

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



UN MODELO BAYESIANO Y NO PARAMÉTRICO
DE REGRESIÓN SOBRE CUANTILES

TESIS

QUE PARA OBTENER EL TÍTULO

LICENCIADO EN MATEMÁTICAS APLICADAS

PRESENTA

CARLOS OMAR PARDO GÓMEZ

ASESOR: DR. JUAN CARLOS MARTÍNEZ OVANDO

México, D.F.

2017

Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada **"UN MODELO BAYESIANO Y NO PARAMÉTRICO DE REGRESIÓN SOBRE CUANTILES"**, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., la autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación.

CARLOS OMAR PARDO GÓMEZ

FECHA

FIRMA

A Mago.

Agradecimientos

¡Muchas gracias a todos!

Prefacio

El centro de esta tesis es describir un modelo de *regresión sobre cuantiles*, debido a las bondades que tiene sobre el comúnmente usado análisis de *regresión sobre la media*. Además, aceptando los axiomas de la Estadística Bayesiana, permite incorporar conocimiento previo del modelador. Por otra parte, el modelo es no paramétrico, aumentando la flexibilidad en su forma.

El capítulo 1 describe la importancia de las aproximaciones distintas a la *regresión sobre la media*, así como la evolución histórica de este tipo de modelos. El capítulo 2 introduce al paradigma bayesiano y sus metodologías generales. El capítulo 3 se centra en los modelos bayesianos tradicionales de regresión, tanto a la media, como sobre cuantiles. El capítulo 4 plantea la especificación no paramétrica específica del modelo de esta tesis, separándolo de los tradicionales. El capítulo 5 introduce las variables latentes y algoritmos necesarios para realizar inferencia y predicción con lo expuesto en los capítulos anteriores. El capítulo 6 muestra algunas aplicaciones del modelo, así como los resultados obtenidos de evaluarlo en diversos conjuntos de datos. Finalmente, el capítulo 7 hace referencia a las conclusiones finales de esta tesis, además de describir el trabajo futuro que se podría desarrollar, retomando las ideas de ésta.

Índice general

1. Introducción	6
2. Paradigma bayesiano	9
2.1. Axiomas	9
2.2. Inferencia	9
2.3. Propiedad conjugada	12
3. Modelos de regresión	13
3.1. Concepto general	13
3.2. Regresión a la media	14
3.2.1. Modelo tradicional	15
3.3. Regresión sobre cuantiles	18
3.3.1. Modelo tradicional	19
Bibliografía	22

Capítulo 1

Introducción

Detrás de cualquier modelo de regresión, la intención es explicar una variable dependiente en función de un conjunto de variables independientes, suponiendo cierto error aleatorio. Ha sido común resumir esta dependencia mediante alguna medida de tendencia central, condicionada a los valores de las covariables.

La medida de tendencia central tradicionalmente usada ha sido la media, dando lugar a los modelos de *regresión sobre la media*, con sus variantes lineal o no lineal, simple o múltiple. Este tipo de modelos tiene un buen número de ventajas, entre las que destacan el bajo costo de calcularlos y la facilidad de interpretación. Sin embargo, como mencionan Hao & Naiman (2007), tienen tres grandes limitaciones.

La primera es que al resumir la relación entre la variable dependiente y las independientes con el valor esperado, no necesariamente se puede extender la inferencia a valores lejanos a la media, que suelen ser de interés en ciertos contextos, como los seguros o las finanzas.

La segunda es que los supuestos de este tipo de modelos no siempre se cumplen en el mundo real. Por ejemplo, el supuesto de homocedasticidad; es decir, la varianza no es constante, sino cambia en sincronía con

distintos valores de las covariables. También es posible que fenómenos de estudio tengan distribuciones de colas pesadas, principalmente en las ciencias sociales. Esto da lugar a valores atípicos, mismos que no suelen ser manejados como se desearía por los modelos de *regresión sobre la media*.

