP4} &= \mu_{Z_{PC_4}(\lambda; IG)} \end{aligned}$$

La solución de este sistema se halla utilizando la función `optim` del R. Para conocer más detalles de la programación ver el Apéndice A.

2.3.2. Aproximación SNP cuando $X \sim \text{Log}N$

La aproximación SNP se evalúa cuando la variable aleatoria X tiene una distribución Log-Normal. Todos los momentos de la distribución Log-Normal existen teóricamente, por lo tanto, la aproximación SNP puede ser planteada a través del método de igualdad de los momentos. Los momentos estandarizados de la distribución Poisson Compuesta cuando $X \sim \text{Log}N(\mu, \sigma)$, $S \sim PC(\lambda; \text{log}N(\mu, \sigma))$, están dados por:

$$\begin{aligned} \mu_{Z_{PC_1}(\lambda; \text{Log}N)} &= 0 \\ \mu_{Z_{PC_2}(\lambda; \text{Log}N)} &= 1 \\ \mu_{Z_{PC_3}(\lambda; \text{Log}N)} &= \frac{\lambda(\exp[3\mu + 9\sigma^2/2])}{(\lambda(\exp[2\mu + 2\sigma^2]))^{\frac{3}{2}}} \end{aligned}$$

$$\mu_{Z_{PC_4}(\lambda; \text{Log}N)} = \frac{\lambda(\exp[4\mu + 8\sigma^2]) + 3\lambda^2(\exp[2\mu + 2\sigma^2])^2}{(\lambda(\exp[2\mu + 2\sigma^2]))^2} \quad (2.13)$$

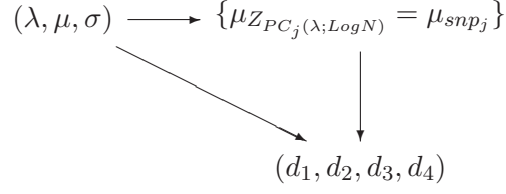


Diagrama 2. Composición del sistema de momentos

El sistema de ecuaciones no lineales a resolver, construido igualando los momentos de la distribución $SNP(d_1, d_2, d_3, d_4)$ denotados μ_{SNP_j} y descritos en las ecuaciones (1.21), con los momentos de la variable estandarizada Z dados en (2.13), cuando $S \sim PC(\lambda; \text{log}N(\mu, \sigma))$, como se expresa en la ecuación (2.11) y se muestra en el diagrama 2, es:

$$\begin{aligned} \mu_{SNP1} &= \mu_{Z_{PC_1}(\lambda; \text{Log}N)} \\ \mu_{SNP2} &= \mu_{Z_{PC_2}(\lambda; \text{Log}N)} \\ \mu_{SNP3} &= \mu_{Z_{PC_3}(\lambda; \text{Log}N)} \\ \mu_{SNP4} &= \mu_{Z_{PC_4}(\lambda; \text{Log}N)} \end{aligned}$$

La solución al sistema de ecuaciones no lineales es resuelta en **R** a través de la función `optim`. Los detalles de programación son especificados en el Apéndice A.

Capítulo 3

Análisis de las Aproximaciones a la Distribución Compuesta

En este capítulo se analiza y compara la calidad del ajuste de las aproximaciones: SNP, Normal Power (NP), Gamma Traslada y Gamma Incompleta a la distribución Poisson Compuesta cuando la variable aleatoria N se distribuye $Poisson(\lambda)$ y X adopta diferentes distribuciones definidas en el capítulo anterior, las cuales son: la distribución Inversa Gaussiana y Log-Normal.

Algunos autores como Chaubey, *et al.* [6] y Gendron & Crepeau [22], advierte que el ajuste de las aproximaciones pierde su calidad cuando la asimetría de los datos de la distribución compuesta supera el valor de 2. En la revisión realizada acerca de las aproximaciones existentes a las distribuciones compuestas, es evidente que la asimetría de la distribución S es un factor determinante en el desempeño y eficiencia de las mismas. Gendron & Crepeau [22] analizan las aproximaciones para tres valores distintos de parámetros de la distribución Inversa Gaussiana, con el fin de obtener diversos valores de asimetría resultante de la distribución Poisson Compuesta. Chaubey, *et al.* [6] asigna diferentes valores a los parámetros de la distribución compuesta, con el fin de que los valores de la asimetría y la kurtosis fueran obtenidos para cada caso, doce escenarios fueron generados y se obtuvo una asimetría de la distribución compuesta en un rango entre: $[0.56, 2.08]$ para evaluar el ajuste.

Por lo tanto, este será uno de los criterios de evaluación, para probar el nivel de ajuste que proporciona la aproximación SNP. La evaluación de las aproximaciones se realiza para cada una de las distribuciones de la variable aleatoria X , en la cual se generan dos casos de asimetría para analizar el desempeño de cada una de las aproximaciones, los casos de asimetría son:

Escenario 1 $\gamma_S < 2$

Escenario 2 $\gamma_S > 2$

El análisis de ajuste de las aproximaciones se concentra en la cola superior de la distribución Poisson Compuesta, ya que ésta es la parte de mayor interés en las aplicaciones de este tipo de distribuciones; un ejemplo de aplicación es en el tema de riesgos, ya que la parte superior de la distribución aporta valiosa información para la administración y gestión. Por lo tanto, las comparaciones se realizan cuando la función de distribución acumulada de la Poisson Compuesta toma valores de probabilidad acumulada superiores a 0.85.

En el capítulo inicial de esta tesis, se presentaron las conclusiones de Klaauw & Koning [26], las cuales indican que a medida que el número de parámetros de la distribución SNP aumenta, los estimadores de densidad SNP presentan mejor desempeño. Por lo tanto, para cada uno de los casos analizados se evalúa el desempeño de la aproximación SNP a medida que el número de parámetros aumenta. El número de parámetros considerados para la distribución SNP son: 4, 6, 8 y 10. Adicionalmente se aumenta el tamaño de muestra para el análisis del desempeño de las aproximaciones.

Como se mencionó anteriormente, la aproximación SNP puede ser construida a través de igualdad de momentos y de estimación por máxima verosimilitud, es de aclarar entonces, que la construcción de igualdad de momentos es sólo evaluada para cuando la distribución SNP tiene cuatro parámetros, ya que al aumentar el número de parámetros aumenta la dificultad de estimación del sistema de ecuaciones conformado por los momentos. Por lo tanto, cuando se menciona que el número de parámetros de la distribución SNP aumenta para construir la aproximación, se está haciendo referencia exclusivamente al método de estimación por máxima verosimilitud.

Muestras aleatorias de la distribución Poisson Compuesta pueden ser halladas a través de un desarrollo en \mathbb{R} , lo que permite establecer un análisis de ajuste comparando las demás aproximaciones mencionadas en esta tesis. Para establecer el nivel de ajuste de cada una de las aproximaciones analizadas se emplean las medidas de distancia definidas en las ecuaciones (1.22) y (1.23), las cuales proporcionan una idea de la diferencia entre las curvas de la distribución Poisson Compuesta y la aproximación que se esté analizando.

El procedimiento para analizar las aproximaciones consiste en plantear conjuntos de valores para los parámetros considerados en la evaluación de la distribución Poisson Compuesta, los cuales son el parámetro de la distribución Poisson y los parámetros de la distribución de la variable aleatoria X ; luego, calcular el valor de la asimetría, estimar cada una de las aproximaciones y calcular las distancias empleadas para la comparación. Adicionalmente, se evalúa la aproximación al aumentar el número de parámetros considerados para la distribución SNP y el tamaño de muestra para la generación de datos con distribución Poisson Compuesta.

Los resultados son presentados gráficamente y a partir de tablas, en las cuales se especifican las medidas de distancia empleadas, para cada una de las aproximaciones comparadas; las escalas de las gráficas son logarítmicas, con el objetivo de poder observar mejor el desempeño de las aproximaciones. Se hace énfasis en la aproximación SNP al presentar tablas y gráficas con los resultados obtenidos al aumentar el número de parámetros empleados y el tamaño de la muestra inicial de la distribución Poisson Compuesta; con el fin de obtener resultados y conclusiones objetivas que permitan sugerir su empleo.

3.1. Resultados de la aproximación SNP a la $PC(\lambda; IG(\mu, \sigma))$

Como se mencionó en la introducción de este capítulo, se presentan en éste análisis dos casos de parámetros, de los cuales una de ellos presenta asimetría inferior al valor de 2 y la otra mayor. En esta sección se considera que X se distribuye Inversa Gaussiana. El tamaño de muestra inicial para los datos de la distribución Poisson Compuesta es de $n = 2000$. Los resultados para cada uno de los casos de asimetría son presentados a continuación.

3.1.1. Caso de Asimetría $\gamma_S < 2$, cuando $S \sim PC(\lambda; IG(\mu, \sigma))$

Para evaluar este caso, se emplean los siguientes valores de los parámetros de la distribución compuesta $PC(\lambda; IG(\mu, \sigma))$: $\lambda = 20$, $\mu = 0.2$ y $\sigma = 2$. Este conjunto de valores arrojan una asimetría de la distribución compuesta $\gamma_S = 0.2577$, la cual puede ser evaluada empleando la ecuación dada en (1.10).

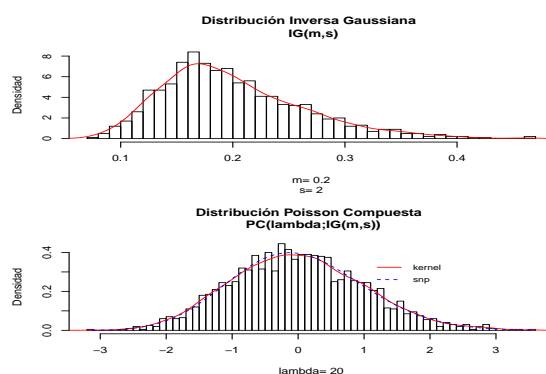


FIGURA 3.1: Gráficas de densidad: Distribución Inversa Gaussiana $IG(0.2, 2)$ y Distribución Compuesta $PC(20; IG(0.2, 2))$, con $\gamma_S < 2$. Ajuste Kernel y SNP_4 .

En la FIGURA 3.1 se muestra el comportamiento de la distribución Inversa Gaussiana con los valores de parámetros considerados, adicionalmente se presenta el histograma de la muestra de la distribución Poisson Compuesta $PC(20; IG(0.2, 2))$ y los ajustes de los estimadores de densidad kernel y SNP. Note que ambas curvas, nombradas como *kernel* y *snp* respectivamente, presentan un ajuste muy similar a los datos de la distribución compuesta.

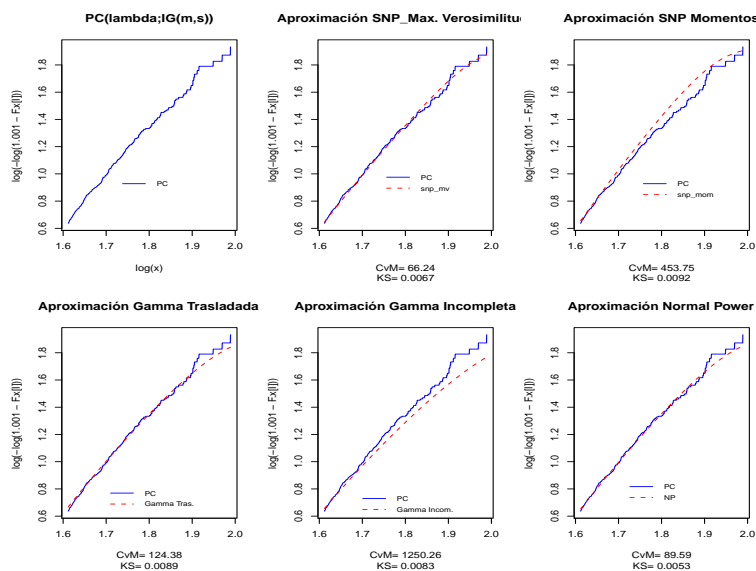


FIGURA 3.2: Comparación de las Aproximaciones a la Distribución Compuesta, cuando $S \sim PC(20; IG(0.2, 2))$ y $\gamma_S < 2$.

