Universidad Nacional Autónoma de México

Facultad de Estudios Superiores Acatlán

Método de ajuste de distribuciones extendido

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A :

José Carlos Del Valle López

TUTOR

Act. Hercilio Barragán Anzures

Facultad de Estudios Superiores AcatlÃąn, Cd. Mx., 2019

Agradecimientos

A mi mamá María del Carmen López Sansalvador, por ser la mujer más sobrealiente, excelsa, fuerte y maravillosa de este mundo, que me ha apollado durante toda mi vida, dado todo el amor, paciencia, cariño y la mejor herencia, mi carrera universitaria.

A mi familia, con quienes he contado incondicionalmente y han estado siempre pendientes de mi.

A mi escuela, la UNAM que me ha dado la más grande oportunidad de mi vida, a los mejores amigos y las mejores experiencias, que me llena de orgullo y que honraré y agradeceré eternamente.

A mis presentes maestros, Hercilio Barragán Anzures, Miguel Ángel Chávez, Gustavo Fuentes Cabrera que no solo me dieron una enseñanza, sino nuevas formas de ver la vida.

A mis antiguos maestros Leonardo Rebollo Pantoja, Fernando Muñoz Razo, Ruíz Murillo Pablo y Rodrigo Gamez Manzo, gracias los ejemplos y aprenizaje que me brindaron, por su pasión por enseñar, desempeño como docentes, como profesionales y como personas.

A mis amigos, Luis Orozco Córdoba, Jesús González Moreno, Alan Sanchez, Ruth Lara Castelán y a todos con quienes siempre he contado, en los buenos y malos momentos.

Índice general

Índice de figuras

Introducción

Divide et impera

-Julio Cesar, Maquiavelo

En 1333 el imperio X sería atacado por tres naciones enemigas, siendo superados 5 a 1, el consejero del emperador solicitó una audiencia para establecer una respuesta ante tal amenaza, al siguiente día liberó a los esclavos y soldados de la primera nación, repitiendo esto durante siete días, diciéndoles que llevaran el mensaje "Su deuda ha sido saldada", la siguiente semana liberó a los esclavos de la segunda nación llevando oculto un soldado entre sus filas cuya etnia de origen era una de las naciones enemigas.

Al llegar los primeros a su ciudad natal las dos naciones aliadas lo notaron, y después de la última caravana fueron acusados de traición y atacados.

La tensión aumentó entre los aliados, el soldado oculto apuñaló por la espalda a uno de los soldados de la nación aliada, gritando el nombre del rey, en el acto, los soldados arremetieron contra los esclavos no quedando ni uno solo.

Sus fuerzas se habían diezmado y después de haberse enfrentado entre ellos, al ejército imperial le tomó poco tiempo sobreponerse contra sus enemigos.

Divide y vencerás en la guerra, en la ciencia y en la vida, implica resolver un problema complejo separándolo en componentes más sencillas tantas veces como sea necesario, por ejemplo:

Aplicamos la descomposición para un análisis de series de tiempo, pues en el proceso separamos la tendencia, la estacionalidad y el ruido blanco, a cada parte aplicamos pruebas distintas que nos dan un mayor entendimiento de lo que sucede a través del tiempo.

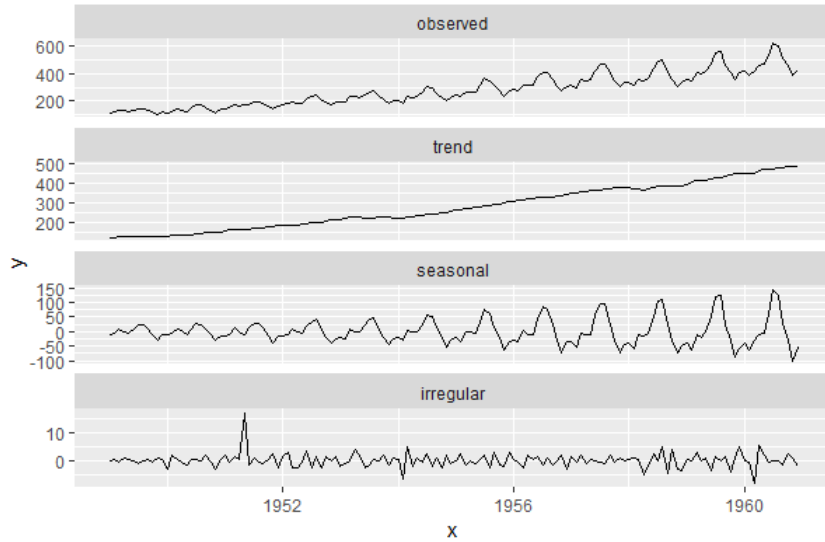

Figura 1: Ejemplo de una serie de tiempo en cada una de sus componentes (Fuente data set AirPassengers)

En el análisis multivariado y ciencias computacionales, aplicamos algoritmos de clasificación como Máquina Vector Soporte, K-Medias, entre otros para entender que partes influyen más a una variable objetivo o para perfilar a nuestra población.

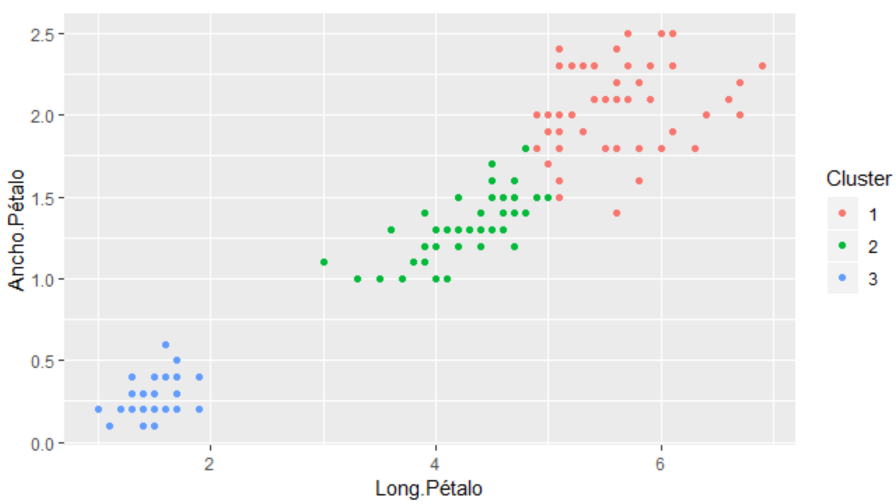

Figura 2: Ejemplo del algoritmo K-Medias en el las medidas del tallo y pétalo de las flores (Fuente data set Iris)

En la ingeiería empleamos Series de Fourier para analizar una onda de sonido por medio de su descomposición en ondas más simples, dándonos un espectro más amplio de lo que sucede. Ante esta idea, no es difícil darse cuenta de que puede realizarse lo mismo con una función de distribución.