La tercera es que no permiten conocer las propiedades y forma de la distribución completa. Por ejemplo, la asimetría es una característica importante en estudios de ingreso, impuestos, esperanza de vida y, en general, en estudios de desigualdad.

Debido a esto, desde mitades del siglo XVIII han surgido alternativas a este tipo de modelos, siendo la primera los modelos de *regresión sobre la mediana*. De nueva cuenta se buscó una medida de tendencia central, pero con otras bondades. Por ejemplo, ser una mejor medida informativa para distribuciones asimétricas y menos susceptible a valores atípicos.

Así como los *modelos de regresión sobre la media* son comúnmente relacionados con la minimización de los errores cuadráticos, los *modelos de regresión sobre la mediana* lo son con la minimización de los errores absolutos. Debido a la no diferenciabilidad, tuvieron que pasar muchos años para que lograran ser viables, hasta que el poder computacional y los algoritmos de Programación Lineal lo permitieron.

Cabe recordar que el cuantil p -ésimo es aquel valor tal que una proporción p de los valores están por debajo de él, y una proporción $1 - p$, por arriba. Así, la mediana es un caso particular de un cuantil, específicamente el 0.5-ésimo. Esto abre la idea de que otros cuantiles también podrían ser modelados en función de las covariables, y no necesariamente tienen que ser una medida de tendencia central.

Los *modelos de regresión sobre cuantiles* fueron introducidos por Koenker & Bassett (1978), y han permitido concentrarse en valores de interés para los modeladores, sin importar que estén alejados de la media. Además, el cálculo de diversos cuantiles para un mismo fenómeno ha

permitido entender mejor la forma y propiedades de las distribuciones condicionales de la variable de respuesta.

En el paradigma bayesiano, el desarrollo de este tipo de modelos ha sido lento. Walker & Mallick (1999), Kottas & Gelfand (2001) y Hanson & Johnson (2002) desarrollaron modelos para la mediana, suponiendo una distribución no paramétrica del error. Yu & Moyeed (2001) y Tsionas (2003) desarrollaron inferencia paramétrica, basados en la distribución asimétrica de Laplace para los errores. Por otro lado, Lavine (1995) y Dunson & Taylor (2005) usaron una perspectiva distinta y propusieron una aproximación de la verosimilitud para cuantiles.

Las limitantes de estos trabajos han sido que, aunque han dado formas flexibles a la distribución del error, han estado basados en funciones lineales para describir la relación entre la variable de respuesta y las covariables, o han tenido que recurrir a estimaciones no probabilísticas o no bayesianas, para resolver alguna parte del problema.

Entendiendo como *modelo no paramétrico* a aquel en el que el número de parámetros no está previamente definido, sino que depende de los datos, esta tesis rescata las ideas de Kottas *et al.* (2007) para proponer un modelo bayesiano totalmente no paramétrico, útil en el contexto de *regresión sobre cuantiles*.

Capítulo 2

Paradigma bayesiano¹

2.1. Axiomas

Esta tesis da como aceptados los axiomas de la Estadística Bayesiana, detallados durante muchos años en la literatura. Por ejemplo, pueden ser encontrados en Fishburn (1986). Por lo tanto, entiende a dicho paradigma como el coherente para hacer estadística, cuando una toma de decisión con incertidumbre es el objetivo final del estudio.

2.2. Inferencia

Un problema clásico de la estadística es el de hacer predicción, utilizando la información de los datos que ya han sido observados. Por ejemplo, es posible pensar que ya se tiene el conjunto de n datos observados $\{y_1, \dots, y_n\}$ y se desea hacer predicción acerca del valor del dato y_{n+1} , que aún no ha sido observado. Para esto, se podría usar la

¹Las ideas de este capítulo son retomadas de Denison (2002).

probabilidad condicional

$$\mathbb{P}(y_{n+1}|y_1, \dots, y_n) = \frac{\mathbb{P}(y_{n+1} \cap \{y_1, \dots, y_n\})}{\mathbb{P}(y_1, \dots, y_n)} = \frac{\mathbb{P}(y_1, \dots, y_n, y_{n+1})}{\mathbb{P}(y_1, \dots, y_n)},$$

pero esto requeriría conocer la función conjunta, misma que puede ser compleja por la estructura de dependencia de los datos.

