

Modelo unimodal simétrico-asimétrico con aplicación al estudio del RNA VIH-1

GUILLERMO MARTÍNEZ-FLÓREZ^a, GERMÁN MORENO-ARENAS^{b,*}

^a Universidad de Córdoba, Departamento de Matemáticas y Estadística, Montería, Colombia.

^b Universidad Industrial de Santander, Escuela de Matemáticas, Bucaramanga, Colombia.

Resumen. Se definen dos nuevas distribuciones de probabilidad: modelo unimodal simétrico con función de riesgo proporcional a la distribución normal y modelo unimodal asimétrico con función de riesgo proporcional a la distribución normal asimétrica. Estos modelos permiten ajustar datos censurados con comportamiento bimodal y altos (o bajos) niveles de curtosis comparado con la curtosis de la distribución normal y altos (o bajos) niveles de asimetría. Además, se estiman los parámetros de los modelos por máxima verosimilitud y se determina la matriz de información observada. La flexibilidad de la nueva distribución se ilustra ajustando un conjunto de datos reales: el número de moléculas de ARN VIH-1 por mililitros de sangre medida en personas con pruebas confirmadas de presencia del VIH.

Palabras claves: Bimodalidad, asimetría, curtosis, función de riesgo proporcional, censura, límite de detección, ARN VIH-1, HAART.

MSC2010: 62F10, 62F12, 62N86.

Uni-bimodal Symmetric-Asymmetric Model with Application to the Study of HIV-1 RNA

Abstract. We define two new probability distributions, unimodal symmetric model with proportional hazard function to the normal distribution and unimodal asymmetric model with proportional hazard function to the skew-normal distribution. These models allow adjust censored data with bimodal behavior and high (or low) levels of kurtosis compared with kurtosis of the normal distribution and high (or low) levels of asymmetry. The model parameters are estimated by maximum likelihood and the observed information matrix is determined. The flexibility of the new distribution is illustrated by adjusting a set of real data, the number of molecules of HIV-1 RNA per milliliter of blood measured in individuals with confirmed test of the presence of HIV.

Keywords: Bimodality, skewness, kurtosis, proportional hazard function, censorship, limit of detection, HIV-1 RNA, HAART.

* Autor para correspondencia: E-mail: gmorenoa@uis.edu.co

Recibido: 29 de enero de 2013. Aceptado: 03 de febrero de 2014.

Para citar este artículo: G. Martínez-Flórez, G. Moreno-Arenas, Modelo unimodal simétrico-asimétrico con aplicación al estudio del RNA VIH-1, *Rev. Integr. Temas Mat.* 32 (2014), no. 1, 1–18.

1. Introducción

Los biomarcadores o marcadores biológicos son características que se miden y se evalúan como un indicador de procesos biológicos normales, procesos patogénicos o respuestas farmacológicas a una intervención terapéutica objetiva [*Biomarkers Definitions Working Group, 2011*]. Los biomarcadores son útiles para indicar el estadio en que se encuentra el desarrollo de una enfermedad, evaluar los resultados del uso de los medicamentos o para detectar el fracaso terapéutico antes de que se desarrollen complicaciones clínicas. En estudios con biomarcadores es común tener límites de detección por debajo de los cuales no es posible determinar los valores específicos de la variable de interés. Por ejemplo, en pacientes que tienen confirmada la presencia del virus de la inmunodeficiencia humana (VIH), la carga viral es medida por la cantidad de moléculas de ARN VIH-1 presentes en la sangre; cuando se usa la prueba *Amplicor*[®] sólo se detectan valores iguales o superiores a 50 unidades por mililitro de sangre y con la prueba *COBAS*[®] *TaqMan*[®] se detectan valores iguales o superiores a 20 copias/mL, siendo esta más sensible.

A los pacientes con VIH se los trata clínicamente con un *coctel* de medicamentos, conocido técnicamente como Terapia Antirretroviral Altamente Activa (HAART, por sus siglas en inglés). El uso de estos medicamentos restringe el desarrollo del síndrome de inmunodeficiencia adquirida (SIDA) y reduce el número de moléculas de ARN VIH-1 hasta niveles indetectables con las pruebas ultrasensibles actuales. Sin embargo, niveles indetectables de carga viral no significan que la replicación viral se ha detenido, y por esto los laboratorios hacen esfuerzos para elaborar pruebas más sensibles, con el propósito de determinar con mayor precisión el número de moléculas de ARN VIH-1 presentes en la sangre. En pacientes con buena respuesta a la HAART, el número de moléculas de ARN VIH-1 se puede reducir a valores por debajo de 20 copias/mL, límite actual de las pruebas ultrasensibles, pero no se puede determinar el número exacto de moléculas cuando está por debajo de este límite de detección, y generalmente se dice que el paciente tiene 20 copias/mL de ARN VIH-1. Este tipo de datos en la literatura estadística se conoce como datos censurados por la izquierda; sólo se sabe que el paciente superó dicho umbral, pero no se sabe cuál es la verdadera medida.

Además de la presencia de datos censurados, en la población de pacientes con VIH se ha observado que la distribución del número de moléculas de ARN VIH-1 tiene un comportamiento bimodal. Por ejemplo, Li et al. [15], en un estudio con pacientes diagnosticados con VIH y bajo la HAART, mostraron que en esta población la cantidad de moléculas de ARN VIH-1 tiene un comportamiento bimodal; Schneider et al. [21] corroboraron el comportamiento bimodal en la distribución del número de moléculas de ARN VIH-1 en otra población de este tipo de pacientes. De tal forma que el pico de la izquierda de la distribución representa los individuos con la más alta reducción en la cantidad de moléculas de ARN VIH-1 usando la HAART, y el pico de la derecha refleja los individuos con respuesta subóptima a los medicamentos de la HAART.

La bimodalidad y la censura están presentes en otras variables biológicas; por ejemplo, Lim et al. [16] mostraron que existe bimodalidad en la distribución del nivel de glucosa medido en milimoles por litro (mmol/L) de sangre en poblaciones con baja prevalencia de diabetes y de obesidad; las pruebas del nivel de glucosa tienen límite de detección mínimo. La bimodalidad también se encuentra en variables climatológicas; Zhang et al. [23] indicaron que la distribución de probabilidad de la cantidad de vapor de agua tropos-

férico medido en los trópicos (X hPa) tiene comportamiento bimodal, y esta bimodalidad implica gradientes agudas entre los regímenes secos y húmedos en el espacio y el tiempo.

En los tres casos expuestos: número de moléculas de ARN VIH-1, nivel de glucosa en la sangre y cantidad de vapor de agua troposférico, los datos fueron modelados con una mixtura de dos distribuciones normales. A pesar de que la mixtura de distribuciones normales es frecuentemente usada para modelar comportamientos bimodales, esta tiene problemas en presencia de distribuciones con altos grados de asimetría y curtosis en las dos subpoblaciones que componen la distribución completa de los datos. En algunos casos, en la mixtura pueden coexistir distribuciones que tienen alto (o bajo) grado de asimetría en cada componente. En consecuencia, el análisis de datos es orientado hacia el uso de distribuciones asimétricas, donde al menos una de las dos componentes de la mixtura o la distribución completa debería ser modelada por una distribución asimétrica.

Algunas distribuciones como la exponencial, gamma, Gumbel, log-normal y Weibull, entre otras, han sido tradicionalmente usadas para modelar datos con comportamientos asimétricos; alternativas recientes como la distribución normal asimétrica introducida por Azzalini [3] o distribución *alfa-power* definida por Durrans [8] también son utilizadas para modelar datos con comportamientos asimétricos; pero todas estas distribuciones son de tipo unimodal.

