

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



# Mezclas de distribución en esperanza-varianza

TESIS

QUE PARA OBTENER EL TÍTULO

LICENCIADO EN ACTUARÍA

PRESENTA

DAVID EDGARDO CASTILLO RODRÍGUEZ

ASESOR: JUAN CARLOS MARTÍNEZ OVANDO

MÉXICO, D.F.

2016

“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 “**TÍTULO DE LA TESIS**”, 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”.

AUTOR

---

FECHA

---

FIRMA

*DEDICATORIA*



# Agradecimientos

¡Muchas gracias a todos!



# Capítulos

- 1. Introducción
- 2. Modelo de mezcla en media
  - 2.1. Modelo de mezcla en varianza
  - 2.3. Modelo de mezcla en media-varianza
- 3. Inferencia estadística
  - 3.1. Noción de distribución tipo mezcla
  - 3.2. Verosimilitud
  - 3.3. Verosimilitud aumentada
  - 3.4. Estimación de parámetros
  - 3.5. Algoritmo
- 4. Caso de estudio
  - 4.1. Contexto
  - 4.2. Fuentes de información y datos
  - 4.3. Estimación
  - 4.4. Implicaciones
- 5. Conclusiones y trabajo futuro





# Introducción



# Modelo de mezcla en media

Se dice que un vector aleatorio  $X$ , de dimensión  $p$ , es una mezcla en media, si dado  $u$  se distribuye normal con vector de medias  $u\mu$ , y matriz de varianza-covarianza  $\Sigma$ . La variable aleatoria  $u$  está definida en  $R_+$ . Entonces la probabilidad de que  $X$  pertenezca a algún subconjunto  $S_p \in R^p$  es:

$$P((X_1, \dots, X_p) \in S_p) = \int_0^{\infty} N_p(X \in S_p | u\mu, \Sigma) f_u(u) du$$

donde  $\mu \in R^p$ ,  $\Sigma \in M_{p \times p}$  y  $u$  es la variable de mezcla.

Si nos enfocamos en la distribución marginal de  $X$  dado  $u$  podemos notar que el valor de  $u$  influye en dónde está centrada la distribución, pues ahora el vector de medias pertenece al segmento dirigido  $L(\mu) = (u\mu | u \in R_+, \mu \in R^p)$ , por lo que si consideramos que  $u$  ha tomado el valor de  $u_i$ , y una muestra aleatoria de  $X$  dado  $u_i$ , entonces esta muestra estará centrada sobre el segmento dirigido  $L(\mu)$  en el punto  $u_i\mu$ , por lo que para valores de  $u$  cercanos a uno, la muestra permanece centrada al rededor de  $\mu$ , mientras que para valores grandes de  $u$ , la muestra se aleja en dirección del vector de medias  $\mu$ .

En la siguiente figura se observa un ejemplo de una distribución tipo mezcla en media con vector de medias  $\mu = (9, 11)$ , matriz de varianza-covarianza  $\Sigma = (5, 1, 1, 4)$ , con variable de mezcla distribuida exponencialmente.

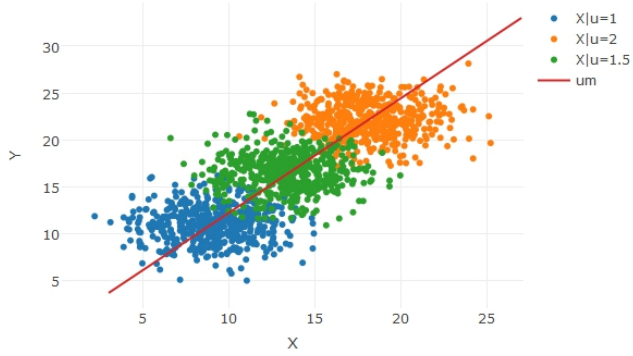


Figura 1: Se asumió que  $u$  tomó los valores 1, 1.5 y 2.

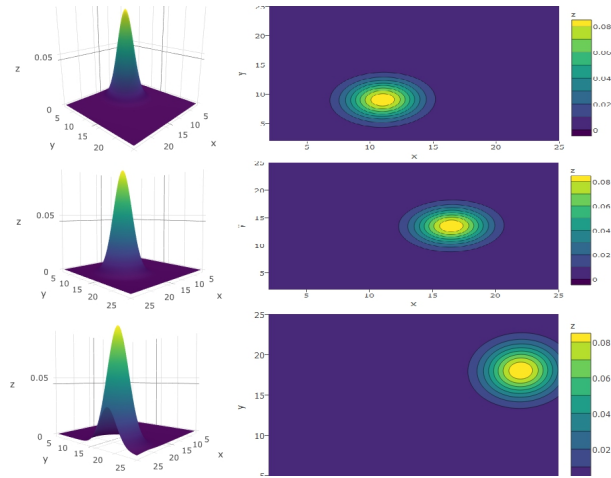


Figura 2:

En la figura 2 se observa cómo se ve afectada por la variable de mezcla  $u$ , tanto la función de densidad de  $X$  dado  $u$ , como los contornos del vector aleatorio del ejemplo anterior.

Ahora, si nos enfocamos en la distribución de  $X$ , tenemos que:

$$E[X] = \int_0^\infty \left( \int_{D_X} X N_p(X|u\mu, \Sigma) f_u(u) du \right) dx =$$

Luego, por el teorema de Fubinni:

$$\int_0^\infty f_u(u) \left( \int_{D_X} X N_p(X|u\mu, \Sigma) f_u(u) dx \right) du = \int_0^\infty f_u(u) E[X|u] du$$

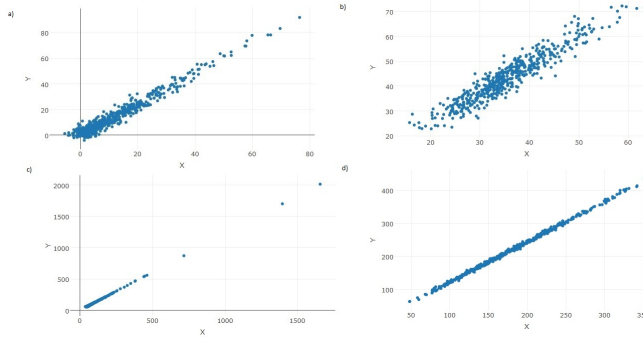
$$\int_0^\infty f_u(u) u \mu du = \mu \int_0^\infty f_u(u) du = \mu E[u]$$

Mientras que por el anexo 4 se tiene que:

$$Cov(X) = E[Cov(X|u)] + Cov[E[X|u]] = E[\Sigma] + Cov[u\mu] = \Sigma + var(u)\mu\mu'$$

Siempre y cuando el segundo momento de  $X$  y  $u$  existan. Así, la variable de mezcla influye en la localización y dispersión del vector aleatorio  $X$ . Este tipo de mezcla nos permite modelar conjuntos de datos que sigan alguna tendencia con una dispersión constante, como se ilustra en la gráfica 1. Por otra parte, una de las mayores dificultades es que, salvo en algunas excepciones, no siempre existe una forma cerrada para la densidad de  $X$ , por lo que tendría que ser aproximada mediante algún método numérico.

Mediante un proceso de simulación se generaron las siguientes gráficas. La gráfica a) proviene de una variable de mezcla exponencial con parámetro  $\lambda = 1,5$ , la gráfica b) proviene de una variable de mezcla gamma con parámetros  $\alpha = 20$ ,  $\beta = 5$ , la gráfica c) proviene de una variable de mezcla pareto con parámetros  $\theta = 5$ ,  $\alpha = 2$ , la gráfica d) proviene de una variable de mezcla Ji-cuadrada con 20 grados de libertad.



## Modelo de mezcla en varianza

Ahora considerando un modelo similar al anterior, pero con la diferencia que la variable de mezcla  $u$  sólo afecta a la matriz de varianza-covarianza  $\Sigma$ , por lo que la distribución del vector aleatorio  $X$   $p$ -dimensional viene dado por:

$$P((X_1, \dots, X_p) \in S_p) = \int_0^{\infty} N_p(X \in S_p | \mu, u\Sigma) f_u(u) du \quad (1)$$

Si nos concentramos en la distribución condicional de  $X$  dado  $u$ , podemos notar que  $u$  influye en la dispersión de  $X$ , por lo que entre mayor sea el valor que tome  $u$ , mayor será la dispersión al rededor del vector de medias  $\mu$ , lo cual se ilustra en la siguiente figura.

En la siguiente figura se observa como se modifica tanto la densidad como los contornos de la densidad de  $X$  dado  $u$ , donde  $u$  tomó los valores del ejemplo anterior.

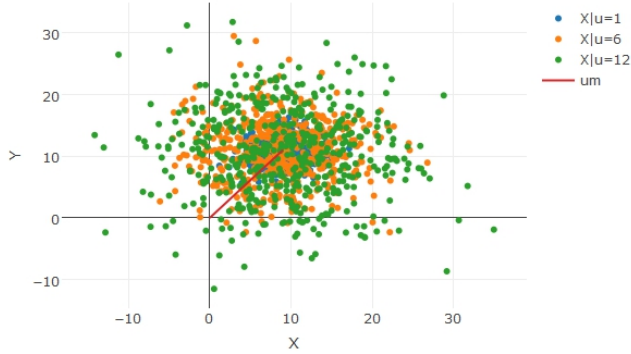


Figura 3: Se consideraron los mismos parámetros que en la mezcla en media, a excepción de que  $u$  tomó los valores 1, 6 y 12 .

Ahora considerando la distribución de  $X$ , mediante el anexo 4, es fácil probar que:

$$E[X] = \mu$$

$$COV[X] = COV_u(E_x(X|u)) + E_x(COV_u(X|u)) = E(u)\Sigma E(u)'$$

Por lo que la distribución de  $X$  estará centrada en el mismo vector de medias que  $X$  dado  $u$ , mientras que la matriz de covarianzas será proporcional a la matriz de varianzas covarianzas de  $X$  dado  $u$ .