No tiene mucho sentido suponer una estructura de independencia entre ellos, porque entonces el conjunto de observaciones $\{y_1, \dots, y_n\}$ no daría información alguna para y_{n+1} . Pero se puede suponer una distribución condicionalmente independiente. Es decir, se supone que cada una de las y_i 's tiene una misma distribución paramétrica, con vector de parámetros θ , y se cumple que

$$\mathbb{P}(y_{k+1}, y_k|\theta) = \mathbb{P}(y_k|\theta) \times \mathbb{P}(y_{k+1}|\theta).$$

Siguiendo el mismo razonamiento, es posible obtener que

$$\mathbb{P}(y_1, \dots, y_n|\theta) = \prod_{i=1}^n \mathbb{P}(y_i|\theta).$$

Dado que se desea hacer inferencia, y al igual que en otros paradigmas, se supone a θ como constante, pero desconocido. Una particularidad del paradigma bayesiano es expresar la incertidumbre que tiene el modelador acerca del valor verdadero mediante la asignación de una distribución a θ , sujeta la información inicial o conocimiento previo que se tenga del fenómeno (H). Es decir, $\mathbb{P}(\theta|H)$. Como una simplificación de la notación, en la literatura normalmente se escribe como $\mathbb{P}(\theta) = \mathbb{P}(\theta|H)$ y se conoce como la *probabilidad inicial* del parámetro.

Regresando al problema inicial, y bajo los supuestos recién mencio-

nados, es importante notar que es posible escribir

$$\mathbb{P}(y_{n+1}|y_1, \dots, y_n) = \int_{\Theta} \mathbb{P}(y_{n+1}|\theta) \mathbb{P}(\theta|y_1, \dots, y_n) d\theta,$$

donde a su vez, usando el **Teorema de Bayes**, se obtiene que

$$\mathbb{P}(\theta|y_1, \dots, y_n) = \frac{\mathbb{P}(y_1, \dots, y_n|\theta) \times \mathbb{P}(\theta)}{\mathbb{P}(y_1, \dots, y_n)},$$

que en el paradigma bayesiano se conoce como la *probabilidad posterior* del parámetro.

Se puede observar que el denominador no depende de θ , por lo que normalmente la probabilidad no se expresa como una igualdad, sino con la proporcionalidad

$$\mathbb{P}(\theta|y_1, \dots, y_n) \propto \mathbb{P}(y_1, \dots, y_n|\theta) \times \mathbb{P}(\theta),$$

y sólo difiere de la igualdad por una constante que permita que al integrar sobre todo el soporte de θ el resultado sea igual a 1.

Cabe resaltar que el factor $\mathbb{P}(y_1, \dots, y_n|\theta)$ es lo que se conoce también en otros paradigmas como *verosimilitud*, y que en caso de independencia condicional puede ser reescrito como

$$\mathbb{P}(y_1, \dots, y_n|\theta) = \prod_{i=1}^n \mathbb{P}(y_i|\theta).$$

Por lo tanto, es posible afirmar que el aprendizaje en el paradigma bayesiano se obtiene como

$$Posterior \propto Verosimilitud \times Inicial,$$

es decir, surge de conjuntar el conocimiento inicial con la información contenida en los datos.

Es importante notar que bajo este enfoque se obtiene una distribución de probabilidad completa para el pronóstico de y_{n+1} . Esta se puede utilizar para el cálculo de estimaciones puntuales o intervalos, que en el caso del paradigma bayesiano son llamados de *probabilidad*, mediante el uso de funciones de utilidad o pérdida, mismas hacen uso de la Teoría de la Decisión.