La función de densidad de probabilidad de una variable aleatoria Z con distribución normal asimétrica está dada por

$$f(z; \lambda) = 2\phi(z)\Phi(\lambda z), \quad \lambda \in \mathbb{R}, \quad (1)$$

donde ϕ y Φ denotan la función de densidad y función de distribución de la distribución normal estándar, respectivamente. Denotamos esta distribución por $Z \sim \text{SN}(\lambda)$. La asimetría es regulada por el parámetro λ ; cuando $\lambda = 0$ se obtiene la distribución normal. La distribución normal asimétrica fue introducida inicialmente por O'Hagan y Leonard [17] como una distribución apriori para estimar el parámetro de localización de la distribución normal. Posteriormente Azzalini [3] investiga sus propiedades fundamentales y propone algunas generalizaciones y extensiones. La densidad (1) ha sido también ampliamente estudiada por [5], [9], [10], [13] y [18].

Durrans [8], en un contexto hidrológico, introdujo la distribución de las estadísticas de orden con función de densidad dada por

$$\varphi_F(z; \alpha) = \alpha f(z) \{F(z)\}^{\alpha-1}, \quad \alpha \in \mathbb{R}^+, \quad (2)$$

donde F es una función de distribución absolutamente continua con función de densidad f y α es el parámetro de forma que controla la asimetría de la distribución. El modelo (2) también es conocido en la literatura como distribución *alfa-power*.

Cuando en (2), $f = \phi$ y $F = \Phi$ tenemos la distribución llamada *power-normal*; además si $\alpha = 1$ obtenemos la distribución normal. Usamos la notación $Z \sim \text{PN}(\alpha)$. En la Figura 1 se puede observar que el parámetro α controla la curtosis de la distribución; cuando $\alpha > 1$ la curtosis es mayor que la curtosis de la distribución normal, y para $0 < \alpha < 1$ ocurre lo contrario.

Nótese que las distribuciones normal asimétrica y *power-normal* ofrecen una alternativa en el caso de datos que presentan mayor (o menor) asimetría o mayor (o menor) curtosis

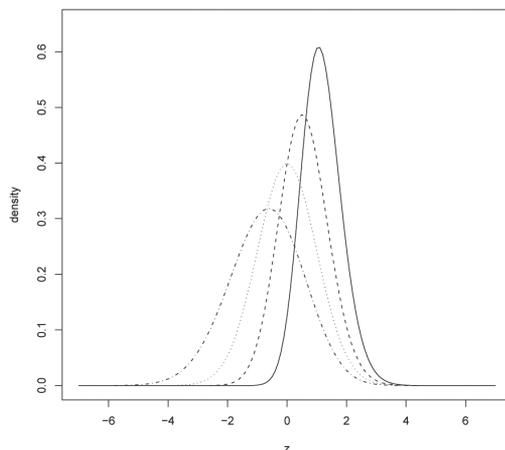


Figura 1. Función de densidad $PN(\alpha)$ para $\alpha = 5$ (línea sólida), $\alpha = 2$ (línea de trazos), $\alpha = 1$ (línea punteada) y $\alpha = 0,5$ (línea de trazos y puntos).

que la distribución normal, pero tienen la característica de ser unimodales. Para modelar comportamientos bimodales tenemos el modelo definido por Kim [14], que es una extensión bimodal del modelo normal asimétrico, llamado normal asimétrico de dos piezas, cuya función de densidad de probabilidad está dada por

$$f(z; \alpha) = k_\alpha \phi(z) \Phi(\alpha|z|), \quad \alpha \in \mathbb{R}, \quad (3)$$

donde $k_\alpha = 2\pi / [\pi + 2 \arctan(\alpha)]$ es una constante de normalización. Denotamos el modelo (3) por $Z \sim \text{TPSN}(\alpha)$. Para $\alpha > 0$, el modelo (3) es bimodal. Una extensión del modelo bimodal de Kim fue presentado por Arnold et al. [2], la función de densidad de probabilidad llamada normal asimétrica extendida de dos piezas está dada por

$$f(z; \alpha, \omega) = 2k_\alpha \phi(z) \Phi(\alpha|z|) \Phi(\omega z), \quad \alpha \in \mathbb{R}, \quad \omega \in \mathbb{R}, \quad (4)$$

donde ω es un parámetro que controla la asimetría de la distribución y k_α es una constante de normalización. El modelo (4) lo denotamos por $Z \sim \text{ETPSN}(\alpha, \omega)$.

Cordeiro y de Castro [7] propusieron una extensión del modelo de Durrans, llamada distribución *Kumaraswamy-G*, la cual está definida mediante la función de densidad de probabilidad

$$\varphi_G(z; a, b) = ab g(z) G^{a-1}(z) \{1 - G^a(z)\}^{b-1}, \quad a \in \mathbb{R}^+, b \in \mathbb{R}^+, \quad (5)$$

donde $G(z)$ es una función de distribución continua arbitraria, $g(z) = dG(z)/dz$ es su correspondiente función de densidad de probabilidad y, a y b son parámetros adicionales de forma. Si Z es una variable aleatoria con densidad (5), escribimos $Z \sim \text{Kw} - G(a, b)$. Cuando $a = 1$ tenemos el modelo

$$\varphi_G(z; b) = b g(z) \{1 - G(z)\}^{b-1}, \quad z \in \mathbb{R}, \quad (6)$$

con función de distribución

$$F_G(z; b) = 1 - \{1 - G(z)\}^b, \quad z \in \mathbb{R},$$

y función de riesgo $h_G(z; b) = b g(z) / \{1 - G(z)\}$. Si Z sigue el modelo (6) lo denotamos por $Z \sim \text{PHF}(b)$ (de *proportional hazard function* en inglés).

Teniendo en cuenta las ideas de [14], [2] y [7], en este trabajo definimos nuevas distribuciones de probabilidad capaces de modelar comportamientos bimodales, sugiriendo que se puede reemplazar la mixtura de dos distribuciones normales por una única distribución bimodal con mayor flexibilidad de ajuste en términos de asimetría y curtosis en cada componente de la distribución y en la distribución completa. Adicionalmente se incluye la metodología de análisis con presencia de datos censurados. Este artículo está organizado de la siguiente forma: en la Sección 2 definimos un nuevo modelo unimodal simétrico con función de riesgo proporcional a la distribución normal, y presentamos la estimación de máxima verosimilitud de los parámetros de esta distribución. En la Sección 3 definimos el modelo unimodal asimétrico como una extensión del modelo *normal asimétrico flexible* estudiado por Gómez et al. [10]. Calculamos los momentos, los estimadores de máxima verosimilitud de los parámetros del modelo y la matriz de información observada. Para fines ilustrativos, en la Sección 4 aplicamos el nuevo modelo para analizar las mediciones de ARN VIH-1 en personas con pruebas confirmadas de presencia del VIH y bajo la HAART, mostrando las diferencias con el ajuste por mixtura de dos distribuciones normales. Por último, en la Sección 5 hacemos algunas observaciones con relación a estos nuevos modelos.

2. Distribución unimodal simétrica

Definimos una variable aleatoria continua Z con distribución unimodal simétrica con parámetro α , si su función de densidad de probabilidad está dada por la expresión

$$\varphi_G(z; \alpha) = \alpha g(z) \{2(1 - G(|z|))\}^{\alpha-1}, \quad \alpha \in \mathbb{R}^+, \quad (7)$$

donde $G(z)$ es una función de distribución absolutamente continua, $g(z)$ es la correspondiente función de densidad simétrica alrededor de cero y función de riesgo proporcional dependiente de la función de distribución $G(z)$. En adelante se denota por $Z \sim \text{UBPHF}(\alpha)$.

