

**TAMAÑO DE MUESTRA NECESARIO PARA QUE LA
ESTIMACIÓN DE INTERVALOS FRECUENTISTA Y BAYESIANO
COINCIDAN DENTRO DE UN ERROR PREDETERMINADO**

Por

DEIVER SUÁREZ GÓMEZ

Tesis sometida en cumplimiento parcial de los requerimientos para el grado de

MAESTRÍA EN CIENCIAS

en

MATEMÁTICAS ESTADÍSTICA

UNIVERSIDAD DE PUERTO RICO
RECINTO UNIVERSITARIO DE MAYAGÜEZ

2018

Aprobada por:

Edgardo Lorenzo González, Ph.D
Miembro, Comité Graduado

Fecha

Dámaris Santana Morant, Ph.D
Miembro, Comité Graduado

Fecha

Wolfgang A. Rolke, Ph.D
Presidente, Comité Graduado

Fecha

Sonia M. Bartolomei Suárez, Ph.D
Representante de Estudios Graduados

Fecha

Olgamary Rivera Marrero, Ph.D
Directora del Departamento

Fecha

Resumen

Por lo general, se acepta que a medida que aumenta el tamaño de un conjunto de datos, el efecto de la distribución previa en el intervalo de credibilidad Bayesiano disminuye y estos intervalos se acercan a los correspondientes intervalos de confianza Frecuentista. En esta tesis, se estudia la pregunta de cuán grande es el tamaño de muestra que se necesita para tener cierta certeza de que los intervalos de confianza y los intervalos de credibilidad sean similares dentro de un error prescrito. Esto, con el propósito de afirmar cual es el tamaño de muestra necesario para que los dos enfoques de la estadística coincidan en la estimación por intervalos. Hay respuestas explícitas para varios casos estándar como la estimación de la media y la desviación estándar de una distribución Normal, la probabilidad de éxito en los ensayos de Bernoulli y las tasas de la distribución de Poisson, Geométrica y Exponencial. Para este trabajo se desarrolló una calculadora en línea que está disponible en https://server-deiver.shinyapps.io/sample_size_frequentist_and_bayesian/. Ésta calcula intervalos de confianza, intervalos de credibilidad y el tamaño de muestra necesario para que estos intervalos sean iguales dentro de un error predeterminado. La calculadora no requiere que el usuario tenga algún conocimiento de R.

Abstract

It is generally accepted that as the size of a dataset increases, the influence of the prior distribution on the Bayesian credible intervals decreases and these intervals approach the corresponding Frequentist confidence intervals. In this thesis, we study how large the sample size needs to be to have some certainty that the confidence intervals and credible intervals are within a prescribed error. There are explicit results for several standard cases such as estimation of the mean and the standard deviation from a Normal distribution, the success probability in Bernoulli trials, and the rates of Poisson, Geometric, and Exponential distributions. For the purpose of this work, we developed a calculator that is available online at https://serverdeiver.shinyapps.io/sample_size_frequentist_and_bayesian/. It calculates confidence intervals, credible intervals and the necessary sample size so that these intervals are equal within a predetermined error. The calculator does not require the user to have any knowledge of R.

Copyright © 2018

por

DEIVER SUÁREZ GÓMEZ

Dedicatoria

A Dios, mi apoyo incondicional, a mi esposa, a la mujer que me dio la vida, mi padre y demás familiares.

AGRADECIMIENTOS

Agradezco enormemente a mi esposa por ser un apoyo incondicional, a todos los profesores que durante este proceso me han brindado sus conocimientos y al Dr. Wolfgang A. Rolke por ser mi guía en este proceso.

Índice general

RESUMEN EN ESPAÑOL	II
ABSTRACT ENGLISH	III
AGRADECIMIENTOS	VI
Índice general	VII
Índice de tablas	X
Índice de figuras	XII
LISTA DE ABREVIATURAS	XXVI
LISTA DE SIMBOLOS	XXVII
1. INTRODUCCIÓN	1
2. Intervalos	6
2.1. Intervalos de confianza	6
2.1.1. Intervalos de confianza para el parámetro p de un modelo Binomial.	7
2.1.2. Intervalos de confianza para el parámetro λ de un modelo Poisson.	7

2.1.3.	Intervalos de confianza para el parámetro p de un modelo Geométrico.	9
2.1.4.	Intervalo de confianza para el parámetro λ de un modelo Exponencial.	9
2.1.5.	Intervalos de confianza para los parámetros μ, σ de un modelo Normal.	9
2.2.	Metodología Bayesiana para la estimación por Intervalos	10
2.2.1.	Selección de la distribución previa.	11
2.2.6.	Intervalos de credibilidad	25
3.	SIMULACIONES	27
3.1.	Distribución Binomial.	29
3.2.	Distribución Poisson.	32
3.3.	Distribución Geométrica.	35
3.4.	Distribución Exponencial.	39
3.5.	Distribución Normal.	41
3.5.1.	Inferencia en μ con σ_0 conocido	41
3.5.2.	Inferencia en σ^2 con μ_0 conocido	45
3.5.3.	Inferencia en μ con σ^2 desconocidos.	49
4.	LA APLICACIÓN WEB INTERACTIVA EN LÍNEA	57
4.1.	Calculate your own intervals.	58
4.2.	Find the sample size required.	79
5.	CONCLUSIONES Y TRABAJOS FUTUROS	102
	Bibliografía	104

APENDICES 106

A. PROPOSICIONES Y CÓDIGOS EN R 107

Índice de tablas

1-1. D_{50} , D_{99} y 1000 distancias obtenidas para el modelo Binomial con $n = 10$ y $p = 0.3$	3
3-1. Tamaño de muestra obtenido para el parámetro $p = 0.1, 0.5, 0.9$ del modelo Binomial.	30
3-2. Tamaño de muestra obtenido para el parámetro $\lambda = 1, 5, 10$ del modelo Poisson.	33
3-3. Tamaño de muestra obtenido para el parámetro $p = 0.1, 0.5, 0.9$ del modelo Geométrico.	36
3-4. Tamaño de muestra obtenido para el parámetro $\lambda = 1, 5, 10$ del modelo exponencial.	39
3-5. Tamaño de muestra obtenido para el parámetro $\mu = 1, 5, 10$ del modelo Normal.	42
3-6. Tamaño de muestra obtenido para el parámetro $\sigma^2 = 1, 5, 10$ del modelo Normal.	46
3-7. Tamaño de muestra obtenido para el parámetro $\mu = 1, 5, 10$ del modelo Normal con σ^2 desconocido, utilizando la distribución previa de Jeffrey para μ y σ^2 (μ y σ^2 independientes).	50

3–8. Tamaño de muestra obtenido para el parámetro $\mu = 1, 5, 10$ del modelo Normal con σ^2 desconocido, utilizando la distribución previa Conjugada para μ y σ^2 (μ y σ^2 independientes).	51
--	----

Índice de figuras

1-1. Histograma de las 1000 distancias simuladas para $n = 10$ y $p = 0.3$, ubicando $D_{50} = 0.1292484$ y $D_{99} = 0.1576000$	4
1-2. Comportamiento de la distancia D_{50} y tamaños de muestra n con el cual D_{50} es menor o igual a D_R cuando $p = 0.3$	5
1-3. Comportamiento de la distancia D_{99} y tamaños de muestra n con el cual D_{99} es menor o igual a D_R cuando $p = 0.3$	5
3-1. Comportamiento del tamaño de muestra n cuando se varía el paráme- tro p del modelo Binomial utilizando la distancia D_{50} y D_{99} para comparar los intervalos de confianza (Cloper-Pearson) con los in- tervalos de credibilidad (generados por las distribuciones previas de Jefrey y Beta Conjugada).	31
3-2. Comportamiento de la distancia en función del parámetro p y dife- rentes tamaños de muestra usando la distribución previa de Jeffrey y Beta Cojugado.	32
3-3. Comportamiento del tamaño de muestra n cuando se varía el paráme- tro λ en el modelo Poisson utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Garwood) y los intervalos de credibilidad (<i>Jeffrey</i> y <i>Gamma</i> (1,1)).	34

3-4. Comportamiento de la distancia en función del parámetro λ y diferentes tamaños de muestra usando la distribución previa Jeffrey y Gamma cojugado.	35
3-5. Comportamiento del tamaño de muestra n cuando se varía el parámetro p del modelo Geometrico utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza descritos en la Subsección 2.1.3 y los intervalos de credibilidad (Jefrey y $Beta(1, 1)$).	37
3-6. Comportamiento de la distancia en función del parámetro p y diferentes tamaños de muestra usando la distribución previa Jeffrey y Beta cojugado.	38
3-7. Tamaño de muestra n cuando se varía el parámetro λ del modelo Exponencial utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.4) y los intervalos de credibilidad (con la distribución $Beta(1, 1)$).	40
3-8. Comportamiento de la distancia en función del parámetro λ y diferentes tamaños de muestra usando la distribución previa $Gamma(1, 1)$	41
3-9. Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma_0 = 1$ conocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa $Normal(0, 1)$ (Ejemplo 2.12).	43

3-10. Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma_0 = 5$ conocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa $Normal(0, 1)$ (Ejemplo 2.12).	43
3-11. Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa $Normal(0, 1)$ en el caso de $\sigma_0 = 1$ conocido.	44
3-12. Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa $Normal(0, 1)$ en el caso de $\sigma_0 = 5$ conocido.	45
3-13. Tamaño de muestra n para el parámetro σ^2 del modelo Normal en el caso de $\mu_0 = 1$ conocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa $Invgamma(1, 1)$ (Ejemplo 2.15).	47
3-14. Tamaño de muestra n para el parámetro σ^2 del modelo Normal en el caso de $\mu_0 = 5$ conocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa $Invgamma(1, 1)$ (Ejemplo 2.15).	47
3-15. Comportamiento de la distancia en función del parámetro σ^2 y diferentes tamaños de muestra usando la distribución previa $Invgamma(1, 1)$ en el caso de $\mu_0 = 1$ conocido.	48

3-16.Comportamiento de la distancia en función del parámetro σ^2 y diferentes tamaños de muestra usando la distribución previa $Invgamma(1, 1)$ en el caso de $\mu_0 = 5$ conocido.	49
3-17.Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma^2 = 1$ desconocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa de Jeffrey (Ejemplo 2.14).	52
3-18.Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma^2 = 5$ desconocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa de Jeffrey (Ejemplo 2.14).	52
3-19.Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma^2 = 1$ desconocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa descrita en el Ejemplo 2.14.	53
3-20.Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma^2 = 5$ desconocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa descrita en el Ejemplo 2.14.	54

3-21.Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa Jeffrey (Ejemplo 2.7) en el caso de $\sigma^2 = 1$	55
3-22.Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa Jeffrey (Ejemplo 2.7) en el caso de $\sigma^2 = 5$	55
3-23.Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa descrita en el Ejemplo 2.14, en el caso de $\sigma^2 = 1$	56
3-24.Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa descrita en el Ejemplo 2.14, en el caso de $\sigma^2 = 5$	56
4-1. Menú de distribuciones muestrales para realizar estimación por intervalos	58
4-2. Selección de la distribución previa para el modelo Binomial	59
4-3. Parámetros de la previa Beta conjugada para el modelo Binomial	59
4-4. Ingrese su propia distribución previa para el modelo Binomial	59
4-5. Formas de ingresar datos en el panel para el modelo Binomial	60
4-6. Formato para ingresar datos del modelo Binomial	60
4-7. Medidas de resumen suficientes del modelo Binomial	60

4-8. Intervalos y gráfica de intervalos del modelo Binomial	61
4-9. Selección de la distribución previa para el modelo Poisson	62
4-10. Parámetros de la previa Gamma conjugada para el modelo Poisson . .	62
4-11. Ingrese su propia distribución previa para el modelo Poisson	62
4-12. Formas de ingresar datos en el panel para el modelo Poisson	63
4-13. Formato para ingresar datos del modelo Poisson	63
4-14. Medidas de resumen suficientes del modelo Poisson	63
4-15. Intervalos y gráfica de intervalos del modelo Poisson	64
4-16. Selección de la distribución previa para el modelo Geométrico	65
4-17. Parámetros de la previa Beta conjugada para el modelo Geométrico . .	65
4-18. Ingrese su propia distribución previa para el modelo Geométrico	66
4-19. Formas de ingresar datos en el panel para el modelo Geométrico	66
4-20. Formato para ingresar datos del modelo Geométrico	66
4-21. Medidas de resumen suficientes del modelo Geométrico	67
4-22. Intervalos y gráfica de intervalos del modelo Geométrico	67
4-23. Selección de la distribución previa para el modelo Exponencial	68

4-24. Parámetros de la previa Gamma conjugada para el modelo Exponencial	68
4-25. Ingrese su propia distribución previa para el modelo Exponencial . . .	68
4-26. Formas de ingresar datos en el panel para el modelo Exponencial . . .	69
4-27. Formato para ingresar datos del modelo Exponencial	69
4-28. Medidas de resumen suficientes del modelo Exponencial	69
4-29. Intervalos y gráfica de intervalos del modelo Exponencial	70
4-30. Inferencia en el modelo Normal	71
4-31. Selección de la distribución previa para el modelo Normal en el caso de μ con σ_0 conocido	71
4-32. Parámetros de la previa Normal conjugada para el modelo Normal el caso de μ con σ_0 conocido	72
4-33. Ingrese su propia distribución previa para el modelo Normal el caso de μ con σ_0 conocido	72
4-34. Formas de ingresar datos en el panel para el modelo Normal el caso de μ con σ_0 conocido	72
4-35. Formato para ingresar datos del modelo Normal el caso de μ con σ_0 conocido	73

4-36. Medidas de resumen suficientes del modelo Normal el caso de μ con σ_0 conocido	73
4-37. Intervalos y gráfica de intervalos del modelo Normal el caso de μ con σ_0 conocido	74
4-38. Selección de la distribución previa para el modelo Normal el caso de σ^2 con μ_0 conocido	74
4-39. Parámetros de la previa Invgamma conjugada para el modelo Normal el caso de σ^2 con μ_0 conocido	75
4-40. Formas de ingresar datos en el panel para el modelo Normal el caso de σ^2 con μ_0 conocido	75
4-41. Formato para ingresar datos del modelo Normal el caso de σ^2 con μ_0 conocido	75
4-42. Medidas de resumen suficientes del modelo Normal el caso de σ^2 con μ_0 conocido	76
4-43. Intervalos y gráfica de intervalos del modelo Normal el caso de σ^2 con μ_0 conocido	76
4-44. Selección de la distribución previa para el modelo Normal el caso de μ con σ^2 desconocido	77
4-45. Parámetros de la previa conjugada para el modelo Normal el caso de μ con σ^2 desconocido	77

4-46. Formas de ingresar datos en el panel para el modelo Normal el caso de μ con σ^2 desconocido	78
4-47. Formato para ingresar datos del modelo Normal el caso de μ con σ^2 desconocido	78
4-48. Medidas de resumen suficientes del modelo Normal el caso de μ con σ^2 desconocido	78
4-49. Intervalos y gráfica de intervalos del modelo Normal el caso de μ con σ^2 desconocido	79
4-50. Menú de distribuciones muestrales para encontrar el tamaño de muestra necesario con el cual la estimación por intervalos frecuentista y bayesiano coincida dentro de un error predeterminado	80
4-51. Selección de la distribución previa para el modelo Binomial	80
4-52. Parámetros de la previa Beta conjugada para el modelo Binomial . . .	81
4-53. Ingrese su propia distribución previa para el modelo Binomial	81
4-54. Parámetro de la distribución Binomial y distancia requerida para aproximar los intervalos	81
4-55. Región de búsqueda para el tamaño de muestra del modelo Binomial . . .	81
4-56. Número de simulaciones en el modelo Binomial	82
4-57. Forma de estimar la distancia simulada en el modelo Binomial	82

4-58.Ingresar el valor del cuantil para estimar la distancia en el modelo Binomial	82
4-59.Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Binomial	83
4-60.Selección de la distribución previa para el modelo Poisson	83
4-61.Parámetros de la previa Gamma conjugada para el modelo Poisson . .	84
4-62.Ingrese su propia distribución previa para el modelo Poisson	84
4-63.Parmetro de la distribucin Poisson y distancia requerida para aproximar los intervalos	84
4-64.Región de búsqueda para el tamaño de muestra del modelo Poisson . .	85
4-65.Número de simulaciones en el modelo Poisson	85
4-66.Forma de estimar la distancia simulada en el modelo Poisson	85
4-67.Ingresar el valor del cuantil para estimar la distancia en el modelo Poisson	86
4-68.Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Poisson	86
4-69.Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Geométrico	86
4-70.Parámetros de la previa Beta conjugada para el modelo Geométrico . .	87

4-71.Ingrese su propia distribución previa para el modelo Geométrico . . .	87
4-72.Parametro de la distribucion Geométrica y distancia requerida para aproximar los intervalos	87
4-73.Región de búsqueda para el tamaño de muestra del modelo Geométrico	88
4-74.Número de simulaciones en el modelo Geométrico	88
4-75.Forma de estimar la distancia simulada en el modelo Geométrico . . .	88
4-76.Ingresar el valor del cuantil para estimar la distancia en el modelo Geométrico	89
4-77.Tamaño de muestra encontrado para aproximar los intervalos de con- fianza y de credibilidad en el modelo Geométrico	89
4-78.Selección de la distribución previa para el modelo Exponencial	89
4-79.Parámetros de la previa Gamma conjugada para el modelo Exponen- cial	90
4-80.Ingrese su propia distribución previa para el modelo Exponencial . .	90
4-81.Parametro de la distribucion Exponencial y distancia requerida para aproximar los intervalos	90
4-82.Región de búsqueda para el tamaño de muestra del modelo Exponencial	91
4-83.Número de simulaciones en el modelo Exponencial	91
4-84.Forma de estimar la distancia simulada en el modelo Exponencial . .	91

4-85.Ingresar el valor del cuantil para estimar la distancia en el modelo Exponencial	92
4-86.Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Exponencial	92
4-87.Inferencia en el modelo Normal	92
4-88.Selección de la distribución previa para el modelo Normal en el caso de μ con σ_0 conocido	93
4-89.Parámetros de la previa Normal conjugada para el modelo Normal el caso de μ con σ_0 conocido	93
4-90.Ingrese su propia distribución previa para el modelo Normal el caso de μ con σ_0 conocido	93
4-91.Parámetros de la distribución Normal y distancia requerida para aproximar los intervalos en el caso de μ con σ_0 conocido	93
4-92.Región de búsqueda para el tamaño de muestra del modelo Normal en el caso de μ con σ_0 conocido	94
4-93.Número de simulaciones en el modelo Normal en el caso de μ con σ_0 conocido	94
4-94.Forma de estimar la distancia simulada en el modelo Normal en el caso de μ con σ_0 conocido	94
4-95.Ingresar el valor del cuantil para estimar la distancia en el modelo Normal en el caso de μ con σ_0 conocido	95

4-96.Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Normal en el caso de μ con σ_0 conocido	95
4-97.Selección de la distribución previa para el modelo Normal en el caso de σ^2 con μ_0 conocido	96
4-98.Parámetros de la previa Invgamma conjugada para el modelo Normal el caso de σ^2 con μ_0 conocido	96
4-99.Parámetros de la distribución Invgamma y distancia requerida para aproximar los intervalos en el caso de σ^2 con μ_0 conocido	96
4-100.Región de búsqueda para el tamaño de muestra del modelo Normal en el caso de σ^2 con μ_0 conocido	97
4-101.Número de simulaciones en el modelo Normal en el caso de σ^2 con μ_0 conocido	97
4-102.Forma de estimar la distancia simulada en el modelo Normal en el caso de σ^2 con μ_0 conocido	97
4-103.Ingresar el valor del cuantil para estimar la distancia en el modelo Normal en el caso de σ^2 con μ_0 conocido	98
4-104.Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Normal en el caso de σ^2 con μ_0 conocido	98

4-105	Selección de la distribución previa para el modelo Normal en el caso de μ con σ^2 desconocido	99
4-106	Parámetros de la previa conjugada para el modelo Normal el caso de μ con σ^2 desconocido	99
4-107	Parámetros de la distribución Normal y distancia requerida para aproximar los intervalos en el caso de μ con σ^2 desconocido	100
4-108	Región de búsqueda para el tamaño de muestra del modelo Normal en el caso de μ con σ^2 desconocido	100
4-109	Número de simulaciones en el modelo Normal en el caso de μ con σ^2 desconocido	100
4-110	Forma de estimar la distancia simulada en el modelo Normal en el caso de μ con σ desconocido	101
4-111	Ingresar el valor del cuantil para estimar la distancia en el modelo Normal en el caso de μ con σ^2 desconocido	101
4-112	Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Normal en el caso de μ con σ^2 desconocido	101

LISTA DE ABREVIATURAS

D_{50}	Distancia promedio.
n_{50}	Tamaño de muestra encontrado a partir de D_{50} .
D_{99}	Distancia cuantil 99.
n_{99}	Tamaño de muestra encontrado a partir de D_{99} .
D_R	Distancia requerida por un investigador.
$MCMC$	Simulación Monte Carlo.
$Invgamma$	Distribución Inversa Gamma.

LISTA DE SIMBOLOS

p	Parámetro de la distribución binomial y Geométrica.
λ	Parámetro de la distribución Poisson y Exponencial.
μ	Media de una distribución Normal.
σ	Desviación estándar de una distribución Normal.
θ_0	Parámetro conocido de un modelo.
θ_1	Parámetro de la distribución previa.
$\pi(\theta)$	Función de distribución previa.
$\pi(\theta X)$	Función de distribución a posteriori.
Γ^{-1}	función cuantil de la distribucion Gamma.
B^{-1}	función cuantil de la distribucion Beta.

Capítulo 1

INTRODUCCIÓN

La estimación de parámetros que caracterizan la distribución de una variable aleatoria es uno de los problemas más comunes que enfrenta la estadística. Una técnica que permite resolver este problema es el cálculo de intervalos. Esta técnica permite dar una región de posibles valores que se esperan de un parámetro, lo cual nos ayuda a observar la incertidumbre vinculada con los hallazgos sobre una estimación ([Cepeda,E.,et,al.2008](#)). Históricamente, la estadística se ha desarrollado a través de dos enfoques: Frecuentista y Bayesiano. El primer enfoque realiza intervalos de confianza y el segundo enfoque realiza intervalos de credibilidad. Los intervalos de confianza dependen de la estimación del parámetro obtenido a partir de la muestra, del tamaño de muestra y del nivel de confianza seleccionado. Los intervalos de credibilidad dependen directamente de la distribución a posteriori, la cual se obtiene de la actualización de una muestra observada con información subjetiva del parámetro (distribución previa). Dicha actualización se fundamenta en el teorema de Bayes.

El problema para un analista es determinar cuál de estos dos enfoques es preferible para realizar una estimación por intervalos. Para esto, se hace un estudio experimental vía simulación con el cual se pueda estimar el tamaño de muestra necesario tal que la distancia calculada (distancia rectilínea, ver Definición 2.2.8) entre los intervalos de confianza y de credibilidad sea menor o igual que una distancia requerida (distancia requerida por un analista, denotada por D_R). Esto, con el objetivo de que una vez encontrado el tamaño de muestra, no importando el enfoque

que se utilice, los dos den intervalos aproximadamente iguales. En este estudio se consideran los modelos: Binomial, Geométrica, Poisson, Exponencial y Normal.

Ejemplo 1.1. *Tamaño de muestra necesario para aproximar intervalos con una confianza y credibilidad del 95 %, para el parámetro $p = 0.3$ en un modelo Binomial.*

A continuación, se varía el tamaño de muestra ($n = 1, 2, 3, \dots$) del número de variables aleatorias Bernoulli con $p = 0.3$ hasta llegar a encontrar el tamaño de muestra adecuado con el cual la distancia calculada sobre los intervalos de confianza (Clopper-Pearson, ver Subsección 2.1.1) e intervalos de credibilidad (con la distribución previa Conjugada Beta(3,3), ver Ejemplo 2.9) sea menor o igual a una distancia requerida $D_R = 0.1, 0.05, 0.01$.

El proceso para llevar a cabo este estudio es el siguiente:

- 1. se fija n (número de ensayos Bernoulli) y se generan 1000 variables aleatorias de la distribución Binomial para n y $p = 0.3$ utilizando la función de R, $x = \text{rbinom}(1000, n, 0.3)$, luego para cada variable aleatoria generada x , se calculan los intervalos de confianza y los intervalos de credibilidad ($IF_x = (LF_x, HF_x)$ y $IB_x = (LB_x, HB_x)$ respectivamente), la distancia para cada par de intervalos $D_x = |LF_x - LB_x| + |HF_x - HB_x|$ y por último se calcula la distancia D_{50} y la distancia D_{99} de las 1000 distancias obtenidas.*

Un ejemplo numérico es: Sea $n = 10$ fijo y $x = 3$ una variable aleatoria generada, entonces los intervalos con un 95 % de confianza y de credibilidad son (ver código en el apéndice A.):

Intervalo de confianza : (0.0667, 0.6525)

Intervalo de credibilidad : (0.0927, 0.6058)

La distancia de estos dos intervalos es:

$$\begin{aligned} D_3 &= |0.0667 - 0.0927| + |0.6525 - 0.6058| \\ &= 0.0727. \end{aligned}$$

Se hace lo anterior, para $n=10$ generando 1000 variables aleatorias x , obteniendo 1000 distancias D_x y en seguida D_{50} y D_{99} .