2.3. Propiedad conjugada

En los casos en los que la probabilidad posterior tiene la misma familia de distribución que la inicial, sólo siendo distintas en el valor de los parámetros, se dice que la distribución inicial y la posterior son **conjugadas**.

Esta propiedad es conveniente, porque permite a la distribución posterior tener forma analítica cerrada, evitando tener que usar métodos numéricos para aproximarla. Además permite ver de forma más clara cómo afectan los datos la actualización respecto a la distribución inicial.

Es demostrable que todas las distribuciones de la familia exponencial para los datos, tienen distribuciones iniciales conjugadas para los parámetros. Algunas de las más conocidas la *Normal-Normal*, *Normal-Gamma* o la *Normal-Gamma Inversa*, donde la primer distribución es la de los datos y la segunda la de los parámetros. También en el caso discreto es popular el uso de la *Bernoulli-Beta* o la *Poisson-Gamma*.

Capítulo 3

Modelos de regresión

3.1. Concepto general

Los modelos de regresión tienen como objetivo describir la distribución de una variable aleatoria $y \in \mathbb{R}$, normalmente conocida como la *variable de respuesta*, condicional a los valores de las variables $x \in \mathbb{R}^n$, conocidas como *covariables* o *variables de entrada*. Visto en términos matemáticos, se puede expresar como

$$y|x \sim \mathbb{P}(y|x).$$

Si bien esta relación se da por hecha y es fija, normalmente es desconocida. Por lo tanto, la intención de estos modelos es realizar alguna aproximación de ella. Dado que es complicado aproximar con exactitud toda la distribución, comúnmente se enfocan en una medición particular, como la media o la mediana.

3.2. Regresión a la media

La *regresión a la media* es el caso particular más usado de los modelos de regresión, tanto en el paradigma bayesiano, como en otros. Esto sucede debido al bajo uso de recursos, además de su capacidad interpretativa.

En notación probabilística, retomando el hecho de que $y|x \sim \mathbb{P}(y|x)$, busca aproximar a la función f , tal que

$$\mathbb{E}(y|x) = f(x).$$

Para hacer esto, normalmente se vale del supuesto que

$$y = f(x) + \varepsilon,$$

con $\varepsilon \in \mathbb{R} \sim \mathcal{N}(0, \sigma^2)$ (denominado comúnmente como el *error aleatorio*), y siendo $f(x) \in \mathbb{R}$ y $\sigma^2 \in \mathbb{R}^+$ desconocidas, de forma que

$$y \sim \mathcal{N}(f(x), \sigma^2).$$

Además, retomando lo visto en el capítulo anterior, se supone independencia condicional. Es decir, sean \bar{x} las covariables asociadas a la variable de respuesta \bar{y} , y \hat{x} las asociadas a \hat{y} , se tiene que $\bar{y}|\bar{x}, f, \sigma^2$ es condicionalmente independiente a $\hat{y}|\hat{x}, f, \sigma^2$. Y, por lo tanto, $\bar{\varepsilon}$ es independiente de $\hat{\varepsilon}$.

3.2.1. Modelo tradicional ¹

La *regresión lineal a la media* es el caso particular más usado en el contexto de *regresión a la media*. Consiste en definir

$$f(x) = \beta^T x,$$

donde $\beta \in \mathbb{R}^n$ se piensa con valores constantes, pero desconocidos, y la tarea es estimarlos, al igual que σ^2 .

Para hacer esto, el enfoque bayesiano le asigna una distribución inicial de probabilidad a ambos parámetros, reflejando la incertidumbre que tiene el modelador acerca de su valor real. Es decir, sea H la hipótesis o el conocimiento previo al que tiene acceso el modelador, se tiene que

$$\beta, \sigma^2 \sim \mathbb{P}(\beta, \sigma^2 | H).$$

A partir de este momento se omitirá escribir la distribución condicional respecto a H por simplificación de la notación, pero es importante no olvidar su existencia.