2.1. Propiedades de la distribución UBPHF

Algunas propiedades de la distribución $\text{UBPHF}(\alpha)$ son:

- (i) **Simetría:** Si $Z \sim \text{UBPHF}(\alpha)$, entonces el modelo (7) es simétrico.
- (ii) **Función de distribución:** Si $Z \sim \text{UBPHF}(\alpha)$, entonces la función de distribución de Z está dada por

$$F_G(z; \alpha) = \begin{cases} \frac{1}{2} \{2G(z)\}^\alpha, & \text{si } z < 0, \\ 1 - \frac{1}{2} \{2(1 - G(z))\}^\alpha, & \text{si } z \geq 0. \end{cases}$$

- (iii) **Generador de números aleatorios:** Si la variable aleatoria Y tiene distribución uniforme en el intervalo $(0, 1)$, denotado por $Y \sim U(0, 1)$, y $g(z)$ es simétrica, entonces mediante el método de inversión podemos obtener una variable aleatoria con distribución $Z \sim \text{UBPHF}(\alpha)$ mediante la expresión

$$Z = \begin{cases} G^{-1} \left[\frac{1}{2}(2Y)^{1/\alpha} \right], & \text{si } Y < 1/2, \\ G^{-1} \left[1 - \frac{1}{2}(2(1-Y))^{1/\alpha} \right], & \text{si } Y \geq 1/2. \end{cases}$$

- (iv) **Familia de localización y escala:** Si $Z \sim \text{UBPHF}(\alpha)$ y $X = \xi + \eta Z$, donde $\xi \in \mathbb{R}$ es el parámetro de localización y $\eta \in \mathbb{R}^+$ es el parámetro de escala, entonces la función de densidad de probabilidad de la variable aleatoria X está dada por

$$\varphi(x; \xi, \eta, \alpha) = \frac{\alpha}{\eta} g \left(\frac{x - \xi}{\eta} \right) \left\{ 2 \left(1 - G \left(\left| \frac{x - \xi}{\eta} \right| \right) \right) \right\}^{\alpha-1}, \quad \alpha \in \mathbb{R}^+. \quad (8)$$

Si una variable aleatoria X sigue el modelo (8), la denotamos $X \sim \text{UBPHF}(\xi, \eta, \alpha)$. Cuando $g = \phi$ y $G = \Phi$, escribimos $X \sim \text{UBPHN}(\xi, \eta, \alpha)$, esto es, distribución unimodal con función de riesgo proporcional a la distribución normal y parámetros ξ , η y α . La estimación de los parámetros del modelo $\text{UBPHN}(\xi, \eta, \alpha)$ se realiza mediante el método de máxima verosimilitud, es decir, maximizando el logaritmo de la función de verosimilitud respecto al vector de parámetros (ξ, η, α) .

Si $\xi = 0$ y $\eta = 1$, denotamos $Z \sim \text{UBPHN}(\alpha)$. La generación de una variable aleatoria $\text{UBPHN}(\alpha)$ se puede hacer mediante el siguiente procedimiento:

1. Sea $Y \sim U(0, 1)$;
2. Si $Y < \frac{1}{2}$, entonces $Z = \Phi^{-1} \left(\frac{1}{2}(2Y)^{1/\alpha} \right)$;
3. En caso contrario, $Z = \Phi^{-1} \left(1 - \frac{1}{2}(2(1-Y))^{1/\alpha} \right)$, donde Φ^{-1} es la inversa de la distribución normal acumulada.

- (v) **Unimodalidad:** Si $Z \sim \text{UBPHN}(\alpha)$, entonces su función de densidad es unimodal para $\alpha \geq 1$ y bimodal para $\alpha < 1$.

Prueba: Derivando $f(z) = \alpha \phi(z) \{2(1 - \Phi(|z|))\}^{\alpha-1}$ con respecto a la variable z , e igualando a cero, tenemos que los puntos donde ocurren los máximos y mínimos son las soluciones de las ecuaciones

$$\begin{cases} (\alpha - 2) \log[1 - \phi(|z|)] + \log(\phi(|z|)) = 0, & \text{si } \alpha \geq 1, \\ (1 - \alpha)\phi(|z|) - |z|(1 - \Phi(|z|)) = 0, & \text{si } \alpha < 1; \end{cases} \quad (9)$$

así, $f(z)$ es unimodal para $\alpha \geq 1$ y bimodal para $\alpha < 1$. Cuando $\alpha = 1$, corresponde a la distribución normal.

2.2. Distribución unimodal simétrica para datos censurados

Cuando los datos son censurados, la distribución resultante es una mezcla entre una distribución discreta y una continua. Esta idea fue definida por Tobin en [22] y el modelo

resultante se conoce como Modelo *Tobit*. Sean $X^* \sim \text{UBPHF}(\alpha)$ y $\{X_1^*, X_2^*, \dots, X_n^*\}$ una muestra aleatoria de tamaño n y punto de censura igual a c (o límite de detección inferior cuando hablamos de las pruebas ultrasensibles con biomarcadores). La variable aleatoria X tomará el valor observado si x^* es mayor que c , y c si x^* es menor o igual que c ; esto es,

$$x_i = \begin{cases} x_i^*, & \text{si } x_i^* > c, \\ c, & \text{si } x_i^* \leq c, \end{cases}$$

para $i = 1, 2, \dots, n$; consecuentemente, $\text{Prob}(x_i = c) = \text{Prob}(x_i^* \leq c)$, y para $x_i^* > c$ la distribución de probabilidad de la variable aleatoria X es igual a la distribución de probabilidad de la variable aleatoria X^* . Entonces X es una variable aleatoria con distribución UBPHF con censura a la izquierda, y la denotamos por $X \sim \text{CUBPHF}(\alpha)$. La generalización con censura a la derecha es trivial.

Tomando $g = \phi$ y $G = \Phi$ llegamos de forma natural a una variable aleatoria censurada con distribución unimodal con función de riesgo proporcional a la distribución normal, la cual denotamos por $X \sim \text{CUBPHN}(\alpha)$. La extensión del modelo $\text{CUBPHN}(\alpha)$ a modelo de localización y escala sigue del caso general. Esta extensión la denotamos por $\text{CUBPHN}(\xi, \eta, \alpha)$. En la Figura 2 se presenta la forma de la función de densidad CUBPHN con punto de censura $c \neq 0$, (a) bimodal para $\alpha < 1$ y (b) unimodal para $\alpha \geq 1$.

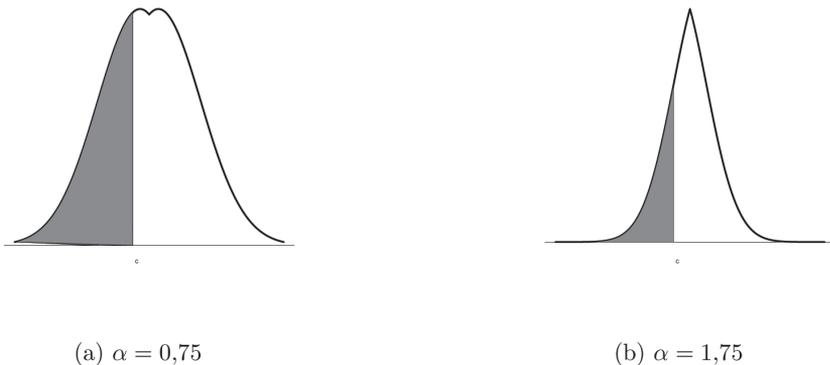


Figura 2. Ejemplos de la función de densidad $\text{UBPHN}(1,5;1;\alpha)$ con censura a la izquierda (zona en color gris).

Los momentos de la variable aleatoria X con distribución CUBPHN son funciones de los momentos incompletos de la variable

$$\mu_r(x) = \alpha \int_x^\infty z^r \phi(z) \{1 - \Phi(z)\}^{\alpha-1} dz, \quad r = 0, 1, 2, \dots \tag{10}$$

y el r -ésimo momento de la variable aleatoria X se da como

$$\mathbb{E}(X^r) = 2^{\alpha-1} \mu_r(0).$$

La generación de una variable aleatoria X con distribución CUBPHN se puede hacer teniendo en cuenta que para el modelo UBPHN(α) la función de distribución acumulada está dada por

$$F_{\Phi}(z; \alpha) = \begin{cases} \frac{1}{2} \{2\Phi(z)\}^{\alpha}, & \text{si } z < 0, \\ 1 - \frac{1}{2} \{2(1 - \Phi(z))\}^{\alpha}, & \text{si } z \geq 0. \end{cases} \quad (11)$$

La extensión de localización y escala del modelo CUBPHN(α) sigue de la transformación $X = \xi + \eta Z$. Así, para un valor $z = c$, tomando $\eta = 1$ y variando x_i , se obtienen diferentes proporciones de censura. La variable generada y los indicadores de la censura se pueden generar a partir de $X_i^* = \min(X_i, c)$ e $I_i = I_{\{X_i \leq c\}}$, indicando si la observación es censurada o no censurada, siendo I_i la función indicadora. La generación de esta variable aleatoria censurada puede hacerse de forma sencilla en cualquier lenguaje de programación; el programa R es una buena opción, dada la sencillez de su lenguaje de programación.