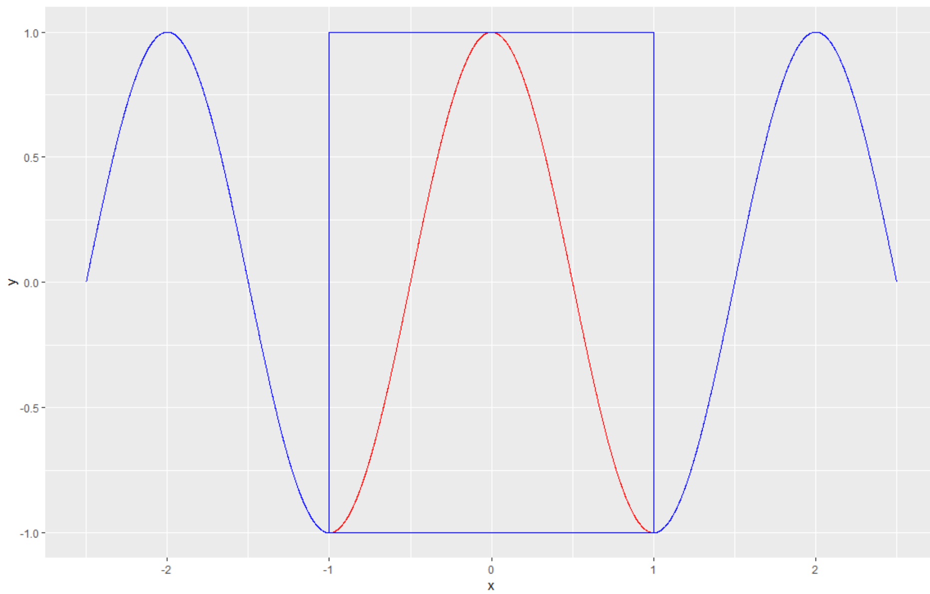

Figura 3: Ejemplo de descomposición de una función de distribución

Si observamos cuidadosamente, al aislar una sección de onda veremos la forma de una función de distribución desplazada en el plano, esa fue la idea inicial por la que nació este escrito, empezando por conocer métodos de descomposición de Series de Fuorier y emplearlos para descomponer una distribución en componentes más simples, buscando formas óptimas de separar y añadiendo el factor aleatoriedad.

Hallar los puntos de acumulción más evidentes y extender un radio a través de estos, si un punto se encontraba en ambos radios, se designaba aleatoriamente a la otra parte de la distribución.

Algunos ejemplos en donde observamos este tipo de comportamientos:

Medicina: Distribución de pacientes con...

Seguros: Si tratamos a los outlayes como parte de otra distribución obtenemos...

Economía: La distribución de los inversores en X acción...

Física: La cantidad de luz en el espectro...

Al pensarlo con detenimiento, contamos con las herramientas para realizar este análisis y es también una respuesta ante la diferencia que existe entre los datos observados y el ajuste a un modelo paramétrico, puesto que en la práctica procedemos si estos fallan, procedemos a calcular estimadores con modelos no paraetricos o a trabajar bajo hipótesis que no se sostienen.

Esta nueva visión es una propuesta al análisis de outlayers, a la separación en panza y cola de una función de densidad o un nuevo camino si cierta distribución no se ajusta a nuestros datos.

Capítulo 1

Descipción de variables aleatorias

1.1. Introducción

Para entender el concepto de variable aleatoria, introduciremos primero los conceptos de experimento aleatorio y un espacio muestral.

Consideremos un experimento cuyo resultado depende completamente del azar, es decir, es desconocido, a este suceso le llamaremos experimento aleatorio, por ejemplo:

1. El resultado de lanzamiento de un dado.
2. El resultado de una canica al girar en la ruleta.
3. El resultado del lanzamiento de una moneda.

Y un espacio muestal es el conjunto de todos los posibles resultados de un experimento aleatorio, siendo los de nuestros ejemplos:

1. S = {1,2,3,4,5,6}
2. S = {1,2,...,38}
3. S = {"Águila","Sol"}

Al realizar un experimento, nos interesamos en los resultados obtenidos al hacerlo una y otra vez, por ejemplo:

1. La suma del lanzamiento de dos dados.

2. El número de veces que cae en la casilla 38.

3. La cantidad de Águilas después de n lanzamientos.

Estas cantidades o números de interés deteminados por un experimento aleatorio son a las que denominamos variables aleatorias.

1.2. Distribuciones Discretas

Decimos que una variable aleatoria X tiene distribución discreta si el rango de X es numerable, es decir, el epacio muestal contiene una cantidad contable de elementos. En la mayoria de los casos dicho rango corresponde a $\mathbb{N}U\{0\}$.

Definimos la función de masa de probabilidades de la variable aleatoria X como: $f(a) = P(X = x)$ es decir, la probabilidad de que X tome el valor a.

Esta función para el espacio muestral $S = \{x_1, x_2, ...\}$ debe cumplir con:

1.-

$$f(x_i) \geq 0$$

$$\forall i \in S$$

2.-

$$\sum_{i=1}^{\infty} f(x_i) = 1$$

Gráficamente puede verse de la siguiente forma:

$$P(X = 1) = .19, P(X = 2) = .23, P(X = 3) = .31, P(X = 4) = .27$$

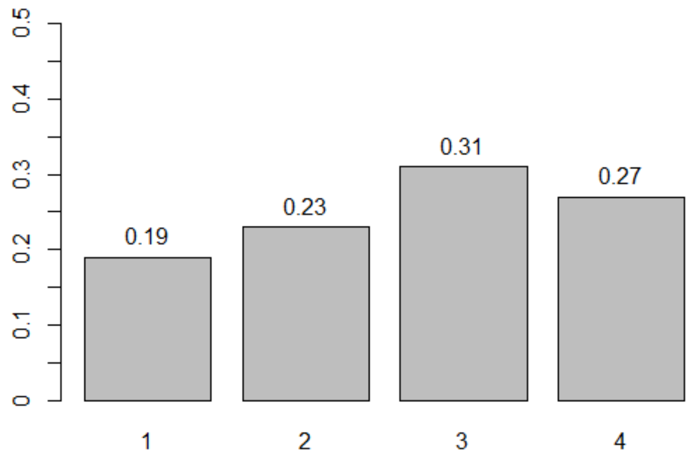

Figura 1.1: Ejemplo Función de Probabilidad Discreta

A continuación, describiremos una de estas funciones

Distribución Poisson.

Sea X una variable aleatoria, decimos que X se distribuye Poisson o:

$X \sim Poisson(\lambda)$ con parámetro $\lambda > 0$

Si su función de densidad se define como:

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

para $x = 0, 1, 2, ...$

Algunas de las características importantes de esta son:

La esperanza o valor esperado de una variable aleatoria se define como:

$$E(X) = \sum_{i=1}^{n} x_i f(x_i)$$

Para $X \sim Poisson(\lambda)$

$$E(X) = \sum_{i=1}^{n} x_i \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \sum_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{(x_i-1)!} = \lambda \sum_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i-1}}{(x_i-1)!}$$

Si realizamos el cambio de variable $z_i = x_i - 1$ tenemos que la parte

$$\sum_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i-1}}{(x_i-1)!} = \sum_{i=1}^{n} \frac{e^{-\lambda}\lambda^{z_i}}{(z_i)!} = 1$$

por lo tanto

$$E(X) = \lambda$$

Varianza:

La varianza puede definirse como:

$$V(X) = E(X^2) - E^2(X)$$

Dado que ya conocemos el valor de $E(X) = \lambda$, la icógnita reside en $E(X^2)$, y empleando la fórmula para la esperanza

$$E(X^2) = E(X^2 - X + X) = E((X(X-1) + X)) = E(X(X-1)) + E(X)$$

$$E(X(X-1)) = \sum_{i=1}^{n} x_i(x_i-1)\frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \lambda^2 \sum_{i=1}^{n} \frac{e^{-\lambda}\lambda^{(x_i-2)}}{(x_i-2)!}$$

Aplicamos nuevamente un cambio de variable con $z_i = x_i - 2$ y llegaremos a que

$$E(X(X-1)) = \lambda^2$$

entonces,

$$E(X^2) = \lambda^2 + \lambda$$

por lo tanto

$$V(X) = \lambda^2 + \lambda \ - \lambda^2 = \lambda$$

Función generadora de momentos:

La función generadora de momentos o función generatriz de momentos de una variable aleatoria X es

$$M_x(t) = E(e^{tX}), t \in \mathbb{R}$$

Para este caso

$$M_x(t) = \sum_{i=1}^{n} e^{tx_i} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = e^{-\lambda} \sum_{i=1}^{n} \frac{(\lambda e^t)^{x_i}}{x_i!}$$

Si recordamos el desarrollo en serie de la función exponencial, veremos que

$$e^t = \sum_{x=0}^{\infty} \frac{t^x}{x!}, \forall t \in \mathbb{R}$$

En este punto, hay que recalcar que para la esperanza

$$\sum_{x=0}^{\infty} x_i f(x_i) = \sum_{x=0}^{n} x_i f(x_i) + \sum_{x=n+1}^{\infty} x_i f(x_i)$$

en dónde

$$\sum_{x=n+1}^{\infty} x_i f(x_i) = 0$$

dado que todos ls valores de $f(x_i) = 0$ para valores que estén fuera del espacio muestral, entonces, al utilizar la forma en serie de la función exponencial llegamos a

$$M_x(t) = e^{-\lambda} e^{\lambda e^t} = exp(\lambda(e^t - 1))$$

Visualización de una variable aleatoria $X \sim Poisson(1)$

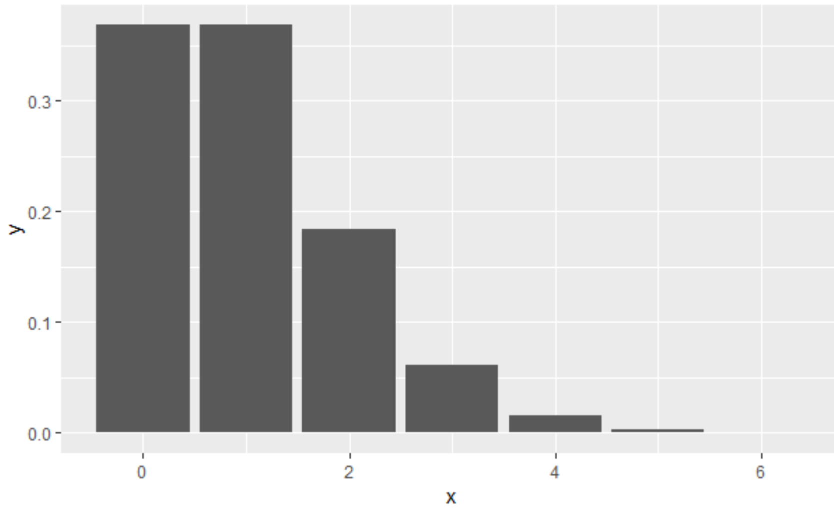

Figura 1.2: Poisson Distribution

Ejemplo de código en R para la función generadora de números aleatorios, función de distribución y función de densidad respectivamente:

```
>set.seed(31109)
>rpois(n = 1,lambda = 2)
[1] 2

>ppois(q = 1,lambda = 1)
[1] 0.7357589

>dpois(x = 1,lambda = 3)
[1] 0.1493612
```

Ejemplificaremos entonces el uso de esta distribución:

Una empresa observa el número de clientes que requieren de sus servicios durante 10 horas, el número de entradas siguie una distribución Poisson y se sabe que el promedio de clientes que reciben es de 15

¿Cuál es la probabilidad de que lleguen 20 clientes?

¿Cuál es la probabilidad de que no lleguen clientes en la primera hora?

¿Cuál es la función de describe que llegue al menos un cliente en cualquier cantidad de horas?

Primero debemos observar que el valor λ corresponde tanto a la esperanza como a la varianza de la distribución, entonces

$$f(x) = \frac{e^{-15} 15^x}{x!}$$

Ahora ya somos capaces de responder la primera pregunta:

$$f(20) = \frac{e^{-15} 15^{20}}{20!} = 0.04181$$

Supongamos que los clientes llegan de forma regular durante esas 9 horas, es decir, que esperamos una media de 5 clientes al cabo de 3 horas, por lo que deberemos tratar con una distribución con parámetro $\lambda_k = 5$, donde k es el número de clientes que habrán llegado a las 3 horas, entonces, nuestra función de densidad toma la siguiente forma:

$$f(x) = \frac{e^{-5} 5^x}{x!}$$

entonces, la probabilidad de que no lleguen clientes es

$$f(0) = \frac{e^{-5} 5^0}{0!} = e^{-5} = 0.00674$$

La última pregunta supone el siguiente plantemiento $P(X > 0)$, es decir, que llegue al menos un cliente, sin embargo, podemos plantearlo como $1 - P(X = 0)$, es decir, que tome todos los valores exceptuando el 0, y sustituyendo, obtenemos:

$$1 - f(0) = 1 - \frac{e^{-\lambda} \lambda^0}{0!} = 1 - e^{-\lambda}$$

Y entonces, la función que buscamos, con $\lambda =$ número de horas es:

$$f(\lambda) = 1 - e^{-\lambda}$$

1.3. Distribuciones Continuas

Estas variables aleatorias tienen que ser definidas de forma distinta, pues no es posible calcular la probabilidad puntual de la variable, es decir, $P(X = a)$, no obstante, es posible definir la probabilidad acumulada hasta cierto valor, es decir, $P(X \leq a)$ la probabilidad de que el valor X sea menor a a.

Distribución Exponencial.

Sea X una variable aleatoria, decimos que X se distribuye Exponencial o:

$X \sim exp(\lambda)$ con parámetro $\lambda > 0$

Si su función de distribución acumulada se define como:

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & \text{si } x \geq 0 \\ 0, & \text{e.o.c} \end{cases}$$

es decir, su función de densidad se define como:

$$f(x) = P(X = x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0 \\ 0, & \text{e.o.c} \end{cases}$$

Esperanza:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Que para este ejemplo resulta:

$$E(X) = \int_{-\infty}^{0} x * 0 dx + \int_{0}^{\infty} x \lambda e^{-\lambda x} dx = \int_{0}^{\infty} x \lambda e^{-\lambda x} dx$$

$$= [-x e^{-\lambda x}]_0^\infty + \int_0^\infty \lambda e^{-\lambda x} dx = 0 + [\frac{-1}{\lambda} e^{-\lambda x}]_0^\infty = 0 + \frac{1}{\lambda}$$

por lo tanto

$$E(X) = \frac{1}{\lambda}$$

Varianza:

La definición de la varianza en función de la esperanza es análoga a la de las variables discretas.

$$V(X) = E(X^2) - E^2(X)$$

Por lo que deberemos hallar el valor de $E(X^2)$ que por definición es:

$$E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = [-x^2 e^{-\lambda x}]_0^\infty + \int_0^\infty 2xe^{-\lambda x} dx$$
$$= 0 + \frac{2}{\lambda} \int_0^\infty xe^{-\lambda x} dx = \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2}$$

entonces

$$V(X) = E(X^2) - E^2(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2}$$

por lo tanto

$$V(X) = \frac{1}{\lambda^2}$$

Función generadora de momentos:

La definición se conserva, es decir:

$$M_x(t) = E(e^{tX}), t \in \mathbb{R}$$

Para este caso

$$M_x(t) = \int_0^\infty exp(tx) \lambda e^{-\lambda x} dx = \lambda \int_0^\infty exp((t - \lambda)x) dx$$
$$\lambda [\frac{exp((t - \lambda)x)}{t - \lambda}]_0^\infty = \frac{\lambda}{\lambda - t}$$

Hay que destacar que para que el valor de la ecuación se haga 0 cuando x tienda a infinito, se debe cumplir que $t - \lambda < 0$, de esa forma el factor al evaluar en infinito se convierte en 0