La FIGURA 3.2 permite analizar el desempeño de las aproximaciones estudiadas para la distribución compuesta. Cada una de las gráficas muestra el ajustes de las aproximaciones: SNP estimada por máxima verosimilitud (snp_{mv}), SNP por el método de los momentos (snp_{mom}), Gamma Traslada (*Gamma Tras.*), Gamma Incompleta (*Gamma Incom.*) y Normal Power (*NP*); adicionalmente cada una de las gráficas presenta el valor de las medidas de distancia empleadas, las cuales son presentadas en la TABLA 3.1. En esta tabla se puede evidenciar que la aproximación SNP, estimada por máxima verosimilitud presenta un buen ajuste, incluso, el mejor ajuste si se analiza la medida de distancia de Cramer-von Mises (CvM); analizando la medida de distancia de Kolmogorov Smirnov (KS) es la segunda aproximación que mejor se ajusta a los datos de la distribución compuesta. En este punto es importante aclarar que las medidas de distancia empleadas no pueden ser comparadas entre sí, ya que su construcción las hace muy distintas.

Con el fin de ampliar el análisis de la aproximación por medio de los estimadores SNP, se procede a aumentar el número de parámetros de la distribución SNP considerados para construir la aproximación. El ajuste de cada una de las distribuciones SNP construidas con 4, 6, 8 y 10 parámetros

TABLA 3.1: Medidas de ajuste de las Aproximaciones a la distribución $PC(20; IG(0.2, 2))$ cuando $\gamma_S < 2$.

Aproximación	CvM	KS
SNP_{LM}	66.24	0.0067
SNP_{MOM}	453.75	0.0091
Gamma Traslada	124.38	0.0089
Gamma Incompleta	1250.26	0.0083
NP	89.59	0.0053

a la distribución Poisson Compuesta es presentado en la FIGURA 3.3, en la cual se compara el ajuste de los estimadores kernel y la estimación SNP; es importante anotar, que a medida que el número de parámetros de la distribución SNP aumenta la estimación SNP se hace muy similar a la estimación por los kernel. Los resultados del ajuste con estas condiciones son presentados en la FIGURA 3.4 y en la TABLA 3.2.

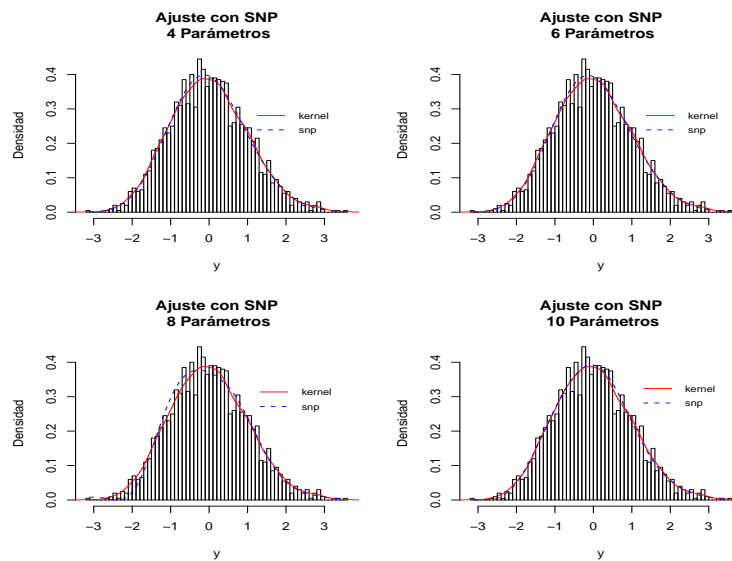


FIGURA 3.3: Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(20; IG(0.2, 2))$, con tamaño de muestra $n = 2000$ y $\gamma_S < 2$.

Note que la diferencia de probabilidad, calculada con la medida de distancia KS es más pequeña a medida que el número de parámetros de la distribución SNP aumenta.

Para analizar la aproximación SNP a medida que el tamaño de muestra de los datos iniciales

TABLA 3.2: Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(20; IG(0.2, 2))$, cuando $\gamma_S < 2$.

Aproximación	CvM	KS
SNP_4	66.24	0.0067
SNP_6	72.94	0.0073
SNP_8	234.58	0.0059
SNP_{10}	50.52	0.0050

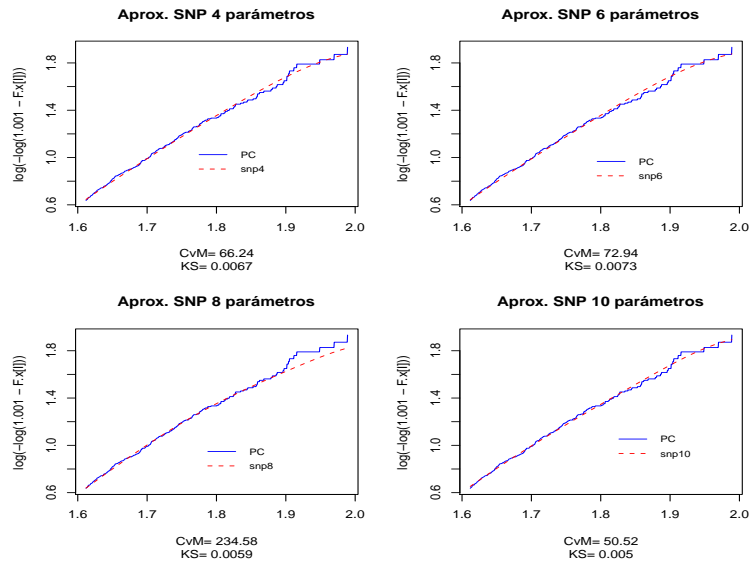


FIGURA 3.4: Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 2000$ y $\gamma_S < 2$.

de la distribución Poisson Compuesta aumenta, se evalúan las aproximaciones con tamaños de muestra $n = 4000$ y $n = 6000$. La TABLA 3.3 presenta los resultados de las medidas de distancia para cada una de las aproximaciones consideradas.

Un punto importante para analizar es que a medida que el tamaño de muestra aumenta, el ajuste parece mejorar para cada una de las aproximaciones consideradas. Las aproximaciones con mejor desempeño en este análisis son: SNP por máxima verosimilitud, la Gamma Traslada y la NP.

Con el fin de continuar con el análisis al incrementar el número de parámetros, se presenta la FIGURA 3.5 que muestra el ajuste de la distribución SNP a los datos simulados de la distribución Poisson Compuesta con tamaño de muestra $n = 6000$.

TABLA 3.3: Medidas de ajuste de las Aproximaciones a la distribución $PC(20; IG(0.2, 2))$, cuando $\gamma_S < 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$.

Tamaño muestra	$n = 4000$		$n = 6000$	
Aproximación	CvM	KS	CvM	KS
SNP_{LM}	359.82	0.0041	693.88	0.0047
SNP_{MOM}	2128.86	0.0095	6326.35	0.0084
Gamma Trasladada	468.87	0.0082	633.01	0.0039
Gamma Incompleta	4431.29	0.0094	7213.60	0.0093
NP	327.30	0.0043	604.41	0.0046

Note que para el tamaño de muestra $n = 6000$, el ajuste kernel y el de los estimadores SNP

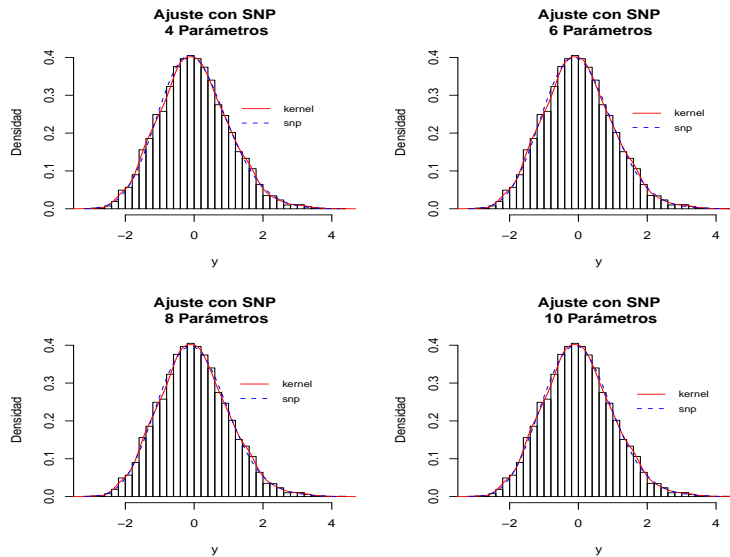


FIGURA 3.5: Gráficas de ajuste de los estimadores SNP con 4, 6, 8 y 10 parámetros de la distribución $PC(20; IG(0.2, 2))$, con tamaño de muestra $n = 6000$ y $\gamma_S < 2$.

son muy similares. Adicionalmente, la FIGURA 3.6 y la TABLA 3.4, dan evidencia que a medida que el tamaño de muestra aumenta, al igual que el número de parámetros, la aproximación SNP a los datos de la distribución Poisson Compuesta mejora su desempeño notablemente.

De la gráfica y la tabla anterior, se puede concluir que la aproximación SNP mejora a medida que el número de parámetros a considerar en esta distribución aumenta, a la igual que el tamaño de muestra; lo que comprueba la conclusión obtenida en Klaauw & Koning [26].

TABLA 3.4: Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(20; IG(0.2, 2))$ cuando $\gamma_S < 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$.

Tamaño muestra	$n = 4000$		$n = 6000$	
Aproximación	CvM	KS	CvM	KS
SNP_4	359.82	0.0041	693.88	0.0047
SNP_6	396.35	0.0056	683.04	0.0036
SNP_8	299.14	0.0062	392.94	0.0045
SNP_{10}	302.42	0.0062	453.99	0.0041

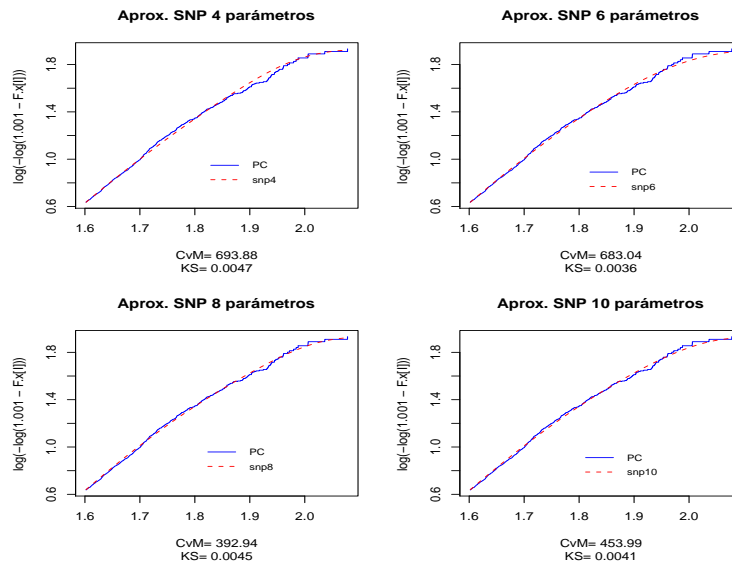


FIGURA 3.6: Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 6000$ y $\gamma_S < 2$.