Ahora se ilustrará, mediante muestras aleatorias, que efecto tiene el condicionamiento en  $u$  de la distribución de  $X$ , adviértase que a diferencia del condicionamiento en media, los datos solo se concentran alrededor del vector de medias  $\mu$ , pues  $E[X] = E[X|u]$ , perdiendo la trayectoria del vector de medias, por lo que intuitivamente podemos pensar que  $u\mu$  del modelo de mezcla en media generará un vector de trayectoria para las observaciones de  $X$ .

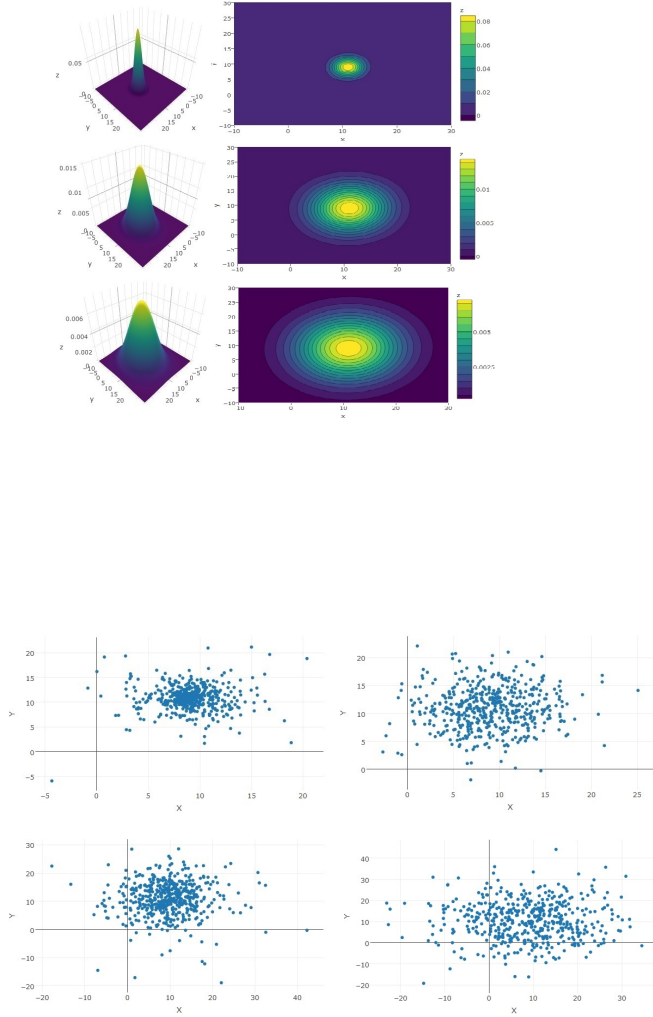


Figura 4: Se tomaron las mismas distribuciones y parámetros que las gráficas de mezcla en esperanza, pero ahora la variable de mezcla solamente afectó a la matriz de varianza-covarianza.



## Modelo de mezcla en esperanza-varianza

Generalizando los dos modelos anteriores se tiene la siguiente definicion:

Se dice que un vector aleatorio  $X \in R^p$  es un vector tipo mezcla p dimensional, si  $X$  dado  $u$  se ditribuye normal con vector de medias  $\mu + u\beta$ , y matrhz de varianza-covarianza  $u\Sigma$ , es decir,  $N(\mu + u\beta, u\Sigma)$ , donde  $u$  es la variable de mezcla con soporte en  $R_+$ ,  $\mu, \beta \in R^p$  y  $\Sigma \in M_{p \times p}$  matriz de varianza-covarianza.

Por lo que la distribución de  $X$  está dada por:

$$f(x) = \int_{S_u} \frac{1}{(2\pi)^{n/2} |u\Sigma|^{n/2}} \exp -\frac{1}{2}(x - \mu - uB)'u\Sigma^{-1}(x - \mu - uB) f(u) du$$

Si nos concentramos en la distribución marginal de  $X$  dado  $u$ , podemos notar que ahora para cada realización  $u_i$ , la densidad de  $X$  dado  $u_i$  estará centrado en el vector  $\mu + u_i\beta$ , y conforme  $u_i$  tome valores más grandes, la densidad de  $X/u_i$  se desplazará en dirección del vector  $u_i\beta$  cada vez con mayor dispersión, lo cual se ilustra en la siguiente figura.

En la distribución de  $X$  podemos considerar a  $\mu$  como un vector de posición, mientras que  $u\beta$  es un vector de dirección o tendencia al rededor del cual estarían las observaciones de  $X$ , mientras que los datos conservarán la estructura de correlación inherenta a  $\Sigma$ , pero además estarán dispersos en proporción a la esperanza de la variable de mezcla  $u$ , más alguna constante que depende de  $u$ . Entonces, considerando el anexo 4, la esperanza y covarianza del vector  $X$  son de la siguiente forma:

$$\begin{aligned} E(X) &= E_u(E_x(x|u)) = E_u(\mu + u\beta) = \mu + E(u)\beta \\ COV(X) &= COV_u(E_x(X|u)) + E_u(COV_x(X|u)) = \\ &COV_u(\mu + u\beta) + E_u(u\Sigma) = Var(u)\beta\beta' + E(u)\Sigma \end{aligned}$$

Una propiedad importante de las variables de mezcla, es que el comportamiento de la cola del vector aleatorio  $X$  dependende del comportamiento de la cola de la variable de mezcla  $u$ , por lo que se podría seleccionar alguna

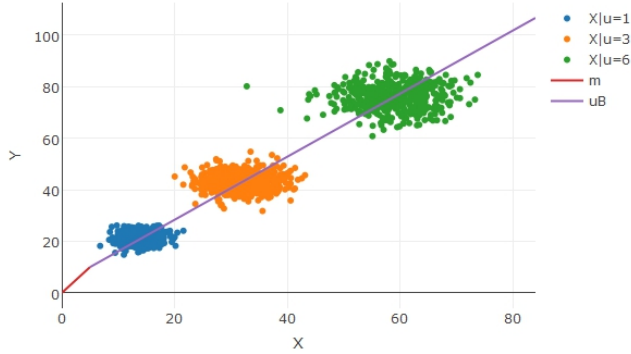


Figura 5: Nuevamente se asumió que  $u$  se distribuyó exponencial tomando los valores 1, 3 y 6, mientras que  $m=(5,10)$ ,  $B=(9,11)$ , y la matriz de varianza covarianza fue la misma que en los ejemplos anteriores.

$u$  arbitraria para que  $X$  tenga colas pesadas.

Y además, como se verá posteriormente, la función generadora de momentos  $M_x(t)$  del vector aleatorio  $X$  es proporcional a  $M_u(t)$ , la función generadora de momentos de la variable de mezcla  $u$ . Y la función característica  $\Phi_x(it)$  es proporcional a la función característica de la variable de mezcla  $u$ ,  $\Phi_u(it)$ .

## Propiedades

1) La función característica de  $X$  es:

$$\Phi_X(t) = \exp(it\mu)\Phi_u(it\beta - \frac{1}{2}t\Delta t)$$

Donde  $\Phi_u(\cdot)$  es la función característica de la variable de mezcla  $u$ .

Demostración:

$$\begin{aligned}\Phi_X(t) &= \int_x \exp(it)f(x)dx = \int_x \exp(it)\left(\int_u f(x)|_u f(u)du\right)dx = \\ &= \int_x \int_u \exp(it)f(x)|_u f(u)dudx\end{aligned}$$

Luego por el Teorema de Fubini (ver algún anexo)

$$\begin{aligned}\int_x \int_u \exp(it)f(x)|_u f(u)dudx &= \int_u \int_x \exp(it)f(x)|_u f(u)dudx = \\ &= \int_u \exp(it(\mu + u\beta) + \frac{1}{2}itu\Delta it)f(u)du\end{aligned}$$

Ya que la función generadora de momentos de una distribución normal  $p$ -variada es:  $\exp(t\mu + t\Sigma t)$ , y además  $X$  dado  $u$  se distribuye  $N(\mu + u\beta, u\Delta)$

Factorizando el exponente de la exponencial se tiene que:

$$\begin{aligned}\int_u \exp(it(\mu + u\beta) + \frac{1}{2}itu\Delta it)f(u)du &= \\ \int_u \exp(it\mu)\exp(u(it\beta - \frac{1}{2}tu\Delta t))f(u)du &= \exp(it\mu)\Phi_u(it\beta - \frac{1}{2}t\Delta t)\end{aligned}$$

Por lo tanto:

$$\Phi_X(t) = \exp(it\mu)\Phi_u(it\beta - \frac{1}{2}t\Delta t)$$

Para comprobar que la función generadora de  $X$  es:

$$M_x(t) = \exp^{t\mu} M_u(t\beta + \frac{1}{2}t\Delta t)$$

basta con sustituir  $t$  por  $it$  en el resultado anterior.

El resultado anterior nos indica que podemos obtener la función generadora de momentos o característica del vector aleatorio  $X$  con tan solo conocer la función característica o generadora de la variable aleatoria  $u$ , lo cual sería de utilidad si solo nos interesan algunos momentos, o también nos ayudaría a identificar la distribución de  $X$ , en caso de reconocer la función característica o generadora de momentos. Inversamente, el resultado anterior también nos indica que si la función característica o generadora de algún vector aleatorio  $Z$  se puede descomponer como el producto de  $\exp it\mu$ , y alguna  $M_y(t)$ , entonces  $Z$  es una variable normal de mezcla en esperanza-varianza, con distribución de mezcla  $y$ .

# Inferencia estadística

En este capítulo se verá lo que es una distribución tipo mezcla en general, así como algunas de sus utilidades, para después enfocarnos en la mezcla que nos es de interés, es decir, la mezcla continua. Luego se ahondará en cómo estimar los parámetros de las mezclas compuestas a través del método de máxima verosimilitud, y la necesidad de extender la verosimilitud. Por último, se propondrá seleccionar el muestreador de Gibbs para encontrar los parámetros de interés de nuestra mezcla.