2.3. Estimación de los parámetros del modelo CUBPHN(ξ, η, α)

Para realizar la estimación de los parámetros del modelo CUBPHN(ξ, η, α) por máxima verosimilitud, vamos a suponer sin pérdida de generalidad que $c = 0$. Así, el aporte a la log-verosimilitud del i -ésimo dato censurado o por debajo del límite de detección es dado por

$$Prob(x_i \leq 0) = 2^{\alpha-1} \left(1 - \Phi\left(\frac{\xi}{\eta}\right) \right)^{\alpha}.$$

Sean $C = \{\text{Conjunto de datos censurados}\}$ y $B = \{\text{Conjunto de datos no censurados}\}$, entonces tenemos que la función de log-verosimilitud para la estimación de los parámetros $\theta = (\xi, \eta, \alpha)^T$, dada $\mathbf{X} = \{X_1, \dots, X_n\}$, está dada por

$$\begin{aligned} \ell(\theta; \mathbf{X}) &= n_1 \log(\alpha) + n \log(2)(\alpha - 1) - n_1 \log(\eta) + \alpha \sum_{i \in C} \log(1 - \Phi(z_0)) \\ &+ \sum_{i \in B} \log(\phi(z_i)) + (\alpha - 1) \sum_{i \in B} \log(1 - \Phi(|z_i|)), \end{aligned}$$

donde $z_i = \frac{x_i - \xi}{\eta}$, $z_0 = \xi/\eta$ y n_1 es el número de observaciones no censuradas. Así, la función de puntuación, definida como la derivada de la función de log-verosimilitud con respecto a cada uno de los parámetros es dada por:

$$\begin{aligned} U(\xi) &= -\frac{\alpha}{\eta} \sum_{i \in C} \frac{\phi(z_0)}{1 - \Phi(z_0)} + \frac{1}{\eta} \sum_{i \in B} z_i - \frac{\alpha - 1}{\eta} \sum_{i \in B} \text{sgn}(z_i) \frac{\phi(|z_i|)}{1 - \Phi(|z_i|)}, \\ U(\eta) &= \frac{\alpha \xi}{\eta^2} \sum_{i \in C} \frac{\phi(z_0)}{1 - \Phi(z_0)} - \frac{n_1}{\eta} + \frac{1}{\eta} \sum_{i \in B} z_i^2 + \frac{\alpha - 1}{\eta} \sum_{i \in B} |z_i| \frac{\phi(|z_i|)}{1 - \Phi(|z_i|)}, \\ U(\alpha) &= \sum_{i \in C} \log(1 - \Phi(z_0)) + \frac{n_1}{\alpha} + n \log(2) + \sum_{i \in B} \log[1 - \Phi(|z_i|)]. \end{aligned}$$

Igualando estas ecuaciones a cero obtenemos las respectivas ecuaciones de puntuación, cuya solución se debe calcular por métodos numéricos iterativos, llevando a los estimadores de máxima verosimilitud de ξ, η y α , denotados por $\hat{\xi}, \hat{\eta}$ y $\hat{\alpha}$.

2.4. Matriz de información observada

Los elementos de la matriz de información observada definidos como menos la segunda derivada de la función de log-verosimilitud con respecto a los parámetros, y denotados por $j_{\xi\xi}$, $j_{\xi\eta}$, $j_{\eta\eta}$, $j_{\xi\alpha}$, $j_{\eta\alpha}$ y $j_{\alpha\alpha}$, se dan mediante las fórmulas

$$\begin{aligned}
 j_{\xi\xi} &= n_0 \frac{\alpha}{\eta^2} w_0 [w_0 - z_0] + \frac{n_1}{\eta^2} - n_1 \frac{\alpha - 1}{\eta^2} \left[\overline{w^2} + \overline{\text{sgn}(z)zw} \right], \\
 j_{\xi\eta} &= n_0 \frac{\alpha}{\eta^2} w_0 [-w_0 z_0 + z_0^2 - 1] + \frac{2n_1}{\eta^2} \bar{z} + n_1 \frac{\alpha - 1}{\eta^2} \left[-\overline{\text{sgn}(z)|z|w^2} + \overline{\text{sgn}(z)z^2w} - \overline{\text{sgn}(z)w} \right], \\
 j_{\eta\eta} &= n_0 \frac{\alpha\xi w_0}{\eta^3} [2 + w_0 z_0 - z_0^2] - \frac{n_1}{\eta^2} + \frac{3n_1}{\eta^2} \bar{z}^2 + n_1 \frac{\alpha - 1}{\eta^2} \left[-2\overline{|z|w} - \overline{z^2w^2} + \overline{|z|^3w} \right], \\
 j_{\xi\alpha} &= \frac{n_0 w_0}{\eta} + \frac{n_1 \overline{\text{sgn}(z)w}}{\eta}, \quad j_{\eta\alpha} = -\frac{n_0 w_0 \xi}{\eta^2} - \frac{n_1 \overline{|z|w}}{\eta}, \quad j_{\alpha\alpha} = \frac{n_1}{\alpha},
 \end{aligned}$$

donde $w_0 = \frac{\phi(z_0)}{1 - \Phi(z_0)}$, $w_i = \frac{\phi(z_i)}{1 - \Phi(|z_i|)}$, $\bar{w} = \frac{\sum_{i \in B} w_i}{n_1}$, $\bar{zw} = \frac{\sum_{i \in B} z_i w_i}{n_1}$, $\overline{w^2} = \frac{\sum_{i \in B} w_i^2}{n_1}$, $\overline{\text{sgn}(z)zw} = \frac{\sum_{i \in B} \text{sgn}(z_i) z_i w_i}{n_1}$, $\overline{z^2 w^2} = \frac{\sum_{i \in B} z_i^2 w_i^2}{n_1}$ y sgn es la función signo. Para $\alpha = 1$ el modelo se reduce al modelo normal censurado.

3. Distribución unimodal asimétrica

Una extensión asimétrica del modelo UBPHN se puede definir aplicando el siguiente lema demostrado por Azzalini en [3].

Lema 3.1. Sean f_0 una función de densidad de probabilidad simétrica alrededor de cero y G una función de distribución tal que G' existe y es una función de densidad simétrica alrededor de cero; entonces

$$f(z; \lambda) = 2f_0(z)G(\lambda z) \tag{12}$$

es una función de densidad para cualquier $\lambda \in \mathbb{R}$.

Usando el Lema 3.1 definimos una extensión asimétrica para la distribución UBPHN(ξ, η, α), la cual llamamos distribución unimodal asimétrica con función de riesgo proporcional a la distribución normal, y la denotamos $X \sim \text{UBPHNA}(\xi, \eta, \alpha, \beta)$, cuya función de densidad de probabilidad está dada por

$$\varphi_{\Phi}(x; \xi, \eta, \alpha, \beta) = 2\frac{\alpha}{\eta} \phi\left(\frac{x - \xi}{\eta}\right) \left\{ 2 \left[1 - \Phi\left(\left|\frac{x - \xi}{\eta}\right|\right) \right] \right\}^{\alpha - 1} \Phi\left(\beta \frac{x - \xi}{\eta}\right), \tag{13}$$

donde $\beta \in \mathbb{R}$ es un parámetro de asimetría, $\xi \in \mathbb{R}$ es el parámetro de localización, $\eta \in \mathbb{R}^+$ es el parámetro de escala y $\alpha \in \mathbb{R}^+$.

