Estimación de parámetros para la distribución Reflected Weibull mediante simulación Monte Carlo usando los paquetes RelDists y gamlss *

Amylkar Urrea Montoya *Universidad de Antioquia*
Carmen Elena Patiño Rodriguez *Universidad de Antioquia*
Freddy Hernández Barajas *Universidad Nacional de Colombia - sede Medellín*
Olga Usuga Manco *Universidad de Antioquia*

Resumen: La distribución Reflected Weibull (`RW`) es empleada en estudios de confiabilidad y de análisis de vida útil, cuya característica principal es que la distribución de los datos es generalmente sesgada. `RW` ofrece una gran flexibilidad para describir la distribución de los datos dado que tiene la capacidad de reflejar su sesgo en el eje vertical. En este trabajo se realiza un estudio de simulación con y sin covariables para la distribución `RW` en el cual se estiman, evaluan y comparan los ajustes de la distribución para distintos tamaños de muestra. El estudio de simulación se realiza mediante el método Monte Carlo haciendo uso de la función generadora de números aleatorios `rRW`, de la familia de distribuciones `RW` contenidas en el paquete **RelDists** y del paquete **gamlss** para realizar el ajuste de los datos. Con este trabajo se logró determinar que los parámetros estimados de `RW` se van pareciendo más a los valores reales a medida que el tamaño de muestra crece, es decir que la estimación mejora considerablemente.

Keywords: Reflected Weibull, Simulaciones Monte Carlo, RelDists, gamlss.

Introducción

Considerando la transformación $Y = -X$, donde X es una variable Weibull aleatoria, Cohen (1973) introdujo la distribución Reflected Weibull (`RW`) con 3 parámetros, posteriormente, Almalki and Nadarajah (2014) reparametrizaron esta distribución con 2 parámetros, uno de localización μ y uno de escala σ. Esta distribución es usada comúnmente en estudios de confiabilidad. La función de distribución de probabilidad (`dRW`), la función de distribución acumulada (`pRW`) y la función de riesgo (`hRW`) estan dadas por las siguientes ecuaciónes, donde y<0 y μ, σ>0:

$$f(y) = \mu\sigma(-y)^{\sigma-1}e^{-\mu(-y)^{\sigma}} \tag{1}$$

$$F(y) = e^{-\mu(-y)^{\sigma}} \tag{2}$$

$$h(y) = \mu\sigma(-y)^{\sigma-1} \tag{3}$$

Estas funciones, junto con la función generadora de números aleatorios `rRW`, la función quantil `qRW` y la familia de distribuciones `RW`, hacen parte del paquete **RelDists**, el cual se ejecuta bajo el entorno **gamlss** (Stasinopoulos et al., 2017) y esta siendo implementado en R Core Team (2019) por los autores. El paqute **RelDists** contiene distribuciones que habian sido desarrolladas pero que no estaban implementadas computacionalmente para su uso.

* **Autor de contacto**: amylkar.urrea@udea.edu.co

Este trabajo se realizó con el objetivo de analizar el comportamiento asintótico de los estimadores de los parámetros de la distribución `RW` mediante un estudio de simulación con y sin variables, en el cual se evalúa y analiza el ajuste de sus parámetros a medida que el tamaño de muestra aumenta.

En la Figura 1 se pueden observar algunas formas que pueden tomar `dRW` y `hRW` para distintos valores de sus parámetros.

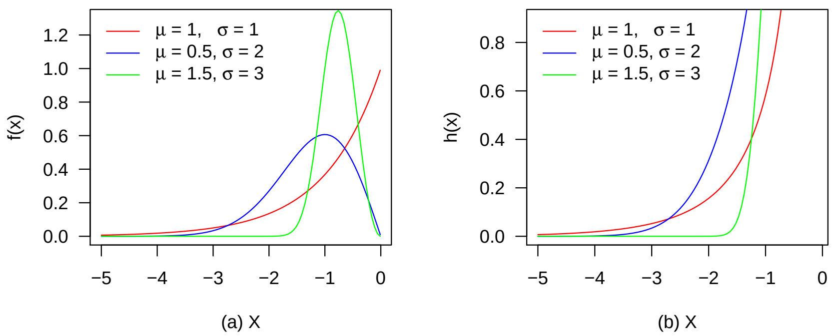

Figure 1: Función de distribución de probabilidad (a) y función de riesgo (b) de la distribución RW para distintos valores de sus parámetros.