## Distribuciones tipo mezcla numerable

Este tipo de distribución consiste en ponderar un conjunto numerable de vectores aleatorios  $X_i$  discretos con el mismo soporte y la misma dimensión con un conjunto de pesos, positivos y menores a uno,  $w_i$ , donde  $i$  proviene de un conjunto  $I$  de índices numerable, y a su vez  $\sum_{i \in I} w_i = 1$ . El índice  $i$  puede ser generado por una distribución de probabilidad discreta, en este caso  $P(i) = w_i$  por lo que a los pesos también se les conoce como probabilidades. Entonces la función de densidad es de la siguiente forma:

$$f_x(x) = \sum_{i \in I} w_i f_{X_i}(x)$$

Donde  $0 < w_i < 1$ ,  $\sum_{i \in I} w_i = 1$ , y  $f_{X_i}(x)$  es un vector aleatorio discreto de dimensión  $p$  para todo  $i$  en  $I$ .

Es común encontrar este tipo de distribuciones en un conjunto de datos que provienen de dos o más distribuciones diferentes, lo cual podría resul-

tar en una distribución multimodal. Si el conjunto de vectores aleatorios provienen de la misma distribución, entonces se dice que es una mezcla de familias paramétricas, mientras que si el vector aleatorio tienen una distribución continua, entonces se le conoce como densidad tipo mezcla continua.

## Densidad tipo mezcla continua

Una densidad tipo mezcla continua resulta de considerar que los pesos  $w$  están dados por una densidad continua  $f_w(w)$  con soporte en  $R_+$ , a  $w$  se le conoce como variable de mezcla. Ahora, el vector aleatorio  $X$  está parametrizado por  $(\theta, w)$ , por lo que ahora la familia de vectores aleatorios queda dada por la familia no numerable  $X$  dado  $w$ , tal que  $w$  se distribuye  $f_w(w)$ , de aquí que la densidad de  $X$  esté dada por:

$$f_x(x) = \int_0^\infty f(x|w)f_w(w)dw$$

Si  $w$  es una variable aleatoria con distribución paramétrica, es común que la distribución de  $X$  dependa de los parámetros de  $u$ , y además ciertas características como la dispersión o la cola de la distribución de  $X$  también dependerán de las características de la variable de mezcla  $w$ .

## Función de verosimilitud y verosimilitud extendida

De entre los distintos métodos de estimación de distribuciones paramétricas se usará el de máxima verosimilitud debido a las propiedades de estos estimadores, y a que actualmente existen métodos computacionales que permiten obtener dichos estimadores sin necesidad de que la distribución del vector aleatorio  $X$  tenga una forma analítica.

Supongamos que tenemos una muestra aleatoria independiente e idénticamente distribuida de tamaño  $n$  del vector aleatorio  $X \in R^p$ , y que además, la distribución de  $X$  puede ser estudiada a partir de una distribución tipo mezcla, con variable de mezcla  $u$ , entonces como los parámetros de  $X$  dependerán de la variable de mezcla  $u$ , la propiedad de invarianza de los estimadores máximo verosímiles nos permite hacer la inferencia sobre

los parámetros de la distribución de mezcla y la variable de mezcla, por lo que la función de verosimilitud está dada de la siguiente forma:

$$P(X_1 = x_1, \dots, X_n) = \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n \int_u f(X_i = x_i | \theta, u) f_u(u | \alpha) du$$

Donde  $\Theta$  son los parámetros correspondientes a la distribución de  $X$  dado  $u$ , mientras que  $\alpha$  son los parámetros de la distribución de mezcla  $u$ . Ahora, suponiendo que se cumplen las condiciones de regularidad se tiene que:

$$Lik(\theta, \alpha) = \int_u \prod_{i=1}^n f(X_i = x_i | \theta, u) f_u(u | \alpha)$$

La última expresión obtenida parece complicada a simple vista, pues además del producto de las funciones de densidad, hay un proceso de integración, lo cual nos lleva a buscar una alternativa para manejar la función de verosimilitud, lo cual se logra a través de extender la verosimilitud.

## Verosimilitud extendida

El proceso de extender la verosimilitud consiste en tratar al vector aleatorio  $X$   $p$  dimensional como una variable tipo mezcla, donde la variable de mezcla  $u$  se considera como latente. Este método pretende encontrar los mejores estimadores para la variable tipo mezcla  $u$ , y a su vez que esta variable de la más verosimil probabilidad de haber observado al vector  $X$ .

Por ejemplo, supongamos que  $X$  es una variable aleatoria Pareto( $1, \beta$ ), de la cual tenemos una realización  $X_0$ , y que nos interesa estimar el valor de  $\beta$ , luego sea  $X$  dado  $u$  una distribución exponencial con parámetro  $\lambda$ , y además supóngase que  $\lambda$  también es una variable exponencial con parámetro  $\beta$ . Entonces,  $X$  se puede expresar como una mezcla de distribuciones exponenciales como sigue:

$$f_X(x) = \int_0^{\infty} f_{x|\lambda}(x) f_{\lambda}(\lambda) d\lambda = \int_0^{\infty} \lambda \exp^{-\lambda x} \beta \exp^{-\beta \lambda} d\lambda = \frac{\beta}{(x + \beta)^2}$$

Ahora, si consideramos la distribución de  $X$  dado  $\lambda_i$ , para cada  $\lambda_i$  tendríamos una densidad diferente. En la siguiente figura se ilustra la densidad de  $X$  dado  $\lambda$ , para algunos valores de  $\lambda$ .

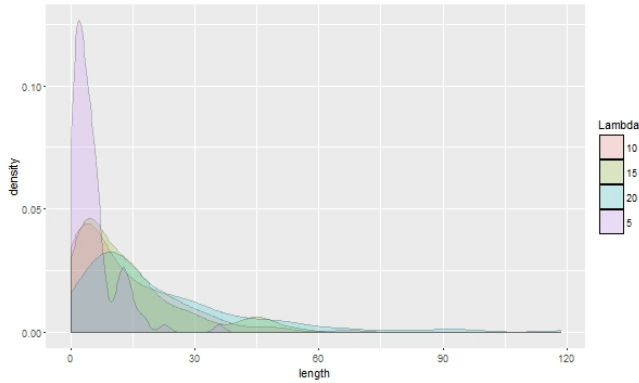


Figura 6: Se graficaron distribuciones exponenciales con parámetro  $\lambda = 5, 10, 15, 20$

De aquí, que nos podríamos plantear la siguiente pregunta ¿Dada una observación  $X_0$ , cual sería la  $\lambda_i$  que le asigne mayor probabilidad o densidad a la ocurrencia de esta observación? Por ejemplo, de la gráfica anterior se puede observar que si  $x_0 = 5$ , el mejor valor de lambda sería  $\lambda = 5$ , mientras que si  $x_0 = 30$ , sería  $\lambda = 15$ . En la misma línea, también sería acertado preguntarse por el mejor estimador de  $\beta$ , donde la probabilidad de que ocurra la  $\lambda_i$  (que maximiza la la probabilidad de ocurrencia de  $x_0$ )



sea máxima. Entonces, si consideramos la función de verosimilitud para  $X_0$ , sería:

$$Lik(\beta|X) = f_x(x) = \frac{\beta}{(x + \beta)^2} = \int_0^{\infty} f_{x|\lambda}(x) f(\lambda) d\lambda$$

Y con el planteamiento anterior, en lugar de buscar el mejor estimador de  $\beta$  en  $Lik(\beta|X_0)$ , podríamos preguntarnos por la mejor  $\lambda_i$  inducida por  $\beta$  tal que  $\beta$  maximice la ocurrencia de  $\lambda_i$ , y a su vez,  $\lambda_i$  maximice la probabilidad de haber observado  $X_0$ , entonces con este enfoque, la función de verosimilitud para  $X_0$  queda expresada de la siguiente manera:

$$Lik(\lambda_i, \beta|X_0) = f_{x|\lambda_i}(x|\beta, \lambda_i) f(\lambda_i)$$

Siguiendo el planteamiento anterior, la verosimilitud extendida se define de la siguiente forma:

Sean  $X_{i=1}^n$  un conjunto de vectores aleatorios independientes e idénticamente distribuidos de dimensión  $p$  con algún vector de parámetros  $\theta$ , y sea  $u$  una variable aleatoria con algún vector de parámetros  $\beta$  y con soporte en  $R_+$ , tal que  $X_i$  se puede expresar como una distribución tipo mezcla con variable de mezcla  $u$ . Entonces  $f_{X_i}(x_i) = \int_0^{\infty} f_{x_i|u}(x_i) f_u(u) du$  para toda  $i$ , con  $X_i$  dado  $u_i$  condicionalmente independiente a  $X_j$  dado  $u_j$ , y  $u_i$  dado  $\beta$  independiente a  $u_j$  dado  $\beta$  para toda  $i$  diferente de  $j$ , entonces:

$$Lik(u_i, \theta, \beta|X) = \prod_{i=1}^n f_{x_i|u_i}(x|\theta, u_i, \beta) f_{u_i}(u_i|\beta)$$

donde  $u_i$  es una variable latente.

Trabajar con la verosimilitud extendida podría parecer más complicado al tener que estimar una variable latente para cada observación de  $X$ , pero en ocasiones resulta más sencillo optimizar la función  $\prod_{i=1}^p f_{x_i|u_i}(x_i) f(u_i)$  que  $f(X)$ , y además, existen algoritmos computacionales, vía cadenas de Markov, o el algoritmo EM, que permiten encontrar los estimadores de los parámetros deseados. Por lo que ahora nuestro objetivo será seleccionar un algoritmo que nos permita trabajar con la verosimilitud extendida.