A continuación podemos ver una representación gráfica de una variable aleatoria

$X \sim exp(1)$

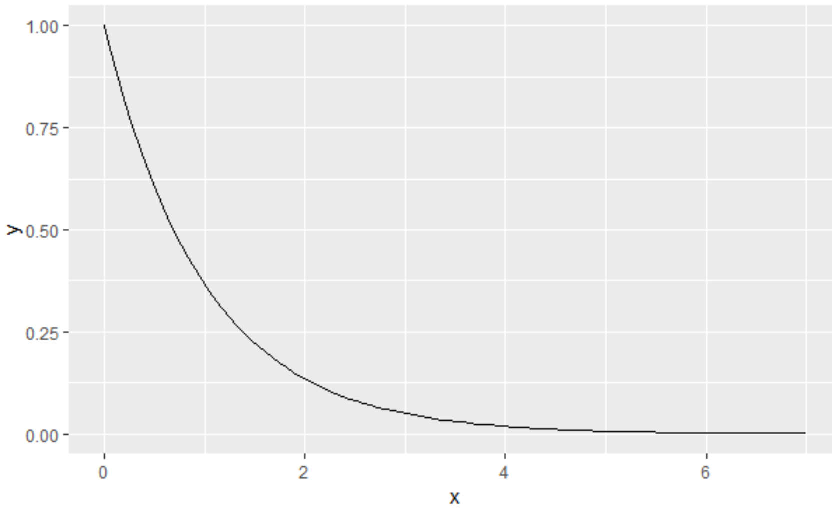

Figura 1.3: Distribucion Exponencial

Algunas aplicaciones para esta distribución son: El tiempo que tarda una proceso de producción en crear un objeto es de 9 minutos, supongamos que este sigue una distribución exponencial.

¿Cuál es la probabilidad de que la máquina tarde menos de 7 minutos?

¿Cuál es la probabilidad de que la máquina tarde más de 15 minutos?

Aquí también observamos que el parámetro $1/\lambda$ corresponde con la media de la distribución, es decir, la función de distribución acumulada se define como:

$$F(x) = \begin{cases} 1 - e^{-x/9}, & \text{si } x \geq 0 \\ 0, & \text{e.o.c} \end{cases}$$

Ahora podemos obtener las repuestas que buscábamos:

$$F(7) = \begin{cases} 1 - e^{-7/9}, & \text{si } x \geq 0 \\ 0, & \text{e.o.c} \end{cases} = 0.5405742$$

Y análogamente:

$$1 - F(15) = \begin{cases} e^{-15/9}, & \text{si } x \geq 0 \\ 0, & \text{e.o.c} \end{cases} = 0.1888756$$

1.4. Distribuciones Bimodales

Ya sentadas las bases, podemos entrar en un terreno de estudio más escabroso.

En la naturaleza es dificil hallar objetos de estudio o comportamientos que se ajusten a una vaibale aleatoria si no se trata de un experimento controlado, es frecuente que nos lleguemos a encontrar con comportamientos más atípicos a los que pareciera que no se le puede ajustaruna distribución pero que sin embargo es necesario hacerlo.

Para estos casos, se suelen atender separando la distribución de los outlaiers que impiden su ajuste y posteriormente añadirlos al análisis, proceo que no siempre es posible debido a la forma de la distribución.

Para ello, son introducidas entoces las distribuciones bimodales, el término bimodal proviende del sufijo bi que significa dos y modal que significa moda, y son aquellas distribuciones que tienen dos modas y que ejemplificamos a continuación:

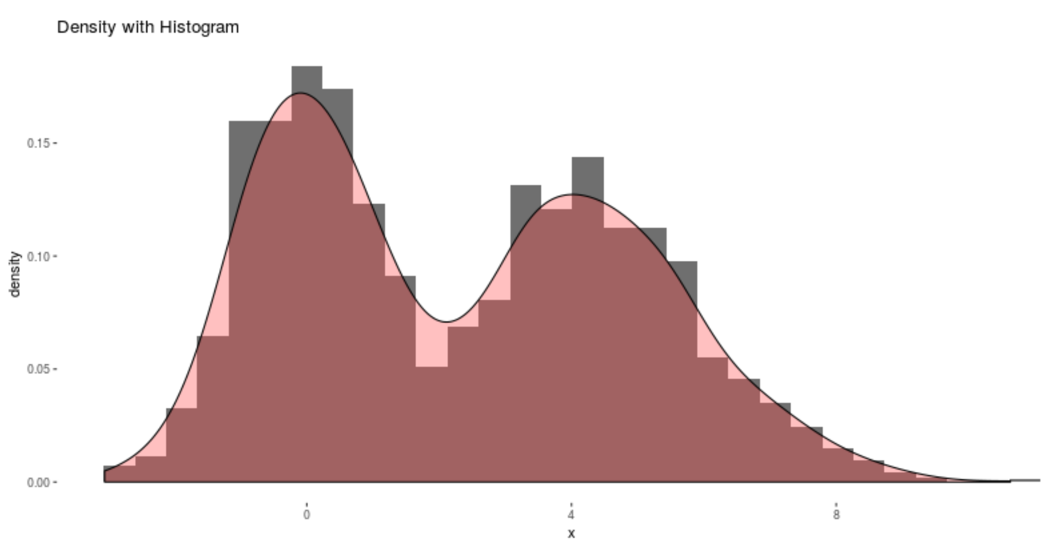

Figura 1.4: Distribución Multimodal

1.4.1. Confirmar bimodalidad

Una distribución de este estilo proveé de datos importantes importantes para la distribución, como que la media no es un parámetro de máxima verosimilitud, que la muestra de datos no es homogénea, las observaciones pueden venir de dos distribuciones empalmadas o que puede haber un error en los instrumentos de medición.

Para el ejemplo generado se utilizaron las siguientes distribuciones empalmadas:

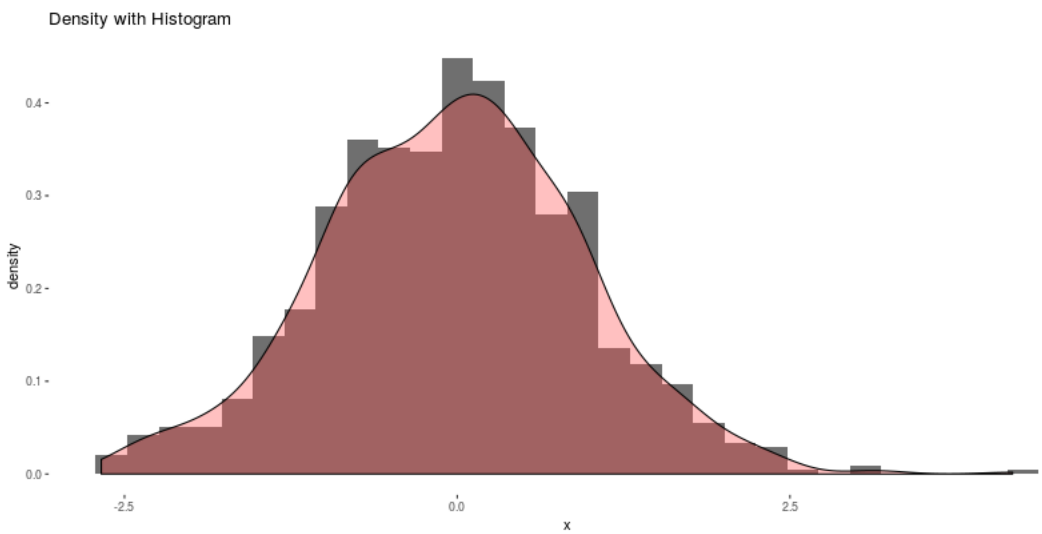

Figura 1.5: Normal(0,1) Distribution

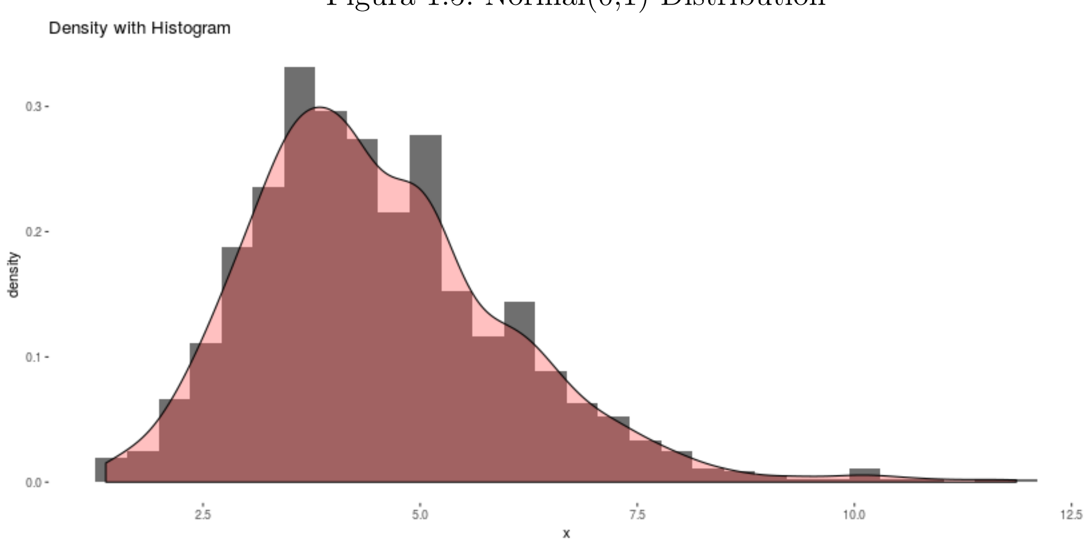

Figura 1.6: Gamma(9,2) Distribution

1.4.2. Distribución Beta-Normal

Mixture Normal Distribution

Esta distribución se caracteriza por tener cuatro parámetros que juntos describen la localización, la escala y la forma

Capítulo 2

Estimación y Ajuste

2.1. Introducción

Hay muchas formas de aproximar una variable aleatoria a obsevaciones que vemos en la naturaleza, muestras de características humanas y en general, conjuntos de datos que puedan ser medidos o contados, en esta sección, exploraremos los enfoques clásicos que nos han ayudado a tener una mayor certidumbre.

Para ello dependeremos de dos conceptos importantes: Hipótesis Nula y p-valor.

Una hipótesis es una afirmación expuesta a ser o no rechazada acerca de una característica de nuestra población, por ejemplo, que la media muestral es igual a 0 lo cual se denota como $H_0 : \mu = 0$

2.2. Estimación de parámetros

Antes de pensar en asignar una distribución a nuestros datos, es necesario conocer el valor de ciertos estadísticos de nuestra muestra que funjan como parámetros para la distribución que deseamos ajustar.

2.2.1. Máxima Verosimilitud

La estimación por máxima verosimilitud (EMV) es un método para estimar los parámetros de una muestra observada y ajustarlos a una funcion de probabilidad

Sea $(x_1, ..., x_n)$ un vector aleatorio cuya distribución depende del parámetro desconocido θ

Definición:

La función de verosimilitud del vector $(x_1, ..., x_n)$ es

$$L(\theta) = f_{X_1,...,X_n}(x_1, ..., x_n; \theta)$$

si $X_1, ..., X_n$ son independientes, entonces

$$L(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i; \theta)$$

y si son identicamente distribuidas

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

El objetivo de este método es encontrar el valor de θ que maximice la función de verosimilitud, que representa la distribución conjunta de nuestro vector aleatorio, a este valor se le llama ëstimador de máxima verosimitituдÿ será el valor que le habremos de asignar al parámetro θ

Recordemos que para encontrar el máximo de una función se debe hallar la derivada de esa función e igualarla a 0, también es válido aplicar transformaciones a la función siempre y cuando estas sean crecientes no afectando el valor máximo de la función, entonces podemos buscar los siguientes resultados:

$$\frac{\partial(L(\theta))}{\partial\theta} = \frac{\partial(\prod_{i=1}^{n} f(x_i; \theta))}{\partial\theta} = 0$$

De igual forma

$$\frac{\partial(log(\prod_{i=1}^{n} f(x_i;\theta)))}{\partial\theta} = \frac{\partial(\sum_{i=1}^{n} log(f(x_i;\theta)))}{\partial\theta} = 0$$

A esta función también se le llama LogVerosimilitud.

Veamos entonces el siguiente ejemplo:

Sea $(x_1, ..., x_n)$ una m.a proveniente de una distribución $X \sim N(\mu, \sigma)$, independientes e identicamente distribuidas, ¿cuál es el EMV para el parámetro μ?

La función de densidad de una distribución normal es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

entonces,

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2}$$

Aplicaremos la derivada a la función de LogVerosimilitud para obtener el estadítico, obteniendo.

$$L(\mu) = (\frac{1}{\sigma\sqrt{2\pi}})^n exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2)$$

entonces,

$$log(L(\mu)) = -nlog(\sigma\sqrt{2\pi}) - (\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2)$$

Ahora derivemos respecto al parámetro que deseamos, en este caso μ e igualemos a 0

$$\frac{\partial log(L(\mu)))}{\partial\theta} = 0$$

$$log(L(\mu)) = -(\frac{1}{2\sigma^2} \sum_{i=1}^{n} -2\,(x_i - \mu)) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

entonces,

$$\sum_{i=1}^{n}(x_i-\mu) = -n\mu + \sum_{i=1}^{n}(x_i) = 0$$

que sicede si y sólo si,

$$\mu = \frac{\sum_{i=1}^{n}(x_i)}{n} = \bar{X}$$

Por lo tanto, el EMV para el parámetro μ es la media de la distribución, que gráficamente se puede ver de la siguiente forma.

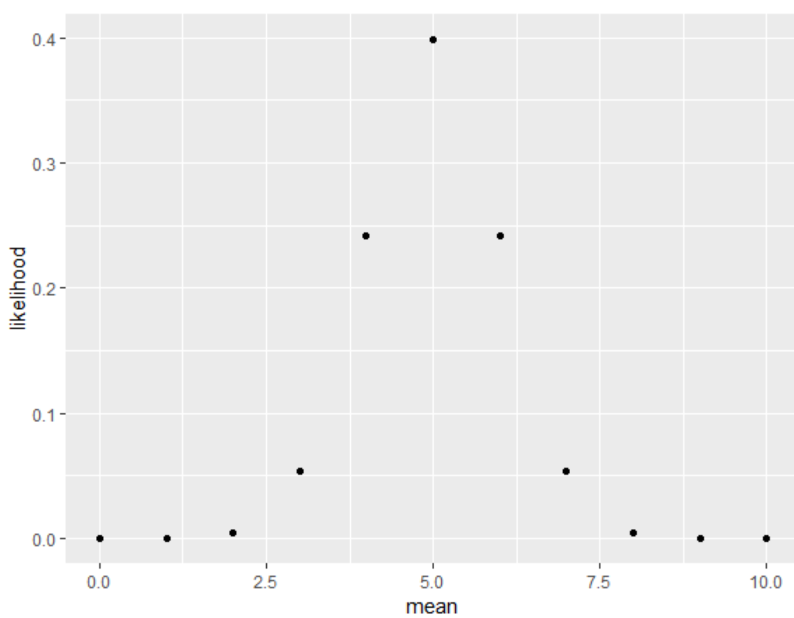

Figura 2.1: Estimador de máxima verosimilitud para una distribución Normal(0,1)

2.2.2. Método de momentos

Si bien el métoo de Máxima Verosimilitud nos ofrece certidumbre sobre el posible valor del parámetro desconocido θ, tambien existe una forma más intuitiva por medio de la función generadora de momentos de una función de densidad.