3.1.2. Caso de asimetría $\gamma_S > 2$, cuando $S \sim PC(\lambda; IG(\mu, \sigma))$

Para evaluar el caso de asimetría alta, es decir, mayor al valor de 2, se consideran los siguientes valores de los parámetros para la distribución Poisson Compuesta $PC(\lambda; IG(\mu, \sigma))$: $\lambda = 10$, $\mu = 4$ y $\sigma = 0.5$; esta distribución tiene una asimetría de $\gamma_S = 2.5415$, la cual se halla a través de la ecuación (1.10).

En la FIGURA 3.7 se presenta la gráfica de la distribución Inversa Gaussiana, la cual evidentemente presenta una asimetría alta; adicionalmente, se presenta la gráfica de la distribución Poisson Compuesta $PC(10; IG(4, 0.5))$ y el ajuste de la estimación kernel y SNP con 4 parámetros. En este caso, el estimador de densidad SNP con 4 parámetros no presenta un buen desempeño comparado

con el estimador kernel.

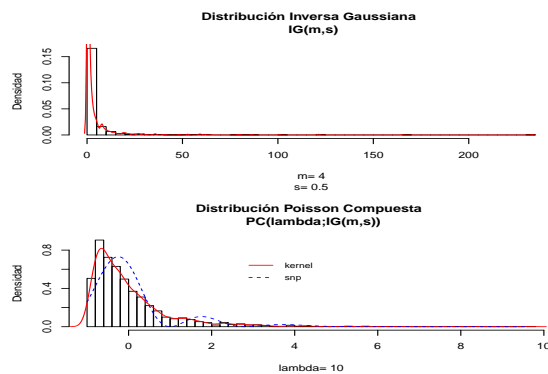


FIGURA 3.7: Gráficas de densidad: Distribución Inversa Gaussiana $IG(4,0.5)$ y Distribución Compuesta $PC(10; IG(4.0.5))$, con $\gamma_S > 2$. Ajuste Kernel y SNP_4 .

El desempeño de las aproximaciones evaluadas presenta dificultades para el caso de asimetría alta, la FIGURA 3.8 evidencia una diferencia alta entre la función de distribución acumulada de la Poisson Compuesta y las aproximaciones. De igual manera a través de la TABLA 3.5, se puede concluir que la aproximación que mejor estima los datos de la distribución Poisson Compuesta, $PC(10; IG(4,0.5))$, es la aproximación Gamma Traslada.

La evaluación de la distribución $PC(10; IG(4.0.5))$ también es realizada a través del aumento

TABLA 3.5: Medidas de ajuste de las Aproximaciones a la distribución $PC(10; IG(4.0.5))$ cuando $\gamma_S > 2$.

Aproximación	CvM	KS
SNP_{LM}	6630.51	0.0287
SNP_{MOM}	23595.20	0.1310
Gamma Traslada	1591.83	0.0069
Gamma Incompleta	5000.79	0.0471
NP	15420.43	0.0524

de parámetros de la distribución SNP; la FIGURA 3.9 presenta el ajuste de la SNP con 4, 6, 8, y 10 parámetros. Es evidente, a través de esta gráfica, que a medida que aumenta el número de parámetros considerados, el ajuste a la distribución Poisson Compuesta mejora.

La anterior conclusión se evidencia también en la FIGURA 3.10, en la cual se puede observar que la línea punteada se ajusta mejor a la línea continua, que corresponde a la distribución Poisson

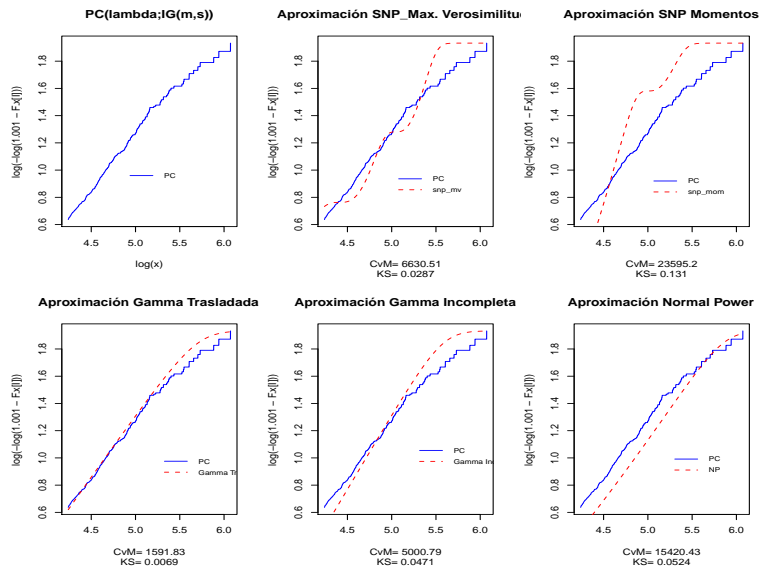


FIGURA 3.8: Comparación de Aproximaciones a las Distribuciones Compuestas, cuando $S \sim PC(10; IG(4.0.5))$ y $\gamma_S > 2$.

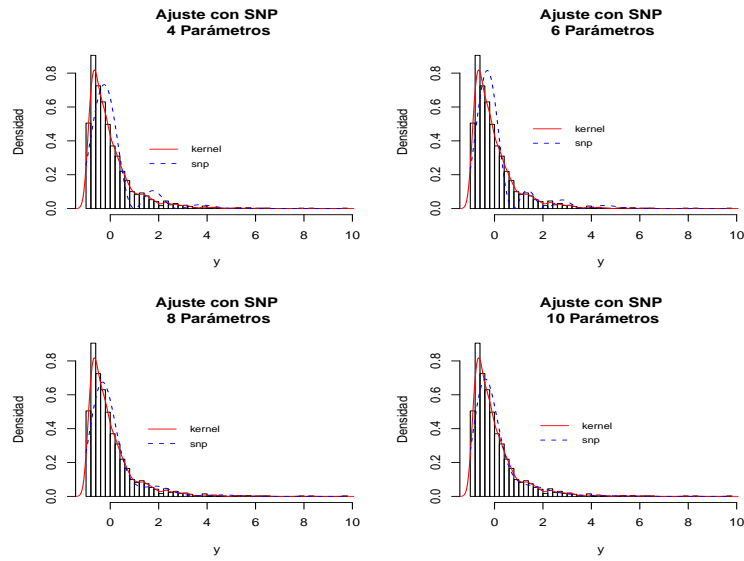


FIGURA 3.9: Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 2000$ y $\gamma_S > 2$.

Compuesta. En la tabla TABLA 3.6, se puede observar cómo el valor del KS disminuye a medida que el número de parámetros aumenta; igualmente, el valor de la medida de distancia CvM tiende a disminuir. Con estos resultados se puede concluir, que la asimetría es un factor determinante para asegurar el buen desempeño de las aproximaciones, ya que mientras mayor valor tenga, mayor

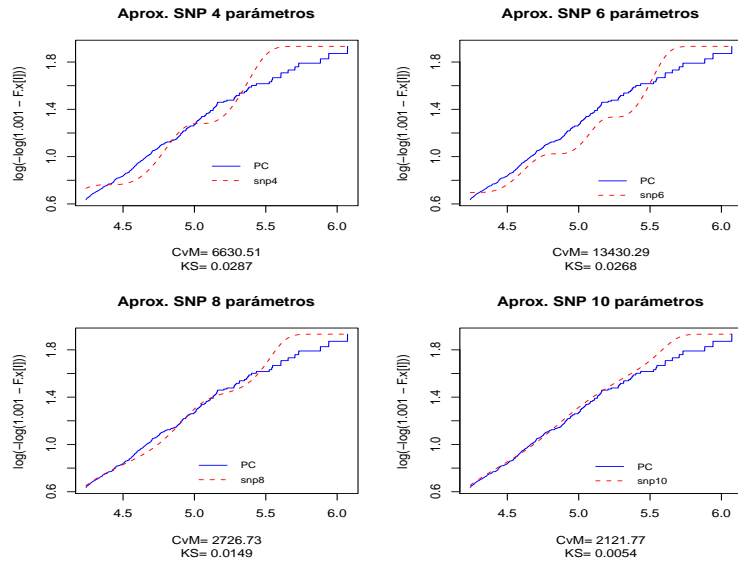


FIGURA 3.10: Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 2000$ y $\gamma_S > 2$.

TABLA 3.6: Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(20; IG(0.2, 2))$, cuando $\gamma_S < 2$.

Aproximación	CvM	KS
SNP_4	6630.51	0.0287
SNP_6	13430.29	0.0268
SNP_8	2726.73	0.0149
SNP_{10}	2121.77	0.0054

dificultad presentan las aproximaciones.

En la TABLA 3.7, se puede observar que las medidas de distancia disminuyen mientras el tamaño de muestra aumenta. Una conclusión similar puede ser obtenida a través FIGURA 3.11 y 3.12, en las cuales se observa que la calidad de ajuste mejora a medida que el número de parámetros aumente en la distribución SNP.

La estimación de la aproximación SNP por el método de los momentos no presenta un buen ajuste, lo que permite concluir que existen problemas de convergencia en el momento de estimación de los parámetros. La alternativa de estimación por máxima verosimilitud, resulta más eficiente. Finalmente, la TABLA 3.8, confirma la conclusión obtenida anteriormente, el número de parámetros de la distribución SNP está completamente relacionado con la eficiencia de los estimadores de densidad

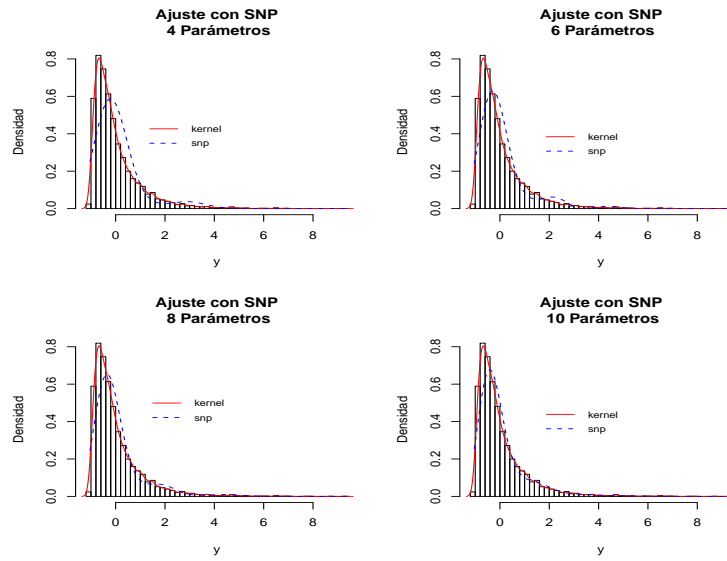


FIGURA 3.11: Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(10; IG(4, 0.5))$, con tamaño de muestra $n = 6000$ y $\gamma_S > 2$.

TABLA 3.7: Medida de ajuste de las Aproximaciones a la distribución $PC(10; IG(4, 0.5))$ cuando $\gamma_S > 2$, y los tamaños de muestra son: $n = 4000$ y $n = 6000$.

Tamaño muestra	$n = 4000$		$n = 6000$	
Aproximación	CvM	KS	CvM	KS
SNP_{LM}	15130.17	0.0474	59082.92	0.0185
SNP_{MOM}	76231.23	0.1071	160339.11	0.0929
Gamma Traslada	2798.65	0.0095	4388.48	0.0085
Gamma Incompleta	13381.82	0.0336	28884.42	0.0264
NP	30738.78	0.0438	78517.15	0.0386

SNP, tal como lo presentan Klaauw & Koning [26].