## Muestreador de Gibbs

El muestreador de Gibbs es un método de simulación, que permite obtener una muestra de una distribución conjunta a partir de las distribuciones marginales. Entre los principales usos de este algoritmo, es obtener una muestra de los parámetros de interés de algún modelo probabilístico, a partir de las distribuciones condicionadas de los parámetros. Este método consiste en construir una cadena de Markov, para así conseguir una muestra de una distribución dada, la construcción de la cadena de Markov es necesaria para asegurar que nuestra muestra proviene efectivamente de la distribución que nos interesaría muestrear. Por lo que en la siguiente parte se verá el planteamiento y la forma de implementar este algoritmo para obtener una muestra de parámetros de interés.

Supongamos que tenemos una distribución de la forma  $f(X, \Theta)$  donde  $X$  es un vector aleatorio y  $\Theta$  algún vector de parámetros del cual se desearía tener una muestra mediante una cadena de Markov, es decir, obtener una muestra de  $f(\theta^k|X)$  a partir de una sucesión de parámetros  $\theta_1^\infty$ , tal que la probabilidad de transición de esta sucesión converga a la distribución de la cual se desearía tener una muestra, esto es,  $\lim_{k \rightarrow \infty} P(\theta^{k+1}|\theta^k, X) = P(\theta|X)$ . Dada la descripción del modelo anterior, la cadena debe ser: estacionaria, irreducible, aperiódica y recurrente positiva. La propiedad de estacionariedad asegurar que  $P(\theta|X)$  no cambie cuando  $k$  tienda a infinito, irreducible para no descartar ningún valor posible de  $\theta^k$  en cada iteración, mientras que aperiodicidad para que la cadena no tenga ciclos, y así permanecer en el estado estacionario. Por último, recurrente positiva, para que se pueda visitar cualquier estado un número infinito de veces para así asegurar que la distribución límite tomó toda la información disponible en consideración.

Ahora bien, para encontrar una cadena que nos permita obtener las características anteriores German & German (citar bien esta parte) demuestra que  $\lim_{k \rightarrow \infty} P(\theta^k|X) = P(\theta|X)$ , donde:

$$P(\theta^k|X) = \prod_{j=1}^t P(\theta_j|X, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_n)$$

donde  $\theta_j$  dado  $X, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_n$  es condicionalmente independiente

para toda  $j$ , para alguna medida de probabilidad arbitraria  $P(\cdot)$ . Entonces si  $\theta_j^k$  se distribulle  $f(\theta_j^k|X, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_n)$ , de aquí sería fácil obtener una muestra de  $\theta_j^k$  para toda  $j$ , si su densidad tiene una forma cerrada, y así se construiría una cadena de markov con las características deseadas,  $[\theta^k]_1^\infty$ , con probabilidad de transición  $P(\theta^k|X)$ .

Entonces los pasos para implementar el Muestreador de Gibbs son los siguientes:

Dar un conjunto de valores iniciales  $\Theta_0$ , no importa que valores se den, la propiedad de estacionareidad, asegurará la convergencia al estado estacionario. El vector  $\theta$  puede ser particionado en  $d$  subconjuntos excluyentes, tal que  $d \leq n$ , con la finalidad de simplificar los cálculos. Estratégicamente se pueden seleccionar los sobconjuntos de modo que los que tienen alta correlación queden juntos.

Para cada  $\theta_j$  obtener su distribución condicionada  $f(\theta_j|X, \dots, \theta_{0,j-1}, \theta_{0,j+1}, \dots, \theta_{0,n})$ , y realizar una simulación de esta. El valor de  $\theta_j$  obtenido reemplazará a los valores iniciales. En caso de que no se pueda tener la distribución de  $\theta|_{X, \theta_{\theta_j}}$  se realiza un procedimiento de aceptación rechazo para obtener una realización de  $\theta_j$ .

Una vez obtenidos todos los valores de  $\theta|_{X, \theta_{\theta_j}}$  se reemplazan por los valores iniciales y se repite el procedimiento hasta que la cadena se estabilice.

computacionalmente el algoritmo se vería de la siguiente forma:

```

 $\Theta_k < -(\theta_{1,0}, \dots, \theta_{d,0})$ 
dowhile  $||\theta^{k+1} - varaux < \epsilon||$ 
     $varaux < -\theta_k$ 
    for (j in 1:d)c
         $\theta_j^{k+1} < -\text{random}(f(\theta|_{X, \theta^k_{\theta_j}}))$ 
    end do while

```

## Inferencia de los parámetros de una distribución Hiperbólica generalizada

Supongamos que tenemos una colección de  $r_n$  datos, donde cada  $r_i \in R^p$ , de los cuales tenemos motivos para creer que fueron generados por una distribución hiperbólica generalizada p-variada. Entonces nos interesaría inferir los valores de los parámetros de dicha distribución, pero antes de ello sería conveniente expresar a nuestra distribución de interés como una distribución tipo mezcla normal, con variable de mezcla  $u$  distribuida gaussiana inversa (GI), por lo que así podemos replantear el problema, enfocandonos en buscar los parámetros de la distribución de mezcla,  $\mu, \beta, \Sigma$ .

Ahora sobre un enfoque bayesiano, podemos plantear la verosimilitud de la distribución aposteriori de  $\mu, \beta, \Sigma$ , es decir:

$$f(\mu, \beta, \Sigma | (r_i)_{i=1}^n) = \prod_{i=1}^n f_{r_i}(r_i | \mu, \beta, \Sigma) \Pi(\mu, \beta, \Sigma)$$

Donde  $\Pi(\mu, \beta, \Sigma)$  es la distribución conjunta a priori de  $\mu, \beta$  y  $\Sigma$ , y además asumiremos que tanto  $\mu, \beta, \Sigma$  son independientes, por lo que:

$$f(\mu, \beta, \Sigma | (r_i)_{i=1}^n) = \prod_{i=1}^n f_{r_i}(r_i | \mu, \beta, \Sigma) \Pi(\mu | \mu_0, \Sigma_0) \Pi(\beta | \mu_\beta, \Sigma_\beta) \Pi(\Sigma | S)$$

Como supuesto adicional, asumiremos que  $\mu$  se distribuye  $N(\mu_0, \Sigma_0)$ ,  $\beta$  se distribuye  $N(\beta_0, \Sigma_\beta)$ , y  $\Sigma$  se distribuye  $W^{-1}(S, n, p)$ . Por lo que nuestra función de verosimilitud adquiere la siguiente forma:

$$Lik(\mu, \mu_0, \beta, \beta_0, \Sigma, \Sigma_0, \Sigma_\beta, S | (r_i)_{i=1}^n) =$$

Luego, expresando a nuestra distribución de interés  $f_{r_i}(r_i)$  como una distribución tipo mezcla normal, con variable de mezcla  $u$ , tenemos que la función de verosimilitud tiene la siguiente forma:

$$\prod_{i=1}^n \int_0^\infty N_p(r_i | \mu, u\beta, u\Sigma) GI(u | \lambda_0, \Psi_0) du N_p(\mu | \mu_0, \Sigma_0) N_p(\beta | \beta_0, \Sigma_\beta) W_{exp}(\Sigma | S)$$

Pro lo que ahora necesitamos conocer también los valores de  $\lambda$  y  $\Psi$  propios de la distribución gaussiana inversa, para ello asumiremos que  $\lambda$  y  $\psi$  son independientes, y además por la gran flexibilidad de la distribución gamma, asumiremos que cada uno de ellos proviene de una distribución gamma con parámetros  $\alpha_\lambda$ ,  $\beta_\lambda$ ,  $\alpha_\Psi$  y  $\beta_\Psi$ , respectivamente, por lo que la dostribución apriori para  $\lambda$  y  $\Psi$  está dada de la siguiente manera:

$$\Pi(\lambda, \Psi) = \Pi(\lambda|\alpha_\lambda, \beta_\lambda)\Pi(\Psi|\alpha_\Psi, \beta_\Psi)$$

Ahora según lo expuesto en el capítulo 3.algo, es conveniente extender la verosimilitud, por lo que ahora tenemos que:

$$Lik(\mu, \beta, \Sigma, (u_i)_{i=1}^n, \lambda, \Psi|(r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi) =$$

$$\prod_{i=1}^n N_p(r_i|\mu, u_i\beta, u_i\Sigma)GI(u_i|\lambda_0, \Psi_0)\Pi(\lambda|\alpha_\lambda, \beta_\lambda)\Pi(\Psi|\alpha_\Psi, \beta_\Psi)$$

$$N_p(\mu|\mu_0, \Sigma_0)N_p(\beta|\mu_\beta, \Sigma_\beta)W_{pxp}^{-1}(\Sigma|S, n, p)$$

Una vez teniendo nuestra función de verosimilitud, ya estamos en condiciones de proponer algún método numérico para encontrar los parámetros de interés; para lo cual, por un lado se utilizará el muestreador de GIBBS para obtener una muestra de los parámetros buscados, mientras que por otro lado se optará por un enfoque bayesiano para tratar la verosimilitud, ya que gracias a este método, podemos aprovechar la información implícita de los datos.

Entonces, con un enfoque bayesiano, será necesario encontrar las distribuciones marginales de los parámetros de interés, para que una vez obtenidas se pueda implementar el muestreador de GIBBS.