Cuando en (13) $\alpha = 1$ obtenemos el modelo de localización y escala de la distribución normal asimétrico, denotado por $X \sim \text{SN}(\xi, \eta, \beta)$. Si $\alpha = 1$ y $\beta = 0$, obtenemos el modelo de localización y escala de la distribución normal, esto es, $X \sim \text{N}(\xi, \eta)$.

3.5. Distribución unimodal asimétrica para datos censurados

De la misma forma como se definió el modelo CUBPHN en la Sección 2.2, se define el modelo UBPHNA para datos censurados, el cual denotamos por $X \sim \text{CUBPHNA}(\xi, \eta, \alpha, \beta)$.

En la Figura 3 se presentan dos ejemplos de la forma de la función de densidad de la distribución CUBPHNA con punto de censura $c \neq 0$: (a) para $\beta = -0,25$ es bimodal con el pico mayor a la izquierda; y (b) para $\beta = 0,25$ es bimodal con el pico mayor a la derecha.

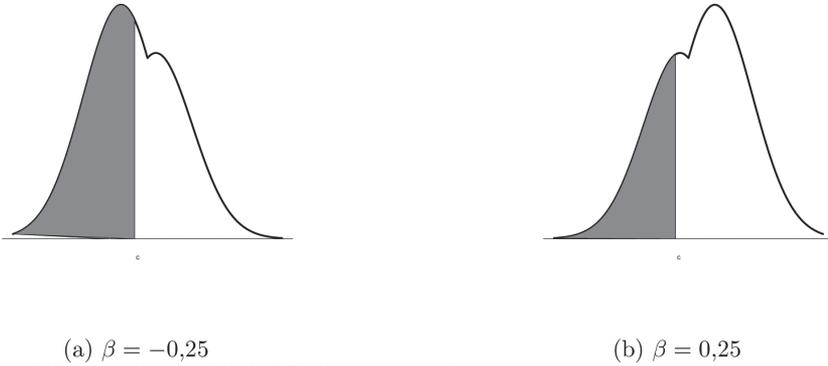


Figura 3. Función de densidad $\text{CUBPHNA}(1,5; 1; 0,5; \beta)$ con punto de censura a la izquierda (zona en color gris).

3.6. Estimación de los parámetros del modelo $\text{CUBPHNA}(\xi, \eta, \alpha, \beta)$

Para realizar la estimación de los parámetros del modelo $\text{CUBPHNA}(\xi, \eta, \alpha, \beta)$ por máxima verosimilitud, nuevamente suponemos sin pérdida de generalidad que $c = 0$. Así, la contribución a la verosimilitud de las observaciones menores o iguales a cero está dada por

$$\Psi(0) = \text{Prob}(x \leq 0) = \alpha 2^\alpha \int_{-\infty}^{-\frac{\xi}{\eta}} \phi(z) \{\Phi(z)\}^{\alpha-1} \Phi(\beta z) dz = \alpha 2^\alpha \varphi(0).$$

Para una muestra aleatoria $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$ con $X_i \sim \text{CUBPHNA}(\xi, \eta, \alpha, \beta)$, la función de log-verosimilitud de $\boldsymbol{\theta} = (\xi, \eta, \alpha, \beta)^\top$ dada \mathbf{X} está dada por

$$\begin{aligned} \ell(\boldsymbol{\theta}; \mathbf{X}) &= \sum_{i \in C} \ell_0 + n_1 \log(\alpha) + n_1 \alpha \log(2) - n_1 \log(\eta) + \sum_{i \in B} \log(\phi(z_i)) \\ &+ (\alpha - 1) \sum_{i \in B} \log(1 - \Phi(|z_i|)) + \sum_{i \in B} \log(\Phi(\beta z_i)), \end{aligned}$$

donde $z_i = \frac{x_i - \xi}{\eta}$ y $\ell_0 = \log(\psi(0))$. Así, la función de puntuación respecto a cada uno

de los parámetros está dada por:

$$\begin{aligned}
 U(\xi) &= \sum_{i \in C} \Lambda_\xi + \frac{1}{\eta} \sum_{i \in B} z_i - \frac{\alpha - 1}{\eta} \sum_{i \in B} \operatorname{sgn}(z_i) \frac{\phi(|z_i|)}{1 - \Phi(|z_i|)} - \frac{\beta}{\eta} \sum_{i \in B} \frac{\phi(\beta z_i)}{\Phi(\beta z_i)}, \\
 U(\eta) &= \sum_{i \in C} \Lambda_\eta + \frac{n_1}{\eta} + \frac{1}{\eta} \sum_{i \in B} z_i^2 + \frac{\alpha - 1}{\eta} \sum_{i \in B} |z_i| \frac{\phi(|z_i|)}{1 - \Phi(|z_i|)} - \frac{\beta}{\eta} \sum_{i \in B} z_i \frac{\phi(\beta z_i)}{\Phi(\beta z_i)}, \\
 U(\alpha) &= \sum_{i \in C} \Lambda_\alpha + \frac{n_1}{\alpha} + n_1 \log(2) + \sum_{i \in B} \log[1 - \Phi(|z_i|)], \\
 U(\beta) &= \sum_{i \in C} \Lambda_\beta + \sum_{i \in B} z_i \frac{\phi(\beta z_i)}{\Phi(\beta z_i)},
 \end{aligned}$$

donde Λ_{θ_j} está definido en el Apéndice 1. Igualando estas ecuaciones a 0 obtenemos las respectivas ecuaciones de puntuación, cuya solución nos ofrece los estimadores de máxima verosimilitud $\hat{\xi}$, $\hat{\eta}$, $\hat{\alpha}$ y $\hat{\beta}$.

3.7. Matriz de información observada

Los elementos de la matriz de información denotados por $j_{\xi\xi}, \dots, j_{\beta\beta}$ están dados por:

$$\begin{aligned}
 j_{\xi\xi} &= \Lambda_{\xi\xi} + \frac{n_1}{\eta^2} - n_1 \frac{\alpha - 1}{\eta^2} \left[\overline{w^2} + \overline{\operatorname{sgn}(z)zw} \right] + n_1 \frac{\beta^2}{\eta^2} \left[\overline{\beta z w_1} + \overline{w_1^2} \right], \\
 j_{\xi\eta} &= \Lambda_{\xi\eta} + \frac{2n}{\eta^2} \bar{z} + n_1 \frac{\alpha - 1}{\eta^2} \left[-\overline{\operatorname{sgn}(z)|z|w^2} + \overline{\operatorname{sgn}(z)z^2w} - \overline{\operatorname{sgn}(z)w} \right] \\
 &\quad + n_1 \frac{\beta}{\eta^2} \left[\overline{\beta^2 z^2 w_1} + \overline{\beta z w_1^2} - \overline{w_1} \right], \\
 j_{\eta\eta} &= \Lambda_{\eta\eta} - \frac{n_1}{\eta^2} + \frac{3n_1}{\eta^2} \bar{z}^2 + n_1 \frac{\alpha - 1}{\eta^2} \left[-2\overline{|z|w} - \overline{z^2 w^2} + \overline{|z|^3 w} \right] - \frac{\beta}{\eta} \overline{z w_1} \\
 &\quad + n_1 \frac{\beta}{\eta^2} \left[\overline{\beta^2 z^3 w_1} + \overline{\beta z^2 w_1^2} - 2\overline{z w_1} \right], \\
 j_{\xi\alpha} &= \Lambda_{\xi\alpha} + \frac{n_1}{\eta} \overline{\operatorname{sgn}(z)w}, \quad j_{\eta\alpha} = \Lambda_{\eta\alpha} - \frac{n_1}{\eta} \overline{|z|w}, \quad j_{\alpha\alpha} = \Lambda_{\alpha\alpha} + \frac{n_1}{\alpha}, \quad j_{\alpha\beta} = \Lambda_{\alpha\beta}, \\
 j_{\beta\eta} &= \Lambda_{\beta\eta} + \frac{n_1}{\eta} \left[\overline{z w_1} - \overline{\beta^2 z^3 w_1} - \overline{\beta z^2 w_1^2} \right], \quad j_{\beta\xi} = \Lambda_{\beta\xi} + \frac{n_1}{\eta} \overline{w_1} - n_1 \frac{\beta}{\eta^2} \left[\overline{\beta z^2 w_1} + \overline{z w_1^2} \right], \\
 j_{\beta\beta} &= \Lambda_{\beta\beta} + n_1 \left[\overline{\beta z^3 w} + \overline{z^2 w_1^2} \right], \quad j_{\beta\beta} = \Lambda_{\beta\beta} + n_1 \left[\overline{\beta z^3 w} + \overline{z^2 w_1^2} \right],
 \end{aligned}$$