Para abordar este método, es necesario recordar dos definiciones importantes,

Sea X una v.a y sea k un entero mayor que 0. El k-ésimo momento de x, si existe, es el número obtenido de $E(X^k)$, también se le llama momento poblacional.

Sea $x_1, ..., x_n$ una m.a. de la distribución $f(x;\theta)$ y sea k un entero mayor que 0, definimos al k-ésimo momento muestral como la variable aleatoria.

$$\frac{\sum_{i=1}^{n}\left(x_i^k\right)}{n}$$

Este método consiste en igualar los momentos muestrales y los poblacionales y resolver el sistema de ecuaciones generado para los parámetros que se desean obtener, igualando tantas ecuaciones como número de parámetros a calcular.

$$E(X^k) = \frac{\sum_{i=1}^{n}\left(x_i^k\right)}{n}$$

Veamos ahora un ejemplo:

Sea X una v.a. con parámetro desconocido $\theta > 0$ y función de densidad

$$f(x) = \begin{cases} \theta x^{\theta-1}, & \text{si } x \in (0,1) \\ 0, & \text{e.o.c} \end{cases}$$

Dado que sólo deseamos obtener un parámetro, será necesario obtener el primer momento y resolver la siguiente ecuación para $k = 1$.

$$E(X) = \frac{\sum_{i=1}^{n}\left(x_i\right)}{n}$$

Es sencillo ver que

$$E(X) = \int_0^1 x\theta x^{\theta-1} = \theta[\frac{x^{\theta+1}}{\theta+1}]_0^1 = \frac{\theta}{\theta+1}$$

entonces, $\frac{\widehat{\theta}}{\widehat{\theta}+1} = \bar{X}$ y depejando el parámetro desconocido $\widehat{\theta}$, llegamos a

$$\widehat{\theta} = \frac{\bar{X}}{\bar{X}+1}$$

2.3. Pruebas de Bondad y Ajuste

Ya que sabemos maneras de estimar los parámetros para una distribución que querramos ajustar, es necesario aplicar pruebas de bondad y ajuste para corroborar la certidumbre de nuestra distribución, estas pruebas describe que tan bien se ajusta una muestra observada respecto a un modelo teórico, para ello deberemos emplear contraste de hipótesis.

2.3.1. Pruebas de hipótesis

Una hipótesis es una proposición que se desea aceptar o rechazar con base en observaciones reales. Es imperativo remarcar que una hipótesis está en constante vrificación, por lo que no se puede estar completamente convencido de que esta es acepada dada la naturaleza aleatoria de nuestros datos. A continuación recapitularemos brevemente las pruebas de hipótesis.

Hipótesis nula

Ésta se denota como H_0, es la proposición que se desea rechazar, en la que se declara un valor y se contrasta con un parámetro o estimador de nuestra muestra observada, por ejemplo:

$$H_0 : \mu = \mu_0$$

$$H_0 : \mu \geq \mu_0$$

Hipotesis Alternativa

La Hipótesis Alternativa, se denota como H_1, esta se puede verificar en base a la evidencia de la muestra y su región debe abarcar el complemento de nuestra hiótesis nula siendo:

$$H_0 : \mu \neq \mu_0$$

$$H_0 : \mu < \mu_0$$

Las hipótesis alternativas del anterior ejemplo respectivamente

Tipos de error

Podemos encontrar cuatro posibles situaciones dados los resultados que arrojan las hipótesis representadas en el siguiente cuadro:

	H_0 Verdadera	H_0 Falsa
Rechazamos H_0	Error Tipo I P(ET I) $= \alpha$	Decisión Correcta
No Rechazamos H_0	Decisión Correcta	Error Tipo II P(ET II) $= \beta$

La Probabilidad de cometer un Error Tipo I se conoce como Nivel de Significancia, se denota como α y es el tamaño de la región de rechazo, este corresponde al conjunto de valores tales que si la prueba estadística cae dentro de este rango, decidimos rechazar la Hipótesis Nula

2.3.2. Kolmogorov Smirnov

Este test nos ayudará a ver las diferencias entre dos distribuciones de probabilidad distintas para determinar si tienen o no la misma distribución planteando las siguientes hipótesis:

$$H_0 : F_X(x) = F_Y(x), \forall x \in R$$

$$H_1 : F_X(x) \neq F_Y(x)$$

En este caso, nos interesa no rechazar la hipótesis nula, es decir, que el p.valor sea mayor a α, al que le asignaremos el valor de 0.05

Primero necesitamos definir la distribución empírica Sea $x_1, ..., x_n$ una m.a., la distribución empírica se define como

$$F_e(x) = \frac{\#\{i|X_i \leq x\}}{n}$$

Es decir, la proporcion de valores observados menores o iguales a x

Por ejemplo, supongamos que tengo los datos observdos $x_1 = 3, x_2 = 5, x_3 = 1$, estos los ordeno de menor a mayor $x_{(1)} = 1, x_{(2)} = 3, x_{(3)} = 5$, entonces

$$F_e(x) = \begin{cases} 0, & \text{si } x < 1 \\ 1/3, & \text{si } 1 \leq x < 3 \\ 2/3, & \text{si } 3 \leq x < 5 \\ 1, & \text{e.o.c} \end{cases}$$

El estadístico de Kolmogorov Smirnov se define para toda $x \in R$ como

$$D = max(F_e(x) - F(x))$$

para distribuciones discretas y para contínuas como:

$$D = sup(F_e(x) - F(x))$$

Esta diferencia se ve representada a continuación

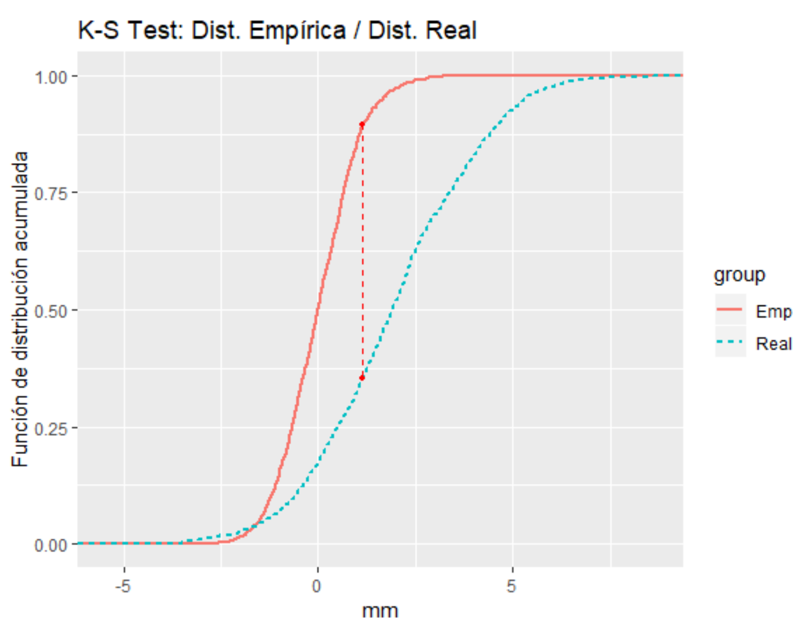

Figura 2.2: Estadístico de KS para dos distribuciones N(0,1) y N(2,2)

Podemos considear las diferencias que se encuentran por encima o por debajo de la distribuciín, denotándolas como:

$$D^+ = sup(F_e(x) - F(x))$$

$$D^- = sup(F(x) - F_e(x))$$

y con ello redefinir el estadístico D

$$D = max\{\frac{j}{n} - F(x_{(j)}), F(x_{(j)}) - \frac{j-1}{n}\}$$

Posteriormente deberemos revisar que el valor D para calcular el p valor respectivo de la siguiente manera:

$$p = P_F(D \geq d)$$

Esto significa que el p valor dependerá de la distribución de D y de F que se esté planteando.