Por lo que ahora necesitamos conocer las distribuciones marginales de:

$$\Pi(\mu|\beta, \Sigma, (u_i)_{i=1}^n, \lambda, \Psi, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi)$$

$$\Pi(\beta|\mu, \Sigma, (u_i)_{i=1}^n, \lambda, \Psi, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi)$$

$$\Pi(\Sigma|\mu, \beta, (u_i)_{i=1}^n, \lambda, \Psi, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi)$$

$$\begin{aligned} & \Pi(u_i|\mu, \beta, \Sigma, \lambda, \Psi, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi) \\ & \Pi(\lambda|\mu, \beta, \Sigma, (u_i)_{i=1}^n, \Psi, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi) \\ & \Pi(\Psi|\mu, \beta, \Sigma, (u_i)_{i=1}^n, \lambda, (r_i)_{i=1}^n, \mu_0, \beta_0, \Sigma_\mu, \Sigma_\beta, S, \alpha_\lambda, \beta_\lambda, \alpha_\Psi, \beta_\Psi) \end{aligned}$$

## Distribuciones marginales

### Distribución marginal de $\mu$

Para encontrar la distribución marginal de  $\mu|\beta, \Sigma, (u_i)_{i=1}^n, (r_i)_{i=1}^n$ , a partir de la función de verosimilitud podemos tomar las densidades donde aparece  $\mu$  e identificar su kernel, es decir, de:

$$\begin{aligned} & \prod_{i=0}^n N_p(r_i|\mu, u_i\beta, u_i\Sigma) N_p(\mu|\mu_0, \Sigma_0) = \\ & \prod_{i=1}^n \frac{\exp(-\frac{1}{2}(r_i - \mu - u_i\beta)'(u_i\Sigma)^{-1}(r_i - \mu - u_i\beta))}{(2u_i\Pi)^{p/2}|\Sigma|^{1/2}} \frac{\exp(-\frac{1}{2}(\mu - \mu_0)'\Sigma_0(\mu - \mu_0))}{(2\Pi)^{p/2}|\Sigma|^{1/2}} \end{aligned}$$

tomando el cambio de variable  $r_i' = r_i - u_i\beta$ , nos quedamos con

$$\prod_{i=1}^n \exp(-\frac{1}{2}(\mu - r_i')'(u_i\Sigma)^{-1}(\mu - r_i')) \exp(-\frac{1}{2}(\mu - \mu_0)'\Sigma_0(\mu - \mu_0))$$

Después, trabajando únicamente con

$$\begin{aligned} & \prod_{i=1}^n \exp(-\frac{1}{2}(\mu - r_i')'(u_i\Sigma)^{-1}(\mu - r_i')) = \exp(-\frac{1}{2} \sum_{i=1}^n (\mu - r_i')'(u_i\Sigma)^{-1}(\mu - r_i')) \\ & \exp(-\frac{1}{2} \sum_{i=1}^n (\mu'(u_i\Sigma)^{-1}\mu - 2\mu(u_i\Sigma)^{-1}r_i' - r_i'(u_i\Sigma)^{-1}r_i')) \\ & \alpha \exp(-\frac{1}{2} \sum_{i=1}^n (\mu'(u_i\Sigma)^{-1}\mu - 2\mu(u_i\Sigma)^{-1}r_i')) = \alpha \exp(-\frac{1}{2}(\mu'(U\Sigma)^{-1}\mu - 2\mu(U\Sigma)^{-1}R)) \end{aligned}$$

Donde :

$$U = (\sum_{i=1}^n \frac{1}{u_i})^{-1}, R = \sum_{i=1}^n \frac{r_i}{u_i}$$

Entonces, según el anexo 4 esta parte del kernel de  $\mu$  le corresponda una distribución normal p variada con vector de medias  $R$ , y matriz de varianza covarianza  $U\Sigma$ . Luego, multiplicando el kernel anterior por la densidad de  $\mu$ , y por el anexo 6 se tiene que  $\mu$  se distribuye normal con vector de medias:

$$(R(U\Sigma)^{-1} + \mu_0\Sigma_0^{-1})(U\Sigma^{-1} + \Sigma_0^{-1})^{-1}$$

y matriz de varianza covarianza:

$$U\Sigma + \Sigma_0$$

### Distribución marginal de $\beta$

Ahora mediante un procedimiento similar a lo descrito anteriormente, encontraremos el Kernel de  $\beta$  dado las demás variables de interés. Entonces, de la función de verosimilitud, nos quedamos con la parte donde sólo aparece  $\beta$ , es decir:

$$\prod_{i=0}^n N_p(r_i|\mu, u_i\beta, u_i\Sigma)N_p(\beta|\mu_\beta, \Sigma_\beta)$$

Luego trabajando con el producto de las n variables normales p variadas correspondientes a  $r_i$ , considerando los términos que únicamente dependen de  $\beta$ , y con el cambio de variable  $\mu_i = r_i - \beta$ , tenemos que:

$$\begin{aligned} \prod_{i=1}^n \exp(-\frac{1}{2}(r_i - \mu - u_i\beta)'(u_i\Sigma)^{-1}(r_i - \mu - u_i\beta)) = \\ \exp(-\frac{1}{2} \sum_{i=1}^n ((u_i\beta - \mu_i)'(u_i\Sigma)^{-1}(u_i\beta - \mu_i)) \end{aligned}$$

Nuevamente desarrollando el exponente de la función exponencial, y quedandonos únicamente con la parte que depende de  $\beta$  tenemos que:

$$\begin{aligned} \exp(-\frac{1}{2} \sum_{i=1}^n ((u_i\beta)'(u_i\Sigma)^{-1}(u_i\beta) - 2(u_i\beta)'(u_i\Sigma)^{-1}\mu_i)) = \\ \exp(-\frac{1}{2}(\beta(V\Sigma)^{-1}\beta - 2\beta(V\Sigma)^{-1}VM)) \end{aligned}$$

Donde:

$$V = \frac{1}{\sum_{i=1}^n u_i}, M = \sum_{i=1}^n (r_i - \mu)$$

Por lo tanto, según el anexo 4, el Kernel del producto de las  $n$  variables normales  $p$  variadas correspondientes a  $\beta$  corresponde a una distribución normal con vector de medias  $VM$ , y matriz de varianza covarianza  $V\Sigma$ . Luego, según el anexo 6, juntando el kernel obtenido anteriormente con la densidad de  $\beta$ , se tiene que  $\beta$  se distribuye normal  $p$  variado con vector de medias:

$$(MV\Sigma)^{-1} + \mu'_\beta \Sigma_\beta^{-1} ((V\Sigma)^{-1} + \Sigma_\beta^{-1})^{-1}$$

Y matriz de varianza covarianza:

$$V\Sigma + \Sigma_\beta$$

## Distribución marginal de $\Sigma$

Para encontrar la distribución marginal de  $\Sigma$ , a partir de la función de verosimilitud, podemos tomar las densidades donde aparece  $\sigma$ , y a partir de ahí, identificar el kernel correspondiente, es decir, tomando:

$$\prod_{i=0}^n N_p(r_i | \mu, u_i \beta, u_i \Sigma) W_{p \times p}^{-1}(\Sigma | S, n, p) =$$

Luego trabajando con el producto de las  $n$  variables normales  $p$ -variadas correspondientes a  $r_i$ , considerando los términos que únicamente dependen de  $\Sigma$ , y con el cambio de variable  $(x_i = r_i - \mu - \beta) / \sqrt{u_i}$ , tenemos que:

$$\begin{aligned} \prod_{i=1}^n \exp\left(-\frac{1}{2}(r_i - \mu - u_i \beta)'(u_i \Sigma)^{-1}(r_i - \mu - u_i \beta)\right) \frac{\exp(-\frac{1}{2}tr(\Sigma^{-1}S))}{|\Sigma|^{n/2}} &= \\ \frac{\exp(-\frac{1}{2} \sum_{i=1}^n x_i \Sigma^{-1} \sum_{i=1}^n x_i)}{|\Sigma|^{\frac{n}{2}}} \frac{\exp(-\frac{1}{2}tr(\Sigma^{-1}S))}{|\Sigma|^{\frac{n}{2}}} &= \\ \frac{\exp(-\frac{1}{2}(X\Sigma^{-1}X + tr(\Sigma^{-1}S)))}{|\Sigma|^n} &= \frac{\exp(-\frac{1}{2}(tr(\Sigma^{-1}XX') + tr(\Sigma^{-1}S)))}{|\Sigma|^n} = \end{aligned}$$



$$\frac{\exp(-\frac{1}{2}(tr(\Sigma^{-1}(XX' + S))))}{|\Sigma|^n}$$

Haciendo  $X = \sum_{i=1}^n x_i$ , y ya que  $X\Sigma^{-1}X = tr(\Sigma^{-1}(XX'))$ . La última expresión corresponde al kernel de una distribución Whishart con matriz de escala  $XX' + S$ , y  $2n$  grados de libertad.

## Distribución marginal de $\lambda$

Para encontrar la distribución marginal de  $\lambda$ , a partir de la función de verosimilitud, podemos tomar las densidades donde aparece  $\lambda$ , y a partir de ahí, identificar el kernel correspondiente, es decir, de la siguiente expresión:

$$\prod_{i=1}^n GI(u_i|\lambda, \Psi)\Pi(\lambda|\alpha_\lambda, \beta_\lambda)$$

nos quedamos con:

$$\prod_{i=1}^n \sqrt{\lambda} \exp\left(-\frac{\lambda(u_i - \Psi)^2}{2\Psi^2 u_i}\right) \lambda^{\beta_\lambda - 1} \exp(-\lambda\alpha_\lambda) \lambda$$

$$\lambda^{n/2 + \beta_\lambda - 1} \exp(-X\lambda)$$

donde  $X = \sum_{i=1}^n \frac{(u_i - \Psi)^2}{2\Psi^2 u_i} - \alpha_\lambda$  Por lo que tenemos que el kernel corresponde a una distribución gamma, por lo tanto  $\lambda$  se distribuye gamma con parámetros  $(\frac{n}{2} + \beta_\lambda), X$ .