A continuación se muestran una tabla con 1000 distancias D_x , un histograma de estas distancias, la distancia D_{50} y la distancia D_{99} (ver código en el apéndice A.).

x	D_x
3	0.1330
1	0.1113
2	0.0981
0	0.1394
7	0.1330
1	0.1113
5	0.1576
4	0.1516
3	0.1330
4	0.1516
3	0.1330
6	0.1516
2	0.0981
5	0.1576
5	0.1576
'data.frame': 1000 obs. of 2 variables...	
D_{50}	D_{99}
0.1292484	0.1576000

Tabla 1–1: D_{50} , D_{99} y 1000 distancias obtenidas para el modelo Binomial con $n = 10$ y $p = 0.3$

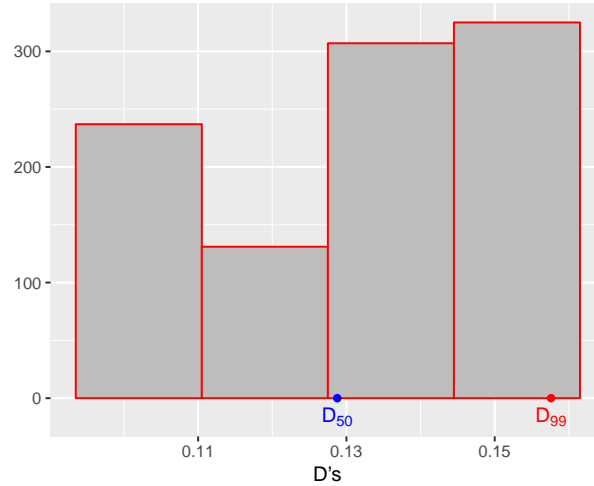


Figura 1-1: Histograma de las 1000 distancias simuladas para $n = 10$ y $p = 0.3$, ubicando $D_{50} = 0.1292484$ y $D_{99} = 0.1576000$.

2. Se repite el proceso realizado en el paso 1. para cada valor de $n = 1, 2, 3, \dots$, fijando una distancia requerida $D_R = 0.1, 0.05, 0.01$, hasta encontrar el valor de n con el cual D_{50} o D_{99} sea menor o igual a D_R .

Las Figuras 1-2 y 1-3, muestran el comportamiento de la distancia D_{50} y D_{99} cuando se varia el tamaño de muestra $n = 1, 2, 3, \dots$ y también exponen los tamaño de muestra encontrados, tal que D_{50} o D_{99} son menores o igual a $D_R = 0.1, 0.05, 0.01$. Note en la Figura 1-2, que para $p = 0.3$, los tamaños de muestra necesarios para que la distancia D_{50} sea menor o igual que una distancia requerida ($D_R = 0.1, 0.05, 0.01$) son respectivamente $n = 14, 32$ y 182 . Es decir, que si un analista tiene una sospecha de que el parámetro p de un modelo binomial es 0.3 y toma una muestra de tamaño 182 , no importa si calcula intervalos de confianza o intervalos de credibilidad, Hay un chance aproximadamente del 50% de que los dos son casi iguales excepto por una distancia $D_R = 0.01$. Un análisis similar se puede realizar para la Figura 1-3.

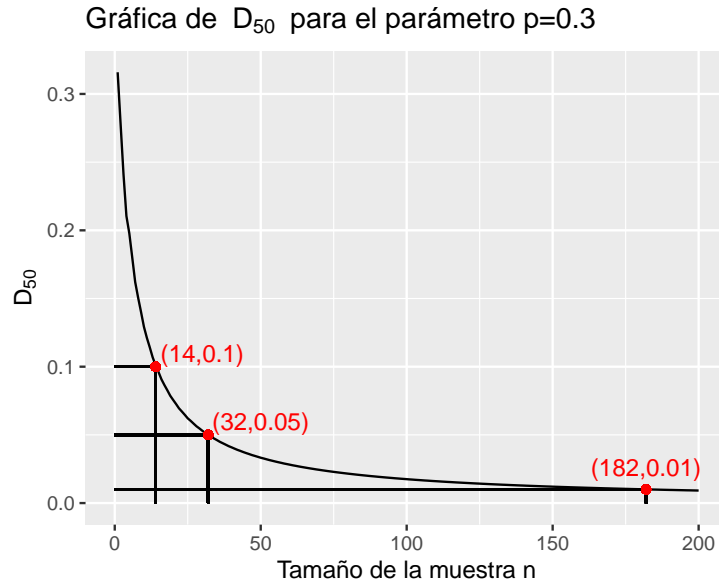


Figura 1-2: Comportamiento de la distancia D_{50} y tamaños de muestra n con el cual D_{50} es menor o igual a D_R cuando $p = 0.3$.

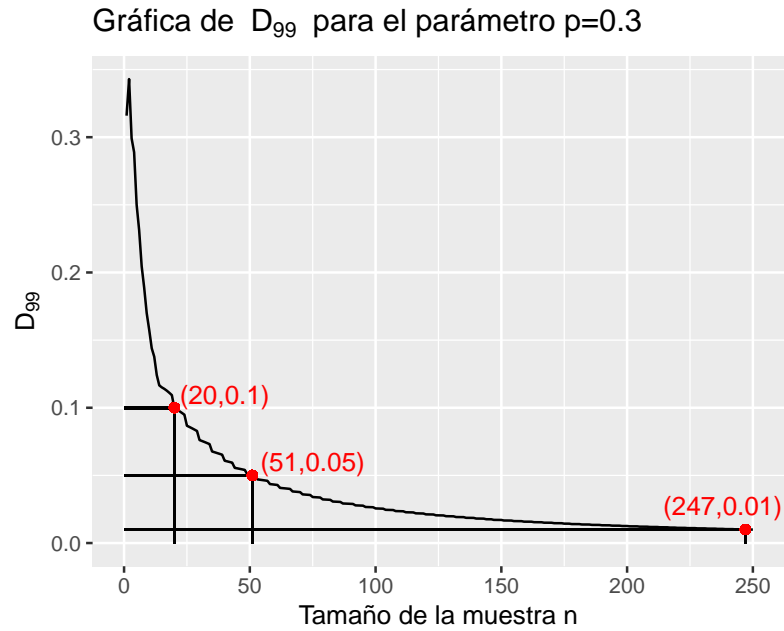


Figura 1-3: Comportamiento de la distancia D_{99} y tamaños de muestra n con el cual D_{99} es menor o igual a D_R cuando $p = 0.3$.

Capítulo 2

INTERVALOS

En este capítulo se estudian los intervalos de confianza y los intervalos de credibilidad que se utilizan para llevar a cabo el estudio experimental descrito en el Capítulo 1 para los modelos: Binomial, Geométrica, Poisson, Exponencial y Normal.

2.1. Intervalos de confianza

En el enfoque frecuentista se considera un parámetro θ , una cantidad fija, aunque desconocida. Un intervalo de confianza de un parámetro de valor real θ es cualquier par de funciones $L(x_1, \dots, x_n)$ y $U(x_1, \dots, x_n)$ de una muestra aleatoria que satisfaga $L(\mathbf{x}) \leq U(\mathbf{x})$, para cualquier valor muestral \mathbf{x} observado. Es decir, si $\mathbf{X} = \mathbf{x}$ es observado, la inferencia que se hace es: $L(\mathbf{x}) \leq \theta \leq U(\mathbf{x})$. El intervalo aleatorio $[L(\mathbf{X}), U(\mathbf{X})]$ se llama intervalo de confianza y la probabilidad de que este contenga θ es de $(1 - \alpha)100\%$. ([Casella and Berger. 1990](#)).

Debido a la variación de la muestra, un intervalo de confianza puede no contener el parámetro θ . Un intervalo de confianza del 95 % significa que si se toma una gran cantidad de muestras y se construyen los intervalos de confianza correspondientes, entonces aproximadamente el 95 % de los intervalos contendrá el parámetro θ . El valor del 95 % definido anteriormente es el nivel de confianza y se denota por $(1 - \alpha)100\%$.

Algunos métodos para encontrar intervalos de confianza son: invertir una prueba de hipótesis, cantidades pivotantes y pivotear una función de distribución acumulada (CDF) ([Casella and Berger. 1990](#)).

2.1.1. Intervalos de confianza para el parámetro p de un modelo Binomial.

El intervalo de confianza que se considera para el parámetro p del modelo Binomial es el intervalo de Clopper-Pearson (Clopper and Pearson. 1934). Este intervalo se construye utilizando el método “Pivoteando una CDF” (Casella and Berger. 1990).

Sean X_1, \dots, X_n variables aleatorias i.i.d. de una población Bernoulli con parámetro p , se define $Y = \sum_{i=1}^n X_i$, donde Y es un estadístico suficiente para p y $Y \sim \text{Binomial}(n, p)$. Al Utilizar el teorema 9.2.14 (Casella and Berger. 1990), si $Y = y$ es observado, entonces el intervalo es de la forma $\{p : p_1 \leq p \leq p_2\}$ con $P(p_1 \leq p \leq p_2) = 1 - \alpha$, donde p_1 y p_2 se obtienen solucionando las siguientes ecuaciones:

$$\sum_{k=0}^y \binom{n}{k} p_2^k (1 - p_2)^{n-k} = \frac{\alpha}{2} \quad \text{y} \quad \sum_{k=y}^n \binom{n}{k} p_1^k (1 - p_1)^{n-k} = \frac{\alpha}{2}. \quad (2.1)$$

Sean $W_1, W_2, \dots, W_n \stackrel{i.i.d.}{\sim} U(0, 1)$ y Y el número de W_i 's menores a p , entonces $Y \sim \text{Binomial}(n, p)$ y el estadístico del orden $W \equiv W_{(y)} \sim \text{Beta}(y, n - y + 1)$ (Casella and Berger. 1990), donde los eventos $Y \geq y$ y $W < p$ son equivalentes. Es decir $P(Y \geq y|p) = P(W < p)$, luego $\alpha/2 = P(Y \geq y|p) = P(W < p)$. Por lo tanto, el límite inferior p_1 es el cuantil $\alpha/2$ de una distribución $\text{Beta}(y, n - y + 1)$. De forma análoga, p_2 es el cuantil $1 - \alpha/2$ de una distribución $\text{Beta}(y + 1, n - y)$ (Lawrence,M.,et,al.2012).

2.1.2. Intervalos de confianza para el parámetro λ de un modelo Poisson.

El intervalo de confianza para el parámetro λ del modelo Poisson es el intervalo de Garwood (Garwood. 1936). Este intervalo se construye utilizando el método “Pivotenado una CDF” (Casella and Berger. 1990).

Sean X_1, \dots, X_n variables aleatorias i.i.d. de una población Poisson con parámetro λ , se define $Y = \sum_{i=1}^n X_i$, donde Y es un estadístico suficiente para λ y $Y \sim \text{Poisson}(n\lambda)$. Al utilizar el teorema 9.2.14 (Casella and Berger. 1990), si $Y = y$

es observado, entonces el intervalo de confianza para λ se obtienen solucionando las siguientes ecuaciones para λ :

$$\sum_{k=0}^y e^{-n\lambda} \frac{(n\lambda)^k}{k!} = \frac{\alpha}{2} \quad \text{y} \quad \sum_{k=y}^{\infty} e^{-n\lambda} \frac{(n\lambda)^k}{k!} = \frac{\alpha}{2}. \quad (2.2)$$

Al Utilizar la relación que existe entre la distribución Gamma y Poisson, se tiene: Si $X \sim \text{Gamma}(\alpha, \beta)$ y $Y \sim \text{Poisson}(x/\beta)$ entonces, $P(X \leq x) = P(Y \geq \alpha)$, usando $\alpha = y + 1$, $\beta = 1$, $x = n\lambda$ tenemos:

$$\begin{aligned} \frac{\alpha}{2} &= P(Y \leq y|\lambda) = P(Y < y + 1|\lambda) \\ &= 1 - P(Y \geq y + 1) = 1 - P(X \leq n\lambda) \\ &= P(X > n\lambda) \end{aligned} \quad (2.3)$$

Así, el límite superior del intervalo para el parámetro λ , es:

$$\lambda = \Gamma^{-1}(\alpha/2, y + 1, 1)/n \quad (2.4)$$

de forma similar se calcula el límite inferior. Por lo tanto, el intervalo de confianza es de la forma:

$$\{\lambda : \Gamma^{-1}(1 - \alpha/2, y, 1)/n \leq \lambda \leq \Gamma^{-1}(\alpha/2, y + 1, 1)/n\} \quad (2.5)$$

o equivalente

$$\{\lambda : qchisq(1 - \alpha/2, 2y)/(2n) \leq \lambda \leq qchisq(\alpha/2, 2(y + 1))/(2n)\} \quad (2.6)$$

2.1.3. Intervalos de confianza para el parámetro p de un modelo Geométrico.

Sean X_1, \dots, X_n variables aleatorias i.i.d. de una población Geométrica con parámetro p , por teorema del límite central, \bar{X} se aproxima asintóticamente a una distribución $\text{Normal}(\mu, \sigma^2)$, donde $\mu = \frac{1}{p}$ y $\sigma^2 = \frac{1-p}{np^2}$.

Sea $g(\bar{X}) = \frac{1}{\bar{X}}$ el estimador de p . Al utilizar el método Delta ([Agresti. 2002](#)), entonces $g(\bar{X}) - g(\mu)$ se distribuye asintóticamente $\text{Normal}(0, \sigma^2[g'(\bar{X})]^2)$, así $\sigma^2[g'(\bar{X})]^2 = \frac{(1-\hat{p})\hat{p}^2}{n}$ y $\frac{g(\bar{X}) - g(\mu)}{\sigma[g'(\bar{X})]}$ tiene asintóticamente una distribución normal estándar y es una cantidad pivotal. por lo tanto, el intervalo de confianza para p es de la forma:

$$\left\{ p : \hat{p} - z_{\alpha/2} \hat{p} \sqrt{\frac{1-\hat{p}}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \hat{p} \sqrt{\frac{1-\hat{p}}{n}} \right\} \quad (2.7)$$

2.1.4. Intervalo de confianza para el parámetro λ de un modelo Exponencial.

Sean X_1, \dots, X_n variables aleatorias i.i.d. de una población Exponencial con parámetro λ , entonces $Y = \sum_{i=1}^n X_i$ es un estadístico suficiente para λ y $Y \sim \text{Gamma}(n, \lambda)$, luego $\lambda Y \sim \text{Gamma}(n, 1)$ es una cantidad pivotal y por lo tanto, si $Y = y$ una observación, el intervalo de confianza es de la forma:

$$\left\{ \lambda : \Gamma^{-1}(1 - \alpha, n, 1)/y \leq \lambda \leq \Gamma^{-1}(\alpha, n, 1)/y \right\} \quad (2.8)$$

2.1.5. Intervalos de confianza para los parámetros μ, σ de un modelo Normal.

Sean X_1, \dots, X_n variables aleatorias i.i.d. de una población Normal con parámetros μ y σ^2 , los intervalos para μ y σ^2 se obtienen de la siguiente manera:

Intervalo para μ

Sea $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ un estimador suficiente para μ y $\bar{X} \sim \text{Normal}(\mu, \sigma^2)$. En el caso en que σ^2 es conocido, tenemos que $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1)$ es una cantidad pivotal y

por lo tanto, si $\bar{X} = \bar{x}$ es un promedio observado, el intervalo de confianza es de la forma:

$$\left\{ \mu : \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\} \quad (2.9)$$

En el caso en que σ^2 es desconocido (σ^2 se aproxima por $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$), tenemos que $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ es una cantidad pivotal y por lo tanto, si $\bar{X} = \bar{x}$ es un promedio observado, el intervalo de confianza es:

$$\left\{ \mu : \bar{x} - t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{1-\alpha/2, n-1} \frac{s}{\sqrt{n}} \right\} \quad (2.10)$$

Intervalo para σ^2

En el caso en que μ es conocido, tenemos que $\frac{nS^2}{\sigma^2} \sim \chi_n^2$ es una cantidad pivotal, donde $S^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$. Por lo tanto, si $S^2 = s^2$ una varianza observada, el intervalo de confianza es de la forma:

$$\left\{ \sigma^2 : \frac{ns^2}{\chi_{n, \alpha/2}^2} \leq \sigma^2 \leq \frac{ns^2}{\chi_{n, 1-\alpha/2}^2} \right\} \quad (2.11)$$

2.2. Metodología Bayesiana para la estimación por Intervalos

En el enfoque Bayesiano, se considera un parámetro θ una cantidad aleatoria, cuya variación puede ser descrita por una distribución de probabilidad (llamada distribución previa). Esta es una distribución subjetiva basada en las creencias de un analista y se define antes de observar los datos. Luego se toma una muestra de una población con parámetro θ y la distribución previa se actualiza con la información de esta muestra. La previa actualizada se llama distribución a posteriori. La distribución a posteriori está fundamentada en el teorema de Bayes ([Casella and Berger. 1990](#)).

La distribución previa se define por $\pi(\theta)$ y la distribución muestral por $f(\mathbf{x}|\theta)$, entonces la distribución condicionada de θ dada una muestra \mathbf{x} es:

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}, \theta)}{m(\mathbf{x})} = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{m(\mathbf{x})}. \quad (2.12)$$

Donde $\pi(\theta|\mathbf{x})$ es la distribución a posteriori y $m(\mathbf{x})$ es la distribución marginal de \mathbf{X} , definida por:

$$m(\mathbf{x}) = \int f(\mathbf{x}|\theta)\pi(\theta)d\theta. \quad (2.13)$$

También, podemos escribir la ecuación (2.1) en términos de la función de máxima verosimilitud:

$$\pi(\theta|x_1, \dots, x_n) = \frac{L(\theta|x_1, \dots, x_n)\pi(\theta)}{m(\mathbf{x})}, \quad m(\mathbf{x}) = \int L(\theta|x_1, \dots, x_n)\pi(\theta)d\theta \quad (2.14)$$

donde $L(\theta|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta)$.

La distribución a posteriori se utiliza para hacer inferencia sobre θ . Por ejemplo, la media de la distribución a posteriori se utiliza como estimador puntual de θ . En nuestro caso, la distribución a posteriori se utiliza para construir intervalos de credibilidad.

2.2.1. Selección de la distribución previa.

En el enfoque Bayesiano, es importante determinar una distribución previa para el parámetro de interés, sobre el que se quiere realizar inferencia.

En esta sección, se mencionan dos tipos de distribuciones previas sobre los parámetros de los modelos que vamos a estudiar, donde estas distribuciones previas son: previa de Jeffrey y previa Conjugada.

Previa de Jeffrey.

En situaciones en donde no hay información previa acerca de los parámetros de un modelo, se define algún tipo de distribución previa no informativa. Es decir, una

distribución previa que intenta no dar información sobre el parámetro de interés. A continuación, definiremos una de las previas no informativas más utilizadas en la inferencia bayesiana ([Robert and Casella.1998](#)).

Definición 2.2.2. *Para una distribución muestral $f(\mathbf{x}|\theta)$, la previa de Jeffrey tiene una densidad proporcional a $\sqrt{I(\theta)}$, donde $I(\theta)$ es la información de Fisher, que está dada por:*

$$I(\theta) = E_{\theta} \left[\frac{d}{d\theta} \log f(\mathbf{X}|\theta) \right]^2 = -E_{\theta} \left[\frac{d^2}{d\theta^2} \log f(\mathbf{X}|\theta) \right]. \quad (2.15)$$

La propiedad más importante de la distribución previa de Jeffrey es la invarianza bajo transformaciones monótonas, ya que dada una transformación $h(\theta)$ monótona y diferenciable de un parámetro escalar θ , se tiene que: $\pi(h(\theta)) = \frac{\pi(\theta)}{|h'(\theta)|}$ ([Jeffreys. 1961](#)). Es decir, que la inferencia no depende de la escala elegida para el parámetro ([Berger. 1985](#)).

En los siguientes ejemplos, se obtienen las distribuciones previas de Jeffrey y la distribución a posteriori correspondiente para cada uno de los modelos que se estudian.

Ejemplo 2.1. *La distribución previa de Jeffrey para el parámetro λ de un modelo Poisson se obtiene utilizando la ecuación (2.4):*

La información de Fisher es:

$$\begin{aligned} I(\lambda) &= -E_{\lambda} \left(\frac{d^2}{d\lambda^2} \log \left(e^{-\lambda} \frac{\lambda^X}{X!} \right) \right) \\ &= -E_{\lambda} \left(\frac{-X}{\lambda^2} \right) \\ &= \frac{1}{\lambda}. \end{aligned} \quad (2.16)$$

Así, $\pi(\lambda) \propto \frac{1}{\sqrt{\lambda}}$

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Poisson, la función de máxima verosimilitud

$$L(\lambda|\mathbf{x}) \propto \lambda^{\sum_{i=1}^n x_i} \exp(-n\lambda). \quad (2.17)$$

Así la distribución a posteriori es:

$$\pi(\lambda|\mathbf{x}) \propto \exp(-n\lambda) \lambda^{\sum_{i=1}^n x_i - \frac{1}{2}} \quad (2.18)$$

luego $\lambda|\mathbf{x} \sim \text{Gamma}(\sum_{i=1}^n x_i + \frac{1}{2}, n)$

Ejemplo 2.2. Distribución previa de Jeffrey para el parámetro λ de un modelo Exponencial.

La información de Fisher es:

$$\begin{aligned} I(\lambda) &= -E_{\lambda} \left(\frac{d^2}{d\lambda^2} \log(\lambda e^{-\lambda X}) \right) \\ &= -E_{\lambda} \left(-\frac{1}{\lambda^2} \right) \\ &= \frac{1}{\lambda^2} \end{aligned} \quad (2.19)$$

Así, $\pi(\lambda) \propto \frac{1}{\lambda}$.

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Exponencial, la función de máxima verosimilitud

$$L(\lambda|\mathbf{x}) \propto \lambda^n \exp \left(-\lambda \sum_{i=1}^n x_i \right) \quad (2.20)$$

Así, la distribución a posteriori es:

$$\pi(\lambda|\mathbf{x}) \propto \lambda^{n-1} \exp \left(-\lambda \sum_{i=1}^n x_i \right) \quad (2.21)$$

luego $\lambda|\mathbf{x} \sim \text{Gamma}(n, \sum_{i=1}^n x_i)$

Ejemplo 2.3. Distribución previa de Jeffrey para el parámetro p de un modelo Bernoulli.

La información de Fisher es:

$$\begin{aligned}
 I(p) &= -E_p \left(\frac{d^2}{dp^2} \log (p^X (1-p)^{1-X}) \right) \\
 &= -E_p \left(\frac{X}{p} - \frac{1-X}{1-p} \right) \\
 &= \frac{1}{p(1-p)}
 \end{aligned} \tag{2.22}$$

Así, $\pi(p) = \sqrt{\frac{1}{p(1-p)}}$, es decir $p \sim \text{Beta}(1/2, 1/2)$.

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Bernoulli, la función de máxima verosimilitud es:

$$L(p|\mathbf{x}) = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \tag{2.23}$$

Así, la distribución a posteriori es:

$$\pi(p|\mathbf{x}) \propto p^{\sum_{i=1}^n x_i + \frac{1}{2} - 1} (1-p)^{n - \sum_{i=1}^n x_i + \frac{1}{2} - 1} \tag{2.24}$$

luego $p|\mathbf{x} \sim \text{Beta}(\sum_{i=1}^n x_i + \frac{1}{2}, n - \sum_{i=1}^n x_i + \frac{1}{2})$

Ejemplo 2.4. Distribución previa de Jeffrey para el parámetro p de un modelo Geométrico.

La información de Fisher es:

$$\begin{aligned}
 I(\sigma) &= -E_\sigma \left(\frac{d^2}{d\sigma^2} \log ((1-p)^{X-1} p) \right) \\
 &= -E_\sigma \left(-\frac{(X-1)}{(1-p)^2} - \frac{1}{p^2} \right) \\
 &= p^{-2} (1-p)^{-1/2}
 \end{aligned} \tag{2.25}$$

Así $\pi(p) = \frac{1}{p(1-p)^{1/2}}$.

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Geométrica, la función de máxima verosimilitud es:

$$L(p|\mathbf{x}) = p^n (1-p)^{\sum_{i=1}^n x_i - n} \tag{2.26}$$

Así, la distribución a posteriori es:

$$\pi(p|\mathbf{x}) \propto p^{n-1}(1-p)^{\sum_{i=1}^n x_i - n + \frac{1}{2} - 1} \quad (2.27)$$

luego $p|\mathbf{x} \sim \text{Beta}\left(n, \sum_{i=1}^n x_i - n + \frac{1}{2}\right)$

Ejemplo 2.5. Distribución previa de Jeffrey para el parámetro μ de un modelo Normal, en el caso de σ_0 conocido.

La información de Fisher es:

$$\begin{aligned} I(\mu) &= -E_\mu \left(\frac{d^2}{d\mu^2} \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(X-\mu)^2}{\sigma^2}} \right) \right) \\ &= -E_\mu \left(\frac{d^2}{d\mu^2} \left(-\log(\sqrt{2\pi\sigma^2}) - \log \left(\frac{1}{2\sigma^2} (X-\mu)^2 \right) \right) \right) \\ &= \frac{1}{\sigma^2} \end{aligned} \quad (2.28)$$

Así, $\pi(\mu) = \text{constante} \propto 1$.

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Normal, en el caso de σ_0 conocido, la función de máxima verosimilitud es:

$$\begin{aligned} L(\mu|\mathbf{x}) &\propto \exp \left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2 \right) \\ &\propto \exp \left(-\frac{n}{2\sigma_0^2} (\bar{x} - \mu)^2 \right) \end{aligned} \quad (2.29)$$

Así, la distribución a posteriori es:

$$\pi(\mu|\mathbf{x}) \propto \exp \left(-\frac{n}{2\sigma_0^2} (\bar{x} - \mu)^2 \right) \quad (2.30)$$

luego $\mu|\mathbf{x} \sim \text{Normal} \left(\bar{x}, \frac{\sigma_0}{\sqrt{n}} \right)$

Ejemplo 2.6. *Distribución previa de Jeffrey para el parámetro σ^2 de un modelo Normal, en el caso de μ_0 conocido.*

La información de Fisher es:

$$\begin{aligned}
 I(\sigma) &= -E_\sigma \left(\frac{d^2}{d\sigma^2} \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(X-\mu)^2}{\sigma^2}} \right) \right) \\
 &= -E_\sigma \left(\frac{d^2}{d\sigma^2} \left(-\log \left(\sqrt{2\pi\sigma^2} \right) - \log \left(\frac{1}{2\sigma^2} (X-\mu)^2 \right) \right) \right) \\
 &= -E_\sigma \left(-\frac{3}{\sigma^4} (X-\mu)^2 + \frac{1}{\sigma^2} \right) \\
 &= \frac{2}{\sigma^2}
 \end{aligned} \tag{2.31}$$

Así, $\pi(\sigma) = \sqrt{\frac{2}{\sigma^2}} \propto \frac{1}{\sigma}$ y, aplicando la propiedad de invarianza bajo transformaciones monótonas de la distribución de Jeffrey, se tiene que $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$.