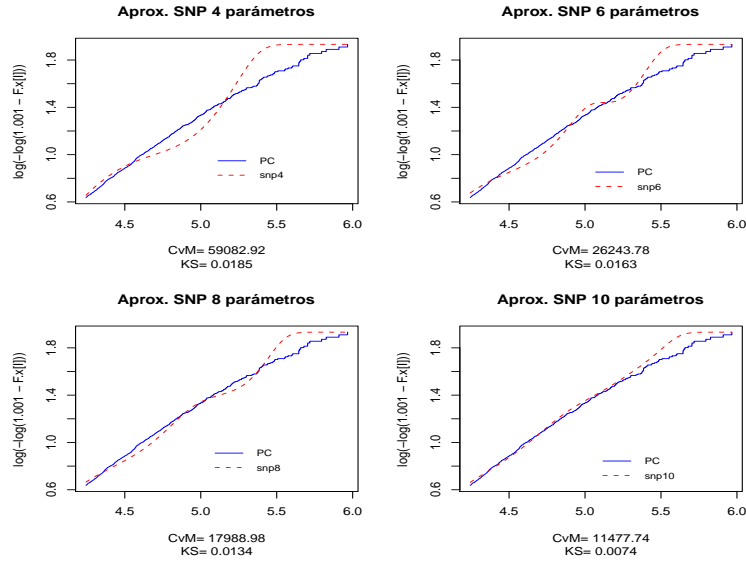


FIGURA 3.12: Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 6000$ y $\gamma_S > 2$.

TABLA 3.8: Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros, a la distribución $PC(10; IG(4, 0.5))$ cuando $\gamma_S < 2$, y los tamaños de muestra son $n = 4000$ y $n = 6000$.

Tamaño muestra	$n = 4000$		$n = 6000$	
Aproximación	CvM	KS	CvM	KS
SNP_4	15130.17	0.0474	59082.92	0.0185
SNP_6	18217.56	0.0325	26243.78	0.0163
SNP_8	5948.26	0.0117	17988.98	0.0134
SNP_{10}	3101.51	0.0047	11477.74	0.0074

3.2. Resultados de la aproximación SNP a la $PC(\lambda; LN(\mu, \sigma))$

3.2.1. Caso de asimetría $\gamma_S < 2$, cuando $S \sim PC(\lambda; LN(\mu, \sigma))$

Con el objetivo de generar un caso particular en el cual se cumpla que $\gamma_S < 2$, cuando $N \sim Poisson(\lambda)$ y $X \sim LN(\mu, \sigma)$, se escogieron los siguientes valores de los parámetros: $\lambda = 10$, $\mu = 1$ y $\sigma = 0.2$; esta distribución genera una asimetría de $\gamma_S = 0.3357$. La FIGURA 3.13 permite observar la distribución $LN(1, 0.2)$ y su ajuste Kernel, adicionalmente, presenta la gráfica de la distribución $PC(10; LN(1, 0.2))$ con los ajustes Kernel y SNP. Las gráficas evidencian el valor pequeño de asimetría en los datos a través de una $PC(10; LN(1, 0.2))$.

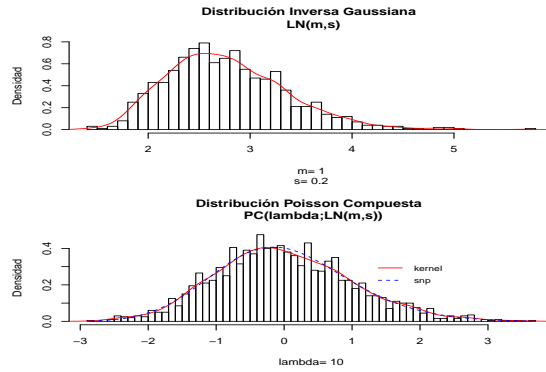


FIGURA 3.13: Gráficas de densidad: Distribución Log-Normal $LN(1,0.2)$ y Distribución Compuesta $PC(10;LN(1,0.2))$, con $\gamma_S < 2$. Ajuste Kernel y SNP_4 .

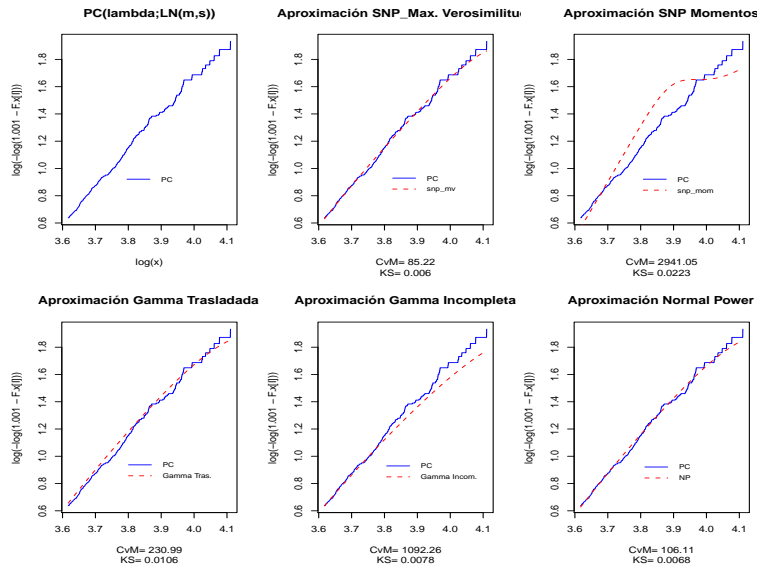


FIGURA 3.14: Comparación de Aproximaciones a las Distribuciones Compuestas, cuando $S \sim PC(10;LN(1,0.2))$ y $\gamma_S < 2$.

La FIGURA 3.14 y la TABLA 3.9 dan evidencia del ajuste de las aproximaciones a la distribución Poisson Compuesta, cuando $X \sim LN$. La aproximación SNP por estimación de máxima verosimilitud, arroja uno de los mejores resultados en el ajuste, en ambas medidas de distancia empleadas. La aproximación SNP por estimación de igualdad de momentos no presenta buenos resultados, lo que permite intuir que esta estimación presenta problemas de convergencia en la solución del sistema de ecuaciones.

Las aproximaciones NP y Gamma Traslada presentan unos buenos resultados en el ajuste de

TABLA 3.9: Medidas de ajuste de las Aproximaciones a la distribución $PC(10; LN(1, 0.2))$ cuando $\gamma_S < 2$.

Aproximación	CvM	KS
SNP_{LM}	85.22	0.0060
SNP_{MOM}	2941.05	0.0223
Gamma Traslada	230.99	0.0106
Gamma Incompleta	1092.26	0.0078
NP	106.11	0.0068

la distribución Poisson Compuesta, cuando la $\gamma_S < 2$.

Continuando con el propósito de analizar el aumento de parámetros en la distribución SNP para evaluar el ajuste a la distribución Poisson Compuesta, se presenta la FIGURA 3.15 en la cual se puede observar que el ajuste de la aproximación SNP por máxima verosimilitud mejora a medida que el número de parámetros considerados en el análisis aumentan; adicionalmente, se observa una vez más que los estimadores Kernel y SNP son muy parecidos para el caso de asimetría baja, es decir, $\gamma_S < 2$.

La TABLA 3.10 y la FIGURA 3.16 verifican una vez más la conclusión obtenida anteriormente. La aproximación SNP para este caso arroja resultados muy buenos de ajuste.

TABLA 3.10: Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(10; LN(1, 0.2))$, cuando $\gamma_S < 2$.

Aproximación	CvM	KS
SNP_4	85.22	0.0060
SNP_6	70.99	0.0048
SNP_8	69.31	0.0050
SNP_{10}	66.81	0.0054

Al aumentar el tamaño de muestra, los resultados de la aproximación SNP mejoran notablemente, y si adicionalmente se aumenta el número de parámetros de la distribución SNP, el ajuste gana mayor precisión.

Adicionalmente, el tamaño de muestra aumenta para analizar los resultados, la TABLA 3.11 presenta las medidas de distancia de las aproximaciones consideradas cuando el tamaño de muestra es $n = 4000$ y $n = 6000$.

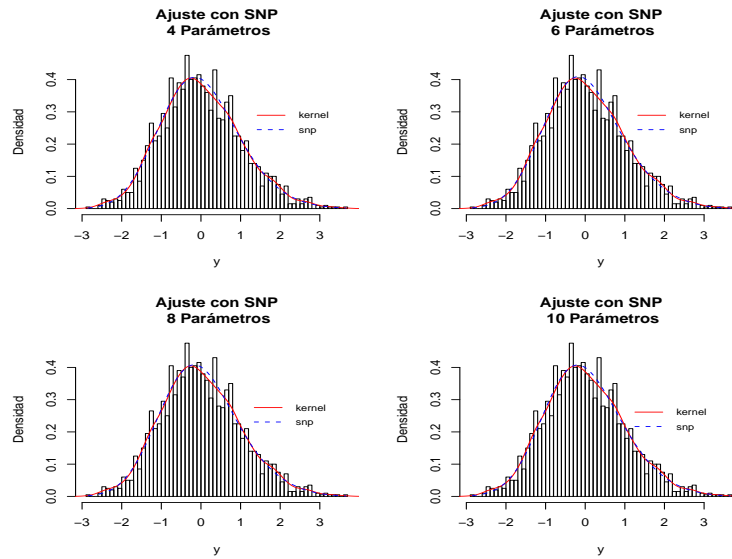


FIGURA 3.15: Gráficas de densidad de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(10; LN(1, 0.2))$, con tamaño de muestra $n = 2000$ y $\gamma_S < 2$.

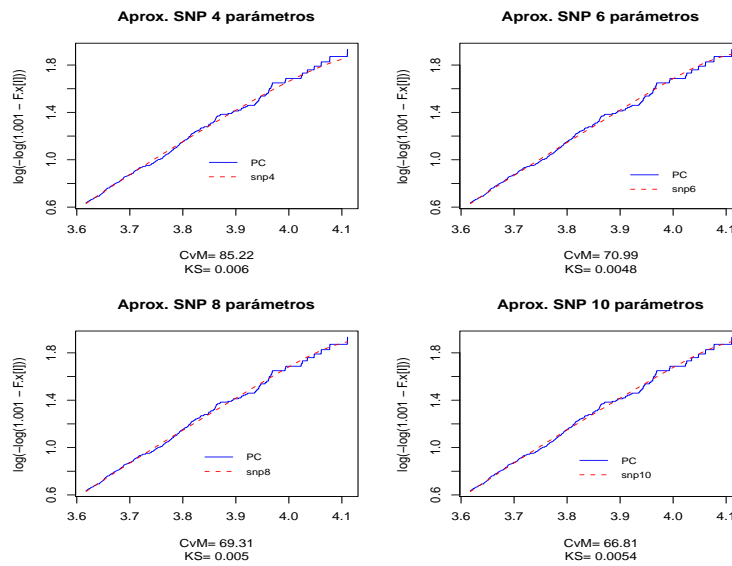


FIGURA 3.16: Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 2000$ y $\gamma_S < 2$.

La FIGURA 3.17 presenta los histogramas de la distribución $PC(10; LN(1, 0.2))$ y el ajuste para cada uno de los parámetros considerados para la distribución SNP. En esta gráfica se observa que la diferencia entre los estimadores Kernel y SNP es muy pequeña.

TABLA 3.11: Medidas de ajuste de las Aproximaciones a la distribución $PC(10; LN(1, 0.2))$, cuando $\gamma_S < 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$.