El documento está organizado como se describe a continuación. Primero se presenta la metodología usada para hacer el estudio de simulación, luego se muestran y analizan los resultados obtenidos y finalmente se dan las conclusiones a las que se llegó con el estudio.

Estudio de simulación

En esta sección se presentan la metodología del estudio de simulación Monte Carlo realizado para explorar la estimación de parámetros sin y con covariables para la distribución `RW`.

En el estudio sin covariables, el desempeño en la estimación de parámetros para la distribución `RW` es analizado con respecto a la media de los parámetros estimados para distintos tamaños de muestra $n = 20, 40, 60, \ldots, 270, 300$. Para realizar este estudio se siguieron los siguientes pasos:

1. Se generaron muestras de distintos tamaños n mediante la función generadora de números aleatorias `rRW` usando los parámetros $\mu = 1$ y $\sigma = 1$.

2. Se obtuvieron y almacenaron los estimadores $\hat{\mu}$ y $\hat{\sigma}$.

3. Se repitieron 1000 veces los pasos 1 y 2.

El código usado para estimar los parámetros fue el siguiente:

```
y <- rRW(n, mu=1, sigma=1)
mod <- gamlss(y~1, sigma.fo=~1, family=RW,
              control=gamlss.control(n.cyc=5000, trace=FALSE))
```

En el estudio con covariables se usó un modelo de regresión con 2 variables `x1` y `x2`, del mismo modo que en el estudio sin covariables, en este estudio se evalúo el desempeño de la estimación a través del MSE, pero con tamaños de muestra $n = 30, 50, 70, \ldots, 470$. Para realizar este estudio se siguieron los siguientes pasos:

1. Se generaron n observaciones de $x_1 \sim U(0.4, 0.6)$ y $x_2 \sim U(0.4, 0.6)$ para ser usadas como covariables.

2. Se generaron n observaciones $y \sim RW(\mu = \exp(1.5 - 1.5x_1), \sigma = \exp(2 - 2x_2))$.

3. Se obtuvieron y almacenaron los coeficientes estimados β_0 y β_1 asociados a μ y los coeficientes estimados γ_0 y γ_1 para σ.
4. Se repitieron 1000 veces los pasos 1, 2 y 3.

El código usado para realizar el modelo de regresión fue el siguiente:

```
x1 <- runif(n, min=0.4, max=0.6)
x2 <- runif(n, min=0.4, max=0.6)
mu <- exp(1.5 - 1.5 * x1)
sigma <- exp(2 - 2 * x2)
x <- rRW(n, mu, sigma)
mod <- gamlss(x~x1, sigma.fo=~x2, family=RW,
              control=gamlss.control(n.cyc=5000, trace=FALSE))
```

Resultados

En esta sección se muestran los resultados a los que se llegó con el estudio de simulación sin y con covariables.

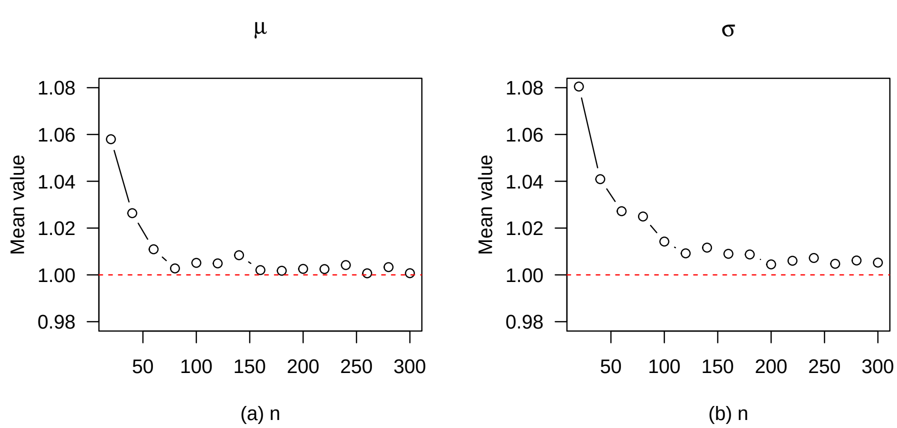

Figure 2: Media para los parámetros estimados $\hat{\mu}$ (a) y $\hat{\sigma}$ (b). Las lineas rojas corresponden a los verdaderos valores μ y σ.

