20321
1

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

ESCUELA NACIONAL DE ESTUDIOS
PROFESIONALES "ACATLAN"

MODELOS BAYESIANOS UNIPARAMETRICOS USUALES

T E S I N A

QUE PARA OBTENER EL TITULO DE :

A C T U A R I O

P R E S E N T A :

ANGELICA AGUIRRE ULLOA

ASESOR: FIS. MAT. JORGE LUIS SUAREZ MADARIAGA

NOVIEMBRE, 2003.

3

Gracias a Dios:

Porque en el transcurso de toda mi vida me ha colmado de tantas bendiciones y momentos maravillosos por los cuales estaré infinitamente agradecida. Por permitirme estar rodeada de personas que han dejado algo muy especial en mí. Porque siempre ha sido y será el poder que tendré para realizar, enfrentar y superar cualquier cosa, por difícil que sea.

Gracias a mis padres:

Porque el gran amor que han puesto en mí y la gran atención por mi formación han logrado un inmenso crecimiento en mi alma y mi ser, por sus grandes sacrificios y esfuerzos. Porque han llenado mi vida de dicha y felicidad. Por fortalecer de sobremanera aquellos lazos que nos unirán para siempre.

Gracias a mis maestros:

Por haber sembrado en mí el hábito de la tenacidad y la dedicación siempre con la visión de llegar hasta el final. Por brindarme la preparación que me dieron para enfrentar el mundo que hoy se presenta. Gracias por haber predicado con su ejemplo y experiencia las cosas que realmente vale la pena formar en un alumno, tanto en el sentido académico como en el sentido ético.

Gracias a todos aquellos que creen en mí.

♡

Índice general

Introducción **7**

1. Elementos de estadística bayesiana **9**
1.1. Modelo general 12
1.2. Elementos de teoría de decisiones y utilidad 15
1.2.1. Elementos Básicos de un Problema de Decisión 16
1.2.2. Criterio de la utilidad máxima esperada 18
1.3. Inferencia estadística 19
1.3.1. Contraste de hipótesis 20
1.3.2. Estimación puntual y estimación por regiones 21
1.3.3. Predicción puntual y predicción por regiones 23

2. Distribuciones probabilísticas **25**
2.1. Distribuciones discretas univariadas 26
2.2. Distribuciones continuas univariadas 35
2.3. Distribuciones discretas multivariadas 49
2.4. Distribuciones continuas multivariadas 52

3. Modelos: Una interacción matemática entre estadística bayesiana y distribuciones de probabilidad **61**
3.1. Modelo Bernoulli 62
3.2. Modelo Poisson 66
3.3. Modelo Binomial Negativa 69
3.4. Modelo Exponencial 71
3.5. Modelo Uniforme 73
3.6. Modelo Normal (con media μ conocida) 76
3.7. Modelo Normal (con precisión λ conocida) 79

Conclusiones 85

Bibliografía 87

A. Propiedades de la función indicadora 89

Introducción

La estadística en un ambiente de metodologías bayesianas, cuyas bases datan desde hace más de dos siglos, es una herramienta que día con día ha tomado mayor fuerza tanto en el campo de la investigación como en ciertas aplicaciones, por ejemplo, para el estudio de datos longitudinales (a través del tiempo), áreas como las ciencias de la salud, de la educación, medicina y algunas de las ciencias sociales, entre otras.

La característica primordial que hace del ánalisis bayesiano una teoría diferente de las demás es la existencia de incertidumbre en la variable de estudio y, en el o los parámetros que definen el comportamiento de dicha variable; de esta manera, a cada una de estas incertidumbres se les asigna una distribución de probabilidad, con el fin de tener las herramientas indicadas para realizar inferencia estadística y pronósticos referentes al problema en cuestión.

El objetivo del presente trabajo de investigación es proporcionar una introducción útil a las herramientas adecuadas para aprender y consultar las técnicas básicas utilizadas en el enfoque estadístico bayesiano así como mostrar el procedimiento específico que antecede a los modelos bayesianos uniparamétricos de mayor uso, basado en una teoría matemática sustentable.

El primer capítulo trata acerca de los antecedentes del enfoque estadístico bayesiano y describe de una manera clara, precisa y digerible la teoría fundamental que genera la creación, así como actualización, de los modelos bayesianos basados en un modelo general que describe la aleatoriedad, tanto de la variable de interés como de sus propios parámetros. También se especifica la instrucción para realizar inferencia estadística establecida por herramientas pertenecientes a teoría de decisiones y, en particular, al criterio de la utilidad máxima esperada.

El segundo capítulo proporciona una recopilación útil de distribuciones de probabilidad con sus respectivas propiedades y características que permiten identificar a cada una de ellas.

El tercer capítulo presenta el desarrollo detallado de ciertos casos especiales referentes a modelos bayesianos que constan de un solo parámetro aleatorio y que tienen mayor recurrencia y utilidad en circunstancias reales. Dicho capítulo tiene como principal propósito el plasmar la aplicación de metodologías bayesianas sobre modelos ya creados, pero no explicados minuciosamente ante el lector.

Capítulo 1

Elementos de estadística bayesiana

A lo largo del tiempo, el ser humano siempre se ha preocupado por saber qué es lo que sucederá a futuro ante cualquier circunstancia o fenómeno. Esta incertidumbre ha propiciado el desarrollo y enriquecimiento de diferentes teorías y metodologías que tienen como fin dominar la aleatoriedad que se presenta a cada momento, o bien, influir en ella.

La técnica estadística conocida como estadística bayesiana es un enfoque alternativo cuyo nombre proviene del Matemático Thomas Bayes.

Aunque no se ha encontrado registro de la fecha exacta de su nacimiento, se sabe que **Thomas Bayes** nació en Londres, Inglaterra, en 1702 y murió el 7 de abril de 1761. Su educación fué privada, un hecho particular para el hijo de un ministro presbiteriano de aquellos tiempos. Bayes fué ordenado ministro presbiteriano y más tarde fué nombrado pastor en Tumbridge Wells (Kent, Inglaterra). Teólogo, matemático y miembro de la Royal Society desde 1742. Fué el primero en utilizar la probabilidad inductivamente y establecer una base matemática para la inferencia probabilística (la manera de calcular, a partir de la frecuencia con la que un acontecimiento ocurre, la probabilidad de que ocurrirá en el futuro). Abordó el problema de las causas a través de los efectos observados y enunció el teorema que lleva su nombre en la publicación Essay towards solving a problem in the doctrine of chances, en 1763; lo cual actualmente, es la base para calcular la probabilidad de la validez de una proposición, o bien, de cierto evento, partiendo de la estimación de probabilidad previa y las evidencias relevantes más recientes.

Esto concede la noción de poder medir la incertidumbre por medio del uso de la probabilidad.

La **teoría de probabilidades** estudia los fenómenos aleatorios o libres de determinación, y los métodos de análisis de éstos, cualquiera que sea el área en que se presenten.

La **probabilidad** representa el riesgo o la posibilidad de que ocurra cierto evento.

Un **fenómeno aleatorio** (fortuito o al azar) es un fenómeno empírico que se caracteriza por la propiedad de que al ser observado bajo determinado conjunto de condiciones, no siempre se obtiene el mismo resultado, de manera que no existe regularidad determinística; pero, al obtener una sucesión muy larga de observaciones realizadas al azar, se acerca a un valor límite estable a medida que el número de observaciones tiende a infinito. Este valor límite de la frecuencia relativa es el que se conoce como **probabilidad del evento aleatorio.**

Sin embargo, al estudiar cierto fenómeno aleatorio, la probabilidad de éste dependerá de dos situaciones importantes:

- El evento incierto
- Las condiciones bajo las cuales se está considerando dicho evento[1]

Al considerar las condiciones bajo las cuales puede tener lugar el fenómeno aleatorio, la probabilidad de ocurrencia es probabilidad condicional debido a que el valor de esta varía en función del conocimiento de determinada información relativa al suceso. Por esta razón, se afirma que **la probabilidad es siempre condicional**, es decir, si se tienen dos eventos A y B; por probabilidad condicional del evento B,dado el evento A, denotada por

$$p\left[B|A\right]$$

se entiende intuitivamente, la probabilidad de que B ocurra bajo la suposición de que A ha ocurrido. En otras palabras, representa la reevaluación de la probabilidad de B bajo la información adicional de que A ha ocurrido, que cumple con la siguiente igualdad

$$p\left[B|A\right] = \frac{p\left[A \cap B\right]}{p\left[B\right]}$$

[1] Si se desea mayor explicación se puede consultar en la revista The Statistician (2000) Vol. 49, págs. 293-337.

y bien, el teorema o regla de Bayes es una simple consecuencia de la definición axiomática de la probabilidad condicional de un suceso dado otro.

Por lo tanto, un mismo evento puede tener diferentes probabilidades de ocurrencia condicionado bajo diferentes circunstancias.

Las diferentes perspectivas que genera la probabilidad condicional están estrechamente relacionadas con la apreciación subjetiva.

Una probabilidad, o bien, **apreciación subjetiva** es una evaluación que una persona, que toma decisiones, hace acerca de la verosimilitud relativa de que ocurra un evento incierto[2] , o sea, representa las "apuestas" que se hacen sobre la ocurrencia de ese evento. Tales apreciaciones son sumamente personales, tanto que, al igual que las diferentes perspectivas en la probabilidad condicional, dos individuos pueden asignar diferentes probabilidades subjetivas al mismo evento debido a que la incertidumbre de uno es diferente a la del otro. No obstante, estas probabilidades subjetivas pueden aprovecharse en la toma de decisiones, de la misma manera que las probabilidades objetivas, ya que en dichas probabilidades se incluye tanto el conocimiento que cierta persona tenga con respecto al evento, como el juicio hecho sobre el riesgo de interés.

Un elemento cardinal con que predominantemente opera el enfoque bayesiano es el manejo subjetivo del concepto de probabilidad.

El interés por el teorema de Bayes trasciende especialmente cuando se amplía a otro contexto en el que la probabilidad no se entiende exclusivamente como la frecuencia relativa de un suceso a largo plazo, sino como el grado de convicción personal acerca de que el suceso ocurra o pueda ocurrir (definición subjetiva de la probabilidad).

Afirmaciones del tipo "es muy probable que el partido X gane las próximas elecciones", "es improbable que cierta persona haya sido quien llamó por teléfono" o "es probable que se encuentre un tratamiento eficaz para el SIDA en los próximos 5 años", normales en el lenguaje común, no pueden cuantificarse formalmente; resultan ajenas, por tanto, a una metodología que se desenvuelva en un marco frecuentista.

Una cuantificación sobre base subjetiva resulta, sin embargo, familiar y fecunda para el enfoque bayesiano. Al admitir un manejo subjetivo de la probabilidad, el analista bayesiano puede emitir juicios de convicción al respecto, tanto antes como después de haber hecho observaciones.

---

[2]Se podría ser más preciso y definir la probabilidad subjetiva en términos de las preferencias de los responsables de la toma de decisiones ante situaciones hipotéticas. Sin embargo, para el presente estudio, la definición intuitiva puede considerarse suficiente.

Si se desea consultar de manera detallada, véase el capítulo 5 del libro Howard Raiffa, Decision Analysis, Ed. Addison-Wesley, 1968.

En su versión mas elemental y en este contexto, el **teorema de Bayes** asume la forma siguiente:

Sean C_1, C_2, ..., C_n n eventos mutuamente exclusivos, y sea B un evento acerca del cual se conocen las probabilidades condicionales $p\left[B|C_i\right]$ de B, dado C_i, y también las probabilidades absolutas $p\left[C_i\right]$. Entonces se puede calcular la probabilidad condicional $p\left[C_i|B\right]$ de cualquiera de los eventos C_i dado B por medio de la fórmula:

$$p\left[C_i|B\right] = \frac{p\left[B|C_i\right]p\left[C_i\right]}{\sum_{j=1}^{n} p\left[B|C_j\right]p\left[C_j\right]}$$

Si los eventos C_i se les llama "causas", entonces la fórmula de Bayes puede considerarse como una fórmula de la probabilidad de que el evento B, que ha ocurrido, sea el resultado de la causa C_i. De esta manera, dicha fórmula se ha interpretado como fórmula de la probabilidad de "causas" o "hipótesis".

En la práctica, la dificultad que surge con esta interpretación es que en muchos casos no se conocen las probabilidades, sobre todo las probabilidades incondicionales $p\left[C_i\right]$ de las "causas", debido a que es difícil encontrar una probabilidad de estas características, sin ningún factor que pueda influir en ellas para ser totalmente incondicionadas.

1.1. Modelo general

Con respecto a la filosofía que afirma describir a la incertidumbre exclusivamente en términos de su propia probabilidad, implementa la idea de requerir la construcción de una distribución de probabilidad para todos los elementos inciertos del fenómeno estudiado que, en su especificación total sería el modelo. Dicho modelo define la percepción de la realidad y está fundamentado en el método bayesiano que consiste escencialmente en lo siguiente:

1. Suponer que el parámetro o parámetros son aleatorios.

2. Describir el comportamiento aleatorio del parámetro mediante una función de probabilidad $p\left[\theta\right]$ llamada **distribución de probabilidad a priori o inicial.**

La distribución a priori refleja el conocimiento que se tiene acerca del parámetro o parámetros, es decir, está formada por toda la información relevante que de cierta manera describa el comportamiento real y disminuya la incertidumbre que genera el fenómeno estudiado.

La especificación de la distribución a priori o inicial puede ser totalmente subjetiva, o bien, puede ser establecida de acuerdo al ingenio que el analista tenga para utilizar conjuntamente toda la información disponible[3].

3. Si el modelo es X con una distribución de probabilidad $p\left[X|\theta\right]$, interpretar dicha distribución como la distribución condicional de X dado que este parámetro aleatorio tome el valor particular θ.

   La probabilidad condicional de los datos dado el valor del parámetro describe, de cierta manera, la variabilidad natural que se presenta en potenciales repeticiones del experimento.

Al tener ambas distribuciones de probabilidad anteriormente mencionadas y utilizando el concepto de probabilidad condicional se obtiene la **distribución de probabilidad predictiva inicial** $P\left[X\right]$ que, resulta ser de la siguiente manera:

$$\int_{\Theta} p(X|\ \theta)p(\ \theta)d\ \theta$$

en caso de que θ sea una variable aleatoria continua, o bien,

$$\sum_{\theta\in\Theta} p(X|\ \theta)p(\ \theta)$$

si θ es variable aleatoria discreta, siendo Θ el conjunto de todos los posibles valores que puede tomar el parámetro.

$P\left[X\right]$ describe el comportamiento de las observaciones futuras basado en la información contenida en $p\left[X|\theta\right]$ y $p\left[\theta\right]$, más sin en cambio es necesario

[3]La especificación de la distribución inicial puede ser totalmente subjetiva o totalmente no informativa. Se puede especificar a través de formas funcionales, familia exponencial, familias conjugadas, información de Fisher, aproximación asintótica normal, reparametrizaciones y modelos jerárquicos; que para mayor detalle se puede consultar en los capítulos 4, 5, 6, 7, 8 y 9 del libro Estadística, Teoría y Métodos de la Facultad de Matemáticas, Ediciones Universidad Católica de Chile que pertenece a la Colección de Textos Universitarios (1ra. ed., abril de 1995).

cuestionar, mejorar y actualizar la efectividad que pudiera tener la distribución inicial, por lo que se recurre a obtener una **distribución de probabilidad a posteriori o final** $P\left[\theta|\mathbf{x}\right]$ que consiste en colectar datos observados[4] , o bien, tomar una muestra aleatoria[5] del fenómeno estudiado e incorporar nueva información al modelo.

Utilizando la regla de Bayes dicha distribución de probabilidad está dada por

$$p\left[\theta|\mathbf{x}\right] = \frac{p(\mathbf{x},\theta)}{p(\mathbf{x})} = \frac{p(X|\ \theta)p(\ \theta)}{\int_{\Theta} p(X|\ \theta)p(\ \theta)d\ \theta}$$

si θ es una variable aleatoria continua, o bien,

$$p\left[\theta|\mathbf{x}\right] = \frac{p(\mathbf{x},\theta)}{p(\mathbf{x})} = \frac{p(X|\ \theta)p(\ \theta)}{\sum\limits_{\theta\in\Theta} p(X|\ \theta)p(\ \theta)}$$

si θ es una variable aleatoria discreta.

Cada vez que se tenga nueva información muestral se podrá realizar el mismo procedimiento y hacer iteraciones. Para cada iteración habrá una distribución a priori y una a posteriori. La distribución a posteriori para la primera iteración será la distribución a priori de la segunda, y así sucesivamente, continuando en una disposición ordenada en cadena.

En caso de que no se cuente con información muestral se podrá utilizar la distribución de probabilidad inicial $p\left[\theta\right]$ para poder calcular el valor del parámetro o parámetros de interés.

De igual manera sucede con la distribución predictiva inicial: con el fin de ejercer una actualización sobre ésta, se puede incorporar información muestral dando lugar a la **distribución de probabilidad predictiva a posteriori o final** que será diferente para cada resultado muestral. Dicha distribución de probabilidad está dada por

$$P\left[X|\mathbf{x}\right] = \int_{\Theta} p(X|\ \theta)p(\ \theta|\mathbf{x})d\ \theta$$

[4] La recolección de datos puede referirse a una muestra aleatoria, o si es permitido, a todos los datos que se tengan sobre el fenómeno aleatorio.

[5] Una muestra aleatoria es el conjunto de medidas repetidas $(x_1, x_2, ..., x_n)$ con respecto a la variable aleatoria X, en la cual sus elementos $x_i, i = 1, ..., n$, son variables aleatorias independientes e idénticamente distribuidas.

en caso de que θ sea una variable aleatoria continua, o bien,

$$P\left[X|\mathbf{x}\right] = \sum_{\theta\in\Theta} p(X|\,\theta)p(\,\theta|\mathbf{x})$$

si θ es variable aleatoria discreta.

En caso de no contar con información muestral, entonces, la herramienta predilecta para el establecimiento de predicciones será la distribución predictiva a priori o inicial $P\left[X\right]$.

El modelo probabilístico especificado es la base para enfrentar cualquier problema desde un punto de vista bayesiano y, puede ser utilizado y actualizado aún con información empírica o subjetiva, sin perder la versatilidad que lo caracteriza con herramientas cien por ciento probabilísticas.

1.2. Elementos de teoría de decisiones y utilidad

A partir de un enfoque bayesiano, la **teoría de decisiones** es una herramienta muy valiosa en el estudio de cierto fenómeno que causa el interés de adquirir un dominio sobre él, ya que proporciona gran ayudantía en el desarrollo de procesos lógicos para la toma de decisiones bajo condiciones de incertidumbre.

Por esta razón, la inferencia estadística conveniente para el estudio del fenómeno o evento se puede desarrollar bajo las condiciones con las que se caracterizan dichos procesos lógicos y se puede resolver como un problema de teoría de decisiones.

El análisis de decisiones es la disciplina que consiste en evaluar alternativas complejas en términos de valores (habitualmente monetarios) y de incertidumbre, proporcionando información sobre las diferencias entre las alternativas definidas y sugerencias de nuevas y mejores alternativas. Es un proceso que le permite al decisor seleccionar una decisión entre un conjunto de opciones posibles cuando existe incertidumbre con respecto al futuro, con el fin de optimizar los próximos resultados en términos de algún tipo de criterio de decisión numérico.

Los **objetivos de la teoría de decisiones** se fundamentan en la idea que especifica a la complejidad de los eventos de interés presentados a cada momento junto con la cantidad de información, incertidumbre y riesgo, como elementos que requieren un marco racional para influir directamente en la toma de decisiones, incorporando orientación, información, discernimiento y estructura al proceso, de tal manera que las decisiones puedan ser mejores y más racionales.

1.2.1. Elementos Básicos de un Problema de Decisión

Un problema de decisión es una estructura determinada por los siguientes elementos:

1. Conjunto de alternativas posibles de decisión, siendo a_i una acción, miembro del conjunto $\mathcal{A}$, que puede ser adoptada por el decisor.

   El conjunto de opciones dede ser:

   Exhaustivo ; es decir, que se agoten todas las posibles acciones que en principio parezcan razonables.

   Excluyente ; es decir, que la elección de una acción $a_i \in \mathcal{A}$, debe excluir a cualquier otra acción posible.

   Una buena decisión depende de un conjunto lo más completo posible que cuente con las alternativas más relevantes y trascendentales.

   Los criterios que se utilizan para restringir el conjunto de acciones que se pueden tomar son los siguientes:

   - Una acción a_i domina a otra acción $a_{i'}$, si:
     - Su consecuencia[6] asignada es mejor o igual que la consecuencia asignada de $a_{i'}$ bajo el mismo escenario.
     - Existe algún estado de la Naturaleza en que la consecuencia asignada a la acción a_i es estrictamente mejor que la consecuencia asignada a la acción $a_{i'}$.

[6] Se hablará más adelante de las consecuencias.

- **Una acción es admisible si no existe otra acción que la domine y viceversa.**
- **Una acción es inadmisible si existe al menos otra acción que la domina y viceversa.**

2. Álgebra de eventos (futuros) posibles[7]

   llamados estados de la Naturaleza, es decir, un conjunto de escenarios posibles.

   Las circunstancias en las cuales se toma una decisión se llaman estados de la Naturaleza y, se identifican y agrupan en el conjunto $\mathcal{E}$ cuyos miembros se denotan como E_j, $j \in \mathbb{N}$. El conjunto $\mathcal{E}$ es un grupo de conjuntos mutuamente excluyentes por lo que solo puede ocurrir un estado de la Naturaleza bajo un mismo criterio de especificidad de estados.

3. Conjunto C de consecuencias, que son los hechos que proceden necesariamente de las acciones; que pueden surgir al haber elegido una acción $a \in \mathcal{A}$ bajo la ocurrencia de un evento $E \in \mathcal{E}$.

4. La existencia de un decisor responsable individual (o bien, conjunto de decisores que lleguen a un mismo acuerdo) que pretenda definir sus preferencias entre las diferentes acciones, elementos de $\mathcal{A}$.

   Para el decisor es desafiante la tarea de comparar varios cursos de acción bajo diferentes estados de la Naturaleza y finalmente seleccionar la acción que se va a realizar; es por esto que, las dificultades de la toma de decisiones están representadas por la complejidad de las alternativas de decisión.