Tamaño muestra	$n = 4000$		$n = 6000$	
Aproximación	CvM	KS	CvM	KS
SNP_{LM}	752.69	0.0033	246.21	0.0022
SNP_{MOM}	21565.48	0.0194	21665.72	0.0193
Gamma Trasladada	1537.43	0.0039	1998.10	0.0058
Gamma Incompleta	9548.52	0.0055	10353.28	0.0079
NP	1647.04	0.0040	135.81	0.0023

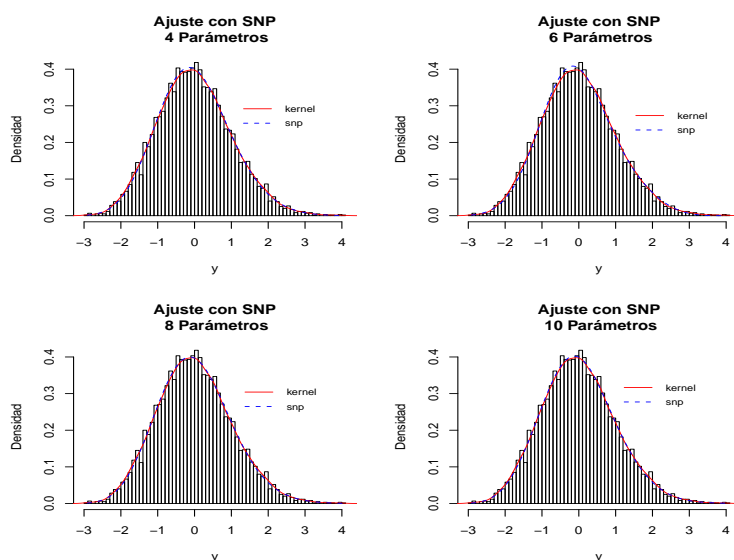


FIGURA 3.17: Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(10; LN(1, 0.2))$, con tamaño de muestra $n = 6000$ y $\gamma_S < 2$.

Finalmente, en la FIGURA 3.18 y la TABLA 3.12 permiten afirmar las conclusiones realizadas anteriormente. Note que en la FIGURA 3.18 cuando la aproximación se construye con 8 parámetros, el ajuste a la distribución Poisson Compuesta es muy bueno.

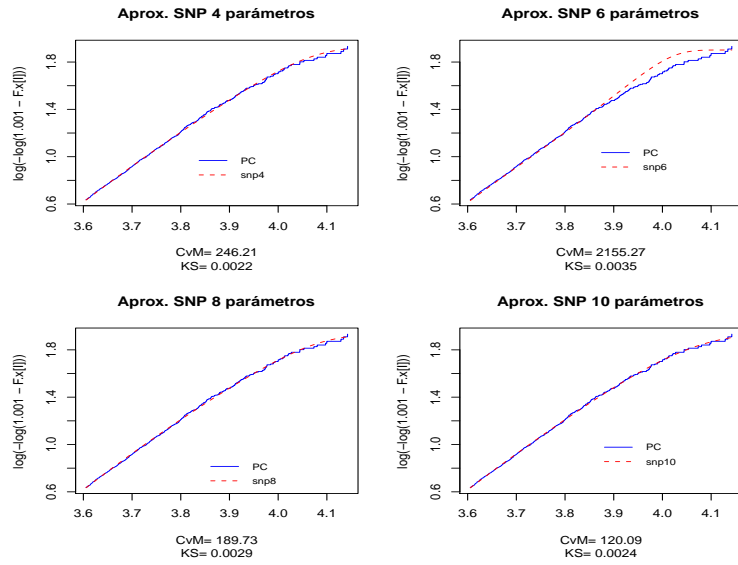


FIGURA 3.18: Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 6000$ y $\gamma_S < 2$.

TABLA 3.12: Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(10; LN(1, 0.2))$ cuando $\gamma_S < 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$.

Tamaño muestra	$n = 4000$		$n = 6000$	
Aproximación	CvM	KS	CvM	KS
SNP_4	752.69	0.0033	246.21	0.0022
SNP_6	643.29	0.0032	2155.27	0.0035
SNP_8	639.19	0.0029	189.73	0.0029
SNP_{10}	186.82	0.0037	120.09	0.0024

3.2.2. Caso de asimetría $\gamma_S > 2$, cuando $S \sim PC(\lambda; LN(\mu, \sigma))$

Como anteriormente se mencionó, el interés de esta revisión práctica es analizar la eficiencia del método de aproximación SNP para valores de asimetría inferiores y superiores al valor de 2. En este caso, se seleccionaron los siguientes valores de los parámetros con el fin de obtener una asimetría de la distribución Poisson Compuesta mayor a 2, $\gamma_S > 2$; por lo tanto, $\lambda = 15$, $\mu = 3$ y $\sigma = 1.3$. La asimetría obtenida con estos parámetros es $\gamma_S = 3.2575$, a través de la ecuación (1.10).

En la FIGURA 3.19 se puede observar la asimetría de los datos de una muestra Log-Normal con éstos parámetros; de igual manera, se presenta el ajuste de los estimadores Kernel y SNP a la distribución Poisson Compuesta $PC(15; LN(3, 1.3))$.

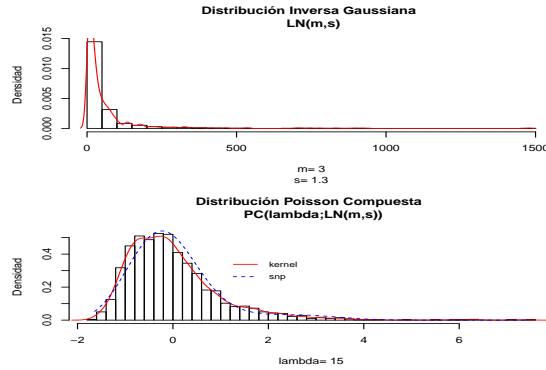


FIGURA 3.19: Gráficas de densidad: Distribución Log-Normal $LN(3, 1.3)$ y Distribución Compuesta $PC(15; LN(3, 1.3))$. Ajuste Kernel y SNP_4 .

El ajuste de las aproximaciones evaluadas son presentadas en FIGURA 3.20 y las medidas de distancia están dadas en TABLA 3.13. Note que las medidas de distancia sugieren dos aproximaciones: la medida de distancia CvM presenta un menor valor de distancia en la aproximación Gamma Incompleta, mientras arroja la mínima distancia para la aproximación SNP estimada por máxima verosimilitud. En general, las aproximaciones no presentan una estimación tan buena, como cuando la asimetría $\gamma_S > 2$.

La revisión de este caso se realiza de forma similar a los anteriores, en los cuales se revisan las aproximaciones de la distribución SNP aumentando los parámetros y el tamaño de muestra. Note de la TABLA 3.14 que a medida que el número de los parámetros aumenta, las medidas de distancia, CvM y KS disminuyen, lo que confirma de nuevo las conclusiones anteriores; gráficamente, FIGURA 3.21 y 3.22, se puede percibir que la aproximación SNP mejora, sin embargo, presenta dificultades para la estimación en la cola de la distribución Poisson Compuesta.

TABLA 3.13: Medidas de ajuste de las Aproximaciones a la distribución $PC(15; LN(3, 1.3))$ cuando $\gamma_S > 2$.

Aproximación	CvM	KS
SNP_{LM}	3633.83	0.0122
SNP_{MOM}	866166.52	0.2032
Gamma Traslada	8952.95	0.0170
Gamma Incompleta	1416.55	0.0183
NP	6104,97	0,0383

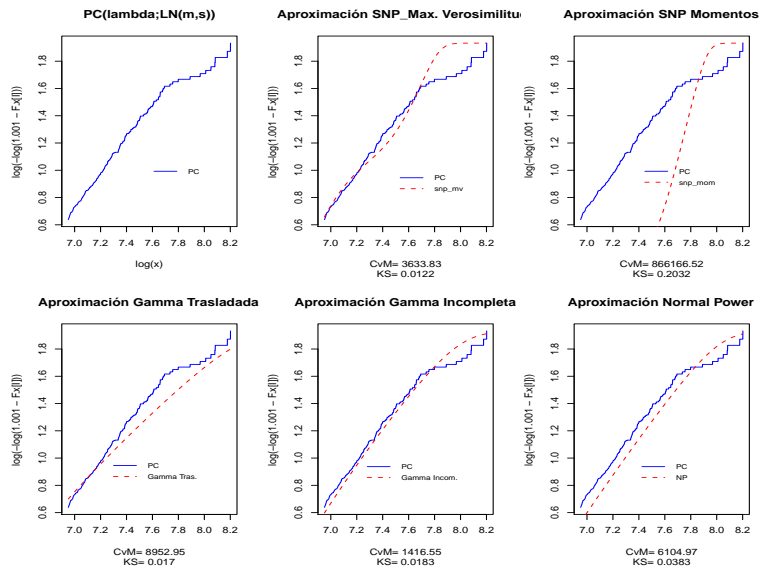


FIGURA 3.20: Comparación de Aproximaciones a las Distribuciones Compuestas, cuando $S \sim PC(15; LN(3, 1.3))$ y $\gamma_S > 2$.

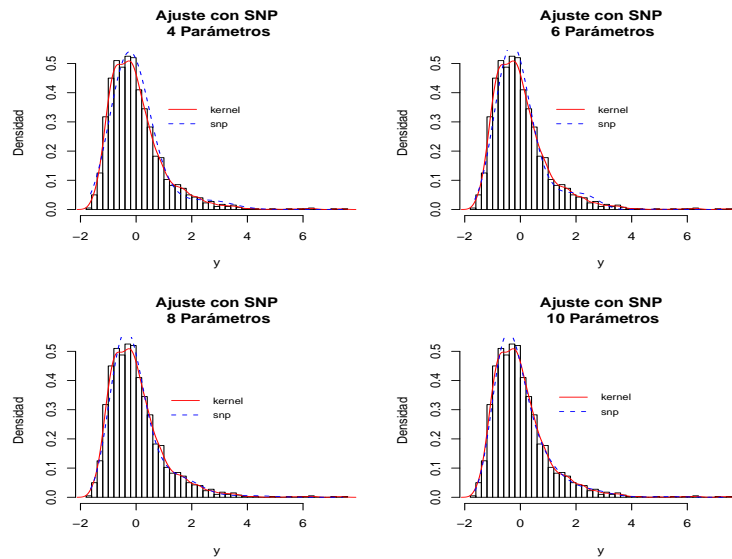


FIGURA 3.21: Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(15; LN(3, 1.3))$, con tamaño de muestra $n = 2000$ y $\gamma_S > 2$.

Al aumentar el tamaño de muestra las medidas de distancia de las aproximaciones empleadas para comparar con la aproximación SNP, no presentan variante significativas que indiquen que se mejora el ajuste al aumentar el tamaño de muestra. Mientras, al aumentar el tamaño de muestra para la aproximación SNP se evidencia un mejor ajuste a las distribuciones Poisson Compuesta,

TABLA 3.14: Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(15; LN(3, 1.3))$. cuando $\gamma_S > 2$.

Aproximación	CvM	KS
SNP_4	3633.83	0.0122
SNP_6	4619.00	0.0098
SNP_8	1536.36	0.0060
SNP_{10}	956.87	0.0055

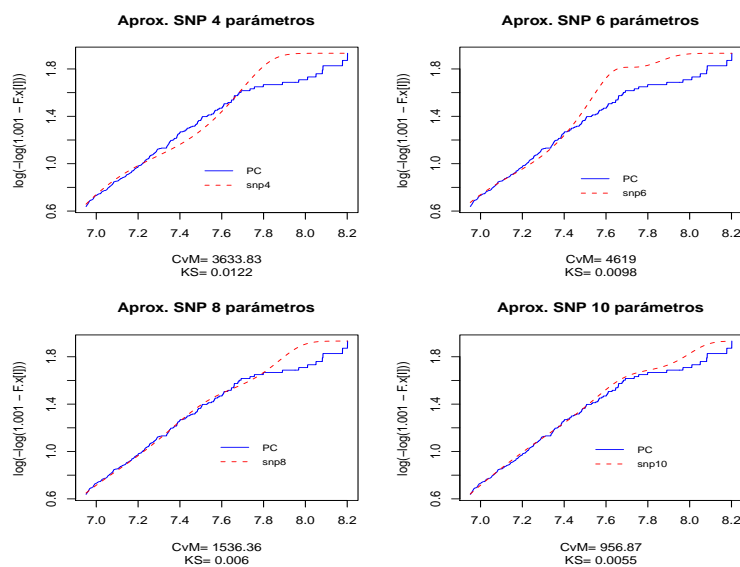


FIGURA 3.22: Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 2000$ y $\gamma_S > 2$.