## Distribución marginal de $\Psi$

Para encontrar la distribución marginal de  $\Psi$ , a partir de la función de verosimilitud, podemos tomar las densidades donde aparece  $\Psi$ , y a partir de ahí identificar el kernel correspondiente, es decir, de la siguiente expresión:

$$\prod_{i=1}^n GI(u_i|\lambda, \Psi)\Pi(\Psi|\alpha_\Psi, \beta_\Psi)$$

nos quedamos con:

$$\prod_{i=1}^n \sqrt{\lambda} \exp\left(-\frac{\lambda(u_i - \Psi)^2}{2\Psi^2 u_i}\right) \Psi^{\beta_\Psi - 1} \exp(-\Psi \alpha_\Psi) \alpha$$

$$\exp\left(-\sum_{i=1}^n \frac{\lambda(u_i - \Psi)^2}{2\Psi^2 u_i}\right) \Psi^{\beta_\Psi - 1} \exp(-\Psi \alpha_\Psi) \alpha$$

El kernel obtenido resulta difícil de identificar, y tal vez sea complicado obtener la densidad condicional de  $\Psi$ , cosa que es necesaria para implementar el muestreador de GIBBS, por lo que se atacará este problema por medio de una herramienta útil en simulación, llamada slice sampler; la ventaja de utilizar el slice sampler es que nos permite obtener una muestra aleatoria de alguna distribución de probabilidad, con tan solo conocer su Kernel, por lo que a continuación se hablará un poco sobre la idea intuitiva que hay detrás del slice sampler.

## Slice Sampler

Los métodos de simulación slice sampler consisten en generar una cadena de markov que converja a la distribución que se desea muestrear. La idea intuitiva de estos métodos consiste en suponer que el vector aleatorio bivariado  $(X, Y)$  se distribuye uniforme en la región que está por debajo de la función de densidad, de la cual obtenríamos una muestra  $(X_0, Y_0)$ , y de esta solo nos quedaríamos con  $X_0$ , siendo esta nuestra variable de interés.

Ahora, para generar una muestra uniforme en dicha región, se utiliza un algoritmo recursivo con el fin de tener una cadena de markov que converga a la muestra deseada.

En términos ilustrativos el procedimiento es como sigue:

1) Supongamos que tenemos el kernel o una densidad de una variable aleatoria univariada  $X$ , de la cual nos interesaría obtener una muestra aleatoria, y que además la gráfica de esta densidad es de la siguiente manera:

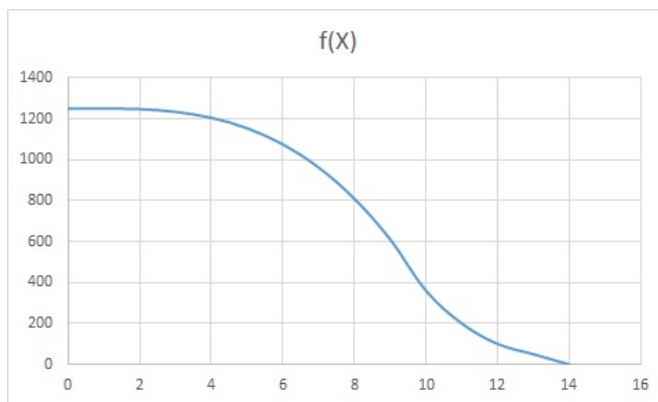


Figura 7:

2) Luego definimos el vector aleatorio bivariado  $(X, Y)$ , el cual suponemos que se distribuye uniforme en la región que está por debajo de la gráfica anterior, y a esta región la denotamos como  $U$ , por lo que  $U$  sería el conjunto de parejas  $(x, y)$  con la propiedad de que  $x$  es menor a  $y$ , siendo  $y = f(x)$ .

Enonces la región de la cual nos interesa obtener una muestra se ve de la siguiente forma:

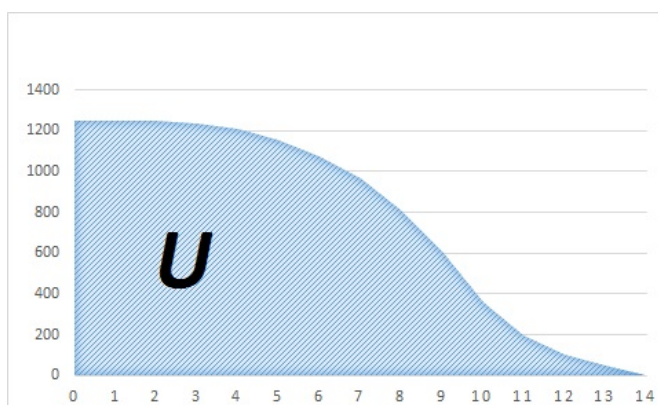


Figura 8:

3) Luego proponemos un valor inicial  $x_0$ , y por el anexo (algo), si un vector aleatorio bivariado se distribuye uniforme en alguna región, entonces para cada valor que tome la variable  $X$ , la distribución marginal de  $Y$  dado  $X = x_0$  se distribuye uniforme dentro del segmento de recta  $L(y) = ((y, x_0)|y \in R)$ , donde  $L(y)$  está dentro de  $U$ . Entonces de esta manera, podríamos generar una realización de  $Y$  dado  $x_0$ , así obtendríamos algún valor  $y_0$ , después fijamos esta valor  $y_0$ , y tendríamos ahora que  $X$  dado  $y_0$  se distribuye uniforme en  $U \cap L(x) = ((x, y_0)|x \in R)$ , por lo que ahora obtendríamos una realización de  $X$  dado  $y_0$  de donde obtenríamos nuevamente un valor  $x_0$ , y así reperiríamos el proceso, pues de esta manera tenemos una cadena de markov que converge a la realización de haber obtenido una muestra uniforme sobre  $U$ , donde solo nos interesa el valor de  $X$ .

En la siguiente gráfica se ilustra el punto 3).

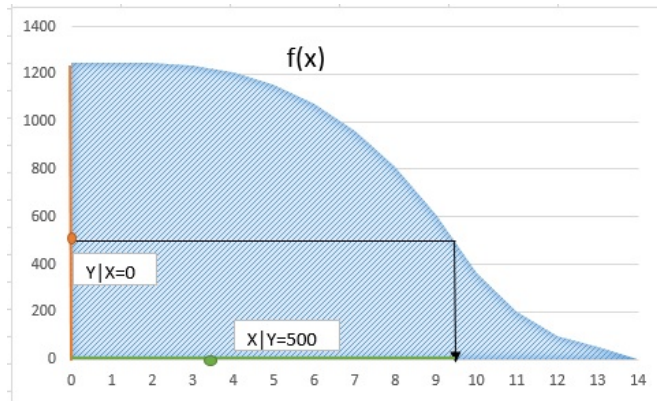


Figura 9: Por simplicidad se asumió que el primer valor que tomó la variable  $X$ , fue 0, luego  $Y$  dado  $x$  igual a cero se distribuye uniforme en el intervalo  $(0,1200)$ , de donde ahora obtenemos una realización de  $Y$  dado  $x$  igual a cero, donde  $Y$  tomó ahora el valor de 500, con esta nueva información actualizamos el valor de  $Y$ , por lo que ahora obtenemos una realización de  $X$  dado  $Y$  igual a 500, y así seguimos actualizando los valores de  $X$  y  $Y$  hasta obtener la muestra deseada.

## Distribución marginal de $u_i$

Para encontrar la distribución marginal de  $u_i$ , a partir de la función de verosimilitud, podemos tomar las densidades donde aparece  $u_i$ , y a partir de ahí identificar el kernel correspondiente, es decir, de la siguiente expresión:

$$\prod_{i=1}^n N_p(r_i | \mu, u_i \beta, u_i \Sigma) GI(u_i | \lambda, \Psi)$$

nos quedamos con:

$$\frac{\exp(-\frac{1}{2}(r_i - \mu - u_i \beta)'(u_i \Sigma)^{-1}(r_i - \mu - u_i \beta))}{(u_i)^{p/2}} \sqrt{\frac{1}{u_i^3}} \exp(-\frac{\lambda(u_i - \Psi)^2}{2\Psi^2 u_i})$$

EL kernel anterior parece muy difícil de identificar, e intentar encontrar la forma anallítica de la distribución condicionada de  $u_i$  puede resultar muy difícil, si no es que imposible, por lo que optaremos por obtener una muestra aleatoria de  $u_i$  por medio del slice sampler.

Una vez descrita la metodología a emplear para inferir los parámetros de nuestra muestra de interés, ya estamos en condiciones de programar el muestreador de GIBBS y emplearlo para estudiar un conjunto de datos, lo cual se llevará acabo en el siguiente capítulo.



# Caso de estudio

## Distribución Gaussiana inversa generalizada

Se dice que  $X$  tiene una distribución gaussiana inversa generalizada  $N(\lambda, \xi, \Psi)$ , si su densidad es de la siguiente forma:

$$f(x) = \frac{\xi^{-\lambda} \sqrt{\xi \Psi}^\lambda x^{\lambda-1} \exp \frac{-1}{2}(\xi x^{-1} + \Psi x)}{2\kappa_\lambda(\sqrt{\xi \Psi})}$$

Donde  $\kappa_{\lambda(\cdot)}$  es una función modificada de Bessel de tercer tipo, y si  $\lambda < 0$ , entonces  $\xi > 0$ ,  $\Psi \geq 0$ ; y si  $\lambda = 0$ , entonces  $\xi > 0$ ,  $\Psi > 0$ , y si  $\lambda > 0$ , entonces  $\xi \geq 0$ ,  $\Psi > 0$ .