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Normal, en el caso de μ_0 conocido y σ^2 desconocido, la función de máxima verosimilitud es:

$$\begin{aligned}
 L(\sigma^2|\mathbf{x}) &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2 \right) \\
 &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left(-\frac{ns^2}{2\sigma^2} \right)
 \end{aligned} \tag{2.32}$$

donde $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$.

Así la distribución a posteriori es:

$$\pi(\sigma^2|\mathbf{x}) \propto (\sigma^2)^{-\frac{n}{2}-1} \exp \left(-\frac{ns^2}{2\sigma^2} \right) \tag{2.33}$$

luego $\sigma^2|\mathbf{x} \sim \text{Invgamma} \left(\frac{n}{2}, \frac{ns^2}{2} \right)$

Ejemplo 2.7. *La distribución previa no informativa para el modelo Normal en el caso en que μ y σ^2 son desconocidos, es la distribución conjunta donde supondremos que la distribución de cada parámetro son independientes ($\pi(\mu, \sigma^2) = \pi(\mu)\pi(\sigma^2)$), donde $\pi(\mu) \propto 1$ (Distribución de Jeffrey del Ejemplo 2.5) y $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$ (Distribución*

de Jeffrey del Ejemplo 2.6). Así obtenemos la distribución conjunta sobre μ y σ^2 :

$$\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \quad (2.34)$$

Esta previa no informativa fue la que recomendó finalmente Jeffrey (1961) (James Berger 1980).

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Normal, la función de máxima verosimilitud es:

$$\begin{aligned} L(\mu, \sigma^2 | \mathbf{x}) &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp \left(-\frac{1}{2\sigma^2} ((n-1)s^2 + n(\bar{x} - \mu)^2) \right) \end{aligned} \quad (2.35)$$

donde $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

Así, la distribución a posteriori es:

$$\begin{aligned} \pi(\mu, \sigma^2 | \mathbf{x}) &\propto (\sigma^2)^{-\frac{n}{2}-1} \exp \left(-\frac{1}{2\sigma^2} ((n-1)s^2 + n(\bar{x} - \mu)^2) \right) \\ &\propto (\sigma^2)^{-\frac{n}{2}-1} \exp \left(-\frac{(n-1)s^2}{2\sigma^2} \right) \exp \left(-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right) \end{aligned} \quad (2.36)$$

luego $\mu, \sigma^2 | \mathbf{x} \sim \text{Normal} \left(\bar{x}, \frac{\sigma^2}{n} \right) \text{Invgamma} \left(\frac{n-1}{2}, \frac{(n-1)}{2} s^2 \right)$.

Previa conjugada

Cuando la información previa sobre un parámetro es bastante limitada, la distribución previa se elige de una familia de distribuciones conjugadas. La motivación principal para usar previas conjugadas es la manejabilidad de las distribuciones a posteriori. Sin embargo, tales elecciones pueden limitar la subjetividad del enfoque Bayesiano. (Robert and Casella.1998).

La idea de la distribución conjugada es que la distribución a posteriori tenga la misma forma algebraica que la distribución previa. Para formalizar esto, se hace uso del concepto de un estadístico suficiente.

Teorema 1. (*Factorización de Neyman*). Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria i.i.d. de una distribución con función de densidad de probabilidad $g(x|\boldsymbol{\theta})$. Se dice que el estadístico $T = u(X_1, \dots, X_n)$, es un estadístico suficiente para $\boldsymbol{\theta}$ si y solo si la función de verosimilitud puede factorizarse de la siguiente forma

$$L(\boldsymbol{\theta}|x_1, \dots, x_n) = f(t|\boldsymbol{\theta})h(x_1, \dots, x_n) \quad (2.37)$$

Para cualquier valor $t = u(x_1, \dots, x_n)$ de T en donde $h(x_1, \dots, x_n)$ no contiene el parámetro θ .

Si se elige una distribución previa $\pi(\theta)$ que tenga la misma estructura de $f(t|\boldsymbol{\theta})$, pensando en esta como función de $\boldsymbol{\theta}$ entonces la distribución a posteriori tendrá la misma forma. A continuación, una definición formal.

Definición 2.2.3. (*Bernardo and Smith. 1994*) La familia conjugada de densidades previas para $\boldsymbol{\theta}$ con respecto a la verosimilitud $L(\boldsymbol{\theta}|\mathbf{x})$ con estadístico suficiente $t = t(\mathbf{x}) = \{n, s(\mathbf{x})\}$ (de dimensión fija k , independiente de \mathbf{x}) es

$$\pi(\boldsymbol{\theta}) = f(\boldsymbol{\theta}|\tau = (\tau_0, \tau_1, \dots, \tau_k) \in \mathbf{I}) \quad (2.38)$$

donde

$$\mathbf{I} = \left\{ \tau : \int f(s = (\tau_1, \dots, \tau_k)|\boldsymbol{\theta}, n = \tau_0) d\boldsymbol{\theta} < \infty \right\} \quad (2.39)$$

y

$$f(\boldsymbol{\theta}|\tau) = \frac{f(s = (\tau_1, \dots, \tau_k)|\boldsymbol{\theta}, n = \tau_0)}{\int f(s = (\tau_1, \dots, \tau_k)|\boldsymbol{\theta}, n = \tau_0) d\boldsymbol{\theta}}. \quad (2.40)$$

Definición 2.2.4. Sea \mathcal{P} una familia de distribuciones previas $\pi(\boldsymbol{\theta})$, para $\boldsymbol{\theta}$ y \mathcal{F} una familia de distribuciones muestrales $f(x|\boldsymbol{\theta})$. Decimos que \mathcal{P} es una familia conjugada a la familia \mathcal{F} si la distribución a posteriori $\pi(\boldsymbol{\theta}|x) \in \mathcal{P}$.

La familia \mathcal{P} debe ser una familia paramétrica específica y suficientemente restringida. Como la mayoría de las distribuciones conocidas pertenecen a la familia exponencial, caracterizaremos las distribuciones previas conjugadas de forma general para los modelos de la familia exponencial (Robert and Casella.1998).

Definición 2.2.5. Sea \mathcal{F} una familia de distribuciones. Si los elementos de \mathcal{F} son de la forma:

$$f(x|\boldsymbol{\theta}) = c(\boldsymbol{\theta})h(x) \exp \left(\sum_{i=1}^k r_i(\boldsymbol{\theta})t_i(x) \right) \quad (2.41)$$

Donde $h(x)$ y $t_1(x), \dots, t_k(x)$ son funciones de valor real de la observación x y $c(\boldsymbol{\theta})$ y $r_1(\boldsymbol{\theta}), \dots, r_k(\boldsymbol{\theta})$ son funciones de valor real del vector de parámetros $\boldsymbol{\theta}$ ($\dim(\boldsymbol{\theta}) = k$). Entonces, decimos que \mathcal{F} es una familia exponencial y denotamos por \mathcal{F}_{exp} (Casella and Berger. 1990).

El siguiente teorema nos garantiza que si una densidad pertenece a una familia \mathcal{F}_{exp} , entonces existe una familia de distribuciones previa conjugada \mathcal{P} para la familia \mathcal{F}_{exp}

Teorema 2. Sea $f(x|\boldsymbol{\theta}) \in \mathcal{F}_{exp}$, entonces la distribución previa $\pi(\boldsymbol{\theta}) = f(\boldsymbol{\theta}|\tau)$ es una previa conjugada de la forma:

$$f(\boldsymbol{\theta}|\tau) = k(\tau)c(\boldsymbol{\theta})^{\tau_0} \exp \left(\sum_{i=1}^k r_i(\boldsymbol{\theta})\tau_i \right) \quad (2.42)$$

y si $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria de $f(X|\boldsymbol{\theta})$ entonces, la distribución a posteriori es: $\pi(\boldsymbol{\theta}|\mathbf{x}) = f(\boldsymbol{\theta}|\tau + t)$, donde $t(\mathbf{x}) = \{n, \sum_{j=1}^n s_1(x_j), \dots, \sum_{j=1}^n s_k(x_j)\}$ es el estadístico suficiente (Bernardo and Smith. 1994).

Demostración. Dada una muestra aleatoria $\mathbf{x} = (x_1, \dots, x_n)$ de $f(x|\boldsymbol{\theta})$, la función de verosimilitud es la forma $L(\boldsymbol{\theta}) \propto c(\boldsymbol{\theta})^{\tau_0+n} \exp\left(\sum_{i=1}^k r_i(\boldsymbol{\theta}) \sum_{j=1}^n s_i(x_j)\right)$. Por hipótesis, la distribución previa es $f(\boldsymbol{\theta}|\tau) = k(\tau)c(\boldsymbol{\theta})^{\tau_0} \exp\left(\sum_{i=1}^k r_i(\boldsymbol{\theta})\tau_i\right)$. Así, la distribución a posteriori es:

$$f(\boldsymbol{\theta}|\tau + t) = c(\boldsymbol{\theta})^{\tau_0+n} \exp\left(\sum_{i=1}^k r_i(\boldsymbol{\theta})(\tau_i + \sum_{j=1}^n s_i(x_j))\right) \quad (2.43)$$

Por lo tanto, $\pi(\boldsymbol{\theta}) = f(\boldsymbol{\theta}|\tau + t)$ y $t(\mathbf{X}) = \{n, \sum_{j=1}^n s_1(x_j), \dots, \sum_{j=1}^n s_k(x_j)\}$ es el estadístico suficiente \square

Ejemplo 2.8. Consideremos el modelo Poisson con parámetro λ , donde la función de densidad se puede escribir:

$$f(x|\lambda) \propto \exp(-\lambda) \exp(x \ln(\lambda)) \quad (2.44)$$

En términos de notación del Teorema 1. $r(\lambda) = \ln(\lambda)$ y $c(\lambda) = \exp(-\lambda)$, entonces la distribución previa conjugada a la distribución exponencial es:

$$\begin{aligned} \pi(\lambda) = f(\lambda|\tau) &\propto \exp(-\lambda\tau_0) \exp(\ln(\lambda)\tau_1) \\ &\propto \lambda^{\tau_1} \exp(-\lambda\tau_0) \end{aligned} \quad (2.45)$$

luego, $\lambda \sim \text{Gamma}(\alpha, \beta)$ donde, $\alpha = \tau_1 + 1$ y $\beta = \tau_0$.

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Poisson, la función de máxima verosimilitud

$$L(\lambda|\mathbf{x}) \propto \lambda^{\sum_{i=1}^n x_i} \exp(-n\lambda) \quad (2.46)$$

donde $t(\mathbf{x}) = \{n, \sum_{i=1}^n x_i\}$ es el estadístico suficiente. Así, la distribución a posteriori está dada por:

$$\begin{aligned}\pi(\lambda|\mathbf{x}) = f(\lambda|\tau + t) &\propto \exp(-\lambda)^{\tau_0+n} \exp(\ln(\lambda)(\tau_1 + \sum_{i=1}^n x_i)) \\ &\propto \lambda^{\tau_1 + \sum_{i=1}^n x_i} \exp(-\lambda(\tau_0 + n))\end{aligned}\quad (2.47)$$

Luego, $\lambda|\mathbf{x} \sim \text{Gamma}(\alpha + \sum_{i=1}^n x_i, \beta + n)$.

Ejemplo 2.9. Consideremos el modelo Binomial con parámetro p , donde la función de densidad puede escribirse:

$$f(x|p) = \binom{n}{x} (1-p)^{n-x} \exp\left(x \log\left(\frac{p}{1-p}\right)\right) \quad (2.48)$$

La distribución previa conjugada es:

$$\begin{aligned}\pi(p|\mathbf{x}) = f(p|\tau) &\propto (1-p)^{\tau_0} \exp\left(\tau_1 \log\left(\frac{p}{1-p}\right)\right) \\ &\propto p^{\tau_1} (1-p)^{\tau_0 - \tau_1}\end{aligned}\quad (2.49)$$

Luego, $p \sim \text{Beta}(\alpha, \beta)$, donde $\alpha = \tau_1 + 1$ y $\beta = \tau_0 - \tau_1 + 1$.

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Bernoulli, la función de máxima verosimilitud

$$L(p|\mathbf{x}) \propto (1-p)^n \exp\left(\sum_{i=1}^n x_i \log\left(\frac{p}{1-p}\right)\right) \quad (2.50)$$

donde $t(\mathbf{x}) = \{n, \sum_{i=1}^n x_i\}$ es el estadístico suficiente. Así, la distribución a posteriori está dada por:

$$\begin{aligned}\pi(p|\mathbf{x}) = f(p|\tau + t) &\propto (1-p)^{n+\tau_0} \exp\left(\left(\sum_{i=1}^n x_i + \tau_1\right) \log\left(\frac{p}{1-p}\right)\right) \\ &\propto p^{\sum_{i=1}^n x_i + \tau_1} (1-p)^{t_0 + n - \tau_1 - \sum_{i=1}^n x_i}\end{aligned}\quad (2.51)$$

Luego, $p|\mathbf{x} \sim \text{Beta}(\sum_{i=1}^n x_i + \alpha, n - \sum_{i=1}^n x_i + \beta)$.

Ejemplo 2.10. Consideremos el modelo Exponencial con parámetro λ , donde la función de densidad puede escribirse:

$$f(x|\lambda) = \lambda \exp(-\lambda x) \quad (2.52)$$

La distribución previa conjugada es:

$$\pi(\lambda) = f(\lambda|\tau) \propto \lambda^{\tau_0} \exp(-\lambda\tau_1) \quad (2.53)$$

Luego, $\lambda \sim \text{Gamma}(\alpha, \beta)$, donde $\alpha = \tau_0 + 1$ y $\beta = \tau_1$

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Bernoulli, la función de máxima verosimilitud

$$L(\lambda|\mathbf{x}) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) \quad (2.54)$$

donde $t(\mathbf{x}) = \{n, \sum_{i=1}^n x_i\}$ es el estadístico suficiente. Así, la distribución a posteriori está dada por:

$$\pi(\lambda|\mathbf{x}) = f(\lambda|\tau + t) \propto \lambda^{n+\tau_0} \exp\left(-\lambda\left(\tau_1 + \sum_{i=1}^n x_i\right)\right) \quad (2.55)$$

Luego, $\lambda|\mathbf{x} \sim \text{Gamma}(\alpha + n, \beta + \sum_{i=1}^n x_i)$.

Ejemplo 2.11. Consideremos el modelo Geométrico con parámetro p , donde la función de densidad puede escribirse:

$$f(x|p) = \frac{p}{1-p} \exp(x \log(1-p)) \quad (2.56)$$

La distribución previa conjugada es:

$$\begin{aligned} \pi(p) = f(p|\tau) &\propto \left(\frac{p}{1-p}\right)^{\tau_0} \exp(\tau_1 \log(1-p)) \\ &\propto p^{\tau_0} (1-p)^{\tau_1 - \tau_0} \end{aligned} \quad (2.57)$$

Luego, $p \sim \text{Beta}(\alpha, \beta)$, donde $\alpha = \tau_0 + 1$ y $\beta = \tau_1 - \tau_0 + 1$

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Geométrica, la función de máxima verosimilitud

$$L(p|\mathbf{x}) = \left(\frac{p}{1-p}\right)^n \exp\left(\sum_{i=1}^n x_i \log(1-p)\right) \quad (2.58)$$

donde $t(\mathbf{x}) = \{n, \sum_{i=1}^n x_i\}$ es el estadístico suficiente. Así, la distribución a posteriori está dada por:

$$\begin{aligned} \pi(p|\mathbf{x}) = f(p|\tau + t) &\propto \left(\frac{p}{1-p}\right)^{\tau_0+n} \exp\left(\left(\sum_{i=1}^n x_i + \tau_1\right) \log(1-p)\right) \\ &\propto p^{\alpha+n-1} (1-p)^{\beta+\sum_{i=1}^n x_i - n-1} \end{aligned} \quad (2.59)$$

Luego, $p|\mathbf{x} \sim \text{Beta}(\alpha + n, \beta + \sum_{i=1}^n x_i - n)$.

Ejemplo 2.12. Consideremos el modelo Normal con parámetro μ y σ_0^2 conocido, donde la función de la densidad se escribe:

$$f(x|\mu) \propto \exp\left(-\frac{\mu^2}{2\sigma_0^2}\right) \exp\left(x \frac{\mu}{\sigma_0^2}\right) \quad (2.60)$$

La distribución previa conjugada es:

$$\begin{aligned} \pi(\mu) = f(\mu|\tau) &\propto \exp\left(-\frac{\tau_0 \mu^2}{2\sigma_0^2}\right) \exp\left(\frac{\mu \tau_1}{\sigma_0^2}\right) \\ &\propto \exp\left(-\frac{\tau_0}{2\sigma_0^2} \left(\mu - \frac{\tau_1}{\tau_0}\right)^2\right) \end{aligned} \quad (2.61)$$

luego $\mu \sim \text{Normal}(\mu_1, \sigma_1^2)$ donde $\mu_1 = \frac{\tau_1}{\tau_0}$ y $\sigma_1^2 = \frac{\sigma_0^2}{\tau_0}$.

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Normal, en el caso de σ_0^2 conocido, la función de máxima verosimilitud es:

$$\begin{aligned} L(\mu|\mathbf{x}) &\propto \exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2\right) \\ &\propto \exp\left(-\frac{n}{2\sigma_0^2} (\bar{x} - \mu)^2\right) \end{aligned} \quad (2.62)$$

Así, la distribución a posteriori está dada por:

$$\begin{aligned}\pi(\mu|\mathbf{x}) &\propto \exp\left(-\frac{1}{2}\left(\frac{(\mu - \mu_1)^2}{\sigma_1^2} + \frac{n(\mu - \bar{x})^2}{\sigma_0^2}\right)\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_1^2} + \frac{n}{\sigma_0^2}\right)\left(\mu - \frac{\frac{1}{\sigma_1^2}\mu_1 + \frac{n}{\sigma_0^2}\bar{x}}{\frac{1}{\sigma_1^2} + \frac{n}{\sigma_0^2}}\right)^2\right)\end{aligned}\quad (2.63)$$

luego, $\mu|\mathbf{x} \sim \text{Normal}\left(\frac{\frac{1}{\sigma_1^2}\mu_1 + \frac{n}{\sigma_0^2}\bar{x}}{\frac{1}{\sigma_1^2} + \frac{n}{\sigma_0^2}}, \left(\frac{1}{\sigma_1^2} + \frac{n}{\sigma_0^2}\right)^{-1}\right)$

Ejemplo 2.13. Consideremos el modelo Normal con parámetro σ^2 y μ_0 conocido, la función de densidad se escribe:

$$f(x|\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu_0)^2}{2\sigma^2}\right) \quad (2.64)$$

La distribución previa conjugada es:

$$\pi(\sigma^2) = f(\sigma^2|\tau) \propto (\sigma^2)^{-\frac{\tau_0}{2}} \exp\left(-\frac{\tau_1}{2\sigma^2}\right) \quad (2.65)$$

luego $\sigma^2 \sim \text{Invgamma}(\alpha, \beta)$ donde $\alpha = \frac{\tau_0}{2} + 1$ y $\beta = \frac{\tau_1}{2}$.

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Normal, en el caso de μ_0 conocido y σ^2 desconocido, la función de máxima verosimilitud es:

$$\begin{aligned}L(\sigma^2|\mathbf{x}) &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2\right) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{ns^2}{2\sigma^2}\right)\end{aligned}\quad (2.66)$$

donde $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$.

Luego, la distribución a posteriori está dada por:

$$\pi(\sigma^2|\mathbf{x}) \propto (\sigma^2)^{-\left(\frac{\tau_0}{2} + \frac{n}{2}\right)} \exp\left(-\frac{1}{\sigma^2} \left(\frac{ns^2}{2} + \frac{\tau_1}{2}\right)\right) \quad (2.67)$$

luego $\sigma^2|\mathbf{x} \sim \text{Invgamma}\left(\alpha + \frac{n}{2}, \beta + \frac{ns^2}{2}\right)$.

Ejemplo 2.14. Consideremos el modelo Normal con parámetros μ y σ^2 desconocidos. La distribución previa que se utilizara en esta situación es la distribución previa conjunta y se suponen que las distribuciones de los parámetros son independientes ($\pi(\mu, \sigma^2) = \pi(\mu)\pi(\sigma^2)$). Donde $\mu \sim \text{Normal}(\mu_1, \sigma_1^2)$ (previa conjugada del Ejemplo 2.12) y $\sigma^2 \sim \text{Invgamma}(\alpha, \beta)$ (previa conjugada del Ejemplo 2.13).

Así, obtenemos la distribución previa conjunta sobre μ y σ^2 es:

$$\pi(\mu, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma_1^2}(\mu - \mu_1)^2\right) (\sigma^2)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma^2}\right) \quad (2.68)$$

Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria i.i.d. de una distribución Normal, la función de máxima verosimilitud es:

$$\begin{aligned} L(\mu, \sigma^2 | \mathbf{x}) &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right)\right) \end{aligned} \quad (2.69)$$

Así, la distribución a posteriori es:

$$\begin{aligned} \pi(\mu, \sigma^2 | \mathbf{x}) &\propto (\sigma^2)^{-(\frac{n}{2} + \alpha) - 1} \exp\left(-\frac{1}{\sigma^2} \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2} + \beta\right)\right) \\ &\quad \exp\left(-\frac{1}{2\sigma_1^2}(\mu - \mu_1)^2 - \frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right). \end{aligned} \quad (2.70)$$

2.2.6. Intervalos de credibilidad

Mediante la metodología bayesiana, es posible construir Intervalos de Credibilidad para parámetros desconocidos θ de una función de probabilidad correctamente definida para la v.a X , $f(x|\theta)$. Específicamente, se desea conocer la región C donde sea más probable encontrar al parámetro de interés θ . La siguiente definición permite determinar una región de credibilidad ([Pacheco, L., 2013](#)) y ([Bernardo and Smith. 1994](#)).

Definición 2.2.7. Una región $C \in \Omega$, donde Ω es el espacio muestral asociado a un experimento aleatorio, tal que: $\int_C p(\theta) = 1 - \alpha$, $0 \leq \alpha \leq 1$, es una región de $(1 - \alpha)\%$ de credibilidad para θ . En el caso de que θ sea un escalar, la región de credibilidad C es usualmente un intervalo $[c_1, c_2]$.

Si $p(\theta)$ es una densidad (previa-a posteriori), entonces C es una región de credibilidad (previa-a posteriori).

Debido a que pueden resultar múltiples regiones de credibilidad para un parámetro desconocido θ , se suele limitar o restringir el espacio solamente a aquellas regiones que cumplan con condiciones particulares tales como: menor volumen o longitud, (Pacheco,L.,2013) y (Jackman. 2009).

Definición 2.2.8. (Región de la Densidad Posterior Más Alta) Una región $C \in \Omega$ es una Región de la Densidad Posterior Mas Alta al $100(1 - \alpha)\%$ para θ bajo $p(\theta)$ si:

1. $P(\theta \in C) = 1 - \alpha$
2. $P(\theta_1) \geq P(\theta_2) \forall \theta_1 \in C \wedge \theta_2 \notin C$.

La distancia que utilizaremos en este trabajo es la distancia conocida como rectilínea.

Definición 2.2.9. Dado dos intervalos $F = (L_F, H_F)$ de confianza y $B = (L_B, H_B)$ de credibilidad, se calcula la distancia de estos dos de la siguiente manera:

$$D(F, B) = |L_F - L_B| + |H_F - H_B|. \quad (2.71)$$

Capítulo 3

SIMULACIONES

Encontrando el tamaño de muestra requerido.

Para comparar los intervalos de confianza con los intervalos de credibilidad (descritos en el Capítulo 2) en términos de encontrar el tamaño de muestra n con el cual los dos intervalos coincidan salvo una distancia requerida D_R , se realizó una simulación en R en la cual se consideran combinaciones (n, θ, D_R) para cada uno de los modelos *Binomomial*, *Geométrico*, *Poisson*, *Exponencial* y *Normal*. En cada combinación D_R, θ se fijan y n se varía ($n = 1, 2, 3, \dots$) tal que para cada valor de n se calcula la distancia sobre los intervalos de confianza y de credibilidad generados utilizando un nivel de confianza y un nivel de credibilidad del 95%. Se repite este proceso 1,000 veces para cada valor de n , obteniendo así una distancia promedio D_{50} (o distancia cuantil 99 D_{99}) de 1,000 distancias simuladas y luego se busca el n con el cual la distancia D_{50} sea menor o igual que la distancia D_R (o la distancia D_{99} sea menor o igual que la distancia D_R). La distancia requerida D_R se considera como un valor dependiente del parámetro de interés θ , los valores para cada θ , son: $D_R = 1\% \theta, 5\% \theta, 10\% \theta$.

Ejemplo 3.1. Simulación para el modelo Normal en el caso de μ, σ^2 desconocidos.

1. Se fijan n, μ, σ^2 empezando con $n = 1$.
2. Se generan n observaciones de la distribución $Normal(\mu, \sigma^2)$.
3. Se calculan los intervalos de confianza (L_F, H_F) y los intervalos de credibilidad (L_B, H_B) .

4. Se calcula $D = |L_F - L_B| + |H_F - H_B|$.
5. Se repiten 1,000 veces los paso 2., 3. y 4.
6. Se calcula D_{50} (o D_{99}).
7. Si $D_{50} \leq D_R$ (o $D_{99} \leq D_R$) pare.
8. De lo contrario, se cambia el valor de n y se repiten los pasos anteriores.

En este estudio de simulación, se utilizaron algunas funciones básicas de R como: rbinom, rbeta, rgamma, rexp, rnorm, qgamma, entre otras, con las cuales se generaron datos y se calcularon los intervalos de credibilidad, calculando los cuantiles 0.025 y 0.975. En el caso donde μ y σ^2 son desconocidos en el modelo Normal, los intervalos de credibilidad se calcularon generando una gran cantidad de muestras de la distribución conjunta a posteriori de (μ, σ^2) obtenida a partir de cada una de las dos distribuciones previas descritas en el Ejemplo 2.7 y el Ejemplo 2.14, empleando la metodología conocida como muestreador de Gibbs (la metodología del muestreador de Gibbs se describe en el Apéndice A), la cual hace parte de los métodos de Markov chain Monte Carlo (MCMC), y a partir de dichas muestras, obtener los intervalos de credibilidad calculando los cuantiles 0.025 y 0.975.