Sea $\{(x_i, y_i) | x_i \in \mathbb{R}^n, y_i \in \mathbb{R}, i \in \{1, \dots, m\}\}$ el conjunto de datos observados, condicionalmente independientes e idénticamente distribuidos, de las variables de respuesta y de las covariables. Es posible representar este mismo conjunto con la notación matricial $\{X, Y | X \in \mathbb{R}^{m \times n}, Y \in \mathbb{R}^m\}$. Sea $\mathcal{E} \in \mathbb{R}^m$ el vector de errores aleatorios, tal que $\mathcal{E} \sim \mathcal{N}(0, \sigma^2 I)$. El modelo se puede reescribir como:

$$Y = X\beta + \mathcal{E} \sim \mathcal{N}(X\beta, \sigma^2 I).$$

¹Algunas ideas de esta subsección son retomadas de Denison (2002) y Bannerjee (2008).

Por el Teorema de Bayes,

$$\begin{aligned}
\mathbb{P}(\beta, \sigma^2 | Y, X) &= \frac{\mathbb{P}(Y|X, \beta, \sigma^2) \times \mathbb{P}(\beta, \sigma^2 | X)}{P(Y|X)} \\
&= \frac{\mathbb{P}(Y|X, \beta, \sigma^2) \times \mathbb{P}(\beta, \sigma^2)}{\mathbb{P}(Y|X)} \\
&\propto \mathbb{P}(Y|X, \beta, \sigma^2) \times \mathbb{P}(\beta, \sigma^2),
\end{aligned}$$

donde $\mathbb{P}(Y|X, \beta, \sigma^2)$ es la verosimilitud de los datos observados y se puede calcular como $\mathbb{P}(Y|X, \beta, \sigma^2) = \mathcal{N}(X\beta, \sigma^2 I) = \prod_{i=1}^m \mathcal{N}(x_i^T \beta, \sigma^2)$. Por otro lado, $\mathbb{P}(\beta, \sigma^2)$ es la distribución inicial de los parámetros.

Por conveniencia analítica, hay una distribución inicial comúnmente usada para los parámetros β y σ debido a que es conjugada respecto a la distribución Normal de los datos. Su nombre es *Normal-Gamma Inversa (NGI)* y se dice que $\beta, \sigma^2 \sim \mathcal{NGI}(M, V, a, b)$, si

$$\begin{aligned}
\mathbb{P}(\beta, \sigma^2) &= \mathbb{P}(\beta | \sigma^2) \times \mathbb{P}(\sigma^2) \\
&= \mathcal{N}(\beta | M, \sigma^2 V) \times \mathcal{GI}(\sigma^2 | a, b) \\
&= \frac{1}{((2\pi)^n |\sigma^2 V|)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\beta - M)^T (\sigma^2 V)^{-1} (\beta - M)\right) \\
&\quad \times \frac{b^a}{\Gamma(a)} (\sigma^2)^{-(a+1)} \exp\left(-\frac{b}{\sigma^2}\right) \\
&= \frac{b^a}{(2\pi)^{\frac{n}{2}} |V|^{\frac{1}{2}} \Gamma(a)} (\sigma^2)^{-(a+(n/2)+1)} \\
&\quad \times \exp\left(-\frac{(\beta - M)^T V^{-1} (\beta - M) + 2b}{2\sigma^2}\right) \\
&\propto (\sigma^2)^{-(a+(n/2)+1)} \exp\left(-\frac{(\beta - M)^T V^{-1} (\beta - M) + 2b}{2\sigma^2}\right),
\end{aligned}$$

donde M es la media inicial de los coeficientes, $\sigma^2 V$ su varianza, y a y b son los parámetros iniciales de forma y medida de σ^2 .

