



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**UN ENFOQUE BAYESIANO PARA DETERMINAR EL
NÚMERO DE USUARIOS SUBSECUENTES DE UN
NUEVO TRATAMIENTO MÉDICO**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

P R E S E N T A :

LAURA NAYELI LILJEHULT LEÓN



**DIRECTOR DE TESIS:
DRA. RUTH SELENE FUENTES GARCÍA
2009**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**A Rosalba, Juan y Dayna,
por su paciencia y cariño**

<p>1. Datos del alumno Liljehult León Laura Nayeli 55 88 59 00 Universidad Nacional Autónoma de México Facultad de Ciencias Actuaría 302129245</p>
<p>2. Datos del tutor Dra. Ruth Selene Fuentes García</p>
<p>3. Datos del sinodal 1 Dra. Silvia Ruíz-Velasco Acosta</p>
<p>4. Datos del sinodal 2 M. en C. Jésica Hernández Rojano</p>
<p>5. Datos del sinodal 3 M. en C. Gerardo Rubio Hernández</p>
<p>6. Datos del sinodal 4 M. en C. Nelson Omar Muriel Torrero</p>
<p>7. Datos del trabajo escrito Un enfoque bayesiano para determinar el número de usuarios subsecuentes de un nuevo tratamiento médico 90 p 2009</p>

Índice general

Introducción	1
1. Preliminares	3
1.1. Muestreo de la distribución Normal	6
1.2. Convergencia de variables aleatorias y Teoremas límite	8
1.3. Intercambiabilidad	15
1.4. Análisis bayesiano	19
1.4.1. Familias conjugadas	22
1.4.2. Teoría de decisión	24
2. Distribución del número de usuarios subsecuentes	27
2.1. Número de usuarios subsecuentes	29
2.2. Tamaño óptimo de la muestra	33
2.3. Distribución del número de usuarios	38
3. Prueba clínica	45
Discusión	59
Tamaño óptimo de la muestra	61
Distribución del número de usuarios	61
Ejemplo	62
Conclusiones	73
A. Distribuciones	77
A.1. Distribución Gamma	77
A.2. Distribución Gamma-Inversa	77
A.3. Modelo Normal-Gamma	79
B. Programas en R	81
B.1. Cálculo de la esperanza	81
B.2. Modelo Gamma	82

Introducción

La motivación del tema de esta tesis fue evidenciar, de una manera particular, la utilidad de un enfoque bayesiano en el análisis de datos. El medio para lograrlo fue una aplicación médica, específicamente, la comparación de dos tratamientos médicos.

El objetivo de discutir la función de distribución del número de usuarios subsecuentes de un nuevo tratamiento médico es determinar si realmente es conveniente su implementación. En caso de que la respuesta sea afirmativa, nos será posible estimar el número esperado de usuarios que preferirán dicho tratamiento. Para este fin utilizamos una técnica bayesiana, donde la distribución de los datos que se propone considerar es Normal lo que permite el empleo de un modelo conjugado.

En el primer capítulo, enunciamos las definiciones y resultados necesarios para el desarrollo de la tesis. Basados en el concepto de intercambiabilidad justificamos el empleo de un enfoque bayesiano para obtener la función de distribución del número de usuarios subsecuentes. También, desarrollamos conceptos básicos de teoría de la decisión, los cuales utilizamos en el segundo capítulo para obtener el tamaño óptimo de la muestra de una prueba, donde el número de usuarios subsecuentes del nuevo tratamiento bajo investigación depende de la evidencia proporcionada por la misma.

Es en el segundo capítulo donde se establece la relación existente entre el número de usuarios subsecuentes, la muestra y la distribución posterior de los parámetros desconocidos en el modelo. Con base en esta relación obtenemos la función de distribución del número de usuarios subsecuentes de un nuevo tratamiento.

En el tercer capítulo presentamos un ejemplo práctico de la técnica desarrollada en esta tesis. El tratamiento bajo investigación es la administración, en diferentes dosis y vías de interferon beta-1a, en el control de esclerosis

múltiple con recaída-remisión. En este caso encontramos el número esperado de usuarios subsecuentes del nuevo tratamiento.

En un último apartado desarrollamos un modelo general donde el estadístico considerado tiene función de distribución F , no necesariamente Normal, y una distribución inicial que puede producir un modelo no conjugado.

Finalmente presentamos un ejemplo, numericamente sencillo, y comparamos los resultados obtenidos mediante el empleo de ambos modelos, con el objetivo de analizar las posibles diferencias y determinar si el supuesto de normalidad genera resultados alejados de la realidad.

Capítulo 1

Preliminares

A lo largo de esta tesis trabajamos con variables aleatorias X_1, X_2, \dots , en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, con Ω el espacio muestral, \mathcal{F} una σ -álgebra y \mathbb{P} una medida de probabilidad.

Comenzamos con conceptos básicos de estadística, área que tiene por tarea analizar e interpretar la información obtenida de una cierta población con la finalidad de concluir sobre la ley de probabilidad que rige un fenómeno aleatorio de interés. Este procedimiento se conoce como inferencia estadística.

En este trabajo proponemos una función de distribución F para los datos analizados que depende de un parámetro θ , es decir, consideramos una familia de distribuciones

$$F^* = \{F(x, \theta) : \theta \in \Theta\}$$

donde Θ es el conjunto de posibles valores del parámetro denominado *espacio paramétrico*.

Al obtener las observaciones con las que se intenta dar un pronóstico sobre la ley de probabilidad del fenómeno en cuestión, se busca que el experimento se realice bajo las mismas condiciones sin importar resultados previos, es decir, obtener observaciones independientes.

Cuando se ha obtenido una muestra de la población objeto de análisis, se tienen n observaciones numéricas (x_1, \dots, x_n) , pero cuando se está planificando el procedimiento de muestreo y diseñando el método para realizar inferencias, aún no se sabe qué valores numéricos se obtendrán, entonces consideramos n variables aleatorias (*v.a*) X_1, X_2, \dots, X_n independientes e idénticamente dis-

tribuidas.

Definición 1. Sea X_1, X_2, \dots, X_n , $n \in \mathbb{N}$ un conjunto de variables aleatorias. Se dice que X_1, \dots, X_n forman una muestra aleatoria (m.a) si X_1, \dots, X_n son independientes e idénticamente distribuidas, es decir, si $\underline{X} = (X_1, \dots, X_n)$ es una muestra aleatoria se satisface:

1. $X_i \sim F(x; \theta) \forall i \in \{1, \dots, n\}$.
2. $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F(x_i; \theta)$.

Definición 2. Sea X_1, X_2, \dots, X_n m.a.. Una estadística es una función de la muestra aleatoria, $T(\underline{X})$, que no depende de algún parámetro desconocido.

Llamamos distribución de una muestra aleatoria a la distribución conjunta de las variables aleatorias que componen dicha muestra.

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F(x_i; \theta).$$

Sea X_1, X_2, \dots, X_n una m.a. de $F(X; \theta)$, tal que $\mathbb{E}[X_i] = \mu$ y $\text{Var}(X_i) = \sigma^2$. Para facilitar la notación definimos las siguientes estadísticas :

1. Media muestral $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$.
2. Varianza muestral $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$.

Es común que surja la pregunta de por qué definir la varianza muestral de esta forma y no como $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$, la razón principal es que el valor esperado de S^2 es igual a la varianza de la población, es decir $\mathbb{E}[S^2] = \sigma^2$ como se demuestra en la siguiente proposición.

Proposición 1.0.1. Sea X_1, X_2, \dots, X_n una m.a. de $F(X; \theta)$, tal que

$$\mathbb{E}[X_i] = \mu \quad y \quad \text{Var}(X_i) = \sigma^2.$$

Se tiene entonces que:

- a) $\mathbb{E}[\bar{X}] = \mu$.
- b) $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$.

c) $Var(S^2) = \sigma^2$.

Demostración. a)

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n}(n\mu) = \mu.$$

b)

$$\begin{aligned} Var(\bar{X}) &= Var\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) \\ &= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}. \end{aligned}$$

c) Primero demostremos la siguiente identidad

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2.$$

Tenemos que

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu)^2 \\ &= \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu)]^2 \\ &= \sum_{i=1}^n \left[(X_i - \bar{X})^2 + 2(X_i - \bar{X})(\bar{X} - \mu) + (\bar{X} - \mu)^2 \right] \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + 2(\bar{X} - \mu)\sum_{i=1}^n (X_i - \bar{X}) + n(\bar{X} - \mu)^2 \end{aligned}$$

como

$$\sum_{i=1}^n (X_i - \bar{X}) = \sum_{i=1}^n X_i - n\bar{X} = \sum_{i=1}^n X_i - n\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = 0,$$

entonces

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2.$$

Utilizando la identidad anterior

$$\begin{aligned}
 \mathbb{E}[S^2] &= \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^n(X_i - \bar{X})^2\right] \\
 &= \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^n(X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] \\
 &= \frac{1}{n-1}\left\{\sum_{i=1}^n\mathbb{E}[(X_i - \mu)^2] - n\mathbb{E}[(\bar{X} - \mu)^2]\right\} \\
 &= \frac{1}{n-1}\left\{\sum_{i=1}^n\sigma^2 - n\text{Var}(\bar{X})\right\} \\
 &= \frac{1}{n-1}\left\{n\sigma^2 - n\frac{\sigma^2}{n}\right\} \\
 &= \frac{1}{n-1}\{(n-1)\sigma^2\} \\
 &= \sigma^2.
 \end{aligned}$$

□

1.1. Muestreo de la distribución Normal

Una variable aleatoria X tiene distribución Normal con parámetros μ y σ^2 , donde $-\infty < \mu < \infty$ y $0 < \sigma^2 < \infty$, si su función de densidad es de la forma:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad -\infty < x < \infty \quad (1.1)$$

Es importante mencionar que si $X \sim N(\mu, \sigma^2)$ y a, b son dos números reales, entonces $aX + b \sim N(b + a\mu, a^2\sigma^2)$.

Si X_1, \dots, X_n son variables aleatorias independientes, tales que

$$X_i \sim N(\mu_i, \sigma_i^2)$$

tenemos que

$$\sum_{i=1}^n X_i \sim N \left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2 \right).$$

Para facilitar los cálculos que se realizarán en el próximo capítulo, demostramos el siguiente resultado.

Proposición 1.1.1. *Sea X_1, \dots, X_n una m.a de la distribución $N(\mu, \sigma^2)$, entonces*

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Demostración. Haremos la demostración a través de la función generadora de momentos

$$\begin{aligned} M_{\bar{X}}(t) &= \mathbb{E} \left[\exp \{ t \bar{X} \} \right] \\ &= \mathbb{E} \left[\exp \left\{ t \frac{\sum_{i=1}^n X_i}{n} \right\} \right] \\ &= \mathbb{E} \left[\prod_{i=1}^n \exp \left\{ \frac{t}{n} X_i \right\} \right] \\ &= \prod_{i=1}^n \mathbb{E} \left[\exp \left\{ \frac{t}{n} X_i \right\} \right] \\ &= \prod_{i=1}^n M_{X_i} \left(\frac{t}{n} \right) \\ &= \prod_{i=1}^n \exp \left\{ \frac{t}{n} \mu + \frac{t^2 \sigma^2}{2n^2} \right\} \\ &= \exp \left\{ t \mu + \frac{t^2 \sigma^2}{2n} \right\} \end{aligned}$$

que es la función generadora de momentos de una *v.a* Normal con media μ y varianza σ^2/n , es decir $\bar{X} \sim N(\mu, \sigma^2/n)$. \square

1.2. Convergencia de variables aleatorias y Teoremas límite

En muchas ocasiones trabajamos con variables aleatorias con una función de distribución que, debido a la complicación en el cálculo de probabilidades, no nos permite hacer inferencias de manera sencilla. Recurrimos entonces a la convergencia de variables aleatorias para obtener resultados asintóticos.

En esta sección exponemos diversos tipos de convergencia, las relaciones entre ellos y al final enunciamos dos teoremas importantes: el “Teorema central del límite” y la “Ley fuerte de los grandes números”.¹

En todo momento trabajamos con sucesiones $\{X_i\}_{i \geq 1}$ de variables aleatorias en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$.

Comenzamos con un tipo de convergencia usualmente conocido.

Definición 3 (Convergencia puntual). *La sucesión de variables aleatorias $\{X_i\}_{i \geq 1}$ converge puntualmente a X en ω , si*

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Si $\{X_i\}_{i \geq 1}$ converge puntualmente a X en ω para todo $\omega \in A$, se dice que $\{X_i\}_{i \geq 1}$ converge puntualmente a X en A .

Definición 4 (Convergencia casi segura). *La sucesión de variables aleatorias $\{X_i\}_{i \geq 1}$ converge casi seguramente a X v.a., si*

$$\mathbb{P} \left[\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} \right] = 1.$$

Notación: $X_n \xrightarrow{c.s.} X$.

La convergencia casi segura nos indica que el conjunto donde la sucesión converge puntualmente a X tiene probabilidad uno, es decir, la probabilidad de no convergencia es cero.

¹Para consultar acerca de este tema ver [3], [4], [5].

Definición 5 (Convergencia en probabilidad). *La sucesión de variables aleatorias $\{X_i\}_{i \geq 1}$ converge en probabilidad a la v.a. X , si para cada $\epsilon > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}[\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}] = 0.$$

Notación: $X_n \xrightarrow{\mathbb{P}} X$.

A continuación exhibimos la relación que existe entre la convergencia casi segura y la convergencia en probabilidad.

Teorema 1.2.1. *Si $X_n \xrightarrow{c.s.} X$, entonces $X_n \xrightarrow{\mathbb{P}} X$.*

*Demostración.*² Sea $\epsilon > 0$. Para $n \in \mathbb{N}$ definimos

$$A_n = \bigcup_{k=n}^{\infty} (|X_k - X| > \epsilon).$$

Esta sucesión es decreciente y su límite es la intersección de todos los eventos A_n . Como $(|X_n - X| > \epsilon) \subseteq A_n$, entonces $\mathbb{P}[|X_n - X| > \epsilon] \leq \mathbb{P}[A_n]$.

Por lo tanto,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| > \epsilon] &\leq \lim_{n \rightarrow \infty} \mathbb{P}[A_n] \\ &= \mathbb{P}\left[\lim_{n \rightarrow \infty} A_n\right] \\ &= \mathbb{P}\left[\bigcap_{n=1}^{\infty} A_n\right] \\ &= \mathbb{P}\left[\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|X_k - X| > \epsilon\}\right] \\ &= \mathbb{P}[\{\forall n \in \mathbb{N} \exists k \geq n : |X_k - X| > \epsilon\}] \\ &= \mathbb{P}[\{\omega : X_n \not\rightarrow X\}] \\ &= 0. \end{aligned}$$

□

²Esta demostración fue tomada de 'Curso intermedio de probabilidad', [2].

El siguiente tipo de convergencia es conocido también como *convergencia débil* y de todas las formas de convergencia que hemos mencionado, es el que impone menos restricciones.

Definición 6 (Convergencia en distribución). *La sucesión de variables aleatorias $\{X_i\}_{i \geq 1}$ converge en distribución a X , si para todo punto x donde la función $F_X(x)$ es continua, se cumple que*

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

Notación: $X_n \xrightarrow{d} X$, o bien, $F_{X_n} \xrightarrow{d} F_X$.

Resulta conveniente aclarar que las variables aleatorias en la definición anterior no necesitan estar definidas en el mismo espacio de probabilidad, aunque usualmente se considera que sí.

En el último capítulo utilizamos resultados específicos de la convergencia de variables aleatorias, los cuales enunciamos a continuación.

Lema 1.2.1 (Lema de Slutsky). *Si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{\mathbb{P}} c$, donde c es constante, entonces*

$$X_n Y_n \xrightarrow{d} cX.$$

Demostración. Sea $c > 0$ y δ tal que $0 < \delta < c$. Como $Y_n \xrightarrow{\mathbb{P}} c$, para todo $\epsilon > 0$ $\lim_{n \rightarrow \infty} \mathbb{P}[|Y_n - c| > \epsilon] = 0$. Podemos encontrar N tal que, para todo $n \geq N$

$$\mathbb{P}[|Y_n - c| > \delta] < \delta.$$

Sea $x \geq 0$,

$$\begin{aligned} \mathbb{P}[X_n Y_n \leq x] &= \mathbb{P}[X_n Y_n \leq x, |Y_n - c| \leq \delta] + \mathbb{P}[X_n Y_n \leq x, |Y_n - c| > \delta] \\ &\leq \mathbb{P}[X_n Y_n \leq x, |Y_n - c| \leq \delta] + \mathbb{P}[|Y_n - c| > \delta]. \end{aligned}$$

La condición $|Y_n - c| \leq \delta$ es equivalente a $-\delta \leq Y_n - c \leq \delta$, así

$$\begin{aligned}
\mathbb{P}[X_n Y_n \leq x] &\leq \mathbb{P}[X_n Y_n \leq x, -\delta \leq Y_n - c \leq \delta] + \delta \\
&= \mathbb{P}[X_n Y_n \leq x, c - \delta \leq Y_n \leq c + \delta] + \delta \\
&= \mathbb{P}\left[X_n \leq \frac{x}{Y_n}, c - \delta \leq Y_n \leq c + \delta\right] + \delta \\
&\leq \mathbb{P}\left[X_n \leq \frac{x}{c - \delta}, c - \delta \leq Y_n \leq c + \delta\right] + \delta \\
&\leq \mathbb{P}\left[X_n \leq \frac{x}{c - \delta}\right] + \delta.
\end{aligned}$$

De manera similar

$$\begin{aligned}
\mathbb{P}[X_n Y_n > x] &\leq \mathbb{P}[X_n Y_n > x, |Y_n - c| \leq \delta] + \delta \\
&\leq \mathbb{P}\left[X_n > \frac{x}{c + \delta}\right] + \delta.
\end{aligned}$$

La última desigualdad implica que

$$\mathbb{P}[X_n Y_n \leq x] = 1 - \mathbb{P}[X_n Y_n > x] \geq 1 - \mathbb{P}\left[X_n > \frac{x}{c + \delta}\right] - \delta = \mathbb{P}\left[X_n \leq \frac{x}{c + \delta}\right] - \delta,$$

es decir,

$$\mathbb{P}\left[X_n \leq \frac{x}{c + \delta}\right] - \delta \leq \mathbb{P}[X_n Y_n \leq x] \leq \mathbb{P}\left[X_n \leq \frac{x}{c - \delta}\right] + \delta.$$

Tomando el límite cuando $n \rightarrow \infty$ y $\delta \downarrow 0$, tenemos

$$\mathbb{P}[X_n Y_n \leq x] \longrightarrow \mathbb{P}\left[X \leq \frac{x}{c}\right] = \mathbb{P}[cX \leq x]$$

si x/c es un punto de continuidad de la función de distribución de X .

Con un argumento similar para $x < 0$, podemos concluir $X_n Y_n \xrightarrow{d} cX$ si $c > 0$. Los casos $c < 0$ y $c = 0$ son similares. □

En la demostración anterior utilizamos el siguiente hecho:

Observación 1.2.1. Sean X variable aleatoria y c constante, entonces x/c es punto de continuidad de la función de distribución de X si y sólo si x es punto de continuidad de la función de distribución de cX .

Demostración. Sean F_X y F_{cX} las funciones de distribución de X y cX respectivamente. Consideremos el caso $c > 0$.

\Rightarrow] Si x/c es punto de continuidad de F_X , entonces

$$\lim_{y \rightarrow x/c} \mathbb{P}[X \leq y] = \lim_{y \rightarrow x/c} F_X(y) = F_X(x/c) = \mathbb{P}[X \leq x/c],$$

es decir,

$$\mathbb{P}[cX \leq x] = \lim_{y \rightarrow x/c} \mathbb{P}[X \leq y] = \lim_{y \rightarrow x/c} \mathbb{P}[cX \leq cy] = \lim_{z \rightarrow x} \mathbb{P}[cX \leq z].$$

Por lo tanto x es punto de continuidad de F_{cX} .

\Leftarrow] Si x es punto de continuidad de F_{cX} , entonces

$$\lim_{y \rightarrow x} \mathbb{P}[cX \leq y] = \lim_{y \rightarrow x} F_{cX}(y) = F_{cX}(x) = \mathbb{P}[cX \leq x]$$

es decir,

$$\mathbb{P}[X \leq x/c] = \lim_{y \rightarrow x} \mathbb{P}[cX \leq y] = \lim_{y \rightarrow x} \mathbb{P}[X \leq y/c] = \lim_{z \rightarrow x/c} \mathbb{P}[X \leq z].$$

Por lo tanto x/c es punto de continuidad de F_X .

Los casos $c > 0$ y $c = 0$ son similares. □

Otro resultado útil es el siguiente:

Proposición 1.2.1. *Sea c constante, si $X_n \xrightarrow{d} c$ entonces $X_n \xrightarrow{\mathbb{P}} c$.*

Demostración. La función de distribución de c es

$$F(x) = \begin{cases} 0 & \text{si } x < c \\ 1 & \text{si } x \geq c \end{cases}$$

donde el único punto de discontinuidad es $x = c$.