Para cada modelo se muestran tablas que presentan los resultados de este estudio. Para una mejor apreciación de las tablas, los nombres de las columnas son:

- θ : Parámetro de interés
- D_R : Distancia requerida.
- D_{50} : Distancia promedio.
- n_{50} : Tamaño de muestra encontrado a partir de D_{50} .
- D_{99} : Distancia cuantil 99.
- n_{99} : Tamaño de muestra encontrado a partir de D_{99} .

Se presenta además una serie de gráficas para cada modelo como una manera de ilustración de la información obtenida.

También, se presentan gráficas describiendo el comportamiento de la distancia D_{50} (y D_{99}), teniendo en cuenta diferentes tamaños de muestras para las distribuciones previas de Jeffrey y Conjugadas.

3.1. Distribución Binomial.

La Tabla 3-1 expone los tamaños de muestra n encontrados para el objetivo del estudio sobre el parámetro p del modelo Binomial en los casos $p = 0.1, 0.5, 0.9$. Se utilizó el intervalo de confianza de Clopper-Pearson (Subsección 2.1.1) y los intervalos de credibilidad utilizando las distribuciones previas de Jeffrey (Ejemplo 2.3) y Beta conjugada (Ejemplo 2.9).

Para una mejor lectura de la Tabla, se describe el siguiente caso: al utilizar la distribución previa $Beta(3, 3)$, si se quiere que la distancia de los dos intervalos (de confianza y de credibilidad) para el parámetro $p = 0.1$ sea menor o igual que una distancia requerida $D_R = 0.001$ ($1\%p$), se necesitan 3,947 observaciones para que haya un chance aproximado del 50% de que la distancia de los dos intervalos sea menor o igual a $D_R = 0.01$ y se necesitan 4,054 observaciones para casi garantizar que la distancia de los dos intervalos es menor o igual a $D_R = 0.001$.

Se observa en la Tabla 3-1 con la distribución previa de Jeffrey, que los valores n_{50} y n_{99} en casi todos los casos (de $p = 0.1, 0.5, 0.9$) son iguales, esto es debido a que la distancia de los límites superiores y la distancia de los límites inferiores siempre es la misma independientemente de la observación que se obtenga. Por otro lado, cuando se utiliza la distribución previa Conjugada los tamaños de muestra n_{50} y n_{99} son diferentes para los casos de $p = 0.1, 0.9$, en el caso de $p = 0.5$ se observa que n_{50} y n_{99} coinciden y además se observar que cuando los valores de los parámetros de la distribución previa Conjugada aumentan, los valores de n_{50} y n_{99} también aumentan.

Distribución Previa	p	D_R	D_{50}	n_{50}	D_{99}	n_{99}
Jeffrey	0.1	0.001	0.001	997	0.001	997
		0.005	0.005	196	0.005	196
		0.01	0.01	96	0.01	96
	0.5	0.005	0.005	197	0.005	197
		0.025	0.025	36	0.025	36
		0.05	0.05	16	0.05	16
	0.9	0.009	0.009	107	0.009	107
		0.045	0.043	20	0.044	24
		0.09	0.085	9	0.09	10
Beta(1,1)	0.1	0.001	0.001	948	0.001	964
		0.005	0.005	173	0.005	188
		0.01	0.01	79	0.01	94
	0.5	0.005	0.005	211	0.005	211
		0.025	0.025	42	0.025	42
		0.05	0.05	20	0.05	20
	0.9	0.009	0.009	89	0.009	104
		0.045	0.045	11	0.045	21
		0.09	0.88	5	0.083	11
Beta(3,3)	0.1	0.001	0.001	3947	0.001	4054
		0.005	0.005	771	0.005	812
		0.01	0.01	375	0.01	404
	0.5	0.005	0.005	256	0.005	257
		0.025	0.024	60	0.024	61
		0.05	0.048	32	0.048	33
	0.9	0.009	0.009	419	0.009	453
		0.045	0.044	70	0.045	80
		0.09	0.089	25	0.088	33

Tabla 3-1: Tamaño de muestra obtenido para el parámetro $p = 0.1, 0.5, 0.9$ del modelo Binomial.

La Figura 3-1 expone el comportamiento del tamaño de muestra n (cuando se varía p de 0.1 a 0.5), con el cual la distancia promedio D_{50} (y D_{99}) calculada sobre los intervalos de confianza y de credibilidad es menor o igual a una distancia Requerida D_R ($D_R = 1\%p, 5\%p, 10\%p$). Se observa con la distribución previa Conjugada $Beta(3, 3)$, que se necesita un tamaño de muestra mucho más grande que cuando se utiliza la distribución previa $Beta(1, 1)$ y Jeffrey. Este resultado se tiene debido a que para casi todos los valores de p excepto para valores muy cercanos a $p = 0.5$, la distribución previa $Beta(3, 3)$ es una previa mala para p y por esta razón se necesita un tamaño de muestra más grande para casi eliminar el efecto de la distribución previa. Esto no pasa en el caso de la distribución previa Jeffrey y $Beta(1, 1)$ debido a que estas no tienen preferencia sobre el valor de p .

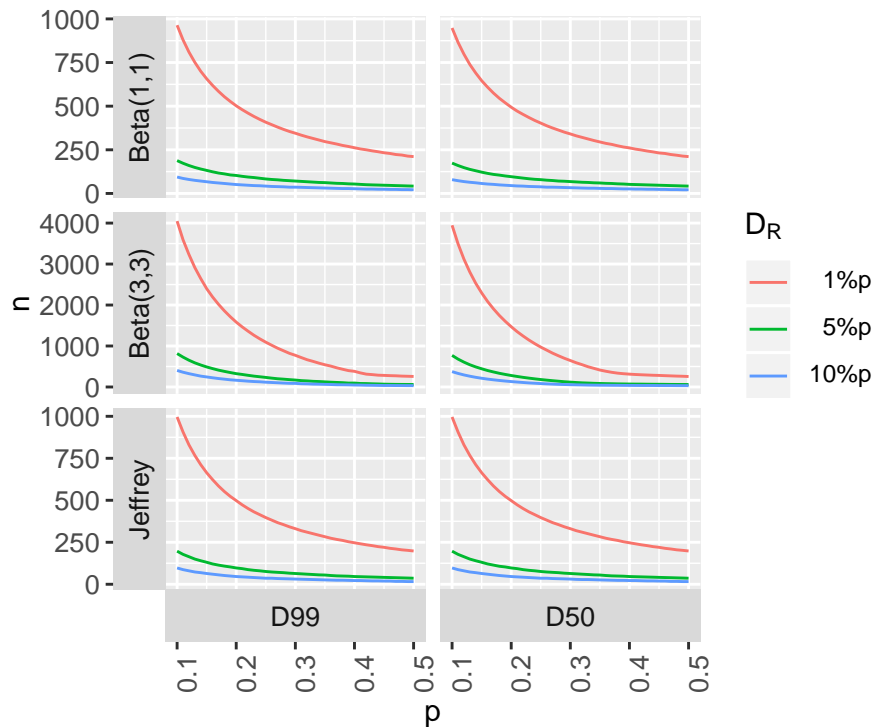


Figura 3-1: Comportamiento del tamaño de muestra n cuando se varía el parámetro p del modelo Binomial utilizando la distancia D_{50} y D_{99} para comparar los intervalos de confianza (Cloper-Pearson) con los intervalos de credibilidad (generados por las distribuciones previas de Jefrey y Beta Conjugada).

La Figura 3-2 muestra el comportamiento de la distancia D_{50} y distancia D_{99} variando el parámetro p para un tamaño de muestra fijo $n = 50, 100, 250, 500, 1000$ usando el intervalo de Clopper-Pearson y las distribuciones previas de Jeffrey, $Beta(1, 1)$ y $Beta(3, 3)$. Se observa simetría alrededor del parámetro $p = 0.5$ para todos los casos, se observa que el comportamiento de la distancia (al utilizar la distribución previa de Jeffrey) es constante para casi todos los valores de p y se observa que cuando n va al infinito la distancia converge a cero.

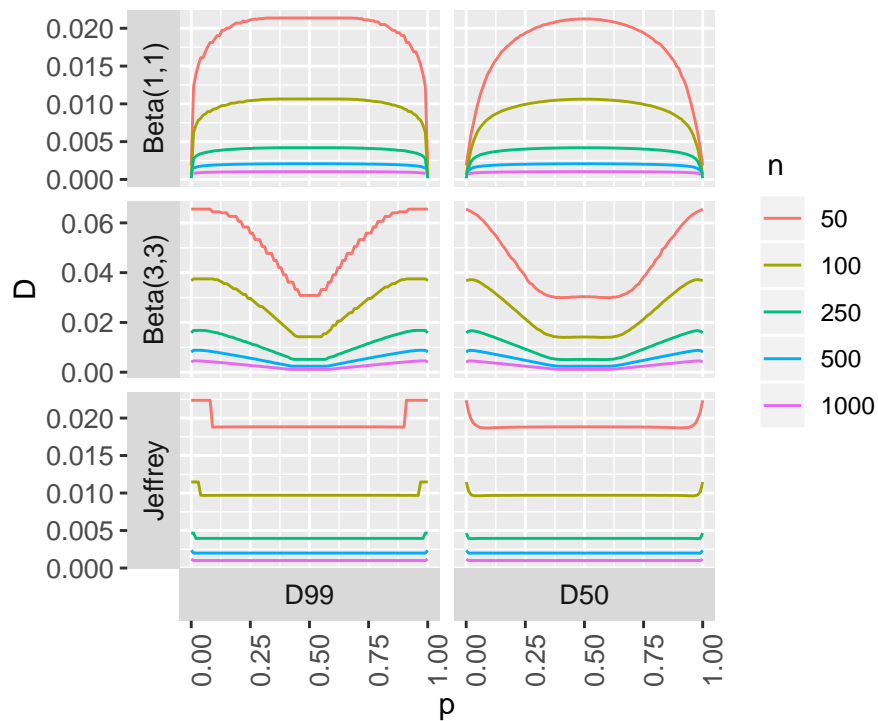


Figura 3–2: Comportamiento de la distancia en función del parámetro p y diferentes tamaños de muestra usando la distribución previa de Jeffrey y Beta Cojugado.

3.2. Distribución Poisson.

La Tabla 3-2 y la Figura 3-3 exponen los tamaños de muestra n encontrados para el objetivo del estudio sobre el parámetro λ del modelo Poisson. Se utilizó el intervalo de confianza de Garwood (Subsección 2.1.2) y el intervalo de credibilidad utilizando las distribuciones previas de Jeffrey (Ejemplo 2.3) y Gamma conjugada (Ejemplo 2.8).

Se observa en la Tabla 3-2, con la distribución previa de Jeffrey, que los valores n_{50} y n_{99} en todos los casos (de $\lambda = 1, 5, 10$) son iguales, esto es debido a que la distancia de los límites superiores y la distancia de los límites inferiores siempre es la misma independientemente de la observación que se obtenga, mientras que con la distribución $Gamma(1, 1)$, se tienen diferentes valores para n_{50} y n_{99} . Esto se evidencia en la Figura 3-3 y la Figura 3-4.

Distribución Previa	λ	D_R	D_{50}	n_{50}	D_{99}	n_{99}
Jeffrey	1	0.01	0.01	99	0.01	99
		0.05	0.0499	19	0.05	19
		0.1	0.0996	9	0.0998	9
	5	0.05	0.05	19	0.05	19
		0.25	0.2497	3	0.2498	3
		0.5	0.4982	1	0.4993	1
	10	0.1	0.1	9	0.1	9
		0.5	0.4994	1	0.4997	1
		1	0.4994	1	0.4997	1
Gamma(1,1)	1	0.01	0.0099	128	0.01	147
		0.05	0.0476	32	0.0478	40
		0.1	0.0999	17	0.0916	24
	5	0.05	0.05	180	0.0497	196
		0.25	0.240	37	0.240	44
		0.5	0.4833	19	0.490	23
	10	0.1	0.0996	190	0.0998	200
		0.5	0.4773	39	0.4966	42
		1	0.9149	20	0.965	22

Tabla 3-2: Tamaño de muestra obtenido para el parámetro $\lambda = 1, 5, 10$ del modelo Poisson.

En la Figura 3-3 se observa con la distribución previa de Jeffrey que el comportamiento del tamaño de muestra n es igual en términos de D_{50} y D_{99} y también que es decreciente a medida que el valor del parámetro aumenta. También se observa con la distribución previa $Gamma(1, 1)$, que el comportamiento del tamaño de muestra n es creciente a medida que el valor del parámetro aumenta y que el tamaño de muestra que se necesita es más grande cuando se utiliza D_{99} que cuando se utiliza D_{50} .

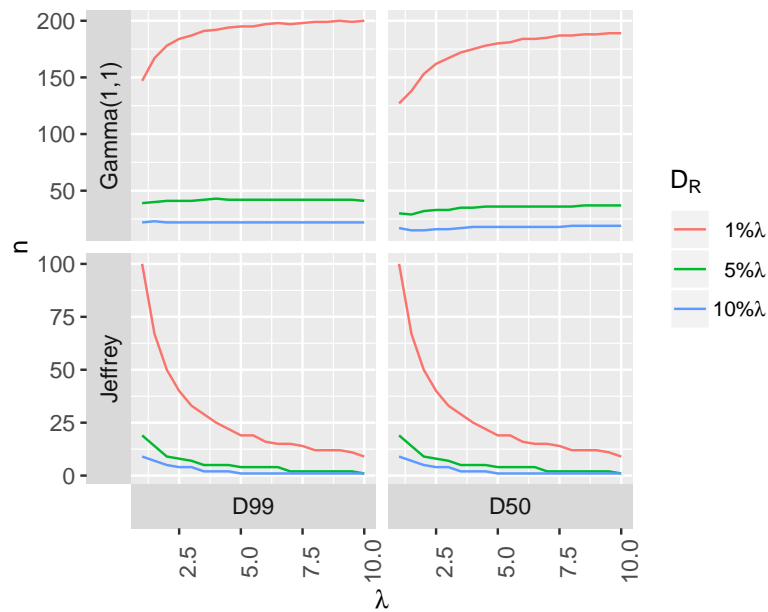


Figura 3-3: Comportamiento del tamaño de muestra n cuando se varía el parámetro λ en el modelo Poisson utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Garwood) y los intervalos de credibilidad (Jeffrey y $Gamma(1, 1)$).

La Figura 3-4 muestra el comportamiento de la distancia D_{50} y distancia D_{99} variando el parámetro λ para un tamaño de muestra fijo $n = 50, 100, 250, 500, 1000$ usando el intervalo de confianza de Garwood y los intervalos de credibilidad con las distribuciones previas de Jeffrey y $Gamma(1, 1)$. Se observa con la distribución previa de Jeffrey que, al fijar n y variar λ , la distancia D_{50} y D_{99} son iguales y constantes para cualquier valor de λ , esto es debido a que la distancia de los límites superiores y la distancia de los límites inferiores de los dos intervalos siempre es la misma, independientemente de la observación que se obtenga. También se observa

con la distribución previa $\text{Gamma}(1, 1)$ que D_{50} y D_{99} crecen aproximadamente de forma lineal cuando el valor del parámetro crece y se observa en cualquiera de los casos, que cuando n va a infinito, la distancia D_{50} y D_{99} convergen a cero.

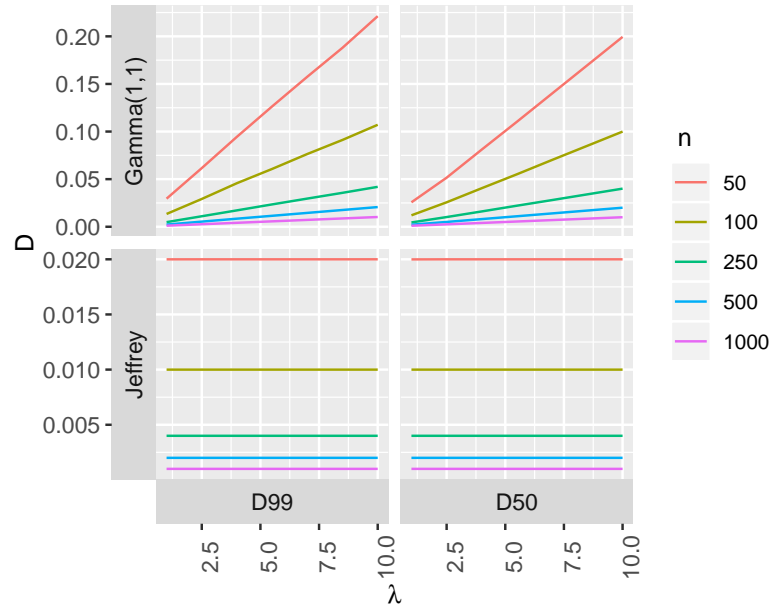


Figura 3-4: Comportamiento de la distancia en función del parámetro λ y diferentes tamaños de muestra usando la distribución previa Jeffrey y Gamma conjugado.

3.3. Distribución Geométrica.

En la Tabla 3-3 y la Figura 3-5, se exponen los tamaños de muestra n encontrados para el objetivo del estudio sobre el parámetro p del modelo Geométrico. Se utilizó el intervalo de confianza descrito en la Subsección 2.1.3 y el intervalo de credibilidad utilizando la distribución previa de Jeffrey (Ejemplo 2.4) y la distribución previa Conjugada $\text{Beta}(1, 1)$ (Ejemplo 2.11).

Se observa en la Tabla 3-3, con la distribución previa Jeffrey y $\text{Beta}(1, 1)$, que el tamaño de muestra n_{50} y n_{99} son diferentes para cada uno de los casos ($p = 0.1, 0.5, 0.9$). También se observa en la mayoría de los casos que los tamaños de muestra n_{50} y n_{99} son más grandes cuando se utiliza la distribución previa $\text{Beta}(1, 1)$, en comparación a cuando se utiliza la distribución previa de Jeffrey.

Distribución Previa	p	D_R	D_{50}	n_{50}	D_{99}	n_{99}
Jeffrey	0.1	0.001	0.001	142	0.00099	177
		0.005	0.0049	29	0.005	35
		0.01	0.01	14	0.0097	19
	0.5	0.005	0.005	55	0.0049	108
		0.025	0.025	14	0.0246	33
		0.05	0.044	9	0.046	22
	0.9	0.009	0.009	241	0.009	271
		0.045	0.045	47	0.045	58
		0.09	0.089	23	0.09	27
Beta(1,1)	0.1	0.001	0.001	312	0.01	340
		0.005	0.005	62	0.005	74
		0.01	0.01	31	0.0098	40
	0.5	0.005	0.005	45	0.005	87
		0.025	0.025	15	0.025	32
		0.05	0.05	9	0.049	22
	0.9	0.009	0.009	309	0.0089	349
		0.045	0.0449	60	0.045	76
		0.09	0.0895	29	0.09	38

Tabla 3-3: Tamaño de muestra obtenido para el parámetro $p = 0.1, 0.5, 0.9$ del modelo Geométrico.

En la Figura 3-5, se observa con la distribución previa $Beta(1,1)$ que el comportamiento del tamaño de muestra n en términos de D_{50} y D_{99} es más grande que cuando se utiliza la distribución previa de Jeffrey. También se observan en los cuatro casos de la figura que el comportamiento del tamaño de muestra es decreciente hasta cierto valor del parámetro p y luego crece.

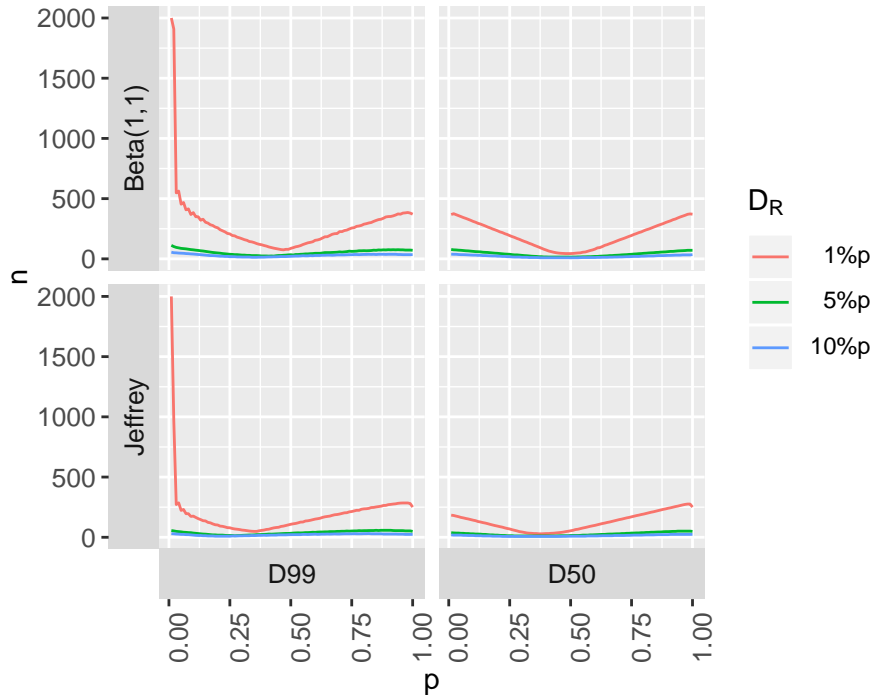


Figura 3-5: Comportamiento del tamaño de muestra n cuando se varía el parámetro p del modelo Geométrico utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza descritos en la Subsección 2.1.3 y los intervalos de credibilidad (Jeffrey y $Beta(1,1)$).

La Figura 3-6 muestra el comportamiento de la distancia D_{50} y distancia D_{99} variando el parámetro p para un tamaño de muestra fijo $n = 50, 100, 250, 500, 1000$ usando el intervalo de confianza descrito en la Subsección 2.1.3 y las distribuciones previas de Jeffrey y la distribución Conjugada $Beta(1,1)$. Se observa un comportamiento extraño de la distancia (D_{50} y D_{99}) cuando se varía el valor del parámetro p . Este comportamiento no se puede explicar de forma explícita debido a que la distancia es calculada por la fórmula $D = |L_F - L_B| + |H_F - H_B|$, donde L_F y H_F son los límites del intervalo de confianza (ver Subsección 2.1.3) y L_B y H_B son los

límites del intervalo de credibilidad (ver Ejemplo 2.4 y Ejemplo 2.11). Estos límites tienen formas muy diferentes para ambos enfoques.

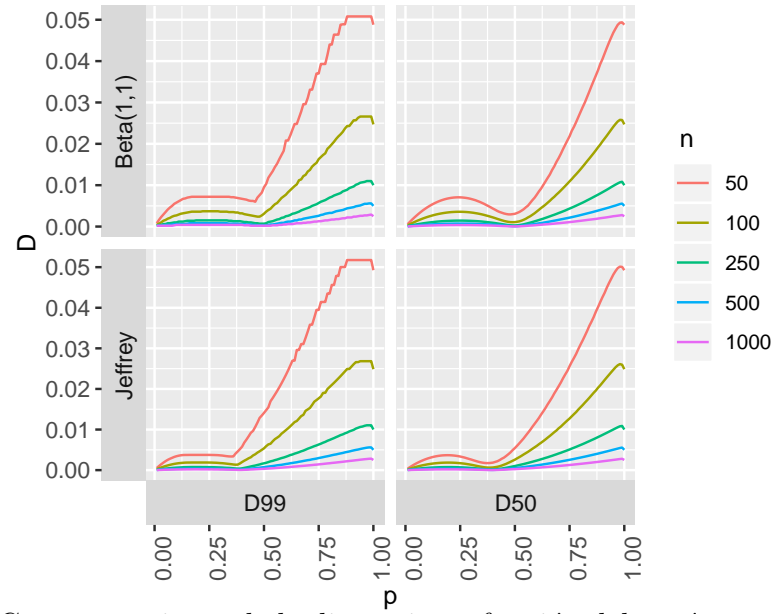


Figura 3–6: Comportamiento de la distancia en función del parámetro p y diferentes tamaños de muestra usando la distribución previa Jeffrey y Beta conjugado.

3.4. Distribución Exponencial.

Para el modelo exponencial no se utilizará la distribución de Jeffrey debido a que con ella los intervalos de credibilidad coinciden con los intervalos de confianza, lo que implica que la distancia calculada siempre es cero para cualquier tamaño de muestra (ver teorema 3, apéndice A).

La Tabla 3-4 y la Figura 3-7 exponen los tamaños de muestra n encontrados para el objetivo del estudio sobre el parámetro λ del modelo Exponencial. Se utilizó el intervalo de confianza descrito en la Subsección 2.1.4 y el intervalo de credibilidad utilizando la distribución previa Gamma Conjugada (Ejemplo 2.10).

Se observa en la Tabla 3-4 que, con la distribución previa $Gamma(1, 1)$, el tamaño de muestra n_{50} y n_{99} son diferentes para cada uno de los casos de $\lambda = 1, 5, 10$. También se observa que cuando el parámetro λ crece el tamaño de muestra también crece.

Distribución Previa	λ	D_R	D_{50}	n_{50}	D_{99}	n_{99}
$Gamma(1, 1)$	1	0.01	0.0097	40	0.01	85
		0.5	0.5	4	0.4496	11
		0.1	0.099	10	0.1	23
	5	0.05	0.0499	800	0.05	946
		0.25	0.25	160	0.25	219
		0.5	0.5	79	0.5	123
	10	0.1	0.0998	1795	0.1	2000
		0.5	0.5	355	0.4923	451
		1	1	174	1	240

Tabla 3-4: Tamaño de muestra obtenido para el parámetro $\lambda = 1, 5, 10$ del modelo exponencial.

En la Figura 3-7, se observa que con la distribución previa Conjugada $Gamma(1, 1)$, el comportamiento del tamaño de muestra n es más grande cuando se utiliza la distancia D_{99} que cuando se utiliza la distancia D_{50} .