Veamos entonces el siguiente ejemplo:

Sea X una muestra aleatoria con los siguientes valores (esta muestra será utilizada para ejercicios posteriores y ya se encuentra ordenada para facilitar los cálculos)

X=(-2.736, -2.445, -1.816, -1.607, -1.358, -1.087, -0.853, -0.832, -0.818, -0.721, -0.613, -0.567, -0.439, -0.402, -0.258, -0.217, -0.16, -0.105, -0.006, 0.133, 0.242, 0.308, 0.33, 0.362, 0.683, 0.791, 0.794, 1.209, 1.284, 1.339)

La obetnción de dicha muestra se realizó mediante el código:

```
set.seed(31109); rnorm(30,0,1)
```

¿Proviene dicha muestra de una distribución Normal(0,1)?

Para este ejemplo el hecho de tener un primer valor muy bajo y al tamaño de la muestra, esto afectará el resultado final, sin embargo procederemos con los calculos.

Muestra Diferences	F. AC	i/N	$\hat{Fn}(X)$
-2.736 0.030	1	0.033	0.003113581
-2.445 0.059	2	0.067	0.007233317
-1.816 0.065	3	0.100	0.034672841

Y la máxima diferencia D es:

24	0.800	0.641	0.159

Para finalizar, ...

Por lo tanto, no rechazamos la hipótesis nula, y la muestra pertenece a una distribución normal con un nivel de confianza del 95

2.3.3. Anderson-Darling

La prueba de Anderson-Darling es otra forma de contrastar valoores observados respecto a una distribución y que comparte las hipótesis de la prueba de Kolmogorov Smirnov:

$$H_0 : F_X(x) = F_Y(x), \forall x \in R$$

$$H_1 : F_X(x) \neq F_Y(x)$$

Es decir, nos ayudará a determinar si una muestra observada proviene de cierta distribución de tamaño N.

El estadístico de prueba se define como:

$A^2 = -N - S$

Donde N es el tamaño de muestra y S:

$$S = \sum_{i=1}^{N} \frac{(2i-1)}{N}[ln(F(Y_i)) + ln(1 - F(Y_{N+1i}))]$$

Y F es la función de distribución acumulada y las Y_i se encuantran ordenadas ascendentemente.

Utilizando la muestra aleatoria del ejercicio anterior, obtendremos los siguientes resultados, mostraremos los 3 primeros renglones:

Dado que la segunda componenete de S corresponde a $N+1-i$, los cálculos serán análogos a colocar los valores de forma descendente, obteniendo entonces:

Los estadísticos resultantes son los siguientes:

$$S = 43.91795$$

$$A^2 = -13.91795338$$

Finalmente, para la distribución Normal contamos con el estadístico de prueba ajustado

$$(1 + \frac{4}{N} - \frac{25}{N^2})A_N^2 = 1.519943$$

Si vemos los niveles de confianza del estadístico para una distribución Normal(0,1), veremos:

Con un nivel de confianza de 1 %, obtenemos

$$(1 + \frac{4}{N} - \frac{25}{N^2})A_N^2 = 1.029$$

Por lo tanto, dado que nuestro estadístico supera dicho valor, tenemos una certeza de al menos 99 % de que nuestros datos provengann de una distribución Normal(0,1).

2.3.4. Aplicación de transformaciones

Otra forma de extender nuestro universo de distribuciones que podemos ajustar es agregando parámetros de escala y desplazamiento a aquellas distribuciones que están limitadas por su función indicadora y aplicando la función inversa de esos parámetros al momento de simular, es deci, aplicarles una transformación lineal $a * X + b$.

Un ejemplo de una de distribución que al aplicarle una transformaión lineal conserva el kernel es la Couchy, pues solo son modificados sus parámetros, lo que nos indica que estos representan desplazamiento y escala.

Por ejemplo, sea X una v.a. contínua con distribución Cauchy(1,0)

¿Como se distribuirá $a * X + b$?

Si $X \sim Cauchy$, entonces su función de densidad está dada por

$$f(x) = (\pi\gamma[1 + (\frac{x - x_0}{\gamma})^2])^{-1}$$

y al sustituir

$$f(x) = (\pi[1 + (x)^2])^{-1}$$

Para aplicar la transformación lineal, deberemos hacer uso de la técnica de la transformación, primero definimos $Y = g(X) = aX + b$, esta transformación al ser lineal, no afecta el rango en el que se encontraba definida, e decir, Y puede tomar valores en R. definimos $X = g^{-1}(X) = \frac{Y-b}{a}$ y el jacobiano como $\frac{\partial X}{\partial Y} = \frac{1}{a}$

Entonces, la funcion de densidad de Y es:

$$f_Y(y) = f_X(g^{-1}(X))|\frac{\partial X}{\partial Y}| = (\pi a[1 + (\frac{Y - b}{a})^2])^{-1}$$

Que sigue una distribución Cauchy con parámetros a y b, de ecala y posición respectivamente

No siempre es posible tener esta peculiaridad con distribuciones que tengan un solo parámetro, pensemos, por ejemplo en una institución financiera recibe ingresos debido a un producto financiero que venden, sin embargo, hay ocasiones en las que dicho producto falla y el dindero se pierde, la forma en la que se gana y se pierde el dinero es igual a una distribución exponencial como se muestra a continuación:

Figura 2.3: Distribución exponencial fuera de su rango usual

Es importante considerar nuevos parámetros de los que podamos disponer, por lo que se aconseja realizar las siguientes transformaciones a partir de la función indicadora que acompañe a la distribución.

Si $x > 0$, como el caso de la distribución exponencial, se debe relizar el ajuste sobre:

$$Y = \frac{X - min(X)}{max(X) - min(X)}$$

si algun valor de X se encuentra por debajo del 0

Si $x \in [0,1]$ como el caso de la distribución Beta o la Kumaraswami

$$Y = X + min(X) + \epsilon$$

si todos los valore de X son mayores que 0

Capítulo 3

Métodos de Clasificación y Multimodalidad

3.1. Introducción

Dentro de nuestro universo de datos, podemos contar con observaciones inusuales, también llamados outlaiers y es de esperar que alguna de estas afecte negativamente el ajustar una distribución y debamos recurrir a modelos no paramétricos para obtener información a partir de nuestros datos como el valor de la media o la distancia interquantilica y deberemos dar un tratamiento especial a estas observaciones por separado.

Las pruebas de bimodalidad o multimodalidad van estrechamente relacionadas a la separación y clasificación de una distribución, por lo que habrá una transición natural a estas pruebas a traves de los primeros métodos de clasificación.

3.2. Separación de outliers

Supongamos entonces que tenemos la siguiente información que representa la estatura de un grupo de personas

Es de esperar que la distribución de estos datos sea Normal sin embargo, como podemos observar, contamos con dos observaciones atípicas.

Este tipo de problemas suele suceder porque al momento de capturar la información esta no es homogénea, dado que se omitieron variables como la edad o el sexo y que de poder clasificarse, las pruebas anteriores pasarían sin problema, observando distribuciones distintas para adultos y niños, (otra forma de atacar el problema es incrementando el tamaño de muestra, probando con distribuciones de colas pesadas como Cauchy o transformar los datos para probar con nuevas distribuciones como Beta o Kumaraswamy).