Como $X_n \xrightarrow{d} c$, $F_{X_n}(x) \rightarrow F(x)$ para $x \neq c$.

Sea $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}[|X_n - c| > \epsilon] &= 1 - \mathbb{P}[|X_n - c| \leq \epsilon] \\
&= 1 - \mathbb{P}[-\epsilon \leq X_n - c \leq \epsilon] \\
&= 1 - \mathbb{P}[c - \epsilon \leq X_n \leq c + \epsilon]
\end{aligned}$$

ya que $\mathbb{P}[X_n = c - \epsilon] = 0$ tenemos que

$$\begin{aligned}
\mathbb{P}[|X_n - c| > \epsilon] &= 1 - [\mathbb{P}[X_n \leq c + \epsilon] - \mathbb{P}[X_n \leq c - \epsilon]] \\
&= 1 - F_{X_n}(c + \epsilon) + F_{X_n}(c - \epsilon)
\end{aligned}$$

tomando límite cuando $n \rightarrow \infty$

$$\begin{aligned}
\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - c| > \epsilon] &= 1 - \lim_{n \rightarrow \infty} F_{X_n}(c + \epsilon) + \lim_{n \rightarrow \infty} F_{X_n}(c - \epsilon) \\
&= 1 - F(c + \epsilon) + F(c - \epsilon) \\
&= 1 - 1 + 0 \\
&= 0.
\end{aligned}$$

Por lo tanto $X_n \xrightarrow{\mathbb{P}} c$.

□

Hasta el momento nos hemos referido a la convergencia de variables aleatorias X_1, X_2, \dots , sin embargo, en el último capítulo será muy útil tener información respecto a la convergencia de $f(X_1), f(X_2), \dots$, para f continua cuando $X_n \xrightarrow{c.s.} X$.

Proposición 1.2.2. *Sea f una función continua.*

a) Si $X_n \xrightarrow{c.s.} X$, entonces $f(X_n) \xrightarrow{c.s.} f(X)$.

b) Si $X_n \xrightarrow{\mathbb{P}} X$, entonces $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

*Demostración.*³

a) Sea

$$D = \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\}$$

³Demostración tomada de 'Probability essentials', [6].

Como $X_n \xrightarrow{c.s.} X$, $\mathbb{P}[D] = 1$.

Por hipótesis f es continua, si $\omega \in D$, entonces

$$\lim_{n \rightarrow \infty} f(X_n(\omega)) = f\left(\lim_{n \rightarrow \infty} X_n(\omega)\right) = f(X(\omega))$$

y esto se cumple $\forall \omega \in D$, es decir

$$\mathbb{P}\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} f(X_n(\omega)) = f(X(\omega))\right\} = 1$$

b) Sea $\epsilon > 0$. Observemos que para cada $K > 0$

$$\begin{aligned} \{|f(X_n) - f(X)| > \epsilon\} &= \{|f(X_n) - f(X)| > \epsilon, |X| \leq K\} \cup \\ &\quad \{|f(X_n) - f(X)| > \epsilon, |X| > K\} \\ &\subset \{|f(X_n) - f(X)| > \epsilon, |X| \leq K\} \cup \\ &\quad \{|X| > K\}. \end{aligned}$$

Como f es continua, es uniformemente continua en el intervalo $[-K, K]$, por lo tanto existe $\delta > 0$ tal que

$$|f(x) - f(y)| \leq \epsilon \quad \text{si } |x - y| < \delta \quad \text{con } x, y \in [-K, K].$$

Lo anterior significa que

$$\begin{aligned} \{|f(X_n) - f(X)| > \epsilon, |X| \leq K\} &\subset \{|X_n - X| > \delta, |X| \leq K\} \\ &\subset \{|X_n - X| > \delta\}. \end{aligned}$$

Entonces,

$$\mathbb{P}[|f(X_n) - f(X)| > \epsilon] \leq \mathbb{P}[|X_n - X| > \delta] + \mathbb{P}[|X| > K].$$

Veamos que

$$0 = \mathbb{P}\left[\bigcap_{K=1}^{\infty} \{|X| > K\}\right] = \mathbb{P}\left[\lim_{K \rightarrow \infty} |X| > K\right] = \lim_{K \rightarrow \infty} \mathbb{P}[|X| > K].$$

Ahora, para $\gamma > 0$ elegimos K tal que $\mathbb{P}[|X| > K] < \gamma$. Cuando K está fijo obtenemos δ de la expresión

$$\{|f(X_n) - f(X)| > \epsilon\} \subset \{|X_n - X| > \delta\} \cup \{|X| > K\}.$$

Por lo tanto

$$\lim_{n \rightarrow \infty} \mathbb{P} [|f(X_n) - f(X)| > \epsilon] \leq \lim_{n \rightarrow \infty} \mathbb{P} [|X_n - X| > \delta] + \gamma = \gamma,$$

y como γ fue arbitrario

$$\lim_{n \rightarrow \infty} \mathbb{P} [|f(X_n) - f(X)| > \epsilon] = 0.$$

□

Para finalizar esta sección, enunciemos dos teoremas importantes.

Teorema 1.2.2 (Teorema central del límite). ⁴ Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas, tales que $\mathbb{E}[X_i] = \mu$ y $\text{Var}(X_i) = \sigma^2 < \infty$. Entonces

$$\frac{1}{\sqrt{n}\sigma} \left(\sum_{i=1}^n X_i - n\mu \right) \xrightarrow{d} N(0, 1).$$

Teorema 1.2.3 (Ley fuerte de los grandes números). ⁵ Sean X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas, tales que $\mathbb{E}[X_i] = \mu$. Entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{c.s} \mu.$$

1.3. Intercambiabilidad

En esta sección introducimos las ideas necesarias para relacionar independencia condicional con el concepto de intercambiabilidad, a partir del cual damos justificación al modelo bayesiano que desarrollamos en el siguiente capítulo.⁶

⁴Para la demostración consultar [1], [2], [4].

⁵Para la demostración consultar [1], [2], [4].

⁶Sobre este tema ver [7].

Definición 7. Las variables aleatorias $\{X_i, 1 \leq i \leq n\}$ son llamadas intercambiables si la distribución conjunta es invariante bajo permutaciones, es decir

$$\{X_1, \dots, X_n\} \stackrel{d}{=} \{X_{\tau(1)}, \dots, X_{\tau(n)}\}$$

para toda τ permutación de $\{1, \dots, n\}$.

Se dice que las variables aleatorias $\{X_i, n \geq 1\}$ son intercambiables si cualquier subconjunto finito de ellas es intercambiable.

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad y \mathcal{G} una sub- σ -álgebra de \mathcal{F} . Para $A, B \in \mathcal{F}$ se define la probabilidad condicional de un evento A dado \mathcal{G} como

$$\mathbb{P}[A|\mathcal{G}] = \mathbb{E}[\mathbb{I}_A|\mathcal{G}],$$

la cual es una función \mathcal{G} -medible en Ω y satisface

$$\int_G \mathbb{P}[A|\mathcal{G}] d\mathbb{P} = \int_G \mathbb{I}_A d\mathbb{P} = \mathbb{P}[AG] \quad G \in \mathcal{G}.$$

La probabilidad condicional de un evento A dado un evento B , cuando $\mathbb{P}[B] > 0$, se define como

$$\mathbb{P}[A|B] = \mathbb{E}[\mathbb{I}_A|\sigma(\mathbb{I}_B)] = \frac{1}{\mathbb{P}[B]} \int_B \mathbb{I}_A d\mathbb{P} = \frac{\mathbb{P}[AB]}{\mathbb{P}[B]}.$$

Definición 8. Sean \mathcal{F}, \mathcal{G} σ -álgebras de eventos. Una medida de probabilidad condicional en \mathcal{F} dado \mathcal{G} es una función $\mathbb{P}[A, \omega]$ definida para $A \in \mathcal{F}$ y $\omega \in \Omega$ tal que

- i. para cada $\omega \in \Omega$, $\mathbb{P}[\cdot, \omega]$ es una medida de probabilidad en \mathcal{F} ,
- ii. para cada $A \in \mathcal{F}$, $\mathbb{P}[A, \cdot]$ es una función \mathcal{G} -medible en Ω que coincide con la probabilidad condicional de A dado \mathcal{G} , es decir, $\mathbb{P}[A, \omega] = \mathbb{P}[A|\mathcal{G}](\omega)$, casi seguramente.

Observemos el siguiente ejemplo.

Ejemplo 1.3.1. Sea (X_1, X_2) un vector aleatorio absolutamente continuo en el espacio de probabilidad $(\mathbb{R}^2, \sigma(X_1, X_2), \mathbb{P})$, es decir que para alguna función de Borel f

$$F(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(s, t) ds dt \quad (x_1, x_2) \in \mathbb{R}^2$$

donde F representa la distribución conjunta de (X_1, X_2) .

Sea $\mathcal{F} = \sigma(X_1, X_2)$, $\mathcal{G} = \sigma(X_2) = \{\mathbb{R} \times B : B \in \mathcal{B}\}$, y para $(x_1, x_2) \in \mathbb{R}^2$ definimos

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, t) dt, \quad f_2(x_2) = \int_{-\infty}^{\infty} f(s, x_2) ds \quad y$$

$$f_1(x_1|x_2) = \begin{cases} f(x_1, x_2)/f_2(x_2) & \text{si } f_2(x_2) > 0 \\ f_1(x_1) & \text{si } f_2(x_2) = 0 \end{cases}$$

Ahora, para $B \in \mathcal{F}$ y $x = (x_1, x_2) \in \mathbb{R}^2$ definimos

$$\mathbb{P}[B, x] = \int_{[s:(s, x_2) \in B]} f_1(s|x_2) ds$$

entonces para cada $x \in \mathbb{R}^2$, $\mathbb{P}[B, x]$ es una medida de probabilidad en \mathcal{F} y para cada $B \in \mathcal{F}$, $\mathbb{P}[B, x]$ es una función de Borel en x_2 y es $\sigma(X_2)$ -medible. Además, para $B \in \mathcal{F}$ y $A_2 = \mathbb{R} \times B_2 \in \sigma(X_2)$

$$\begin{aligned} \int_{A_2} \mathbb{P}[B, x] d\mathbb{P} &= \int_{B_2} \int_{-\infty}^{\infty} \mathbb{P}[(B, (s, t))] f(s, t) ds dt \\ &= \int_{B_2} \int_{-\infty}^{\infty} \left[\int_{[u:(u, t) \in B]} f_1(u|t) du \right] f(s, t) ds dt \\ &= \int_{B_2} \int_{[u:(u, t) \in B]} f_1(u|t) f_2(t) du dt \\ &= \int_{B_2} \int_{[u:(u, t) \in B]} f(u, t) du dt \\ &= \int_{B_2} \int_{-\infty}^{\infty} \mathbb{I}_B f(u, t) du dt \\ &= \mathbb{P}[BA_2]. \end{aligned}$$

De acuerdo con la definición de probabilidad condicional de un evento dada una sub- σ -álgebra, $\mathbb{P}[B, x] = \mathbb{P}[B|\sigma(X_2)](x)$ c.s. para cada $B \in \mathcal{F}$.

A $f_1(x_1|x_2)$ se le denomina la función de densidad condicional de X_1 dado $X_2 = x_2$.

La siguiente definición ayuda a establecer la relación entre intercambiabilidad e independencia condicional para variables aleatorias idénticamente distribuidas.

Definición 9. Sea \mathcal{G} una σ -álgebra de eventos y $\{\mathcal{G}_n, n \geq 1\}$ una sucesión de clases de eventos. La sucesión $\{\mathcal{G}_n, n \geq 1\}$ es condicionalmente independiente dada \mathcal{G} si para cualquier elección de $A_m \in \mathcal{G}_{k_m}$, donde $k_i \neq k_j$ para $i \neq j$, $m = 1, 2, \dots, n$ y $n = 2, 3, \dots$,

$$\mathbb{P}[A_1 A_2 \dots A_n | \mathcal{G}] = \prod_{i=1}^n \mathbb{P}[A_i | \mathcal{G}], \quad c.s.$$

Así, una sucesión de variables aleatorias $\{X_n, n \geq 1\}$ es llamada condicionalmente independiente dada \mathcal{G} si la sucesión de clases de eventos \mathcal{G}_n tal que $\mathcal{G}_n = \sigma(X_n)$ para $n \geq 1$, es condicionalmente independiente dada \mathcal{G} .

Observación 1.3.1. Las variables aleatorias X_1, \dots, X_n en $(\Omega, \mathcal{F}, \mathbb{P})$ son condicionalmente independientes dada una σ -álgebra de eventos \mathcal{G} , si y sólo si para todo $(x_1, \dots, x_n) \in \mathbb{R}^n$

$$\mathbb{P}\left[\bigcap_{i=1}^n [X_i < x_i] \middle| \mathcal{G}\right] = \prod_{i=1}^n \mathbb{P}[X_i < x_i | \mathcal{G}], \quad c.s.$$

Supongamos que $\{X_i\}_{i=1}^{\infty}$ son variables aleatorias condicionalmente independientes dada $\sigma(\Theta)$, con función de densidad $f(x|\theta)$ dado $\Theta = \theta$ y que Θ tiene función de distribución Π_{Θ} . Entonces la distribución conjunta de cualquier subvector $(X_{i_1}, \dots, X_{i_n})$ es

$$f_{X_{i_1}, \dots, X_{i_n}}(x_1, \dots, x_n) = \int \prod_{i=1}^n f(x_i | \theta) d\Pi_{\Theta}.$$

Si Π_{Θ} tiene función de densidad π_{Θ} la expresión se vería de la forma

$$f_{X_{i_1}, \dots, X_{i_n}}(x_1, \dots, x_n) = \int \prod_{i=1}^n f(x_i | \theta) \pi_{\Theta}(\theta) d\theta.$$

En ambos casos, la distribución conjunta de $(X_{i_1}, \dots, X_{i_n})$ no depende del orden de i_1, \dots, i_n . Obsevemos así, que la condición de independencia condicional e idéntica disitribución implica intercambiabilidad (en el caso finito).

En general se tiene el siguiente resultado

Teorema 1.3.1. *Las variables aleatorias $\{X_n\}_{n=1}^\infty$ en $(\Omega, \mathcal{F}, \mathbb{P})$ son intercambiables, si y solo si, son condicionalmente independientes e idénticamente distribuidas dada alguna σ -álgebra de eventos \mathcal{G} .*

Enunciemos el siguiente resultado.

Teorema 1.3.2 (Teorema de de Finetti). *Si $\{X_n, n \geq 1\}$ son variables aleatorias intercambiables, entonces existe una σ -álgebra de eventos \mathcal{G} tal que para todo $m \geq 1$*

$$\mathbb{P}[X_1 < x_1, \dots, X_m < x_m] = \int \prod_{j=1}^m \mathbb{P}[X_1 < x_j | \mathcal{G}] d\mathbb{P} \quad (1.2)$$

En el caso de modelos paramétricos, si consideramos variables aleatorias X_1, X_2, \dots condicionalmente independientes e idénticamente distribuidas dada la σ -álgebra generada por el parámetro, lo que el Teorema anterior asegura es que $\mathbb{P}[X_1 < x_1, \dots, X_m < x_m]$ puede verse como la integral del producto de las distribuciones condicionales de cada X_i dado que el parámetro toma un valor, respecto a una cierta medida. En estadística bayesiana esta medida es denominada *distribución inicial* o *distribución a priori*.

1.4. Análisis bayesiano

El análisis bayesiano representa la combinación entre la información que la muestra aleatoria (\underline{X}) nos proporciona y la información a priori ($\pi(\theta)$), en la llamada distribución posterior $\pi(\theta | X)$, sobre la cual son realizadas todas las inferencias.

Una manera usual de simplificar los cálculos en la inferencia estadística es plantear un modelo que describa el comportamiento de un fenómeno específico. Para simplificar este procedimiento, consideramos modelos paramétricos. En un enfoque bayesiano, los parámetros de este modelo se consideran variables aleatorias. Una ventaja de este enfoque es que si tenemos información respecto al fenómeno de investigación, puede ser incorporada al modelo mediante la distribución inicial de los parámetros.

Cuando no se cuenta con información previa acerca del fenómeno de estudio, surge la discusión del por qué utilizar un enfoque bayesiano y en base a qué elegir la distribución inicial. Surgen así, las denominadas distribuciones

iniciales no informativas. Estas distribuciones se obtienen a partir de la muestra aleatoria, ya que es la única información a la que se tiene acceso.

Existen diversos tipos de distribuciones iniciales no informativas, de acuerdo con la técnica empleada para obtenerlas. Como ejemplo podemos mencionar: *Inicial de Laplace, inicial invariante, inicial de Jeffreys.*⁷

Al momento de extraer una muestra aleatoria de la población que estamos estudiando, adquirimos cierta información, misma que se verá reflejada en la distribución posterior de los parámetros. Es decir, proponemos un modelo paramétrico con una distribución inicial de Θ , y cuando adquirimos información mediante una muestra, actualizamos dicho modelo al considerar ahora que nuestro parámetro tiene la distribución condicional $\pi(\theta|X)$.

Si la distribución inicial propuesta no puede ser completamente determinada por la información que tenemos entonces contendrá información que arbitrariamente suponemos cierta. Es importante evaluar la influencia de la parte arbitraria, en cantidades posteriores, y verificar que ésta no tenga una influencia predominante. A este proceso de evaluación lo llamamos *análisis de sensibilidad*.

Expuesto lo anterior, determinemos cómo calcular la distribución posterior.

Teorema 1.4.1 (Teorema de Bayes). *La distribución posterior de $\Theta|X$ se puede obtener como*

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} \propto f(x|\theta)\pi(\theta).$$

Demostración.

$$\pi(\theta|x)f(x) = f(x, \theta) = f(x|\theta)\pi(\theta)$$

es decir

$$\pi(\theta|x)f(x) = f(x|\theta)\pi(\theta)$$

si $f(x) \neq 0$

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{f(x)}$$

además

$$f(x) = \int f(x, \theta)d\theta = \int f(x|\theta)\pi(\theta)d\theta$$

⁷Ver [8].

así,

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}.$$

Observemos que la información necesaria para obtener $\pi(\theta|x)$ la tenemos con $f(x|\theta)\pi(\theta)$, además $\int f(x|\theta)\pi(\theta)d\theta$ es la constante de normalización, que hace que $\int \pi(\theta|x)d\theta = 1$. Por lo tanto

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} \propto f(x|\theta)\pi(\theta).$$

□

Ejemplo 1.4.1. Sea $X \sim N(\theta, \sigma^2)$, donde θ es desconocida, pero σ^2 es conocida. Si la distribución a priori de θ es una densidad $N(\mu, \tau^2)$, donde μ y τ^2 son conocidas. Entonces la función de densidad conjunta de (X, θ) es

$$h(x; \theta) = \pi(\theta)f(x|\theta) = (2\pi\sigma\tau)^{-1} \exp \left\{ -\frac{1}{2} \left[\frac{(\theta - \mu)^2}{\tau^2} + \frac{(x - \theta)^2}{\sigma^2} \right] \right\}.$$

Para encontrar la función de densidad marginal de X se integra $h(x; \theta)$ con respecto a θ

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} (2\pi\sigma\tau)^{-1} \exp \left\{ -\frac{1}{2} \left[\frac{(\theta - \mu)^2}{\tau^2} + \frac{(x - \theta)^2}{\sigma^2} \right] \right\} d\theta \\ &= (2\pi\sigma\tau)^{-1} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[\frac{\theta^2 - 2\theta\mu - \mu^2}{\tau^2} + \frac{x^2 - 2\theta x + \theta^2}{\sigma^2} \right] \right\} d\theta \\ &= (2\pi\sigma\tau)^{-1} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[\theta^2 \left(\frac{1}{\tau^2} + \frac{1}{\sigma^2} \right) - 2\theta \left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2} \right) \right] \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} \left(\frac{\mu^2}{\tau^2} + \frac{x^2}{\sigma^2} \right) \right\} d\theta \\ &= (2\pi\sigma\tau)^{-1} \exp \left\{ -\frac{1}{2} \left(\frac{\mu^2}{\tau^2} + \frac{x^2}{\sigma^2} \right) \right\} \\ &\quad \times \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left(\frac{\tau^2 + \sigma^2}{\tau^2\sigma^2} \right) \left[\theta^2 - 2\theta \left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2} \right) \left(\frac{\tau^2\sigma^2}{\tau^2 + \sigma^2} \right) \right] \right\} d\theta \end{aligned}$$

$$\begin{aligned}
&= (2\pi\sigma\tau)^{-1} \exp \left\{ -\frac{1}{2} \left(\frac{\mu^2}{\tau^2} + \frac{x^2}{\sigma^2} \right) \right\} \\
&\quad \times \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[\frac{\tau^2 + \sigma^2}{\tau^2\sigma^2} \left(\theta - \left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2} \right) \frac{\tau^2\sigma^2}{\tau^2 + \sigma^2} \right)^2 \right] \right\} \\
&\quad \times \exp \left\{ \frac{1}{2} \left[\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2} \right)^2 \frac{\tau^2\sigma^2}{\tau^2 + \sigma^2} \right] \right\} d\theta \\
&= \frac{1}{2\pi\sigma\tau} \exp \left\{ -\frac{1}{2} \left(\frac{\mu^2}{\tau^2} + \frac{x^2}{\sigma^2} \right) + \frac{1}{2} \left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2} \right)^2 \frac{\tau^2\sigma^2}{\tau^2 + \sigma^2} \right\} \sqrt{\frac{\tau^2\sigma^2}{\tau^2 + \sigma^2}} \sqrt{2\pi} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \exp \left\{ \frac{1}{2} \left[\frac{\tau^2\sigma^2}{\tau^2 + \sigma^2} \left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2} \right)^2 - \frac{\tau^2 + \sigma^2}{\tau^2 + \sigma^2} \left(\frac{\mu^2}{\tau^2} + \frac{x^2}{\sigma^2} \right) \right] \right\} \\
&= \exp \left\{ \frac{1}{2(\tau^2 + \sigma^2)} \left[\tau^2\sigma^2 \left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2} \right)^2 - (\tau^2 + \sigma^2) \left(\frac{\mu^2}{\tau^2} + \frac{x^2}{\sigma^2} \right) \right] \right\} \\
&\quad \times \frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \\
&= \exp \left\{ \frac{1}{2(\tau^2 + \sigma^2)} \left[\tau^2\sigma^2 \left(\frac{\mu^2}{\tau^4} + \frac{2\mu x}{\tau^2\sigma^2} + \frac{x^2}{\sigma^4} \right) - \left(\frac{\tau^2\mu^2}{\tau^2} + \frac{\tau^2x^2}{\sigma^2} + \frac{\mu^2\sigma^2}{\tau^2} + \frac{\sigma^2x^2}{\sigma^2} \right) \right] \right\} \\
&\quad \times \frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\tau^2 + \sigma^2}} \exp \left\{ -\frac{1}{2(\tau^2 + \sigma^2)} (\mu - x)^2 \right\}
\end{aligned}$$

es decir $X \sim N(\mu, \tau^2 + \sigma^2)$.