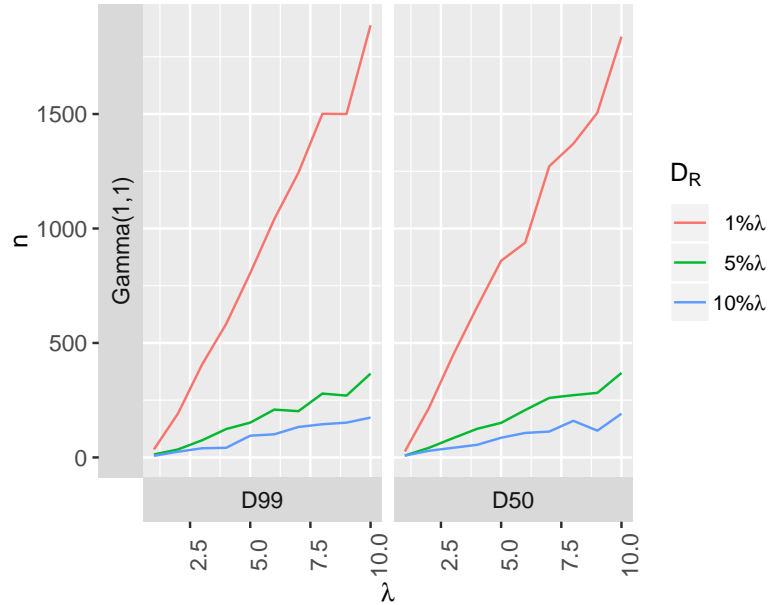


Figura 3-7: Tamaño de muestra n cuando se varía el parámetro λ del modelo Exponencial utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.4) y los intervalos de credibilidad (con la distribución $Beta(1, 1)$).

La Figura 3-8 muestra el comportamiento de la distancia D_{50} y la distancia D_{99} a medida que se varía el parámetro λ para un tamaño de muestra fijo $n = 50, 100, 250, 500, 1000$ utilizando la distribución previa $Gamma(1, 1)$.

Se observa que la distancia D_{99} crece más rápido que la distancia D_{50} a medida que λ varía de 1 a 10. Se observa que la distancia D_{50} y D_{99} crece cuando el valor del parámetro crece y se fija el tamaño de muestra y también se observa que cuando n va a infinito, la distancia converge a cero.

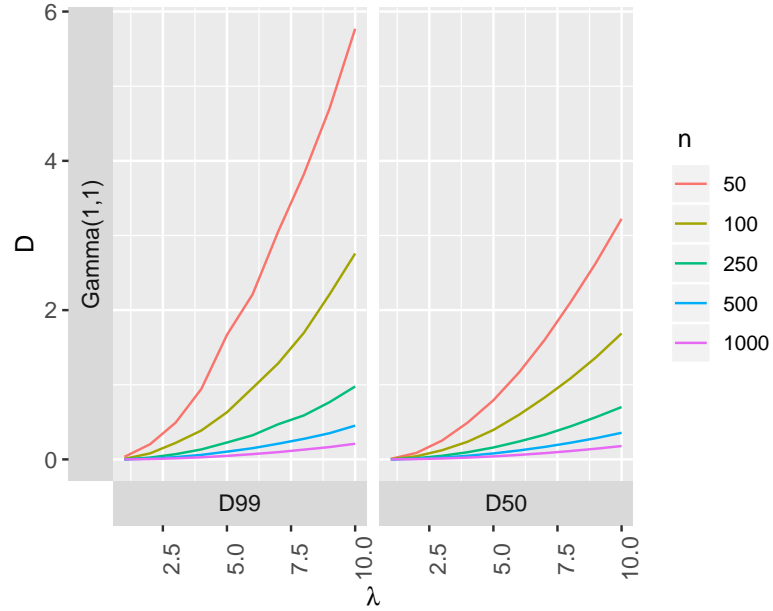


Figura 3–8: Comportamiento de la distancia en función del parámetro λ y diferentes tamaños de muestra usando la distribución previa $Gamma(1, 1)$

3.5. Distribución Normal.

3.5.1. Inferencia en μ con σ_0 conocido

En este caso, no se utilizará la distribución de Jeffrey debido a que con esta los intervalos de confianza y de credibilidad son equivalentes, lo que implica que la distancia calculada sobre los intervalos siempre es cero para cualquier tamaño de muestra (ver teorema 2 del apéndice A).

La Tabla 3-5, la Figura 3-9 y la Figura 3-10 exponen los tamaños de muestra n encontrados para el objetivo del estudio sobre el parámetro μ en el caso $\sigma_0 = 1, 5$ conocido del modelo Normal. Se utilizó el intervalo de confianza descrito en la Subsección 2.1.5 y el intervalo de credibilidad utilizando la distribución previa Normal conjugada (Ejemplo 2.12). Se observa en la Tabla 3-5, con la distribución previa $Normal(0, 1)$ que el tamaño de muestra n_{50} y n_{99} son diferentes para cada uno de los casos $\mu = 1, 5, 10$, cuando se fija D_R y σ_0 .

	μ	σ_0	D_R	D_{50}	n_{50}	D_{99}	n_{99}
<i>Normal</i> (0, 1)	1	1	0.01	0.01	199	0.01	229
			0.05	0.05	39	0.0499	52
			0.1	0.0997	19	0.0987	28
		5	0.01	0.01	4969	0.01	5740
			0.05	0.0497	977	0.05	1286
			0.1	0.1	475	0.0999	689
		5	0.05	0.05	199	0.05	205
			0.25	0.25	39	0.2495	42
			0.5	0.4993	19	0.5	21
		5	0.05	0.05	4975	0.05	5135
			0.25	0.2497	976	0.25	1046
			0.5	0.5	474	0.4991	524
	10	1	0.1	0.1	199	0.10	202
			0.5	0.4997	39	0.5	40
			1	0.9994	19	0.9963	20
		5	0.1	0.1	4975	0.1	5054
			0.5	0.4998	975	0.5	1009
			1	1	475	1	500

Tabla 3-5: Tamaño de muestra obtenido para el parámetro $\mu = 1, 5, 10$ del modelo Normal.

En Figura 3-9, ($\sigma_0 = 1$ conocido) y la Figura 3-10 ($\sigma_0 = 5$ conocido), se observa que cuando se utiliza D_{50} , el tamaño de muestra es constante para cualquier valor de μ , mientras que con D_{99} se observa que el tamaño de muestra decrece a medida que μ crece. Esto ocurre en los casos donde $D_R = 1\% \mu, 5\% \mu, 10\% \mu$. También se

observa que el tamaño de muestra es más grande cuando $\sigma_0 = 5$ comparado con el caso donde $\sigma_0 = 1$.

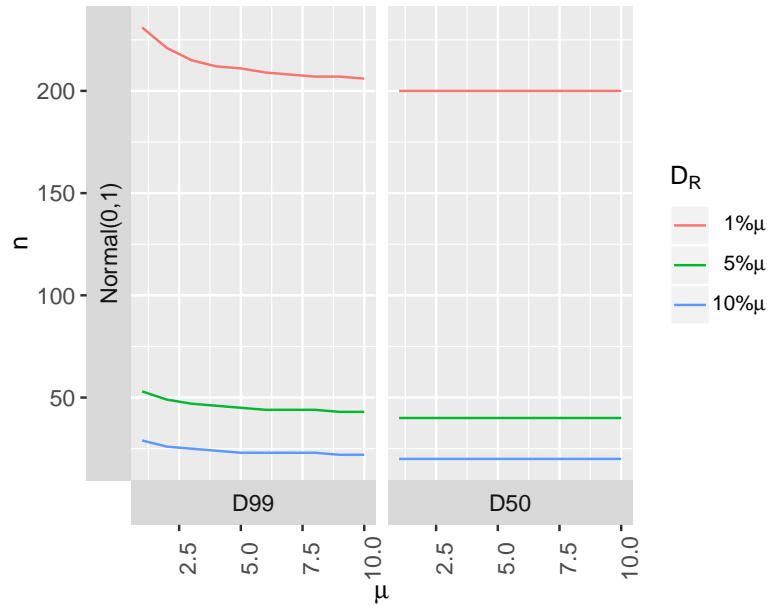


Figura 3–9: Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma_0 = 1$ conocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa $Normal(0, 1)$ (Ejemplo 2.12).

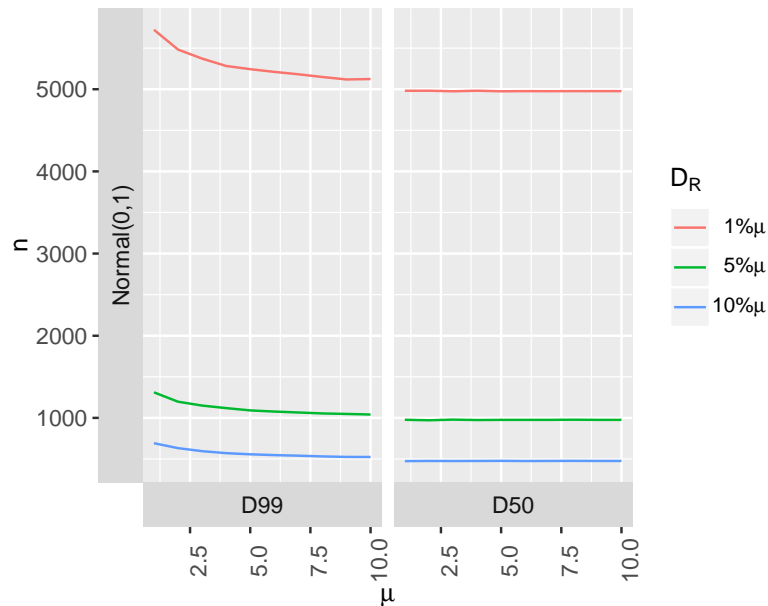


Figura 3–10: Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma_0 = 5$ conocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa $Normal(0, 1)$ (Ejemplo 2.12).

La Figura 3-11 ($\sigma_0 = 1$ conocido) y la Figura 3-12 ($\sigma_0 = 5$ conocido), muestra el comportamiento de la distancia D_{50} y la distancia D_{99} a medida que se varía el parámetro μ para un tamaño de muestra fijo $n = 50, 100, 250, 500, 1000$ utilizando la distribución previa $Normal(0, 1)$. Se observa que cuando $\sigma_0 = 5$ la distancia D_{50} y D_{99} crece de forma lineal más rápido que cuando $\sigma_0 = 1$.

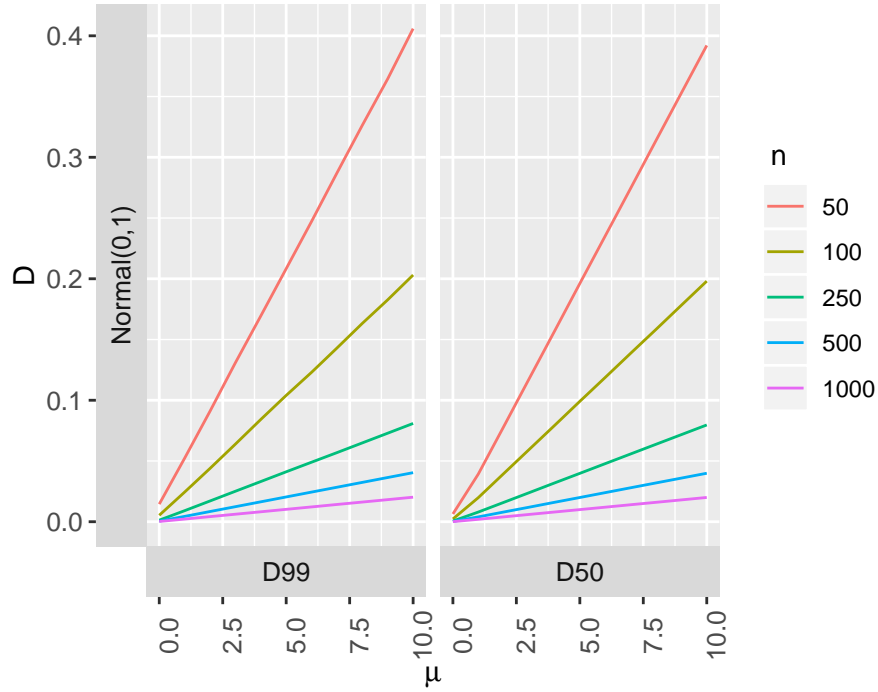


Figura 3–11: Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa $Normal(0, 1)$ en el caso de $\sigma_0 = 1$ conocido.

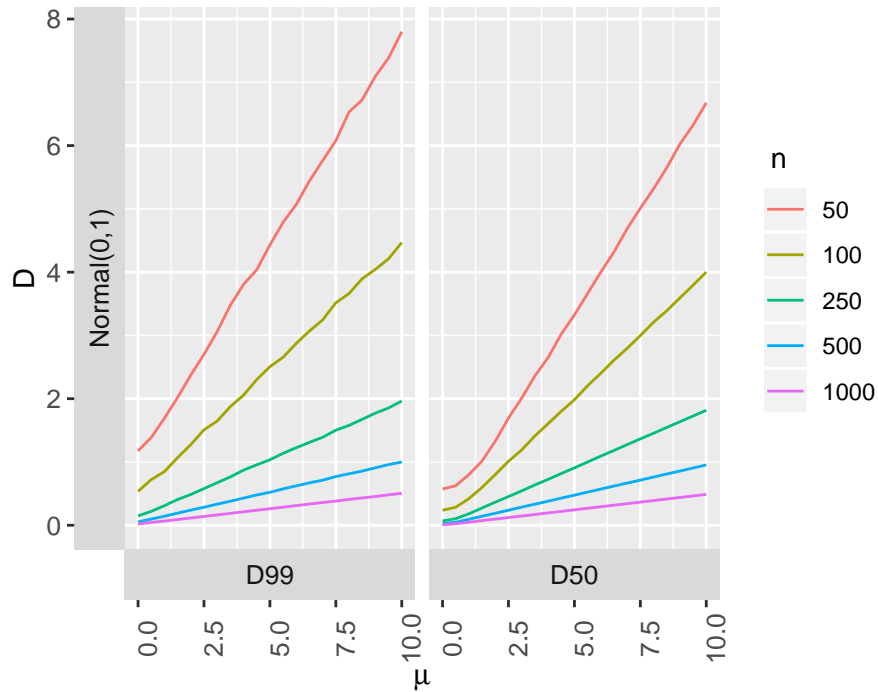


Figura 3–12: Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa $Normal(0, 1)$ en el caso de $\sigma_0 = 5$ conocido.

3.5.2. Inferencia en σ^2 con μ_0 conocido

En este caso, no se utilizará la distribución previa de Jeffrey debido a que con esta los intervalos de confianza y de credibilidad son equivalentes, lo que implica que la distancia calculada sobre los intervalos siempre es cero para cualquier tamaño de muestra (ver teorema 4 del apéndice A).

La Tabla 3-6, la Figura 3-13 y la Figura 3-14 exponen los tamaños de muestra n encontrados para el objetivo del estudio sobre el parámetro σ^2 en el caso $\mu_0 = 1, 5$ es conocido en el modelo Normal. Se utilizó el intervalo de confianza descrito en la Subsección 2.1.5 y el intervalo de credibilidad utilizando la distribución previa Invgamma conjugada (Ejemplo 2.13).

Se observa en la Tabla 3-6, con la distribución previa $Invgamma(1, 1)$ que el tamaño de muestra n_{50} y n_{99} son diferentes para cada uno de los casos de $\sigma^2 = 1, 5, 10$ fijando D_R y μ_0 . También se observa que el tamaño de muestra n_{50} y n_{99} no cambia de forma significativa cuando μ_0 cambia de 1 a 5.

Distribución Previa	σ^2	μ_0	D_R	D_{50}	n_{50}	D_{99}	n_{99}
Invgamma(1,1)	1	1	0.01	0.01	81	0.0098	149
			0.05	0.0482	31	0.0496	56
			0.1	0.1	21	0.1	37
		5	0.01	0.0098	81	0.01	143
			0.05	0.0471	31	0.05	55
			0.1	0.1	20	0.1	38
		5	0.05	0.05	331	0.0498	400
			0.25	0.25	74	0.25	102
			0.5	0.4986	41	0.4956	60
		5	0.05	0.05	331	0.05	401
			0.25	0.2492	74	0.25	101
			0.5	0.4972	41	0.4913	60
	10	1	0.1	0.1	370	0.0999	440
			0.5	0.5	81	0.492	109
			1	0.99	45	0.9	62
		5	0.1	0.0998	371	0.099	438
			0.5	0.5	81	0.5	108
			1	0.98	45	0.9993	63

Tabla 3-6: Tamaño de muestra obtenido para el parámetro $\sigma^2 = 1, 5, 10$ del modelo Normal.

Se observa en la Figura 3-12 ($\mu_0 = 1$ conocido) y la Figura 3-13 ($\mu_0 = 5$ conocido) que el tamaño de muestra es creciente cuando se utiliza D_{50} y D_{99} a medida que σ^2 crece además, se observa que el cambio de μ_0 de 1 a 5 no afecta el comportamiento del tamaño de muestra cuando se utiliza D_{50} y D_{99} .

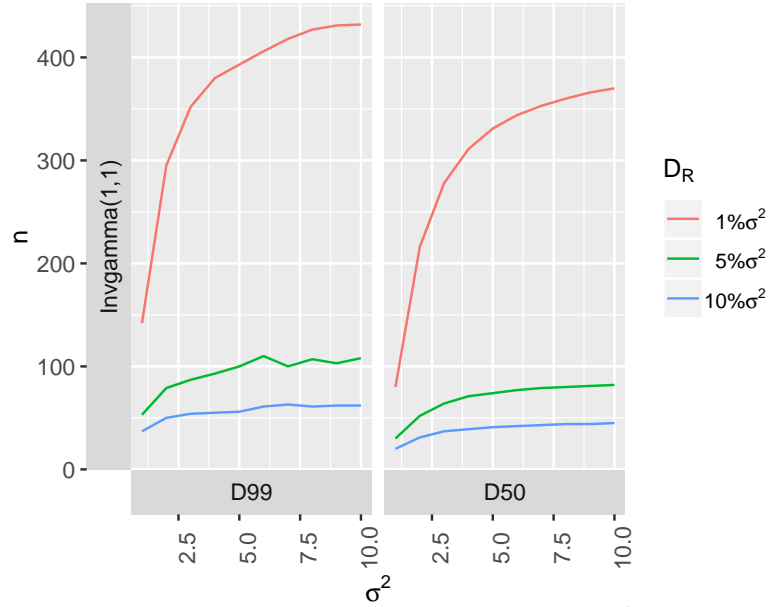


Figura 3-13: Tamaño de muestra n para el parámetro σ^2 del modelo Normal en el caso de $\mu_0 = 1$ conocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa $Invgamma(1, 1)$ (Ejemplo 2.15).

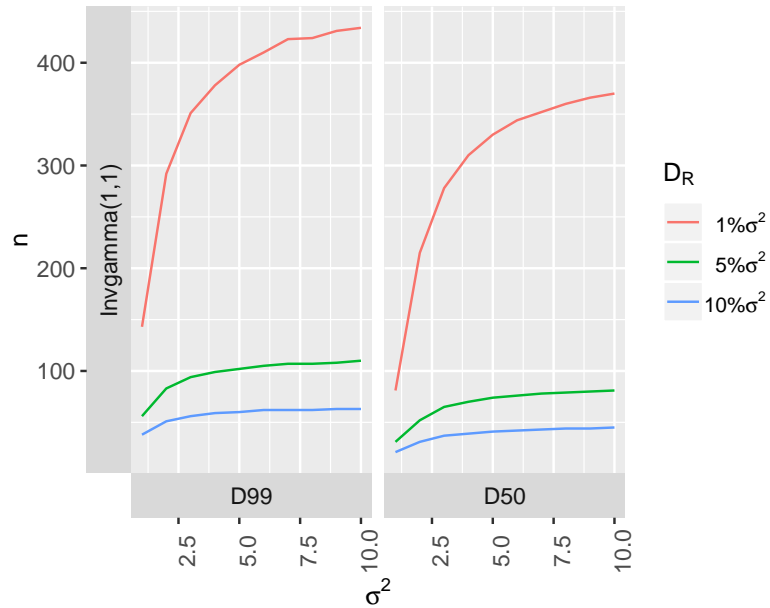


Figura 3-14: Tamaño de muestra n para el parámetro σ^2 del modelo Normal en el caso de $\mu_0 = 5$ conocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa $Invgamma(1, 1)$ (Ejemplo 2.15).

La Figura 3-14 ($\mu_0 = 1$ conocido) y la Figura 3-15 ($\mu_0 = 5$ conocido) muestra el comportamiento de la distancia a medida que se varía el parámetro σ^2 para un tamaño de muestra fijo $n = 50, 100, 250, 500, 1000$ utilizando la distribución previa $Invgamma(1, 1)$.

Se observa que la distancia D_{99} crece de forma lineal más rápido que la distancia D_{50} a medida que σ^2 varia de 1 a 10. Se observa que la distancia crece cuando el valor del parámetro crece y se fija el tamaño de muestra. También se observa que cuando n va a infinito, la distancia converge a cero.

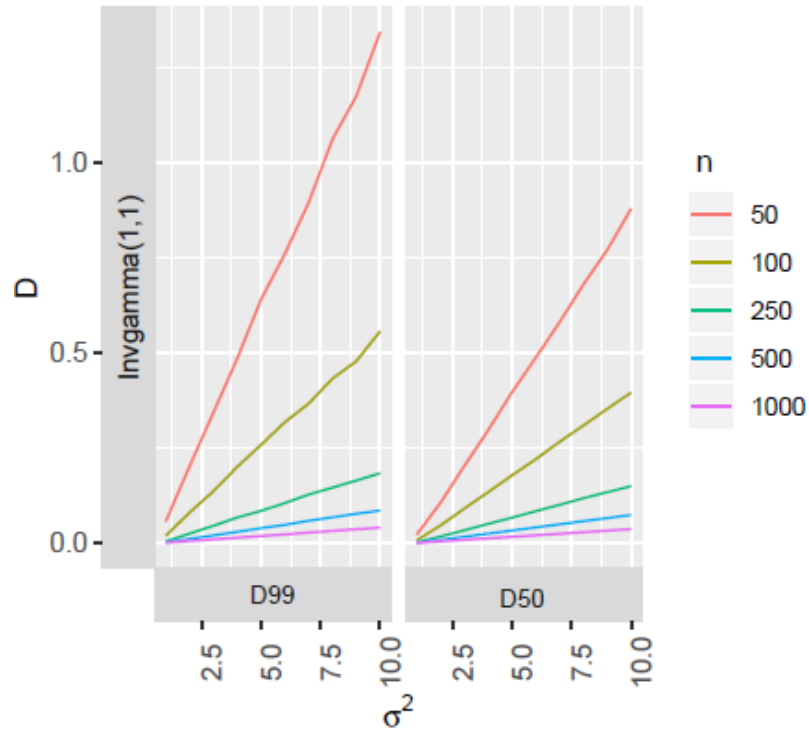


Figura 3-15: Comportamiento de la distancia en función del parámetro σ^2 y diferentes tamaños de muestra usando la distribución previa $Invgamma(1, 1)$ en el caso de $\mu_0 = 1$ conocido.

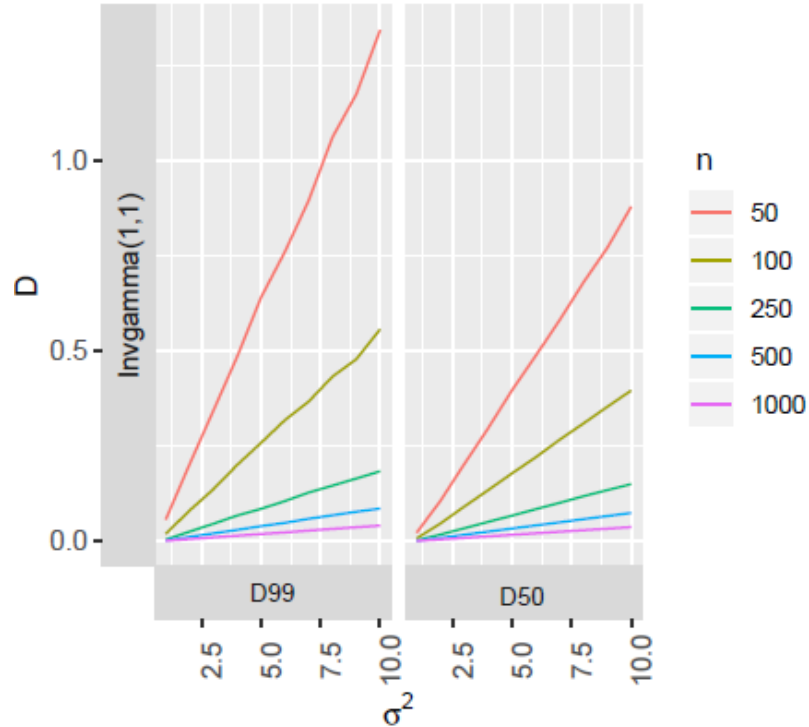


Figura 3-16: Comportamiento de la distancia en función del parámetro σ^2 y diferentes tamaños de muestra usando la distribución previa $Invgamma(1,1)$ en el caso de $\mu_0 = 5$ conocido.

3.5.3. Inferencia en μ con σ^2 desconocidos.

La Tabla 3-7, la Tabla 3-8, la Figura 3-17, la Figura 3-18, la Figura 3-19 y la Figura 3-20 exponen los tamaños de muestra n encontrados para el objetivo del estudio sobre el parámetro μ en el caso σ^2 desconocido. Se utilizó el intervalo de confianza descrito en la Subsección 2.1.5 y los intervalos de credibilidad utilizando las distribuciones previas del Ejemplo 2.7 y el Ejemplo 2.14.

Se observa en la tabla 3-7 que cuando μ es 5 y 10, los valores de n_{50} y n_{99} tienden a ser iguales, lo que implica para estos casos que la distancia de los límites superiores y la distancia de los límites inferiores de estos dos intervalos tienden a ser constantes.

Se observa en la tabla 3-8 que los valores de n_{50} y n_{99} son diferentes para cada uno de los casos de $\mu = 1, 5, 10$. También se observa que el tamaño de muestra n_{50}

tiende a ser constante para los valores de $\mu = 1, 5, 10$ cuando se fija σ^2 y un valor de $D_R = 1\% \mu, 5\% \mu, 10\% \mu$.