Si  $X$  se distribuye  $N(\lambda, \xi, \Phi)$ , entonces su función generadora de momentos es:

$$\begin{aligned} \Phi(it) &= \int_{-\infty}^{\infty} \exp itx \frac{\xi^{-\lambda} \sqrt{\xi \Psi}^\lambda x^{\lambda-1} \exp \frac{-1}{2}(\xi x^{-1} + \Psi x)}{2\kappa_\lambda(\sqrt{\xi \Psi})} dx = \\ &= \frac{\sqrt{\xi \Psi}^\lambda}{2\kappa_\lambda(\sqrt{\xi \Psi})} \frac{2\kappa_\lambda(\sqrt{\xi(2it + \Psi)})}{\sqrt{\xi(2it + \Psi)}^\lambda} \int_{-\infty}^{\infty} \xi^{-\lambda} \frac{\sqrt{\xi(2it + \Psi)}^\lambda}{2\kappa_\lambda(\sqrt{\xi(2it + \Psi)})} \exp \frac{-1}{2}(\xi x^{-1} + (2it + \Psi)x) \end{aligned}$$

Donde la última integral vale uno por ser una densidad  $N(\lambda, \xi, 2it + \psi)$  integrada sobre su soporte, por lo que:

$$\Phi(it) = \frac{\sqrt{\xi \Psi}^\lambda}{\kappa_\lambda(\sqrt{\xi \Psi})} \frac{\kappa_\lambda(\sqrt{\xi(2it + \Psi)})}{\sqrt{\xi(2it + \Psi)}^\lambda}$$

Si  $X$  se distribuye  $N(\lambda, \xi, \Phi)$ , entonces su  $r$ -ésimo momento es:

$$E[X^r] = \int_{-\infty}^{\infty} x^r \frac{\xi^{-\lambda} \sqrt{\xi \Psi}^{\lambda} x^{\lambda-1} \exp \frac{-1}{2} (\xi x^{-1} + \Psi x)}{2\kappa_{\lambda}(\sqrt{\xi \Psi})} dx =$$

$$\frac{x^r 2\kappa_{\lambda+r}(\sqrt{\xi \Psi})}{\sqrt{\xi \Psi}^r 2\kappa_{\lambda}(\sqrt{\xi \Psi})} \int_{-\infty}^{\infty} \frac{\xi^{-\lambda-r} \sqrt{\xi \Psi}^{\lambda+r} x^{\lambda+r-1} \exp \frac{-1}{2} (\xi x^{-1} + \Psi x)}{2\kappa_{\lambda+r}(\sqrt{\xi \Psi})} dx$$

Por lo tanto:

$$E[X^r] = \left(\frac{\xi}{\Psi}\right)^{\frac{r}{2}} \frac{\kappa_{\lambda+r}(\sqrt{\xi \Psi})}{\sqrt{\kappa_{\lambda}(\sqrt{\xi \Psi})}}$$

## Distribución Gaussiana inversa

Se dice que el vector aleatorio  $X$  de dimención  $p$  tiene una distribución Gaussiana inversa  $GI(\Psi, \lambda)$ , si su función de densidad es de la siguiente forma:

$$f_x(X) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x - \Psi)^2}{2\Psi^2 x}\right)$$

Para  $\lambda$ ,  $\mu$  y  $X$  mayores a cero.

## Distribución Hiperbólica Generalizada

Se dice que el vector  $p$  variado  $X$  tiene una distribución hiperbólica generalizada  $GH_p(\lambda, \xi, \Psi, \mu, \Sigma, \beta)$  si su densidad es de la siguiente forma:

$$f(x) = c \frac{\kappa_{\lambda-d/2}(\sqrt{\xi + (x - \mu)' \Sigma^{-1} (x - \mu) (\Psi + \beta' \Sigma^{-1} \beta)})}{\xi + (x - \mu)' \Sigma^{-1} (x - \mu) (\Psi + \beta' \Sigma^{-1} \beta)^{\frac{d}{2} - \lambda}} \exp(x - \mu)' \Sigma^{-1} \beta$$

$$\text{donde } c = \frac{\sqrt{\xi} \lambda^{-\lambda} \Psi^{\lambda} (\Psi + \beta' \Sigma^{-1} \beta)^{\frac{d}{2} - \lambda}}{(2\Pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} \kappa_{\lambda}(\sqrt{\xi \Psi})}.$$

Una característica importante de la distribución hiperbólica generalizada es que puede descomponerse como una variable de mezcla, donde  $X|u$  se distribuye  $N(\mu + u\beta, u\Sigma)$ , y  $u$  se distribuye  $N(\lambda, \xi, \Psi)$ .



$$f(x) = \int_{-\infty}^{\infty} \frac{\exp(x - \mu)' \Sigma^{-1} (x - \mu)}{(2\Pi)^{d/2} |\Sigma|^{1/2} u^{d/2}} \exp - \frac{(x - \mu)' \Sigma^{-1} (x - \mu)}{2u} - \frac{\beta \Sigma^{-1} \beta}{2/u} f(u) du$$

Donde:

$$f(u) = \frac{\xi^{-\lambda} \sqrt{\xi \Psi}^\lambda u^{\lambda-1} \exp \frac{-1}{2} (\xi u^{-1} + \Psi u)}{2\kappa_\lambda(\sqrt{\xi \Psi})}$$

Sea  $a = (x - \mu)' \Sigma^{-1} (x - \mu)$ , y  $b = \beta \Sigma^{-1} \beta$ , entonces de aquí se sigue que:

$$f(x) = \frac{\exp(x - \mu)' \Sigma^{-1} \beta \xi^{-\lambda} \sqrt{\xi \Psi}^\lambda}{(2\Pi)^{d/2} |\Sigma|^{1/2} 2\kappa_\lambda(\sqrt{\xi \Psi})} \int_{-\infty}^{\infty} u^{\lambda-d/2-1} \exp -1/2 (au^{-1}bu + \xi u^{-1} + \Psi u) du$$

Trabajando solamente con el integrando se tiene que:

$$\begin{aligned} \int_{-\infty}^{\infty} u^{\lambda-d/2-1} \exp -1/2 (au^{-1} + bu + \xi u^{-1} + \Psi u) du = \\ \int_{-\infty}^{\infty} u^{\lambda-d/2-1} \exp -1/2 ((a + \xi)u^{-1} + (b + \Psi)u) du \end{aligned}$$

Ahora, sea  $\lambda' = \lambda - 1/2$ ,  $\xi' = a + \xi$ ,  $\Psi' = b + \Psi$ , entonces:

$$\int_{-\infty}^{\infty} u^{\lambda-d/2-1} \exp -1/2 ((a + \xi)u^{-1} + (b + \Psi)u) du = \int_{-\infty}^{\infty} u^{\lambda'-1} \exp -1/2 (\xi' u^{-1} + \Psi' u) du$$

Luego,

$$\int_{-\infty}^{\infty} u^{\lambda'-1} \exp -1/2 (\xi' u^{-1} + \Psi' u) du = \frac{2\kappa_{\lambda'}(\sqrt{\xi' \Psi'})}{\xi'^{-\lambda'} (\sqrt{\xi' \Psi'})^{\lambda'}} \int_{-\infty}^{\infty} \frac{\xi^{-\lambda'} (\sqrt{\xi \Psi})^{\lambda'}}{2\kappa_{\lambda'}(\sqrt{\xi \Psi})} du$$

$$u^{\lambda'-1} \exp -1/2(\xi'u^{-1} + \Psi'u)du$$

Donde la última integral vale uno por ser una distribución  $N(\lambda, \xi, \Psi)$  integrada sobre su soporte, por lo que tenemos que:

$$f(x) = \frac{\exp(x - \mu)' \Sigma^{-1} \beta \xi^{-\lambda} \sqrt{\xi} \Psi^\lambda}{(2\Pi)^{d/2} |\Sigma|^{1/2} 2k_\lambda(\sqrt{\xi} \Psi)} \frac{2k_{\lambda'}(\sqrt{\xi'} \Psi')}{\xi^{-\lambda'} (\sqrt{\xi'} \Psi')^{\lambda'}}$$

Por último, sustituyendo los valores de  $a$ ,  $b$ ,  $\xi$ ,  $\Psi$ , y agrupando algunos términos tenemos que efectivamente  $f(x)$  se distribuye  $GH_p(\lambda, \xi, \Psi, \mu, \Sigma, \beta)$ .

Ahora para encontrar los momentos de la distribución  $GH_p$ , como puede ser expresada como una variable de mezcla en esperanza-varianza, la propiedad 1) nos dice que  $\Phi_X(t) = \exp(it\beta\mu)\Phi_u(it\beta - \frac{1}{2}t\Sigma t)$ , donde  $\Phi_u(\cdot)$  es la función característica del la variable de mezcla  $u$  que se distribuye  $N(\lambda, \xi, \Psi)$ .

La función característica nos permite calcular fácilmente la distribución de una transformación lineal del vector p-variado  $X$ , pues si ahora consideramos el vector aleatorio  $Y = aX + b$ , donde  $a \in M_{r \times p}$ ,  $b \in R^p$  entonces:

$$\begin{aligned} \Phi_Y(t) &= \exp(it\beta\mu)\Phi_X(at) = \exp(it\beta\mu)\Phi_u(iat\beta - \frac{1}{2}at\Sigma at) = \\ &\exp(it(b + \beta\mu))\Phi_u(iat\beta - \frac{1}{2}at\Sigma at) \end{aligned}$$

Lo cual nos indica que  $Y$  se distribuye  $GH_p(\lambda, \xi, \Psi, b + \beta\mu, a\Sigma a', a\beta)$ , y a su vez que.

También a partir de la función característica es fácil ver que la distribución marginal de  $X_i$  es  $GH_p(\lambda, \xi, \Psi, \mu_i, \Sigma_i, \beta_i)$ , donde  $\Sigma_i$  es el  $i$ -ésimo componente del primer renglón de  $\Sigma$ . Las dos propiedades anteriores nos dicen que si  $X$  se distribuye  $GH_p$ , entonces es cerrado bajo convoluciones.

# Anexos

## 0.1. Distribución normal p variada

Se dice que el vector aleatorio  $X \in R^p$  tiene una distribución normal p-variada  $N_p(X|\mu, \Sigma)$  si su densidad es de la siguiente forma:

$$f(X|\mu, \Sigma) = \frac{1}{(2\Pi)^{p/2}|\Sigma|^{1/2}} \exp(-1/2(X - \mu)'\Sigma^{-1}(X - \mu))$$

Donde  $\mu \in R^p$  es el vector de medias, y  $\Sigma \in M_{p \times p}$  es la matriz de varianza-covarianza.