Observemos entonces cómo se ajusta una distribución normal con y sin outlayers.

La forma de distinguir y separar estas observaciones también ha sido afrontada en el pasado mediante los siguientes métodos:

3.2.1. Método z

3.2.2. Rango intercuantílico

3.2.3. Pruebas de multimodalidad

Existen pruebas para la detección de la multimodalidad en una función de distribución como el exceso de masa y ancho de banda crítico

Exceso de masa

Ancho de banda crítico

Para un número de modas $k \in N$, el ancho de banda crítico es el menor de los anchos de banda tal que la densidad del kernel tiene al menos k modas

Recordemos que el ancho de banda se define como Y la densidad del kernel corresponde a

Entonces:

$$h_k = inf\{h : M(f_h) \geq k\}$$

En donde $M(f_h)$ es el número de modas de f_h y f_h representa el estimador de la densidad del kernel de una muestra aleatoria $X = (X_1, ..., X_n)$.

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

3.2.4. Otros métodos de clasificación

K-medias

También podemos introducir modelos de aprendizaje no supervizado

Originalmente, estos modelos son utilizados para datos multivariantes

K-vecinos

3.2.5. ...

Capítulo 4

Construcción del modelo

4.1. Introducción

Ahora que hemos definido metodologías para estimar parámetros, ajustar distribuciones y clasificar observaciones es momento de aplicarlo.

Observemos el caso anterior en el que la solución fue seprar las observaciones y ajustar la distribución, ¿Qué sucedería si aplicamos dicho proceso de forma más general?, es decir, que de un conjunto de datos al que las distribuciones usuales fueron rechazadas y que ahora decidimos clasificar nuestras observaciones de tal forma que los conjuntos por separado sean nuevamente examinados para una nueva prueba de bondad y ajuste.

4.2. Forma de estudio

Paso 1 Tomar una muestra suficientemente grande de datos

Paso 2 Estimar parámetros y ajustar los datos

Los tenores de esta parte incluyen aplicar transformaciones, distintos métodos de estimación de parámetros y diversos test de ajuste de distribuciones

En caso de que el Paso 2 falle, es preciso separar la distribución y aplicar de forma recursiva el paso 1 y 2 hasta que todas las particiones tengan una distribución asignada que haya pasado las pruebas.

4.3. Comparativa entre modelos

Serán tomadas las siguientes consideraciones:

Aquellos casos en los que la distribución multimodal se componga de dos distribuciones que se hallen muy separadas entre sí serán omitidas, pues el punto de separación de la distribución e trivial.

Se ponderá el p.valor y el criterio de Akaike, de tal forma que se obtenga el mejor modelo.

El p.valor será aplicable para los test de Kolmogorov Smirnov y Anderson Darling, tomado dos criterios distintos, el primero será si pasó uno de ellos o si pasó ambos y se tomará un $\alpha = 0.05$ por defecto.

Se aplicarán los distintos métodos de clasificación en caso de que no ajuste la distribución.

Se aplicarán métodos de ajuste de distribuciones multimodales según sea el caso.

Se realizará con 500 muestras distintas de tamaño 5000 cada una para asegurar la consistencia de los resultados y que sean generalizables

Trabajaremos con los datos generados a partir de la siguiente expresión en el software estadístico R:

Primero generamos muestras de tamaño 10000 para los siguientes modelos:

Beta-Normal

R<-c(rbeta(5000,1,1/2),rnorm(5000,-1,3/4))

Se esperan los siguientes resultados:

Resultados obtenidos

Normal-Normal

R<-c(rnorm(5000,0,1),rnorm(5000,3,1))

LogNormal-Gamma

4.4. Ventajas

Gana precisión al momento de estimar una distribución.

Es una generalización del los modelos contemporáneos y la solución no se basa en fuerza bruta, es decir, no es necesario calcular cientos de distribuciones distintas para elegir alguna que ajuste.

Se puede simplificar fácilmente a un modelo de ajuste clásico.

Es sencillo obtener más de un modelo multinomial que ajuste.

Puede ayudarnos a visualizar si la información que estamos evaluando necesita un análisis previo.

Amplía el panorama para una mejor evaluación de estadísticos locales y no uno global que pudiera tener un mayor error.

Hay más de una forma de escribir una variable aleatoria multimodal, podemos verla como una combinación lineal de dos o más distribuciones, por ejemplo

$Y = Norm * I + Gamma * I$

En donde la función indicadora nos facilita el análisis del total de la distribución y al ser una suma, los estadístios como la esperanza y varianza pueden ser estimados de la misma forma que en el método clásico.

Se pueden combinar distribuciones de colas pesadas, añadiendo más familias con las que es posible trabajar.

Se ofrece un modelo paramétrico para información a la que es casi imposible ajustar o cuyos p.valores son insuficientes.

4.5. Desventajas y Soluciones

Al tener un mayor número de parámetros, la función de distribución multimodal se ve afectada por el criterio de Akaike y Bias, por lo que se recomienda su uso buscando en principio distribuciones de un solo parámetro (como la exponencial), y un número pequeño de particiones para la distribución, es decir buscar que sea bimodal o trimodal .

Dependiendo del método de corte o división de la variable aletoria, se podrán obtener diferentes modelos para una sola muestra, la recendación es emplear el modelo con diferentes muestras de la distribuciónom llegar a aquel que más se repita o elegir las distribuciones más simples, es decir, realizar un ejercicio de bootstraping no para elegir un parámetro sino un modelo.

Por el momento sólo y está enfocado en modelos univariantes, al añadir más dimensiones al modelo, el proceso se vuelve más robusto, debiendo añadir más criterios para los algoritmos de separación e introducir cópulas para no perder la correlación entre marginales.

La definición formal de un modelo multimodal es difícil de definir debido a los problemas de distribuciones "demasiado combinadas"

Capítulo 5

Aplicaciones

5.1. Introducción

Las distibuciones multimodales se pueden encontrar en... Y la información a la que aplicaríamos un modelo no paramétrico sería...

5.2. Seguros

Supongamos que los montos de las reclamaciones son obtenidos por la siguiente distribución y los siguientes outlayers...

5.3. Epidemiología

Tomemos la distribución por edad de personas que han adquirido la enfermedad...

5.4. Cobranza

Capítulo 6

Conclusiones

El criterio princpal por el que se debe empezar a tratar un conjunto de datos es por medio de sus variables discretas, pues los errores de ajuste se deben principalmente a que la información no ha sido tratada correctamente.

La separación de la información es una buena práctica aún si parece no haber razón para dividir, en caso de contar con un conjunto de datos de gran tamaño se puede optar por reducir dimensiones o priorizar sólo las variables más significativas.

Otra ventaja de este trabajo es que puede

Una de las principales desventajas es el sobreajuste, en caso de que una muestra muy grande sea dividida en muchos pedazos pequeños indica que no será bueno generalizando, pues la cantidad de parámetros que tendrá la funcón afectan muy fuertemente por medidas como el criterio de Akaike

Antes de ejecutar el algoritmo, siempre hay que verificar que no haya variables categóricas que podamos usar a nuestro favor

Capítulo 7

Anexo Código en R.

Se ha realizado un paquete en R exclusivo para este trabajo, disponible con el siguiente código: install.packages("FitUltD")

FDist que ajusta una distribución a una muestra de datos dada con la ventaja de que ajusta la función indicadora de esta a la forma en la que están dados los datos

FDistUlt emplea la función anterior para ajustar una distribución, en caso de no haber ajuste separa la muestra utilizando el método de K-Medias y continúa de forma recursiva hasta hallar las distribuciones que simulan a la muestra.

Todos los gráficos que fueron hechos en R se encuentran disponibles en la siguiente dirección:

https://github.com/jcval94/Tesis