Aprovechando la propiedad conjugada, es posible escribir la probabilidad posterior de los parámetros como:

$$\begin{aligned}\mathbb{P}(\beta, \sigma^2 | Y, X) &\propto \mathbb{P}(Y | X, \beta, \sigma^2) \times \mathbb{P}(\beta, \sigma^2), \\ &\propto (\sigma^2)^{-(\bar{a} + (n/2) + 1)} \exp \left(-\frac{(\beta - \bar{M})^T \bar{V}^{-1} (\beta - \bar{M}) + 2\bar{b}}{2\sigma^2} \right),\end{aligned}$$

donde

$$\begin{aligned}\bar{M} &= (V^{-1} + X^T X)^{-1} (V^{-1} M + X^T Y), \\ \bar{V} &= (V^{-1} + X^T X)^{-1}, \\ \bar{a} &= a + n/2, \\ \bar{b} &= b + \frac{\bar{M}^T V^{-1} M + Y^T Y - \bar{M}^T \bar{V}^{-1} \bar{M}}{2}.\end{aligned}$$

Es decir, la distribución posterior de (β, σ^2) es *Normal - Gamma Inversa*, con parámetros $\mathcal{NGI}(\bar{M}, \bar{V}, \bar{a}, \bar{b})$.

Si se tiene una nueva matriz de covariables X_* y se desea hacer predicción de las respectivas variables de salida Y_* , es posible hacer inferencia con los datos observados de la siguiente manera:

$$\begin{aligned}\mathbb{P}(Y_* | X_*, Y, X) &= \int \int \mathbb{P}(Y_* | X_*, \beta, \sigma^2) \times \mathbb{P}(\beta, \sigma^2 | Y, X) d\sigma^2 d\beta \\ &= \int \int \mathcal{N}(Y_* | X_* \beta, \sigma^2 I) \times \mathbb{P}(\beta, \sigma^2 | Y, X) d\sigma^2 d\beta.\end{aligned}$$

Particularmente, si se continúa con el modelo conjugado *Normal - Gamma Inversa / Normal*, es posible encontrar la solución analítica:

$$\begin{aligned}\mathbb{P}(Y_* | X_*, Y, X) &= \int \int \mathcal{N}(Y_* | X_* \beta, \sigma^2 I) \times \mathbb{P}(\beta, \sigma^2 | Y, X) d\sigma^2 d\beta \\ &= \int \int \mathcal{N}(Y_* | X_* \beta, \sigma^2 I) \times \mathcal{NGI}(\beta, \sigma^2 | \bar{M}, \bar{V}, \bar{a}, \bar{b}) d\sigma^2 d\beta \\ &= MVSt_{2\bar{a}} \left(X_* \bar{M}, \frac{\bar{b}}{\bar{a}} \left(I + X_* \bar{V} X_*^T \right) \right),\end{aligned}$$

donde $MVSt$ es la distribución t -Student multivariada, y cuya definición se describe a continuación.

Definición 1. Sea $X \in \mathbb{R}^p$ un vector aleatorio, con media, mediana y moda μ , matriz de covarianzas Σ , y ν grados de libertad, entonces $X \sim MVSt_\nu(\mu, \Sigma)$ si y sólo si su función de densidad es:

$$f(x|\mu, \sigma, \nu) = \frac{\Gamma((\nu + p)/2)}{\Gamma(\nu/2)\nu^{p/2}\pi^{p/2}|\Sigma|^{1/2}} \left[1 + \frac{1}{\nu}(x - \mu)^T \Sigma^{-1}(x - \mu) \right]^{-\frac{\nu+p}{2}}.$$

3.3. Regresión sobre cuantiles

Definición 2. El cuantil p -ésimo de la variable aleatoria Y es aquel valor q_p tal que

$$F_Y(q_p) = p.$$

Equivalentemente, la función que regresa el cuantil p -ésimo de la variable aleatoria Y se escribe

$$q_p(Y) = F_Y^{-1}(p),$$

cuando F_Y^{-1} está bien definida.