   La capacidad que tiene un decisor de procesar información limitada es un factor de exigencia ya cuando se consideran las implicancias de un solo curso de acción, sin embargo en muchas decisiones, se deben visualizar y comparar las implicancias de varios cursos de acción. Además, hay factores desconocidos que se inmiscuyen en la situación problemática y rara vez se

[7] Se dice que $\mathcal{E}$ es un álgebra de subconjuntos de Ω si:

i) $\mathcal{E}$ es no vacío

ii) Para toda E perteneciente a $\mathcal{E}$, el complemento de E tambien pertenece a $\mathcal{E}$

iii) Si $E_1, E_2, \ldots, E_n$ pertenecen a $\mathcal{E}$, entonces la unión de $E_1, E_2, \ldots, E_n$ también pertenece a $\mathcal{E}$.

Sin embargo, usualmente los eventos reales se presentan ante un solo escenario o estado E y cabe aclarar que si el espacio de estados $\mathcal{E}$ es finito, siempre es posible encontrar solución al problema de decisión, en cambio, si es infinito, puede o no existir solución.

conoce con certeza el resultado. Estos detalles orillan al decisor, algunas veces, a posponer la elección lo más posible y después decidir sin intentar considerar todas las implicancias de su decisión.

La toma de decisión, fundamentalmente, tiene que ver con combinar información sobre probabilidades e información sobre deseos e interés. Y, el decisor debe de tener cierto conocimiento de los estados de la Naturaleza para poder predecir las probabilidades de cada estado. De lo contrario no podrá tomar una buena decisión que sea razonable y defendible. También, siempre que un decisor tengo cierto conocimiento sobre los estados de la Naturaleza podrá asignar una probabilidad subjetiva a la ocurrencia de cada estado, o bien, puede tomar información relevante de especialistas para decidir correctamente.

1.2.2. Criterio de la utilidad máxima esperada

Las distintas consecuencias posibles generadas por las preferencias de quien ha de tomar una decisión (decisor) bajo distintos escenarios se pueden cuantificar por medio de una función de utilidad, o bien de pérdida, dependiendo de lo que se esté especificando.

A la medida en que una alternativa satisface un criterio se le denomina utilidad. Hay otras palabras que tienen escencialmente el mismo significado como valor ganancia, valor psicológico y "satisfactoriedad". Un modelo de utilidad es un modelo gráfico o matemático que puede usarse para estimar la utilidad de un concepto o de una alternativa. Esencialmente, un modelo de utilidad transforma una descripción de un concepto o alternativa en una evaluación o descripción numérica de preferencias existentes que permita ordenar las eventualidades lo más certeramente posible.

Regularmente, la manera más usual de formular un problema de decisión es usando una matriz de consecuencias, o bien de beneficios, que consiste en asignar a cada fila una acción posible y a cada columna un escenario posible. Se analiza y se estudia la ocurrencia de cada acción y, se le asigna a cada una de ellas, bajo los distintos escenarios existentes, un valor o beneficio generado por la función de utilidad, o bien de pérdida, según los intereses del tomador de decisiones; así como también, se le asigna cierta probabilidad de ocurrencia a cada posible estado de la Naturaleza.

Una vez definida la estructura junto con los valores y las probabilidades se procede a calcular la utilidad esperada para cada una de las acciones y, se toma la acción correspondiente a la utilidad esperada más alta, o bien, correspondiente a la pérdida esperada más baja, según la prespectiva con la que se analice la situación.

Es claro notar que al tomar una decisión basada en este procedimiento tal vez no se obtengan los resultados que realmente se esperan, sin embargo se puede tener un considerable acercamiento a ellos.

Este procedimiento es conocido como criterio de la utilidad esperada máxima (criterio de Bayes) cuya coherencia y optimalidad están respaldadas por una teoría matemática bien fundamentada y totalmente específica[8].

Las decisiones pueden verse afectadas por la racionalidad subjetiva de las personas y por la manera en que a cada problema de decisión se hace frente, por ejemplo, algunas personas tienen la tendencia a evitar el riesgo cuando hay perspectivas de ganancia, y buscan el riesgo cuando las perspectivas son de pérdida. Es por esto que, abordar las decisiones como si fueran "apuestas" es la base de la teoría de la decisión. Sin embargo, pueden existir errores en decisiones arriesgadas debido a presunciones falsas, no tener estimación exacta de las probabilidades, depender de la espectativa, dificultades en medir la función de utilidad y los errores de pronóstico.

EVENTOS

↓

ACCIONES ⟶ MODELO MATEMÁTICO ⟶ CONSECUENCIAS

1.3. Inferencia estadística

La inferencia juega un papel muy valioso e importante dentro del campo de la Estadística ya que por medio de ella se define cual será la mejor decisión que se tomará ante las circunstancias que se estén analizando. La información estadística se recopila y analiza no solamente con el propósito de añadirla al conocimiento científico, sino también para ayudar en la toma de decisiones, por

[8] Véase detalladamente en el libro Bernardo José M. y Smith Adrián F. M., Bayesian Theory, Ed. Wiley, 1994.

eso es que la Estadística se ha llegado a definir como un grupo de métodos encaminados a tomar decisiones correctas ante la incertidumbre y como se ha mencionado en la sección anterior, la inferencia estadística se puede resolver como un problema de teoría de decisiones en un ambiente o enfoque bayesiano.

La inferencia estadística bayesiana se lleva a cabo mediante el contraste de hipótesis, estimación puntual, estimación por regiones y predicción.

1.3.1. Contraste de hipótesis

El contraste de hipótesis constituye el proceso relacionado con aceptar o rechazar declaraciones acerca de los parámetros de cierto evento.

El problema de inferencia pura, o bien, contraste de hipótesis puede ser descrito como el método o procedimiento en el que se busca saber **cual de los elementos**, de cierto conjunto de hipótesis, teorías, estados de la Naturaleza o modelos paramétricos mutuamente exclusivos H_j; **es verdadero**, es decir, cual de todos los escenarios planteados es el que tiene mayor posibilidad de ocurrir. Generalmente el conjunto de hipótesis es una partición finita[9] de posibles escenarios dentro del espacio de estados de la Naturaleza $\mathcal{E}$ y el objetivo es contrastar

$$H_1 : \theta \in \Theta_1, H_2 : \theta \in \Theta_2, ..., H_m : \theta \in \Theta_m$$

siendo $\Theta_1, \Theta_2, ..., \Theta_m$ subconjuntos disjuntos de Θ.

El contraste se lleva a cabo mediante el cálculo de la probabilidad que cada hipótesis tiene y se utiliza el criterio óptimo llamado criterio de la utilidad máxima esperada (criterio de Bayes) para escoger la hipótesis de mayor conveniencia.

En este caso, el espacio de estados relevantes es $\{\Theta_1, \Theta_2, ..., \Theta_m\}$ o indistintamente $\{H_1, H_2, ..., H_m\}$, el espacio de acciones es $\mathcal{A} = \{a_1, ..., a_m\}$ en donde el significado de a_i será actuar como si H_i ocurriera y

$$p\Big[H_j\Big] = p(\theta \in \Theta_j) = \int_{\Theta_j} p(\theta) d\theta,$$

o bien,

[9] Si un conjunto de tamaño N se puede dividir en n subconjuntos ordenados tales que el primero tenga tamaño k_1, el segundo k_2, y así sucesivamente de tal manera que la suma de todos los k_i, con i=1, ..., n, sea igual a N; entonces a cada subconjunto d$k_i, i = 1, ..., n$, se le llama partición finita.

$$p\left[H_j\right] = p(\theta \in \Theta_j) = \sum_{\theta \in \Theta_j} p(\theta),$$

o, en caso de existir información muestral $\mathbf{x} = (x_1, x_2, ..., x_n)$

$$p\left[H_j\right] = p(\theta \in \Theta_j) = \int_{\Theta_j} p(\theta|\mathbf{x}) d\theta,$$

o bien,

$$p\left[H_j\right] = p(\theta \in \Theta_j) = \sum_{\theta \in \Theta_j} p(\theta|\mathbf{x}).$$

1.3.2. Estimación puntual y estimación por regiones

Se recurre a la estimación de parámetros dentro de la inferencia estadística debido a que no se conoce el valor real de dichos parámetros y, en vez de su valor real se hace uso de un estimador que es una aproximación tentativa para tener resultados óptimos en la toma de decisiones.

Las estimaciones pueden ser de dos tipos:

Puntuales ; las cuales se caracterizan por ser un solo valor numérico único.

y

Por regiones ; las cuales se caracterizan por contar con un límite inferior y un límite superior que delimitan el intervalo en donde se encuentra el "verdadero" valor del parámetro.

ESTIMACIÓN PUNTUAL

En estadística bayesiana, la estrategia básica utilizada, en caso de tener interés por hacer estimación puntual, está dada por los siguientes cursos de acción:

a) En primer lugar, se necesita utilizar el espacio de acciones $\mathcal{A}$ como el espacio de todos los posibles valores que podría tomar el parámetro, es decir, $\mathcal{A} \equiv \Theta$, de tal manera que cada acción, que se pueda realizar, se defina como si el valor de θ fuera cierto.

b) Se debe definir una función de utilidad o pérdida clara llamada $u\left[\hat{\theta}, \theta\right]$ o $l\left[\hat{\theta}, \theta\right]$ donde $\hat{\theta}$ es la acción de actuar como si θ fuera cierta.

c) Una vez definida $u\left[\hat{\theta}, \theta\right]$, o bien, $l\left[\hat{\theta}, \theta\right]$, se aplica el Criterio de Bayes, utilizando la distribución de probabilidad de θ, ya sea a priori $p\left[\theta\right]$ (en caso de no existir información muestral), o bien, a posteriori $p\left[\theta|\mathbf{x}\right]$ (en caso de contar con información muestral).

El estimador puntual bayesiano será aquel número que minimize la pérdida esperada, o bien, maximize la utilidad esperada y se puede interpretar como un simple resumen de la distribución de θ ya que toda la información necesaria está contenida en dicha distribución.

Los estimadores puntuales bayesianos más utilizados se mencionan a continuación con su respectiva función de utilidad:

$\hat{\theta}^*$	$u(\hat{\theta}, \theta)$
$-(\hat{\theta} - \theta)^2$	$E(\theta)$
$-\lvert\hat{\theta} - \theta\rvert$	$Med(\theta)$
$-(\frac{\hat{\theta} - \theta}{\hat{\theta}})^2$	$\frac{E(\theta^{[illegible]})}{E(\theta)}$

ESTIMACIÓN POR REGIONES

En ocasiones la descripción de la información sobre θ a través de $p\left[\theta\right]$ o de $p\left[\theta|\mathbf{x}\right]$ resulta complicada para aquellos que no están familiarizados con el suceso. Esta situación necesita contar con resultados más accesibles, por lo que se recurre simplemente al reporte de ciertas regiones contenidas en Θ, con cierta probabilidad dada, que puedan contener el valor real del parámetro.

Sin embargo, es claro notar que para cierta probabilidad α dada, por lo general existe una infinidad de regiones dentro de Θ. Por esta razón, al hacer estimación por regiones de cierto parámetro se buscan otras características a la par, como obtener el intervalo cuya distancia sea mínima entre el límite superior e inferior y obtener la máxima densidad dentro del intervalo.

Básicamente el procedimiento para llevar a cabo la estimación por regiones de θ está dado bajo las condiciones mencionadas, haciendo uso de la distribución de probabilidad de θ, ya sea a priori, o bien, a posteriori y, la complejidad del cálculo de la región estimada dependerá de dicha distribución.

1.3.3. Predicción puntual y predicción por regiones

Al modelar probabilísticamente un fenómeno aleatorio causa interés saber cuál será el próximo comportamiento que tendrá el parámetro analizado θ (como ya se ha hecho incapié), pero también, ocurre con frecuencia el valioso interés por hacer inferencias sobre observaciones o repeticiones futuras del fenómeno aleatorio.

Bajo estas condiciones se utiliza la distribución de probabilidad del fenómeno, ya sea a priori o a posteriori, $p\left[X\right]$ o $p\left[X|\mathrm{x}\right]$ respectivamente y, **se aplican las mismas técnicas** de contraste de hipótesis, estimación puntual y estimación por regiones para realizar la inferencia de θ, aplicadas a la inferencia sobre X.

Aunque las bases de estadística bayesiana datan de hace más de dos siglos, no es hasta fechas recientes cuando empieza a asistirse a un uso creciente de este enfoque en el ámbito de la investigación. Una de las razones que explican esta realidad y que a la vez anuncian un impetuoso desarrollo futuro, es la absoluta necesidad del cálculo computarizado para la resolución de algunos problemas de cierta complejidad. **La existencia de software disponible hace posible operar con estas técnicas** debido a que los procedimientos bayesianos constituyen una tecnología emergente de procesamiento y análisis de información para la que cabe esperar una presencia cada vez más intensa en el campo de la aplicación estadística.

Capítulo 2

Distribuciones probabilísticas

Debido a las condiciones y ambiente en que la estadística es considerada desde un punto de vista bayesiano, la aleatoriedad tanto del evento de interés como de los parámetros involucrados es conveniente considerarla en términos de una distribución de probabilidad. Por esta razón, el conocimiento de las distribuciones probabilísticas es de gran utilidad para la realización de aplicaciones ya que para describir un fenómeno aleatorio con resultados numéricos, es necesario y suficiente especificar su función de probabilidades.

Se tienen a la mano representaciones distintas pero equivalentes del mismo concepto matemático, que es posible llamar ley de probabilidades (o distribución de probabilidades) del fenómeno aleatorio con resultados numéricos.

Una de dichas representaciones es la ley de probabilidades llamada discreta, la cual corresponde a una función de distribución discreta y describe comportamientos o eventos que solo pueden tomar valores enteros.

La segunda representación es la ley de probabilidades continua que corresponde a una función de distribución continua y describe comportamientos que pueden tomar valores correspondientes a un intervalo sin interrupciones.

Sin embargo, aunque las mencionadas representaciones forman parte de una principal clasificación de ley de probabilidades, también se pueden encontrar distribuciones con características mixtas, es decir, que dentro de su comportamiento probabilístico se involucran tanto comportamientos discretos como comportamientos continuos; y también, dentro de la anterior clasificación se podrán encontrar distribuciones que describen a una sola variable aleatoria (distribuciones univariadas), o bien, distribuciones que describan a un vector de variables aleatorias llamado vector aleatorio que definirá su comportamiento con respecto a vectores o matrices de parámetros (distribuciones multivariadas).

A continuación se presentará una recopilación útil de distribuciones de probabilidad tanto discretas como continuas en los contextos univariante y multivariante, a las cuales se recurre con gran frecuencia al realizar aplicaciones bayesianas e inequivocamente cuando existe interés en la aleatoriedad y probabilidades de algún evento en particular independientemente de un enfoque bayesiano. Cada distribución de probabilidad contará con ciertas propiedades y características que las definen claramente y de manera única.

2.1. Distribuciones discretas univariadas

Distribución Bernoulli

Una cantidad aleatoria discreta x tiene distribución Bernoulli con parámetro θ, donde $0 < \theta < 1$, si su función de probabilidad $Br(x|\theta)$ está dada por

$$Br(x|\theta) = \theta^x(1-\theta)^{1-x},$$

para $x = 0, 1$

La distribución Bernoulli representa un experimento con dos posibles resultados, los que por convención se llaman fracaso y éxito, donde $x(\text{fracaso}) = 0$, $x(\text{éxito}) = 1$ y $Br(x = 0) = 1 - \theta$, $Br(x = 1) = \theta$.

La media y varianza son $E[x] = \theta$, y $V[x] = \theta(1-\theta)$

Distribución Binomial

Una cantidad aleatoria discreta x tiene distribución binomial con parámetros θ y n $(0 < \theta < 1, n = 1, 2, ...)$ si su función de probabilidad $Bi(x|\theta, n)$ está dada por

$$Bi(x|\theta, n) = \binom{n}{x}\theta^{x}(1-\theta)^{1-x},$$

donde $x = 0, 1, ..., n$.

La distribución binomial representa la distribución del número total de éxitos de n ensayos independientes para los que θ es la probabilidad de éxito de cada ensayo individual.

La media y varianza son $E[x] = n\theta$, y $V[x] = n\theta(1-\theta)$. Y, el dato modal es conseguido por el entero más grande $M[x]$ el cual no debe exceder de $x_m = (n+1)\theta$; si x_m es un entero, entonces ambos x_m y $x_m - 1$ son modas.

Si $n = 1$, se dice que x tiene una distribución Bernoulli con función de probabilidad denotada por $Br(x|\theta)$.

La suma de k cantidades aleatorias binomiales independientes con parámetros (θ, n_i), $i = 1, ..., k$, es una cantidad aleatoria binomial con parámetros θ y $n_1 + ... + n_k$.

Distribución Binomial - Beta[1]

Una cantidad aleatoria discreta x tiene una distribución binomial - beta con parámetros α, β y n $(\alpha > 0, \beta > 0, n = 1, 2, ...)$ si su función de probabilidad $Bb(x|\alpha, \beta, n)$ está dada por

$$Bb(x|\alpha, \beta, n) = c\ \binom{n}{x}\Gamma(\alpha + x)\Gamma(\beta + n - x),$$

para $x = 0, ..., n$,

[1] Para la comprensión de esta distribución es necesario conocer la distribución de la variable aleatoria continua beta.

donde $c = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)}$.

La distribución es generada por la combinación de distribuciones

$$Bb(x|\alpha,\beta,n) = \int_0^1 Bi(x|\theta,n)\ Be(\theta|\alpha,\beta)\ d\theta.$$

La media y la varianza están dadas por

$$E[x] = n\frac{\alpha}{\alpha+\beta} \quad \text{and} \quad V[x] = \frac{n\alpha\beta}{(\alpha+\beta)^2}\frac{(\alpha+\beta+n)}{(\alpha+\beta+1)}.$$

Y, la moda es adquirida como el entero más alto $M[x]$ el cual no excede

$$x_m = \frac{(n+1)(\alpha-1)}{\alpha+\beta-2};$$

si x_m es un entero, ambos x_m y $x_m - 1$ son modas.

Si $\alpha = \beta = 1$ se obtendrá una distribución uniforme discreta, asignando en gran cantidad $(n+1)^{-1}$ a cada posible x.

Distribución Binomial Negativa

Una cantidad aleatoria discreta x tiene una distribución binomial negativa con parámetros θ y r $(0 < \theta < 1, r = 1, 2, ...)$ si su función de probabilidad $Bn(x|r,\theta)$ es

$$Bn(x|r,\theta) = c\ \binom{r+x-1}{r-1}(1-\theta)^x,$$

donde $x = 1, 2, ...$ y $c = \theta^r$.

La distribución binomial negativa se utiliza al observar los resultados de ensayos independientes, cada uno con probabilidad θ de que el resultado se considere un éxito, hasta que suceda el r-ésimo fracaso. El número de éxitos obtenidos es la variable aleatoria que se llama binomial negativa.

La media y la varianza son $E[x] = r\frac{(1-\theta)}{\theta}$ y $V[x] = r\frac{(1-\theta)}{\theta^2}$.

Si $r(1-\theta) > 1$ el dato modal $M[x]$ es el mínimo entero no menor que $r\frac{(1-\theta)}{\theta}$; si $r(1-\theta) = 1$ existen dos modas en $x = 0$ y $x = 1$; y si $r(1-\theta) < 1$ entonces la moda $M[x] = 0$.

Si $r = 1$ se dice que x tiene una distribución Pascal, o bien, distribución geométrica. Más aún, la suma de k variables aleatorias independientes con distribución binomial negativa con parámetros (θ, r_i), $i = 1, ..., k$, es una variable aleatoria con distribución binomial negativa con parámetros θ y $r_1 + ... + r_k$.

Distribución Binomial Negativa - Beta[2]

Una cantidad aleatoria discreta x tiene una distribución binomial negativa - beta con parámetros α, β y r $(\alpha > 0, \beta > 0, r = 1, 2, ...)$ si su función de probabilidad $Bnb(x|\alpha, \beta, r)$ es

$$Bnb(x|\alpha,\beta,r) = c\binom{r+x-1}{r-1}\frac{\Gamma(\beta+x)}{\Gamma(\alpha+\beta+x+r)},$$

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

donde $c = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)}{\Gamma(\alpha)\Gamma(\beta)}$.

La distribución es generada por la combinación de distribuciones

$$Nbb(x|\alpha,\beta,r) = \int_0^1 Nb(x|\theta,r)\ \ Be(\theta|\alpha,\beta)\ \ d\theta.$$

La media es $E[x] = \frac{r\beta}{\alpha-1}, \alpha > 1$, y la varianza está dada por

[2] Para la comprensión de esta distribución es necesario conocer la distribución de la variable aleatoria continua beta.

$$V[x] = \frac{r\beta}{\alpha - 1}\left[\frac{\alpha + \beta + r - 1}{\alpha - 2} + \frac{r\beta}{(\alpha - 1)(\alpha - 2)}\right], \qquad \alpha > 2.$$

Distribución Geométrica

Una cantidad aleatoria discreta x tiene una distribución geométrica con parámetro θ con $0 < \theta < 1$ si su función de probabilidad $Geo(x|\theta)$ es

$$Geo(x|\theta) = \theta(1 - \theta)^x - 1,$$

para $x = 1, 2, \ldots,$

Un caso particular de la distribución binomial negativa lo constituye la distribución geométrica del número de éxitos hasta el primer fracaso.

La media y la varianza están dadas por $E[x] = \dfrac{\theta}{1-\theta}$ y $V[x] = \dfrac{\theta}{1-\theta}$.

Si $(1-\theta) > 1$ el dato modal $M[x]$ es el mínimo entero no menor que $\dfrac{(1-\theta)}{\theta}$; si $(1-\theta) = 1$ existen dos modas en $x = 0$ y $x = 1$; y si $(1-\theta) < 1$ entonces la moda $M[x] = 0$.