1.4.1. Familias conjugadas

Para realizar predicciones o actualizar nuestro modelo posterior a la toma de una muestra, realizamos inferencia sobre la distribución posterior del parámetro en cuestión, al calcular de manera exacta dicha distribución nos encontramos con integrales de la forma:

$$\int_{\Theta} g(\theta) f(x|\theta) \pi(\theta) d\theta$$

que en muchas ocasiones no pueden calcularse de manera directa y es necesario emplear métodos de aproximación, como son la integración numérica o los métodos de Monte Carlo. Nuestro análisis se centrará en una familia particular de distribuciones, con la cual el cálculo de la distribución posterior no tiene mayor complicación.

Definición 10. Una familia \mathcal{F} de distribuciones de probabilidad para Θ es llamada conjugada (o cerrada bajo muestreo) para una función de verosimilitud $f(x|\theta)$ si, para toda $\pi \in \mathcal{F}$, la distribución posterior $\pi(\theta|x)$ también pertenece a \mathcal{F} . Es decir, si

$$\pi \in \mathcal{F}$$

entonces

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} \in \mathcal{F}.$$

Ejemplo 1.4.2. Consideremos como distribución inicial una Normal con parámetros conocidos μ y τ^2 . Comprobemos que esta familia es conjugada para la función de verosimilitud

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\} \quad \mathbb{I}_{\mathbb{R}}(x)$$

donde σ^2 es conocida y θ es desconocida.

$$\begin{aligned} \pi(\theta|x) &= \frac{f(X|\theta)\pi(\theta)}{\int f(X|\theta)\pi(\theta)d\theta} \\ &\propto f(X|\theta)\pi(\theta) \\ &= (2\pi\sigma\tau)^{-1} \exp\left\{-\frac{1}{2}\left[\frac{(\theta-\mu)^2}{\tau^2} + \frac{(x-\theta)^2}{\sigma^2}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\frac{\theta^2 - 2\mu\theta + \mu^2}{\tau^2} + \frac{x^2 - 2x\theta + \theta^2}{\sigma^2}\right]\right\} \end{aligned}$$

$$\begin{aligned}
&= \exp \left\{ -\frac{1}{2} \left[\frac{\sigma^2\theta^2 - 2\sigma^2\mu\theta + \sigma^2\mu^2 + \tau^2x^2 - 2\tau^2x\theta + \tau^2\theta^2}{\tau^2\sigma^2} \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[\frac{\theta^2(\sigma^2 + \tau^2) - 2\theta(\sigma^2\mu + \tau^2x)}{\tau^2\sigma^2} \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[\frac{\theta^2 - 2\theta((\sigma^2\mu + \tau^2x)/(\sigma^2 + \tau^2)) + ((\sigma^2\mu + \tau^2x)/(\sigma^2 + \tau^2))^2}{\tau^2\sigma^2/(\sigma^2 + \tau^2)} \right] \right\} \\
&= \exp \left\{ -\frac{1}{2} \left[\frac{\left(\theta - \frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2} \right)^2}{\tau^2\sigma^2/(\sigma^2 + \tau^2)} \right] \right\} \\
&\propto N \left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\tau^2\sigma^2}{\sigma^2 + \tau^2} \right).
\end{aligned}$$

Es decir, la distribución $N(\mu, \tau^2)$ es conjugada para la función de verosimilitud $N(\theta, \sigma^2)$. Y la distribución posterior resulta ser $Normal(\mu', \tau'^2)$ donde

$$\mu' = \frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2} \quad y \quad \tau'^2 = \frac{\tau^2\sigma^2}{\sigma^2 + \tau^2}.$$

Elegir una distribución inicial, tal que el modelo resulte conjugado, puede justificarse con la siguiente idea. Si suponemos una distribución inicial $\pi(\cdot)$ del parámetro Θ , al obtener una muestra y calcular la distribución posterior, esperamos que la estructura del parámetro no se modifique por completo. Con un modelo conjugado, la estructura del parámetro posterior a obtener una muestra, conserva la misma ley de probabilidad con alguna modificación en los parámetros.

1.4.2. Teoría de decisión

Cuando debemos realizar una acción, surge la interrogante de cuál es la mejor. En este tipo de problemas es necesario un proceso lógico de decisión que nos permita evaluar las consecuencias de la misma.

Un problema de decisión Bayesiano está compuesto por los siguientes elementos:

- Distribución inicial de los parámetros, $\pi(\theta)$ $\theta \in \Omega$.
- Distribución de muestreo o modelo, $f(\cdot|\theta)$ con soporte en \mathbb{X} .
- Conjunto de acciones potenciales del cual tendremos que escoger, \mathcal{A} . ($a \in \mathcal{A}$ es una acción).
- Conjunto de posibles decisiones, $d \in \mathcal{D}$; $d: \mathbb{X} \rightarrow \mathcal{A}$.
- Función de pérdida, $l(\theta, d): \Omega \times \mathcal{D} \rightarrow \mathbb{R}^+$

En el contexto estadístico, las acciones y decisiones son consideradas iguales ($d(x) = a$).

La función de pérdida puede ser interpretada como el costo que implica elegir la acción a cuando el parámetro toma el valor θ . En el contexto de estimación puntual es el costo por utilizar d para estimar θ , en este caso, el proceso de decisión d toma la interpretación de estimador de θ .

Definición 11 (Riesgo frecuentista). *El riesgo frecuentista es la media de la función de pérdida dado un valor de θ , es decir,*

$$R(\theta, d) = \mathbb{E}_\theta[l(\theta, d(x))] = \int_{\mathbb{X}} l(\theta, d(x))f(x|\theta)dx$$

donde $d(x)$ denota la regla de decisión para $x \sim f(x|\theta)$.

Con esta definición no podemos comparar qué decisión implica un riesgo mayor, ya que depende del valor que tome θ . Alternativamente podemos considerar la siguiente definición.

Definición 12 (Riesgo integrado). *El riesgo integrado se define como*

$$r(\pi, d) = \mathbb{E}_\pi[R(\theta, d)] = \int_{\Omega} R(\theta, d)\pi(\theta)d\theta.$$

Esta nueva definición asigna un número real a cada decisión, lo que nos permitirá comparar acciones (decisiones).

El enfoque bayesiano de teoría de decisión se basa en la **pérdida posterior esperada** definida de la siguiente manera

$$\rho(\pi, d|x) = \mathbb{E}_\pi[l(\theta, d)|x] = \int_{\Omega} l(\theta, d)\pi(\theta|x)d\theta.$$

Podemos observar que es una función de la muestra. Tenemos entonces dos posibles problemas:

- * Dado $X = x$, elegir la acción con menor riesgo, es decir, escoger d que minimice la pérdida posterior esperada, $\rho(\pi, d|x)$, ó
- * En la etapa de planeación, elegir un $d(X)$ que minimice $r(\pi, d)$

Mediante el siguiente resultado, observemos que ambos problemas tienen la misma solución

$$\begin{aligned}
 r(\pi, d) &= \int_{\Omega} R(\theta, d)\pi(\theta)d\theta \\
 &= \int_{\Omega} \left[\int_{\mathbb{X}} l(\theta, d)f(x|\theta)dx \right] \pi(\theta)d\theta \\
 &= \int_{\mathbb{X}} \int_{\Omega} l(\theta, d)f(x|\theta)\pi(\theta)d\theta dx \\
 &= \int_{\mathbb{X}} \left[\int_{\Omega} l(\theta, d)\pi(\theta|x)d\theta \right] f(x)dx \\
 &= \int_{\mathbb{X}} \rho(\pi, d|x)f(x)dx.
 \end{aligned}$$

Así que un estimador que minimice $r(\pi, d)$ se puede obtener seleccionando, para cada $x \in \mathbb{X}$, el valor que minimice la pérdida posterior esperada $\rho(\pi, d|x)$, ya que

$$r(\pi, d) = \int_{\mathbb{X}} \rho(\pi, d|x)f(x)dx.$$

Observación 1.4.1. *Dada una función de pérdida existe una función de utilidad asociada*

$$l(\theta, d) = -U(\theta, d)$$

*Si en lugar de trabajar con una función de pérdida trabajamos con una función de utilidad, nuestro objetivo será maximizar el **beneficio posterior esperado**.*

$$\int_{\mathbb{X}} \int_{\Omega} U(\theta, d)\pi(\theta|x)d\theta f(x)dx.$$

Capítulo 2

Un enfoque bayesiano para determinar el número de usuarios subsecuentes de un nuevo tratamiento médico

En el área de la salud es usual encontrarse con situaciones en las cuales, por diversas razones, debe considerarse la opción de implementar un nuevo tratamiento. Antes de tomar esta decisión es necesario evaluar su eficiencia y verificar que exista una diferencia significativa en comparación con el tratamiento existente.

En este capítulo describimos un modelo para determinar el número de usuarios subsecuentes de un nuevo tratamiento médico, utilizando un enfoque bayesiano. Este modelo está desarrollado en el artículo de Hamid Pezeshk y John Gittins titulado 'Bayesian approach to determine the number of subsequent users of a new treatment'[11].

Para ello requerimos realizar una prueba clínica que nos proporcione información respecto al beneficio que se obtendrá si se decide implementar el nuevo tratamiento.

Una cuestión importante en la planeación de una prueba para un nuevo tratamiento es el tamaño de la misma, es decir, cuántos usuarios tanto del tratamiento existente como del nuevo serán observados. Deseamos que el tamaño óptimo de la prueba implique el menor costo posible y a la vez podamos obtener la mayor cantidad de información.

En este trabajo consideramos la teoría de decisión, basada en una función de utilidad, para determinar el tamaño de una prueba en la que el número de usuarios subsecuentes del nuevo tratamiento bajo investigación, y por lo tanto el beneficio total resultante de la prueba, depende de la evidencia proporcionada por la misma.

La información de cada individuo que la prueba nos proporcione la llamaremos resultado clínico. El modelo considera que cada resultado es independiente del obtenido por cualquier otro individuo y del tratamiento que se esté utilizando.

Denotamos con X_i al resultado clínico, en una escala apropiada, del paciente i utilizando el nuevo tratamiento y con Y_j el resultado del paciente j utilizando el tratamiento actual. Suponemos que

$$X_i \sim N(\theta + \delta, \eta^2) \quad \text{y} \quad Y_j \sim N(\theta, \eta^2)$$

y que todos los resultados son independientes.

Sabemos que

$$\bar{X}_n = \frac{(\sum_{i=1}^n X_i)}{n} \sim N(\theta + \delta, \eta^2/n) \quad \text{y} \quad \bar{Y}_m = \frac{(\sum_{i=1}^m Y_i)}{m} \sim N(\theta, \eta^2/m).$$

Suponiendo que los dos grupos tienen el mismo tamaño de muestra, $n = m$ definimos $\bar{Z}_n = \bar{X}_n - \bar{Y}_n$ de esta forma se tiene que $\bar{Z}_n \sim N(\delta, 2\eta^2/n)$, donde $\sigma^2 = 2\eta^2$ se supone conocida por experiencia previa y δ es desconocida.

De acuerdo con nuestra hipótesis de Normalidad con varianza conocida, $\bar{Z}_n = \bar{X}_n - \bar{Y}_n$ es una estadística suficiente para el parámetro δ .

Proponemos como distribución inicial de δ una $N(\mu, \tau^2)$, con μ y τ^2 conocidas. La elección de esta densidad a priori se hace con la finalidad de obtener un modelo conjugado. Nuestra pretensión es inferir sobre los resultados posteriores a la obtención de una muestra y al conocer la distribución posterior de los parámetros, este proceso se simplifica considerablemente.

2.1. Número de usuarios subsecuentes de un nuevo tratamiento

El objetivo principal de este análisis es calcular la función de distribución del número de usuarios subsecuentes de un nuevo tratamiento, el cual suponemos función de \bar{Z}_n y lo denotamos como $m(\bar{Z}_n)$ o simplemente como m .

Para que usuarios potenciales acepten un nuevo tratamiento, es necesario que al comparar el desarrollo de éste con el del tratamiento en curso, exista una mejora significativa.¹ En el modelo esta evidencia se ve reflejada en δ , que es la diferencia de medias entre los resultados clínicos de ambos tratamientos.

Otro aspecto que los usuarios consideran importante es el costo que tendrá el nuevo tratamiento, y cómo se compara con el costo del tratamiento en curso, debemos entonces tomar en cuenta dicho aspecto. Sea

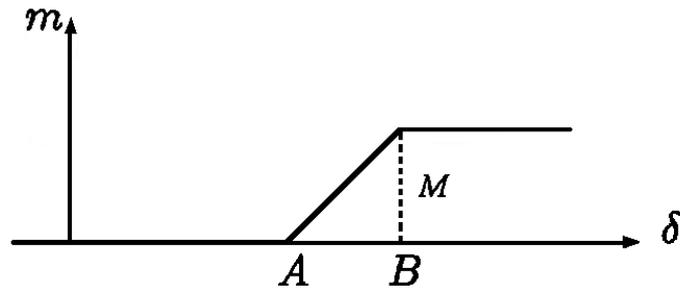
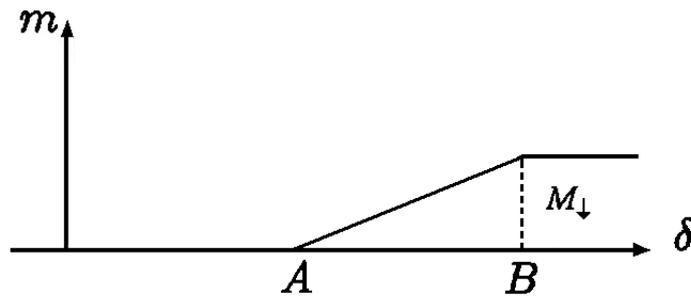
$$\Delta C = \text{costo esperado del nuevo tratamiento} - \text{costo esperado del tratamiento en curso.}$$

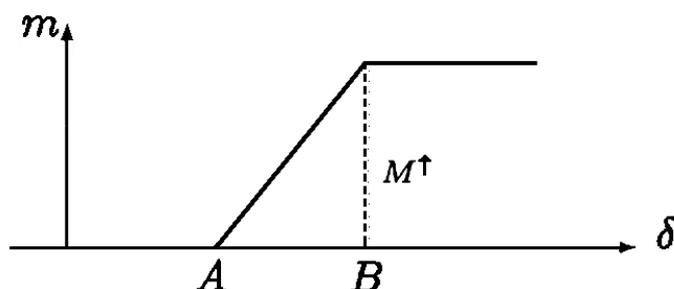
Como primer caso, supongamos $\Delta C = 0$. Por simplicidad, la relación que consideraremos entre m el número de usuarios subsecuentes y δ , será la que se muestra en la figura 2.1. Donde M es el número total esperado de usuarios cuando $\Delta C = 0$ y existe una diferencia significativa entre el desarrollo de ambos tratamientos. La figura 2.1 implica que si la magnitud de mejoría es menor que un cierto valor A , entonces nadie cambia al nuevo tratamiento; si está ente A y B , algunos pacientes cambian al nuevo tratamiento; y si es mayor que B , todos cambian al nuevo tratamiento.

Ahora suponemos $\Delta C > 0$. La pendiente cambia como se muestra en la figura 2.2. Para algunas personas un incremento en el costo no implica diferencia alguna en su elección, así que A no cambia. Para algunos otros el costo extra es prohibitivo, entonces el número esperado de usuarios subsecuentes M decrece. Y como algunos consideran que el costo extra significa que la mejoría debe ser aparentemente mayor que cuando $\Delta C = 0$, entonces B aumenta.

Finalmente, suponemos que $\Delta C < 0$. La pendiente se muestra en la figura 2.3. Comparado con el caso $\Delta C = 0$, observamos que B no cambia, M

¹How large should a clinical trial be? [12].

Figura 2.1: $\Delta C = 0$.Figura 2.2: $\Delta C > 0$.

Figura 2.3: $\Delta C < 0$.

aumenta y porque ahora algunos pacientes, monetariamente, tendrán acceso al nuevo tratamiento, A decrece (A podría ser negativa).

Con lo anterior estamos considerando que la magnitud de la mejora resultante al cambiar al nuevo tratamiento es conocida. En el caso que nosotros manejamos no contamos con esta información, así que la relación entre m y \bar{Z}_n estará dada en términos de la esperanza y la varianza de la distribución posterior de δ . Esto corresponde a suponer que los individuos cambian, por voluntad, al nuevo tratamiento en virtud de que la probabilidad de que la diferencia entre los dos tratamientos exceda un límite personal, sea mayor o igual que cierto valor mínimo.

El número de usuarios subsecuentes del nuevo tratamiento no será grande, a menos que este tratamiento sea, en el sentido estadístico y clínico, significativamente mejor que el existente.

Finalmente, suponemos que el número de usuarios subsecuentes del nuevo tratamiento depende de $\mu'(\bar{z}_n)$ y τ' (media y desviación estándar de la distribución posterior de δ) como se muestra en la figura 2.4.

En términos algebraicos, la función es

$$m(\bar{z}_n) = \begin{cases} 0 & \mu'(\bar{z}_n) < A + 1.5\tau' \\ \frac{M}{B-A}[\mu'(\bar{z}_n) - A - 1.5\tau'] & A + 1.5\tau' < \mu'(\bar{z}_n) < B + 1.5\tau' \\ M & B + 1.5\tau' < \mu'(\bar{z}_n) \end{cases} \quad (2.1)$$

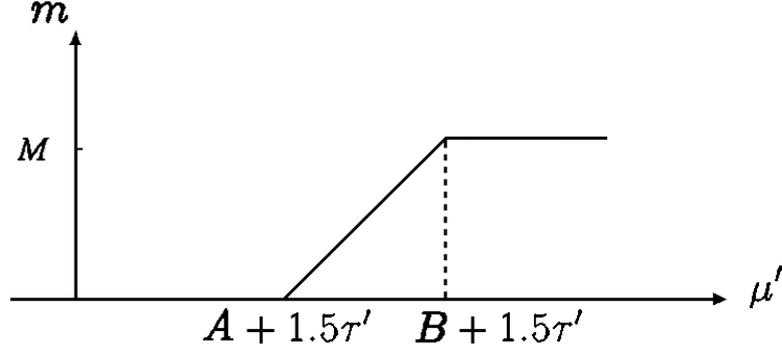


Figura 2.4: Número de usuarios subsecuentes.

Esta función nos indica que si la magnitud de la mejora esperada al cambiar al nuevo tratamiento, dada la muestra, es inferior a $A + 1.5\tau'$, entonces nadie cambia al nuevo tratamiento; si se encuentra entre $A + 1.5\tau'$ y $B + 1.5\tau'$, algunos pacientes cambian y si es superior a $B + 1.5\tau'$, entonces todos los pacientes cambian al nuevo tratamiento.