Distribución Previa	μ	σ^2	D_R	D_{50}	n_{50}	D_{99}	n_{99}
Jeffrey	1	1	0.01	0.009	29	0.009	120
			0.05	0.048	5	0.049	13
			0.1	0.09	4	0.097	8
		5	0.01	0.0097	107	0.01	560
			0.05	0.047	10	0.048	35
			0.1	0.09	6	0.09	14
		5	0.05	0.0497	5	0.048	13
			0.25	0.249	3	0.25	7
			0.5	0.25	3	0.31	4
		5	0.05	0.049	10	0.05	34
			0.25	0.24	4	0.249	8
			0.5	0.48	3	0.47	4
	10	1	0.1	0.084	4	0.08	8
			0.5	0.3	4	0.46	4
			1	0.0746	4	0.2239	4
		5	0.1	0.0842	8	0.0882	12
			0.5	0.1664	4	0.2694	5
			1	0.3744	4	0.4677	4

Tabla 3–7: Tamaño de muestra obtenido para el parámetro $\mu = 1, 5, 10$ del modelo Normal con σ^2 desconocido, utilizando la distribución previa de Jeffrey para μ y σ^2 (μ y σ^2 independientes).

Distribución Previa	μ	σ	D_R	D_{50}	n_{50}	D_{99}	n_{99}
$Normal(0, 1)$ $\times Invgamma(1, 1)$	1	1	0.01	0.0096	209	0.01	450
			0.05	0.492	44	0.5	71
			0.1	0.095	24	0.097	40
		5	0.01	0.01	991	0.0099	2178
			0.05	0.0498	197	0.0465	330
			0.1	0.0988	99	0.0969	168
		5	0.05	0.0496	206	0.05	261
			0.25	0.2413	45	0.2471	68
			0.5	0.47	25	0.49	37
		5	0.05	0.05	999	0.0495	1212
			0.25	0.25	200	0.25	250
			0.5	0.48	101	0.499	132
	10	1	0.1	0.098	206	0.096	254
			0.5	0.49	46	0.483	64
			1	0.95	30	0.89	41
		5	0.1	0.1	997	0.099	1134
			0.5	0.5	201	0.499	247
			1	0.99	104	1	131

Tabla 3–8: Tamaño de muestra obtenido para el parámetro $\mu = 1, 5, 10$ del modelo Normal con σ^2 desconocido, utilizando la distribución previa Conjugada para μ y σ^2 (μ y σ^2 independientes).

Se observa en la Figura 3-17 y la Figura 3-18 que se necesita un mayor tamaño de muestra cuando el valor de $\sigma^2 = 5$ que cuando $\sigma^2 = 1$. También se observa que

cuando μ es mayor que 5, el tamaño de muestra tiende a estabilizarse cerca de uno para cualquier valor de D_R .

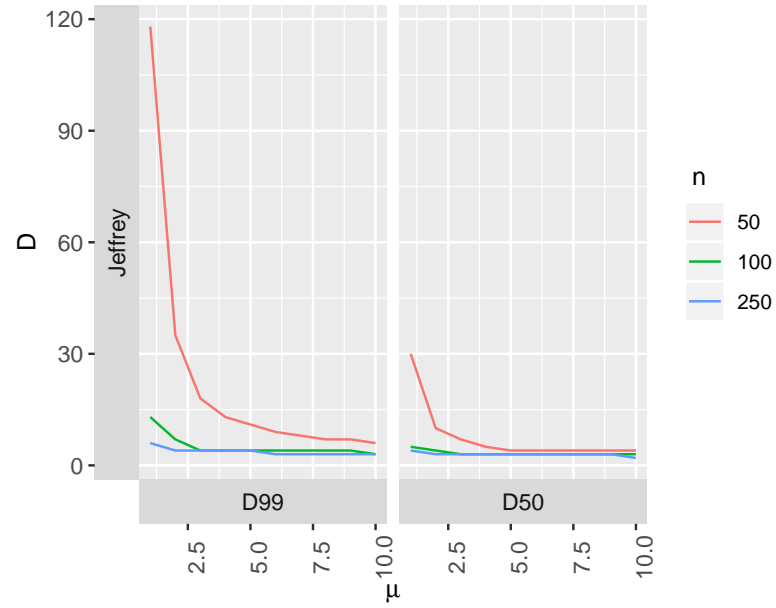


Figura 3–17: Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma^2 = 1$ desconocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa de Jeffrey (Ejemplo 2.14).

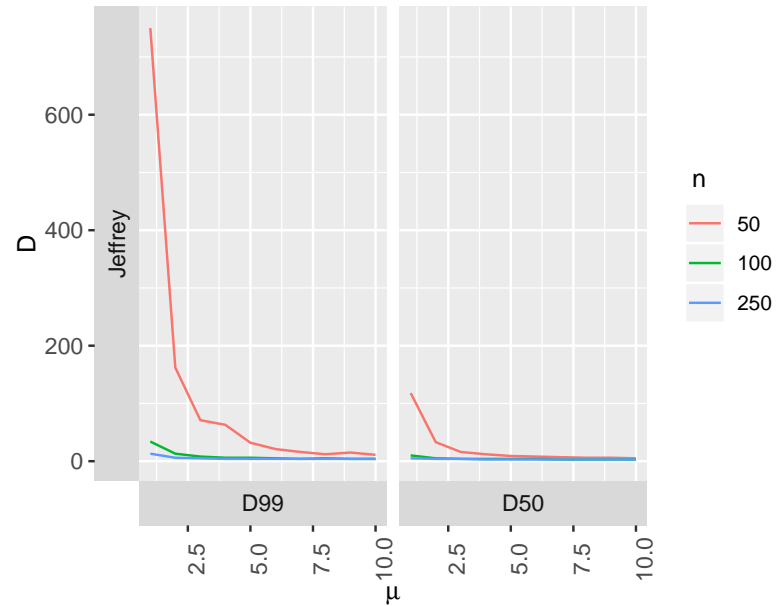


Figura 3–18: Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma^2 = 5$ desconocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa de Jeffrey (Ejemplo 2.14).

Se observa en la Figura 3-19 y la Figura 3-20 que se necesita un mayor tamaño de muestra cuando el valor de $\sigma^2 = 5$ comparado cuando $\sigma^2 = 1$. También se observa que cuando se utiliza D_{50} , el comportamiento del tamaño de muestra tiende a ser constante y cuando se utiliza D_{99} y $D_R = 1\%\mu$ el tamaño de muestra tiende a decrecer y luego a estabilizarse después de $\mu = 5$, en los demás valores de D_R el tamaño de muestra tiende a ser constante.

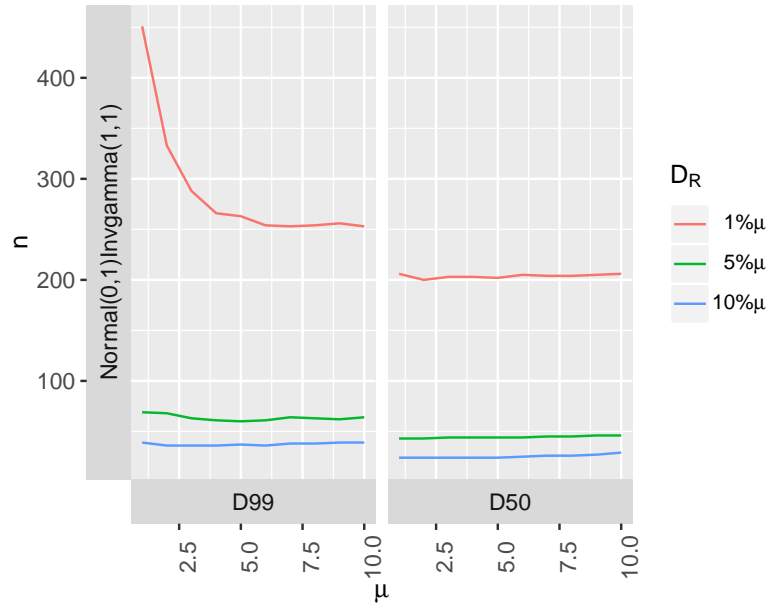


Figura 3–19: Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma^2 = 1$ desconocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa descrita en el Ejemplo 2.14.

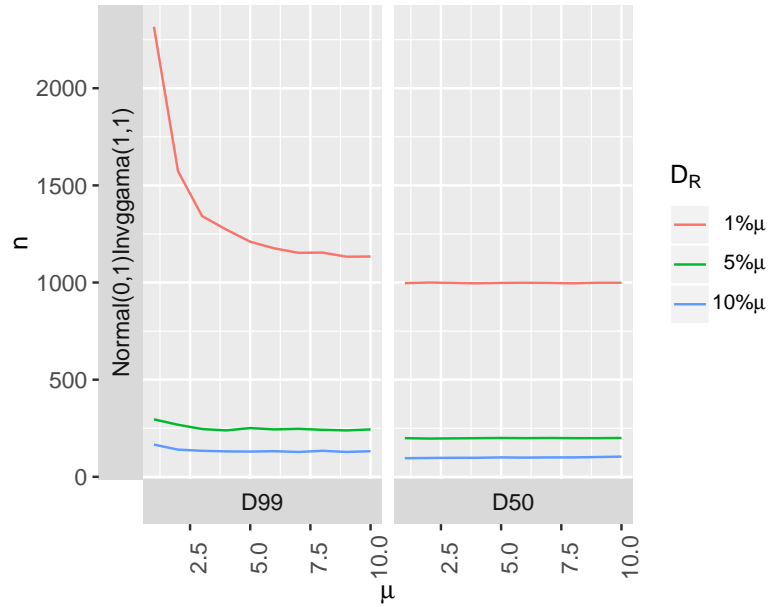


Figura 3-20: Tamaño de muestra n para el parámetro μ del modelo Normal en el caso de $\sigma^2 = 5$ desconocido, utilizando la Distancia D_{50} y D_{99} para comparar los intervalos de confianza (Subsección 2.1.5) y los intervalos de credibilidad utilizando la distribución previa descrita en el Ejemplo 2.14.

La Figura 3-21, la Figura 3-22, la Figura 3-23 y la Figura 3-24 muestran el comportamiento de la distancia D_{50} y D_{99} a medida que se varía el parámetro μ para un tamaño de muestra fijo $n = 50, 100, 250, 500, 1000$ utilizando la distribución previa descrita en el Ejemplo 2.7 y el Ejemplo 2.14.

Se observa en la Figura 3-21 y la Figura 3-22, que la distancia D_{50} y D_{99} tienden a ser constantes cuando se fija n y $\sigma^2 = 1, 5$. También se observa que cuando n va a infinito, la distancia converge a cero.

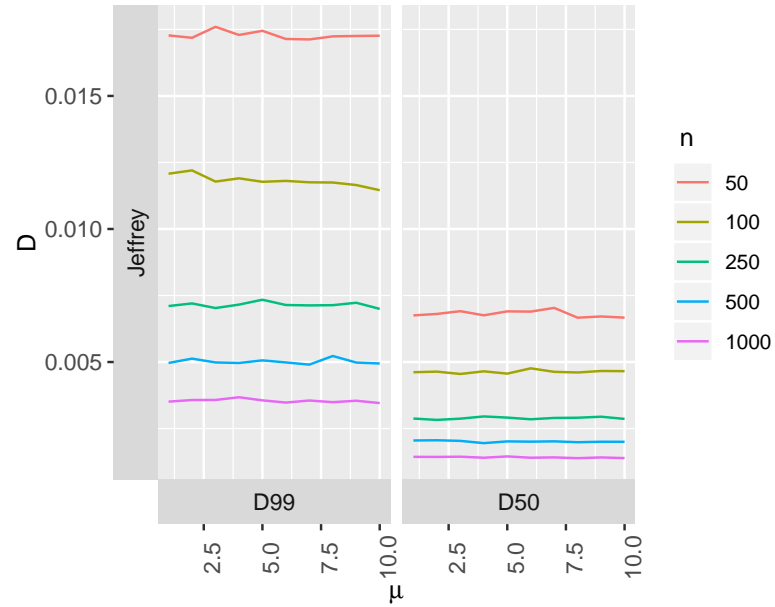


Figura 3–21: Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa Jeffrey (Ejemplo 2.7) en el caso de $\sigma^2 = 1$.

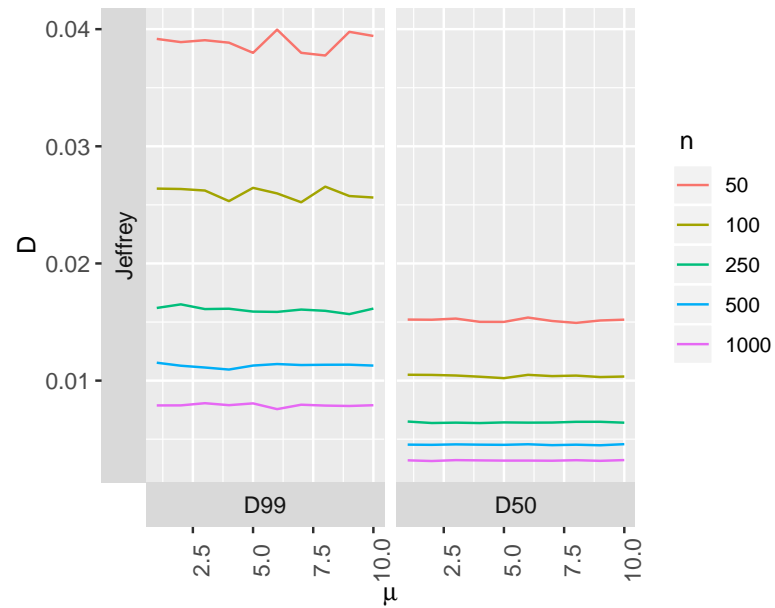


Figura 3–22: Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa Jeffrey (Ejemplo 2.7) en el caso de $\sigma^2 = 5$.

Se observa en la figura 3-23 y la figura 3-24 que la Distancia D_{99} crece más rápido que la distancia D_{50} a medida que μ varia de 1 a 10. Se observa que la

distancia crece cuando el valor del parámetro crece y se fija el tamaño de muestra n . También se observa que cuando n va a infinito, la distancia converge a cero.

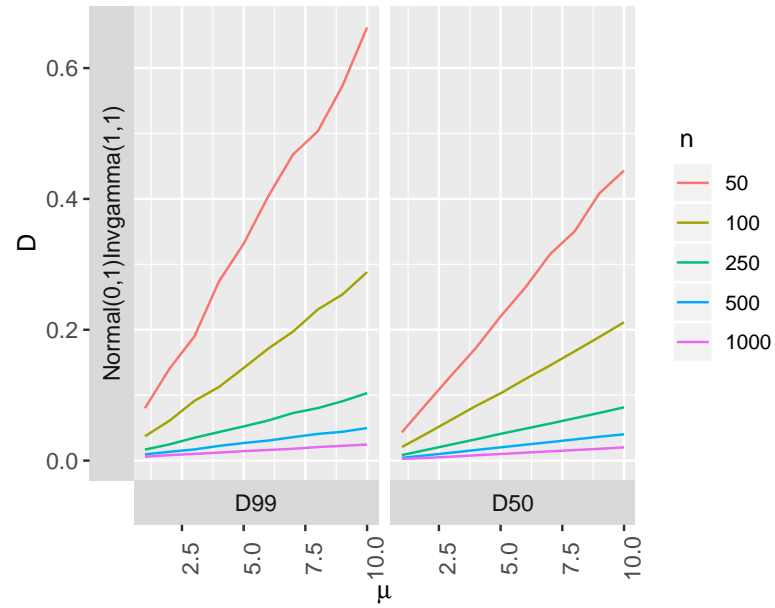


Figura 3–23: Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa descrita en el Ejemplo 2.14, en el caso de $\sigma^2 = 1$.

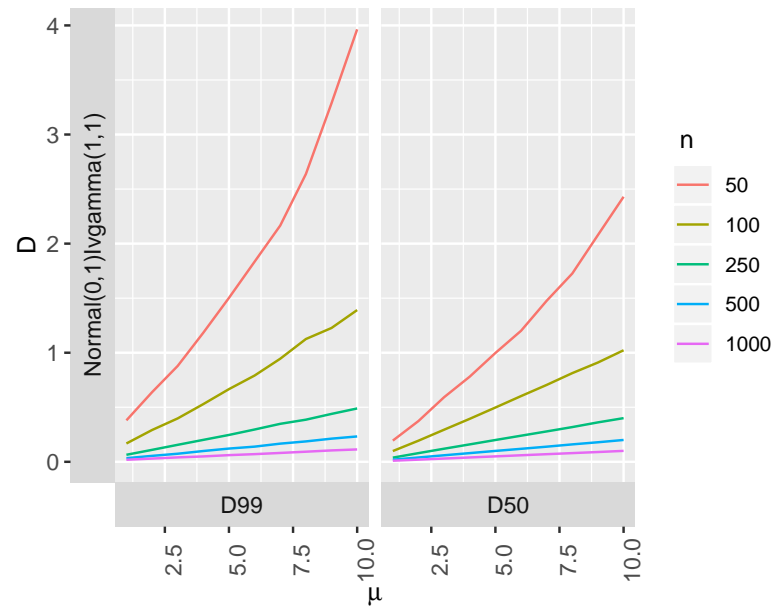


Figura 3–24: Comportamiento de la distancia en función del parámetro μ y diferentes tamaños de muestra usando la distribución previa descrita en el Ejemplo 2.14, en el caso de $\sigma^2 = 5$.

Capítulo 4

LA APLICACIÓN WEB INTERACTIVA EN LÍNEA

[R Shiny](#) es una herramienta que permite crear aplicaciones web interactivas utilizando R, en donde un usuario puede interactuar directamente con la aplicación siguiendo una secuencia de pasos sencillos sin tener la necesidad de conocer el código fuente creado en R. Las aplicaciones creadas en Shiny pueden ser publicadas en una página web, lo cual permite que cualquier usuario pueda acceder a la aplicación.

En el presente capítulo, describimos una aplicación creada en shiny que se encuentra disponible en la siguiente dirección: https://server-deiver.shinyapps.io/sample_size_frequentist_and_bayesian/, la cual hemos utilizado para obtener los resultados de las tablas del Capítulo 3.

Esta aplicación, en la vista principal, tiene dos opciones para escoger:

1. Calculate your own intervals.
2. Find the sample size required.

En la primera parte, se pueden calcular intervalos de confianza e intervalos de credibilidad para los modelos Binomial, Poisson, Geometric, Exponential y Normal. En la segunda parte, se puede realizar una comparación vía simulación entre los intervalos de confianza y los intervalos de credibilidad para los modelos mencionados en la primera parte, encontrando el tamaño de muestra n con el cual la distancia (métrica del taxista, ver Definición 2.2.8) entre los intervalos de confianza e intervalos

de credibilidad sea aproximadamente igual a una distancia requerida (ingresada por el usuario). El nivel de confianza y de credibilidad utilizado es del 95 %.

Los intervalos de confianza que utiliza el App son los intervalos descritos en la Sección 2.1.

A continuación, una descripción en detalle.

4.1. Calculate your own intervals.

En esta parte, se puede seleccionar uno de los cinco modelos (**Binomial**, **Poisson**, **Geometric**, **Exponential** y **Normal**) en el cual se quiere calcular intervalos de confianza y de credibilidad: “**Choose Data Distribution:**”

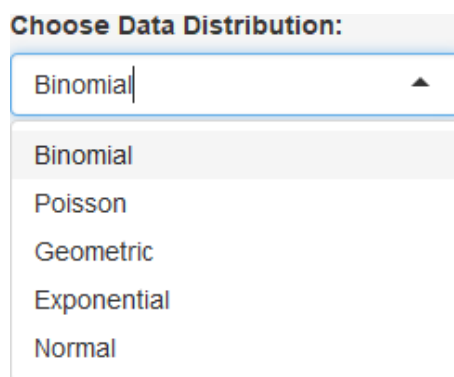
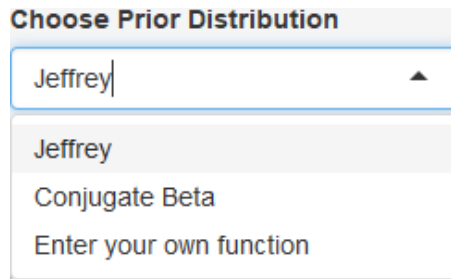


Figura 4–1: Menú de distribuciones muestrales para realizar estimación por intervalos

✓ **Binomial.**

En esta parte, se puede seleccionar la distribución previa que se utiliza para calcular los intervalos de credibilidad sobre el parámetro p en el modelo Binomial. Se presentan tres formas de seleccionar la distribución previa: “**Choose Prior Distribution:**”



Choose Prior Distribution

Jeffrey

Conjugate Beta

Enter your own function

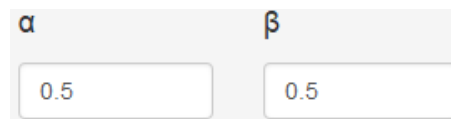
Figura 4–2: Selección de la distribución previa para el modelo Binomial

✓ Jeffrey

Aquí, se utiliza la distribución previa de Jeffrey para el parámetro p del modelo Binomial (ver Ejemplo 2.3).

✓ Conjugate Beta

Aquí, se utiliza la distribución previa Conjugate Beta para el parámetro p del modelo Binomial: “ $Beta(\alpha, \beta)$ (ver Ejemplo 2.9)”, donde α y β pueden ser ingresados por el usuario:



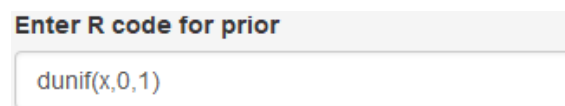
α β

0.5 0.5

Figura 4–3: Parámetros de la previa Beta conjugada para el modelo Binomial

✓ Enter your own function

Aquí, el usuario puede ingresar su propia distribución previa para el parámetro p del modelo Binomial: “Enter your own function:”



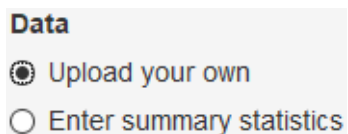
Enter R code for prior

dunif(x,0,1)

Figura 4–4: Ingrese su propia distribución previa para el modelo Binomial

Una vez seleccionada la distribución previa, podemos ingresar un conjunto de datos o ingresar información de un conjunto de datos para calcular los intervalos de confianza y de credibilidad sobre el parámetro p del modelo binomial. Tenemos dos

formas de ingresar los datos en el panel: “Data”



Data

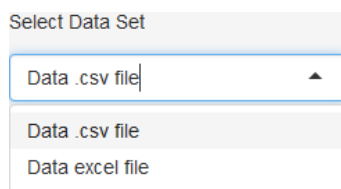
☒ Upload your own

☐ Enter summary statistics

Figura 4–5: Formas de ingresar datos en el panel para el modelo Binomial

☒ Upload your own

En la parte inferior del panel se despliega “Select Data Set”. En esta parte es posible importar datos provenientes de un modelo Bernoulli desde un archivo (.csv) o un archivo (.xlsx). La forma de importar los datos debe ser en un archivo con los datos organizados en una columna.



Select Data Set

Data .csv file

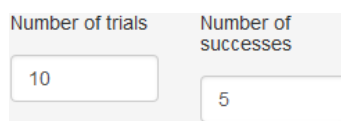
Data .csv file

Data excel file

Figura 4–6: Formato para ingresar datos del modelo Binomial

☐ Enter summary statistics

Como alternativa al importar los datos, es posible también ingresar las medidas de resumen de los datos para calcular los intervalos de confianza y de credibilidad, que para el caso del parámetro p en el modelo Binomial son:



Number of trials	Number of successes
10	5

Figura 4–7: Medidas de resumen suficientes del modelo Binomial

Number of trials: Es el número de observaciones(n) Bernoulli

Number of successes: número de éxitos en las n observaciones Bernoulli.

Con el botón **CALCULATE**, se calculan los intervalos de confianza y de credibilidad. Aparecerán en la parte derecha de la pantalla los intervalos, el cálculo de la distancia de dichos intervalos (utilizando la métrica del taxista) y una gráfica de los mismos.

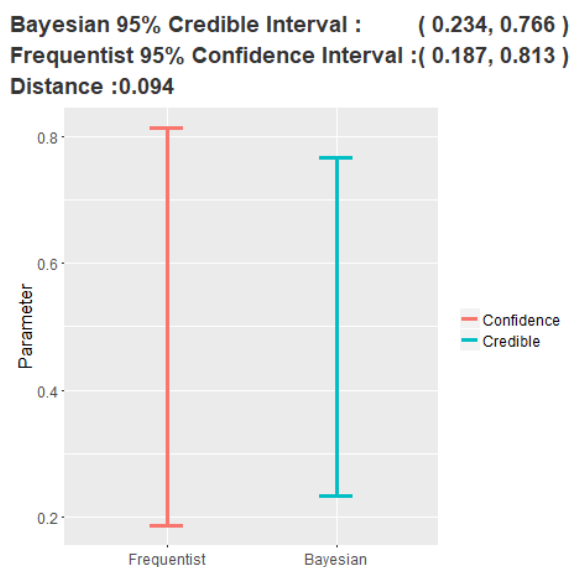
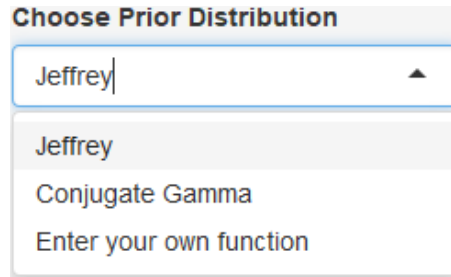


Figura 4-8: Intervalos y gráfica de intervalos del modelo Binomial

✓ Poisson.

En esta parte, se puede seleccionar la distribución previa que se utiliza para calcular los intervalos de credibilidad sobre el parámetro λ en el modelo Poisson. Se presentan tres formas de seleccionar la distribución previa: “Choose Prior Distribution:”



Choose Prior Distribution

Jeffrey

Conjugate Gamma

Enter your own function

Figura 4–9: Selección de la distribución previa para el modelo Poisson

✓ Jeffrey

Aquí, se utiliza la distribución previa de Jeffrey para el parámetro λ del modelo Poisson (ver Ejemplo 2.1).