TABLA 3.15: Medidas de ajuste de las Aproximaciones a la distribución $PC(15; LN(3, 1.3))$, cuando $\gamma_S > 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$.

Tamaño muestra	$n = 4000$		$n = 6000$	
	CvM	KS	CvM	KS
SNP_{LM}	11489.62	0.0147	30916.62	0.0119
SNP_{MOM}	3120932.14	0.2041	6505350.82	0.2057
Gamma Trasladada	33724.04	0.0258	48955.32	0.0174
Gamma Incompleta	3306.08	0.0104	17851.03	0.0151
NP	16705.68	0.0250	98495.85	0.0359

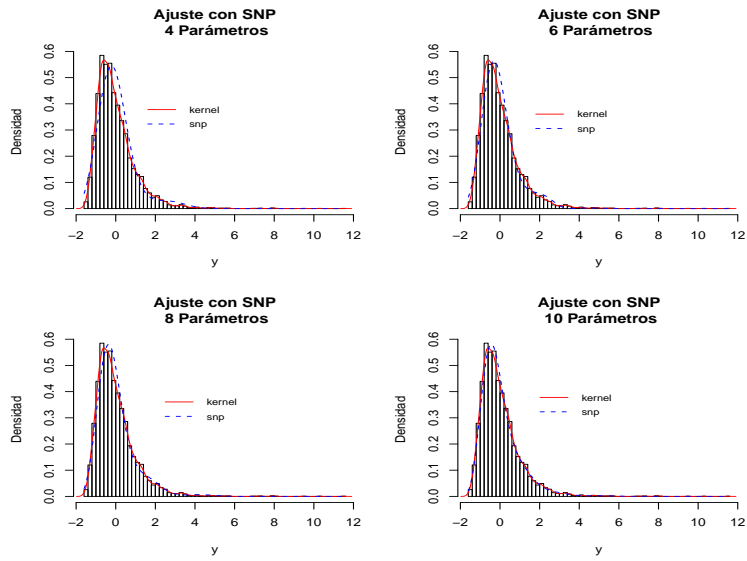


FIGURA 3.23: Gráficas de ajuste de los estimadores Kernel y SNP con 4, 6, 8 y 10 parámetros a la $PC(15; LN(3, 1.3))$, con tamaño de muestra $n = 6000$ y $\gamma_S > 2$.

vea FIGURA 3.24 y TABLA 3.16.

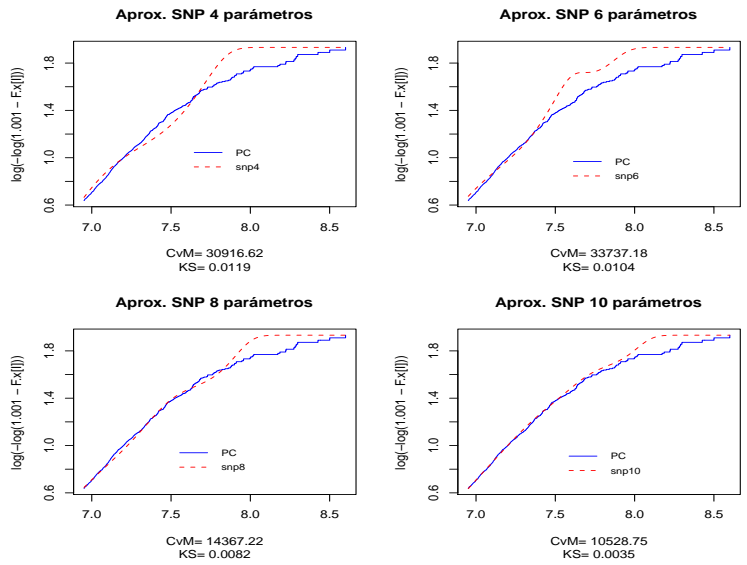


FIGURA 3.24: Gráficas de aproximación de los estimadores de densidad SNP con 4, 6, 8 y 10 parámetros, con tamaño de muestra $n = 6000$ y $\gamma_S > 2$.

TABLA 3.16: Medidas de ajuste de las Aproximaciones SNP con 4, 6, 8 y 10 parámetros a la distribución $PC(15; LN(3, 1.3))$ cuando $\gamma_S < 2$ y los tamaños de muestra son $n = 4000$ y $n = 6000$.

Tamaño muestra	$n = 4000$		$n = 6000$	
Aproximación	CvM	KS	CvM	KS
SNP_4	11489.62	0.0147	30916.62	0.0119
SNP_6	7684.45	0.0115	33737.18	0.0104
SNP_8	4350.69	0.0080	14367.22	0.0082
SNP_{10}	2519.90	0.0037	10528.75	0.0035

Capítulo 4

Aplicaciones

La distribución Poisson Compuesta ha sido empleada en temas relacionados con el análisis de riesgos y de seguros, que comunmente involucran un periodo de tiempo para el análisis de las posibles ocurrencias, siniestralidades o eventos. Las aplicaciones mencionadas en este capítulo están relacionadas con el estudio de los riesgos operacionales y la aplicación en seguros. Sin embargo, otras aplicaciones se encuentran dirigidas al análisis de supervivencia, Aalen [1].

Panjer [32], se concentra en la aplicación y modelación del riesgo operacional, incluyendo la aplicación de las distribuciones compuestas y exponiendo los detalles de la metodología por él propuesta.

En las siguientes secciones se presenta la aplicación de la distribución Poisson Compuesta al tema de riesgo operacional y al modelo de riesgo colectivo.

4.1. Aplicación al Riesgo Operacional

Las corporaciones financieras y la industria en general, se encuentran expuestas a riesgos de operación que puede fácilmente llevar a la quiebra a cualquier compañía. Algunas entidades financieras de los países más desarrollados se acogen al Acuerdo de Basilea II, que define los riesgos operacionales a los que se encuentran expuestas las entidades bancarias y sugieren algunos tratamientos para este riesgo.

En el Acuerdo de Capitales de Basilea II ¹, el riesgo operacional se define como: «El riesgo de

¹El Comité de Supervisión Bancaria de Basilea fue creado en 1975 por los dirigentes de los bancos centrales del Grupo de los Diez (G10), y está compuesto por autoridades de supervisión bancaria. Dicho Comité construye acuerdos con el fin de dar recomendaciones sobre la legislación y regulación bancaria, proponiendo un referente base

pérdida resultante por fallas en los procesos internos, humanos y de los sistemas o por eventos externos. Esta definición incluye el riesgo legal ², dejando por fuera el riesgo estratégico y reputacional»³.

Según la definición de Basilea II, este tipo de riesgo consta de siete grandes categorías de eventos, los cuales son considerados como las principales causas de las pérdidas operacionales en las entidades financieras:

- Fraude interno
- Fraude externo
- Prácticas de empleo, salud y seguridad en el trabajo
- Prácticas con clientes, productos, y de negocio
- Daños en activos físicos: hacen referencia a pérdidas o daños en activos físicos de la entidad, originados por desastres naturales u otros sucesos.
- Interrupción del negocio y fallas en los sistemas
- Ejecución, entrega y gestión de los procesos

Para efectos de clasificar las actividades del sector financiero, el Comité de Basilea II considera las siguientes líneas de negocio:

- Finanzas Corporativas
- Negociación y Ventas
- Banca Minorista
- Banca Comercial
- Pagos y Liquidaciones

a las entidades con objeto de establecer los requerimientos de capital necesarios para asegurar la protección frente a los riesgos financieros y operativos. Por lo tanto, el Acuerdo de Basilea II publicado en Junio de 2004, se ha vuelto un referente para los entes reguladores de las entidades.

» ²Riesgo legal: Es la posibilidad de pérdida en que incurre una entidad al ser sancionada u obligada a indemnizar daños como resultado del incumplimiento de normas o regulaciones y obligaciones contractuales.

³Riesgo Reputacional: Es la posibilidad de pérdida en que incurre una entidad por desprestigio, mala imagen, publicidad negativa, cierta o no, respecto de la institución y sus prácticas de negocios, que cause pérdida de clientes, disminución de ingresos o procesos judiciales. La Superintendencia Financiera de Colombia incluye este riesgo en la definición o concepción del riesgo operacional.

- Servicios de Agencia
- Administración de Activos
- Intermediación Minorista

Las anteriores definiciones de 7 tipos de eventos y 8 líneas de negocios, son establecidas con el fin de dar consistencia y claridad en la recolección de los datos y de estandarizarlos para posibles comparaciones.

El tratamiento y la cuantificación del riesgo operacional ha tomado mayor fuerza en los últimos años y varios autores han propuesto metodologías que permiten exponer estadísticamente el comportamiento de las pérdidas obtenidas por este tipo de riesgo. Embrechts, *et al* [11] exponen una discusión acerca de los métodos para cuantificar el riesgo operacional en los bancos y mencionan posibles metodologías para su tratamiento, entre ellas: teoría de valores extremos, colas pesadas y la probabilidad de ruina. Moscadelli [30] menciona en su artículo la importancia de tener en una entidad expuesta al riesgo operacional una adecuada administración y un buen sistema de medición, y analiza dos metodologías: las distribuciones actuariales y los modelos basados en la Teoría de Valor Extremo. Otros autores como Panjer [32], han propuesto metodologías para tratar los datos provenientes de este tipo de riesgo. Chavez- Demoulin, *et al.* [7] presenta un ejemplo de la aplicación de la distribución Poisson Compuesta para aplicación de modelos avanzados mencionados en Basilea II.

Consideramos que la aproximación SNP puede ser una alternativa buena para aproximar la distribución de los riesgos operaciones de cualquier entidad, siempre y cuando los valores de pérdida no sean muy extremos, ya que de ser así, el mejor tratamiento sería el propuesto por Moscadelli [30]: particionar el comportamiento de las pérdidas en dos distribuciones.

4.2. Modelos de Riesgo Colectivo

El modelo básico de Riesgo Colectivo o de Cramer-Lundberg se define como un proceso estocástico en tiempo continuo, $R(t), t \geq 0$, con valores en \mathbf{R} , definido mediante las siguientes variables:

- $(N(t), t \geq 0)$: proceso Poisson con tasa λ .
- $(X_j)_{j=1,2,\dots}$: sucesión de variables aleatorias i.i.d., con distribución $F(x) = P(X_j \leq x)$, independientes de $N(t)$.
- $S(t) = \sum_{j=1}^{N(t)} X_j$: suma aleatoria de variables aleatorias, que conforman la distribución compuesta.

El Modelo de Riesgo Colectivo se define como

$$R(t) = u + \Pi t - S(t), \quad t \geq 0 \quad (4.1)$$

donde $u, \Pi > 0$ son constantes dadas.

Se han definido varias extensiones de este modelo, entre las cuales se pueden mencionar el modelo de Sparre Andersen, que tiene la misma forma del modelo: $R(t) = u + \Pi t - \sum_{j=1}^{N(t)} X_j$, pero donde $\{N(t)\}_{t \geq 0}$ es un proceso en tiempo continuo generado por una sucesión de variables aleatorias i.i.d. no negativas, denominadas “tiempo entre arribo” $\{T_j\}_{j \geq 0}$, de tal forma que

$$N(t) = \inf\{j \geq 0 : T_0 + \dots + T_j \geq t\},$$

Se asume que las variables aleatorias $\{T_j\}_{j \geq 1}$ y $\{X_j\}_{j \geq 1}$ son independientes.