Es decir, que cada individuo tiene un umbral de diferencia a y cambiará al nuevo tratamiento, siempre que la probabilidad de que la diferencia entre los dos tratamientos supere ese umbral sea de al menos 0.93.

Lo anterior está basado en que la probabilidad de que una variable aleatoria con distribución Normal estándar sea menor que 1.5 es 0.93, entonces la probabilidad de que la diferencia entre ambos tratamientos dada la muestra sea mayor que $\mu' - 1.5\tau'$ es al menos 0.93.

Observemos que

$$\mathbb{P}[N(0, 1) > a] = 0.93$$

implica $a = -1.5$. Así que,

$$\begin{aligned} 0.93 &= \mathbb{P}\left[\frac{\delta - \mu'}{\tau'} > -1.5 \mid \underline{X}\right] \\ &= \mathbb{P}[\delta - \mu' > -1.5\tau' \mid \underline{X}] \\ &= \mathbb{P}[(\delta > \mu' - 1.5\tau' \mid \underline{X})] \end{aligned}$$

e inicialmente consideramos A cota inferior para δ . Si $A < \mu' - 1.5\tau'$,

entonces

$$\mathbb{P}[A < \delta | \underline{X}] \geq 0.93.$$

Bajo el mismo argumento si $\mu' - 1.5\tau' < B$, entonces

$$\mathbb{P}[\delta > B | \underline{X}] \leq 0.93,$$

o equivalentemente

$$\mathbb{P}[\delta \leq B | \underline{X}] \geq 0.07.$$

2.2. Tamaño óptimo de la muestra para una prueba clínica

Realizar pruebas clínicas para determinar si un tratamiento cumple con los requerimientos mínimos para ser implementado genera un gasto. Suena lógico pensar que entre mayor sea el tamaño de la muestra para dicha prueba mayor será el costo de la misma, sin embargo, obtendríamos más información que con una muestra más pequeña.

En esta sección determinamos el tamaño de la muestra para la prueba que nos proporciona la mayor cantidad de información al menor costo. La herramienta que empleamos en este proceso es la teoría de decisión.

Con n pacientes en cada tratamiento, la función objetivo (o beneficio neto esperado) estará definida por

$$r(n) = \text{beneficio total esperado del cambio de pacientes al nuevo tratamiento} - \text{costo de la prueba.}$$

Para modelar el beneficio por paciente se puede considerar una función lineal de δ . Nosotros nos restringimos al caso donde el beneficio por paciente es una constante b independiente de δ . De esta forma el beneficio tiene una relación con el número de pacientes utilizando el nuevo tratamiento, pero no con la evolución o desarrollo del mismo.

Sea $cn + d$ el costo por realizar una prueba con n pacientes en cada tratamiento si $n > 0$ y 0 si $n = 0$. De esta manera d es un costo inicial, independiente de n . Podemos proceder a calcular el óptimo de la prueba n^*

considerando $d = 0$. La prueba puede ser realizada sólo si $r(n^*) > 0$, donde $r(n^*)$ ahora incluye el costo inicial. Para simplificar nuestra discusión, vamos a suponer a partir de ahora que $d = 0$. La función objetivo puede ser escrita como

$$r(n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m(\bar{z}_n) b \Pi^n(\delta | \bar{z}_n) d\delta f(\bar{z}_n) d\bar{z}_n - cn \quad (2.2)$$

donde $\Pi^n(\delta | \bar{z}_n)$ es la distribución posterior del parámetro desconocido δ dado \bar{z}_n . En este caso es $N(\mu'(\bar{z}_n), \tau'^2)$, donde

$$\mu'(\bar{z}_n) = \frac{\sigma^2 \mu + n\tau^2 \bar{z}_n}{\sigma^2 + n\tau^2} \quad \text{y} \quad \tau'^2 = \frac{\sigma^2 \tau^2}{\sigma^2 + n\tau^2}.$$

Resolviendo la integral con respecto a δ en la ecuación (2.2) tenemos que

$$r(n) = \int_{-\infty}^{\infty} m(\bar{z}_n) b f(\bar{z}_n) d\bar{z}_n - cn$$

y $f(\bar{z}_n)$ es la densidad a priori de \bar{Z}_n .

Sabemos que $\bar{Z}_n \sim N(\mu, \tau^2 + \sigma^2/n)$ de acuerdo a lo expuesto en los preliminares.

Siguiendo el desarrollo del artículo de Gittins y Pezeshk[12] encontramos el tamaño óptimo de la prueba que maximiza el beneficio total esperado.

Como su nombre lo indica, el beneficio total esperado está determinado para una muestra de tamaño n , pero conviene expresar el problema en términos de variables escaladas, pues de esta manera, obtendríamos el beneficio esperado de manera individual $R(n) = r(n)/Mb$ y el costo $C = c/Mb$ por realizar la prueba a cada individuo.

$$\begin{aligned}
R(n) &= \frac{r(n)}{Mb} \\
&= \int_{-\infty}^{\infty} \frac{m}{M} f(\bar{z}_n) d\bar{z}_n - \frac{cn}{Mb} \\
&= 0 + \int_{H_1(A,n)}^{H_2(B,n)} \frac{M}{(B-A)M} (\mu'(\bar{z}_n) - A - 1.5\tau') f(\bar{z}_n) d\bar{z}_n + \int_{H_2(B,n)}^{\infty} f(\bar{z}_n) d\bar{z}_n \\
&\quad - Cn \\
&= \int_{H_1(A,n)}^{H_2(B,n)} \frac{1}{B-A} (\mu'(\bar{z}_n) - A - 1.5\tau') f(\bar{z}_n) d\bar{z}_n + \int_{H_2(B,n)}^{\infty} f(\bar{z}_n) d\bar{z}_n - Cn
\end{aligned}$$

donde $H_1(A, n)$ es el valor de \bar{z}_n para el cual $A + 1.5\tau' = \mu'(\bar{z}_n)$ y $H_2(B, n)$ es valor de \bar{z}_n para el cual $B + 1.5\tau' = \mu'(\bar{z}_n)$, pero

$$\begin{aligned}
A + 1.5\tau' &= \mu'(\bar{z}_n) \\
\Leftrightarrow A + 1.5\tau' &= \frac{\sigma^2\mu + n\tau^2\bar{z}_n}{\sigma^2 + n\tau^2} \\
\Leftrightarrow \frac{(A + 1.5\tau')(\sigma^2 + n\tau^2) - \sigma^2\mu}{n\tau^2} &= \bar{z}_n
\end{aligned}$$

entonces,

$$\begin{aligned}
H_1(A, n) &= \frac{(A + 1.5\tau')(\sigma^2 + n\tau^2) - \sigma^2\mu}{n\tau^2} \\
H_2(B, n) &= \frac{(B + 1.5\tau')(\sigma^2 + n\tau^2) - \sigma^2\mu}{n\tau^2}
\end{aligned}$$

utilicemos el siguiente cambio de variable

$$u = \frac{\bar{z}_n - \mu}{\sqrt{\sigma^2/n + \tau^2}}$$

de esta forma $U \sim N(0, 1)$ y reescribimos $R(n)$ de la siguiente manera

$$\begin{aligned}
R(n) &= \frac{1}{B-A} \int_{h_1(A,n)}^{h_2(B,n)} \left(\frac{\sigma^2\mu + n\tau^2(u\sqrt{\tau^2 + \sigma^2/n} + \mu)}{\sigma^2 + n\tau^2} - A - 1.5\tau' \right) \\
&\quad \times \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du + \int_{h_2(B,n)}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du - Cn.
\end{aligned}$$

Observemos que

$$\begin{aligned}
\frac{\sigma^2\mu + n\tau^2(u\sqrt{\tau^2 + \sigma^2/n} + \mu)}{\sigma^2 + n\tau^2} &= \frac{\sigma^2\mu + n\tau^2u\sqrt{\tau^2 + \sigma^2/n} + n\tau^2\mu}{\sigma^2 + n\tau^2} \\
&= \frac{\mu(\sigma^2 + n\tau^2) + n\tau^2u\sqrt{\tau^2 + \sigma^2/n}}{\sigma^2 + n\tau^2} \\
&= \mu + \frac{n\tau^2u\sqrt{\tau^2 + \sigma^2/n}}{\sigma^2 + n\tau^2} \\
&= \mu + \frac{\sqrt{n}\tau^2u\sqrt{n\tau^2 + \sigma^2}}{\sigma^2 + n\tau^2} \\
&= \mu + un^{1/2}\tau\sqrt{\frac{\tau^2}{\sigma^2 + n\tau^2}} \\
&= \mu + n^{1/2}\tau\left(\frac{\tau^2}{\sigma^2 + n\tau^2}\right)^{1/2}u \\
&= \mu + n^{1/2}\tau\left(\frac{\sigma^2 + \tau^2n}{\tau^2}\right)^{-1/2}u \\
&= \mu + n^{1/2}\tau\left(\frac{\sigma^2}{\tau^2} + n\right)^{-1/2}u.
\end{aligned}$$

Además

$$\begin{aligned}
\tau' &= \sqrt{\frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}} \\
&= \sigma\left(\frac{\sigma^2 + n\tau^2}{\tau^2}\right)^{-1/2} \\
&= \sigma\left(\frac{\sigma^2}{\tau^2} + n\right)^{-1/2},
\end{aligned}$$

entonces,

$$\begin{aligned}
R(n) &= \frac{1}{B-A} \int_{h_1(A,n)}^{h_2(B,n)} \left(\mu + n^{1/2} \tau \left(\frac{\sigma^2}{\tau^2} + n \right)^{-1/2} u \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\
&\quad - \frac{A}{B-A} \int_{h_1(A,n)}^{h_2(B,n)} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\
&\quad - 1.5 \frac{1}{B-A} \sigma \left(\frac{\sigma^2}{\tau^2} + n \right)^{-1/2} \int_{h_1(A,n)}^{h_2(B,n)} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\
&\quad + \int_{h_2(B,n)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du - Cn
\end{aligned}$$

donde $h_1(A, n)$ es el valor de u para el cual $\bar{z}_n = H_1(A, n)$

$$\bar{z}_n = \frac{(A + 1.5\tau')(\sigma^2 + n\tau^2) - \sigma^2\mu}{n\tau^2}.$$

Lo anterior implica que

$$u = \left(\frac{(A + 1.5\tau')(\sigma^2 + n\tau^2) - \sigma^2\mu}{n\tau^2} - \mu \right) \frac{1}{\sqrt{\frac{\sigma^2}{n} + \tau^2}}.$$

Calculando la diferencia y factorizando obtenemos

$$\begin{aligned}
u &= \frac{(A + 1.5\tau')(\sigma^2 + n\tau^2) - \sigma^2\mu - n\tau^2\mu}{n\tau^2} \left(\frac{n}{\sigma^2 + n\tau^2} \right)^{1/2} \\
&= \frac{(A + 1.5\tau')(\sigma^2 + n\tau^2) - \mu(\sigma^2 + n\tau^2)}{n\tau^2} \left(\frac{n}{\sigma^2 + n\tau^2} \right)^{1/2} \\
&= \frac{(A + 1.5\tau')(\sigma^2 + n\tau^2)^{1/2} - \mu(\sigma^2 + n\tau^2)^{1/2}}{n^{1/2}\tau^2} \\
&= \frac{(A + 1.5\tau' - \mu) \left(\frac{\sigma^2 + n\tau^2}{\tau^2} \right)^{1/2}}{n^{1/2}\tau} \\
&= \frac{(A + 1.5\tau' - \mu) \left(\frac{\sigma^2}{\tau^2} + n \right)^{1/2}}{n^{1/2}\tau}.
\end{aligned}$$

Recordando que

$$\tau' = \sigma \left(\frac{\sigma^2 + n\tau^2}{\tau^2} \right)^{-1/2},$$

concluimos que

$$h_1(A, n) = \frac{[A + 1.5\sigma(\sigma^2/\tau^2 + n)^{1/2} - \mu](\sigma^2/\tau^2 + n)^{1/2}}{n^{1/2}\tau}$$

$$h_2(B, n) = \frac{[B + 1.5\sigma(\sigma^2/\tau^2 + n)^{1/2} - \mu](\sigma^2/\tau^2 + n)^{1/2}}{n^{1/2}\tau}.$$

2.3. Distribución del número de usuarios subsecuentes de un nuevo tratamiento

En esta sección, calculamos la función de distribución inicial del número de usuarios subsecuentes de un nuevo tratamiento. La motivación es que a partir de esta distribución podemos decidir, respecto a implementar o no el nuevo tratamiento. Por ejemplo, si la prueba está siendo realizada por una compañía farmacéutica, es de vital importancia conocer el número esperado de usuarios subsecuentes para determinar si se obtendrá el nivel esperado de ganancia.

Si la prueba fuera realizada por el sector de salud pública, poseer esta información les permitiría planificar la producción o distribución del tratamiento.

En general podremos conocer la demanda esperada, en base a ella decidir si es conveniente implementar el nuevo tratamiento, y en caso de serlo, diseñar la estrategia óptima para hacerlo.

Ahora estamos listos para demostrar el resultado principal de este trabajo.

Teorema 2.3.1 (Resultado principal). *La función de distribución inicial del número m de usuarios subsecuentes de un nuevo tratamiento es*

$$F_m(y) = \begin{cases} 0 & y < 0 \\ \Phi \left(\left[\frac{y}{M\sigma}(B - A) + \frac{A}{\sigma} + 1.5(T^{-2} + n)^{-1/2} - D \right] \frac{(T^{-2} + n)^{1/2}}{T\sqrt{n}} \right) & 0 \leq y < M \\ 1 & M \leq y, \end{cases}$$

donde $T = \tau/\sigma$, $D = \mu/\sigma$ y $\Phi(x) = \int_{-\infty}^x (1/\sqrt{2\pi})e^{(-1/2)z^2} dz$.

Demostración. Observemos que $0 \leq y < M$ si y solo si,

$$A + 1.5\tau' < \mu'(\bar{z}_n) < B + 1.5\tau'$$

y el número de usuarios subsecuentes en este intervalo es igual a

$$\frac{M}{B - A}[\mu' - A - 1.5\tau'],$$

entonces podemos escribir

$$\begin{aligned} F_m(y) &= \mathbb{P}[m \leq y] \\ &= \mathbb{P}\left[\frac{M}{B - A}[\mu'(\bar{z}_n) - A - 1.5\tau'] \leq y\right] \\ &= \mathbb{P}\left[\mu'(\bar{z}_n) - A - 1.5\tau' \leq \frac{y}{M}(B - A)\right] \\ &= \mathbb{P}\left[\mu'(\bar{z}_n) \leq \frac{y}{M}(B - A) + A + 1.5\tau'\right] \end{aligned}$$

$$\begin{aligned}
&= \mathbb{P} \left[\frac{\sigma^2 \mu + n\tau^2 \bar{z}_n}{\sigma^2 + n\tau^2} \leq \frac{y}{M}(B-A) + A + 1.5\tau' \right] \\
&= \mathbb{P} \left[\bar{z}_n \leq \frac{(\sigma^2 + n\tau^2) \left(\frac{y}{M}(B-A) + A + 1.5\tau' \right) - \sigma^2 \mu}{n\tau^2} \right] \\
&= \mathbb{P} \left[\frac{\bar{z}_n - \mu}{\sqrt{\tau^2 + \frac{\sigma^2}{n}}} \leq \left\{ \frac{(\sigma^2 + n\tau^2) \left(\frac{y}{M}(B-A) + A + 1.5\tau' \right) - \sigma^2 \mu}{n\tau^2} - \mu \right\} \frac{1}{\sqrt{\tau^2 + \frac{\sigma^2}{n}}} \right].
\end{aligned}$$

Observemos que

$$\frac{\bar{z}_n - \mu}{\sqrt{\tau^2 + \frac{\sigma^2}{n}}}$$

es una variable aleatoria con distribución Normal de media cero y varianza uno. Considerando el conjunto al que le estamos calculando su probabilidad concentrémonos en el lado derecho de la desigualdad

$$\begin{aligned}
K &= \left[\frac{(\sigma^2 + n\tau^2) \left(\frac{y}{M}(B-A) + A + 1.5\tau' \right) - \sigma^2 \mu}{n\tau^2} - \mu \right] \frac{1}{\sqrt{\tau^2 + \frac{\sigma^2}{n}}} \\
&= \frac{(\sigma^2 + n\tau^2) \left(\frac{y}{M}(B-A) + A + 1.5\tau' \right) - \sigma^2 \mu}{n\tau^2 \sqrt{\frac{n\tau^2 + \sigma^2}{n}}} - \frac{\mu \sqrt{n}}{\sqrt{n\tau^2 + \sigma^2}} \\
&= \frac{(\sigma^2 + n\tau^2)}{n\tau^2} \sqrt{\frac{n}{\sigma^2 + n\tau^2}} \left(\frac{y}{M}(B-A) + A + 1.5\tau' \right) \\
&\quad - \mu \left(\frac{\sigma^2}{n\tau^2} \sqrt{\frac{n}{\sigma^2 + n\tau^2}} + \sqrt{\frac{n}{\sigma^2 + n\tau^2}} \right)
\end{aligned}$$

$$\begin{aligned}
&= \sqrt{\frac{\sigma^2 + n\tau^2}{n}} \frac{1}{\tau^2} \left(\frac{y}{M}(B - A) + A + 1.5\tau' \right) - \mu \sqrt{\frac{n}{\sigma^2 + n\tau^2}} \left(\frac{\sigma^2}{n\tau^2} + 1 \right) \\
&= \sqrt{\frac{\sigma^2 + n\tau^2}{n}} \frac{1}{\tau^2} \left(\frac{y}{M}(B - A) + A + 1.5\tau' \right) - \mu \sqrt{\frac{n}{\sigma^2 + n\tau^2}} \left(\frac{\sigma^2 + n\tau^2}{n\tau^2} \right) \\
&= \sqrt{\frac{\sigma^2 + n\tau^2}{n}} \frac{1}{\tau^2} \left(\frac{y}{M}(B - A) + A + 1.5\tau' \right) - \mu \sqrt{\frac{\sigma^2 + n\tau^2}{n}} \frac{1}{\tau^2} \\
&= \left(\frac{y}{M}(B - A) + A + 1.5\tau' - \mu \right) \sqrt{\frac{\sigma^2 + n\tau^2}{n}} \frac{1}{\tau^2} \\
&= \left(\frac{y}{M}(B - A) + A + 1.5\tau' - \mu \right) \frac{\sqrt{\sigma^2 + n\tau^2}}{\tau} \frac{1}{\tau\sqrt{n}} \\
&= \left(\frac{y}{M}(B - A) + A + 1.5\tau' - \mu \right) \frac{\left(\frac{\sigma^2 + n\tau^2}{\tau^2} \right)^{1/2}}{\tau\sqrt{n}}
\end{aligned}$$

como $\left(\frac{\tau^2\sigma^2}{\sigma^2 + n\tau^2} \right)^{1/2}$

$$\begin{aligned}
K &= \frac{1}{\sigma} \left[\frac{y}{M}(B - A) + A + 1.5 \left(\frac{\tau^2\sigma^2}{\sigma^2 + n\tau^2} \right)^{1/2} - \mu \right] \frac{\sigma \left(\frac{\sigma^2 + n\tau^2}{\tau^2} \right)^{1/2}}{\tau\sqrt{n}} \\
&= \left[\frac{y}{M\sigma}(B - A) + \frac{A}{\sigma} + 1.5 \left(\frac{\sigma^2 + n\tau^2}{\tau^2} \right)^{-1/2} - \frac{\mu}{\sigma} \right] \frac{\sigma \left(\frac{\sigma^2 + n\tau^2}{\tau^2} \right)^{1/2}}{\tau\sqrt{n}} \\
&= \left[\frac{y}{M\sigma}(B - A) + \frac{A}{\sigma} + 1.5 \left(\frac{\sigma^2}{\tau^2} + n \right)^{-1/2} - \frac{\mu}{\sigma} \right] \frac{\left(\frac{\sigma^2}{\tau^2} + n \right)^{1/2}}{\frac{\tau}{\sigma}\sqrt{n}} \\
&= \left[\frac{y}{M\sigma}(B - A) + \frac{A}{\sigma} + 1.5 (T^{-2} + n)^{-1/2} - D \right] \frac{(T^{-2} + n)^{1/2}}{T\sqrt{n}}
\end{aligned}$$

donde $T = \tau/\sigma$, $D = \mu/\sigma$.