Distribución Hipergeométrica

Una cantidad aleatoria discreta x tiene una distribución hipergeométrica con parámetros enteros N, M y n ($n \leq N + M$) si su función de probabilidad $Hy(x|N, M, n)$ está dada por

$$Hy(x|N, M, n) = c \binom{N}{x}\binom{M}{n - x},$$

donde $max\{0, n - M\} \leq x \leq min\{n, M\}$, y

$$c = \binom{N + M}{n}^{-1}.$$

La distribución hipergeométrica se utiliza al tener una población con N elementos, M de los cuales cumplen con cierta propiedad (los $N-M$ restantes no la cumplen). Se extrae una muestra de tamaño n sin reemplazo; entonces, el número de elementos en la muestra que cumplen con la propiedad es una variable aleatoria que se distribuye como una variable hipergeométrica.

La media y la varianza son $E[x] = \dfrac{nN}{N+M}$ y $V[x] = \dfrac{nMN}{(N+M)^2}\dfrac{(N+M-n)}{(N+M-1)}$.

El dato modal se obtiene como el máximo entero $M[x]$ que no excede

$$x_m = \frac{(n+1)(N+1)}{M+N+2};$$

si x_m es un entero, entonces x_m y $x_m - 1$ son modas.

Distribución Hipergeométrica Negativa

Una cantidad aleatoria discreta x tiene distribución hipergeométrica negativa con parámetros enteros r, M y N si su función de probabilidad $Hn(x|r, M, N)$ es

$$Hn(x|r, M, N) = c \binom{x+r-1}{x}\binom{N-r-x}{M-x},$$

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

donde $c = \binom{N}{M}^{-1}$.

La distribución hipergeométrica negativa se utiliza al observar una población con N elementos, M de los cuales poseen una cierta propiedad ($N-M$ no la cumplen, los restantes). Se extrae de la población sin reemplazo; entonces, el número de elementos que tiene la propiedad hasta que se tenga el r-ésimo que no la tiene, es una variable aleatoria que se distribuye como una variable hipergeométrica negativa.

La media y la varianza están dadas por

$$E[x] = \frac{Mr}{N - M + 1} \text{ y } V[x] = \frac{Mr(N - M - r + 1)(N + 1)}{(N - M + 1)^2(N - M + 2)}.$$

La distribución hipergeométrica negativa es también una distribución binomial negativa con parámetros r, $N - M - R + 1$ y M.

Distribución Pascal

Una cantidad aleatoria discreta x tiene distribución Pascal con parámetros r y θ (r entero positivo y $0 < \theta < 1$) si su función de probabilidad $Ps(x|r, \theta)$ es

$$Ps(x|r, \theta) = \binom{x - 1}{r - 1}(1 - \theta)^r \theta^x,$$

para $x = r, r + 1, \ldots$

La distribución Pascal se utiliza al observar los resultados de ensayos independientes consecutivos, cada uno con probabilidad θ de que el resultado sea un éxito, hasta que suceda el r-ésimo fracaso. El número de ensayos es una variable aleatoria con distribución Pascal.

Esta distribución puede confundirse con la binomial negativa, donde al realizar el experimento también se está en espera del r-ésimo fracaso, pero donde la variable aleatoria es el número de éxitos obtenidos antes del evento esperado, en tanto que en la distribución Pascal la variable aleatoria es el número de ensayos reuqeridos hasta obtener el evento esperado.

La media y la varianza están dadas por

$$E[x] = \frac{r}{1 - \theta} \text{ y } V[x] = \frac{r\theta}{(1 - \theta)^2}.$$

La distribución Pascal con parámetros r y θ es también una distribución geométrica con parámetro θ y $Ps(x|r, \theta) = Bn(x - r|r, \theta)$.

Distribución Poisson

Una cantidad aleatoria discreta x tiene distribución Poisson con parámetro λ ($\lambda > 0$) si su función de probabilidad $Pn(x|\lambda)$ es

$$Pn(x|\lambda) = c \ \frac{\lambda^x}{x!},$$

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

donde $c = e^{-\lambda}$.

Esta distribución puede considerarse naturalmente como la distribución del número de eventos que ocurren en el intervalo de tiempo $(0, 1)$ cuando los eventos ocurren a razón de λ por unidad de tiempo.

La media y la varianza están dadas por $E[x] = V[x] = \lambda$.

El dato modal $M[x]$ es el máximo entero de los cuales no exceden λ. Si λ es un entero, entonces λ y $\lambda - 1$ son modas.

La suma de k cantidades aleatorias independientes Poisson con parámetros λ_i, $i = 1, ..., k$, es una cantidad aleatoria Poisson con parámetro $\lambda_1 + ... + \lambda_k$.

Distribución Poisson - Gamma[3]

Una cantidad aleatoria discreta x tiene distribución Poisson - gamma con parámetros α β y n ($\alpha > 0, \beta > 0, n = 1, 2, ...$) si su función de probabilidad $Pg(x|\alpha, \beta, n)$ es

$$Pg(x|\alpha, \beta, n) = c \ \frac{\Gamma(\alpha + x)}{x!} \frac{n^x}{(\beta + n)^{\alpha + x}},$$

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

[3]Para la comprensión de esta distribución es necesario conocer la distribución de la variable aleatoria continua gamma.

$$\text{donde } c = \frac{\beta^\alpha}{\Gamma(\alpha)}.$$

La distribución es generada por la combinación

$$Pg(x|\alpha,\beta,n) = \int_0^1 Pn(x|n\lambda)\ \ Ga(\lambda|\alpha,\beta)\ \ d\lambda.$$

Una distribución Poisson compuesta es, en efecto, una generalización de la distribución binomial negativa $Bn(x|\alpha, \frac{\beta}{(\beta+n)})$, solamante para enteros α.

La media es $E[x] = \frac{n\alpha}{\beta}$, y la varianza es $V[x] = \frac{n\alpha(\beta+n)}{\beta^2}$.

Si $\alpha n > \beta+n$, existe una moda en el mínimo entero no menor a $(\frac{n(\alpha-1)}{\beta})-1$; si $\alpha n = \beta+n$, existen dos modas en $x=0$ y $x=1$; y, si $\alpha n < \beta+n$, $M[x]=0$.

Distribución Uniforme

Una cantidad aleatoria discreta x tiene distribución uniforme con parámetro n $(n = 1, 2, ...)$ si su función de probabilidad $Und(x|n)$ es

$$Und(x|n) = \frac{1}{n+1},$$

para $x = x_0, x_1, ..., x_n$.

La media y varianza estan dadas por

$$E[x] = \frac{\sum_{i=0}^{n} x_i}{n+1}$$

y

$$V[x] = \frac{n\sum_{i=0}^{n} x_i^2 - 2\sum_{i<j}^{n} x_i x_j}{(n+1)^2}.$$

La distribución uniforme carece de moda.

2.2. Distribuciones continuas univariadas

Distribución Beta

Una variable aleatoria continua x tiene distribución beta con parámetros α y β $(\alpha > 0, \beta > 0)$ si su función de densidad $Be(x|\alpha, \beta)$ es

$$Be(x|\alpha,\beta) = cx^{\alpha-1}(1-x)^{\beta-1},$$

para $0 < x < 1$,

donde

$$c = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

y $\Gamma(x) = \int_0^\infty t^{x-1}\ e^{-t}\ dt$; valores enteros y semi-enteros de la función gamma son fácilmente de encontrar utilizando la relación recursiva $\Gamma(x+1) = x\Gamma(x)$, y los valores de $\Gamma(1) = 1$ y $\Gamma(1/2) = \sqrt{\pi} \approx 1{,}7725$.

La aplicación sistemática de la integral beta,

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\ dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

se obtiene que

$$E[x] = \frac{\alpha}{\alpha+\beta} \qquad \text{y} \qquad V[x] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Si $\alpha > 1$ y $\beta > 1$, existe una única moda en $\dfrac{\alpha-1}{\alpha+\beta-2}$. Si x tiene densidad de probabilidad $Be(x|\alpha,\beta)$, entonces $y = 1-x$ tiene densidad de probabilidad $Be(y|\alpha,\beta)$. Si $\alpha = \beta = 1$, se dice que x tiene distribución uniforme $Un(x|0,1)$ en el intervalo $(0,1)$.

Al considerar la transformación $y = a + x(b-a)$, donde x tiene densidad de probabilidad $Be(x|\alpha,\beta)$, la distribución beta puede ser generalizada para cualquier intervalo finito (a,b).

En particular, la distribución uniforme $Un(y|a,b)$ en el intervalo (a,b),

$$Un(y|a,b) = (b-a)^{-1}, \qquad a < y < b,$$

tiene media $E[x] = (a+b)/2$ y varianza $V[y] = (b-a)^2/12$.

Distribución Cauchy

Una variable aleatoria x tiene distribución Cauchy con parámetros α y β $(-\infty < \alpha < \infty, \beta > 0)$ si su función de densidad de probabilidad $Ca(x|\alpha,\beta)$ es

$$Ca(x|\alpha,\beta) = \frac{\beta}{\pi}\ \frac{1}{\beta^2 + (x-\alpha)^2},$$

donde $-\infty < x < \infty$.

La función de densidad de probabilidad de Cauchy es simétrica con respecto a α que es la mediana de la variable, y ni la media ni la varianza existe para esta distribución.

Distribución Exponencial

Una variabla aleatoria x tiene distribución exponencial con parámetro θ $(\theta > 0)$ si su finción de densidad $Exp(x|\theta)$ es

$$Exp(x|\theta) = \theta\ e^{-\theta x},$$

donde $x > 0$.

La media y la varianza están dadas por

$$E[x] = \frac{1}{\theta} \qquad \text{y} \qquad V[x] = \frac{1}{\theta^2}.$$

La moda de una distribución exponencial está localizada en $x = 0$.

Distribución Exponencial Negativa

Una variable aleatoria x tiene distribución exponencial negativa con parámetros α y β $(-\infty < \alpha < \infty, \beta > 0)$ si su función de densidad de probabilidad $ExpNeg(x|\alpha,\beta)$ es

$$ExpNeg(x|\alpha,\beta) = \beta \; e^{-\beta(x-\alpha)},$$

para $\alpha < x < \infty$.

La media y la varianza están dadas por

$$E[x] = \frac{\alpha\beta - 1}{\beta} \quad \text{y} \quad V[x] = \frac{\alpha^2\beta^2 + 1}{\beta^2}.$$

En particular, $ExpNeg(0,\beta) = Ga(1, \frac{1}{\beta}) = W(\beta, 1)$[4].

Distribución F (o de Snedecor)

Una variable aleatoria continua x tiene distribuciín F o Snedecor con parámetros α y β (grados de libertad) $(\alpha > 0, \beta > 0)$ si su densidad de probabilidad $Fs(x|\alpha,\beta)$ es

$$Fs(x|\alpha,\beta) = c \; \frac{x^{(\alpha/2)-1}}{(\beta + \alpha x)^{(\alpha+\beta)/2}},$$

para $x > 0$,

donde

$$c = \frac{\Gamma((\alpha+\beta)/2)}{\Gamma(\alpha/2)\Gamma(\beta/2)} \alpha^{\alpha/2} \beta^{\beta/2}.$$

Si $\beta > 2$, $E[x] = \dfrac{\beta}{\beta - 2}$ y existe una única moda en $\left[\dfrac{\beta}{\beta+2} \dfrac{\alpha-2}{\alpha}\right]$; más aún, si $\beta > 4$,

[4]Distribución Weibull

$$V[x] = 2\frac{\beta^2}{\alpha(\beta-4)}\frac{(\alpha+\beta-2)}{(\beta-2)^2}.$$

Si x y y son cantidades aleatorias independientes, ambas con distribución ji-cuadrada, con v_1 y v_2 grados de libertad respectivos, entonces

$$z = \frac{(x/v_1)}{(y/v_2)}$$

tiene distribución Snedecor con v_1 y v_2 grados de libertad.

Distribución Gamma

Una variable aleatoria continua x tiene distribución gamma con parámetros α y β $(\alpha > 0, \beta > 0)$ si su función de densidad $Ga(x|\alpha,\beta)$ es

$$Ga(x|\alpha,\beta) = c\ x^{\alpha-1}\ e^{-\beta x},$$

donde $x > 0$ y donde $c = \beta^\alpha/\Gamma(\alpha)$.

De la aplicación sistemática de la integral gamma

$$\int_0^\infty x^{\alpha-1}\ e^{-\beta x}\ dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$$

se obtiene que $E[x] = \dfrac{\alpha}{\beta}$ y $V[x] = \dfrac{\alpha}{\beta^2}$.

Si $\alpha > 1$, existe una única moda en $\dfrac{\alpha-1}{\beta}$ y si $\alpha < 1$ no existen modas (la densidad es ilimitada).

Si $\alpha = 1$, se dice que x tiene distribución exponencial $Exp(x|\beta)$, y si también $\beta = 1$ se dice que x tiene distribución exponencial estandarizada.

Si $\beta = 1$, se dice que x tiene distribución Erlang con parámetro α, o bien, se dice que tiene distribución gamma estandarizada.

Si $\alpha = \frac{v}{2}$ y $\beta = \frac{1}{2}$, se dice que x tiene distribución ji-cuadrada (χ^2) (central) con parámetro v (frecuentemente conocido como grados de libertad) y densidad denotada como $\chi^2(x|v)$ o $\chi^2_v(x)$.

Considerando la transformación $y = a+x$ o $z = b-x$, donde x tiene densidad $Ga(x|\alpha,\beta)$, la distribución gamma puede ser generalizada para los rangos (a,∞) o $(-\infty,b)$. Más aún, la suma de k variables aleatorias con distribución gamma independientes con parámetros (α_i,β), $i = 1,...,k$ es una variable aleatoria con función de densidad gamma con parámetros $\alpha_1 + ... + \alpha_k$ y β.

Distribución Gamma - Gamma

Una variable aleatoria continua x tiene distribución gamma-gamma con parámetros α, β y n ($\alpha > 0, \beta > 0, n = 1,2,...$) si su función de probabilidad $Gg(x|\alpha,\beta,n)$ es

$$Gg(x|\alpha,\beta,n) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+n)}{\Gamma(n)} \frac{x^{n-1}}{(\beta+x)^{\alpha+n}},$$

para $x > 0$.

La distribución es generada por la combinación

$$Gg(x|\alpha,\beta,n) = \int_0^\infty Ga(x|n,\lambda) \; Ga(\lambda|\alpha,\beta) \; d\lambda.$$

La media y la varianza están dadas por

$$E[x] = n \frac{\beta}{\alpha-1}, \quad \alpha > 1,$$

$$V[x] = \frac{\beta^2(n+n(\alpha-1))}{(\alpha-1)^2(\alpha-2)}, \quad \alpha > 2.$$

Distribución Gamma Invertida

Una variable aleatoria continua x tiene distribución gamma invertida con parámetros α y β $(\alpha > 0, \beta > 0)$ si su función de densidad $Gi(x|\alpha,\beta)$ es

$$Gi(x|\alpha,\beta) = c\ x^{-(\alpha+1)}\ e^{-\beta/x},$$

para $x > 0$, donde $c = \dfrac{\beta^\alpha}{\Gamma(\alpha)}$.

La media y la varianza son respectivamente

$$E[x] = \frac{\beta}{\alpha-1}, \qquad \alpha > 1,$$

$$V[x] = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \qquad \alpha > 2.$$

En la distribución gamma invertida solo existe una moda en $x = \beta/(\alpha+1)$.

El término gamma invertida se deriva del dato fácilmente establecido, el cual dice que si y tiene densidad gamma $Ga(y|\alpha,\beta)$, entonces $x = y^{-1}$ tiene densidad gamma invertida $Gi(x|\alpha,\beta)$.

Si x tiene distribución gamma invertida con $\alpha = v/2$, $\beta = 1/2$, entonces se dice que x tiene distribución χ^2_v-invertida.

Una variable aleatoria continua x tiene distribución raíz cuadrada gamma invertida $Ga^{-1/2}(y|\alpha,\beta)$, si $x = y^{-2}$ tiene densidad gamma $Ga(x|\alpha,\beta)$.

Distribución Ji - cuadrada

Una variable aleatoria continua x tiene distribución ji-cuadrada con parámetro v (grados de libertad) si su función de densidad $\chi^2(x|v) = \chi^2_v$ es

$$c\ x^{(v/2)-1}\ e^{-x/2},$$

para $x > 0$, donde $c = \dfrac{(1/2)^{v/2}}{\Gamma(v/2)}$.

La media y la varianza están dadas por

$$E[x] = v \quad y \quad V[x] = 2v$$

La distribución es unimodal y la moda se encuentra en el valor de $M[x]$ tal que $\chi^2(M[x]|v)$ sea igual a $\chi^2(M[x]|v-2)$.

Distribución Ji - cuadrada no central

Una variable aleatoria continua x tiene distribución χ^2 no central con parámetros v (grados de libertad) y λ (no centralidad) ($v > 0, \lambda > 0$) si su función de densidad $\chi^2(x|v,\lambda)$ es

$$\chi^2(x|v,\lambda) = \sum_{i=0}^{\infty} Pn(i|\lambda/2) \;\; \chi^2(x|v+2i),$$

para $x > 0$; es decir, la combinación de distribuciones χ^2 centrales con pesos Poisson.

La distribución ji-cuadrada no central se reduce a una distribución ji-cuadrada central cuando $\lambda = 0$.

La media y la varianza están dadas por $E[x] = v + \lambda$ y $V[x] = 2(v + 2\lambda)$; y, la distribución es unimodal y la moda se encuentra en el vaor de $M[x]$tal que $\chi^2(M[x]|v,\lambda)$ sea igual a $\chi^2(M[x]|v-2,\lambda)$.

Si $x_1, ..., x_k$ son variables aleatorias mutuamente independientes y con distribución normal $N(x_i|\mu_i, 1)$ entonces $z = \sum_{i=1}^{k} x_i^2$ tiene distribución χ^2 no central,

$$\chi^2(z|k, \sum_{i=1}^{k} \mu_i^2).$$

La suma de k distribuciones χ^2 no centrales independientes con parámetros (v_i, λ_i) tiene distribución χ^2 no central con parámetros $v_1 + ... + v_k$ y $\lambda_1 + ... + \lambda_k$.

Distribución Laplace

Una variable aleatoria continua x tiene distribución Laplace con parámetros α y β $(-\infty < \alpha < \infty, \beta > 0)$ si su función de densidad $Lap(x|\alpha, \beta)$ es

$$Lap(x|\alpha, \beta) = \frac{\beta}{2}\ e^{-\beta|x-\alpha|},$$

para $-\infty < x < \infty$.

La media y la varianza están dadas por $E[x] = \alpha$ y $V[x] = \frac{2}{\beta^2}$.

La función de densidad es simétrica con respecto a α.

Distribución Logística

Una variable aleatoria continua x tiene distribución logística con parámetros α y β $(\alpha \in \Re, \beta > 0)$ si su función de densidad $Lo(x|\alpha, \beta)$ es

$$Lo(x|\alpha, \beta) = c\ \frac{e^{-\frac{x-\alpha}{\beta}}}{\left[1 + e^{-\frac{x-\alpha}{\beta}}\right]^2},$$

donde $x \in \Re$ y $c = \beta^{-1}$.

Una expresión alternativa para la función de densidad es

$$\frac{1}{4\beta} sech^2 \left\{ \frac{1}{2}\ \left(\frac{x-\alpha}{\beta} \right) \right\}.$$

por lo que la distribución logística es llamada también distribución *sech*-cuadrada.

La distribución logística es más simple de expresar en términos de su función de distribución,

$$F_x(x) = \left[1 + e^{-\frac{x-\alpha}{\beta}}\right]^{-1}.$$

La distribución es simétrica con respecto a $x = \alpha$. La media y la moda están dadas por $E[x] = M[x] = \alpha$, y la varianza es $V[x] = \beta^2\pi^2/3$.

Distribución Lognormal

Una variable aleatoria continua x tienen distribución lognormal con parámetros μ y λ $(-\infty < \mu < \infty, \lambda > 0)$ si su función de densidad $LogN(x|\mu, \lambda)$ es

$$LogN(x|\mu, \lambda) = c \ \ e^{-\frac{\lambda}{2}(log x-\mu)^2},$$

para $x > 0$, donde $c = \left(\frac{\lambda}{2\pi x^2}\right)^{1/2}$

La media y la varianza están dadas por

$$E[x] = e^{\mu+\frac{1}{2\lambda}}$$

y

$$V[x] = e^{2\mu}\left(e^{\frac{2}{\lambda}} - e^{\frac{1}{\lambda}}\right)$$

La distribución de probabilidad es simétrica con respecto a $x = \mu$, por lo que $E[x] = M[x]$.

Distribución Maxwell

Una variable aleatoria continua x tiene distribución Maxwell con parámetros μ y λ $(-\infty < \mu < \infty), \lambda > 0)$ si su función de densidad $Max(x|\mu, \lambda)$ es

$$Max(x|\mu, \lambda) = c \ \ (x-\mu)^2 \ \ e^{-\frac{\lambda}{2}(x-\mu)^2}$$

para $\mu < x < \infty$, donde $c = \left(\frac{2\lambda^3}{\pi}\right)^{1'2}$.

La media y la varianza están dadas por

$$E[x] = \mu + 2\left(\frac{2}{\lambda\pi}\right)^{1'2}$$

y

$$V[x] = \frac{3\pi - 8}{\lambda\pi}$$

Distribución Normal

Una variable aleatoria continua x tiene distribución normal con parámetros μ y λ ($\mu \in \Re, \lambda > 0$) si su función de densidad $N(x|\mu, \lambda)$ es

$$N(x|\mu, \lambda) = c\ e^{-\frac{\lambda}{2}(x-\mu)^2},$$

para $x \in \Re$, donde $c = \left(\frac{\lambda}{2\pi}\right)^{1/2}$.

La distribución es simétrica con respecto a $x = \mu$. La media y la moda son $E[x] = M[x] = \mu$ y la varianza es $V[x] = \lambda^{-1}$, tal que λ representa la precisión de la distribución. Alternativamente, $N(x|\mu, \lambda)$ es denotada por $N(x|\mu, \sigma^{-2})$, donde $\sigma^2 = V[x]$ es la varianza.

Si $\mu = 0$ y $\lambda = 1$, se dice que x tiene distribución normal estndar, con función de distribución

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\ dt.$$

Si $y = \lambda^{1/2}(x - \mu) = (x - \mu)/\sigma$, donde x tiene densidad de probabilidad normal $N(x|\mu, \lambda)$, entonces y tiene función de densidad normal estándar $N(x|0, 1)$.