✓ Conjugate Gamma

Aquí, se utiliza la distribución previa Conjugate Gamma para el parámetro λ del modelo Poisson: “ $\text{Gamma}(\alpha, \beta)$ (ver Ejemplo 2.8)”, donde α y β pueden ser ingresados por el usuario:



α

1

β

2

Figura 4–10: Parámetros de la previa Gamma conjugada para el modelo Poisson

✓ Enter your own function

Aquí, el usuario puede ingresar su propia distribución previa para el parámetro p del modelo Poisson: “Enter your own function:”



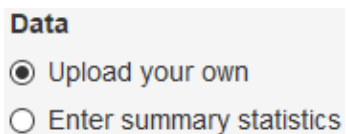
Enter R code for prior

dgamma(x,1,2)

Figura 4–11: Ingrese su propia distribución previa para el modelo Poisson

Una vez seleccionada la distribución previa, podemos ingresar un conjunto de datos o ingresar información de un conjunto de datos para calcular los intervalos de confianza y de credibilidad sobre el parámetro λ de un modelo Poisson. Tenemos

dos formas de ingresar los datos en el panel: “Data”



Data

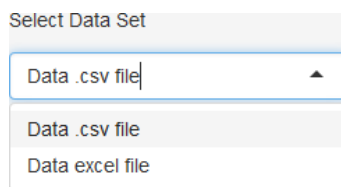
☒ Upload your own

☐ Enter summary statistics

Figura 4–12: Formas de ingresar datos en el panel para el modelo Poisson

⊙ Upload your own

En la parte inferior del panel se despliega “Select Data Set”. En esta parte es posible importar datos provenientes de un modelo Poisson desde un archivo (.csv) o un archivo (.xlsx). La forma de importar los datos debe ser en un archivo con los datos organizados en una columna.



Select Data Set

Data .csv file

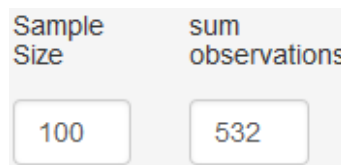
Data .csv file

Data excel file

Figura 4–13: Formato para ingresar datos del modelo Poisson

⊙ Enter summary statistics

Como alternativa al importar los datos, es posible también ingresar las medidas de resumen de los datos para calcular los intervalos de confianza y de credibilidad, que para el caso del parámetro λ en el modelo Poisson son:



Sample Size	sum observations
100	532

Figura 4–14: Medidas de resumen suficientes del modelo Poisson

Sample Size: Es el número de observaciones (n) provenientes de un modelo Poisson.

Sum observations: Suma de n observaciones provenientes de un modelo Poisson.

Con el botón **CALCULATE**, se calculan los intervalos de confianza y de credibilidad. Aparecerán en la parte derecha de la pantalla los intervalos, el cálculo de la distancia de dichos intervalos (utilizando la métrica del taxista) y una gráfica de los mismos.

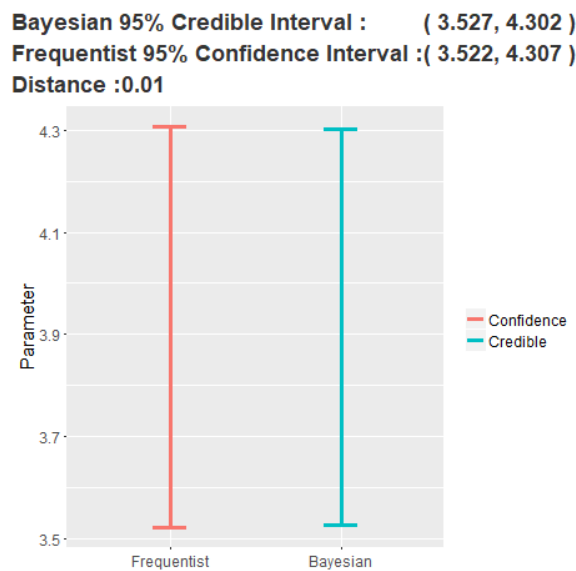


Figura 4–15: Intervalos y gráfica de intervalos del modelo Poisson

✓ Geometric.

En esta parte, se puede seleccionar la distribución previa que se utiliza para calcular los intervalos de credibilidad sobre el parámetro p en el modelo Geometric. Se presentan tres formas de seleccionar la distribución previa: “Choose Prior Distribution:”

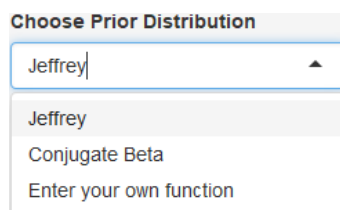


Figura 4–16: Selección de la distribución previa para el modelo Geométrico

✓ Jeffrey

Aquí, se utiliza la distribución previa de Jeffrey para el parámetro p del modelo Geometric (ver Ejemplo 2.4)

✓ Conjugate Beta

Aquí, se utiliza la distribución previa Conjugate Beta para el parámetro p del modelo Geometric: “ $Beta(\alpha, \beta)$ (ver Ejemplo 2.11)”, donde α y β pueden ser ingresados por el usuario:

Figura 4–17: Parámetros de la previa Beta conjugada para el modelo Geométrico

✓ Enter your own function

Aquí, el usuario puede ingresar su propia distribución previa para el parámetro p del modelo Geometric: “Enter your own function:”

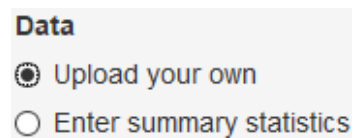


Enter R code for prior

dbeta(x,1,1)

Figura 4–18: Ingrese su propia distribución previa para el modelo Geométrico

Una vez seleccionada la distribución previa, podemos ingresar un conjunto de datos o ingresar información de un conjunto de datos para calcular los intervalos de confianza y de credibilidad para el parámetro p de un modelo Geometric. Tenemos dos formas de ingresar los datos en el panel: “Data”



Data

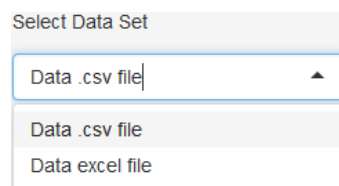
☒ Upload your own

☐ Enter summary statistics

Figura 4–19: Formas de ingresar datos en el panel para el modelo Geométrico

☒ Upload your own

En la parte inferior del panel se despliega “Select Data Set”. En esta parte es posible importar datos provenientes de un modelo Geometric desde un archivo (.csv) o un archivo (.xlsx). La forma de importar los datos debe ser en un archivo con los datos organizados en una columna.



Select Data Set

Data .csv file

Data .csv file

Data excel file

Figura 4–20: Formato para ingresar datos del modelo Geométrico

☐ Enter summary statistics

Como alternativa al importar los datos, es posible también ingresar las medidas de resumen de los datos para calcular los intervalos de confianza y de credibilidad, que para el caso del parámetro p en el modelo Geometric son:

Sample Size	sum observations
100	314

Figura 4-21: Medidas de resumen suficientes del modelo Geométrico

Sample Size: Es el número de observaciones(n) provenientes de un modelo Geometric.

Sum observations: Suma de n observaciones provenientes de un modelo Geometric.

Con el botón **CALCULATE**, se calculan los intervalos de confianza y de credibilidad. Aparecerán en la parte derecha de la pantalla los intervalos, el cálculo de la distancia de dichos intervalos (utilizando la métrica del taxista) y una gráfica de los mismos.

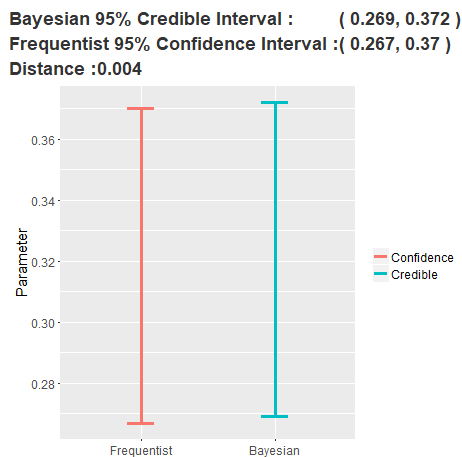


Figura 4-22: Intervalos y gráfica de intervalos del modelo Geométrico

✓ Exponetial.

En esta parte, se puede seleccionar la distribución previa que se utiliza para calcular los intervalos de credibilidad sobre el parámetro λ en el modelo Exponential. Se presentan tres formas de seleccionar la distribución previa: “Choose Prior Distribution:”

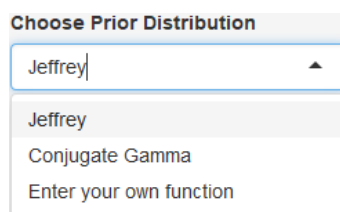


Figura 4–23: Selección de la distribución previa para el modelo Exponencial

✓ Jeffrey

Aquí, se utiliza la distribución previa de Jeffrey para el parámetro λ del modelo Exponential (ver Ejemplo 2.2)

✓ Conjugate Gamma

Aquí, se utiliza la distribución previa Conjugate Gamma para el parámetro λ del modelo Exponential: “ $\text{Gamma}(\alpha, \beta)$ (ver Ejemplo 2.10)”, donde α y β pueden ser ingresados por el usuario:

α	β
1	1

Figura 4–24: Parámetros de la previa Gamma conjugada para el modelo Exponencial

✓ Enter your own function

Aquí, el usuario puede ingresar su propia distribución previa para el parámetro λ del modelo Exponential: “Enter your own function:”

Enter R code for prior

```
dgamma(x,0.5,0.5)
```

Figura 4–25: Ingrese su propia distribución previa para el modelo Exponencial

Una vez seleccionada la distribución previa, podemos ingresar un conjunto de datos o ingresar información de un conjunto de datos para calcular los intervalos de confianza y de credibilidad sobre el parámetro λ de un modelo Exponential. Tenemos dos formas de ingresar los datos en el panel: “Data”

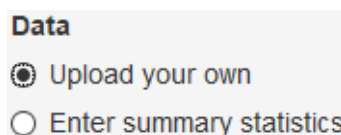


Figura 4–26: Formas de ingresar datos en el panel para el modelo Exponencial

⊙ Upload your own

En la parte inferior del panel se despliega “Select Data Set”. En esta parte, es posible importar datos provenientes de un modelo Exponential desde un archivo (.csv) o un archivo (.xlsx). La forma de importar los datos debe ser en un archivo con los datos organizados en una columna.

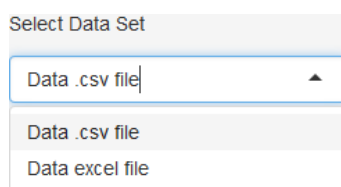


Figura 4–27: Formato para ingresar datos del modelo Exponencial

⊙ Enter summary statistics

Como alternativa al importar los datos, es posible también ingresar las medidas de resumen de los datos para calcular los intervalos de confianza y de credibilidad, que para el parámetro λ en el modelo Exponential son:

Sample Size	sum observations
20	46.93

Figura 4–28: Medidas de resumen suficientes del modelo Exponencial

Sample Size: Es el número de observaciones(n) provenientes de un modelo Exponential.

Sum observations: Suma de n observaciones provenientes de un modelo Exponential.

Con el botón **CALCULATE**, se calculan los intervalos de confianza y de credibilidad. Aparecerán en la parte derecha de la pantalla los intervalos, el cálculo de la distancia de dichos intervalos (utilizando la métrica del taxista) y una gráfica de los mismos.

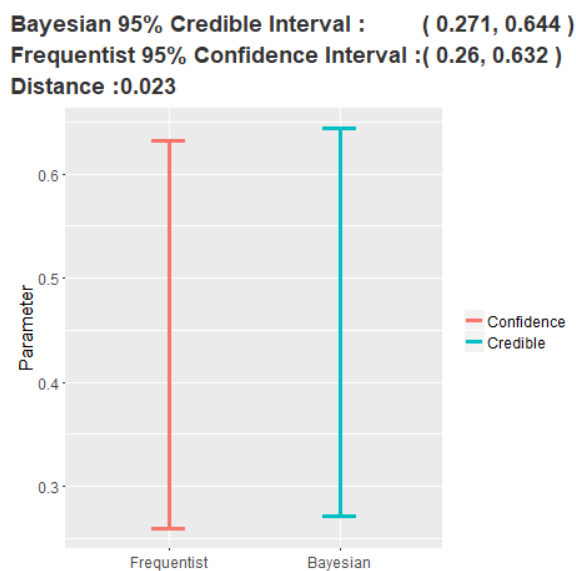


Figura 4-29: Intervalos y gráfica de intervalos del modelo Exponencial

✓ Normal.

En esta parte, tenemos tres casos para realizar intervalos de confianza y de credibilidad para el modelo Normal: “Select the parameter”

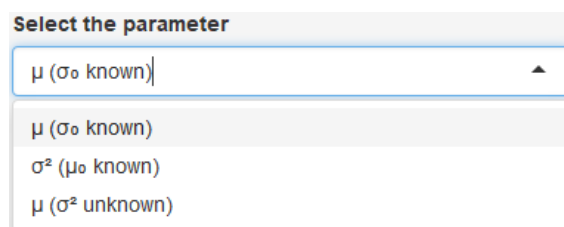


Figura 4–30: Inferencia en el modelo Normal

✓ $\mu(\sigma_0 \text{ known})$

En esta parte, se puede ingresar el valor de “ $\sigma_0(known)$ ”: y se puede seleccionar la distribución previa que se utiliza para calcular los intervalos de credibilidad en el modelo Normal. En el caso de $\mu(\sigma_0 \text{ known})$ se presentan tres formas de seleccionar la distribución previa de μ : “Choose Prior Distribution:”

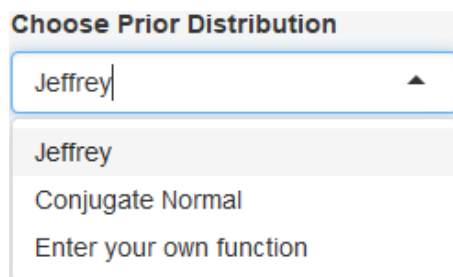


Figura 4–31: Selección de la distribución previa para el modelo Normal en el caso de μ con σ_0 conocido

✓ Jeffrey

Aquí, se utiliza la distribución previa de Jeffrey para el parámetro μ del modelo Normal con σ conocido (ver Ejemplo 2.5)

✓ Conjugate Normal

Aquí, se utiliza la distribución previa Conjugate Normal para el parámetro μ del modelo Normal con σ conocido: “ $Norm(\mu_1, \sigma_1)$ ”(ver Ejemplo 2.12)”, donde μ_1 y

σ_1 pueden ser ingresados por el usuario:

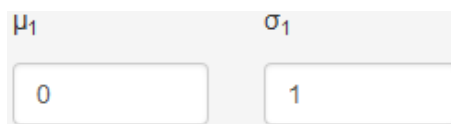


Figura 4–32: Parámetros de la previa Normal conjugada para el modelo Normal el caso de μ con σ_0 conocido

✓ Enter your own function

Aquí, el usuario puede ingresar su propia distribución previa para el parámetro μ del modelo Normal con σ conocido: “Enter your own function:”



Figura 4–33: Ingrese su propia distribución previa para el modelo Normal el caso de μ con σ_0 conocido

Una vez seleccionada la distribución previa, podemos ingresar un conjunto de datos o ingresar información de un conjunto de datos para calcular los intervalos de confianza y de credibilidad para el parámetro μ del modelo Normal en el caso de que σ es conocido. Tenemos dos formas de ingresar los datos en el panel: “Data”

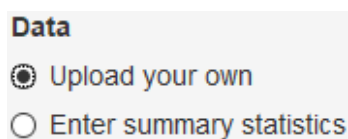
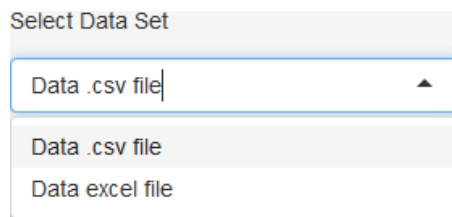


Figura 4–34: Formas de ingresar datos en el panel para el modelo Normal el caso de μ con σ_0 conocido

⊙ Upload your own

En la parte inferior del panel se despliega “Select Data Set”. En esta parte es posible importar datos provenientes de un modelo Normal desde un archivo (.csv) o un archivo (.xlsx). La forma de importar los datos debe ser en un archivo con los datos organizados en una columna.



Select Data Set

Data .csv file

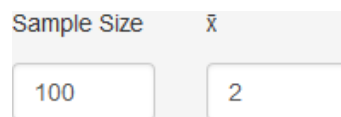
Data .csv file

Data excel file

Figura 4–35: Formato para ingresar datos del modelo Normal el caso de μ con σ_0 conocido

⊙ Enter summary statistics

Como alternativa al importar los datos, es posible también ingresar las medidas de resumen de los datos para calcular los intervalos de confianza y de credibilidad, que para el parámetro μ , modelo Normal con σ conocido, son:



Sample Size \bar{x}

100 2

Figura 4–36: Medidas de resumen suficientes del modelo Normal el caso de μ con σ_0 conocido

Sample Size: Es el número de observaciones (n) provenientes de un modelo Normal.

\bar{x} : Promedio de n observaciones provenientes de un modelo Normal.

Con el botón **CALCULATE**, se calculan los intervalos de confianza y de credibilidad. Aparecerán en la parte derecha de la pantalla los intervalos, el cálculo de la distancia de dichos intervalos (utilizando la métrica del taxista) y una gráfica de los mismos.

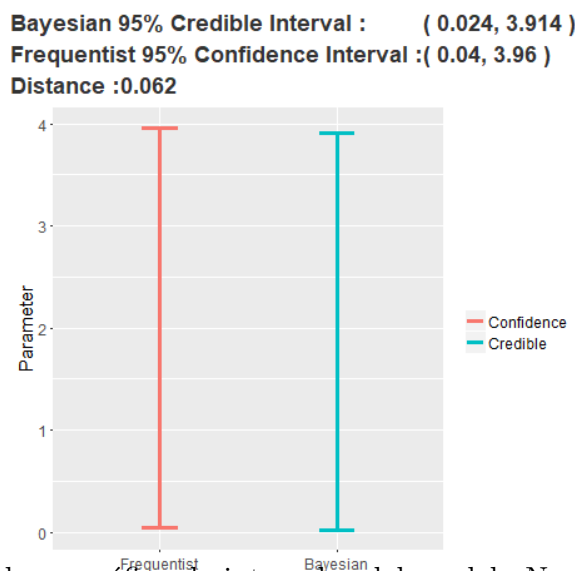


Figura 4–37: Intervalos y gráfica de intervalos del modelo Normal el caso de μ con σ_0 conocido

✓ $\sigma^2(\mu_0 \text{ known})$

En esta parte, se puede seleccionar la distribución previa que se utiliza para calcular los intervalos de credibilidad en el modelo Normal. En el caso de $\sigma^2(\mu_0 \text{ known})$ se presentan dos formas de seleccionar la distribución previa de σ^2 : “Choose Prior Distribution:”

Choose Prior Distribution

Jeffrey

Jeffrey

Conjugate Ivgamma

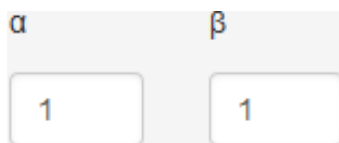
Figura 4–38: Selección de la distribución previa para el modelo Normal el caso de σ^2 con μ_0 conocido

✓ Jeffrey

Aquí, se utiliza la distribución previa de Jeffrey para el parámetro σ^2 del modelo Normal con μ_0 conocido (ver Ejemplo 2.6)

✓ Conjugate Invgamma

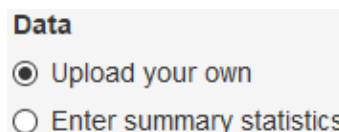
Aquí, se utiliza la distribución previa Conjugate Ivgamma para el parámetro σ^2 del modelo Normal con μ conocido: “*Ivgamma*(α, β) (ver Ejemplo 2.13)”, donde α y β pueden ser ingresados por el usuario:



The image shows a form with two input fields. The first field is labeled with the Greek letter alpha (α) and contains the number 1. The second field is labeled with the Greek letter beta (β) and also contains the number 1.

Figura 4–39: Parámetros de la previa Invgamma conjugada para el modelo Normal el caso de σ^2 con μ_0 conocido

Una vez seleccionada la distribución previa, podemos ingresar un conjunto de datos o ingresar información de un conjunto de datos para calcular los intervalos de confianza y de credibilidad para el parámetro σ^2 del modelo Normal en el caso de que μ es conocido. Tenemos dos formas de ingresar los datos en el panel: “Data”

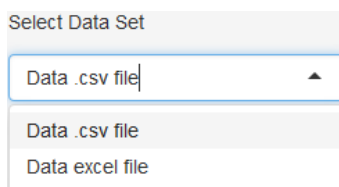


The image shows a section titled "Data" with two radio button options. The first option, "Upload your own", is selected with a filled radio button. The second option, "Enter summary statistics", is unselected with an empty radio button.

Figura 4–40: Formas de ingresar datos en el panel para el modelo Normal el caso de σ^2 con μ_0 conocido

⊙ Upload your own

En la parte inferior del panel se puede ingresar el valor de “ $\mu_0(know)$ ” : . También se despliega “Select Data Set”. En esta parte es posible importar datos provenientes de un modelo Normal desde un archivo (.csv) o un archivo (.xlsx). La forma de importar los datos debe ser en un archivo con los datos organizados en una columna.



The image shows a dropdown menu titled "Select Data Set". The selected option is "Data .csv file". Below the dropdown, there are two more options: "Data .csv file" and "Data excel file".

Figura 4–41: Formato para ingresar datos del modelo Normal el caso de σ^2 con μ_0 conocido

⊙ Enter summary statistics

Como alternativa al importar los datos, es posible también ingresar las medidas de resumen de los datos para calcular los intervalos de confianza y de credibilidad, que para el parámetro σ^2 , modelo Normal con μ_0 conocido, son:

Sample Size	s^2 (use μ_0)
100	1

Figura 4–42: Medidas de resumen suficientes del modelo Normal el caso de σ^2 con μ_0 conocido

Sample Size: Es el número de observaciones (n) provenientes de un modelo Normal.

s^2 : varianza de n observaciones provenientes de un modelo Normal, calculada utilizando μ .

Con el botón **CALCULATE**, se calculan los intervalos de confianza y de credibilidad. Aparecerán en la parte derecha de la pantalla los intervalos, el calculo de la distancia de dichos intervalos (utilizando la métrica del taxista) y una gráfica de los mismos.

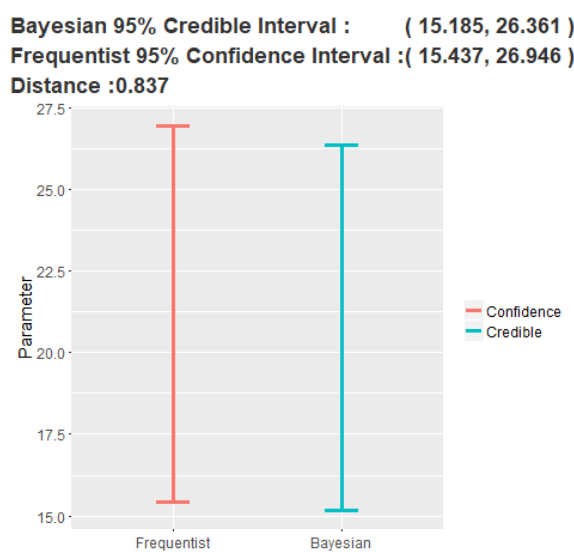


Figura 4–43: Intervalos y gráfica de intervalos del modelo Normal el caso de σ^2 con μ_0 conocido

✓ $\mu(\sigma^2 \text{ unknown})$

En esta parte, se pueden seleccionar las distribuciones previas que se utilizan para calcular los intervalos de credibilidad en el modelo Normal. En el caso de μ (σ^2 unknown) se presentan dos formas de seleccionar la distribución previa de μ y σ^2 : “Choose Prior Distribution:”

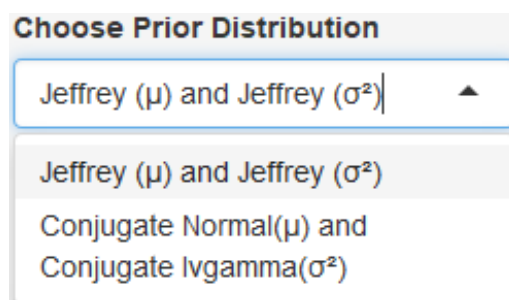


Figura 4–44: Selección de la distribución previa para el modelo Normal el caso de μ con σ^2 desconocido

✓ Jeffrey (μ) and Jeffrey (σ^2)

Aquí, se utilizan las distribuciones previas de Jeffrey para los parámetros μ y σ^2 del modelo Normal, con μ y σ^2 desconocidos e independientes (ver Ejemplo 2.7)

✓ Conjugate Normal (μ) and Conjugate Ivgamma (σ^2)

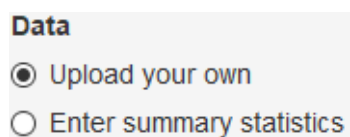
Aquí, se utilizan las distribuciones previas conjugadas para los parámetros μ y σ^2 del modelo Normal, con μ y σ^2 desconocidos e independientes: “ $Norm(\mu_1, \sigma_1^2)$ y $Ivgamma(\alpha, \beta)$ (ver Ejemplo 2.14)”, donde μ_1 , σ_1^2 , α y β pueden ser ingresados por el usuario.

μ_1 (Normal)	σ_1 (Normal)	α (Ivgamma)	β (Ivgamma)
0	1	0.5	0.5

Figura 4–45: Parámetros de la previa conjugada para el modelo Normal el caso de μ con σ^2 desconocido

Una vez seleccionada la distribución previa, podemos ingresar un conjunto de datos o ingresar información de un conjunto de datos para calcular los intervalos de

confianza y de credibilidad para el parámetro μ del modelo Normal en el caso de que σ^2 sea desconocido. Tenemos dos formas de ingresar los datos en el panel: “Data”



Data

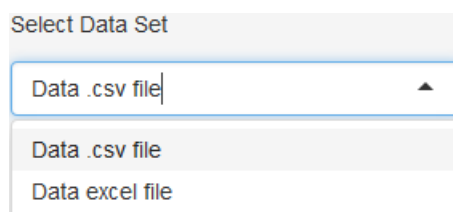
☒ Upload your own

☐ Enter summary statistics

Figura 4–46: Formas de ingresar datos en el panel para el modelo Normal el caso de μ con σ^2 desconocido

☒ Upload your own

En la parte inferior del panel se despliega “Select Data Set”. En esta parte es posible importar datos provenientes de un modelo Normal desde un archivo (.csv) o un archivo (.xlsx). La forma de importar los datos debe ser en un archivo con los datos organizados en una columna.