Dicho en otras palabras, si se tiene un conjunto grande de realizaciones de una variable aleatoria Y , se esperará que el $p \times 100\%$ esté por debajo de $q_p(Y)$ y el $(1 - p) \times 100\%$ esté por arriba. Por ejemplo, la mediana es un caso particular de un cuantil, específicamente el 0.5-ésimo.

La *regresión sobre cuantiles* es una alternativa que se ha desarrollado recientemente y que permite enfocarse a otros aspectos de la distribución, como lo que pasa en las colas. Además relaja supuestos de la *regresión a la media*, como la simetría inducida por el error normal.

En notación probabilística, busca aproximar a la función f , tal que

$$q_p(y|x) = f(x),$$

para $p \in (0, 1)$ fijo arbitrario.

Para hacer esto, normalmente se vale del supuesto que

$$y = f(x) + \varepsilon,$$

con $\varepsilon \in \mathbb{R} \sim E(\theta)$, de manera que E es una variable aleatoria con vector de parámetros θ , tal que $q_p(E) = 0$.

Es importante aclarar que $f(x) \in \mathbb{R}$ y θ son desconocidos. Asimismo, al igual que con la *regresión a la media*, se supone independencia condicional entre observaciones.

3.3.1. Modelo tradicional

Cuando surgió entre la comunidad estadística el problema de *regresión sobre cuantiles*, inicialmente fue modelado bajo un enfoque no bayesiano, como se describe en Yu & Moyeed (2001). Posteriormente, Koenker & Bassett (1978) retomaron esas ideas, y las aplicaron en el paradigma bayesiano.

Al igual que en la *regresión a la media*, el primer y más popular modelo ha sido el lineal. Es decir, para $p \in (0, 1)$ fijo arbitrario, se define

$$f(x) = x^T \beta_p,$$

donde β_p es el vector de coeficientes, dependiente de p .

Definición 3. *Se define a la función*

$$\rho_p(u) = u \times [pI_{(u>0)} - (1-p)I_{(u<0)}].$$

Se dice que una variable aleatoria U sigue una distribución asimétrica de Laplace ($U \sim AL_p(\sigma)$) si su función de densidad se escribe como

$$f_p(u|\sigma) = \frac{p(1-p)}{\sigma} \exp \left[-\rho_p \left(\frac{u}{\sigma} \right) \right],$$

con σ parámetro de escala.

Es posible demostrar que si $E \sim AL_p(\sigma)$, entonces $q_p(E) = 0$. Recordando que esta es la única característica necesaria para la distribución del error aleatoria, entonces se definirá

$$\varepsilon \sim AL_p(\sigma).$$

El modelo se puede reescribir como:

$$y|x, \beta_p, \sigma \sim AL_p(y - x^T \beta_p | \sigma).$$

Sea $\{X, Y | X \in \mathbb{R}^{m \times n}, Y \in \mathbb{R}^m\}$ el conjunto de datos observados. Por el Teorema de Bayes,

$$\mathbb{P}(\beta_p, \sigma | Y, X) \propto \mathbb{P}(Y | X, \beta_p, \sigma) \times \mathbb{P}(\beta_p, \sigma),$$

donde $\mathbb{P}(Y | X, \beta_p, \sigma)$ es la verosimilitud de los datos observados y se puede calcular como

$$\mathbb{P}(Y | X, \beta_p, \sigma) = \prod_{i=1}^m AL_p(y_i - x_i^T \beta_p | \sigma).$$

Por otro lado, $\mathbb{P}(\beta_p, \sigma^2)$ es la distribución inicial de los parámetros, para los que normalmente se usa

$$\beta_p, \sigma \sim \mathcal{NGI}(M, V, a, b).$$

A diferencia de la *regresión a la media*, este modelo no es conjugado y, de hecho, no es posible obtener una distribución posterior analítica. Por lo tanto se requieren métodos computacionales (como los que serán descritos en el capítulo 5) para aproximarla.