En general, si $y = a + \sum_{i=1}^{k} b_i x_i$, donde las x_i son independientes entre s con densidades $N(x_i|\mu_i, \lambda_i)$, entonces y tiene densidad de probabilidad normal,

$N(y|a + \sum_{i=1}^{k} b_i\mu_i, \lambda)$ donde $\lambda = (\sum_{i=1}^{k} \frac{b_i^2}{\lambda_i})^{-1}$, una media armónica ponderada de precisiones individuales.

Si $x_1, ..., x_k$ son variables aleatorias independientes y con distribución normal estándar, entonces $z = \sum_{i=1}^{k} x_i^2$ tienen distribución χ_k^2 central.

Distribución Pareto

Una variable aleatoria continua x tiene distribución Pareto con parámetros α y β ($\alpha > 0, \beta > 0$) si su función de densidad $Pa(x|\alpha, \beta)$ es

$$Pa(x|\alpha, \beta) = c\ x^{-(\alpha+1)},$$

para $x \geq \beta$, donde $c = \alpha\beta^\alpha$.

La media y la varianza están dadas por

$$E[x] = \frac{\alpha\beta}{\alpha - 1}, \quad si\ \alpha > 1$$

$$V[x] = \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}, \quad si\ \alpha > 2.$$

La moda es $M[x] = \beta$.

La distribución es generada por la combinación

$$Pa(x|\alpha, \beta) = \int_0^\infty Exp(x - \beta|\theta)\ Ga(\theta|\alpha, \beta)\ d\theta.$$

Una variable aleatoria continua y tiene densidad Pareto invertida $Pi(y|\alpha, \beta)$ si $x = y^{-1}$ tiene densidad $Pa(x|\alpha, \beta)$.

Distribución Rayleigh

Una variable aleatoria continua x tiene distribución Rayleigh con parámetro β ($\beta > 0$) si su función de densidad $Ray(x|\beta)$ es

$$Ray(x|\beta) = c\ x\ e^{-\frac{x^2\beta^2}{2}},$$

para $x > 0$, donde $c = \beta^2$.

La media y la varianza están dadas por

$$E[x] = \frac{1}{\beta}\sqrt{\frac{\pi}{2}}$$

y

$$V[x] = \frac{4-\pi}{2\beta^2}.$$

Distribución Triangular

Una variable aleatoria continua x tiene distribución triangular con parámetros a, b y c ($a < b < c$) si su función de densidad $Tri(x|a,b,c)$ es

$$Tri(x|a,b,c) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & a < x < b \\ \dfrac{2(c-x)}{(c-b)(c-a)} & b < x < c \end{cases}$$

Su media y su varianza están dadas por

$$E[x] = \frac{a+b+c}{3}$$

y

$$V[x] = \frac{a^2+b^2+c^2-ab-ac-bc}{18}.$$

La moda se obtiene cuando $x = b$ y es igual a $\dfrac{2}{b-a}$.

Distribución T - Student

Una variable aleatoria continua x tiene distribución T-student con parámetros μ, λ y α ($\mu \in \Re, \lambda > 0, \alpha > 0$) si su finción de densidad $St(x|\mu, \lambda, \alpha)$ es

$$St(x|\mu, \lambda, \alpha) = c \left[1 + \frac{\lambda}{\alpha}(x - \mu)^2\right]^{-(\alpha+1)/2},$$

para $x \in \Re$

donde

$$c = \frac{\Gamma((\alpha+1)/2)}{\Gamma(\alpha/2)\Gamma(1/2)} \left(\frac{\lambda}{\alpha}\right)^{1/2}.$$

La distribución es simétrica con respecto a $x = \mu$, y tiene su única moda en $M[x] = \mu$.

La media y la varianza son

$$E[x] = \mu, \qquad si \;\; \alpha > 1,$$

$$V[x] = \frac{1}{\lambda} \frac{\alpha}{(\alpha - 2)}, \qquad si \;\; \alpha > 2.$$

El parámetro α usualmente se refiere a los grados de libertad de la distribución.

La distribución es generada por la combinación

$$St(x|\mu, \lambda, \alpha) = \int_0^\infty N(x|\mu, \lambda y) \;\; Ga\left(y|\frac{\alpha}{2}, \frac{\alpha}{2}\right) \; dy.$$

e incluye a la distribución normal como un caso de límite, tal que

$$N(x|\mu, \lambda) = \lim_{\alpha \to \infty} St(x|\mu, \lambda, \alpha).$$

Si $y = \lambda^{1/2}$, donde x tiene densidad $St(x|\mu, \lambda, \alpha)$, entonces y tiene densidad T-student estándar $St(y|0, 1, \alpha)$. Si $\alpha = 1$ se dice que x tiene distribución Cauchy con densidad $Ca(x|\mu, \lambda)$.

Si x tiene distribución normal estándar, y tiene distribución χ^2_v, y si x y y son mutuamente independientes, entonces

$$z = \frac{x}{(y/v)^{1/2}}$$

tiene densidad T-student estándar $St(z|0, 1, v)$.

Distribución Uniforme

Una variable aleatoria continua x tiene distribución uniforme con parámetros a y b ($a \in \Re, b \in \Re, a < b$) si su función de densidad $U(x|a, b)$ es

$$U(x|a, b) = \frac{1}{b-a},$$

para $a < x < b$.

La media y la varianza están dadas por $E[x] = \dfrac{a+b}{2}$ y $V[x] = \dfrac{b-a}{12}$.

La distribución uniforme carece de moda.

Distribución Weibull

Una variable aleatoria continua x tiene distribución Weibull con parámetros a y b ($a > 0, b > 0$) si su función de densidad $W(x|a, b)$ es

$$W(x|a, b) \;=\; a \;.\; b \; x^{b-1} \; e^{-ax^b},$$

para $x > 0$.

La media y la varianza están dadas por

$$E[x] = \left(\frac{1}{a}\right)^{\frac{1}{b}} \Gamma\left(\frac{1}{b+1}\right)$$

y

$$V[x] = \left(\frac{1}{a}\right)^{\frac{2}{b}} \left(\Gamma\left(\frac{2}{b+1}\right) - \Gamma^2\left(\frac{1}{b+1}\right)\right).$$

Si $b = 1$ entonces $W(x|a, 1) = Ga(x|1, 1/a) = Exp(x|1/a)$.

Si $a = 1/\beta$ y $b = 1$ entonces $W(x|1/\beta, 1) = Exp(x|\beta)$.

Si $a = 1$ y $b = 1$ entonces $W(x|1, 1) = Ga(x|1, 1) =$ Exponencial estandarizada.

2.3. Distribuciones discretas multivariadas

Distribución Multinomial

Un vector aleatorio discreto $\mathbf{x} = (x_1, ..., x_k)$ tiene una distribución multinomial de dimensión k, con parámetros $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ y n $(0 < \theta_i < 1, \sum_{i=1}^{n} \theta_i < 1, n = 1, 2, ...)$ si su función de probabilidad $Mu_k(\mathbf{x}|\boldsymbol{\theta}, n)$, para $x_i = 0, 1, 2, ...$, con $\sum_{i=1}^{n} x_i \leq n$, es

$$Mu_k(\mathbf{x}|\boldsymbol{\theta}, n) = \frac{n!}{\prod_{i=1}^{k} x_i!(n - \sum_{i=1}^{n} x_i)!} \prod_{i=1}^{k} \theta_i^{x_i} (1 - \sum_{i=1}^{k} \theta_i)^{n - \sum_{i=1}^{n} x_i}.$$

El vector de medias y la matriz de covarianzas están dados por

$$E[x_i] = n\theta_i, \qquad V[x_i] = n\theta_i(1 - \theta_i), \qquad C[x_i, x_j] = -n\theta_i\theta_j.$$

La moda (s) de la distribución es (son) localizada cerca de $E[\mathbf{x}]$, satisfaciendo

$$n\theta_i < M[x_i] \leq (n+k-1)\theta_i, \quad i = 1, ..., k;$$

estas desigualdades, con la condición $\sum_{i=1}^{n} x_i \leq n$, restringen las posibles modas relativamente en algunos puntos.

La distribución marginal de $\mathbf{x}^{(m)} = (x_1, ..., x_m)$, $m < k$, es la ditribución multinomial $Mu_m(\mathbf{x}^{(m)}|\theta_1, ..., \theta_m, n)$.

La distribución condicional de $\mathbf{x}^{(m)}$, dado el recordatorio de saber que cada $x_i's$ tiene también distribución multinomial y que dicha distribución depende de las $x_i's$ solo a través de su suma $s = \sum_{i=m+1}^{n} x_i$; especificamente es

$$p(\mathbf{x}^{(m)}|x_{m+1}, ..., x_k) = Mu_k\left(\mathbf{x}^{(m)}\middle|\frac{\theta_1}{\sum_{i=1}^{m}\theta_j}, ..., \frac{\theta_m}{\sum_{i=1}^{m}\theta_j}, n-s\right).$$

Si $\mathbf{x} = (x_1, ..., x_k)$ tiene densidad $Mu_k(\mathbf{x}|\boldsymbol{\theta}, n)$, entonces $\mathbf{y} = (y_1, ..., y_t)$ donde

$$y_1 = x_1 + ... + x_{i1}, ..., y_t = x_{i_{t-1}+1} + ... + x_{ik}, \qquad 1 \leq t < k,$$

tiene densidad $Mu_t(\mathbf{y}|\boldsymbol{\phi}, n)$, donde

$$\phi_1 = \theta_1 + ... + \theta_{i1}, ..., \phi_t = \theta_{i_{t-1}+1} + ... + \theta_{ik}.$$

Si $\mathbf{z}$ es la suma de m vectores aleatorios independientes teniendo densidades multinomiales con parámetros $(\boldsymbol{\theta}, n_i)$, $i = 1, ..., m$, entonces $\mathbf{z}$ también tiene densidad multinomial con parámetros $\boldsymbol{\theta}$ y $(n_1 + ... + n_m)$. Si $k = 1$, $Mu_k(\mathbf{x}|\boldsymbol{\theta}, n)$ se reduce a una función de densidad binomial $Bi(x|\theta, n)$.

Si $x_1, ..., x_k$ son k cantidades aleatorias independientes con distribución de probabilidad Poisson $Pn(x_i|\lambda_i)$, entonces la distribución conjunta de $\mathbf{x} = (x_1, ..., x_k)$, dado que $\sum_{i=1}^{n} x_i = n$, es multinomial $Mu(x|\boldsymbol{\theta}, n)$, con $\theta_i = \frac{\lambda_i}{\sum_{i=1}^{n}\lambda_j}$.

Distribución Multinomial - Dirichlet[5]

Un vector aleatorio discreto $\mathbf{x} = (x_1, ..., x_k)$ tiene distribución multinomial - Dirichlet de dimensión k, con parámetros $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_{k+1})$ y n donde $\alpha_i > 0$, y $n = 1, 2, ...$, si su función de probabilidad $Dm(\mathbf{x}|\boldsymbol{\alpha}, n)$, para $x_i = 0, 1, 2, ...$, con $\sum_{i=1}^{n} x_i \leq n$, es

$$Dm(\mathbf{x}|\boldsymbol{\alpha}, n) = c \prod_{i=1}^{k+1} \left(\frac{\alpha_j^{[x_j]}}{x_j!} \right)$$

donde $\alpha^{[s]} = \prod_{j=1}^{s} (\alpha + j - 1)$ define la función factorial ascendente, con $x_{k+1} = n - \sum_{j=1}^{k} x_j$ y

$$c = \frac{n!}{\sum_{j=1}^{k+1} \alpha_j^{[n]}}.$$

El vector de medias y la matriz de covarianzas están dadas por

$$E[x_i] = np_i, \qquad p_i = \frac{\alpha_i}{\sum_{j=1}^{k+1} \alpha_j}$$

$$V[x_i] = \frac{n + \sum_{j=1}^{k+1} \alpha_j}{1 + \sum_{j=1}^{k+1} \alpha_j} np_i(1 - p_i).$$

$$C[x_i, x_j] = -\frac{n + \sum_{j=1}^{k+1} \alpha_j}{1 + \sum_{j=1}^{k+1} \alpha_j} np_i p_j.$$

[5] Para la comprensión de esta distribución es necesario conocer la distribución de la variable aleatoria continua multivariada Dirichlet.

La distribución marginal del subconjunto $\{x_1, ..., x_s\}$ es una multinomial - Dirichlet con parámetros $\{\alpha_1, ..., \alpha_s, \sum_{j=1}^{k+1} \alpha_j - \sum_{j=1}^{s} \alpha_j\}$ y n. En particular, la distribución marginal de x_i es binomial - beta $Bb(x_i|\alpha_i, \sum_{j=1}^{k+1} \alpha_j - \alpha_i)$. Más aún, la distribución condicional de $\{x_{s+1}, ..., x_k\}$ dado $\{x_1, ..., x_s\}$ es también multinomial - Dirichlet, con parámetros $\{\alpha_{s+1}, ..., \alpha_k, \sum_{j=1}^{k+1} \alpha_j - \sum_{j=s+1}^{k} \alpha_j\}$ y $n - \sum_{j=1}^{s} x_j$.

2.4. Distribuciones continuas multivariadas

Distribución Dirichlet

Un vector aleatorio continuo $\mathrm{x} = (x_1, ..., x_k)$ tiene distribución Dirichlet de dimensión k, con parámetros $\alpha = (\alpha_1, ..., \alpha_{k+1})$ $(\alpha_i > 0, i = 1, ..., k+1)$ si su densidad de probabilidad $Di(\mathrm{x}|\alpha)$, $0 < x_i < 1$ y $x_1 + ... + x_k < 1$, es

$$Di_k(\mathrm{x}|\alpha) = c\ x_i^{\alpha_1 - 1} ... x_k^{\alpha_k - 1}(1 - \sum_{i=1}^{k} x_i)^{\alpha_{k+1} - 1},$$

donde

$$c = \frac{\Gamma(\sum_{i=1}^{k+1} \alpha_i)}{\prod_{i=1}^{k+1} \Gamma(\alpha_i)}.$$

Si $k = 1$, $Di(\mathrm{x}|\alpha)$ se reduce a una densidad de probabilidad beta $Be(x|\alpha_1, \alpha_2)$.

En el caso general, el vector de medias y la matriz de covarianzas están dados por

$$E[x_i] = \frac{\alpha_i}{\sum_{j=1}^{k+1} \alpha_j}, \qquad V[x_i] = \frac{E[x_i](1 - E[x_i])}{1 + \sum_{j=1}^{k+1} \alpha_j}, \qquad C[x_i, x_j] = \frac{-E[x_i]E[x_j]}{1 + \sum_{j=1}^{k+1} \alpha_j}.$$

Si $\alpha_i > 1$, $i = 1, ..., k$, se define la moda dada por

$$M[x_i] = \frac{\alpha_i - 1}{\sum\limits_{j=1}^{k+1} \alpha_j - k - 1}.$$

La distribución marginal de $\mathbf{x}^{(m)} = (x_1, ..., x_m)$, $m < k$, es distribución Dirichlet

$$p(\mathbf{x}^{(m)}) = Di_m(\mathbf{x}^{(m)} | \alpha_1, ..., \alpha_m, \sum_{j=m+1}^{k+1} \alpha_j).$$

La distribución condicional, dado $x_{m+1}, ..., x_k$, de

$$x_i' = \frac{x_i}{1 - \sum\limits_{j=m+1}^{k} x_j}, \qquad i = 1, ..., m$$

es también Dirichlet, $Di_m(x_1', ..., x_m | \alpha_1, ..., \alpha_m, \alpha_{k+1})$. En particular,

$$p(x_i' | x_{m+1}, ..., x_k) = Be(x_i' | \alpha_i, \sum_{j=1}^{m} \alpha_j + \alpha_{k+1} - \alpha_i), \qquad i = 1, ..., m.$$

Más aún, si $\mathbf{x} = (x_1, ..., x_k)$ tiene densidad de probabilidad $Di_k(\mathbf{x}|\boldsymbol{\alpha})$, entonces $\mathbf{y} = (y_1, ..., y_t)$ donde

$$y_1 = x_1 + ... + x_{i_1}, ..., y_t = x_{i_{t-1}+1} + ... + x_k, \qquad 1 \leq t < k,$$

tiene densidad de probabilidad $Di_t(\mathbf{y}|\boldsymbol{\beta})$, donde

$$\beta_1 = \alpha_1 + ... + \alpha_{i_1}, ..., \beta_t = \alpha_{i_{t-1}+1} + ... + \alpha_k, \beta_{t+1} = \alpha_{k+1}.$$

Distribución Normal - Gamma

Un vector aleatorio continuo bivariado (x, y) tiene distribución normal - gamma, con parámetros μ, λ, α y β, $(\mu \in \Re, \lambda > 0, \alpha > 0, \beta > 0)$ si su densidad de probabilidad $Ng(x, y|\mu, \lambda, \alpha, \beta)$ es

$$Ng(x, y|\mu, \lambda, \alpha, \beta) = N(x|\mu, \lambda y) \; Ga(y|\alpha, \beta), \qquad x \in \Re, y > 0,$$

donde las densidades de probabilidad normal y gamma son definidas en la sección 2.2.

Las medias y varianzas están dadas por

$$E[x] = \mu, \quad E[y] = \frac{\alpha}{\beta} \qquad V[x] = \frac{\beta}{\lambda(\alpha-1)} \quad V[y] = \frac{\alpha}{\beta^2}$$

Es claro, de acuerdo a la función dada, que la función de densidad condicional de x dado y es $N(x|\mu, \lambda y)$ y que la función de densidad marginal de y es $Ga(y|\alpha, \beta)$. Ms aún, la función de densidad marginal de x es $St(x|\mu, \frac{\lambda\alpha}{\beta}, 2\alpha)$.

Distribución Normal - Gamma Multivariada

Un vector aleatorio continuo $\mathbf{x} = (x_1, ..., x_n)$ y una cantidad aleatoria y tienen una distribución normal - gamma conjunta multivariada de dimensión k, con parámetros $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$, α, β ($\boldsymbol{\mu} \in \Re^k$, $\boldsymbol{\lambda}$ una matriz simétrica positiva definida de k x k, $\alpha > 0$ y $\beta > 0$) si la densidad de probabilidad conjunta de $\mathbf{x}$ y y, $Ng(\mathbf{x}, y|\boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha, \beta)$ es

$$Ng(\mathbf{x}, y|\boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha, \beta) = N_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\lambda} y) \;\; Ga(y|\alpha, \beta),$$

donde las densidades de probabilidad multivariadas gamma y normal han sido ya definidas en la sección 2.2.

El vector de medias, la matriz de covarianzas para $\mathbf{x}$ y la varianza de y están dadas por

$$E[\mathbf{x}, y] = (\boldsymbol{\mu}, \frac{\alpha}{\beta}), \quad V[\mathbf{x}] = \frac{\beta}{\boldsymbol{\lambda}(\alpha-1)}, \quad V[y] = \frac{\alpha}{\beta^2}.$$

De la función de densidad anterior, se tiene que, la densidad condicional de $\mathbf{x}$ dado y es $N_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\lambda} y)$ y la densidad de probabilidad marginal de y es $Ga(y|\alpha, \beta)$. Más aún, la densidad de probabilidad marginal de $\mathbf{x}$ es $St_k(\mathbf{x}|\boldsymbol{\mu}, \frac{\beta\boldsymbol{\lambda}}{\alpha}, 2\alpha)$.

Distribución Normal Multivariada

Un vector aleatorio continuo $\mathbf{x} = (x_1, ..., x_k)$ tiene distribución normal multivariada de dimensión k, con parámetros $\boldsymbol{\mu} = (\mu_1, ..., \mu_k)$ y $\boldsymbol{\lambda}$, donde $\boldsymbol{\mu} \in \Re^k$ y $\boldsymbol{\lambda}$ es una matriz simétrica positiva definida de $k \times k$, si su densidad de probabilidad $N_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\lambda})$ es

$$N_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\lambda}) = c\ e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\boldsymbol{\lambda}(\mathbf{x}-\boldsymbol{\mu})}, \qquad \mathbf{x} \in \Re^k,$$

donde $c = (2\pi)^{-\frac{k}{2}}|\boldsymbol{\lambda}|^{-\frac{1}{2}}$.

Si $k = 1$, como $\boldsymbol{\lambda}$ es un escalar, λ, $N_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\lambda})$ se reduce a una densidad de probabilidad normal univariada $N(x|\mu, \lambda)$.

En le caso general, $E[x_i] = \mu_i$, y, con $\Sigma = \boldsymbol{\lambda}^{-1}$ para cualquier elemento σ_{ij}, $V[x_i] = \sigma_{ii}$ y $C[x_i, x_j] = \sigma_{ij}$, tal que $V[\mathbf{x}] = \boldsymbol{\lambda}^{-1}$. El parámetro $\boldsymbol{\mu}$ por lo tanto etiqueta el vector de medias y el parámetro $\boldsymbol{\lambda}$ la matriz de precisiones (la matriz inversa a la matriz de covarianzas Σ).

Si $\mathbf{y} = \mathbf{A}\mathbf{x}$, donde $\mathbf{A}$ es una matriz de $m \times k$ de números reales tal que $\mathbf{A}\Sigma\mathbf{A}^t$ es no singular[6] , entonces $\mathbf{y}$ tiene densidad de probabilidad $N_m(\mathbf{y}|\mathbf{A}\boldsymbol{\mu}, (\mathbf{A}\Sigma\mathbf{A}^t)^{-1})$.

En particular, la densidad de probabilidad marginal para cualquier subvector de $\mathbf{x}$ es normal (multivariada), de dimensión apropiada, con vector de medias y matriz de covarianzas dados correspondientemente por el subvector $\boldsymbol{\mu}$ y la submatriz $\boldsymbol{\lambda}^{-1}$. Más aún, si $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ es una partición de $\mathbf{x}$, con $\mathbf{x}_i$ de dimensión k_i, y $k_1 + k_2 = k$, y las particiones correspondientes de $\boldsymbol{\mu}$ y $\boldsymbol{\lambda}$ son

$$\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix}\boldsymbol{\lambda}_{11} & \boldsymbol{\lambda}_{12}\\ \boldsymbol{\lambda}_{21} & \boldsymbol{\lambda}_{22}\end{pmatrix},$$

entonces la densidad de probabilidad condicional de $\mathbf{x}_1$ dado $\mathbf{x}_2$ es también normal (multivariada), de dimensión k_1 con vector de medias y matriz de precisiones dados, respectivamente, por

$$\boldsymbol{\mu}_1 - \boldsymbol{\lambda}_{11}^{-1}\boldsymbol{\lambda}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \ y \ \boldsymbol{\lambda}_{11}.$$

[6]Una matriz cuya inversa existe es no singular.