Así que $F_m(y)$ puede ser escrita como

$$\mathbb{P} \left[\frac{\bar{Z}_n - \mu}{\sqrt{\sigma^2 + n\tau^2/n}} \leq \frac{[(y/M\sigma)(B - A) + A/\sigma + 1.5(T^{-2} + n)^{-1/2} - D](T^{-2} + n)^{1/2}}{T\sqrt{n}} \right],$$

por lo tanto,

$$F_m(y) = \Phi \left(\left[\frac{y}{M\sigma}(B - A) + \frac{A}{\sigma} + 1.5(T^{-2} + n)^{-1/2} - D \right] \frac{(T^{-2} + n)^{1/2}}{T\sqrt{n}} \right)$$

para $0 \leq y < M$. □

Con base en la distribución anterior, calculemos la función de densidad. Para ello, observemos que

$$\begin{aligned} \mathbb{P}[m = 0] &= \mathbb{P}[\mu'(\bar{z}_n) < A + 1.5\tau'] \\ &= \mathbb{P} \left[\frac{\sigma^2\mu + n\tau^2\bar{z}_n}{\sigma^2 + n\tau^2} < A + 1.5\tau' \right] \\ &= \mathbb{P} \left[\frac{\bar{z}_n - \mu}{\sqrt{\tau^2 + \frac{\sigma^2}{n}}} < \left(\frac{(A + 1.5\tau')(\sigma^2 + n\tau^2) - \sigma^2\mu}{n\tau^2} - \mu \right) \frac{1}{\sqrt{\tau^2 + \frac{\sigma^2}{n}}} \right]. \end{aligned}$$

Por los cálculos antes revisados podemos concluir que

$$\begin{aligned} \mathbb{P}[m = 0] &= \mathbb{P} \left[\frac{\bar{Z}_n - \mu}{\sqrt{\sigma^2 + n\tau^2/n}} < \left[\frac{A}{\sigma} + 1.5(T^{-2} + n)^{-1/2} - D \right] \frac{(T^{-2} + n)^{1/2}}{T\sqrt{n}} \right] \\ &= \Phi(h_1(A, n)) \end{aligned}$$

donde

$$h_1(A, n) = \frac{[A/\sigma + 1.5(T^{-2} + n)^{-1/2} - D](T^{-2} + n)^{1/2}}{T\sqrt{n}}$$

y de manera similar

$$\mathbb{P}[m = M] = \mathbb{P}\left[\frac{\bar{Z}_n - \mu}{\sqrt{\sigma^2 + n\tau^2/n}} \geq h_2(B, n)\right] = 1 - \Phi(h_2(B, n))$$

donde

$$h_2(B, n) = \frac{[B/\sigma + 1.5(T^{-2} + n)^{-1/2} - D](T^{-2} + n)^{1/2}}{T\sqrt{n}}.$$

En resumen, obtenemos que la función de densidad a priori de m es de la forma

$$f_m(y) = \begin{cases} \mathbb{P}[m = y] = \Phi(h_1(A, n)) & y = 0 \\ \phi\left(\left[\frac{y}{M\sigma}(B - A) + \frac{A}{\sigma} + 1.5(T^{-2} + n)^{-1/2} - D\right] \frac{(T^{-2} + n)^{1/2}}{T\sqrt{n}}\right) \\ \quad \times \frac{B - A}{M\sigma T\sqrt{n}}(T^{-2} + n)^{1/2} & 0 < y < M \\ \mathbb{P}[m = y] = 1 - \Phi(h_2(B, n)) & y = M \end{cases}$$

$$\text{donde } \phi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}.$$

Capítulo 3

Prueba clínica

El objetivo de este capítulo es analizar los datos publicados en el “*Journal of the Neurological Sciences, 2005*” [13], con base en lo desarrollado en el capítulo anterior.

En este artículo se estudia la administración de interferon beta-1a en el control de esclerosis múltiple con recaída-remisión. Los tratamientos que consideramos son dos modalidades de administración.

	Tratamientos	
	Actual	Nuevo
Cantidad	30mcg	44mcg subcutáneo
Vía	intramuscular	subcutánea
Periodicidad	una vez a la semana	tres veces a la semana

De un total de 674 pacientes, 337 recibieron el tratamiento actual y 337 recibieron el tratamiento nuevo. El periodo considerado es lo que en el artículo denominan fase comparativa, con una duración de 24 semanas. Al final de dicho periodo la información obtenida fue la siguiente: 329 de los 337 pacientes que recibieron el tratamiento actual continuaron en el estudio, y 330 de 337 pacientes que recibieron el tratamiento nuevo continuaron con el estudio.

Los resultados clínicos obtenidos son considerados de la siguiente manera

* 1 si el paciente respondió positivamente al tratamiento, es decir, si continuó en la siguiente etapa.

* 0 si respondió negativamente al tratamiento, es decir, si no continuó en la siguiente etapa.

De esta manera

$$\sum_{i=1}^n X_i = 330 \quad \text{y} \quad \sum_{i=1}^n Y_i = 329.$$

Definimos los parámetros

$$p_1 = \mathbb{P}[\text{responder favorablemente}|\text{tratamiento nuevo}]$$

$$p_2 = \mathbb{P}[\text{responder favorablemente}|\text{tratamiento actual}].$$

El análisis realizado en el capítulo anterior supone que los resultados clínicos provienen de una muestra de variables aleatorias con función de distribución Normal, y considera la estadística \bar{Z}_n . En esta prueba clínica, los resultados obtenidos se distribuyen Bernoulli con cierto parámetro p . Pero no todo está perdido, definimos una nueva estadística T_n que contiene toda la información sobre la muestra, y demostramos que dicha estadística cumple

$$T_n \xrightarrow{d} N(\delta, \sigma^2/n)$$

donde δ es desconocida y σ^2 es conocida.

Enunciamos este argumento en el siguiente resultado que será demostrado posteriormente.

Lema 3.0.1. Sean $\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n$ colecciones de variables aleatorias tales que $X_i \sim \text{Bernoulli}(p_1)$ y $Y_i \sim \text{Bernoulli}(p_2)$. Sea

$$T_n = \ln \left(\frac{\frac{1}{n} \sum_{i=1}^n X_i}{1 - \frac{1}{n} \sum_{i=1}^n X_i} \right) - \ln \left(\frac{\frac{1}{n} \sum_{i=1}^n Y_i}{1 - \frac{1}{n} \sum_{i=1}^n Y_i} \right)$$

entonces $T_n \xrightarrow{d} N \left(\delta, \frac{\sigma^2}{n} \right)$.

Donde

$$\delta = \ln \left(\frac{p_1}{1-p_1} \right) - \ln \left(\frac{p_2}{1-p_2} \right) \quad \text{y} \quad \sigma^2 = 2.$$

Notemos que $X = \sum_{i=1}^n X_i \sim \text{Bin}(n, p_1)$. Por el Teorema central del límite, sabemos que

$$\frac{X - \mathbb{E}[X]}{\sqrt{\text{var}(X)}} \xrightarrow{d} N(0, 1)$$

o lo que es equivalente

$$\sqrt{n} \left(\frac{X}{n} - p_1 \right) \xrightarrow{d} N(0, p_1(1 - p_1)). \quad (3.1)$$

Observación 3.0.1. Sea $[a, b]$ un intervalo tal que $\bar{X}_n, p_1 \in [a, b]$, f una función continua en $[a, b]$ y diferenciable en (a, b) , con f' continua y distinta de cero en p . Si

$$\sqrt{n} (\bar{X}_n - p_1) \xrightarrow{d} N(0, p_1(1 - p_1))$$

entonces

$$\sqrt{n} (f(\bar{X}_n) - f(p_1)) \xrightarrow{d} N(0, p_1^2(1 - p_1)^2 [f'(p_1)]^2).$$

Demostración. Si $\bar{X}_n = p_1$, entonces $f(\bar{X}_n) - f(p_1)$ sería idéntica a la constante cero.

Supongamos $\bar{X}_n > p_1$, entonces f es continua en $[p_1, \bar{X}_n]$ y diferenciable en (p_1, \bar{X}_n) . Por el teorema del valor medio existe $\theta_n \in (p_1, \bar{X}_n)$ tal que

$$f'(\theta_n) = \frac{f(\bar{X}_n) - f(p_1)}{\bar{X}_n - p_1}$$

o lo que es equivalente

$$f(\bar{X}_n) - f(p_1) = f'(\theta_n)(\bar{X}_n - p_1) \quad (3.2)$$

multiplicando ambos lados de la igualdad por \sqrt{n}

$$\sqrt{n} [f(\bar{X}_n) - f(p_1)] = \sqrt{n} [f'(\theta_n)(\bar{X}_n - p_1)]. \quad (3.3)$$

Recordemos que si $S_n \xrightarrow{\mathbb{P}} c$ y g es una función continua, entonces $g(S_n) \xrightarrow{\mathbb{P}} g(c)$.¹

¹Proposición 1.2.2.

Si adicionalmente tenemos que $W_n \xrightarrow{d} W$, sabemos que $S_n W_n \xrightarrow{d} cW$.²

Por hipótesis $f'(p_1)$ es continua y por la ley fuerte de los grandes números podemos asegurar que

$$\bar{X}_n \xrightarrow{c.s.} \mathbb{E}[X_i] = p_1.$$

En particular

$$\theta_n \xrightarrow{\mathbb{P}} p_1,$$

entonces

$$f'(\theta_n) \xrightarrow{\mathbb{P}} f'(p_1).$$

Con lo anterior en mente y por la expresión (3.3), podemos concluir que

$$\sqrt{n} [f(\bar{X}_n) - f(p_1)] \xrightarrow{d} f'(p_1)N(0, p(1-p)),$$

es decir,

$$\sqrt{n} [f(\bar{X}_n) - f(p_1)] \xrightarrow{d} N(0, p(1-p)[f'(p_1)]^2).$$

Si $\bar{X}_n < p_1 \Rightarrow f$ es continua en $[\bar{X}_n, p_1]$ y diferenciable en (\bar{X}_n, p_1) , por el teorema del valor intermedio existe $\tilde{\theta}_n \in (\bar{X}_n, p_1)$ tal que

$$f'(\tilde{\theta}_n)(p_1 - \bar{X}_n) = f(p_1) - f(\bar{X}_n)$$

multiplicando por (-1) ambos lados de la igualdad

$$f'(\tilde{\theta}_n)(\bar{X}_n - p_1) = f(\bar{X}_n) - f(p_1)$$

equivalente a la expresión (4.2) y así concluye la prueba. \square

Antes de proceder a la demostración del Lema 3.0.1, hagamos las siguientes aclaraciones. Si

$$\sum_{i=1}^n X_i = 0$$

entonces podemos concluir que el tratamiento actual es muy malo y ningún paciente responde favorablemente, si

$$\sum_{i=1}^n X_i = n$$

²Lema 1.2.1.

el tratamiento sería muy bueno, ya que todos los pacientes responderían favorablemente.

Si alguna de estas dos situaciones se presentara, nuestro análisis no sería de mucho interés, es por esta razón que suponemos

$$1 \leq \sum_{i=1}^n X_i \leq n - 1.$$

De la misma forma no tendría sentido suponer que $p_1 = 0$ ó $p_1 = 1$.

Establecido lo anterior, la estadística T_n está bien definida.

Demostración del Lema 3.0.1. Observemos que la función $\ln(\cdot)$ es continua y diferenciable en $(0, \infty)$.

Ahora, si $f(x) = \ln(x/(1-x))$, entonces

$$f'(x) = \left(\frac{1-x}{x} \right) \left(\frac{1-x+x}{(1-x)^2} \right) = \frac{1}{x(1-x)}$$

la cual es distinta de cero y continua en $x \neq 0, 1$. Entonces se cumplen las hipótesis de la observación 3.0.1, y

$$\sqrt{n} \left[\ln \left(\frac{\bar{X}_n}{1-\bar{X}_n} \right) - \ln \left(\frac{p_1}{1-p_1} \right) \right] \xrightarrow{d} N(0, 1),$$

es decir,

$$\ln \left(\frac{\bar{X}_n}{1-\bar{X}_n} \right) \xrightarrow{d} N \left(\ln \left(\frac{p_1}{1-p_1} \right), \frac{1}{n} \right).$$

De manera análoga y estableciendo para $\sum_{i=1}^n Y_i$ y p_2 las mismas condiciones que para $\sum_{i=1}^n X_i$ y p_1 , obtenemos

$$\ln \left(\frac{\bar{Y}_n}{1-\bar{Y}_n} \right) \xrightarrow{d} N \left(\ln \left(\frac{p_2}{1-p_2} \right), \frac{1}{n} \right).$$

Así concluimos que

$$T_n \xrightarrow{d} N \left(\ln \left(\frac{p_1}{1-p_1} \right) - \ln \left(\frac{p_2}{1-p_2} \right), \frac{2}{n} \right).$$

□

Con base en el resultado obtenido sobre la estadística T_n , hacemos uso de las técnicas desarrolladas en el capítulo anterior para obtener el número esperado de usuarios subsecuentes del nuevo tratamiento.

Ya que carecemos de información respecto a la distribución inicial de la diferencia entre ambos tratamientos δ y, para fines ilustrativos, consideramos la media

$$\mu \approx 2.48490665$$

que es el logaritmo natural de la razón de probabilidades cuando la probabilidad de éxito es de 0.75 con el tratamiento actual y de 0.20 con placebo.

También pediremos que δ se encuentre en el intervalo $[0, 2\mu]$ con probabilidad 0.95, más aún

$$\mathbb{P}[\delta < 0] = \mathbb{P}[\delta > 2\mu] = 0.025.$$

Sabemos que

$$\mathbb{P} \left[\frac{\delta - \mu}{\tau} < -1.96 \right] = 0.025 \quad \text{y} \quad \mathbb{P} \left[\frac{\delta - \mu}{\tau} > 1.96 \right] = 0.025$$

de esta forma

$$-1.96\tau + \mu = 0 \quad \text{y} \quad 1.96\tau + \mu = 2\mu$$

lo cual implica que $\tau \approx \frac{\mu}{2} = 1.242453325$.

Suponemos $A = 0.8\mu = 1.98792532$ y $B = 1.2\mu = 2.98188798$. De esta manera la distribución inicial de δ es simétrica con media $\mu = (A + B)/2$.

En nuestro caso, no tenemos acceso al número de usuarios potenciales subsecuentes del nuevo tratamiento. Tampoco tenemos información sobre el costo o beneficio de la prueba así que consideraremos que el tamaño óptimo

de la prueba es justamente el número de pacientes que participaron en la investigación $n^* = 337$ y que el número de usuarios potenciales subsecuentes es ($M = 10000$).

De acuerdo a lo desarrollado en el capítulo anterior

$$T = \tau/\sigma \approx 0.8785472 \quad \text{y} \quad D = \mu/\sigma \approx 1.757094,$$

entonces la función de distribución del número de usuarios subsecuentes es la siguiente

$$F_m(y) = \begin{cases} 0 & y < 0 \\ \Phi([7.028377 \times 10^{-5}y - 0.2698653] 0.003371101) & 0 \leq y < 10000 \\ 1 & 10000 \leq y \end{cases}$$

De tal manera que la función de densidad está determinada por

$$f_m(y) = \begin{cases} \mathbb{P}[m = y] = 0.3791317 & y = 0 \\ \phi([7.028377 \times 10^{-5}y - 0.2698653] 0.003371101) 8.015363^{-5} & 0 < y < 10000 \\ \mathbb{P}[m = y] = 0.3107328 & y = 10000 \end{cases}$$

En base a la función de densidad calculamos el valor esperado de usuarios subsecuentes que es aproximadamente $\mathbb{E}[m] \approx 4639.15^3$.

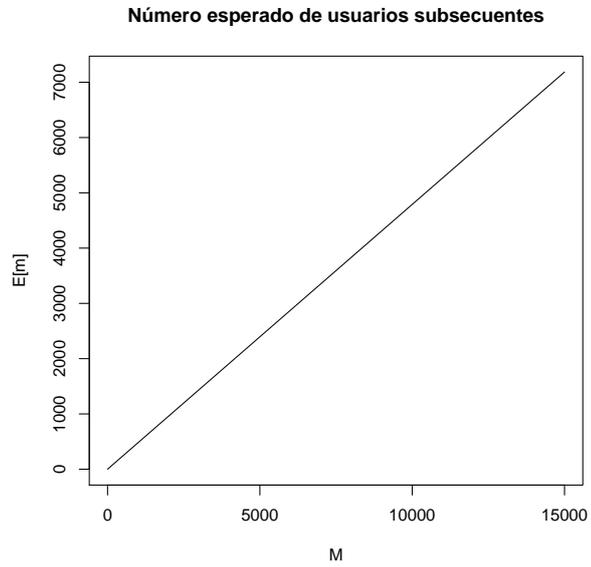
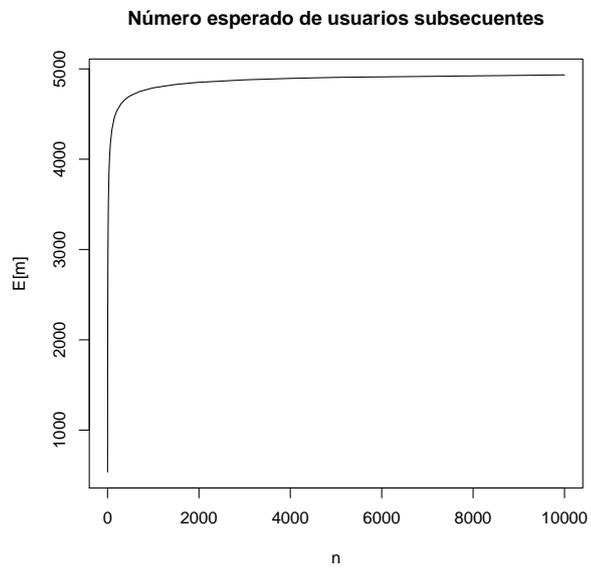
Para ejemplificar nuestra aplicación asumimos ciertos parámetros. Veamos qué sucede con el valor esperado de los usuarios subsecuentes cuando variamos estos parámetros.

Supongamos $n=1000$, si hacemos tender M a infinito, entonces $\mathbb{E}[m]$ tiende a infinito como se observa en la figura 3.1.

Si $M = 10000$ y hacemos tender n a infinito, entonces $\mathbb{E}[m] \rightarrow M/2$, figura 3.2.

En la figura 3.3 se describe el comportamiento de $\mathbb{E}[m]$ como función de μ .

³El cálculo de esta esperanza fue realizada con R, consultar apéndice ??.

Figura 3.1: $n = 1000$.Figura 3.2: $M = 10000$.

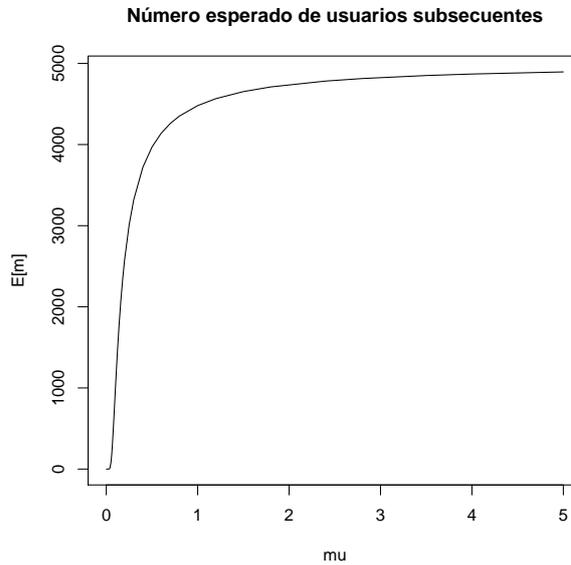


Figura 3.3: $n = 1000$, $M = 10000$.

Si fijásemos condiciones mínimas para considerar que el nuevo tratamiento es significativamente mejor que el actual, con los resultados obtenidos podríamos decidir implementar o no el nuevo tratamiento. De aquí la utilidad de conocer explícitamente la función de distribución del número de usuarios subsecuentes.

Consideramos la hipótesis σ^2 conocida, basados en el supuesto de que la institución que realizó la prueba poseía información sobre la respuesta de los usuarios del tratamiento en curso.

El caso σ^2 desconocida puede ser resuelto proponiendo distribuciones $\pi_2(\lambda^2)$, y $\pi_1(\delta|\lambda^2)$, donde $\lambda^2 = \sigma^2/n$ y considerando el modelo donde la distribución inicial $\pi(\delta)$ es descompuesta de la siguiente manera

$$\pi(\delta) = \int_{\Theta_1} \pi_1(\delta|\lambda^2)\pi_2(\lambda^2)d\lambda^2.$$

A este tipo de modelos bayesianos se les denomina jerárquicos y al parámetro λ^2 se le llama hiperparámetro de primer orden.

Esto es distinto a proponer una distribución inicial para $(\delta, \sigma^2/n)$ donde σ^2 tiene una influencia directa en el modelo. En nuestro caso sería más conve-

niente utilizar un modelo jerárquico que incluya la información de σ^2 , considerada subjetiva, a través de su relación con δ , que es la principal información requerida, puesto que mide la diferencia entre un tratamiento y el otro.

En general tenemos la siguiente definición.