donde $w_{1i} = \frac{\phi(\beta z_i)}{\Phi(\beta z_i)}$, $\overline{w_1} = \frac{\sum_{i \in B} w_{1i}}{n_1}$, $\overline{w_1^2} = \frac{\sum_{i \in B} w_{1i}^2}{n_1}$ y $\Lambda_{\theta_j \theta_j}$, son dados en el Apéndice 1.

La matriz de información de Fisher se obtiene calculando los valores esperados de las expresiones anteriores, y se denota como $I(\theta)$. Se puede demostrar que $|I(\theta)| \neq 0$; así pues, la matriz de información de Fisher es no singular. Por lo tanto, para muestras grandes el EMV $\hat{\theta}$ de $\theta = (\xi, \eta, \alpha, \beta)$ es asintóticamente normal, esto es,

$$\widehat{\theta} \xrightarrow{A} N_4(\theta, I(\theta)^{-1}), \quad (14)$$

resultando que la varianza asintótica de $\widehat{\theta}$ es la inversa de $I(\theta)$.

En consecuencia, con el resultado dado en la Ecuación 14 se pueden calcular los intervalos de confianza para cada uno de los parámetros del modelo.

4. Aplicación

En esta sección ajustamos el modelo CUBPHNA usando un conjunto de datos reales. La base de datos fue proporcionada por la Secretaría de Salud del Departamento de Santander, Colombia, y corresponde a personas con pruebas confirmadas de presencia del VIH que han sido reportadas al Sistema Nacional de Vigilancia en Salud Pública (SIVIGILA). Esta base de datos mantiene absoluta confidencialidad de la identificación del paciente y contiene, entre otras variables, la edad, el sexo, la fecha de ingreso al SIVIGILA, presencia o ausencia del tratamiento HAART, recuentos de CD-4, CD-8 y carga viral medida como número de copias de ARN VIH-1 realizadas con la prueba *Amplikor*[®] que tiene un límite de detección inferior de 50 copias/mL.

La base de datos contiene información de 1275 personas infectadas con el VIH, reportadas oficialmente al SIVIGILA hasta junio del 2012. En la base de datos existen pacientes que están en diferentes etapas de la enfermedad, pero sólo 681 pacientes reciben los medicamentos de la HAART y tienen mediciones de ARN VIH-1, 206 mujeres y 475 hombres. Para la ilustración, consideramos un subgrupo de 106 mujeres y 263 hombres infectados con el VIH que han sido tratados por lo menos un año con la HAART, dado que ya deben tener alguna respuesta de mejoría con el uso de esos medicamentos.

Inicialmente ajustamos los datos de carga viral de las 106 mujeres. La proporción de mujeres por debajo del límite de detección de 50 copias/mL es 34,90 %. La prueba de multimodalidad, definida por Hartigan y Hartigan en [11] e implementada computacionalmente en Hartigan [12], nos genera el valor $D = 0,0686$ de la estadística de prueba con un $p_{valor} = 0,0133$, rechazándose la hipótesis nula de que la distribución de la carga viral es unimodal. En el histograma presentado en la Figura 4-(a) se evidencia la existencia de bimodalidad en la distribución del conjunto de datos. Cuando la hipótesis nula es rechazada a un nivel de confianza del 95 %, la bimodalidad es confirmada si y sólo si se pueden observar dos picos “obvios” en la distribución de los datos (véase Zhang et al. [23], pág. 2851).

Para el ajuste del modelo CUBPHNA usamos el método de optimización *optim* del paquete estadístico *R*, el *optim* del *R* sirve para maximizar la función de verosimilitud. Los estimadores de máxima verosimilitud con sus respectivos errores estándar entre paréntesis, son: $\hat{\xi} = 1,502$ (0,127), $\hat{\eta} = 0,456$ (0,101), $\hat{\alpha} = 0,117$ (0,042) y $\hat{\beta} = -0,144$ (0,127).

La proporción estimada de datos censurados con el modelo CUBPHNA es 33,9% evidenciando un buen ajuste del modelo, dado que el valor en la base de datos es 34,90 %. Las Figuras 4-(b) y 4-(c) muestran el *qqplot* y la función de distribución acumulada del modelo estimado. Se puede observar en todas las gráficas el ajuste adecuado del modelo CUBPHNA.

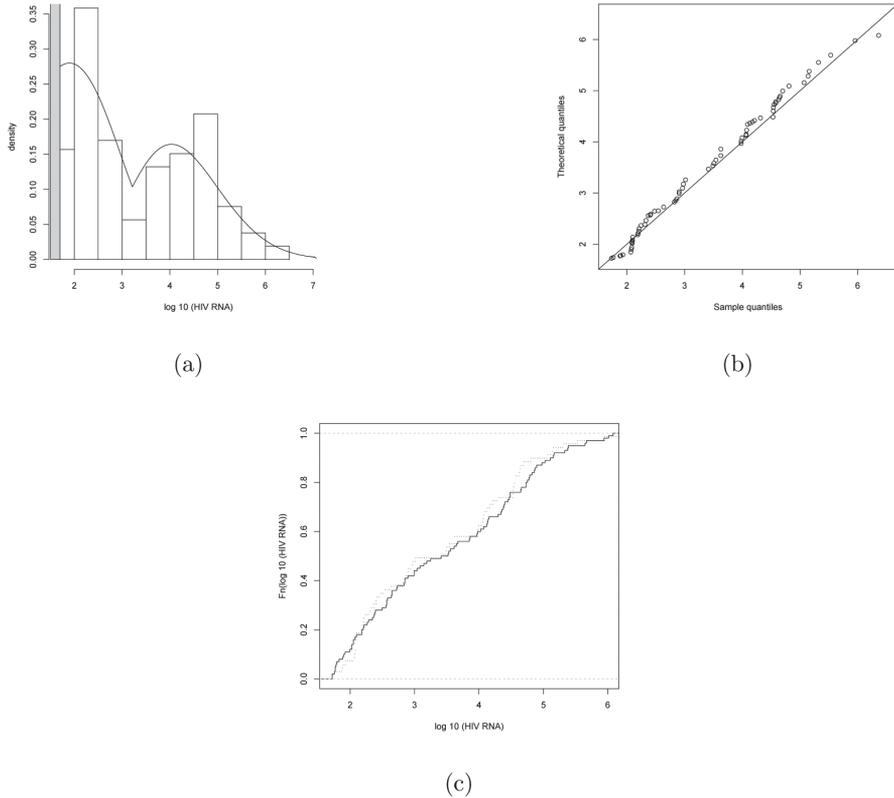


Figura 4. (a) Histograma de la carga viral para mujeres y modelo CUBPHAN ajustado, (b) *qqplot* y (c) fda.

Usando los datos de la carga viral de los hombres, en la Tabla 1 están calculados los estadísticos descriptivos para las observaciones por encima del límite de detección. Los valores $\sqrt{b_1}$ y b_2 corresponden a los coeficientes de asimetría y curtosis de los datos por encima de $\log_{10}(50)$, respectivamente. Indicando que el conjunto de datos presenta alta asimetría positiva y baja curtosis comparada con el modelo normal, así que el modelo normal censurado puede no ser la mejor elección para el ajuste de los datos de carga viral.