La cantidad aleatoria $y = (\mathbf{x} - \boldsymbol{\mu})^t\lambda(\mathbf{x} - \boldsymbol{\mu})$ tienes densidad de probabilidad $\chi^2(y|k)$. Y, también es preciso notar que, de la forma de la función de densidad se puede deducir que

$$\int_{\Re^k} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\lambda(\mathbf{x}-\boldsymbol{\mu})}\, d\mathbf{x} = \frac{(2\pi)^{\frac{k}{2}}}{|\lambda|^{\frac{1}{2}}}.$$

Distribución Normal - Wishart Multivariada

Un vector aleatorio continuo $\mathbf{x}$ y una matriz simétrica positiva definida de cantidades aleatoria $\mathbf{y}$ tienen una distribución conjunta multivariada normal-Wishart multivariada de dimensión k, con parámetros $\boldsymbol{\mu}$, λ, α, $\boldsymbol{\beta}$ ($\boldsymbol{\mu} \in \Re^k$, $\lambda > 0$, $2\alpha > k - 1$ entero, y $\boldsymbol{\beta}$ una matriz simétrica no singular de k x k), si la densidad de probabilidad de $\boldsymbol{x}$ y de los $\frac{k(k+1)}{2}$ elementos distintos de $\boldsymbol{y}$, $N_{w_k}(\mathbf{x}, \mathbf{y}|\boldsymbol{\mu}, \lambda, \alpha, \boldsymbol{\beta})$ es

$$Nw_k(\mathbf{x}, \mathbf{y}|\boldsymbol{\mu}, \lambda, \alpha, \boldsymbol{\beta}) = N_k(\mathbf{x}|\boldsymbol{\mu}, \lambda\mathbf{y}) \;\; Wi_k(\mathbf{y}|\alpha, \boldsymbol{\beta}),$$

donde la densidad de probabilidad normal multivariada ha sido ya definida y la densidad de probabilidad Wishart está definida en páginas posteriores de la presente sección.

Las medias y las varianzas están definidas respectivamente por

$$E[\mathbf{x}, \mathbf{y}] = \{\boldsymbol{\mu}, \frac{\alpha}{\boldsymbol{\beta}}\} \quad y \quad V[\mathbf{x}] = \frac{\boldsymbol{\beta}}{\lambda(\alpha - 1)}$$

De la función de densidad de probabilidad se tiene que la densidad de probabilidad condicional de $\mathbf{x}$ dado $\mathbf{y}$ es $N_k(\mathbf{x}|\boldsymbol{\mu}, \lambda\mathbf{y})$ y la densidad de probabilidad marginal de $\mathbf{y}$ es $W_{i_k}(\mathbf{y}|\alpha, \boldsymbol{\beta})$. Más aún, la densidad marginal de $\mathbf{x}$ es $St_k(\mathbf{x}|\boldsymbol{\mu}, \frac{\lambda\alpha}{\boldsymbol{\beta}}, 2\alpha)$.

Distribución Pareto Bilateral

Un vector aleatorio bivariado continuo (x, y) tiene distribución Pareto bilateral con parámetros β_0, β_1, y α ($\{\beta_0, \beta_1\} \in \Re^2, \beta_0 < \beta_1, \alpha > 0$) si su función de densidad de probabilidad $Pa_2(x, y|\alpha, \beta_0, \beta_1)$ es

$$Pa_2(x,y|\alpha,\beta_0,\beta_1) = c\ (y-x)^{-(\alpha+2)} \quad , x \leq \beta_0\ ,\ y \geq \beta_1,$$

donde $c = \alpha(\alpha+1)(\beta_1-\beta_0)^\alpha$

La medias y varianzas están dadas por

$$E[x] = \frac{\alpha\beta_0-\beta_1}{\alpha-1}, \quad E[y] = \frac{\alpha\beta_1-\beta_0}{\alpha-1}, \quad si \quad \alpha > 1,$$

$$V[x] = V[y] = \frac{\alpha(\beta_1-\beta_0)^2}{(\alpha-1)^2(\alpha-2)}, \quad si \quad \alpha > 2,$$

y la correlación entre x y y es $-\alpha^{-1}$.

Las distribuciones marginales de $t_1 = \beta_1 - x$ y $t_2 = y - \beta_0$ son ambas $Pa(t|\beta_1-\beta_0,\alpha)$.

Distribución T - Student Multivariada

Un vector aleatorio continuo $\mathbf{x} = (x_1,...,x_k)$ tiene distribución T-student multivariada de dimensión k, con parámetros $\boldsymbol{\mu} = (\mu_1,...,\mu_k)$, $\boldsymbol{\lambda}$ y α ($\boldsymbol{\mu} \in \Re^k$, $\boldsymbol{\lambda}$ una matriz simétrica positiva definida de dimensión k x k, $\alpha > 0$) si su densidad de probabilidad $St(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\lambda},\alpha)$ es

$$St_k(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\lambda},\alpha) = c\ \left[1+\frac{1}{\alpha}(\mathbf{x}-\boldsymbol{\mu})^t\boldsymbol{\lambda}(\mathbf{x}-\boldsymbol{\mu})\right]^{-(\alpha+k)/2},$$

para $\mathbf{x} \in \Re^k$,

donde

$$c = \frac{\Gamma\left(\frac{(\alpha+k)}{2}\right)}{\Gamma(\frac{\alpha}{2})(\alpha\pi)^{k/2}}|\boldsymbol{\lambda}|^{1/2}.$$

Si $k=1$, tal que $\boldsymbol{\lambda}$ es un escalar, λ, entonces $St_k(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\lambda},\alpha)$ se reduce a una densidad de probabilidad univariada $St(x|\mu,\lambda,\alpha)$.

En el caso general, $E[\mathbf{x}] = \boldsymbol{\mu}$ y $V[\mathbf{x}] = \boldsymbol{\lambda}^{-1}(\alpha/(\alpha-2))$. Aunque no exactamente igual a la matriz inversa de matriz de covarianzas, el parámetro $\boldsymbol{\lambda}$ a menudo referida como la matriz de presición de la distribución. Si $\mathbf{y} = \mathbf{A}\mathbf{x}$, donde $\mathbf{A}$ es una matriz de dimensión m x k ($m \leq k$) de números reales tal que la matriz $\mathbf{A}\boldsymbol{\lambda}^{-1}\mathbf{A}^t$ sea no singular, entonces $\mathbf{y}$ tiene densidad $St_m(\mathbf{y}|\mathbf{A}\boldsymbol{\mu}, (\mathbf{A}\boldsymbol{\lambda}^{-1}\mathbf{A}^t)^{-1}, \alpha)$. En paticular, la densidad marginal para cualquier subvector de $\mathbf{x}$ es T-student (multivariada)., con su dimensión apropiada, con vector de medias y matriz de presición dados por el correspondiente subvector de $\boldsymbol{\mu}$ y submatriz de $\boldsymbol{\lambda}$. Más aún, si $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ es una partición de $\mathbf{x}$ y las particiones correspondientes de $\boldsymbol{\mu}$ y $\boldsymbol{\lambda}$ están dadas por

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda}_{11} & \boldsymbol{\lambda}_{12} \\ \boldsymbol{\lambda}_{21} & \boldsymbol{\lambda}_{22} \end{pmatrix},$$

entonces la densidad condicional de $\mathbf{x}_1$, dado $\mathbf{x}_2$ es también T-student (multivariada), de dimensión k_1, con $\alpha + k_2$ grados de libertad, y el vector de medias y la matriz de presición, respectivamente, están dados por

$$\boldsymbol{\mu}_1 - \boldsymbol{\lambda}_{11}^{-1}\boldsymbol{\lambda}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2),$$

y

$$\boldsymbol{\lambda}_{11}\left[\frac{\alpha + k_2}{\alpha + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^t(\boldsymbol{\lambda}_{22} - \boldsymbol{\lambda}_{21}\boldsymbol{\lambda}_{11}^{-1}\boldsymbol{\lambda}_{12})^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)}\right].$$

Distribución Wishart

Una matriz simétrica positiva definida $\mathbf{x}$ de variables aleatorias tal que $x_{ij} = x_{ji}$, para $i = 1, ..., k$, $j = 1, ..., k$, tiene distribución Wishart de dimensión k, con parámetros α y $\boldsymbol{\beta}$ (con $2\alpha > k - 1$ y $\boldsymbol{\beta}$ una matriz simétrica no singular de k x k), si la densidad de probabilidad $Wi_k(\mathbf{x}|\alpha, \boldsymbol{\beta})$ para los distintos $\frac{k(k+1)}{2}$ vectores aleatorios dimensionales de $\mathbf{x}$ es

$$Wi_k(\mathbf{x}|\alpha, \boldsymbol{\beta}) = c|\mathbf{x}|^{\alpha-(k+1)/2} \ . \ e^{-tr(\boldsymbol{\beta}\mathbf{x})},$$

donde $c = |\boldsymbol{\beta}|^{\alpha}/\Gamma_k(\alpha)$,

$$\Gamma_k(\alpha) = \pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\Big(\frac{2\alpha+1-i}{2}\Big)$$

es la función gamma generalizada y $tr(.)$, denota la traza del argumento de la matriz[7].

Si $k = 1$, tal que $\boldsymbol{\beta}$ es un escalar, β, entonces $W_k(\mathbf{x}|\alpha, \boldsymbol{\beta})$ se reduce a una densidad de probabilidad gamma $Ga(x|\alpha, \beta)$.

Si $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ es una muestra aleatoria de tamaño $n > 1$ de la distribución normal multivariada $N_k(\mathbf{x}_i|\boldsymbol{\mu}, \boldsymbol{\lambda})$, y $\bar{\mathbf{x}} = n^{-1} \sum_{i=1}^{n} \mathbf{x}_i$, entonces $\bar{\mathbf{x}}$ tiene densidad de probabilidad $N_k(\bar{\mathbf{x}}|\boldsymbol{\mu}, n\boldsymbol{\lambda})$, y

$$S = \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t$$

es independiente de $\bar{\mathbf{x}}$, y tiene distribución Wishart $Wi_k(S|\frac{1}{2}(n-1), \frac{1}{2}\boldsymbol{\lambda})$.

Algunas de las características que definen una distribución Wishart son las siguientes: $E\mathbf{x}] = \alpha\boldsymbol{\beta}^{-1}$ y $E[\mathbf{x}^{-1}] = (\alpha - (k+1)/2)^{-1}\boldsymbol{\beta}$; si $\mathbf{y} = \mathbf{A}\mathbf{x}\mathbf{A}^t$ donde $\mathbf{A}$ es una matriz de m x k ($m \leq k$) de números reales, entonces $\mathbf{y}$ tiene distribución Wishart de dimensión m con parámetros α y $(\mathbf{A}\boldsymbol{\beta}^{-1}\mathbf{A}^t)^{-1}$, si es que existe; en particular, si $\mathbf{x}$ y $\boldsymbol{\beta}^{-1}$ son particionadas conformablemente en

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} \\ \mathbf{x}_{21} & \mathbf{x}_{22} \end{pmatrix}, \quad \boldsymbol{\beta}^{-1} = \begin{pmatrix} \boldsymbol{\sigma}_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\sigma}_{22} \end{pmatrix},$$

donde $\mathbf{x}_{11}$, $\boldsymbol{\sigma}_{11}$ son matrices cuadradas de h x h ($1 \leq h \leq k$), entonces $\mathbf{x}_{11}$ tiene distribución Wishart de dimensión h con parámetros α y $(\boldsymbol{\sigma}_{11})^{-1}$. Más aún, si $\mathbf{x}_1, ..., \mathbf{x}_s$ son matrices aleatorias independientes de k x k, cada una con distribución Wishart, con parámetros α_i, $\boldsymbol{\beta}$, $i = 1, ..., s$, entonces $\mathbf{x}_1 + ... + \mathbf{x}_s$ también tiene distribución Wishart, con parámetros $\alpha_1 + ... + \alpha_s$ y $\boldsymbol{\beta}$.

Es notable que, de la forma de la densidad de probabilidad Wishart se deduce que

[7]La traza de una matriz cuadrada$k$ x k es la suma de los elementos de la diagonal.

$$\int |\mathbf{x}|^{\alpha-(k+1)/2} \cdot e^{-tr(\beta \mathbf{x})} \; d\mathbf{x} = c^{-1},$$

siendo entendido que la integración es con respecto a los $\dfrac{k(k+1)}{2}$ elementos distintos de la matriz $\mathbf{x}$.

Capítulo 3

Modelos: Una interacción matemática entre estadística bayesiana y distribuciones de probabilidad

La aplicación de metodologías bayesianas para la construcción de modelos referentes a problemas reales y de la vida cotidiana, resulta un tanto laboriosa debido a la gran diversidad de comportamientos y a la complejidad de manipulación en operaciones para obtener las distribuciones de probabilidad deseables. Sin embargo, existen modelos a los que comúnmente se acude para hacer inferencia estadística y, para proporcionar alguna idea sobre el funcionamiento de las técnicas bayesianas, solo se hará referencia a algunos casos muy simples y ya definidos.

Estos modelos son creados a través de dos distribuciones de probabilidad dadas: la distribución de probabilidad $p(x|\theta)$, y la distribución de probabilidad paramétrica a priori, es decir, $p(\theta)$. A partir de dichas distribuciones se genera completamente un modelo particular y el desarrollo de cada modelo en particular se basa en los procedimientos explicados en el Capítulo I referentes al modelo general.

No obstante, es necesario hacer incapié en el procedimiento que antecede dichos modelos, ya que es necesario poder visualizar la aplicación de la teoría fundamental que sustenta al enfoque bayesiano, así como comprobar los resultados una vez ya establecidos.

Los modelos que en particular se desglosarán a detalle, serán modelos inferenciales uniparamétricos, es decir, los que únicamente constan de un solo parámetro aleatorio; y por medio de dicho desarrollo se establecerán, en el presente orden y para cada uno de los modelos, la distribución de probabilidad predictiva inicial o a priori, la distribución de probabilidad paramétrica final o a posteriori y la distribución predictiva final o a posteriori.

Los modelos serán los siguientes:

- Modelo Bernoulli
- Modelo Poisson
- Modelo Binomial Negativa
- Modelo Exponencial
- Modelo Uniforme
- Modelo Normal con precisión λ conocida
- Modelo Normal con media μ conocida

3.1. Modelo Bernoulli

Sea la muestra aleatoria $x_1, x_2, ..., x_n$ tal que $x_j \sim Bernoulli(\theta)$ con θ desconocida $\Rightarrow f_X \in \mathcal{P} = \{p(x|\theta) = \theta^x(1-\theta)^{1-x} \ \mathbb{I}_{\{0,1\}}^{(x)} : \theta \in \Theta =]0,1[\}$.

Sea la distribución inicial propuesta para θ tal que $\theta \sim Beta(\alpha, \beta)$

$\Rightarrow$

$$p(x) = \int_{\Theta} p(x|\tilde{\theta}) \ p(\tilde{\theta}) \ d\tilde{\theta}$$

$$= \int_{\Theta} \tilde{\theta}^{x} \; (1-\tilde{\theta})^{1-x} \; \mathbb{I}_{\{0,1\}}^{(x)} \; \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\;\Gamma(\beta)} \; \tilde{\theta}^{\alpha-1} \; (1-\tilde{\theta})^{\beta-1} \; \mathbb{I}_{]0,1[}^{(\theta)} \; d\tilde{\theta} \; =$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\;\Gamma(\beta)} \int_0^1 \tilde{\theta}^{x+\alpha-1} \; (1-\tilde{\theta})^{\beta-x} \; \mathbb{I}_{\{0,1\}}^{(x)} \; d\tilde{\theta} \; =$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\;\Gamma(\beta)} \int_0^1 \tilde{\theta}^{(\alpha+x)-1} \; (1-\tilde{\theta})^{(\beta-x+1)-1} \; \mathbb{I}_{\{0,1\}}^{(x)} \; d\tilde{\theta}$$

Para resolver la integral anterior se tiene que

$$\int_0^1 \frac{\Gamma(\alpha+x+\beta-x+1)}{\Gamma(\alpha+x)\;\Gamma(\beta-x+1)} \; \tilde{\theta}^{(\alpha+x)-1} \; (1-\tilde{\theta})^{(\beta-x+1)-1} \; d\tilde{\theta} \; = \; 1$$

$$\Rightarrow \int_0^1 \tilde{\theta}^{(\alpha+x)-1} \; (1-\tilde{\theta})^{(\beta-x+1)-1} \; d\tilde{\theta} \; = \; \frac{\Gamma(\alpha+x)\;\Gamma(\beta-x+1)}{\Gamma(\alpha+x+\beta-x+1)}$$

$$\Rightarrow \; p(x) \; = \; \frac{\Gamma(\alpha+\beta)\;\Gamma(\alpha+x)\;\Gamma(\beta-x+1)}{\Gamma(\alpha)\;\Gamma(\beta)\;\Gamma(\alpha+\beta+1)} \; \mathbb{I}_{\{0,1\}}^{(x)}$$

$$\therefore \quad x \sim Binomial - Beta(\alpha, \beta, 1)$$

Por otra parte y considerando que las observaciones provienen de una muestra aleatoria, la función de verosimilitud está dada de la siguiente manera:

$$p(\mathbf{x}|\theta) \; = \; \theta^{\sum_{i=1}^{n} x_i} \; (1-\theta)^{n-\sum_{i=1}^{n} x_i} \; \prod_{i=1}^{n} \mathbb{I}_{\{0,1\}}^{(x)}$$

$$\Rightarrow \; p(\theta|\mathbf{x}) \; = \; \frac{p(\mathbf{x}|\theta)\;p(\theta)}{\int_{\Theta} p(\mathbf{x}|\tilde{\theta})\;p(\tilde{\theta})\;d\tilde{\theta}} \; =$$

$$= \frac{\theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \mathbb{I}_{\{0,1\}}^{(x_i)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \mathbb{I}_{]0,1[}^{(\theta)}}{\int_0^1 \tilde{\theta}^{\sum_{i=1}^{n} x_i} (1-\tilde{\theta})^{n-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \mathbb{I}_{\{0,1\}}^{(x_i)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)} \tilde{\theta}^{\alpha-1} (1-\tilde{\theta})^{\beta-1} d\tilde{\theta}}$$

$$= \frac{\theta^{\sum_{i=1}^{n} x_i+\alpha-1} (1-\theta)^{n-\sum_{i=1}^{n} x_i+\beta-1} \mathbb{I}_{]0,1[}^{(x)}}{\int_0^1 \tilde{\theta}^{(\alpha+\sum_{i=1}^{n} x_i)-1} (1-\theta)^{(\beta+n-\sum_{i=1}^{n} x_i)-1} d\tilde{\theta}}$$

Para dar solución a la integral del denominador se tiene que

$$\int_0^1 \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\sum_{i=1}^{n} x_i)\ \Gamma(\beta+n-\sum_{i=1}^{n} x_i)} \tilde{\theta}^{(\alpha+\sum_{i=1}^{n} x_i)-1} \cdot$$

$$\cdot\ (1-\theta)^{(\beta+n-\sum_{i=1}^{n} x_i)-1} d\tilde{\theta} = 1$$

$$\Rightarrow \int_0^1 \tilde{\theta}^{(\alpha+\sum_{i=1}^{n} x_i)-1} (1-\theta)^{(\beta+n-\sum_{i=1}^{n} x_i)-1} d\tilde{\theta} =$$

$$= \frac{\Gamma(\alpha+\sum_{i=1}^{n} x_i)\ \Gamma(\beta+n-\sum_{i=1}^{n} x_i)}{\Gamma(\alpha+\beta+n)}$$

$$\Rightarrow p(\theta|\mathbf{x}) = \frac{\theta^{\sum_{i=1}^{n} x_i+\alpha-1} (1-\theta)^{n-\sum_{i=1}^{n} x_i+\beta-1} \mathbb{I}_{]0,1[}^{(x)}}{\frac{\Gamma(\alpha+\sum_{i=1}^{n} x_i)\ \Gamma(\beta+n-\sum_{i=1}^{n} x_i)}{\Gamma(\alpha+\beta+n)}}$$

$$= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\sum_{i=1}^{n} x_i)\ \Gamma(\beta+n-\sum_{i=1}^{n} x_i)} \theta^{\sum_{i=1}^{n} x_i+\alpha-1} (1-\theta)^{n-\sum_{i=1}^{n} x_i+\beta-1} \mathbb{I}_{]0,1[}^{(x)}$$

$$\therefore \quad \theta|\mathbf{x} \sim Beta\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i\right)$$

Finalmente,

$$p(x|\mathbf{x}) = \int_{\Theta} p(x|\tilde{\theta}) \; p(\tilde{\theta}|\mathbf{x}) \; d\tilde{\theta} =$$

$$= \int_{\Theta} \tilde{\theta}^{x} \; (1-\tilde{\theta})^{1-x} \; \mathbb{I}_{\{0,1\}}^{(x)} \; \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\sum_{i=1}^{n} x_i) \; \Gamma(\beta+n-\sum_{i=1}^{n} x_i)} \cdot$$

$$\cdot \; \tilde{\theta}^{\sum_{i=1}^{n} x_i+\alpha-1} \; (1-\tilde{\theta})^{n-\sum_{i=1}^{n} x_i+\beta-1} \; \mathbb{I}_{]0,1[}^{(x)} \; d\tilde{\theta}$$

$$= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\sum_{i=1}^{n} x_i) \; \Gamma(\beta+n-\sum_{i=1}^{n} x_i)} \cdot$$

$$\cdot \int_0^1 \tilde{\theta}^{(\alpha+\sum_{i=1}^{n} x_i+x)-1} \; (1-\tilde{\theta})^{(\beta+n-\sum_{i=1}^{n} x_i+x)-1} \; \mathbb{I}_{\{0,1\}}^{(x)} \; d\tilde{\theta}$$

$$= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\sum_{i=1}^{n} x_i) \; \Gamma(\beta+n-\sum_{i=1}^{n} x_i)} \cdot$$

$$\cdot \; \frac{\Gamma(\alpha+\sum_{i=1}^{n} x_i+x) \; \Gamma(\beta-\sum_{i=1}^{n} x_i - x + n + 1)}{\Gamma(\alpha+\beta+n+1)} \; \mathbb{I}_{\{0,1\}}^{(x)}$$

$$\therefore \quad x|\mathbf{x} \sim Binomial-Beta\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i, 1\right)$$

□

3.2. Modelo Poisson

Sea la muestra aleatoria $x_1, x_2, ..., x_n$ tal que $x_j \sim Poisson(\lambda)$ con λ desconocida $\Rightarrow f_X \in \mathcal{P} = \{p(x|\lambda) = e^{-\lambda}\ \frac{\lambda^x}{x!}\ \mathbb{I}_{\{0,1,...\}}^{(x)} : \lambda \in \Lambda =]0, \infty[\}$.