Definición 13. ⁴Un modelo jerárquico de Bayes es un modelo estadístico bayesiano, $(f(x|\theta), \pi(\theta))$, donde la distribución inicial $\pi(\theta)$ es descompuesta en distribuciones condicionales

$$\pi_1(\theta|\theta_1), \pi_1(\theta_1|\theta_2), \dots, \pi_1(\theta_{n-1}|\theta_n)$$

y una distribución marginal $\pi_{n+1}(\theta_n)$, tal que

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_n} \pi_1(\theta|\theta_1), \pi_1(\theta_1|\theta_2), \dots, \pi_1(\theta_{n-1}|\theta_n) d\theta_1 \dots d\theta_n.$$

Los parámetros θ_i son llamados hiperparámetros de orden i , $1 \leq i \leq n$.

Ejemplo 3.0.1. Consideremos el modelo jerárquico $(f(x|\delta), \pi(\delta))$, donde la distribución inicial $\pi(\delta)$ es descompuesta en la distribución condicional

$$\pi_1(\delta|\lambda^2) = \frac{1}{\sqrt{2\pi\tau\lambda^2}} \exp \left\{ -\frac{(\delta - \mu)^2}{2\tau\lambda^2} \right\}$$

y la distribución marginal

$$\pi_2(\lambda^2) = \frac{(\beta/2)^{v/2}}{\Gamma(v/2)} (\lambda^{-2})^{\frac{v}{2}+1} \exp \left\{ -\frac{\beta\lambda^{-2}}{2} \right\}$$

En este caso la distribución inicial $\pi(\delta)$ está determinada por

$$\begin{aligned} \pi(\delta) &= \int_0^\infty \frac{1}{\sqrt{2\pi\tau\lambda^2}} \exp \left\{ -\frac{(\delta - \mu)^2}{2\tau\lambda^2} \right\} \frac{(\beta/2)^{v/2}}{\Gamma(v/2)} (\lambda^{-2})^{\frac{v}{2}+1} \exp \left\{ -\frac{\beta\lambda^{-2}}{2} \right\} d\lambda^2 \\ &= \frac{1}{\sqrt{2\pi\tau}} \frac{(\beta/2)^{v/2}}{\Gamma(v/2)} \int_0^\infty (\lambda^{-2})^{1/2} (\lambda^{-2})^{\frac{v}{2}+1} \exp \left\{ -\lambda^{-2} \left[\frac{1}{2} \left(\frac{(\delta - \mu)^2}{\tau} + \beta \right) \right] \right\} d\lambda^2 \\ &= \frac{1}{\sqrt{2\pi\tau}} \frac{(\beta/2)^{v/2}}{\Gamma(v/2)} \int_0^\infty (\lambda^{-2})^{\frac{v+1}{2}+1} \exp \left\{ -\lambda^{-2} \left[\frac{1}{2} \left(\frac{(\delta - \mu)^2}{\tau} + \beta \right) \right] \right\} d\lambda^2 \end{aligned}$$

⁴The bayesian Choice', [8].

multiplicando por la constante adecuada que completa una función de densidad Gamma-Inversa⁵

$$IG\left(\frac{v+1}{2}, \frac{1}{2}\left[\frac{(\delta-\mu)^2}{\tau} + \beta\right]\right)$$

obtenemos

$$\begin{aligned} \pi(\delta) &= \frac{\left(\frac{1}{2}\right)^{1/2} \left(\frac{\beta}{2}\right)^{v/2} \Gamma((v+1)/2)}{\sqrt{\tau\pi}\Gamma(v/2) \left(\frac{1}{2}\right)^{\frac{v+1}{2}} \left[\frac{(\delta-\mu)^2}{\tau} + \beta\right]^{\frac{v+1}{2}}} \\ &= \frac{\Gamma((v+1)/2) \left(\frac{1}{\tau}\right)^{1/2} \left(\frac{1}{2}\right)^{1/2} \left(\frac{\beta}{2}\right)^{v/2}}{\Gamma(v/2) \sqrt{\pi} \left(\frac{1}{2}\right)^{v/2} \left(\frac{1}{2}\right)^{1/2} \left[\frac{(\delta-\mu)^2 + \beta\tau}{\tau}\right]^{\frac{v+1}{2}}} \\ &= \frac{\Gamma((v+1)/2) \left(\frac{1}{\tau}\right)^{1/2} \beta^{v/2} \left(\frac{1}{2}\right)^{v/2}}{\Gamma(v/2) \sqrt{\pi} \left(\frac{1}{2}\right)^{v/2} \left(\frac{1}{\tau}\right)^{v/2} \left(\frac{1}{\tau}\right)^{1/2} [(\delta-\mu)^2 + \beta\tau]^{v/2}} \\ &= \frac{\Gamma((v+1)/2) \beta^{v/2} \tau^{v/2}}{\Gamma(v/2) \sqrt{\pi} [(\delta-\mu)^2 + \beta\tau]^{\frac{v+1}{2}}} \\ &= \frac{\Gamma((v+1)/2)}{\Gamma(v/2) \sqrt{\pi\beta\tau} \left[\frac{(\delta-\mu)^2}{\beta\tau} + 1\right]^{\frac{v+1}{2}}} \\ &= \frac{\Gamma((v+1)/2)}{\Gamma(v/2) \sqrt{\pi\beta\tau} \left[\frac{(\delta-\mu)^2}{\beta\tau} + 1\right]^{-\frac{v+1}{2}}} \end{aligned}$$

En el ejemplo anterior elegimos esa descomposición para la distribución inicial ya que le familia Gamma-Inversa es conjugada para la función de verosimilitud Normal ⁶.

⁵Ver apéndice ??.

⁶Ver apéndice ??

Otra forma de resolver el caso σ^2 desconocida es proponer una distribución inicial para el vector $(\delta, \sigma^2/n)$. Como ya hemos mencionado, un modelo de este tipo considera que σ^2 tiene una influencia directa en él.

Ejemplo 3.0.2. Consideremos el modelo donde $f(x|\delta, \sigma^2/n) = N(\delta, \sigma^2/n)$ con δ y σ^2 ambos desconocidos y la distribución inicial de $(\delta, \sigma^2/n) = (\delta, \lambda^2)$ dada por

$$\pi(\delta, \lambda^2) = N(\delta; \gamma, \lambda^2/(2\alpha - 1)) IG(\lambda^2; \alpha - 2, \beta).$$

El beneficio de elegir esta distribución inicial es que $\pi(\delta, \sigma^2/n)$ es conjugada para la función de verosimilitud $N(\delta, \sigma^2/n)$ como se demuestra en la siguiente observación.

Observación 3.0.2. Sea

$$f(x|\delta, \lambda) = N(\delta, \sigma^2/n)$$

y

$$\pi(\delta, \lambda^2) = N(\delta; \gamma, \lambda^2/(2\alpha - 1)) IG(\lambda^2; \alpha - 2, \beta)$$

entonces $\pi(\delta, \lambda^2)$ es una familia conjugada para la función de verosimilitud $f(x|\delta, \lambda^2)$.

Demostración. Calculemos la distribución posterior de (δ, λ^2)

$$\begin{aligned} \pi(\delta, \lambda^2|x) &\propto \frac{1}{\sqrt{2\pi\lambda^2}} \exp\left\{-\frac{(x-\delta)^2}{2\lambda^2}\right\} \frac{1}{\sqrt{2\pi(\lambda^2/(2\alpha-1))}} \exp\left\{-\frac{(\delta-\gamma)^2}{2(\lambda^2/(2\alpha-1))}\right\} \\ &\quad \times \frac{\beta^{\alpha-2}}{\Gamma(\alpha-2)(\lambda^2)^{\alpha-1}} \exp\left\{-\frac{\beta}{\lambda^2}\right\} \\ &\propto (\lambda^{-2})^\alpha \exp\left\{-\frac{(x-\delta)^2}{2\lambda^2}\right\} \exp\left\{-\frac{(\delta-\gamma)^2(2\alpha-1)}{2\lambda^2}\right\} \exp\left\{-\frac{\beta}{\lambda^2}\right\} \\ &\propto (\lambda^{-2})^\alpha \exp\left\{-\lambda^{-2} \left[\frac{(x-\delta)^2 + (\delta-\gamma)^2(2\alpha-1) + 2\beta}{2}\right]\right\} \\ &\propto (\lambda^{-2})^\alpha \exp\left\{-\lambda^{-2} \left[\frac{x^2 - 2x\delta + \delta^2 + (2\alpha-1)(\delta^2 - 2\gamma\delta + \gamma^2) + 2\beta}{2}\right]\right\} \end{aligned}$$

$$\propto \exp \left\{ -\lambda^{-2} \left[\frac{\delta^2 (2\alpha) - 2\delta (x + \gamma (2\alpha - 1)) + x^2 + (2\alpha - 1) \gamma^2 + 2\beta}{2} \right] \right\} \\ \times (\lambda^{-2})^\alpha.$$

Trabajemos con el numerador de la fracción en el argumento de la exponencial

$$\begin{aligned} v &= \delta^2 (2\alpha) - 2\delta (x + \gamma (2\alpha - 1)) + x^2 + (2\alpha - 1) \gamma^2 + 2\beta \\ &= \delta^2 (2\alpha) - 2\delta (x + \gamma (2\alpha - 1)) + \frac{(x + \gamma(2\alpha - 1))^2}{2\alpha} + x^2 + (2\alpha - 1) \gamma^2 + 2\beta \\ &\quad - \frac{(x + \gamma(2\alpha - 1))^2}{2\alpha} \\ &= 2\alpha \left(\delta^2 - 2\delta \frac{x + \gamma(2\alpha - 1)}{2\alpha} + \frac{(x + \gamma(2\alpha - 1))^2}{(2\alpha)^2} \right) + (x^2 + (2\alpha - 1) \gamma^2 + 2\beta) \\ &\quad - \frac{(x + \gamma(2\alpha - 1))^2}{2\alpha} \\ &= 2\alpha \left(\delta - \frac{x + \gamma(2\alpha - 1)}{2\alpha} \right)^2 + (x^2 + (2\alpha - 1) \gamma^2 + 2\beta) - \frac{(x + \gamma(2\alpha - 1))^2}{2\alpha}. \end{aligned}$$

Si definimos

$$a = \alpha + \frac{1}{2}$$

$$g = \frac{x + \gamma(2\alpha - 1)}{2\alpha}$$

$$b = \frac{1}{2} \left[(x^2 + (2\alpha - 1) \gamma^2 + 2\beta) - \frac{(x + \gamma(2\alpha - 1))^2}{2\alpha} \right]$$

entonces

$$v = (2a - 1)(\delta - g)^2 + 2b.$$

De esta manera

$$\begin{aligned}
 \pi(\delta, \lambda^2 | x) &\propto (\lambda^{-2})^{a-\frac{1}{2}} \exp \left\{ -\lambda^{-2} \left[\frac{(2a-1)(\delta-g)^2 + 2b}{2} \right] \right\} \\
 &\propto \left(\frac{1}{\lambda^2} \right)^{\frac{1}{2}} \left(\frac{2a-1}{2\pi} \right)^{\frac{1}{2}} \exp \left\{ -\frac{(\delta-g)^2}{2(\lambda^2/(2a-1))} \right\} \\
 &\quad \times \frac{b^{a-2}}{\Gamma(a-2) (\lambda^2)^{a-1}} \exp \left\{ -\frac{b}{\lambda^2} \right\} \\
 &\propto N(\delta; g, \lambda^2/(2a-1)) IG(\lambda^2; a-2, b).
 \end{aligned}$$

□

En los ejemplos anteriores, tenemos una solución al planteamiento de un modelo cuando σ^2 es desconocida, una posible dificultad en el ejemplo 3.0.1 es el cálculo de la distribución posterior de δ , lo cuál podría complicar la obtención de las cotas que determinan la función del número de usuarios subsecuentes, así como el cálculo de la probabilidad de ciertos eventos, necesaria para especificar la función de distribución del número de usuarios subsecuentes.

Discusión

En este apartado planteamos un modelo donde la distribución de los resultados clínicos no necesariamente es Normal concientes de las dificultades que pueda presentar el cálculo de la distribución del número de usuarios subsecuentes de un nuevo tratamiento.

En el capítulo 2 consideramos que los resultados clínicos de los pacientes con el tratamiento nuevo y los pacientes con el tratamiento actual seguían una distribución $X_i \sim N(\theta + \delta, \eta^2)$ y $Y_j \sim N(\theta, \eta^2)$ respectivamente. A partir de esta información pudimos determinar la distribución de $\bar{Z}_n = \bar{X}_n - \bar{Y}_n$.

En el planteamiento de un modelo más general, donde $X_i \sim F_1$ y $Y_j \sim F_2$, no siempre es fácil determinar la distribución de $\bar{Z}_n = \bar{X}_n - \bar{Y}_n$, entonces partiremos del hecho de que $\bar{Z} \sim F$ con media δ desconocida y varianza σ^2 conocida (en algunos casos la varianza estará en función del tamaño de la muestra). Para nuestro parámetro desconocido δ proponemos una distribución inicial con función de densidad $\pi(\delta)$, cuyos parámetros son conocidos. Siguiendo la notación del capítulo 2 denotamos con $m(\bar{z}_n) = m$ el número de usuarios subsecuentes del nuevo tratamiento y M el número de usuarios potenciales subsecuentes del nuevo tratamiento.

Seguimos considerando que la relación entre m y δ es la que se muestra en la figura 4.1.

Para calcular el número de usuarios subsecuentes siguiendo la idea antes desarrollada consideramos el percentil a que acumula 0.93 en la distribución de $(\delta - \mu'(\bar{z}_n))/\tau'(\bar{z}_n)$ dada la muestra \underline{X} , donde $\mu'(\bar{z}_n)$ y $\tau'(\bar{z}_n)$ son la media y la desviación estándar de la distribución posterior de δ^7 . Es decir buscamos

⁷Observemos que estamos considerando F y $\pi(\delta)$ tal que la distribución posterior tiene primer y segundo momento finitos.

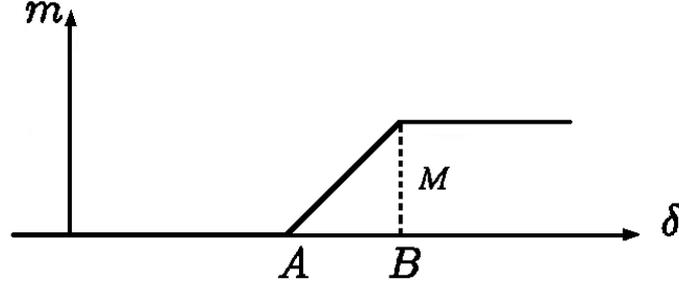


Figura 4.1

a tal que

$$0.93 = \mathbb{P} \left[\frac{\delta - \mu'(\bar{z}_n)}{\tau'(\bar{z}_n)} > a \mid \underline{X} \right] = \mathbb{P} [\delta > \mu'(\bar{z}_n) + a\tau'(\bar{z}_n) \mid \underline{X}].$$

Consideramos A y B cotas iniciales para δ . Si $A < \mu'(\bar{z}_n) + a\tau'(\bar{z}_n)$, entonces

$$\mathbb{P} [A < \delta \mid \underline{X}] \geq \mathbb{P} [\mu'(\bar{z}_n) + a\tau'(\bar{z}_n) < \delta \mid \underline{X}] = 0.93.$$

Bajo el mismo argumento, si $\mu'(\bar{z}_n) + a\tau'(\bar{z}_n) < B$, entonces

$$\mathbb{P} [\delta > B \mid \underline{X}] \leq \mathbb{P} [\delta > \mu'(\bar{z}_n) + a\tau'(\bar{z}_n) \mid \underline{X}] = 0.93$$

o equivalentemente

$$\mathbb{P} [\delta \leq B \mid \underline{X}] \geq 0.07.$$

Entonces el número de usuarios subsecuentes del nuevo tratamiento es

$$m(\bar{z}_n) = \begin{cases} 0 & \mu'(\bar{z}_n) < A - a\tau'(\bar{z}_n) \\ \frac{M}{B-A} [\mu'(\bar{z}_n) - A + a\tau'(\bar{z}_n)] & A - a\tau'(\bar{z}_n) < \mu'(\bar{z}_n) < B - a\tau'(\bar{z}_n) \\ M & B - a\tau'(\bar{z}_n) < \mu'(\bar{z}_n). \end{cases}$$

Tamaño óptimo de la muestra

Para calcular el tamaño óptimo de la muestra nuevamente consideramos que el beneficio por paciente al realizar la prueba es una constante b independiente de δ . Entonces la función objetivo es

$$r(n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m(\bar{z}_n) b \Pi^n(\delta | \bar{z}_n) d\delta f(\bar{z}_n) d\bar{z}_n - cn$$

donde $\Pi^n(\delta | \bar{z}_n)$ es la distribución posterior del parámetro δ dado \bar{z}_n .

Resolviendo la integral respecto a δ obtenemos

$$r(n) = \int_{-\infty}^{\infty} m(\bar{z}_n) b f(\bar{z}_n) d\bar{z}_n - cn$$

donde $f(\bar{z}_n)$ es la función de densidad a priori de \bar{Z}_n ⁸. En términos de variables escaladas $R(n) = r(n)/Mb$ y $C = c/Mb$ tenemos

$$R(n) = \int_{H_1(A,n)}^{H_2(B,n)} \frac{1}{B-A} (\mu'(\bar{z}_n) - A + a\tau'(\bar{z}_n)) f(\bar{z}_n) d\bar{z}_n \\ + \int_{H_2(B,n)}^{\infty} f(\bar{z}_n) d\bar{z}_n - Cn$$

donde $H_1(A, n)$ es el valor de (\bar{z}_n) para el cual $\mu'(\bar{z}_n) = A - a\tau'(\bar{z}_n)$ y $H_2(B, n)$ es el valor de (\bar{z}_n) para el cual $\mu'(\bar{z}_n) = B - a\tau'(\bar{z}_n)$. Para obtener el tamaño óptimo de la muestra bastaría maximizar la función $R(n)$.

Distribución del número de usuarios

Una vez que obtuvimos el tamaño de la muestra, procedemos a calcular la función de distribución del número de usuarios subsecuentes. Observamos que para $y < 0$ tenemos $\mathbb{P}[m \leq y] = 0$. El caso $0 \leq y < M$ sucede si y solo si

$$A - a\tau'(\bar{z}_n) < \mu'(\bar{z}_n) < B - a\tau'(\bar{z}_n)$$

y en este intervalo

$$m(\bar{z}_n) = \frac{M}{B-A} [\mu'(\bar{z}_n) - A + a\tau'(\bar{z}_n)],$$

⁸En el caso Normal, $\bar{Z}_n \sim N(\mu, \tau^2 + \sigma^2/n)$, pero en nuestro caso tendríamos que calcular $f(\bar{z}_n) = \int_{-\infty}^{\infty} f(\bar{z}_n | \delta) \pi(\delta) d\delta$.

entonces

$$\begin{aligned}
 F_m(y) &= \mathbb{P} \left[\frac{M}{B-A} [\mu'(\bar{z}_n) - A + \tau'(\bar{z}_n)] \leq y \right] \\
 &= \mathbb{P} \left[\mu'(\bar{z}_n) - A + a\tau'(\bar{z}_n) \leq y \frac{B-A}{M} \right] \\
 &= \mathbb{P} \left[\mu'(\bar{z}_n) + a\tau'(\bar{z}_n) \leq \frac{y}{M}(B-A) + A \right].
 \end{aligned}$$

Entonces la función de distribución el número de usuarios subsecuentes de un nuevo tratamiento se expresa de la siguiente forma

$$F_m(y) = \begin{cases} 0 & y < 0 \\ \mathbb{P} \left[\mu'(\bar{z}_n) + a\tau'(\bar{z}_n) \leq \frac{y}{M}(B-A) + A \right] & 0 \leq y < M \\ 1 & y \geq M, \end{cases}$$

A diferencia del caso donde \bar{Z}_n tiene distribución Normal, la distribución de m queda en términos de la distribución de una función de \bar{Z}_n , que tiene que ver con la media y desviación estándar de la distribución posterior de δ . Si τ' no depende de \bar{Z}_n , entonces queda simplemente en términos de la distribución de $\mu'(\bar{Z}_n)$.

Ejemplo

Consideramos un ejemplo donde la distribución de Z_n es $Gamma(n, \lambda)$, obtenemos la distribución de m mediante ambos modelos y comparamos $\mathbb{E}[m]$ en cada caso.