Las variables definidas tienen las siguientes interpretaciones, en el contexto de seguros de daños:

- $(N(t), t \geq 0)$: el número de reclamos hasta el tiempo t .
- $(X_j)_{j=1,2,\dots}$: el costo de cada reclamo, con $\mu = E(X_j)$.
- $S(t) = \sum_{j=1}^{N(t)} X_j$: el costo total acumulado hasta t .
- u : reserva inicial de capital, $R(0) = s$.
- Π : prima anual.
- $u + \Pi t - S(t) =$ “ingresos - gastos”.
- $R(t) = u + Pt - S(t)$: superávit hasta el tiempo t .

En el caso especial en el que las X_j se distribuyen exponencialmente y $N(t)$ se distribuye Poisson tiene una expresión explícita para la distribución compuesta de $S(t)$. El modelo de riesgo colectivo en las aplicaciones se utiliza para un período anual con $t = 1$, con el fin de analizar el cálculo de Π dado el nivel de reservas u y la distribución de X_j . También se utiliza en tiempo continuo $t \geq 0$ con el fin de analizar la probabilidad del evento de insolvencia ($R(t) < 0$) en un período $t \in [0, T]$. El interés en este trabajo es aproximar la distribución compuesta $S(t)$ en un período de tiempo anual $t = 1$. En adelante se denotaran las variables del modelo sin referencia al tiempo, por ejemplo S , en lugar de $S(t)$, etc..

Escalante & Arango [13] presentan de manera más detalla la teoría que soporta los aspectos básicos del modelo de riesgo colectivo.

Capítulo 5

Conclusiones

En el desarrollo de esta tesis se comprueba que los estimadores de densidad SNP pueden ser empleados para estimar las distribuciones compuestas, siempre y cuando cumpla con las condiciones que se especifican en la sección 2.1.1. Adicionalmente, la implementación de esta aproximación no resulta muy compleja para desarrollar (en el apéndice A se presentan los programas con los cuales se ejecutó la aproximación SNP).

La aproximación SNP propuesta, estimada por máxima verosimilitud, presenta muy buenos resultados al compararla con algunas de las aproximaciones mencionadas. La aproximación SNP construida a partir del método de igualdad de momentos no presenta los mismos resultados, ya que su solución es más compleja; por esta razón, no se desarrolló la aproximación para más parámetros, ya que la solución del sistema de ecuaciones, conformado por los momentos de las distribuciones compuestas y SNP, es más complejo a medida que se vinculan a la estimación más parámetros. Por lo tanto, se sugiere emplear el método de estimación por máxima verosimilitud para construir la aproximación SNP a las distribuciones compuestas.

A través de los resultados gráficos, se puede concluir que los estimadores de densidad Kernel y SNP son muy similares, esta conclusión se obtuvo por Fenton & Gallant [16]. Este resultado le da impulso teórico a la aproximación propuesta en esta tesis, ya que se comprueba que los estimadores SNP presentan un buen desempeño para plantear aproximaciones.

Adicionalmente, se verifican las conclusiones obtenidas por Klaauw & Koning [26], las cuales señalan que los estimadores de densidad SNP mejoran su desempeño y ajuste, a medida que el número de parámetros considerados en la distribución aumentan; la mayoría de los casos prácticos desarrollados permiten verificar estas conclusiones. Además, el tamaño de muestra presenta en ocasiones un efecto favorable en la estimación de las distribuciones.

Al comparar los resultados obtenidos en el análisis de las aproximaciones Gamma Traslada, Gamma Incompleta, NP y SNP, con las conclusiones obtenidas por las referencias, concluimos:

- La asimetría es un factor determinante en la exactitud del ajuste de las aproximaciones, ya que los resultados al aplicar las aproximaciones estudiadas demuestran que se obtiene mejor desempeño cuando γ_S es menor a 2 y se presentan dificultades al estimar la parte superior de la cola derecha de la distribución Poisson Compuesta cuando $\gamma_S > 2$. Esta conclusión coincide con los resultados obtenidos por Gendron & Crepeau [22].
- Los resultados no coinciden con las conclusiones obtenidas por Seal [40], ya que es evidente que el desempeño de la aproximación Gamma Incompleta es dependiente del valor de la asimetría, tal como lo concluye Pentikainen [33].
- Se coincide también con Gendron & Crepeau [22] en la conclusión que hace referencia a que la aproximación NP subestima el valor exacto de la distribución Poisson Compuesta.
- En la ejecución práctica de las aproximaciones se observa que el desempeño de las aproximaciones es muy similar al de la aproximación SNP, sin embargo, la aproximación SNP mejora al aumentar el número de parámetros considerados. Esta conclusión es más explícita cuando el valor de la asimetría es alto.

Finalmente, esta tesis brinda una nueva aproximación a las distribuciones compuestas, la cual es fácil de calcular y por medio de la cual se obtienen buenas aproximaciones, sobretodo cuando el número de parámetros considerados en la distribución SNP es alto.

Futuros avances de esta aproximación podrían estar dirigidos al análisis con otras distribuciones que expliquen las variables aleatorias N y X , a realizar aplicaciones en los temas relacionados con actuaría y el estudio de riesgos, entre otros temas.

Apéndice A

Documentación y Código de las Funciones en R

Por simplicidad sólo se exponen las funciones de la distribución SNP con cuatro parámetros.

`fd.snp4`

Función de densidad de la SNP

Descripción

Calcula la densidad de la distribución SNP.

Uso

$fd_snp(x, d)$

Argumentos

x Vector de cuantiles.

d Vector de parámetros de la distribución SNP. Para este caso es de longitud 4.

Resultado

Vector con valores de la densidad de la distribución $SNP(x; d)$.

Código

```
fdp.snp4 = function(p,zt){
```

```

d1 = p[1];
d2 = p[2];
d3 = p[3];
d4 = p[4];
h1 = zt;
h2 = (zt^2-1)/sqrt(2);
h3 = zt*(zt^2-3)/sqrt(6);
h4 = (zt^4-6*zt^2+3)/(2*sqrt(6));
res = ((1+d1*h1+d2*h2+d3*h3+d4*h4)^2)*exp(-zt^2/2)/sqrt(2*pi)/(1+d1^2+d2^2+d3^3+d4^2)
return(res)
}

```

fda_snp

Función de Distribución Acumulada de la $SNP(d)$

Descripción

Calcula la función de distribución acumulada de la distribución $SNP(d)$ con 4 parámetros.

Uso

$fda_snp(d, x)$

Argumentos

d Vector de parámetros de la distribución SNP.

x Vector de cuantiles.

Resultado

Un vector con las probabilidades acumuladas. Hay una forma sencilla de calcular la función de densidad acumulada que es por medio de la integración; existe otra manera precisa de calcular los valores de la FDA, que es la ecuación exacta de la función de distribución acumulada, que no se presenta por lo extensa.

rand_pc_ig

Función generadora muestras aleatorias de la Poisson Compuesta cuando X se distribuye Inversa Gaussiana

Descripción

Esta función genera muestras aleatorias de la distribución Poisson Compuesta cuando la distribución de X es Inversa Gaussiana con parámetros (μ, σ) .

Uso

`rpc_ig(n, lambda, mu, sigma)`

Argumentos

`n` Tamaño de la muestra que se desea calcular.

`lambda` Tasa de ocurrencia de los eventos, parámetro de la distribución Poisson.

`mu` Parámetro de la distribución Inversa Gaussiana positivo.

`sigma` Parámetro de precisión de la Inversa Gaussiana.

Resultado

Un vector que contienen la muestra aleatoria de la distribución $PC(\lambda; IG(\mu, \sigma))$.

Código

```
rand_pc_ig = function(n,lambda,mu,sigma){
rt = double(n)
  for( j in 1:n){
    k = rpois(1,lambda);
    if( k > 0 ) rt[j] = sum(rinvgauss(k,mu,sigma))
    else rt[j] = 0;
  }
return(rt)
}
```

rand_pc_ln

Función generadora muestras aleatorias de la Poisson Compuesta cuando X se distribuye Log-Normal

Descripción

Esta función genera muestras aleatorias de la distribución Poisson Compuesta cuando la distribución de X es Log-Normal con parámetros (μ, σ) .

Uso

rpc_ln(n, lambda, mu, sigma)

Argumentos

n Tamaño de la muestra que se desea calcular.

lambda Tasa de ocurrencia de los eventos, parámetro de la distribución Poisson.

mu La media de la distribución Log-Normal.

sigma La desviación estándar de la distribución Log-Normal.

Resultado

Un vector que contienen la muestra aleatoria de la distribución $PC(\lambda; \text{Log}N(\mu, \sigma))$.

Código

```
rand_pc_ln = function(n,lambda,mu,sigma){
rt = double(n)
  for( j in 1:n){
    k = rpois(1,lambda);
    if( k > 0 ) rt[j] = sum(rlnorm(k,mu,sigma))
    else rt[j] = 0;
  }
return(rt)
}
```

sis_mom

Estimación por el Método de los Momentos de los Parámetros de la Distribución SNP

Descripción

Esta función estima, por medio del método de igualdad de momentos, los parámetros de la distribución SNP. Esta función está definida para estimar los valores de los 4 parámetros de la distribución SNP. La función es definida para ser utilizada por la función de optimización *optim*.

Uso

sis_mom

Para estimación:

`optim(par = p0, fn = sis_mom, method = "BFGS")`.

Argumentos

`d` Valores de los parámetros de la SNP.

`p0` Valores iniciales para estimar los valores de los parámetros.

`l` Parámetro λ de la distribución Poisson.

`m` Parámetro μ de la distribución Log-Normal.

`s` Parámetro σ de la distribución Log-Normal.

Resultado

Valores de los cuatro parámetros de la distribución SNP.

Código

```
sis_mom <- function(d,l,m,s){
dd<-(1+d[1]^2+d[2]^2+d[3]^2+d[4]^2)
d1 <- d[1]; d2 <- d[2]; d3 <- d[3]; d4 <- d[4]
v_res <- rep(NA, 4)
v_res[1] <- -2*d1+2*sqrt(2)*d1*d2+2*sqrt(3)*d2*d3+4*d3*d4-0*dd
v_res[2] <- -1+2*sqrt(2)*d2+3*d1^2+2*sqrt(6)*d1*d3+5*d2^2+4*sqrt(3)*d2*d4+7*d3^2+9*d4^2-1*dd
v_res[3] <- -6*d1+12*sqrt(2)*d1*d2+18*sqrt(3)*d2*d3+48*d3*d4+2*sqrt(6)*d3+4*sqrt(6)*d1*d4-
(1*(exp(3*m+9*s^2/2))/(1^(3/2)*(exp(2*m+2*s^2))^(3/2)))*dd
v_res[4] <- -3+12*sqrt(2)*d2+15*d1^2+20*sqrt(6)*d1*d3+39*d2^2+56*sqrt(3)*d2*d4+75*d3^2+123*d4^2+
4*sqrt(6)*d4-((1*(exp(4*m+8*s^2))+3*1^2*(exp(2*m+2*s^2))^2)/(1^2*(exp(2*m+2*s^2))^2))*dd
sum(v_res^2)
}
optim(d_0, sis_mom, l=lambda, m=a, s=b, method="BFGS", hessian=T, control=list(maxit=20000))
```

El sistema de momentos se construye de forma similar para cuando $X \sim IG$.

snplogpdf_4

Estimación por el Método de Máxima Verosimilitud de los Parámetros de la Distribución SNP

Descripción

Esta función estima por el método de máxima verosimilitud los parámetros de la distribución SNP.

La estimación se hace por medio de un proceso de optimización y empleando la función `optim`.

Uso

`snplogpdf4`

Para estimación:

`optim(par = p0, fn = snplogpdf4, method = "BFGS", x = z)`

Argumentos

`d` Valores de los parámetros de la SNP.

`p0` Valores iniciales para estimar los valores de los parámetros.