Sea la distribución inicial propuesta para λ tal que $\lambda \sim Gamma(\alpha, \beta)$

$\Rightarrow$

$$p(x) = \int_\Lambda p(x|\tilde{\lambda})\ p(\tilde{\lambda})\ d\tilde{\lambda}$$

$$= \int_0^\infty e^{-\tilde{\lambda}}\ \frac{\tilde{\lambda}^x}{x!}\ \frac{\beta^\alpha}{\Gamma(\alpha)}\ \tilde{\lambda}^{\alpha-1}\ e^{-\beta\tilde{\lambda}}\ \mathbb{I}_{\{0,1,...\}}^{(x)}\ d\tilde{\lambda}$$

$$= \frac{\beta^\alpha}{x!\ \Gamma(\alpha)} \int_0^\infty \tilde{\lambda}^{x+\alpha-1}\ e^{-\tilde{\lambda}(\beta+1)}\ \mathbb{I}_{\{0,1,...\}}^{(x)}\ d\tilde{\lambda}$$

Ahora, para resolver

$$\int_0^\infty \tilde{\lambda}^{x+\alpha-1}\ e^{-\tilde{\lambda}(\beta+1)}\ \mathbb{I}_{\{0,1,...\}}^{(x)}\ d\tilde{\lambda}$$

se tiene que

$$\int_0^\infty \frac{(\beta+1)^{x+\alpha}}{\Gamma(x+\alpha)}\ \tilde{\lambda}^{x+\alpha-1}\ e^{-\tilde{\lambda}(\beta+1)}\ d\tilde{\lambda} = 1$$

$$\Rightarrow \int_0^\infty \tilde{\lambda}^{x+\alpha-1}\ e^{-\tilde{\lambda}(\beta+1)}\ d\tilde{\lambda} = \frac{\Gamma(x+\alpha)}{(\beta+1)^{x+\alpha}}$$

$$\Rightarrow \quad p(x) = \frac{\beta^\alpha\ \Gamma(x+\alpha)}{x!\ \Gamma(\alpha)\ (\beta+1)^{x+\alpha}}\ \mathbb{I}_{\{0,1,...\}}^{(x)}$$

$$\therefore \quad x \sim Poisson - Gamma(\alpha, \beta, 1)$$

Por otra parte y considerando que las observaciones provienen de una muestra aleatoria, la función de verosimilitud está dada de la siguiente manera:

$$p(\mathbf{x}|\lambda) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} \mathbb{I}_{\{0,1,\ldots\}}^{(x_i)}$$

$$\Rightarrow p(\lambda|\mathbf{x}) = \frac{p(\mathbf{x}|\lambda)\ p(\lambda)}{\int_{\Lambda} p(\mathbf{x}|\tilde{\lambda})\ p(\tilde{\lambda})\ d\tilde{\lambda}}$$

$$= \frac{e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} \mathbb{I}_{\{0,1,\ldots\}}^{(x_i)} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1}\ e^{-\beta\lambda}\ \mathbb{I}_{\mathbb{R}^+}^{(\lambda)}}{\int_0^{\infty} e^{-n\tilde{\lambda}} \frac{\tilde{\lambda}^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} \mathbb{I}_{\{0,1,\ldots\}}^{(x_i)} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \tilde{\lambda}^{\alpha-1}\ e^{-\beta\tilde{\lambda}}\ d\tilde{\lambda}}$$

$$= \frac{\lambda^{\alpha+(\sum_{i=1}^{n} x_i)-1}\ e^{-\lambda(\beta+n)}\ \mathbb{I}_{\mathbb{R}^+}^{(\lambda)}}{\int_0^{\infty} \tilde{\lambda}^{\alpha+(\sum_{i=1}^{n} x_i)-1}\ e^{-\tilde{\lambda}(\beta+n)}\ d\tilde{\lambda}}$$

Para dar solución a la integral del denominador se tiene que

$$\int_0^{\infty} \frac{(\beta+n)^{\alpha+\sum_{i=1}^{n} x_i}}{\Gamma(\alpha+\sum_{i=1}^{n} x_i)} \tilde{\lambda}^{\alpha+(\sum_{i=1}^{n} x_i)-1}\ e^{-\tilde{\lambda}(\beta+n)}\ d\tilde{\lambda} = 1$$

$$\Rightarrow \int_0^{\infty} \tilde{\lambda}^{\alpha+(\sum_{i=1}^{n} x_i)-1}\ e^{-\tilde{\lambda}(\beta+n)}\ d\tilde{\lambda} = \frac{\Gamma(\alpha+\sum_{i=1}^{n} x_i)}{(\beta+n)^{\alpha+\sum_{i=1}^{n} x_i}}$$

$$\therefore\ p(\lambda|\mathbf{x}) = \frac{(\beta+n)^{\alpha+\sum_{i=1}^{n} x_i}}{\Gamma(\alpha+\sum_{i=1}^{n} x_i)} \lambda^{\alpha+(\sum_{i=1}^{n} x_i)-1}\ e^{-\lambda(\beta+n)}\ \mathbb{I}_{\mathbb{R}^+}^{(\lambda)}$$

$$\therefore \quad \lambda|\mathbf{x} \sim Gamma\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n\right)$$

Finalmente,

$$p(x|\mathbf{x}) \;=\; \int_{\Lambda} p(x|\tilde{\lambda}) \;\; p(\tilde{\lambda}|\mathbf{x}) \;\; d\tilde{\lambda} \;=$$

$$= \int_{\Lambda} e^{-\tilde{\lambda}} \;\; \frac{\tilde{\lambda}^{x}}{x!} \;\; \mathbb{I}_{\{0,1,\ldots\}}^{(x)} \;\; \frac{(\beta+n)^{\alpha+\sum_{i=1}^{n} x_i}}{\Gamma(\alpha+\sum_{i=1}^{n} x_i)} \;\; \tilde{\lambda}^{\alpha+(\sum_{i=1}^{n} x_i)-1} \;\; e^{-\tilde{\lambda}(\beta+n)} \;\; \mathbb{I}_{\mathbb{R}^+}^{(\lambda)} \;\; d\tilde{\lambda}$$

$$= \frac{(\beta+n)^{\alpha+\sum_{i=1}^{n} x_i}}{\Gamma(\alpha+\sum_{i=1}^{n} x_i)} \;\; \int_{0}^{\infty} \tilde{\lambda}^{x+\alpha+(\sum_{i=1}^{n} x_i)-1} \;\; e^{-\tilde{\lambda}(\beta+n+1)} \;\; \mathbb{I}_{\{0,1,\ldots\}}^{(x)} \;\; d\tilde{\lambda}$$

Para dar solución a la integral anterior se sabe que

$$\frac{(\beta+n+1)^{x+\alpha+\sum_{i=1}^{n} x_i}}{\Gamma(x+\alpha+\sum_{i=1}^{n} x_i)} \;\; \int_{0}^{\infty} \tilde{\lambda}^{x+\alpha+(\sum_{i=1}^{n} x_i)-1} \;\; e^{-\tilde{\lambda}(\beta+n+1)} \;\; d\tilde{\lambda} \;=\; 1$$

$$\Rightarrow \quad \int_{0}^{\infty} \tilde{\lambda}^{x+\alpha+(\sum_{i=1}^{n} x_i)-1} \;\; e^{-\tilde{\lambda}(\beta+n+1)} \;\; d\tilde{\lambda} \;=\; \frac{\Gamma(x+\alpha+\sum_{i=1}^{n} x_i)}{(\beta+n+1)^{x+\alpha+\sum_{i=1}^{n} x_i}}$$

$$\therefore \quad p(x|\mathbf{x}) \;=\; \frac{(\beta+n)^{\alpha+\sum_{i=1}^{n} x_i} \;\; \Gamma(x+\alpha+\sum_{i=1}^{n} x_i)}{\Gamma(\alpha+\sum_{i=1}^{n} x_i) \;\; (\beta+n+1)^{x+\alpha+\sum_{i=1}^{n} x_i}} \;\; \mathbb{I}_{\{0,1,\ldots\}}^{(x)}$$

$$\therefore \quad x|\mathbf{x} \sim Poisson-Gamma\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n, 1\right)$$

□

3.3. Modelo Binomial Negativa

Sea la muestra aleatoria $x_1, x_2, ..., x_n$ tal que $x_j \sim$ *Binomial Negativa*(θ, r) con θ desconocida $\Rightarrow f_X \in \mathcal{P} = \{p(x|\theta) = \binom{r+x-1}{r-1}\ \theta^r\ (1-\theta)^x\ \mathbb{I}_{\{0,1,...\}}^{(x)} : \theta \in \Theta =]0,1[\}$.

Sea la distribución inicial propuesta para θ tal que $\theta \sim Beta(\alpha, \beta)$

$\Rightarrow$

$$p(x) = \int_\Theta p(x|\tilde{\theta})\ p(\tilde{\theta})\ d\tilde{\theta}$$

$$= \int_\Theta \binom{r+x-1}{r-1}\ \tilde{\theta}^r\ (1-\theta)^x\ \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)}\ \tilde{\theta}^{\alpha-1}\ (1-\tilde{\theta})^{\beta-1}\ \mathbb{I}_{\{0,1,...\}}^{(x)}\ \mathbb{I}_{[0,1]}^{(\theta)}\ d\tilde{\theta}$$

$$= \binom{r+x-1}{r-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)} \int_0^1 \tilde{\theta}^{\alpha+r-1}\ (1-\theta)^{x+\beta-1}\ \mathbb{I}_{\{0,1,...\}}^{(x)}\ d\tilde{\theta}$$

como,

$$\int_0^1 \frac{\Gamma(\alpha+r+\beta+x)}{\Gamma(\alpha+r)\ \Gamma(x+\beta)}\ \tilde{\theta}^{\alpha+r-1}\ (1-\theta)^{x+\beta-1}\ d\tilde{\theta} = 1$$

$$\Rightarrow \int_0^1 \tilde{\theta}^{\alpha+r-1}\ (1-\theta)^{x+\beta-1}\ d\tilde{\theta} = \frac{\Gamma(\alpha+r)\ \Gamma(x+\beta)}{\Gamma(\alpha+r+\beta+x)}$$

$$\Rightarrow p(x) = \binom{r+x-1}{r-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)}\ \frac{\Gamma(\alpha+r)\ \Gamma(x+\beta)}{\Gamma(\alpha+r+\beta+x)}\ \mathbb{I}_{\{0,1,...\}}^{(x)}$$

$$\therefore \quad x \sim Binomial\ Negativa - Beta(\alpha, \beta, r)$$

Ahora,

$$
\begin{aligned}
p(\theta|\mathbf{x}) &= \frac{p(\mathbf{x}|\theta)\ p(\theta)}{\int_{\Theta} p(\mathbf{x}|\tilde{\theta})\ p(\tilde{\theta})\ d\tilde{\theta}} = \\
&= \frac{\prod_{i=1}^{n} \binom{r+x_i-1}{r-1}\ \theta^{nr}\ (1-\theta)^{\sum_{i=1}^{n} x_i}\ \prod_{i=1}^{n} \mathbb{I}_{\{0,1,\ldots\}}^{(x_i)}}{\prod_{i=1}^{n} \binom{r+x_i-1}{r-1}\ \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)}} \cdot \\
&\cdot \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)}\ \theta^{\alpha-1}\ (1-\theta)^{\beta-1}\ \mathbb{I}_{]0,1[}^{(\theta)}}{\int_0^1 \tilde{\theta}^{nr}\ (1-\tilde{\theta})^{\sum_{i=1}^{n} x_i}\ \tilde{\theta}^{\alpha-1}\ (1-\tilde{\theta})^{\beta-1}\ \prod_{i=1}^{n} \mathbb{I}_{\{0,1,\ldots\}}^{(x_i)}\ d\tilde{\theta}} \\
&= \frac{\theta^{\alpha+nr-1}\ (1-\theta)^{\sum_{i=1}^{n} x_i+\beta-1}\ \mathbb{I}_{]0,1[}^{(\theta)}}{\dfrac{\Gamma(\alpha+nr)\ \Gamma(\beta+\sum_{i=1}^{n} x_i)}{\Gamma(\alpha+nr+\beta+\sum_{i=1}^{n} x_i)}} \\
&= \frac{\Gamma(\alpha+nr+\beta+\sum_{i=1}^{n} x_i)}{\Gamma(\alpha+nr)\ \Gamma(\beta+\sum_{i=1}^{n} x_i)}\ \theta^{\alpha+nr-1}\ (1-\theta)^{\sum_{i=1}^{n} x_i+\beta-1}\ \mathbb{I}_{]0,1[}^{(\theta)}
\end{aligned}
$$

$$
\therefore\ \ \theta|\mathbf{x} \sim Beta\left(\alpha+nr, \beta+\sum_{i=1}^{n} x_i\right)
$$

Por último,

$$
p(x|\mathbf{x}) = \int_{\Theta} p(x|\tilde{\theta})\ p(\tilde{\theta}|\mathbf{x})\ d\tilde{\theta} =
$$

$$
= \int_{\Theta} \binom{r+x-1}{r-1}\ \tilde{\theta}^{r}\ (1-\tilde{\theta})^{x}\ \mathbb{I}_{\{0,1,\ldots\}}^{(x)}\ \frac{\Gamma(\alpha+nr+\beta+\sum_{i=1}^{n} x_i)}{\Gamma(\alpha+nr)\ \Gamma(\beta+\sum_{i=1}^{n} x_i)} \cdot
$$

$$\cdot\ \tilde{\theta}^{\alpha+nr-1}\ (1-\tilde{\theta})^{\sum_{i=1}^{n} x_i+\beta-1}\ \mathbb{I}_{]0,1[}^{(\theta)}\ d\tilde{\theta}$$

$$= \binom{r+x-1}{r-1} \frac{\Gamma(\alpha+nr+\beta+\sum_{i=1}^{n} x_i)}{\Gamma(\alpha+nr)\ \Gamma(\beta+\sum_{i=1}^{n} x_i)}\ .$$

$$\cdot \int_0^1 \tilde{\theta}^{\alpha+nr+r-1}\ (1-\tilde{\theta})^{\sum_{i=1}^{n} x_i+\beta+x-1}\ \mathbb{I}_{\{0,1,\ldots\}}^{(x)}\ d\tilde{\theta}$$

$$= \binom{r+x-1}{r-1} \frac{\Gamma(\alpha+nr+\beta+\sum_{i=1}^{n} x_i)}{\Gamma(\alpha+nr)\ \Gamma(\beta+\sum_{i=1}^{n} x_i)}\ .$$

$$\cdot \frac{\Gamma(\alpha+nr+r)\ \Gamma(\beta+\sum_{i=1}^{n} x_i+x)}{\Gamma(\alpha+nr+r+\beta+\sum_{i=1}^{n} x_i+x)}\ \mathbb{I}_{\{0,1,\ldots\}}^{(x)}$$

$$\therefore\ x|\mathrm{x} \sim Binomial\ Negativa-Beta\left(\alpha+nr,\beta+\sum_{i=1}^{n} x_i, r\right)$$

□

3.4. Modelo Exponencial

Sea la muestra aleatoria $x_1, x_2, ..., x_n$ tal que $x_j \sim Exponencial(\theta)$ con θ desconocida $\Rightarrow f_X \in \mathcal{P} = \{p(x|\theta) = \theta\ \ e^{-\theta x}\ \ \mathbb{I}_{\mathbb{R}^+}^{(x)} : \theta \in \Theta =]0, \infty[\}$.

Sea la distribución inicial propuesta para θ tal que $\theta \sim Gamma(\alpha, \beta)$

$\Rightarrow$

$$p(x) = \int_\Theta p(x|\tilde{\theta})\ \ p(\tilde{\theta})\ \ d\tilde{\theta}$$

$$= \int_0^\infty \tilde{\theta}\ e^{-\tilde{\theta}x}\ \frac{\beta^\alpha}{\Gamma(\alpha)}\ \tilde{\theta}^{\alpha-1}\ e^{-\beta\tilde{\theta}}\ \mathbb{I}_{]0,\infty[}^{(x)}\ d\tilde{\theta}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \tilde{\theta}^\alpha\ e^{-\tilde{\theta}(\beta+x)}\ \mathbb{I}_{]0,\infty[}^{(x)}\ d\tilde{\theta}$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)}\ \frac{\Gamma(\alpha+1)}{(\beta+x)^{\alpha+1}}\ \mathbb{I}_{]0,\infty[}^{(x)}$$

$$\therefore \quad x \sim Gamma-Gamma(\alpha,\beta,1)$$

Por otra parte,

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)\ p(\theta)}{\int_\Theta p(\mathbf{x}|\tilde{\theta})\ p(\tilde{\theta})\ d\tilde{\theta}} =$$

$$= \frac{\theta^n\ e^{-\theta\sum_{i=1}^n x_i}\ \frac{\beta^\alpha}{\Gamma(\alpha)}\ \theta^{\alpha-1}\ e^{-\beta x}\ \prod_{i=1}^n \mathbb{I}_{]0,\infty[}^{(x_i)}\ \mathbb{I}_{\mathbb{R}^+}^{(\theta)}}{\frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \tilde{\theta}^{\alpha+n-1}\ e^{-\tilde{\theta}(\beta+\sum_{i=1}^n x_i)}\ \prod_{i=1}^n \mathbb{I}_{]0,\infty[}^{(x_i)}\ d\tilde{\theta}}$$

$$= \frac{\theta^{\alpha+n-1}\ e^{-\theta(\beta+\sum_{i=1}^n x_i)}\ \mathbb{I}_{\mathbb{R}^+}^{(\theta)}}{\frac{\Gamma(\alpha+n)}{(\beta+\sum_{i=1}^n x_i)^{\alpha+n}}}$$

$$\frac{(\beta+\sum_{i=1}^n x_i)^{\alpha+n}}{\Gamma(\alpha+n)}\ \theta^{\alpha+n-1}\ e^{-\theta(\beta+\sum_{i=1}^n x_i)}\ \mathbb{I}_{\mathbb{R}^+}^{(\theta)}$$

$$\therefore \quad \theta|\mathbf{x} \sim Gamma\left(\alpha+n, \beta+\sum_{i=1}^n x_i\right)$$

Por último,

$$p(x|\mathbf{x}) = \int_{\Theta} p(x|\tilde{\theta})\ p(\tilde{\theta}|\mathbf{x})\ d\tilde{\theta} =$$

$$= \int_{\Theta} \tilde{\theta}\ e^{-\tilde{\theta}x}\ \mathbb{I}_{\mathbb{R}^+}^{(x)} \frac{(\beta + \sum_{i=1}^{n} x_i)^{\alpha+n}}{\Gamma(\alpha+n)}\ \tilde{\theta}^{\alpha+n-1}\ e^{-\tilde{\theta}(\beta+\sum_{i=1}^{n} x_i)}\ \mathbb{I}_{\mathbb{R}^+}^{(\theta)}\ d\tilde{\theta}$$

$$= \frac{(\beta + \sum_{i=1}^{n} x_i)^{\alpha+n}}{\Gamma(\alpha+n)} \int_0^{\infty} \tilde{\theta}^{\alpha+n}\ e^{-\tilde{\theta}(\beta+\sum_{i=1}^{n} x_i+x)}\ d\tilde{\theta}$$

$$= \frac{(\beta + \sum_{i=1}^{n} x_i)^{\alpha+n}}{\Gamma(\alpha+n)} \cdot \frac{\Gamma(\alpha+n+1)}{(\beta + \sum_{i=1}^{n} x_i)^{\alpha+n+1}}\ \mathbb{I}_{\mathbb{R}^+}^{(x)}$$

$$\therefore x|\mathbf{x} \sim Gamma - Gamma\left(\alpha+n, \beta + \sum_{i=1}^{n} x_i, 1\right)$$

□

3.5. Modelo Uniforme

Sea la muestra aleatoria $x_1, x_2, ..., x_n$ tal que $x_j \sim Uniforme(0, \theta)$ con θ desconocida $\Rightarrow f_X \in \mathcal{P} = \{p(x|\theta) = \frac{1}{\theta}\ \mathbb{I}_{[0,\theta]}^{(x)} : \theta \in \Theta =]0, \infty[\}$.