Sea $\bar{Z}_n \sim Gamma(n, \lambda)$, entonces

$$\mathbb{E}[\bar{Z}_n] = \frac{n}{\lambda} \quad y \quad Var(\bar{Z}_n) = \frac{n}{\lambda^2}.$$

Proponemos como distribución inicial para λ una exponencial con media μ conocida. Entonces la distribución inicial de δ es $IG(1, n/\mu)$ ⁹. Con esta información, calculamos la distribución posterior

$$\begin{aligned} f_{\delta|\bar{z}_n}(\delta|\bar{z}_n) &\propto \text{Gama}(n, n/\delta)IG(1, n/\mu) \\ &\propto \left(\frac{n}{\delta}\right)^n \frac{\bar{z}_n^{n-1}}{\Gamma(n)} \exp\{-\bar{z}_n n/\delta\} \frac{n}{\mu\delta^2} \exp\left\{-\frac{n}{\mu\delta}\right\} \\ &\propto \frac{1}{\delta^{n+2}} \exp\left\{-\frac{1}{\delta}\left(n\bar{z}_n + \frac{n}{\mu}\right)\right\} \\ &\propto IG(n+1, n(\bar{z}_n + 1/\mu)). \end{aligned}$$

Por lo tanto la esperanza y la desviación estándar de la distribución posterior son

$$\mu'(\bar{z}_n) = \bar{z}_n + \frac{1}{\mu} \quad \text{y} \quad \tau'(\bar{z}_n) = \frac{\bar{z}_n + 1/\mu}{\sqrt{n-1}}.$$

Procedemos a calcular el percentil a para determinar al número de usuarios, es decir a tal que

$$\begin{aligned} 0.93 &= \mathbb{P}\left[\frac{\delta - \mu'(\bar{z}_n)}{\tau'(\bar{z}_n)} > a \mid \bar{Z}_n\right] \\ &= \mathbb{P}\left[\frac{\delta - (\bar{z}_n + 1/\mu)}{(\bar{z}_n + 1/\mu)/\sqrt{n-1}} > a \mid \bar{Z}_n\right] \\ &= \mathbb{P}\left[\delta > \bar{z}_n + \frac{1}{\mu} + a \frac{\bar{z}_n + 1/\mu}{\sqrt{n-1}} \mid \bar{Z}_n\right], \end{aligned}$$

entonces $\bar{z}_n + \frac{1}{\mu} + a \frac{\bar{z}_n + 1/\mu}{\sqrt{n-1}}$ es el percentil que acumula 0.93 en una función de distribución $IG(n+1, n(\bar{z}_n + 1/\mu))$.

⁹Ver apéndice A.2.1.

En tal caso tenemos que

$$\begin{aligned}
& \mathbb{P} \left[\mu'(\bar{z}_n) + a\tau'(\bar{z}_n) \leq \frac{y}{M}(B - A) + A \right] \\
&= \mathbb{P} \left[\bar{z}_n + \frac{1}{\mu} \leq \frac{y}{M}(B - A) + A - a \frac{\bar{z}_n + 1/\mu}{\sqrt{n-1}} \right] \\
&= \mathbb{P} \left[\bar{z}_n + a \frac{\bar{z}_n}{\sqrt{n-1}} \leq \frac{y}{M}(B - A) + A - \frac{1}{\mu} - a \frac{1/\mu}{\sqrt{n-1}} \right] \\
&= \mathbb{P} \left[\bar{z}_n \left(1 + \frac{a}{\sqrt{n-1}} \right) \leq \frac{y}{M}(B - A) + A - \frac{1}{\mu} \left(1 + \frac{a}{\sqrt{n-1}} \right) \right] \\
&= \mathbb{P} \left[\bar{z}_n \leq \left[\frac{y}{M}(B - A) + A \right] \frac{\sqrt{n-1}}{\sqrt{n-1} + a} - \frac{1}{\mu} \right] \\
&= F_{\bar{Z}_n} \left(\left[\frac{y}{M}(B - A) + A \right] \frac{\sqrt{n-1}}{\sqrt{n-1} + a} - \frac{1}{\mu} \right).
\end{aligned}$$

De tal manera, la función de distribución del numero de usuarios subsecuentes es

$$F_m(y) = \begin{cases} 0 & y < 0 \\ F_{\bar{Z}_n} \left(\left[\frac{y}{M}(B - A) + A \right] \frac{\sqrt{n-1}}{\sqrt{n-1} + a} - \frac{1}{\mu} \right) & 0 \leq y < M \\ 1 & y \geq M. \end{cases}$$

Observemos que

$$\mathbb{P}[m = 0] = F_m(0) \lim_{h \rightarrow 0^+} F_m(0 - h) = F_{\bar{Z}_n} \left(\frac{A\sqrt{n-1}}{\sqrt{n-1} + a} - \frac{1}{\mu} \right)$$

y

$$\mathbb{P}[m = M] = F_m(M) - \lim_{h \rightarrow 0^+} F_m(M - h) = 1 - F_{\bar{Z}_n} \left(\frac{B\sqrt{n-1}}{\sqrt{n-1} + a} - \frac{1}{\mu} \right).$$

Por lo tanto la función de densidad del número de usuarios subsecuentes es la siguiente:

$$f_m(y) = \begin{cases} \mathbb{P}[m = y] = F_{\bar{Z}_n} \left(\frac{A\sqrt{n-1}}{\sqrt{n-1} + a} - \frac{1}{\mu} \right) & y = 0 \\ f_{\bar{Z}_n} \left(\left[\left(\frac{y}{M}(B-A) + A \right) \frac{\sqrt{n-1}}{\sqrt{n-1} + a} - \frac{1}{\mu} \right] \right) \\ \times \frac{\sqrt{n-1}(B-A)}{M\sqrt{n-1} + 1} & 0 < y < M \\ \mathbb{P}[m = M] = 1 - F_{\bar{Z}_n} \left(\frac{B\sqrt{n-1}}{\sqrt{n-1} + a} - \frac{1}{\mu} \right) & y = M. \end{cases}$$

Observemos que la distribución del m queda en términos de a y de la distribución inicial de \bar{Z}_n que podemos obtener de la distribución conjunta de \bar{Z}_n y δ

$$f_{\bar{Z}_n, \delta}(z, \delta) = f_{\bar{Z}_n | \delta}(z | \delta) \pi(\delta) = \left(\frac{n}{\delta} \right)^n \frac{z^{n-1}}{\Gamma(n)} \exp \{ -\bar{z}_n n / \delta \} \frac{n}{\mu \delta^2} \exp \left\{ -\frac{n}{\mu \delta} \right\},$$

entonces

$$\begin{aligned} f_{\bar{Z}_n}(z) &= \int_0^\infty f_{\bar{Z}_n, \delta}(z, \delta) d\delta \\ &= \frac{n^{n+1} z^{n-1}}{\mu \Gamma(n)} \int_0^\infty \frac{1}{\delta^{n+2}} \exp \left\{ -\frac{1}{\delta} \left(nz + \frac{n}{\mu} \right) \right\} d\delta \\ &= \frac{n^{n+1} z^{n-1}}{\mu \Gamma(n)} \times \frac{\Gamma(n+1)}{\left(nz + (n/\mu) \right)^{n+1}} \\ &= \frac{nz^{n-1}}{\mu \left(z + \frac{1}{\mu} \right)^{n+1}}. \end{aligned}$$

A partir de la función de densidad calculamos la función de distribución

$$F_{\bar{Z}_n}(z) = \int_0^z f_{\bar{Z}_n}(u) du = \int_0^z \frac{nu^{n-1}}{\mu \left(u + \frac{1}{\mu} \right)^{n+1}} du,$$

utilizando el cambio de variable

$$v = \frac{u}{u + \frac{1}{\mu}}, \quad dv = \frac{1/\mu}{\left(u + \frac{1}{\mu}\right)^2} du$$

$$\begin{aligned} F_{\bar{Z}_n}(z) &= \frac{n}{\mu} [\mu \int v^{n-1} dv] \\ &= n \left[\frac{v}{n-1} \right] \\ &= \left(\frac{u}{u + \frac{1}{\mu}} \right)^n \Big|_0^z \\ &= \left(\frac{z}{z + \frac{1}{\mu}} \right)^n. \end{aligned}$$

Ahora, para calcular a debemos conocer el percentil 0.93 de una distribución Gamma-Inversa de parámetros n y $n(\bar{z}_n + (1/\mu))$.

Observemos que si

$$X \sim IG(n, n(\bar{z}_n + (1/\mu))) \implies Y = 1/X \sim \text{Gamma}(n, n(\bar{z}_n + (1/\mu))),$$

entonces

$$F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[X^{-1} \leq y] = \mathbb{P}[X \geq 1/y] = 1 - F_X(1/y).^{10}$$

Por otro lado, si aplicamos el modelo desarrollado en el capítulo 2, definimos la estadística $T_n = \ln(\bar{Z}_n/n)$ cuya distribución se aproxima a una $N(\ln(1/\lambda), 1/n)$. Como $\bar{Z}_n \sim \text{Gamma}(n, \lambda)$, el Teorema central del límite asegura que

¹⁰Numericamente es más sencillo calcular a a partir de la distribución de una variable aleatoria Gamma.

$$\begin{aligned} \frac{\bar{Z}_n - (n/\lambda)}{(\sqrt{n}/\lambda)} &\xrightarrow{d} N(0, 1) && \iff \\ \frac{\bar{Z}_n - (n/\lambda)}{\sqrt{n}} &\xrightarrow{d} N(0, 1/\lambda^2) && \iff \\ \sqrt{n} \left(\frac{\bar{Z}_n}{n} - \frac{1}{\lambda} \right) &\xrightarrow{d} N(0, 1/\lambda^2) \end{aligned}$$

Utilizando una versión más general de la observación 3.0.1¹¹, podemos afirmar que si $f'(1/\lambda)$ existe y es distinta de cero, entonces

$$\sqrt{n} \left(f \left(\frac{\bar{Z}_n}{n} \right) - f \left(\frac{1}{\lambda} \right) \right) \xrightarrow{d} N \left(0, \frac{1}{\lambda^2} [f'(1/\lambda)]^2 \right),$$

es decir,

$$\begin{aligned} \sqrt{n} \left(\ln \left(\frac{\bar{Z}_n}{n} \right) - \ln \left(\frac{1}{\lambda} \right) \right) &\xrightarrow{d} N(0, 1) && \iff \\ \ln \left(\frac{\bar{Z}_n}{n} \right) - \ln \left(\frac{1}{\lambda} \right) &\xrightarrow{d} N \left(0, \frac{1}{n} \right) \end{aligned}$$

Entonces,

$$T_n = \ln \left(\frac{\bar{Z}_n}{n} \right) \xrightarrow{d} N \left(\delta, \frac{\sigma^2}{n} \right)$$

donde $\delta = \ln \left(\frac{1}{\lambda} \right)$ y $\sigma^2 = 1$.

A continuación presentamos las gráficas donde comparamos la esperanza del número de usuarios subsecuentes, calculada en base a $\bar{Z}_n \sim \text{Gamma}(n, \lambda)$ y en base a $T_n = \ln(\bar{Z}_n/n)$ ¹². Al modelo donde consideramos a \bar{Z}_n lo llamamos modelo 1 y aquel en el que consideramos a T_n modelo 2.

Primero analizamos $\mathbb{E}[m]$ como función de M y obtuvimos que cuando $M \rightarrow \infty$ la esperanza crece en forma lineal con ambos modelos, figura 4.2.

Aparentemente no existe diferencia entre un modelo y otro.

¹¹A este resultado se le denomina “Metodo delta”.

¹²Todos los cálculos fueron realizados con el programa R, ver apéndice B.2.

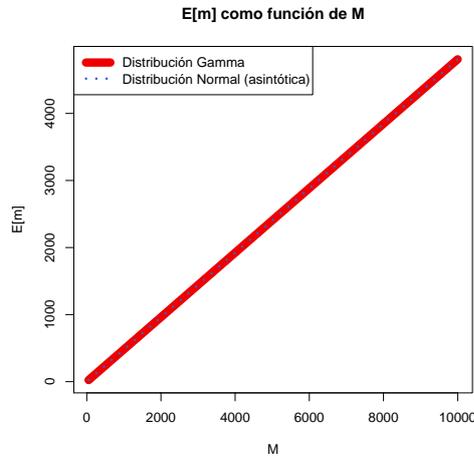


Figura 4.2: $\mu = 115$, $n = 50$.

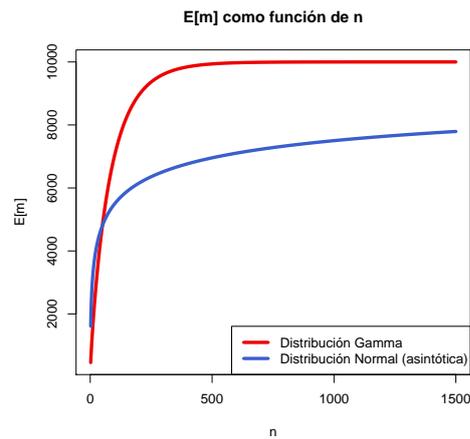
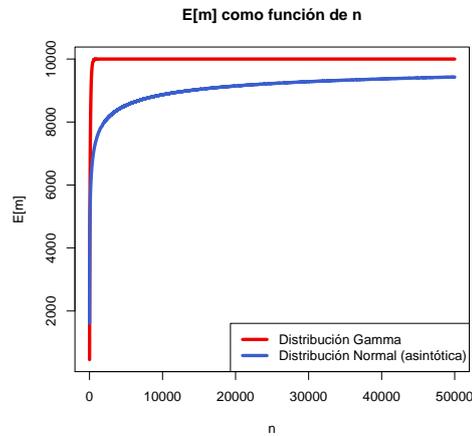
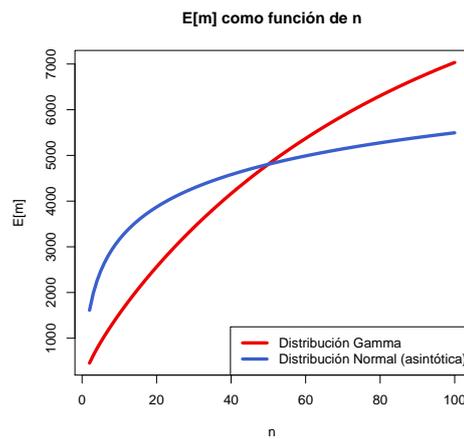


Figura 4.3: $\mu = 115$, $M = 10000$.

Después consideramos $\mathbb{E}[m]$ como función de n , y pudimos observar que la esperanza con el modelo 1 crece más rápidamente, conforme $n \rightarrow \infty$, que con el modelo 2, figura 4.3.

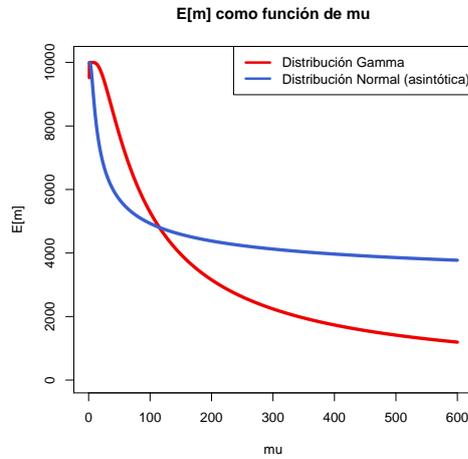
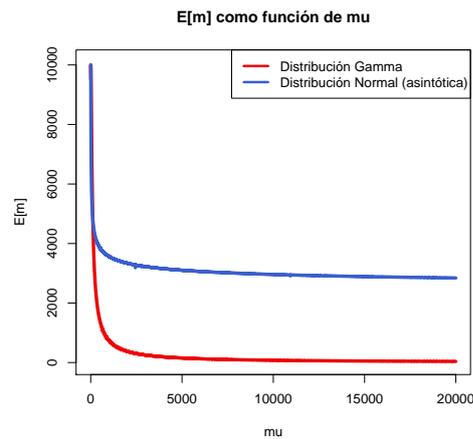
A pesar de que con ambos modelos $\mathbb{E}[m]$ se aproxima a M , figura 4.4, la velocidad de convergencia es distinta y para n entre 1 y 100, la diferencia entre modelos es notable, figura 4.5.

Figura 4.4: $\mu = 115$, $M = 10000$.Figura 4.5: $\mu = 115$, $M = 10000$.

Por último analizamos $\mathbb{E}[m]$ como función de μ , y observamos que el comportamiento es completamente distinto de acuerdo al modelo considerado; mientras con el modelo 1 decrece rápidamente a cero, con el modelo 2 si μ es del orden 10^4 , entonces $\mathbb{E}[m] > 2000$, figura 4.7.

Es importante observar que a pesar de considerar los mismos parámetros, el valor obtenido de $\mathbb{E}[m]$ es muy diferente entre un modelo y otro.

Para los primeros dos casos ($\mathbb{E}[m]$ como función de M y n) el valor que

Figura 4.6: $n = 50$, $M = 10000$.Figura 4.7: $n = 50$, $M = 10000$.

utilizamos para μ fue aquel donde la diferencia entre modelos es mínima para un valor fijo de M y n ($\mu = 115$), figura 4.8.

La razón por la que realizamos el anterior ejemplo fue tratar de contestar la interrogante de si el supuesto de normalidad está justificado, fue por esta razón que definimos una estadística cuya distribución se aproxima a una Normal, y con los resultados obtenidos podemos afirmar que la condición de normalidad asintótica no es suficiente para que nuestro modelo sea lo sufi-

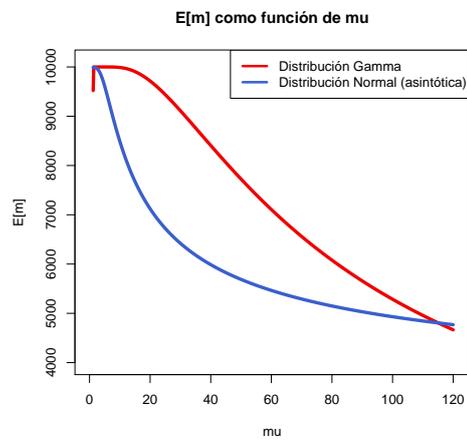


Figura 4.8: $n = 50$, $M = 10000$.

cientemente cercano a la realidad, es decir, podemos aplicarlo sólomente en aquellos casos donde los resultados clínicos, efectivamente siguen una distribución Normal.

Conclusiones

El desarrollo de este modelo bayesiano para determinar el número de usuarios subsecuentes de un nuevo tratamiento fue realizado suponiendo que la entidad que realiza la prueba clínica tiene información que le permite proponer una distribución inicial, con parámetros conocidos, para la diferencia entre el desarrollo de ambos tratamientos. Supusimos que los resultados clínicos tenían una distribución Normal.

Consideramos como hipótesis lógica que la estructura del parámetro δ no debería cambiar posterior a la toma de una muestra aleatoria, fue así, que propusimos un modelo conjugado.

A partir de la función de distribución obtuvimos explícitamente la función de densidad del número de usuarios subsecuentes. La razón fue, poder calcular el número esperado de usuarios que volverán a utilizar el nuevo tratamiento, con el fin de planear una estrategia óptima de producción o distribución y principalmente evaluar si existe una diferencia benéfica al implementar este nuevo tratamiento en comparación con el actual.

Aplicamos el modelo a los datos obtenidos en una prueba clínica realizada a pacientes con esclerosis múltiple, donde se comparaban dos modalidades de administración de interferon beta-1a. En este caso los resultados clínicos tenían distribución Bernoulli y el modelo no consideraba este supuesto. Definimos una nueva estadística para nuestro análisis, que asintóticamente era Normal, permitiéndonos de esta manera analizar los datos con base en el modelo propuesto en el capítulo 2.

Debido a la falta de información de las condiciones en que se realizó la prueba, propusimos ciertos valores con el fin de ejemplificar nuestra aplicación. Posteriormente analizamos cómo cambia el número esperado de usuarios subsecuentes cuando variamos estos parámetros.

Al final del capítulo 3 mencionamos dos propuestas para el caso donde la varianza de los resultados clínicos es desconocida: un modelo jerárquico donde la varianza es un hiperparámetro de primer orden o una distribución conjunta inicial para la media y la varianza de \bar{Z}_n . La segunda propuesta implica que la varianza de los resultados clínicos tiene una influencia directa en el modelo.

La complicación en que σ^2 sea desconocida comienza desde el momento en que debemos calcular el percentil 0.93 de la distribución posterior necesario para determinar la función del número de usuarios subsecuentes, es decir, determinar la función de distribución del número de usuarios subsecuentes es más complicado y la elección de los parámetros queda restringida a que los cálculos sean computacionalmente posibles.

A pesar de que el modelo desarrollado en el capítulo 2 resulta sencillo, es conveniente analizar la hipótesis de normalidad, para no cometer el error de considerar resultados que se alejan de la realidad del problema por el hecho de facilitar los cálculos.

En un último apartado discutimos el planteamiento de un modelo más general, donde suprimimos la hipótesis de normalidad y de que la elección de la distribución inicial sea tal que el modelo resulte conjugado. Obtuvimos que la distribución del número de usuarios subsecuentes, queda en términos de la media y la desviación estándar posteriores, entonces es posible analizar cualquier prueba donde los resultados clínicos tienen una distribución arbitraria, basándonos en la distribución de \bar{Z}_n y una distribución inicial para la esperanza de dicha estadística, con la restricción de que la posterior resultante tenga primer y segundo momento finito. Es necesario tener en cuenta las dificultades numéricas que con este planteamiento podemos tener, ya que al trabajar con una función de distribución sin expresión analítica tendremos que obtenerla mediante aproximaciones y estaremos limitados a la capacidad del programa de computadora que estemos utilizando.