En el caso de la predicción, si se tiene una nueva matriz de covariables $X_* \in \mathbb{R}^{r \times n}$, la inferencia con los datos observados se realiza de la siguiente manera:

$$\begin{aligned}\mathbb{P}(Y_*|X_*, Y, X) &= \int \int \mathbb{P}(Y_*|X_*, \beta_p, \sigma) \times \mathbb{P}(\beta_p, \sigma|Y, X) d\sigma d\beta_p \\ &= \int \int \prod_{i=1}^r AL_p(y_i - x_i^T \beta_p | \sigma) \times \mathbb{P}(\beta_p, \sigma|Y, X) d\sigma d\beta_p,\end{aligned}$$

que tampoco tiene solución analítica.

Si bien este modelo representa un gran avance, aún queda la posibilidad de retomar estas ideas y crear modelos más precisos. La intención de esta tesis es encontrar un modelo para la *regresión sobre cuantiles* que sea completamente bayesiano y no paramétrico, con la intención de poder representar distribuciones más complejas.

Bibliografía

- Bannerjee, S. 2008. *Bayesian Linear Models: The Gory Details*. Descargado de <http://www.biostat.umn.edu/ph7440/>.
- Denison, D.G.T. 2002. *Bayesian Methods for Nonlinear Classification and Regression*. Wiley Series in Probability and Statistics. Wiley.
- Dunson, D.B., & Taylor, J.A. 2005. Approximate Bayesian inference for quantiles. *Journal of Nonparametric Statistics*, **17**, 385–400.
- Fishburn, Peter C. 1986. The Axioms of Subjective Probability. *Statistical Science*, **1**(3), 335–345.
- Hanson, T., & Johnson, W.O. 2002. Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*, **97**, 1020–1033.
- Hao, L., & Naiman, D.Q. 2007. *Quantile Regression*. Quantile Regression, no. 149. SAGE Publications.
- Koenker, Roger, & Bassett, Gilbert. 1978. Regression Quantiles. *Econometrica*, **46**(1), 33–50.
- Kottas, A., & Gelfand, A.E. 2001. Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, **96**, 1458–1468.

- Kottas, A., & Krnjajic, M. 2005. *Bayesian Nonparametric Modeling in Quantile Regression*. Technical Report AMS 2005-06. University of California, Santa Cruz.
- Kottas, A., Krnjajic, M., & Taddy, M. 2007. Model-Based Approaches to Nonparametric Bayesian Quantile Regression. *Pages 1137–1148 of: Proceedings of the 2007 Joint Statistical Meetings*.
- Lavine, M. 1995. On an approximate likelihood for quantiles. *Biometrika*, **82**, 220–222.
- Pavlidis, Marios G., & Wellner, Jon A. 2012. Nonparametric estimation of multivariate scale mixtures of uniform densities. *Journal of Multivariate Analysis*, **107**, 71–89.
- Rasmussen, C.E., & Williams, C.K.I. 2006. *Gaussian Processes for Machine Learning*. Adaptive computation and machine learning series. University Press Group Limited.
- Schervish, M.J. 1996. *Theory of Statistics*. Springer Series in Statistics. Springer New York.
- Teh, Yee Whye. 2010. Dirichlet Process. *Pages 280–287 of: Sammut, C, & Webb, GI (eds), Encyclopedia of Machine Learning*. Springer.
- Tsionas, E.G. 2003. Bayesian quantile inference. *Journal of Statistical Computation and Simulation*, **73**, 659–674.
- Walker, S.G., & Mallick, B.K. 1999. A bayesian semiparametric accelerated failure time model. *Biometrics*, **55**(2), 477–483.
- Yu, K., & Moyeed, Rana A. 2001. Bayesian quantile regression. *Statistics & Probability Letters*, **54**(4), 437–447.