Sea la distribución inicial propuesta para θ tal que $\theta \sim Pareto(\alpha, \beta)$

$\Rightarrow$

$$p(x) = \int_{\Theta} p(x|\tilde{\theta})\ p(\tilde{\theta})\ d\tilde{\theta}$$

$$= \int_{\theta} \frac{1}{\tilde{\theta}}\ \mathbb{I}_{[0,\theta]}^{(x)}\ \alpha\beta^{\alpha}\ \tilde{\theta}^{-(\alpha+1)}\ \mathbb{I}_{[\beta,\infty[}^{(\theta)}\ d\tilde{\theta}$$

$$= \alpha\beta^{\alpha} \int_{\Theta} \tilde{\theta}^{-(\alpha+2)} \; \mathbb{I}_{[0,\theta]}^{(x)} \; \mathbb{I}_{[\beta,\infty[}^{(\theta)} \; d\tilde{\theta}$$

$$= \alpha\beta^{\alpha} \int_{max\{x,\beta\}}^{\infty} \tilde{\theta}^{-(\alpha+2)} \; d\tilde{\theta} \; = \; \alpha\beta^{\alpha} \; \frac{\theta^{-(\alpha+1)}}{-(\alpha+1)} \Bigg|_{max\{x,\beta\}}^{\infty}$$

$$= \frac{\alpha}{\alpha+1} \cdot \frac{1}{\beta} \; \mathbb{I}_{\{x \leq \beta\}} \; + \; \frac{\alpha\beta^{\alpha}}{\alpha+1} \cdot x^{-(\alpha+1)} \; \mathbb{I}_{\{x>\beta\}}$$

$$\therefore \quad x \sim \frac{\alpha}{\alpha+1} Uniforme(0,\beta) \;\; s.q. \;\; x \leq \beta \;\; ó \;\; x \sim \frac{1}{\alpha+1} Pareto(\alpha,\beta) \;\; s.q. \;\; x > \beta$$

Ahora bien, al considerar que las observaciones forman parte de una muestra aleatoria, la función de verosimilitud referente al modelo es la siguiente:

$$p(\mathbf{x}|\theta) \;=\; \frac{1}{\theta^{n}} \; \prod_{i=1}^{n} \mathbb{I}_{]0,\theta[}^{(x_i)}$$

Entonces,

$$p(\theta|\mathbf{x}) \;=\; \frac{p(\mathbf{x}|\theta) \;\; p(\theta)}{\int_{\Theta} p(\mathbf{x}|\tilde{\theta}) \;\; p(\tilde{\theta}) \;\; d\tilde{\theta}} \;=$$

$$= \frac{\frac{1}{\theta^{n}} \; \prod_{i=1}^{n} \mathbb{I}_{]0,\theta[}^{(x_i)} \;\; \alpha\beta^{\alpha} \;\; \theta^{-(\alpha+1)} \;\; \mathbb{I}_{[\beta,\infty[}^{(\theta)}}{\int_{max\{x_1,x_2,\ldots,x_n,\beta\}}^{\infty} \frac{1}{\tilde{\theta}^{n}} \;\; \alpha\beta^{\alpha} \;\; \tilde{\theta}^{-(\alpha+1)} \;\; d\tilde{\theta}}$$

$$= \frac{\alpha\beta^{\alpha} \;\; \theta^{-(\alpha+n+1)} \; \prod_{i=1}^{n} \mathbb{I}_{]0,\theta[}^{(x_i)} \;\; \mathbb{I}_{[\beta,\infty[}^{(\theta)}}{\alpha\beta^{\alpha} \int_{max\{x_1,x_2,\ldots,x_n,\beta\}}^{\infty} \tilde{\theta}^{-(\alpha+n+1)} \;\; d\tilde{\theta}}$$

$$= \frac{\theta^{-(\alpha+n+1)} \prod_{i=1}^{n} \mathbb{I}_{]0,\theta[}^{(x_i)} \ \mathbb{I}_{[\beta,\infty[}^{(\theta)}}{\left.\frac{-\tilde{\theta}^{-(\alpha+n)}}{\alpha+n}\right|_{max\{x_1,x_2,\ldots,x_n,\beta\}}^{\infty}}$$

$$= \ (\alpha+n) \ \beta_n^{\alpha+n} \ \theta^{-(\alpha+n+1)} \ \prod_{i=1}^{n} \mathbb{I}_{]0,\theta[}^{(x_i)} \ \mathbb{I}_{[\beta,\infty[}^{(\theta)}$$

donde $\beta_n = max\{x_1, x_2, ..., x_n, \beta\}$

$$\therefore \quad \theta|\mathbf{x} \sim Pareto(\alpha+n, \beta_n)$$

donde $\beta_n = max\{x_1, x_2, ..., x_n, \beta\}$

Finalmente,

$$p(x|\mathbf{x}) \ = \ \int_{\Theta} p(x|\tilde{\theta}) \ p(\tilde{\theta}|\mathbf{x}) \ d\tilde{\theta} \ =$$

$$\int_{\Theta} \frac{1}{\tilde{\theta}} \ (\alpha+n) \ \beta_n^{\alpha+n} \ \tilde{\theta}^{-(\alpha+n+1)} \ \prod_{i=1}^{n} \mathbb{I}_{]0,\theta[}^{(x_i)} \ \mathbb{I}_{[\beta,\infty[}^{(\theta)} \ \mathbb{I}_{]0,\theta[}^{(x)} \ d\tilde{\theta}$$

$$= \ (\alpha+n) \ \beta_n^{\alpha+n} \ \int_{max\{x_1,x_2,\ldots,x_n,\beta,x\}}^{\infty} \tilde{\theta}^{-(\alpha+n+2)} \ d\tilde{\theta}$$

$$= \ (\alpha+n) \ \beta_n^{\alpha+n} \ . \ \left.\frac{\theta^{-(\alpha+n+1)}}{-(\alpha+n+1)}\right|_{max\{x_1,x_2,\ldots,x_n,\beta,x\}}^{\infty}$$

$$= \ \frac{\alpha+n}{\alpha+n+1} \ . \ \frac{1}{\beta_n} \ \mathbb{I}_{\{x \le \beta_n\}} \ + \ \frac{(\alpha+n) \ \beta_n^{\alpha+n}}{\alpha+n+1} \ . \ x^{-(\alpha+n+1)} \ \mathbb{I}_{\{x>\beta\}}$$

$$\therefore \quad x|\mathbf{x} \sim \frac{\alpha+n}{\alpha+n+1} \ Uniforme(0, \beta_n) \ \ s.q. \ \ x \le \beta$$

ó

$$x|\mathbf{x} \sim \frac{1}{\alpha+n+1}\ Pareto(\alpha+n, \beta_n)\ \ s.q.\ \ x > \beta$$

$\square$

3.6. Modelo Normal (con media μ conocida)

Sea la muestra aleatoria $x_1, x_2, ..., x_n$ tal que $x_j \sim Normal(\mu, \theta)$ con μ conocida y θ desconocida $\Rightarrow f_X \in \mathcal{P} = \{p(x|\mu,\theta) = \sqrt{\frac{\theta}{2\pi}}\ e^{-\frac{1}{2}\theta(x-\mu)^2}\ \mathbb{I}_{]-\infty,\infty[}^{(x)} : \theta \in \Theta =]0, \infty[\}$.

Sea la distribución inicial propuesta para θ tal que $\theta \sim Gamma(\alpha, \beta)$

$\Rightarrow$

$$p(x) = \int_{\Theta} p(x|\tilde{\theta})\ p(\tilde{\theta})\ d\tilde{\theta}$$

$$= \int_0^{\infty} \sqrt{\frac{\tilde{\theta}}{2\pi}}\ e^{-\frac{1}{2}\tilde{\theta}(x-\mu)^2}\ \frac{\beta^{\alpha}}{\Gamma(\alpha)}\ \tilde{\theta}^{\alpha-1}\ e^{-\beta\tilde{\theta}}\ \mathbb{I}_{]-\infty,\infty[}^{(x)}\ d\tilde{\theta}$$

$$= \sqrt{\frac{1}{2\pi}} \cdot \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} e^{-\tilde{\theta}(\frac{1}{2}(x-\mu)^2+\beta)}\ \tilde{\theta}^{\alpha+\frac{1}{2}-1}\ \mathbb{I}_{]-\infty,\infty[}^{(x)}\ d\tilde{\theta}$$

Para dar solución a la integral anterior, se tiene que

$$\int_0^{\infty} \frac{(\frac{1}{2}(x-\mu)^2+\beta)^{\alpha+\frac{1}{2}}}{\Gamma(\alpha+\frac{1}{2})}\ .\ e^{-\tilde{\theta}(\frac{1}{2}(x-\mu)^2+\beta)}\ \tilde{\theta}^{\alpha+\frac{1}{2}-1}\ d\tilde{\theta} = 1$$

$$\Rightarrow \int_0^{\infty} e^{-\tilde{\theta}(\frac{1}{2}(x-\mu)^2-\beta)}\ \tilde{\theta}^{\alpha+\frac{1}{2}-1}\ d\tilde{\theta} = \frac{\Gamma(\alpha+\frac{1}{2})}{(\frac{1}{2}(x-\mu)^2+\beta)^{\alpha+\frac{1}{2}}}$$

Entonces,

$$p(x) = \sqrt{\frac{1}{2\pi}} \cdot \frac{\beta^{\alpha}}{\Gamma(\alpha)}\ .\ \Gamma(\alpha+\frac{1}{2})\ .\ (\frac{1}{2}(x-\mu)^2+\beta)^{-(\alpha+\frac{1}{2})}\ \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$= \sqrt{\frac{1}{2\beta\pi}} \cdot \frac{\Gamma(\alpha+\frac{1}{2})}{\Gamma(\alpha)} \cdot \left(\frac{1}{2\beta}(x-\mu)^2+1\right)^{-(\alpha+\frac{1}{2})} \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$\therefore \quad x \sim T-Student\left(\mu, \frac{\alpha}{\beta}, 2\alpha\right)$$

Por otra parte, se sabe que la función de verosimilitud referente al modelo normal con μ conocida está dada por

$$p(\mathbf{x}|\theta) = \left(\sqrt{\frac{\theta}{2\pi}}\right)^n e^{-\frac{1}{2}\theta\sum_{i=1}^{n}(x_i-\mu)^2} \prod_{i=1}^{n} \mathbb{I}_{]-\infty,\infty[}^{(x_i)}$$

$$\Rightarrow p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)\ p(\theta)}{\int_\Theta p(\mathbf{x}|\tilde{\theta})\ p(\tilde{\theta})\ d\tilde{\theta}} =$$

$$\frac{\left(\sqrt{\frac{\theta}{2\pi}}\right)^n e^{-\frac{1}{2}\theta\sum_{i=1}^{n}(x_i-\mu)^2} \prod_{i=1}^{n} \mathbb{I}_{]-\infty,\infty[}^{(x_i)} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \mathbb{I}_{\mathbb{R}^+}^{(\theta)}}{\int_0^\infty \left(\sqrt{\frac{\tilde{\theta}}{2\pi}}\right)^n e^{-\frac{1}{2}\tilde{\theta}\sum_{i=1}^{n}(x_i-\mu)^2} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \tilde{\theta}^{\alpha-1} e^{-\beta\tilde{\theta}} \prod_{i=1}^{n} \mathbb{I}_{]-\infty,\infty[}^{(x_i)}\ d\tilde{\theta}}$$

$$= \frac{e^{-\theta(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\beta)} \theta^{\alpha+\frac{n}{2}-1} \mathbb{I}_{\mathbb{R}^+}^{(\theta)}}{\int_0^\infty e^{-\tilde{\theta}(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\beta)} \tilde{\theta}^{\alpha+\frac{n}{2}-1}\ d\tilde{\theta}}$$

Resolviendo la integral del denominador se tiene que

$$p(\theta|\mathbf{x}) = \frac{(\beta+\frac{1}{2}(x_i-\mu)^2)^{\alpha+\frac{n}{2}}}{\Gamma(\alpha+\frac{n}{2})} e^{-\theta(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\beta)} \theta^{\alpha+\frac{n}{2}-1} \mathbb{I}_{\mathbb{R}^+}^{(\theta)}$$

$$\therefore \quad \theta|\mathbf{x} \sim Gamma\left(\alpha+\frac{1}{2}n, \beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

Ahora,

$$p(x|\mathbf{x}) = \int_{\Theta} p(x|\tilde{\theta})\ p(\tilde{\theta}|\mathbf{x})\ d\tilde{\theta} =$$

$$= \int_0^{\infty} \sqrt{\frac{\tilde{\theta}}{2\pi}}\ e^{-\frac{1}{2}\tilde{\theta}(x-\mu)^2}\ \frac{(\beta + \frac{1}{2}(x_i - \mu)^2)^{\alpha+\frac{n}{2}}}{\Gamma(\alpha + \frac{n}{2})} \ .$$

$$.\ e^{-\tilde{\theta}(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\beta)}\ \tilde{\theta}^{\alpha+\frac{n}{2}-1}\ \mathbb{I}^{(x)}_{]-\infty,\infty[}\ d\tilde{\theta}$$

$$= \sqrt{\frac{1}{2\pi}}\ .\ \frac{\left(\beta + \frac{1}{2}(x_i - \mu)^2\right)^{\alpha+\frac{n}{2}}}{\Gamma(\alpha + \frac{n}{2})} \ .$$

$$.\ \int_0^{\infty} e^{-\tilde{\theta}(\frac{1}{2}(x-\mu)^2+(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\beta)}\ \tilde{\theta}^{\alpha+\frac{n}{2}+\frac{1}{2}-1}\ \mathbb{I}^{(x)}_{]-\infty,\infty[}\ d\tilde{\theta}$$

Pero

$$\int_0^{\infty} e^{-\tilde{\theta}(\frac{1}{2}(x-\mu)^2+(\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\beta)}\ \tilde{\theta}^{\alpha+\frac{n}{2}+\frac{1}{2}-1}\ d\tilde{\theta} =$$

$$= \frac{\Gamma(\alpha + \frac{n}{2} + \frac{1}{2})}{\left(\frac{1}{2}(x-\mu)^2 + (\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 + \beta\right)^{\alpha+\frac{n}{2}+\frac{1}{2}}}$$

$$\Rightarrow\ p(x|\mathbf{x}) =$$

$$= \sqrt{\frac{1}{2\pi}}\ .\ \frac{\left(\beta + \frac{1}{2}(x_i - \mu)^2\right)^{\alpha+\frac{n}{2}}}{\Gamma(\alpha + \frac{n}{2})} \ .$$

$$.\ \frac{\Gamma(\alpha + \frac{n}{2} + \frac{1}{2})}{\left(\frac{1}{2}(x-\mu)^2 + (\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 + \beta\right)^{\alpha+\frac{n}{2}+\frac{1}{2}}}\ \mathbb{I}^{(x)}_{]-\infty,\infty[}$$

$$= \frac{\Gamma(\alpha + \frac{n}{2} + \frac{1}{2})}{\Gamma(\alpha + \frac{n}{2})}\ .\ \sqrt{\frac{\frac{1}{\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\beta}}{2\pi}} \ .$$

$$\cdot \left(1 + \frac{(x-\mu)^2}{\sum_{i=1}^{n}(x_i-\mu)^2 + 2\beta}\right)^{-(\alpha+\frac{n}{2}+\frac{1}{2})} \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$\therefore \quad x|\mathbf{x} \sim T-Student\left(\mu, \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2, 2\alpha+n\right)$$

□

3.7. Modelo Normal (con precisión λ conocida)

Sea la muestra aleatoria $x_1, x_2, ..., x_n$ tal que $x_j \sim Normal(\theta, \lambda)$ con λ conocida y θ desconocida $\Rightarrow f_X \in \mathcal{P} = \{p(x|\theta,\lambda) = \sqrt{\frac{\lambda}{2\pi}}\ e^{-\frac{1}{2}\lambda(x-\theta)^2}\ \mathbb{I}_{]-\infty,\infty[}^{(x)} : \theta \in \Theta =]0,\infty[\}$.

Sea la distribución inicial propuesta para θ tal que $\theta \sim Normal(\mu_0, \lambda_0)$

$\Rightarrow$

$$p(x) = \int_{\Theta} p(x|\tilde{\theta})\ p(\tilde{\theta})\ d\tilde{\theta}$$

$$= \int_{-\infty}^{\infty} \sqrt{\frac{\lambda}{2\pi}}\ e^{-\frac{1}{2}\lambda(x-\tilde{\theta})^2}\ \mathbb{I}_{]-\infty,\infty[}^{(x)} \cdot \sqrt{\frac{\lambda_0}{2\pi}}\ e^{-\frac{1}{2}\lambda_0(\tilde{\theta}-\mu_0)^2}\ d\tilde{\theta}$$

$$= \int_{-\infty}^{\infty} \frac{\sqrt{\lambda\lambda_0}}{2\pi}\ e^{-\frac{1}{2}\lambda(x-\tilde{\theta})^2-\frac{1}{2}\lambda_0(\tilde{\theta}-\mu_0)^2}\ \mathbb{I}_{]-\infty,\infty[}^{(x)}\ d\tilde{\theta}$$

$$= \int_{-\infty}^{\infty} \frac{\sqrt{\lambda\lambda_0}}{2\pi}\ e^{-\frac{1}{2}\lambda(x^2-2x\tilde{\theta}+\tilde{\theta}^2)-\frac{1}{2}\lambda_0(\tilde{\theta}^2-2\tilde{\theta}\mu_0+\mu_0^2)}\ \mathbb{I}_{]-\infty,\infty[}^{(x)}\ d\tilde{\theta}$$

$$= \frac{\sqrt{\lambda\lambda_0}}{2\pi} \cdot e^{-\frac{1}{2}(\lambda x^2+\lambda_0\mu_0^2)} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((\lambda+\lambda_0)\tilde{\theta}^2+2(\lambda x+\lambda_0\mu_0)\tilde{\theta}\right)}\ \mathbb{I}_{]-\infty,\infty[}^{(x)}\ d\tilde{\theta}$$

Para dar solución a la integral anterior es preciso notar que

$$\int_{-\infty}^{\infty} \sqrt{\frac{\lambda+\lambda_0}{2\pi}} \ . \ e^{-\frac{\lambda+\lambda_0}{2}\left(\tilde{\theta}-\left(\frac{\lambda x+\lambda_0\mu_0}{\lambda+\lambda_0}\right)^2\right)} \quad d\tilde{\theta} \ =$$

$$\int_{-\infty}^{\infty} \sqrt{\frac{\lambda+\lambda_0}{2\pi}} \ . \ e^{-\frac{\lambda+\lambda_0}{2}\left(\tilde{\theta}^2-2\tilde{\theta}\left(\frac{\lambda x+\lambda_0\mu_0}{\lambda+\lambda_0}\right)+\left(\frac{\lambda x+\lambda_0\mu_0}{\lambda+\lambda_0}\right)^2\right)} \quad d\tilde{\theta} \ = \ 1$$

$$\Rightarrow \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((\lambda+\lambda_0)\tilde{\theta}^2+2(\lambda x+\lambda_0\mu_0)\tilde{\theta}\right)} \ d\tilde{\theta} \ =$$

$$= \sqrt{\frac{2\pi}{\lambda+\lambda_0}} \ . \ e^{\frac{1}{2}\left(\frac{(\lambda x+\lambda_0\mu_0)^2}{\lambda+\lambda_0}\right)}$$

$$\Rightarrow \ p(x) \ = \ \frac{\sqrt{\lambda\lambda_0}}{2\pi} \ . \ \frac{\sqrt{2\pi}}{\sqrt{\lambda+\lambda_0}} \ e^{-\frac{1}{2}\left(\lambda x^2+\lambda_0\mu_0^2-\frac{(\lambda x+\lambda_0\mu_0)^2}{\lambda+\lambda_0}\right)} \ \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$= \sqrt{\frac{\frac{\lambda\lambda_0}{\lambda+\lambda_0}}{2\pi}} \ e^{-\frac{1}{2}\cdot\frac{\lambda\lambda_0}{\lambda+\lambda_0}\left(\frac{(\lambda+\lambda_0)x^2}{\lambda_0}+\frac{(\lambda+\lambda_0)\mu_0^2}{\lambda}-\frac{(\lambda x+\lambda_0\mu_0)^2}{\lambda\lambda_0}\right)} \ \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$= \sqrt{\frac{\frac{\lambda\lambda_0}{\lambda+\lambda_0}}{2\pi}} \ e^{-\frac{1}{2}\cdot\frac{\lambda\lambda_0}{\lambda+\lambda_0}\left(\frac{\lambda(\lambda+\lambda_0)x^2+\lambda_0(\lambda+\lambda_0)\mu_0^2 \quad (\lambda x+\lambda_0\mu_0)^2}{\lambda\lambda_0}\right)} \ \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$= \sqrt{\frac{\frac{\lambda\lambda_0}{\lambda+\lambda_0}}{2\pi}} \ e^{-\frac{1}{2}\cdot\frac{\lambda\lambda_0}{\lambda+\lambda_0}\left(\frac{\lambda^2x^2+\lambda\lambda_0x^2+\lambda_0\lambda\mu_0^2+\lambda_0^2\mu_0^2 \quad \lambda^2x^2 \quad 2\lambda x\lambda_0\mu_0 \quad \lambda_0^2\mu_0^2}{\lambda\lambda_0}\right)} \ \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$= \sqrt{\frac{\frac{\lambda\lambda_0}{\lambda+\lambda_0}}{2\pi}} \ e^{-\frac{1}{2}\cdot\frac{\lambda\lambda_0}{\lambda+\lambda_0}\left(\frac{\lambda\lambda_0}{\lambda\lambda_0}(x^2-x\mu_0+\mu_0^2)\right)} \ \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$= \sqrt{\frac{\frac{\lambda\lambda_0}{\lambda+\lambda_0}}{2\pi}} \ e^{-\frac{1}{2}\cdot\frac{\lambda\lambda_0}{\lambda+\lambda_0}(x-\mu_0)^2} \ \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$\therefore \ x \sim Normal\left(\mu_0, \frac{\lambda\lambda_0}{\lambda+\lambda_0}\right)$$