## 0.2. Distribución Wishart

Se dice que una matriz  $S \in M_{p \times p}$ , simétrica y positiva definida, se distribuye Whisart con n grados de libertad,  $W(\Sigma, n, p)$ , si su densidad es de la siguiente forma:

$$f_S(S) = (2^{np/2} \pi^{p(p-1)/4} \prod_{i=1}^p \gamma(\frac{n+1-i}{2}))^{-1} \frac{|S|^{(n-p-1)/2}}{|\Sigma|^{n/2}} \exp(-\frac{1}{2}tr(\Sigma^{-1}S))$$

Donde  $\Sigma \in M_{p \times p}$  es simétrica y positiva definida, y se le conoce como matriz de escala, y con  $n$  mayor a  $p$ .

## 0.3. Distribución inversa Wishart

Se dice que una matriz  $G \in M_{p \times p}$ , simétrica y positiva definida, se distribuye whishart inversa con n grados de libertad,  $W^{-1}(K, n, p)$ , si su

densidad es de la siguiente forma:

$$f_G(G) = (2^{(n-p-1)p/2} \pi^{p(p-1)/4} \prod_{i=1}^p \gamma(\frac{n-p-i}{2}))^{-1} \frac{|K|^{(n-p-1)/2}}{|G|^{n/2}} \exp(-\frac{1}{2} \text{tr}(G^{-1}K))$$

Donde a  $K$  se le conoce como matriz de escala, y es simétrica y definida positiva. Es importante notar que si alguna matriz aleatoria  $\Sigma \in M_{p \times p}$  se distribuye Whisart con matriz de escala  $S$ , y con  $n$  grados de libertad, entonces  $\Sigma^{-1}$  se distribuye inversa whisart con matriz de escala  $S^{-1}$ , y con  $n + p + 1$  grados de libertad.

## 0.4. Distribución tipo mezla normal p variada

Se dice que el vector aleatorio  $X \in R^p$  con densidad  $f(x)$  puede ser expresado como una distribución de mezcla normal  $N(X|\mu, \beta, \Sigma)$  con variable de mezcla  $u$  con densidad  $f(u)$  y soporte en  $R^+$ , si:

$$f(X|\mu, \beta, \Sigma) = \int_0^\infty N_p(X|\mu, u\beta, u\Sigma) f_u(u) du$$

Donde,  $N_p(X|\mu, u\beta, u\Sigma)$  es una distribución normal p-variada con vector de medias  $\mu + u\beta$ , y matriz de varianza-covarianza  $u\Sigma$

## 0.5. Covarianza de un vector aleatorio p variado

Se define la covarianza de un vector aleatorio p variado como:

$$COV(X) = E((X - \mu)(X - \mu)')$$

Siempre y cuando el vector de medias (del vector aleatorio  $X$ )  $\mu$  exista.

## 0.6. Covarianza de una distribución tipo mezla

Sea  $X$  dado  $u$  una distribución tipo mezcla de dimención  $p$ , y  $u$  una variable de mezcla con soporte en  $R_+$ , entonces la covarianza de  $X$  se puede calcular como:

$$COV(X) = E_u(COV(X|u)) - COV_u(E(X|u))$$

Demostración: De la definición de  $COV(X)$  tenemos que:

$$\begin{aligned}
COV(X) &= E_x((X - \mu)(X - \mu)') = E_x(XX' - 2X\mu + \mu\mu') = \\
&= E_x(XX') - 2E(X)\mu + \mu\mu' = E_x(XX') - E_x(X)E_x(X)' = \\
&= E_x(XX') - E(X)_x E_x(X)' + E(X)_x E(X)'_x - 2E(X)_x E(X)'_x + E(X)_x E(X)_x = \\
&= E_u(E_x(XX'|u) - E_x(X|u)E_x(X|u)') + E_u(E_x(X|u)E_x(X|u)') - \\
&= 2E_u(E_x(X|u)E(X)') + E_u(E(X|u))E_u(E(X|U)) = \\
&= E_u(E_x(XX'|u) - 2E_x(X|u)E_x(X|u)' + E_x(X|u)E_x(X|u)') + \\
&= E_u(E_x(X|u)E_x(X|u)' - 2E_x(X|u)E_u(E_x(X|u)) + E_uE_x(X|u)E_uE_x(X|u)') = \\
&= E_u(E_x(XX' - 2XE_x(X|u) + E_x(X|u)E_x(X|U)) + \\
&= E_u((E_x(X|u) - E_u(E_x(X|u)))(E_x(X|u) - E_u(E_x(X|u)))') = \\
&= E_u((E_x(X) - E_x(X|u))(E_x(X) - E_x(X|u))') + COV_u(E_x(X|u)) = \\
&= E_u(COV_x(X|u)) + COV_u(E_x(X|u)).
\end{aligned}$$

## 0.7. Kernel de una distribución normal p variada

El kernel un vector aleatorio es la parte de la función de densidad que únicamente depende de dicho vector, y a su vez nos permite identificar de que familia proviene la distribución. Por ejemplo, en el caso de la distribución normal p variada

$$f(X|\mu, \Sigma) = \frac{1}{(2\Pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x'\Sigma^{-1}x - x'\Sigma^{-1}\mu - \mu'\Sigma^{-1}x + \mu'\Sigma^{-1}\mu)\right)$$

Tenemos que el Kernel correspondiente es:

$$\exp\left(-\frac{1}{2}(x'\Sigma^{-1}x - 2x'\Sigma^{-1}\mu)\right)$$

donde  $\mu'\Sigma^{-1}x = x'\Sigma^{-1}\mu$ , ya que es una forma cuadrática.

## 0.8. Kernel de una distribución Wishart

Análogamente al caso normal p variado, si nos concentramos en la parte de la densidad Wishart, con matriz de escala  $S$  y con  $n$  grados de libertad, que únicamente depende de  $\Sigma$ , tenemos que el correspondiente kernel es:

$$\frac{\exp(-\frac{1}{2}\text{tr}(\Sigma^{-1}S))}{|\Sigma|^{\frac{n}{2}}}$$

## 0.9. Kernel del producto de $n$ distribuciones normales p variadas con mismo vector de medias y misma matriz de varianza covarianza

Supongamos que  $X_i$  se distribuye normal p variada con vector de media  $\mu$  y matriz de varianza covarianza  $\Sigma$ , entonces la función de interés es de la siguiente manera:

$$\prod_{i=1}^n N(X_i|\mu, \Sigma) = \prod_{i=1}^n \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp(-\frac{1}{2}(X_i - \mu)' \Sigma^{-1}(X_i - \mu))$$

De aquí, trabajando únicamente con el exponente de la función exponencial tenemos que:

$$\prod_{i=1}^n \exp(-\frac{1}{2}(X_i - \mu)' \Sigma^{-1}(X_i - \mu)) = \exp(-\frac{1}{2} \sum_{i=1}^n (X_i - \mu)' \Sigma^{-1}(X_i - \mu)) =$$

$$\exp(-\frac{1}{2} \sum_{i=1}^n (X_i' \Sigma^{-1} X_i - 2X_i' \Sigma^{-1} \mu + \mu' \Sigma^{-1} \mu)) = \exp(-\frac{1}{2} (X' \Sigma^{-1} X - 2X' \Sigma^{-1} \mu + n\mu' \Sigma^{-1} \mu))$$

Donde la  $j$ -ésima coordenada del vector  $X$  es  $\sum_{i=1}^n X_{i,j}$ . Por último, como sólo nos interesan los términos donde aparece  $X_i$ , llegamos a que el kernel es:

$$\exp(-\frac{1}{2}(X' \Sigma^{-1} X - 2X' \Sigma^{-1} \mu))$$

## 0.10. Kernel del vector de medias de una distribución normal p variada multiplicada por la distribución del vector de medias

En este caso tenemos que  $X$  se distribuye normal con vector de medias  $\mu$  y matriz de varianza covarianza  $\Sigma$ , mientras que  $\mu$  se distribuye normal con vector de medias  $\mu_0$  y matriz de varianza covarianza  $\Sigma_0$ , y ahora nos interesa conocer el Kernel correspondiente a  $\mu$ , entonces tendríamos que el producto de las funciones de densidad es:

$$f_X(X|\mu, \Sigma) f_\mu(\mu|\mu_0, \Sigma_0) = \frac{\exp(-1/2(X - \mu)' \Sigma^{-1} (X - \mu))}{(2\Pi)^{p/2} |\Sigma|^{1/2}} \frac{\exp(-1/2(\mu - \mu_0)' \Sigma_0^{-1} (\mu - \mu_0))}{(2\Pi)^{p/2} |\Sigma_0|^{1/2}}$$

Luego de cada función de densidad tomamos lo que dependa de  $\mu$ , para así obtener sus respectivos kerneles según el anexo 4, lo cual implica que:

$$\begin{aligned} \text{Kernel} &= \exp(-\frac{1}{2}(\mu' \Sigma^{-1} \mu - 2x' \Sigma^{-1} \mu)) \exp(-\frac{1}{2}(\mu' \Sigma_0^{-1} \mu - 2\mu_0' \Sigma_0^{-1} \mu)) = \\ &= \exp(-\frac{1}{2}(\mu' \Sigma^{-1} \mu - 2x' \Sigma^{-1} \mu + \mu' \Sigma_0^{-1} \mu - 2\mu_0' \Sigma_0^{-1} \mu)) = \\ &= \exp(-\frac{1}{2}(\mu' (\Sigma^{-1} + \Sigma_0^{-1}) \mu - 2(x' \Sigma^{-1} + \mu_0' \Sigma_0^{-1}) \mu)) = \\ &= \exp(-\frac{1}{2}(\mu' (\Sigma^{-1} + \Sigma_0^{-1}) \mu - 2(x' \Sigma^{-1} + \mu_0' \Sigma_0^{-1}) (\Sigma^{-1} + \Sigma_0^{-1})^{-1} (\Sigma^{-1} + \Sigma_0^{-1}) \mu)) \end{aligned}$$

De aquí se tiene que  $\mu$  se distribuye normal p variada con matriz de varianza covarianza  $\Sigma + \Sigma_0$ , y vector de medias  $(x' \Sigma^{-1} + \mu_0' \Sigma_0^{-1}) (\Sigma^{-1} + \Sigma_0^{-1})^{-1}$





# Bibliografía