La Figura 2 describe la evolución de la media de los parámetros estimados $\hat{\mu}$ y $\hat{\sigma}$ para distintos tamaño de muestra n. Se puede observar que a medida que este tamaño aumenta, el valor promedio de $\hat{\mu}$ y $\hat{\sigma}$ se van pareciendo más a los verdaderos parámetros $\mu = 1$ y $\sigma = 1$. Es decir, que, entre más información contenga la muestra, `RW` es más consistente en la estimación de los valores de μ y σ reales que sigue la distribución, por tanto, la inferencia que se haga apartir de sus diferentes funciones se realiza sobre el verdadero comportamiento de los datos.

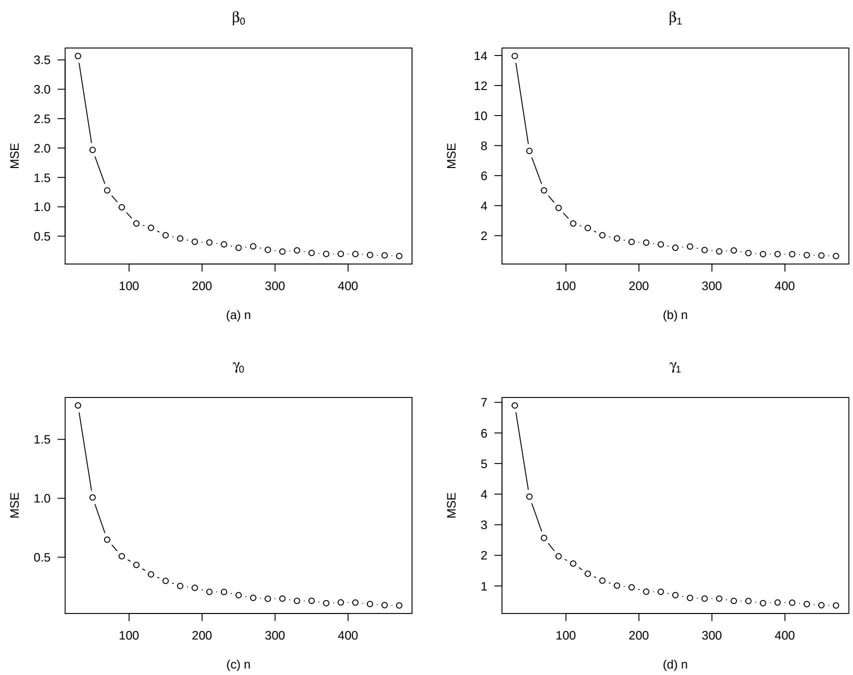

Figure 3: MSE para los parámetros estimados $\hat{\mu}$ (a y b) y $\hat{\sigma}$ (c y d)

En la Figura 3 se observa la tendencia que tiene el `MSE` para distintos valores de `n` en los parámetros estimados del modelo de regresión β_0, β_1, γ_0 y γ_1. En las 4 figuras es evidente como a medida que el tamaño de muestra aumenta, el valor correspondiente al `MSE` disminuye acercandose a cero, es decir que el error entre el valor real de los parámetros y los estimados es cada vez menor. Este comportamiento indica que el modelo de regresión ofrece la posibilidad de obtener predicciones más precisas a medida que se cuenta con más información.

Conclusiones

La simulación Monte Carlo demostró ser una técnica realmente útil para evaluar la estimación de los parámetros de la distribución `RW` dentro del entorno gamlss. Los resultados obtenidos en las Figuras 2 y 3 permitieron establecer que los parámetros estimados de `RW` se van pareciendo más a los valores reales a medida que el tamaño de muestra crece, es decir que la estimación mejora considerablemente. La inferencia que se haga a partir de distribuciónes de datos y de modelos de regresión para `RW`, resulta ser efectiva ya que esta distribución es capaz de estimar parámetros que reflejan el verdadero comportamiento de los datos.

References

Almalki, Saad J and Saralees Nadarajah. 2014. "Modifications of the Weibull distribution: A review." *Reliability Engineering & System Safety* 124:32–55.

Cohen, C. 1973. "The Reflected Weibull Distribution." *Technometrics* 15(4):867–873.
URL: *https://www.jstor.org/stable/1267396*

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
URL: *https://www.R-project.org/*

Stasinopoulos, Mikis D, Robert A. Rigby, Gillian Z. Heller, Vlasios Voudouris and Fernanda Bastiani. 2017. *Flexible Regression and Smoothing Using GAMLSS in R.* New York, NY: CRC Press.