Por otra parte y considerando que las observaciones provienen de una muestra aleatoria, la función de verosimilitud está dada de la siguiente manera:

$$\left(\sqrt{\frac{\lambda}{2\pi}}\right)^n e^{-\frac{\lambda}{2}\sum_{i=1}^n (x_i-\theta)^2} \prod_{i=1}^n \mathbb{I}_{]-\infty,\infty[}^{(x_i)}$$

$$\Rightarrow \; p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)\; p(\theta)}{\int_\Theta p(\mathbf{x}|\tilde\theta)\; p(\tilde\theta)\; d\tilde\theta} =$$

$$= \frac{\left(\sqrt{\frac{\lambda}{2\pi}}\right)^n e^{-\frac{\lambda}{2}\sum_{i=1}^n (x_i-\theta)^2} \prod_{i=1}^n \mathbb{I}_{]-\infty,\infty[}^{(x_i)} \sqrt{\frac{\lambda_0}{2\pi}}\; e^{-\frac{\lambda_0}{2}(\theta-\mu_0)^2}\; \mathbb{I}_{]-\infty,\infty[}^{(\theta)}}{\left(\sqrt{\frac{\lambda}{2\pi}}\right)^n \sqrt{\frac{\lambda_0}{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{n\lambda}{2}\sum_{i=1}^n (x_i-\tilde\theta)^2}\; e^{-\frac{\lambda_0}{2}(\tilde\theta-\mu_0)^2} \prod_{i=1}^n \mathbb{I}_{]-\infty,\infty[}^{(x_i)}\; d\tilde\theta}$$

$$= \frac{e^{-\frac{1}{2}\left(\lambda(\sum_{i=1}^n x_i - 2\sum_{i=1}^n x_i^2\theta + n\theta^2) + \lambda_0(\theta^2 - 2\mu_0\theta + \mu_0^2)\right)}\; \mathbb{I}_{]-\infty,\infty[}^{(\theta)}}{e^{-\frac{1}{2}(\lambda\sum_{i=1}^n x_i^2 + \lambda_0\mu_0^2)} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((n\lambda+\lambda_0)\tilde\theta^2 - 2(\lambda\sum_{i=1}^n x_i + \lambda_0\mu_0)\tilde\theta\right)}\; d\tilde\theta}$$

$$= \frac{e^{-\frac{1}{2}\left((n\lambda+\lambda_0)\theta^2 - 2(\lambda\sum_{i=1}^n x_i + \lambda_0\mu_0)\theta\right)} \;.\; e^{-\frac{1}{2}\left(\lambda\sum_{i=1}^n x_i^2 + \lambda_0\mu_0^2\right)}\; \mathbb{I}_{]-\infty,\infty[}^{(\theta)}}{e^{-\frac{1}{2}(\lambda\sum_{i=1}^n x_i^2 + \lambda_0\mu_0^2)} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((n\lambda+\lambda_0)\tilde\theta^2 - 2(\lambda\sum_{i=1}^n x_i + \lambda_0\mu_0)\tilde\theta\right)}\; d\tilde\theta}$$

$$= \frac{e^{-\frac{n\lambda+\lambda_0}{2}\left(\theta - \frac{\lambda\sum_{i=1}^n x_i + \lambda_0\mu_0}{n\lambda+\lambda_0}\right)^2} \;.\; e^{-\frac{n\lambda+\lambda_0}{2}\left(-\left(\frac{\lambda\sum_{i=1}^n x_i + \lambda_0\mu_0}{n\lambda+\lambda_0}\right)^2\right)}\; \mathbb{I}_{]-\infty,\infty[}^{(\theta)}}{\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((n\lambda+\lambda_0)\tilde\theta^2 - 2(\lambda\sum_{i=1}^n x_i + \lambda_0\mu_0)\tilde\theta\right)}\; d\tilde\theta}$$

Para dar solución a la integral ubicada en el denominador se sabe que

$$\int_{-\infty}^{\infty} \sqrt{\frac{n\lambda+\lambda_0}{2\pi}}\; e^{-\frac{(n\lambda+\lambda_0)}{2}\left(\tilde\theta^2 - \frac{2(\lambda\sum_{i=1}^n x_i + \lambda_0\mu_0)\tilde\theta}{n\lambda+\lambda_0} + \left(\frac{\lambda\sum_{i=1}^n x_i + \lambda_0\mu_0}{n\lambda+\lambda_0}\right)^2\right)}\; d\tilde\theta = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((n\lambda+\lambda_0)\tilde{\theta}^2-2(\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0)\tilde{\theta}\right)}\ d\tilde{\theta} =$$

$$= \int_{-\infty}^{\infty} e^{-\frac{(n\lambda+\lambda_0)}{2}\left(\tilde{\theta}^2-\frac{2(\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0)\tilde{\theta}}{(n\lambda+\lambda_0)}\right)}\ d\tilde{\theta} =$$

$$\sqrt{\frac{2\pi}{n\lambda+\lambda_0}}\ e^{\frac{(n\lambda+\lambda_0)}{2}\left(\frac{\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0}{n\lambda+\lambda_0}\right)^2}$$

$$\therefore\ p(\theta|\mathbf{x}) = \frac{e^{-\frac{n\lambda+\lambda_0}{2}\left(\theta-\frac{\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0}{n\lambda+\lambda_0}\right)^2}\ .\ e^{-\frac{n\lambda+\lambda_0}{2}\left(-\left(\frac{\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0}{n\lambda+\lambda_0}\right)^2\right)}\ \mathbb{I}_{]-\infty,\infty[}^{(\theta)}}{\sqrt{\frac{2\pi}{n\lambda+\lambda_0}}\ e^{\frac{(n\lambda+\lambda_0)}{2}\left(\frac{\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0}{n\lambda+\lambda_0}\right)^2}}$$

$$= \int_{-\infty}^{\infty}\sqrt{\frac{n\lambda+\lambda_0}{2\pi}}\ e^{-\frac{(n\lambda+\lambda_0)}{2}\left(\theta-\frac{(\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0)}{n\lambda+\lambda_0}\right)^2}\ \mathbb{I}_{]-\infty,\infty[}^{(\theta)}$$

$$\therefore\quad \theta|\mathbf{x} \sim Normal\left(\frac{(\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0)}{n\lambda+\lambda_0}, n\lambda+\lambda_0\right)$$

Por último,

$$p(x|\mathbf{x}) = \int_{\Theta} p(x|\tilde{\theta})\ p(\tilde{\theta}|\mathbf{x})\ d\tilde{\theta} =$$

$$= \int_{\Theta}\sqrt{\frac{\lambda}{2\pi}}\ e^{-\frac{\lambda}{2}(x-\tilde{\theta})^2}\ \mathbb{I}_{]-\infty,\infty[}^{(x)}\ .$$

$$.\ \sqrt{\frac{n\lambda+\lambda_0}{2\pi}}\ e^{-\frac{(n\lambda+\lambda_0)}{2}\left(\tilde{\theta}-\frac{(\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0)}{n\lambda+\lambda_0}\right)^2}\ \mathbb{I}_{]-\infty,\infty[}^{(\theta)}\ d\tilde{\theta}$$

$$= \frac{\sqrt{\lambda(n\lambda+\lambda_0)}}{2\pi}\ .$$

$$\cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\lambda(x^2-2x\tilde{\theta}+\tilde{\theta}^2)+(n\lambda+\lambda_0)\left(\tilde{\theta}^2-2\frac{\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0}{n\lambda+\lambda_0}\tilde{\theta}+\frac{(\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0)^2}{(n\lambda+\lambda_0)^2}\right)\right)} \mathbb{I}_{]-\infty,\infty[}^{(x)}\, d\tilde{\theta}$$

$$= \frac{\sqrt{\lambda(n\lambda+\lambda_0)}}{2\pi} \ .$$

$$\cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((\lambda+n\lambda+\lambda_0)\tilde{\theta}^2-2(\lambda x+\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0)\tilde{\theta}+\lambda x^2+\frac{(\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0)^2}{n\lambda+\lambda_0}\right)} \mathbb{I}_{]-\infty,\infty[}^{(x)}\, d\tilde{\theta}$$

$$= \frac{\sqrt{\lambda(n\lambda+\lambda_0)}}{2\pi}\ e^{-\frac{1}{2}\left(\lambda x^2+\frac{(\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0)^2}{n\lambda+\lambda_0}\right)} \ .$$

$$\cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((\lambda+n\lambda+\lambda_0)\tilde{\theta}^2-2(\lambda x+\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0)\tilde{\theta}\right)} \mathbb{I}_{]-\infty,\infty[}^{(x)}\, d\tilde{\theta}$$

Para dar solución a la integral anterior se sabe que

$$\int_{-\infty}^{\infty} \sqrt{\frac{\lambda+n\lambda+\lambda_0}{2\pi}} \ . \ e^{-\frac{\lambda+n\lambda+\lambda_0}{2}\left(\theta^2-2\frac{(\lambda x+\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0)\theta}{\lambda+n\lambda+\lambda_0}+\left(\frac{\lambda x+\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0}{\lambda+n\lambda+\lambda_0}\right)^2\right)}\ d\theta = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left((\lambda+n\lambda+\lambda_0)\tilde{\theta}^2-2(\lambda x+\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0)\tilde{\theta}\right)}\ d\tilde{\theta} =$$

$$= \sqrt{\frac{2\pi}{\lambda+n\lambda+\lambda_0}} \ . \ e^{\frac{1}{2}\left(\frac{(\lambda x+\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0)^2}{\lambda+n\lambda+\lambda_0}\right)}$$

$$\Rightarrow p(x|\mathbf{x}) = \frac{\sqrt{\lambda(n\lambda+\lambda_0)}}{2\pi}\ e^{-\frac{1}{2}\left(\lambda x^2+\frac{(\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0)^2}{n\lambda+\lambda_0}\right)} \ .$$

$$\cdot \sqrt{\frac{2\pi}{\lambda+n\lambda+\lambda_0}} \ . \ e^{\frac{1}{2}\left(\frac{(\lambda x+\lambda\sum_{i=1}^{n} x_i+\lambda_0\mu_0)^2}{\lambda+n\lambda+\lambda_0}\right)} \mathbb{I}_{]-\infty,\infty[}^{(x)} =$$

$$= \sqrt{\frac{\frac{\lambda(n\lambda+\lambda_0)}{\lambda+n\lambda+\lambda_0}}{2\pi}} \; .$$

$$. \; e^{-\frac{\lambda(n\lambda+\lambda_0)}{2(\lambda+n\lambda+\lambda_0)}\left(\frac{(\lambda+n\lambda+\lambda_0)x^2}{n\lambda+\lambda_0}+\frac{(\lambda+n\lambda+\lambda_0)(\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0)^2}{\lambda(n\lambda+\lambda_0)^2}-\frac{(\lambda x+\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0)^2}{\lambda(n\lambda+\lambda_0)}\right)} \; \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$\text{Haciendo } \mu_n = \frac{\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0}{n\lambda+\lambda_0} \text{ y } \lambda_n = n\lambda+\lambda_0$$

$$\Rightarrow \; p(x|\mathrm{x}) = \sqrt{\frac{\frac{\lambda\lambda_n}{\lambda+\lambda_n}}{2\pi}} \; e^{-\frac{\lambda\lambda_n}{2(\lambda+\lambda_n)}\left(\frac{(\lambda+\lambda_n)x^2}{\lambda_n}+\frac{(\lambda+\lambda_n)\mu_n^2}{\lambda}-\frac{(\lambda x+\mu_n\lambda_n)^2}{\lambda\lambda_n}\right)} \; \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$= \sqrt{\frac{\frac{\lambda\lambda_n}{\lambda+\lambda_n}}{2\pi}} \; e^{-\frac{\lambda\lambda_n}{2(\lambda+\lambda_n)}\left(\frac{\lambda(\lambda+\lambda_n)x^2+\lambda_n(\lambda+\lambda_n)\mu_n^2-(\lambda x+\mu_n\lambda_n)^2}{\lambda\lambda_n}\right)} \; \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$= \sqrt{\frac{\frac{\lambda\lambda_n}{\lambda+\lambda_n}}{2\pi}} \; e^{-\frac{\lambda\lambda_n}{2(\lambda+\lambda_n)}\left(\frac{\lambda^2x^2+\lambda\lambda_n x^2+\lambda_n\lambda\mu_n^2+\lambda_n^2\mu_n^2-\lambda^2x^2-2\lambda x\mu_n\lambda_n-\mu_n^2\lambda_n^2}{\lambda\lambda_n}\right)} \; \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$= \sqrt{\frac{\frac{\lambda\lambda_n}{\lambda+\lambda_n}}{2\pi}} \; e^{-\frac{\lambda\lambda_n}{2(\lambda+\lambda_n)}\left(\frac{\lambda\lambda_n}{\lambda\lambda_n}(x^2-2x\mu_n+\mu_n^2)\right)} \; \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$= \sqrt{\frac{\frac{\lambda\lambda_n}{\lambda+\lambda_n}}{2\pi}} \; e^{-\frac{\lambda\lambda_n}{2(\lambda+\lambda_n)}(x-\mu_n)^2} \; \mathbb{I}_{]-\infty,\infty[}^{(x)}$$

$$\therefore \; x|\mathrm{x} \sim Normal\left(\mu_n, \frac{\lambda\lambda_n}{\lambda+\lambda_n}\right)$$

$$\text{donde } \mu_n = \frac{\lambda\sum_{i=1}^{n}x_i+\lambda_0\mu_0}{n\lambda+\lambda_0} \text{ y } \lambda_n = n\lambda+\lambda_0$$

□

Conclusiones

En problemas reales, es de suma importancia generar, o bien, tener a la mano el modelo que mejor se ajusta al comportamiento del evento observado. Sin embargo, considero que este es el paso que más requiere del discernimiento y buen juicio del analista, así como de su experiencia y conocimiento, tanto del fenómeno en cuestión como de las herramientas matemáticas, ya que pasar de la realidad a la abstracción puede ser verdaderamente difícil. No obstante, de acuerdo al presente trabajo, se sabe ya que para abstraer un evento real en distribuciones de probabilidad y así generar el modelo estadístico bayesiano; solo se necesitan establecer las funciones de probabilidad $p(x|\theta)$ y $p(\theta)$, con información preliminar ya recabada que de cierta manera conceptualiza el modelo.

Toda esta situación, aunada al desarrollo tecnológico y computacional de diferentes tipos de software que facilitan cada vez más el manejo de la complejidad de cálculos matemáticos, brindan alternativas oportunas para dar solución a problemas reales; lo cual se considera la principal razón de pensar, a buen juicio, que la estadística bayesiana es una teoría en vanguardia que persigue satisfacer hasta el control de variables mas ambicioso que se pueda presentar, siempre y cuando exista información disponible; y, por otra parte, es preciso concluir que por medio de los modelos ya desarrollados se puede iniciar un buen proyecto, o bien, se pueden adquirir los métodos suficientes para innovar cualquier modelo que se requiera adaptar en una situación real.

TESIS CON
FALLA DE ORIGEN

Bibliografía

* Lindley, D.V., *Bayesian Statistics, a Review*, University College London, impreso para The Society for Industrial and Applied Mathematics por J. V. Arrowsmith Ltd., Bristol 3, Inglaterra, 1972.

* Bernardo, José M. & Smith, Adrian F. M., *Bayesian Theory*, Ed. John Wiley & Sons, Chichester, Inglaterra, 1994.

* Gelman, A., Crlin, J. B., Stern, H. S. & Rubin, D. B., *Bayesian Data Analysis*, Ed. Chapman & Hall, London, 1995.

* Lindley, Dennis V., *The philosophy of statistics*, The Statistician, Vol. 49 págs. 293-337, Minehead, UK, 2000.

* Pino Manresa, Guido del, *Estadística. Teoría y Métodos*, Ediciones Universidad Católica de Chile, Colección de Textos Univarsitarios. Facultad de Matemáticas, 1ra. ed., 1995.

* Bergel, J. O., *Statistical Decision Theory and Bayesian Analysis*, Ed. Springer Verlag, New York, 1985.

* Lindley, D. V., *An intorduction to Probability and Statistics from a Bayesian Viewpoint*, Vol. 2 Inference, Ed. Cambridge University Press, Cambridge, 1965.

* Gutiérrez Peña, Eduardo, *Análisis Bayesiano de Modelos Jerárquicos Lineales*, Serie Monografías, Vol. 7 No. 16, IIMAS - UNAM, 1ra. ed., México, 1998.

* Parzen, Emanuel, *Teoría Moderna de Probabilidades y sus Aplicaciones*, Ed. Limusa, Grupo Noriega Editores, México, 1979.

★ White, Douglas John, *Teoría de la decisión*, Ed. Alianza, versión española de García Molina, José Luis, Madrid, 1979.

★ Keeney, Ralph L. & Raiffa, Howard & Meyer, Richard F., *Decision with Multiply Objetives: Prefernces and Value Tradeoffs*, Ed. Wiley, Wiley Series in Probability and Mathematical Statistics, EUA, 1976.

★ Lindley, D. V., *Principios de la Teoría de la Decisión*, Ed. Vicens - Vives, 1ra. ed., España, 1977

★ Jones, J. Morgan, *Introduction to Decision Theory*, Ed. Richard D. Irvin, Inc., 1ra. ed., Illinois, EUA, 1977.

★ Evans, Merran & Hastings, Nocholas & Peacock, Brian, *Statistical distributions*, Ed. Wiley, Wiley Series in Probability and Statistics, 3ra. ed., EUA, 2000.

★ Lamport, Leslie, *LaTeX: A Document Preparation System*. Ed. Addison-Wesley. Reading. Massachusetts. 2da. ed., 1994.

★ Knuth, Donald E., *The TeX book*, Tomo A de *Computers and Typesetting*, Ed. Addison-Wesley Publishing Company, 1984.

★ Goossens, Michel & Mittelbach, Frank & Samarin, Alexander, *The LaTeXCompanion*, Ed. Addison-Wesley, Reading, Massachussets, 1994.

★ Bautista, Tomás & Oetiker, Tobias & Partl, Hubert & Hyna, Irene & Schlegl, Elisabeth, *Una Descripción de LaTeX2ε*, Manual del *Centro de Microelectrónica Aplicada* de la Universidad de Las Palmas de G. C., versión 0.4b, 11 de noviembre de 1998.

Apéndice A

Propiedades de la función indicadora

El método de las funciones indicadoras consiste en interpretar operaciones con eventos en términos de operaciones artiméticas.

Dado un evento A, en un espacio de descripciones muestrales S, se define la función indicadora del evento, denotada por $\mathbb{I}_{\{A\}}$, como una función de variable real definida en S, con valor en cualquier descripción s, denotada por $\mathbb{I}_{\{A\}}^{(s)}$, e igual a 1 o 0, dependiendo de que la descripción s pertenezca o no a A, es decir,

$$\mathbb{I}_{\{A\}}^{(s)} = \begin{cases} 1 & si \quad s \in A \\ 0 & si \quad s \in A^c \end{cases}$$

Las dos propiedades básicas de las funciones indicadoras, que permiten hacer operaciones con ellas, son las siguientes.

Primera, un producto de funciones indicadoras siempres se puede reducir a una sola función indicadora; de manera más precisa, para cualesquiera eventos $A_1, A_2, ..., A_n$ se tiene que,

$$\mathbb{I}_{\{A_1\}}\mathbb{I}_{\{A_2\}}...\mathbb{I}_{\{A_n\}} = \mathbb{I}_{\{A_1 \cap A_2 \cap ... \cap A_n\}}$$

o bien,

$$\mathbb{I}_{\cap_{i=1}^n A_i} = \prod_{i=1}^{n} \mathbb{I}_{\{A_i\}}$$

La habilidad para escribir expresiones de este tipo para las funciones indicadoras de eventos compuestos, en términos de las funciones indicadoras de eventos que los componen, tiene importancia debido a lo siguiente:

Una ecuación que contenga únicamente sumas y diferencias (pero no productos) de funciones indicadoras, conduce inmediatamente a una ecuación correspondiente a probabilidades; esta relación se obtiene al reemplazar $\mathbb{I}(.)$ por $P(.)$. Por ejemplo, si se hace esa sustitución en $\mathbb{I}_{\{A \cup B\}} = \mathbb{I}_{\{A\}} + \mathbb{I}_{\{B\}} - \mathbb{I}_{\{A \cap B\}}$, se obtiene la bien conocida fórmula $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

De esta gran importancia, característica de las funciones indicadoras, se tiene que

$$E(\mathbb{I}_{\{A\}}) = P(A)$$

A partir de las dos propiedades básicas de las funciones indicadoras, se pueden generar propiedades secundarias de las mismas, que también pueden describir (con más detalle y de manera particular) la caracterización de las funciones indicadoras y así, crear un conocimiento de mayor profundidad sobre ellas.

A continuación se especificarán algunas de las propiedades que, en particular, pueden brindar una ayuda mayor al manejarlas en cierto contexto de interés:

a) $\mathbb{I}_{\{A \cap B\}} = f(\mathbb{I}_{\{A\}}, \mathbb{I}_{\{B\}}) = \begin{cases} 1 & si\ ocurre \quad A \cap B \\ 0 & si\ ocurre \quad (A \cap B)^c \end{cases} = \mathbb{I}_{\{A\}}\mathbb{I}_{\{B\}}$

b) $\mathbb{I}_{\{A \cup B\}} = \mathbb{I}_{\{A\}} + \mathbb{I}_{\{B\}} - \mathbb{I}_{\{A \cap B\}} = \mathbb{I}_{\{A\}} + \mathbb{I}_{\{B\}} - \mathbb{I}_{\{A\}}\mathbb{I}_{\{B\}}$

c)
$$\begin{aligned} \mathbb{I}_{\{\cup_{i=1}^{n} A_i\}} &= \sum_{i=1}^{n} \mathbb{I}_{\{A_i\}} - \sum_{i<j} \mathbb{I}_{\{A_i A_j\}} + \sum_{i<j<k} \mathbb{I}_{\{A_i\}}\mathbb{I}_{\{A_j\}}\mathbb{I}_{\{A_k\}} \\ &\quad - \dots + \sum_{i=1}^{n} (-1)^{n+1}\, \mathbb{I}_{\{A_1\}}\, \mathbb{I}_{\{A_2\}} \dots \mathbb{I}_{\{A_n\}} \\ &= \sum_{i=1}^{n} (-1)^{n+1} \sum_{j_1<j_2<\dots<j_r} \mathbb{I}_{\{\cap_{i=1}^{n} A_{j_k}\}} \end{aligned}$$

d) $\mathbb{I}_{\{A^c\}} = 1 - \mathbb{I}_A = \begin{cases} 1 & si\ ocurre \quad A^c \\ 0 & si\ ocurre \quad (A^c)^c = A \end{cases}$

e) $\mathbb{I}_{\{A/B\}} = \mathbb{I}_{\{A-B\}} = \mathbb{I}_{\{A \cap B^c\}} = \mathbb{I}_{\{A\}}\mathbb{I}_{\{B^c\}} = \mathbb{I}_{\{A\}}(1 - \mathbb{I}_{\{B\}})$

$$= \mathbb{I}_{\{A\}} - \mathbb{I}_{\{A\}}\mathbb{I}_{\{B\}} = \mathbb{I}_{\{A\}} - \mathbb{I}_{\{A\cap B\}}$$

f) $\mathbb{I}_{\{A\triangle B\}} = \mathbb{I}_{\{A\}} + \mathbb{I}_{\{B\}} - 2\mathbb{I}_{\{A\}}\mathbb{I}_{\{B\}}$

g) $\left(\mathbb{I}_{\{A\}}\right)^n = \mathbb{I}_{\{A\}},\ n \in \mathbb{N}$

h) $\int_A \mathbb{I}_{\{B\}}^{(u)}\, du = \int_{A\cap B} du$