El ejemplo que presentamos al final de esta tesis nos demuestra que si los resultados clínicos no siguen una distribución Normal y decidimos entonces considerar una estadística cuya distribución sea asintóticamente Normal, no podemos asegurar que los resultados obtenidos con este modelo respondan, de manera aceptable la pregunta planteada.

Podemos decir que el modelo propuesto en el capítulo 2 tiene como prin-

principal limitación que está dirigido a aquellas pruebas clínicas donde la distribución de los datos es normal, pero si contamos con esta hipótesis obtener la distribución del número de usuarios subsecuentes es relativamente sencillo. El modelo planteado en el último apartado no tiene esta limitación, pero no siempre es computacionalmente posible obtener la distribución del número de usuarios subsecuentes.

Apéndice A

Distribuciones

A.1. Distribución Gamma

$X \sim \text{Gamma}(r, \lambda)$ con $r > 0$, $\lambda > 0$.

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \mathbb{I}_{(0, \infty)}(x).$$

$$\mathbb{E}[X] = \frac{r}{\lambda}.$$

$$\text{Var}(X) = \frac{r}{\lambda^2}.$$

Si $r = 1$ se dice que X se distribuye exponencial de parámetro λ ,

$$X \sim \text{Exp}(\lambda).$$

A.2. Distribución Gamma-Inversa

$X \sim \text{IG}(r, \lambda)$ con $r > 0$, $\lambda > 0$.

$$f(x) = \frac{\lambda^r}{\Gamma(r)x^{r+1}} e^{-\lambda/x} \mathbb{I}_{(0, \infty)}(x).$$

$$\mathbb{E}[X] = \frac{\lambda}{r-1} \text{ para } r > 1.$$

$$\text{Var}(X) = \frac{\lambda^2}{(r-1)^2(r-2)} \text{ para } r > 2.$$

Proposición A.2.1. Sea X v.a. Gamma(r, λ).

a) Si $Y = g(X) = 1/X$, entonces $Y \sim IG(r, \lambda)$.

b) Si $r = 1$ y para $n \in \mathbb{N}$ se tiene que $W = g(X) = n/X$, entonces $W \sim IG(1, n\lambda)$.

Demostración. a) Sea $g^{-1}(y)$ el valor de x tal que $g(x) = y$. Como $g(x)$ es estrictamente monótona y diferenciable, sabemos que

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & \text{si } y = g(x) \text{ para alguna } x \\ 0 & \text{si } y \neq g(x) \text{ para toda } x. \end{cases}$$

Veamos que $g^{-1}(y) = 1/y$, entonces

$$\left| \frac{d}{dy} g^{-1}(y) \right| = \left| -\frac{1}{y^2} \right| = \frac{1}{y^2}.$$

De tal manera que

$$\begin{aligned} f_Y(y) &= \left(\frac{1}{y^2} \right) \frac{\lambda^r}{\Gamma(r)} \left(\frac{1}{y} \right)^{r-1} e^{-\lambda(1/y)} \\ &= \frac{\lambda^r}{\Gamma(r) y^{r+1}} e^{-\lambda/y}. \end{aligned}$$

Por lo tanto $Y \sim IG(r, \lambda)$.

b) Sea $g^{-1}(w)$ el valor de x tal que $g(x) = w$. Como $g(x)$ es estrictamente monótona y diferenciable, sabemos que

$$f_W(w) = \begin{cases} f_X(g^{-1}(w)) \left| \frac{d}{dw} g^{-1}(w) \right| & \text{si } w = g(x) \text{ para alguna } x \\ 0 & \text{si } w \neq g(x) \text{ para toda } x. \end{cases}$$

Veamos que $g^{-1}(w) = n/w$, entonces

$$\left| \frac{d}{dw} g^{-1}(w) \right| = \left| -\frac{n}{w^2} \right| = \frac{n}{w^2}.$$

De tal manera que

$$\begin{aligned} f_W(w) &= \left(\frac{n}{w^2}\right) \lambda e^{-\lambda(n/w)} \\ &= \frac{n\lambda}{w^2} e^{-n\lambda/w}. \end{aligned}$$

Por lo tanto $W \sim IG(1, n\lambda)$.

□

A.3. Modelo Normal-Gamma

El modelo bayesiano Normal-Gamma está compuesto por la función de verosimilitud $\text{Normal}(\mu, \tau\sigma^2)$ y la distribución inicial $\pi(\sigma^2) = IG(v/2, \beta/2)$.

Proposición A.3.1. Sean $f(\theta|\sigma^2)$ función de verosimilitud y $\pi(\sigma^2)$ distribución inicial tal que

$$f(\theta|\sigma^2) = \frac{1}{\sqrt{2\pi\tau\sigma^2}} \exp\left\{-\frac{(\theta - \mu)^2}{2\tau\sigma^2}\right\}$$

y

$$\pi(\sigma^2) = \frac{\left(\frac{\beta}{2}\right)^{v/2}}{\Gamma(v/2)} (\sigma^{-2})^{\frac{v}{2}+1} \exp\left\{-\frac{\beta\sigma^{-2}}{2}\right\}.$$

Entonces la familia $\pi(\sigma^2)$ es conjugada para la función de verosimilitud $f(\theta|\sigma^2)$.

Demostración. Calculemos la distribución posterior

$$\begin{aligned} \pi(\sigma^2|\theta) &\propto f(\theta|\sigma^2)\pi(\sigma^2) \\ &= \frac{1}{\sqrt{2\pi\tau\sigma^2}} \exp\left\{-\frac{(\theta - \mu)^2}{2\tau\sigma^2}\right\} \frac{\left(\frac{\beta}{2}\right)^{v/2}}{\Gamma(v/2)} (\sigma^{-2})^{\frac{v}{2}+1} \exp\left\{-\frac{\beta\sigma^{-2}}{2}\right\} \\ &\propto \exp\left\{-\frac{(\theta - \mu)^2}{2\tau}\sigma^{-2}\right\} (\sigma^{-2})^{1/2} (\sigma^{-2})^{\frac{v}{2}+1} \exp\left\{-\frac{\beta\sigma^{-2}}{2}\right\} \end{aligned}$$

$$\begin{aligned} &= (\sigma^{-2})^{\frac{v+1}{2}+1} \exp \left\{ -\frac{1}{2} \left(\frac{(\theta - \mu)^2}{\tau} + \beta \right) \sigma^{-2} \right\} \\ &\propto IG \left(\frac{v+1}{2}, \frac{1}{2} \left(\frac{(\theta - \mu)^2}{\tau} + \beta \right) \right). \end{aligned}$$

Por lo tanto, $\pi(\sigma^2)$ es conjugada para la función de verosimilitud $f(\theta|\sigma^2)$. \square

Apéndice B

Programas en R

B.1. Cálculo de la esperanza

```
library(MASS)

##### función de distribución de m
library(MASS)
peme<-function(y,M,n,sig,mu,tau){
  A=.8*mu # A
  B=1.2*mu # B
  tau=mu/2 # desviación estándar de la distribución inicial
  T=tau/sig # T
  D=mu/sig # D

  return(pnorm((y*(B-A)/(M*sig)+(A/sig)+1.5*(T^(-2)+n)^(-.5)-D)
              *(T^(-2)+n)^(.5)/(T*sqrt(n)))) )
}

##### esperanza de m
esp<-function(M,n,sig,mu,tau){
  A=.8*mu # A
  B=1.2*mu # B
  tau=mu/2 # desviación estándar de la distribución inicial
  T=tau/sig # T
  D=mu/sig # D

  f=as.function( alist(x=,a=A,b=B,c=M,d=n,e=mu,j=tau,1-pnorm((x*(B-A)/(M*sig)
    +(A/sig)+1.5*(T^(-2)+n)^(-.5)-D)*(T^(-2)+n)^(.5)/(T*sqrt(n))))))
```

```
return(integrate(f,0,M)$value)
}
```

```
esperanza=esp(10000,337,sqrt(2),2.48490665,1.242453325)
```

B.2. Modelo Gamma

El programa utilizado para realizar los cálculos del ejemplo desarrollado en el último apartado es el siguiente:¹

```
library(MASS)
##### modelo general #####
##### función de distribución de Z_n
pz_in<-function(z,n,mu){ #z mayor igual a 0
{return((z/(z+(1/mu)))^n)}
}
##### densidad gamma inversa (par1<170)
dinversgama<-function(w,par1,par2)
{
if(is.nan( par2^par1/gamma(par1)/w^(par1+1)*exp(-par2/w))==TRUE)
{return(0)}
if(is.nan( par2^par1/gamma(par1)/w^(par1+1)*exp(-par2/w))==FALSE)
{return(par2^par1/gamma(par1)/w^(par1+1)*exp(-par2/w))}
}
##### función de distribución gamma inversa
pinvgama<-function(u,par1,par2)
{
return(1-pgamma(1/u,shape=par1,scale=(1/par2)))
}
##### percentil #####
##### para encontrar los extremos a, b del intervalo donde
##### va a comenzar la función encuentra
tope<-function(u,par1,par2)
{
if(u>1||u<0){return("u no se encuentra en (0,1)")}
i=1
mus=par2/(par1-1)
while(u>pinvgama(mus*i,par1,par2)){i=i+1}
```

¹Las funciones “tope” y “encuentra” utilizadas en este programa son parte del proyecto de tesis de Arrigo Coen Coria.

```

return(i)
}
##### Encuentra el percentil u de una distribución
encuentra<-function(u,a,b,eps,par1,par2)
{
if(b-a<eps){return((a+b)/2)}
if(b-a>=eps)
{
if(u>=pinvgama((a+b)/2,par1,par2))
{return(encuentra(u,(a+b)/2,b,eps,par1,par2))}
if(u<pinvgama((a+b)/2,par1,par2))
{return(encuentra(u,a,(a+b)/2,eps,par1,par2))}
}
}
##### Función de distribución del número de usuarios
##### subsecuentes (modelo general)
pgam<-function(v,z,A,B,M,n,mu,percentil_u)
{
w=(percentil_u-z-(1/mu))*sqrt(n-1)/(z+(1/mu)) # percentil "a"
vprima=((v/M)*(B-A)+A)*sqrt(n-1)/(sqrt(n-1)+w)-(1/mu)
return(pz_in(vprima,n,mu))
}
##### Esperanza del número de usuarios subsecuentes (modelogeneral)
esperanza_m<-function(z,A,B,M,n,mu,percentil_u)
{
w=(percentil_u-z-(1/mu))*sqrt(n-1)/(z+(1/mu)) # percentil "a"
f=as.function(alist(x=,a=n,b=mu,1-(pz_in(((x/M)*(B-A)+A)
  *sqrt(n-1)/(sqrt(n-1)+w))-(1/mu),n=a,mu=b))))
return( integrate(f,0,M)$value)
}
##### normalidad asintótica #####
##### esperanza de ln(1/lambda)
dunolambda<-function(y,mu)
{
return((1/mu)*exp(-exp(-y)/mu-y))
}
eunolambda<-function(mu)
{
f=as.function(alist(x=,a=mu,x*dunolambda(x,mu=a)))
return( integrate(f,-Inf,Inf)$value )
}

```

```
##### Varianza de ln(1/lambda)
vunolambda<-function(mu)
{
f=as.function(alist(x=,a=mu,(x^2)*dunolambda(x,mu=a)))
return( integrate(f,-Inf,Inf)$value)
}
##### función de densidad de m
library(MASS)
peme<-function(y,A,B,M,n,mudos,taudos)
{
Ados=log(A/n) # A
Bdos=log(B/n) # B
sig=1
T=taudos/sig # T
D=mudos/sig # D
return(pnorm((y*(Bdos-Ados)/(M*sig)+(Ados/sig)+
1.5*(T^(-2)+n)^(-.5)-D)*(T^(-2)+n)^(.5)/(T*sqrt(n))))
}
##### esperanza de m
esp<-function(A,B,M,n,mudos,taudos)
{
Ados=log(A/n) # A
Bdos=log(B/n) # B
sig=1
T=taudos/sig # T
D=mudos/sig # D
f=as.function(alist(x=,a=A,b=B,c=M,d=n,e=mudos,j=taudos,
1-pnorm((x*(Bdos-Ados)/(M*sig)+(Ados/sig)
+1.5*(T^(-2)+n)^(-.5)-D)*(T^(-2)+n)^(.5)/(T*sqrt(n))))))
return(integrate(f,0,M)$value)
}
##### con diferentes parámetros #####
n=50 # parámetro1=n <=68 si mu=1.32(n<=77 si mu=5)(n<=73 si mu=3)
mu=115 # parámetro2=mu
u=.93
z=.1
par1=n+1 # 1er parámetro de la distribución posterior de delta
par2=n*(z+(1/mu)) # 2do parámetro de la distribución posterior de delta
mus=par2/(par1-1) # media de la distribución posterior de delta
a=(tope(u,par1,par2)-1)*mus
b=tope(u,par1,par2)*mus
```

```

eps=.0001
percentil_u=encuentra(u,a,b,eps,par1,par2)
A=.5
B=1.2
M=10000 # número de usuarios potenciales
mudos=eunolambda(mu) # media de la distribución inicial
taudos=sqrt(vunolambda(mu)) # des. estándar de distribución inicial
esperanza_m(z,A,B,M,n,mu,percentil_u) # esperanza de m con el modelo general
esp(A,B,M,n,mudos,taudos) # esperanza de m con normalidad asintótica
#####
##### para compararlas en una gráfica #####
#### E[m] con modelo general, como función de mu
# para mu<1.2 la esperanza se hace extremadamente negativa
laura=0
t=6
#Te=3000
Te=100000
#Te=600
for(i in t:Te ){
n=50
u=.93
z=.1
mu=i/5
par1=n+1
par2=n*(z+(1/mu))
mus=par2/(par1-1)
a=(tope(u,par1,par2)-1)*mus
b=tope(u,par1,par2)*mus
eps=.0001
percentil_u=encuentra(u,a,b,eps,par1,par2)
A=.5
B=1.2
M=10000
laura[i-t+1]=esperanza_m(z,A,B,M,n,mu,percentil_u)
}
vmu=0
for(i in t:Te ){
vmu[i-t+1]=i/5
}
plot(x=vmu,y=laura,ylim=c(0,10100),xlim=c(0,vmu[Te-t+1]+.5),type="l",
lwd=4,col="red2",xlab="mu",ylab="E[m]",main="E[m] como función de mu")

```

```

#plot(x=vmu,y=laura,ylim=c(4000,10100),xlim=c(0,vmu[Te-t+1]+.5),type="l",
lwd=4,col="red2",xlab="mu",ylab="E[m]",main="E[m] como función de mu")
#####
laus=0
for(i in t:Te ){
n=50
mu=i/5
A=.5
B=1.2
M=10000
mudos=eunolambda(mu)
taudos=sqrt(vunolambda(mu))
laus[i-t+1]=esp(A,B,M,n,mudos,taudos)
}
lines(x=vmu,y=laus,ylim=c(0,10100),xlim=c(0,vmu[Te-t+1]+.5),
type="l",lwd=4,col="royalblue3")
#lines(x=vmu,y=laus,ylim=c(4000,10100),xlim=c(0,vmu[Te-t+1]+.5),
type="l",lwd=4,col="royalblue3")
#legend("topright",xjust=1,c("Distribución Gamma","Distribución Normal
(asintótica)",lwd=c(3,3),lty=c(1,1),col=c("red2","royalblue3"))
legend(par("usr")[2],par("usr")[4],xjust=1,c("Distribución Gamma",
"Distribución Normal (asintótica)",lwd=c(3,3),lty=c(1,1),
col=c("red2","royalblue3"))
which.min(abs(laura-lauss))
#####
##### E[m] con modelo general, como función de M #####
la=0
t=50
Te=10000
for(i in t:Te ){
n=50
u=.93
z=.1
mu=115
M=i
par1=n+1
par2=n*(z+(1/mu))
mus=par2/(par1-1)
a=(tope(u,par1,par2)-1)*mus
b=tope(u,par1,par2)*mus
eps=.0001

```

```

percentil_u=encuentra(u,a,b,eps,par1,par2)
A=.5
B=1.2
la[i-t+1]=esperanza_m(z,A,B,M,n,mu,percentil_u)
}
vM=0
for(i in t:Te ){
vM[i-t+1]=i
}
plot(x=vM,y=la,type="l",lty=1,lwd=10,col="red2",xlab="M",ylab="E[m]",
main="E[m] como función de M")
#####
lau=0
for(i in t:Te ){
n=50
mu=115
M=i
A=.5
B=1.2
mudos=eunolambda(mu)
taudos=sqrt(vunolambda(mu))
lau[i-t+1]=esp(A,B,M,n,mudos,taudos)
}
lines(x=vM,y=lau,type="l",lty=3,lwd=3,col="royalblue3")
legend("topleft",xjust=1,c("Distribución Gamma","Distribución Normal
(asintótica)"),lwd=c(10,3),lty=c(1,3),col=c("red2","royalblue3"))
#plot(x=vM,y=la,ylim=c(0,1000),xlim=c(49,2000),type="l",lty=1,lwd=1,col="red2",
xlab="M",ylab="E[m]",main="E[m] como función de M")
#lines(x=vM,y=lau,type="l",lty=3,lwd=3,col="royalblue3")

plot(x=vM,y=la,ylim=c(23,27),xlim=c(49.8,53),type="o",lty=1,lwd=.1,col="black",
xlab="M",ylab="E[m]",main="E[m] como función de M")
lines(x=vM,y=lau,type="o",lty=1,lwd=.1,col="violetred3")
legend("topleft",xjust=1,c("Distribución Gamma","Distribución Normal
(asintótica)"),lwd=c(.1,.1),lty=c(1,1),col=c("black","violetred3"))
##### E[m] con modelo general, como función de n #####
lauri=0
t=2
Te=50000
#t=2
#Te=1500

```

```

for(i in t:Te ){
n=i
u=.93
z=.1
mu=115
M=10000
par1=n+1
par2=n*(z+(1/mu))
mus=par2/(par1-1)
a=(tope(u,par1,par2)-1)*mus
b=tope(u,par1,par2)*mus
eps=.0001
percentil_u=encuentra(u,a,b,eps,par1,par2)
A=.5
B=1.2
lauri[i-t+1]=esperanza_m(z,A,B,M,n,mu,percentil_u)
}
vn=0
for(i in t:Te ){
vn[i-t+1]=i
}
plot(x=vn,y=lauri,type="l",lty=1,lwd=4,col="red2",xlab="n",
ylab="E[m]",main="E[m] como función de n")
#####
lauris=0
for(i in t:Te ){
n=i
mu=115
M=10000
A=.5
B=1.2
mudos=eunolambda(mu)
taudos=sqrt(vunolambda(mu))
lauris[i-t+1]=esp(A,B,M,n,mudos,taudos)
}
lines(x=vn,y=lauris,type="l",lty=1,lwd=4,col="royalblue3")
legend("bottomright",xjust=1,c("Distribución Gamma","Distribución
Normal (asintótica)"),lwd=c(4,4),lty=c(1,1),col=c("red2","royalblue3"))

```

Bibliografía

- [1] Gut, Allan. Probability: A graduate course. Springer Texts in Statistics, 2005.
- [2] Rincón, Luis. Curso intermedio de probabilidad 1^a Edición. Las prensas de Ciencias, 2007.
- [3] Grimmett G.R., Stirzaker D.R. Probability and random processes 3ed Edition. Oxford, 2001.
- [4] Resnick S.I. A probability path. Birkhauser, 1999.
- [5] Billingsley P. Convergence of probability measures 2nd Edition. Wiley, 1999.
- [6] Jacod J., Protter P. Probability essentials 2nd Edition. Springer, 2003.
- [7] Chow, Yuan Shih and Teicher, Henry. Probability Theory. Independence, Interchangeability, Martingales. Third edition.
- [8] Robert, Christian P. The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation 2nd Edition. Springer Text in Statistic.
- [9] Lehmann E.L. Elements of Large-Sample Theory. Springer-Verlag, New York 1999.
- [10] Mood, Alexander M., Graybill, Franklin A. and Boes, Duane C. Introduction to the Theory of Statistics Third Edition. McGraw-Hill Series in Probability and Statistics.
- [11] Pezeshk, Hamid and Gittins, John. Bayesian approach to determine the number of subsequent users of a new treatment. *Statistical Methods in Medical Research* 2006;15: 585–592.

- [12] Pezeshk, Hamid and Gittins, John. How Large Should a Clinical Trial Be?. *The Statistician* 2000; Vol. 49, No. 2, pp. 177-187.
- [13] Hillel Panitch, Douglas Goodin, Gordon Fransis, Peter Chang, Patricia Coyle, Paul O'Connor, David Li, Brian Weinschenker. Benefits of high-dose, high-frequency interferon beta-1a in relapsing-remitting multiple sclerosis are sustained to 16 months: Final comparative results of EVIDENCE trial. *Journal of the Neurological Sciences*, 239(2005)67-74.