`z` vector de muestra aleatoria de la distribución Poisson Compuesta, los datos deben ser estandarizados.

Resultado

Valores de los cuatro parámetros de la distribución SNP.

Código

```
snplogpdf_4 = function(p,y){
d1 = p[1]
d2 = p[2]
d3 = p[3]
d4 = p[4]

m = length(y)

h1 = y
h2 = (y^2-1)/sqrt(2)
h3 = y*(y^2-3)/sqrt(6)
h4 = (y^4-6*y^2+3)/(2*sqrt(6))

loglik = sum(log((1+d1*h1+d2*h2+d3*h3+d4*h4)^2))-sum(y^2/2)-m*log(1+d1*d1+d2*d2+d3*d3+d4*d4)
return(-loglik)
}
res = optim(d_0,snplogpdf_4,y=y,method="BFGS",hessian=T,control=list(maxit=2000))
```

Esta función puede ser generalizada para cuando se requiere la estimación por máxima verosimilitud de la distribución SNP con 6, 8 y 10 parámetros. En la librería `actuar` del R se encuentran

programadas las aproximaciones Normal, Normal Power y Panjer. A continuación se presentan los programa empleados para calcular estas aproximaciones.

`apr_gamma`

Aproximación Gamma Traslada

Descripción

Esta líneas de programación calculan la aproximación Gamma Traslada a las distribuciones Poisson Compuestas. Para ver más detalles de la propuesta de aproximación Gamma Traslada vea Embrechts [11].

Argumentos

Los argumentos empleados para esta función son los parámetros de la Poisson cuando $X \sim LN(\mu, \sigma)$.

Resultado

Los valores de la aproximación de la Gamma Traslada.

Código

```
(P = LN.Gamma(lambda,mu,sigma))
k = P[1]
alfa = P[2]
beta = P[3]
g1 = P[4]
y.gama = pgamma(x-k,shape=alfa,rate=beta)

LN.Gamma = function(lambda,mu,sigma){

E = double(4)
E[1] = exp(mu+sigma^2/2)
E[2] = exp(2*mu+2*sigma^2)
E[3] = exp(3*mu+9*sigma^2/2)
E[4] = exp(4*mu+8*sigma^2)

# parametros de la gamma trasladada
k = lambda*E[1] - 2*lambda*E[2]^2/E[3]
a = 4*lambda*E[2]^3/E[3]^2
b = 2*E[2]/E[3]

# asimetria de la PC(1,LN(mu,sigma))
g1 = lambda*E[3]/sqrt(lambda*E[2])^3
p = c(k,a,b,g1)
return(p)
```


}

Esta función puede ser implementada para cuando $X \sim IG(\mu, \sigma)$.

apr_gammaInc

Aproximación Gamma Incompleta

Descripción

Calcula la Aproximación Gamma Incompleta

Argumentos

Los parámetros de la distribución Poisson Compuesta, cuando $X \sim LN(\mu, \sigma)$ para calcular el valor de la asimetría denotado por g_1 y obtenido por la función anterior.

Resultado

Los valores de la aproximación de la Gamma Incompleta.

Código

```
alfa.B = 4/g1  
y.gama.B = pgamma(alfa.B + sqrt(alfa.B)*t, shape=alfa.B, scale=1 )
```

apr_NP

Aproximación Normal Power

Descripción

Calcula la Aproximación Normal Power

Argumentos

Los parámetros de la distribución Poisson Compuesta, cuando $X \sim LN(\mu, \sigma)$ y el valor de la asimetría denotado por $gamma$ y obtenido por la función anterior.

Resultado

Los valores de la aproximación NP

Código

```
fda.np = function(x,m,s,gamma){  
z = (x-m)/s
```

```
g = gamma/6
xo = -sqrt(7/4)
eo = ifelse(xo - z,1,0)
y = ifelse(x > 1, sqrt(1+1/(4*g^2)+z/g)-1/(2*g),z-g*(z^-1)+g^2*(4*z^3-7*z)*eo)
F = pnorm(y,0,1)
return(F)
}
```

Bibliografía

- [1] Aalen, O.O. 1992. Modeling Heterogeneity in Survival Analysis by the Compound Poisson Distribution. *The Annals of Applied Probability*, 2, No. 4, pp. 951-972.
- [2] Barbour, A.D. and Chryssaphinou, O. 2001. Compound Poisson Approximation: A User's Guide. *The Annals of Applied Probability*, 11, No. 3, pp. 964-1002.
- [3] Bowers, N.L. 1966. Expansion of probability density functions as a sum of Gamma densities with applications in risk theory. *Transactions of Society of Actuaries*, 18, No. 52, pp. 125-137.
- [4] Bowers, N.L., Hickman, J.C., Nesbitt, C.J., Jones, D.A. & Gerber, H.U. 1997. Actuarial Mathematics. Actuarial Mathematics (2th edition). The Society of Actuaries, Schamburg, Illinois.
- [5] Brunner, A. 1992. Conditional Asymmetries in Real GNP: A Semionparametric Approach. *Journal of Business & Economic Statistics*, 10, No. 1, pp. 65-72 .
- [6] Chaubey, Y.P., Garrido, J. & S. Trudeau. 1998. On the computation of aggregate claims distributions: some new approximations. *Insurance: Mathematics and Economics*, 23, pp. 215-230.
- [7] Chavez-Demoulin, V., Embrechts, P. & Neslehova, J. 2005. Quantitative models for operational risk: extremes, dependence and aggregation. Preprint, ETH Zurich.
- [8] Chen, X. 2006. Large sample sieve estimation of semi-nonparametric models. Editors: Heckman, J.J & Leamer, E.E. *Handbook of Econometrics*, Volumen 6, Chapter 76, North-Holland.
- [9] Consul, P.C., Jain, G.C. 1973. A Generalization of the Poisson Distribution. *Technometrics*, 15, No. 4, pp. 791-799.
- [10] Coppejans, M. & Gallant, R. 2002. Cross-validated SNP density estimates. *Journal of Econometrics*, 110, pp. 27-65.
- [11] Embrechts, P., Kaufmann, R. & Samorodnisky, G. 2004. Ruin theory Revisited: Stochastic Models for Operational Risk. Technical report <http://www.math.ethz.ch/~embrechts/>

- [12] Embrechts, P. & Puccetti, G. 2006. Aggregating risk capital, with an application to operational risk. *The Geneva Papers on Risk and Insurance Theory*, 31, 2, pp. 71-90.
- [13] Escalante, C., Arango, G. 2004. Aspectos básicos del modelo de riesgo colectivo. *Matemáticas: Enseñanza Universitaria*, XII, No. 2, pp. 3-15.
- [14] Escalante, C. 2006. Distribuciones clase (a,b) y algoritmo de Panjer. *Matemáticas: Enseñanza Universitaria*, XIV, No. 2, pp. 3-17.
- [15] Fenton, V. & Gallant, R. 1996a. Convergence Rates of SNP Density Estimators. *Econometrica*, 64, No. 3, pp. 719-727.
- [16] Fenton, V. & Gallant, R. 1996b. Qualitative and asymptotic performance of SNP density estimators. *Journal of Econometrics*, 74, pp. 77-118.
- [17] Folks, J.L. & Chhikara, R.S. 1978. The inverse Gaussian distribution and its statistical application - a review. *Journal Royal Statistical Society*, series B, No. 40, pp. 261-258.
- [18] Gabler, S., Laisney, F. & Lechner, M. 1993. Semiparametric estimation of binary choice models with an application to labor-force participation. *Journal of Business & Economic Statistics*, 11, No. 1, pp. 61-80.
- [19] Gallant, R.A. & Nychka, D.W. 1987. Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica*, 55 (2), pp. 363-390.
- [20] Gallant, R.A., Rossi, P.E. & Tauchen, G. 1993. Nonlinear dynamic structures. *Econometrica*, 61, pp. 871-907.
- [21] Gallant, R.A. & Tauchen, G. 1989. Semiparametric estimation of conditionally constrained heterogeneous process: asset pricing application. *Econometrica*, 57, No. 5, pp. 1091-1120.
- [22] Gendron, M. & Crepeau, H. 1989. On the computation of the aggregate claim distribution when individual claims are Inverse Gaussian. *Insurance: Mathematics and Economics*, 8, pp. 251-258.
- [23] Hogg, R. & Klugman, S. 1984. *Loss Distribution*. John Wiley, New York.
- [24] Hurlimann, W. 2003. A Gaussian Exponential Approximation to some Compound Poisson Distributions. *Astin Bulletin*, 33, No. 1, pp. 41-55.
- [25] Khuri, A.I. 2003. *Advanced Calculus with Applications in Statistics*. Wiley-Interscience. New Jersey.
- [26] Klaauw, B., Koning, R.H. 1998. Some applications of Semi-Nonparametric Maximum Likelihood Estimation. *Tinbergen Institute Discussion Papers from Tinbergen Institute*, No. 96-161/7.

- [27] Klugman, S.A., Panjer, H.H. & Willmot, G.E. 2004. Loss Models From Data to Decisions (2th edition). John Wiley & Sons, Hoboken, New Jersey.
- [28] León, A., Mencía, J. & Sentana, E. 2005. Parametric properties of Semi-Nonparametric Distribution, with Applications to Option Valuation, CEMFI Working Paper No. 0509. www.cemfi.es
- [29] McNeil, A.J., Freyand, R. & Embrechts, P. (2005). Quantitative Risk Management: Concepts, Techniques and Tools. Princeton Univeristy Press, Princeto and Oxford.
- [30] Moscadelli, M. 2004. The Modelling of Operaciotional Risk: the Experience with the Analysis of the Data Collected by the Basel Committee. Banca d'Italia, Temi di Discussione del Servizio Studi, Number 517.
- [31] Panjer, H.H. & Willmot, G.E. 1992. Insurance Risk Models. The Society of Actuaries, Schaumburg, Illinois.
- [32] Panjer, H.H. 2006. Operational Risk, modeling analytics. Wiley interscience, New Jersey.
- [33] Pentikainen, T. 1977. On the approximation of the total amount of claims. *Astin Bulletin*, No. 9, pp. 281-289.
- [34] Pentikainen, T. 1980. The Theory of the Risk and Some Applications. *The Journal of Risk and Insurance*, 47, No. 1, pp. 16-43.
- [35] Pentikainen, T. 1987. Aproximate evaluation of the distribution function of aggregate claims. *Astin Bulletin*, 17 (1), pp. 15-40.
- [36] Phillips, P. 1983. ERA's: A new approach to small sample theory. *Econometrica*, 51, 1505-1525.
- [37] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [38] Ramsay, C.M. (1991). A Note of the Normal Power Approximation. Short Contributions. *Astin Bulletin*, 21, No. 1 pp. 147-150.
- [39] Rolski, T. , Schmidli, H., Schmidt, V. and Teugels, J. (1999) *Stochastic Processes for Insurance and Finance*, John Wiley and Sons, Ltd, Chichester.
- [40] Seal, H.L. 1977. Approximations to risk theory's $F(x, t)$ by means of the gamma distribution. *Astin Bulletin*, No. 9, pp. 213-218.
- [41] Seal, H.L. 1978. From aggregate claims distribution to probability of ruin. *Astin Bulletin*, 10, pp. 47-53.

- [42] Seshadri, V. 1983. The Inverse Gaussian Distribution: Some Properties and Characterizations. *The Canadian Journal of Statistics - La Revue Canadienne de Statistique*, 11, No. 2., pp. 131-136.
- [43] Sundt, B. 2002. Review: Recursive Evaluation of Aggregate Claims Distributions. *Insurance: Mathematics and Economics*, 30, pp. 297-322.
- [44] Willmot, G. & Lin, X. 1994. Lundberg bounds on the tails of compound distributions. *Journal of Applied Probability*, 31, 743-756.