\bar{y}	s_y^2	$\sqrt{b_1}$	b_2
1,7350	1,7397	0,5258	2,1346

Tabla 1. Estadísticos descriptivos para las observaciones de carga viral de los hombres por encima de $\log_{10}(50)$.

Para realizar un estudio más completo con los datos de la carga viral de los hombres ajustamos también los modelos: normal para datos censurados (CN), normal asimétrica para datos censurados (CSN) y CUBPHNA. En la Tabla 2 están los estimadores de los parámetros del modelo con sus respectivas desviaciones estándar.

Para comparar los modelos utilizamos el criterio de Akaike (AIC), a saber,

$$AIC = -2 * \widehat{\ell}(\cdot) + 2p,$$

donde p es el número de parámetros para el modelo considerado. El mejor modelo es el que tiene el menor AIC. En este caso, según este criterio el mejor modelo es el modelo CUBPHNA.

Parámetro	CN	CSN	CUBPHNA
AIC	833,615	835,587	818,682
μ	0,477(0,137)	1,689(1,147)	1,698(0,080)
σ	1,978(0,121)	2,362(0,767)	0,370(0,137)
α		-0,861 (1,013)	0,060(0,039)
β			-0,142(0,055)

Tabla 2. Parámetros estimados (desviación estándar) de los modelos ajustados de la carga viral de los hombres.

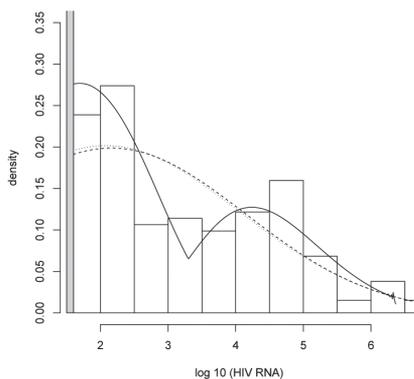


Figura 5. Histograma de la carga viral para hombres y modelos ajustados: CN (línea punteada), CSN (línea de guiones) y CUBPHNA (línea continua).

El histograma de la Figura 5 ofrece una fuerte evidencia de que el comportamiento de la variable ARN VIH-1 es bimodal. Los modelos CN y CSN sugieren ajustar adecuadamente la asimetría presente en la distribución de los datos, pero no la naturaleza bimodal, que parece ser mejor captada por el modelo CUBPHNA.

Otro test para probar la hipótesis de bimodalidad, dado que estos modelos son encajados, es

$$H_0 : \beta = 0 \quad \text{versus} \quad H_1 : \beta \neq 0,$$

el cual compara el modelo CSN contra el modelo CUBPHNA (el modelo es bimodal cuando $\beta \neq 0$). Podemos hacer uso de la razón de verosimilitudes,

$$\Lambda = \frac{\ell_{\text{CSN}}(\hat{\theta})}{\ell_{\text{CUBPHNA}}(\hat{\theta})};$$

de la Tabla 2 obtenemos

$$-2 \log(\Lambda) = -2(405,341 - 415,793) = 20,904,$$

que es mayor que el valor crítico de la distribución ji-cuadrado al 5% con un grado de libertad, a saber, $\chi^2_{1;5\%} = 3,8414$, rechazándose la hipótesis nula y aceptando la hipótesis alternativa de que la distribución de la carga viral de los hombres es bimodal, dando evidencia de que el modelo CUBPHNA es una buena alternativa para el ajuste de datos de ARN HIV-1 que tienen comportamiento bimodal.

Generalmente, varios autores utilizan la mixtura de dos distribuciones normales (MN) para ajustar este tipo de datos. La función de densidad de esta distribución está dada por

$$f(x; \mu_1, \sigma_1, \mu_2, \sigma_2, p) = pf_1(x, \mu_1, \sigma_1) + (1 - p)f_2(x; \mu_2, \sigma_2),$$

donde f_j corresponde a la función de densidad de la distribución normal con parámetros (μ_j, σ_j) , $j = 1, 2$ y $0 < p < 1$. Para datos censurados, denotamos el modelo de mixtura de dos distribuciones normales por $\text{CMN}(\mu_1, \sigma_1, \mu_2, \sigma_2, p)$.

El modelo ajustado usando la librería *gamlss.mx* del paquete *gamlss* creado por Stasinopoulos y Rigby [20] del programa *R* es

$$\text{CMN}(1,4780; 0,903; 4,4817; 0,921; 0,714),$$

con $\text{AIC}=819,8578$.

Como se puede observar, este modelo tiene AIC mayor que el AIC del modelo CUBPHNA, así que el modelo CUBPHNA ajusta mejor los datos de las mediciones de ARN HIV-1 del subgrupo de los hombres comparado con el modelo CMN. En la Figura 6-(a) se presenta el histograma con las gráficas de los modelos ajustados CMN y CUBPHNA.

Se resalta que el porcentaje de datos censurados es 39,92% de la muestra bajo estudio. El porcentaje estimado de observaciones censuradas bajo el modelo CUBPHNA es 39,64% mostrando indicios de una nueva opción para ajustar distribuciones de datos censurados con comportamientos bimodales. En la Figura 6-(b) se presenta el *qqplot* usando el modelo CUBPHNA, el cual indica un excelente ajuste para la mayoría de las observaciones.

5. Conclusiones

A pesar de que la mixtura de dos distribuciones normales es frecuentemente usada para ajustar datos con comportamientos bimodales, esta tiene restricciones para ajustar comportamientos asimétricos en uno o los dos componentes de la distribución. En este trabajo

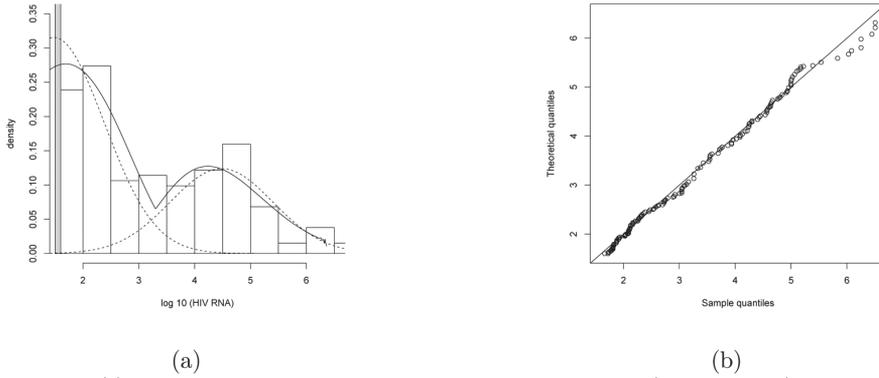


Figura 6. (a) Modelos ajustados para la carga viral de los hombres: CMN (línea punteada) y CUBPHNA (línea sólida), (b) *qqplot* calculado con el modelo CUBPHNA.

presentamos dos nuevos modelos bimodales: modelo unimodal simétrico con función de riesgo proporcional a la distribución normal y modelo unimodal asimétrico con función de riesgo proporcional a la distribución normal asimétrica. Estas distribuciones pueden ajustar asimetría y curtosis por fuera de los rangos establecidos con las otras distribuciones bimodales competentes. Además, estos nuevos modelos pueden tener comportamiento unimodal y bimodal. La inclusión de un nuevo parámetro puede explicar simultáneamente el comportamiento asimétrico y bimodal de una gran familia de distribuciones normales, lo que permite tener un modelo más flexible que la mixtura de distribuciones normales. La aplicación con datos reales de personas infectadas con VIH evidencian que esta propuesta ajusta mejor los datos de la carga viral que otros modelos ya existentes, y las pruebas de hipótesis así lo confirmaron. En resumen, el modelo CUBPHNA tiene el potencial de ajustar adecuadamente datos censurados con comportamientos bimodales, con alta o baja asimetría y con alta o baja curtosis.

Apéndice 1

$$\Lambda_{\xi} = -\frac{1}{\eta}w_{01}, \quad \Lambda_{\eta} = \frac{\xi}{\eta^2}w_{01}, \quad \Lambda_{\alpha} = \frac{1}{\alpha} + \log(2) + w_{021}, \quad \Lambda_{\beta} = w_{031}, \quad (15)$$

donde

$$w_{01} = \frac{\phi\left(\frac{\xi}{\eta}\right)\left\{\Phi\left(-\frac{\xi}{\eta}\right)\right\}^{\alpha-1}\Phi\left(-\beta\frac{\xi}{\eta}\right)}{\varphi(0)}, \quad w_{03k} = \frac{1}{\varphi(0)}\int_{-\infty}^{-\frac{\xi}{\eta}}z^k\phi(z)\{\Phi(z)\}^{\alpha-1}\phi(\beta z)dz.$$

$$y \quad w_{02k} = \frac{1}{\varphi(0)}\int_{-\infty}^{-\frac{\xi}{\eta}}\phi(z)\{\log[\Phi(z)]\}^k\{\Phi(z)\}^{\alpha-1}\Phi(\beta z)dz.$$

$$\Lambda_{\xi\xi} = \frac{1}{\eta^2}\left[w_{01}^2 - \beta w_{04} - \frac{\xi}{\eta}w_{01} - (\alpha - 1)w_{00}w_{01}\right],$$

$$\Lambda_{\eta\eta} = \frac{\xi}{\eta^4} \left[\xi\beta w_{04} + \xi w_{01}^2 - \left(\frac{\xi^2}{\eta} + \xi(\alpha - 1)w_{00} - 2\eta \right) w_{01} \right],$$

$$\Lambda_{\alpha\alpha} = \frac{1}{\alpha^2} + w_{021}^2 + w_{022}, \quad \Lambda_{\beta\beta} = w_{031}^2 + \beta w_{033},$$

$$\Lambda_{\xi\eta} = \frac{1}{\eta^4} [\beta\xi\eta w_{04} - \xi\eta w_{01}^2 + [\xi\eta(\alpha - 1)w_{00} - \eta^2 + \xi^2]w_{01}],$$

$$\Lambda_{\xi\alpha} = \frac{w_{01}}{\eta} [u_0 - w_{021}], \quad \Lambda_{\xi\beta} = \frac{1}{\eta^2} [-\xi w_{04} - \eta w_{01} w_{031}],$$

$$\Lambda_{\eta\alpha} = \frac{\xi}{\eta^2} w_{01} [w_{021} - u_0], \quad \Lambda_{\eta\beta} = \frac{\xi}{\eta^3} [\eta w_{031} + \xi w_{04}], \quad \Lambda_{\alpha\beta} = w_{021} w_{031} - w_{05},$$

donde $u_0 = \log \left[\Phi \left(-\frac{\xi}{\eta} \right) \right]$ y $w_{05} = \frac{1}{\varphi(0)} \int_{-\infty}^{-\frac{\xi}{\eta}} z\phi(z) \{\log[\Phi(z)]\} \{\Phi(z)\}^{\alpha-1} \phi(\beta z) dz$.

Agradecimientos

Los datos en este artículo fueron recolectados por la Secretaría de Salud del Departamento de Santander, de acuerdo con los reportes realizados por las Unidades Primarias Generadoras de Datos y Unidades Notificadoras Municipales. Esta base de datos es parte integral del programa del Sistema Nacional de Vigilancia en Salud Pública (SIVIGILA) del Departamento de Santander, Colombia. El contenido de la publicación es responsabilidad de los autores y no representa las opiniones oficiales de la Secretaría de Salud del Departamento de Santander. Los autores agradecen a los tres árbitros por las revisiones y sugerencias que hicieron para lograr esta versión final del artículo. Germán Moreno-Arenas agradece al Programa de Movilidad de Investigadores de la Vicerrectoría de Investigación de la Universidad Industrial de Santander.

Referencias

- [1] Amplicor HIV-1 MONITOR[®]. *Test, version 1.5, Branchburg, NJ: Roche Molecular Systems, Inc., 2002.*
- [2] Arnold B., Gómez H. and Salinas H., “On multiple constraint skewed models”, *Statistics* 43 (2009), no. 3, 279–293.
- [3] Azzalini A. “A class of distributions which includes the normal ones”, *Scandinavian Journal of Statistics* 12 (1985), 171–178.
- [4] Biomarkers Definitions Working Group: Biomarkers and surrogate endpoints. Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69 (2001), 89–95.
- [5] Chiogna M., “Some results on the scalar skew-normal distribution”, *Journal of the Italian Statistical Society* 1 (1998), 1–14.
- [6] COBAS[®] AmpliPrep/COBAS[®] TaqMan[®] HIV-1 Test, version 2.0. Branchburg, NJ: Roche Molecular Systems, Inc., 2010.

- [7] Cordeiro G. and de Castro M., “A new family of generalized distributions”, *J. Statist. Comput. Simul.* 81 (2012), 883–898.
- [8] Durrans S.R., “Distributions of fractional order statistics in hydrology”, *Water Resources Research* 28 (1992), no. 6, 1649–1655.
- [9] Gómez H., Venegas O., and Bolfarine H., “Skew-symmetric distributions generated by the distribution function of the normal distribution”, *Environmetrics* 18 (2007), 395–407.
- [10] Gómez H., Elal-Olivero D., Salinas H., and Bolfarine H., “Bimodal extension based on the Skew-Normal distribution with application to pollen data”, *Environmetrics* 22 (2009), 50–62.
- [11] Hartigan J.A. and Hartigan P.M., “The dip test of unimodality”, *Ann. Stat.* 13 (1985), 70–84.
- [12] Hartigan P.M., “Computation of the dip statistics to test for unimodality”, *Appl. Stat.* 34 (1985), 320–325.
- [13] Henze N., “A probabilistic representation of the skew-normal distribution”, *Scandinavian Journal of Statistics* 13 (1986), 271–275.
- [14] Kim H., “On a class of two-piece skew-normal distribution”, *Statistic* 39 (2005), no. 6, 537–553.
- [15] Li X., Chu H., Gallant J., Hoover D., Mack W., Chmiel J., and Muñoz A., “Bimodal virological response to antiretroviral therapy for HIV infection: an application using a mixture model with left censoring”, *Journal Epidemiol. Community Health* 60 (2006), no. 9, 811–818.
- [16] Lim T., Bakri R., Morad Z., and Hamid M., “Bimodality in blood glucose distribution. Is it universal?”, *Diabetes Care* 25 (2002), 2212–2217.
- [17] O’Hagan A. and Leonard T., “Bayes estimation subject to uncertainty about parameter constraints”, *Biometrika* 63 (1976), 201–203.
- [18] Pewsey A., “Problems of inference for Azzalini’s skew-normal distribution”, *Journal of Applied Statistics* 27 (2000), no. 7, 859–870.
- [19] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012.
- [20] Stasinopoulos M.D. and Rigby R.A., “Generalized Additive Models for Location Scale and Shape (GAMLSS) in R”, *Journal of Statistical Software* 23 (2007), no. 7, 1–46
- [21] Schneider M., Margolick J., Jacobson L., Reddy S., Martinez–Maza O., and Muñoz A., “Improved estimation of the distribution of suppressed plasma HIV-1 RNA in men receiving effective antiretroviral therapy”, *Journal Acquir. Immune. Defic. Syndr.* 59 (2012), no. 4, 389–392.
- [22] Tobin J., “Estimation of relationships for limited dependent variables”, *Econometrica* 26 (1958), no. 1, 24–36.
- [23] Zhang C., Mapes B.E., and Soden B.J., “Bimodality in tropical water vapour”, *Q. J. R. Meteorol. Soc.* 129 (2003), 2847–2866.