Select Data Set

Data .csv file

Data .csv file

Data excel file

Figura 4–47: Formato para ingresar datos del modelo Normal el caso de μ con σ^2 desconocido

☐ Enter summary statistics

Como alternativa al importar los datos, es posible también ingresar las medidas de resumen de los datos para calcular los intervalos de confianza y de credibilidad, que para el parámetro μ , modelo Normal con σ desconocidos, son:



Sample Size	\bar{x}	s^2
100	10	2

Figura 4–48: Medidas de resumen suficientes del modelo Normal el caso de μ con σ^2 desconocido

Sample Size: Es el número de observaciones (n) provenientes de un modelo Normal.

\bar{x} : promedio de n observaciones provenientes de un modelo Normal.

s^2 : varianza de n observaciones provenientes de un modelo Normal.

Con el botón **CALCULATE**, se calculan los intervalos de confianza y de credibilidad. Aparecerán en la parte derecha de la pantalla los intervalos, el cálculo de la distancia de dichos intervalos (utilizando la métrica del taxista) y una gráfica de los mismos.

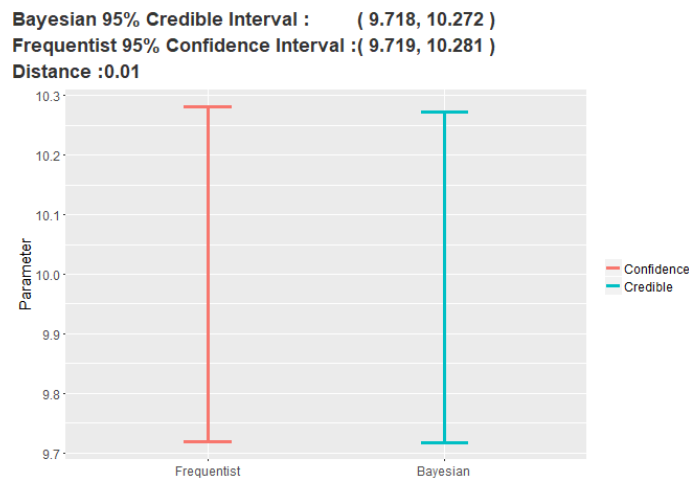


Figura 4–49: Intervalos y gráfica de intervalos del modelo Normal el caso de μ con σ^2 desconocido

4.2. Find the sample size required.

En esta parte, se puede seleccionar uno de los cinco modelos (**Binomial**, **Poisson**, **Geometric**, **Exponential**, **Normal**), para el cual se quiere encontrar el tamaño de muestra n para que la distancia (calculada usando la métrica del taxista, Definición 2.2.8) sobre los intervalos de confianza y de credibilidad sea aproximadamente una distancia requerida. El cálculo del tamaño de muestra n se realizará vía simulación.

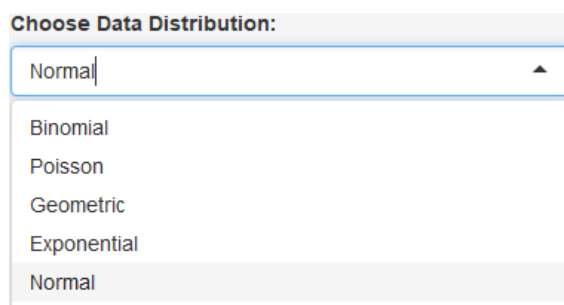


Figura 4–50: Menú de distribuciones muestrales para encontrar el tamaño de muestra necesario con el cual la estimación por intervalos frecuentista y bayesiano coincida dentro de un error predeterminado

✓ Binomial.

En esta parte, se puede seleccionar la distribución previa que se utiliza para simular los intervalos de credibilidad sobre el parámetro p en el modelo Binomial. Se presentan tres formas de seleccionar la distribución previa: “Choose Prior Distribution:”

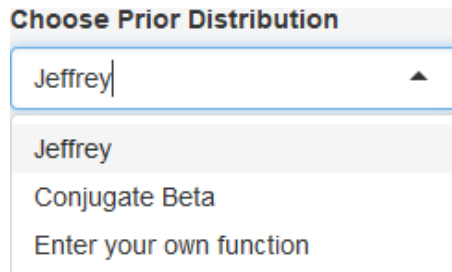


Figura 4–51: Selección de la distribución previa para el modelo Binomial

✓ Jeffrey

Aquí, se utiliza la distribución previa de Jeffrey para el parámetro p del modelo Binomial (Ver ejemplo 2.3)

✓ Conjugate Beta

Aquí, se utiliza la distribución previa Conjugate Beta para el parámetro p del modelo Binomial: “ $Beta(\alpha, \beta)$ (ver Ejemplo 2.4)”, donde α y β pueden ser ingresados por el usuario:

α	β
0.5	0.5

Figura 4–52: Parámetros de la previa Beta conjugada para el modelo Binomial

✓ Enter your own function

Aquí, el usuario puede ingresar su propia distribución previa para el parámetro p del modelo Binomial: “Enter your own function:”

Enter R code for prior

```
dunif(x,0,1)
```

Figura 4–53: Ingrese su propia distribución previa para el modelo Binomial

Una vez seleccionada la distribución previa, se puede ingresar el valor para el parámetro p del modelo Binomial y el valor de **distancia requerida** para la cual se quiere encontrar el tamaño de muestra.

Parameter (p)	Required Distance
0.5	0.1

Figura 4–54: Parámetro de la distribución Binomial y distancia requerida para aproximar los intervalos

Se establece el rango de valores para el tamaño de muestra n , donde el programa buscará el tamaño de muestra adecuado, tal que la distancia entre los intervalos sea menor o igual a la distancia requerida.

Search region for n:

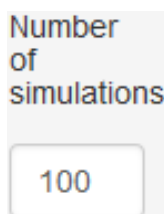
min	max
1	100

Figura 4–55: Región de búsqueda para el tamaño de muestra del modelo Binomial

min: valor mínimo del rango de búsqueda para n

max: valor máximo del rango de búsqueda para n

Se determina el número de simulaciones, el cual es el número de veces que se generan muestras binomiales, para estimar la distancia entre los intervalos en cada iteración del n en el método de bisección.

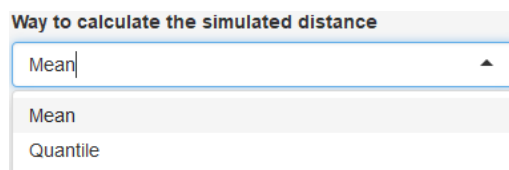


Number of simulations

100

Figura 4-56: Número de simulaciones en el modelo Binomial

A continuación, se determina la forma de estimar la distancia: “Way to estimate the simulated distance”



Way to calculate the simulated distance

Mean

Mean

Quantile

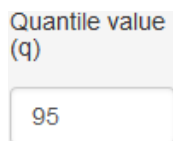
Figura 4-57: Forma de estimar la distancia simulada en el modelo Binomial

✓ Mean

Se estima la distancia calculando la media de las distancias simuladas.

✓ Quantile

Se estima la distancia calculando el cuantil q de las distancias simuladas, donde el valor cuantil puede ser ingresado por el usuario: “Quantile value (q)”



Quantile value (q)

95

Figura 4-58: Ingresar el valor del cuantil para estimar la distancia en el modelo Binomial

Con el botón **CALCULATE**, se calcula el tamaño de muestra n con el cual la estimación de la distancia (media o cuantil de las distancias simuladas) entre los intervalos de confianza y de credibilidad para el parámetro p del modelo Binomial sea menor o igual a la distancia requerida (ingresada por el usuario) y se muestran las medias de los intervalos simulados para este tamaño de muestra calculado.

Expectation bayesian 95% Credible Interval: (0.1573, 0.788)
Expectation Frequentist 95% Confidence Interval:(0.1127, 0.8408)
Estimated distance:0.0976
Sample size required:6

Figura 4–59: Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Binomial

✓ Poisson.

En esta parte, se puede seleccionar la distribución previa que se utiliza para simular los intervalos de credibilidad sobre el parámetro λ en el modelo Poisson. Se presentan tres formas de seleccionar la distribución previa: “Choose Prior Distribution:”

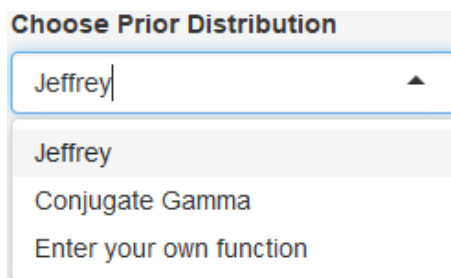


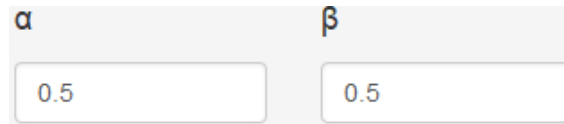
Figura 4–60: Selección de la distribución previa para el modelo Poisson

✓ Jeffrey

Aquí, se utiliza la distribución previa de Jeffrey para el parámetro λ del modelo Poisson (ver Ejemplo 2.1)

✓ Conjugate Beta

Aquí, se utiliza la distribución previa Conjugate Beta para el parámetro λ del modelo Poisson: “ $Gamma(\alpha, \beta)$ (ver Ejemplo 2.1)”, donde α y β pueden ser ingresados por el usuario:



α	β
0.5	0.5

Figura 4–61: Parámetros de la previa Gamma conjugada para el modelo Poisson

✓ Enter your own function

Aquí, el usuario puede ingresar su propia distribución previa para el parámetro λ del modelo Poisson: “Enter your own function:”

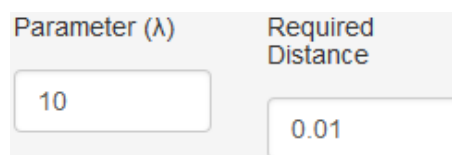


Enter R code for prior

`dgamma(x,1,2)`

Figura 4–62: Ingrese su propia distribución previa para el modelo Poisson

Una vez seleccionada la distribución previa, se puede ingresar el valor para el parámetro λ del modelo Poisson y el valor de **distancia requerida** para la cual se quiere encontrar el tamaño de muestra.



Parameter (λ)	Required Distance
10	0.01

Figura 4–63: Parámetro de la distribución Poisson y distancia requerida para aproximar los intervalos

Se establece el rango de valores para el tamaño de muestra n , donde el programa buscará el tamaño de muestra adecuado, tal que la distancia entre los intervalos sea menor o igual a la distancia requerida.

Search region for n:

min	max
1	1000

Figura 4-64: Región de búsqueda para el tamaño de muestra del modelo Poisson

min: valor mínimo del rango de búsqueda para n

max: valor máximo del rango de búsqueda para n

Se determina el número de simulaciones, el cual es el número de veces que se generan muestras Poisson, para estimar la distancia entre los intervalos en cada iteración del n en el método de bisección.

Number of simulations

100

Figura 4-65: Número de simulaciones en el modelo Poisson

A continuación se determina la forma de estimar la distancia: “Way to estimate the simulated distance”

Way to estimate the simulated distance

Mean

Mean

Quantile

Figura 4-66: Forma de estimar la distancia simulada en el modelo Poisson

✓ Mean

Se estima la distancia calculando la media de las distancias simuladas.

✓ Quantile

Se estima la distancia calculando el cuantil q de las distancias simuladas, donde el valor cuantil puede ser ingresado por el usuario: “Quantile value (q)”

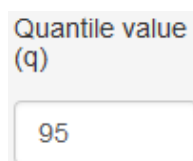


Figura 4-67: Ingresar el valor del cuantil para estimar la distancia en el modelo Poisson

Con el botón **CALCULATE**, se calcula el tamaño de muestra n con el cual la estimación de la distancia (media o cuantil de las distancias simuladas) entre los intervalos de confianza y de credibilidad para el parámetro λ del modelo Poisson, sea menor o igual a la distancia requerida (ingresada por el usuario) y se muestran las medias de los intervalos simulados para este tamaño de muestra calculado.

Expectation bayesian 95% Credible Interval: (9.8081, 10.1998)
Expectation Frequentist 95% Confidence Interval:(9.8125, 10.2054)
Estimated distance:0.0102
Sample size required:1001

Figura 4-68: Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Poisson

✓ Geometric.

En esta parte, se puede seleccionar la distribución previa que se utiliza para simular los intervalos de credibilidad sobre el parámetro p en el modelo Geometric. Se presentan tres formas de seleccionar la distribución previa: “Choose Prior Distribution:”

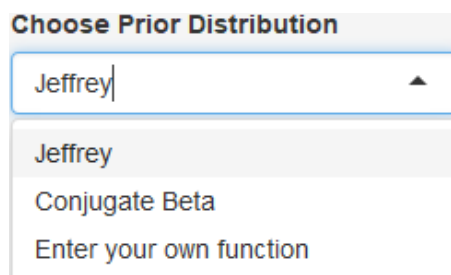


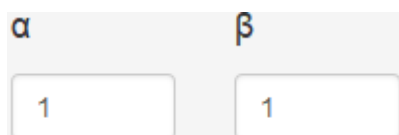
Figura 4-69: Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Geométrico

✓ Jeffrey

Aquí, se utiliza la distribución previa de Jeffrey para el parámetro p del modelo Poisson (ver Ejemplo 2.4)

✓ Conjugate Beta

Aquí, se utiliza la distribución previa Conjugate Beta para el parámetro p del modelo Geometric: “ $Beta(\alpha, \beta)$ (ver Ejemplo 2.11)”, donde α y β pueden ser ingresados por el usuario:

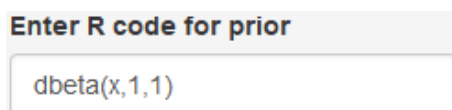


α	β
1	1

Figura 4–70: Parámetros de la previa Beta conjugada para el modelo Geométrico

✓ Enter your own function

Aquí, el usuario puede ingresar su propia distribución previa para el parámetro p del modelo Geometric: “Enter your own function:”

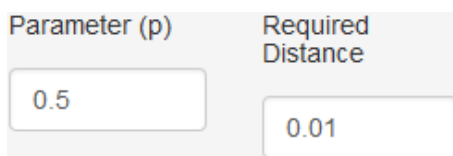


Enter R code for prior

`dbeta(x,1,1)`

Figura 4–71: Ingrese su propia distribución previa para el modelo Geométrico

Una vez seleccionada la distribución previa, se puede ingresar el valor para el parámetro p del modelo Geometric y el valor de **distancia requerida** para la cual se quiere encontrar el tamaño de muestra.

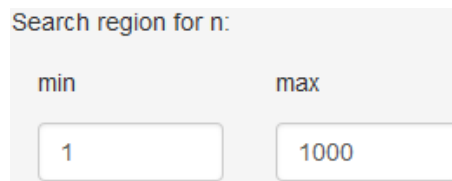


Parameter (p)	Required Distance
0.5	0.01

Figura 4–72: Parámetro de la distribución Geométrica y distancia requerida para aproximar los intervalos

Se establece el rango de valores para el tamaño de muestra n , donde el programa buscará el tamaño de muestra adecuado, tal que la distancia entre los intervalos

sea menor o igual a la distancia requerida.



Search region for n:

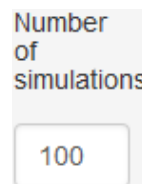
min	max
1	1000

Figura 4–73: Región de búsqueda para el tamaño de muestra del modelo Geométrico

min: valor mínimo del rango de búsqueda para n

max: valor máximo del rango de búsqueda para n

Se determina el número de simulaciones, el cual es el número de veces que se generan muestras Geométrica, para estimar la distancia entre los intervalos en cada iteración del n en el método de bisección.

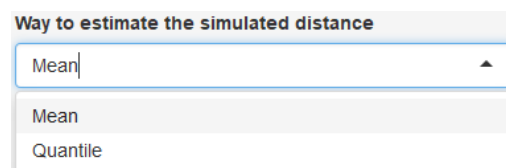


Number of simulations

100

Figura 4–74: Número de simulaciones en el modelo Geométrico

A continuación, se determina la forma de estimar la distancia: “Way to estimate the simulated distance”



Way to estimate the simulated distance

Mean

Mean

Quantile

Figura 4–75: Forma de estimar la distancia simulada en el modelo Geométrico

✓ Mean

Se estima la distancia calculando la media de las distancias simuladas.

✓ Quantile

Se estima la distancia calculando el cuantil q de las distancias simuladas. Donde el

valor cuantil puede ser ingresado por el usuario: “Quantile value (q)”

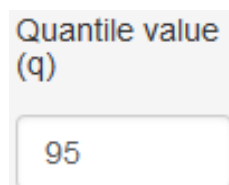


Figura 4-76: Ingresar el valor del cuantil para estimar la distancia en el modelo Geométrico

Con el botón **CALCULATE**, se calcula el tamaño de muestra n con el cual la estimación de la distancia (media o cuantil de las distancias simuladas) entre los intervalos de confianza y de credibilidad para el parámetro p del modelo Geometric, sea menor o igual a la distancia requerida (ingresada por el usuario) y se muestran las medias de los intervalos simulados para este tamaño de muestra calculado.

Expectation bayesian 95% Credible Interval: (9.8081, 10.1998)
Expectation Frequentist 95% Confidence Interval:(9.8125, 10.2054)
Estimated distance:0.0102
Sample size required:1001

Figura 4-77: Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Geométrico

✓ Exponential.

En esta parte, se puede seleccionar la distribución previa que se utiliza para simular los intervalos de credibilidad sobre el parámetro λ en el modelo Geometric. Se presentan dos formas de seleccionar la distribución previa: “Choose Prior Distribution:”

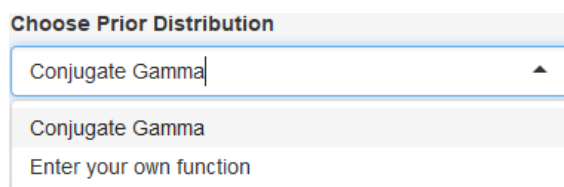
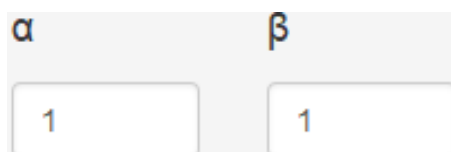


Figura 4-78: Selección de la distribución previa para el modelo Exponencial

✓ Conjugate Gamma

Aquí, se utiliza la distribución previa Conjugate Gamma para el parámetro λ del modelo Exponential: “ $\text{Gamma}(\alpha, \beta)$ (ver Ejemplo 2.10)”, donde α y β pueden ser ingresados por el usuario:




The image shows a form with two input fields. The first field is labeled with the Greek letter alpha (α) and contains the value 1. The second field is labeled with the Greek letter beta (β) and also contains the value 1.

Figura 4–79: Parámetros de la previa Gamma conjugada para el modelo Exponencial

✓ Enter your own function

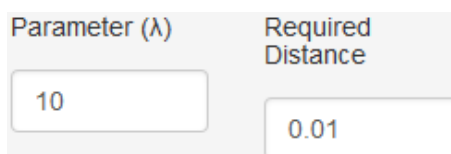
Aquí, el usuario puede ingresar su propia distribución previa para el parámetro λ del modelo Exponential: “Enter your own function:”



The image shows a form with a single input field. Above the field is the label “Enter R code for prior”. Inside the field, the R code `dgamma(x,0.5,0.5)` is entered.

Figura 4–80: Ingrese su propia distribución previa para el modelo Exponencial

Una vez seleccionada la distribución previa, se puede ingresar el valor para el parámetro λ del modelo Exponential y el valor de **distancia requerida** para la cual se quiere encontrar el tamaño de muestra.



The image shows a form with two input fields. The first field is labeled “Parameter (λ)” and contains the value 10. The second field is labeled “Required Distance” and contains the value 0.01.

Figura 4–81: Parámetro de la distribución Exponencial y distancia requerida para aproximar los intervalos

Se establece el rango de valores para el tamaño de muestra n , donde el programa buscará el tamaño de muestra adecuado, tal que la distancia entre los intervalos sea menor o igual a la distancia requerida.

Search region for n :

min	max
1	1000

Figura 4–82: Región de búsqueda para el tamaño de muestra del modelo Exponencial

min: valor mínimo del rango de búsqueda para n

max: valor máximo del rango de búsqueda para n

Se determina el número de simulaciones, el cual es el número de veces que se generan muestras Exponenciales, para estimar la distancia entre los intervalos en cada iteración del n en el método de bisección.

Number of simulations

100

Figura 4–83: Número de simulaciones en el modelo Exponencial

A continuación, se determina la forma de estimar la distancia: “Way to estimate the simulated distance”

Way to estimate the simulated distance

Mean

Mean

Quantile

Figura 4–84: Forma de estimar la distancia simulada en el modelo Exponencial

✓ Mean

Se estima la distancia calculando la media de las distancias simuladas.

✓ Quantile

Se estima la distancia calculando el cuantil q de las distancias simuladas, donde el valor cuantil puede ser ingresado por el usuario: “Quantile value (q)”

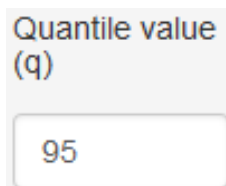


Figura 4–85: Ingresar el valor del cuantil para estimar la distancia en el modelo Exponencial

Con el botón **CALCULATE**, se calcula el tamaño de muestra n con el cual la estimación de la distancia (media o cuantil de las distancias simuladas) entre los intervalos de confianza y de credibilidad para el parámetro λ del modelo Exponential, sea menor o igual a la distancia requerida (ingresada por el usuario) y se muestran las medias de los intervalos simulados para este tamaño de muestra calculado.

Expectation bayesian 95% Credible Interval: (0.6872, 1.4821)
Expectation Frequentist 95% Confidence Interval:(0.6851, 1.4887)
Estimated distance:0.0094
Sample size required:26

Figura 4–86: Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Exponencial

✓ Normal.

En esta parte, tenemos tres casos para el modelo Normal: “Select the parameter”

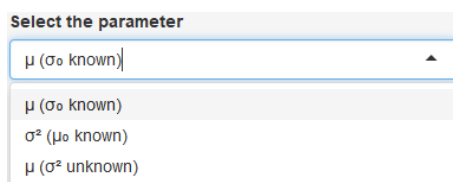


Figura 4–87: Inferencia en el modelo Normal

✓ $\mu(\sigma_0 \text{ known})$

En esta parte, se presentan dos formas de seleccionar la distribución previa de μ :
 “Choose Prior Distribution:”

Figura 4–88: Selección de la distribución previa para el modelo Normal en el caso de μ con σ_0 conocido

✓ Conjugate Normal

Aquí, se utiliza la distribución previa Conjugate Normal para el parámetro μ del modelo Normal con σ conocido: “ $Norm(\mu_1, \sigma_1)$ ”(ver Ejemplo 2.12)”, donde μ_0 y σ_0 pueden ser ingresados por el usuario:

Figura 4–89: Parámetros de la previa Normal conjugada para el modelo Normal el caso de μ con σ_0 conocido

✓ Enter your own function

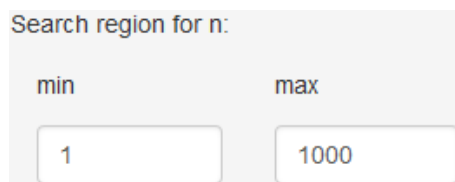
Aquí, el usuario puede ingresar su propia distribución previa para el parámetro μ del modelo Normal: “Enter your own function:”

Figura 4–90: Ingrese su propia distribución previa para el modelo Normal el caso de μ con σ_0 conocido

Una vez seleccionada la distribución previa, se puede ingresar el valor para el parámetro μ , $\sigma_0(Know)$ del modelo Normal y el valor de la **distancia requerida** para la cual se quiere encontrar el tamaño de muestra.

Figura 4–91: Parámetros de la distribución Normal y distancia requerida para aproximar los intervalos en el caso de μ con σ_0 conocido

Se establece el rango de valores para el tamaño de muestra n , donde el programa buscará el tamaño de muestra adecuado, tal que la distancia entre los intervalos sea menor o igual a la distancia requerida.



Search region for n:

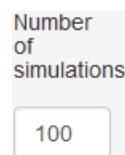
min	max
1	1000

Figura 4–92: Región de búsqueda para el tamaño de muestra del modelo Normal en el caso de μ con σ_0 conocido

min: valor mínimo del rango de búsqueda para n

max: valor máximo del rango de búsqueda para n

Se determina el número de simulaciones, el cual es el número de veces que se generan muestras Normales, para estimar la distancia entre los intervalos en cada iteración del n en el método de bisección.

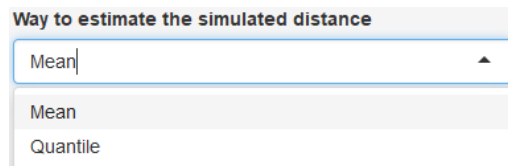


Number of simulations

100

Figura 4–93: Número de simulaciones en el modelo Normal en el caso de μ con σ_0 conocido

A continuación, se determina la forma de estimar la distancia: “Way to estimate the simulated distance”



Way to estimate the simulated distance

Mean

Mean

Quantile

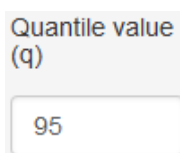
Figura 4–94: Forma de estimar la distancia simulada en el modelo Normal en el caso de μ con σ_0 conocido

✓ Mean

Se estima la distancia calculando la media de las distancias simuladas.

✓ Quantile

Se estima la distancia calculando el cuantil q de las distancias simuladas, donde el valor cuantil puede ser ingresado por el usuario: “Quantile value (q)”



Quantile value (q)

95

Figura 4–95: Ingresar el valor del cuantil para estimar la distancia en el modelo Normal en el caso de μ con σ_0 conocido

Con el botón **CALCULATE**, se calcula el tamaño de muestra n con el cual la estimación de la distancia (media o cuantil de las distancias simuladas) entre los intervalos de confianza y de credibilidad para el parámetro μ del modelo Normal, sea menor o igual a la distancia requerida (ingresada por el usuario) y se muestran las medias de los intervalos simulados para este tamaño de muestra calculado.

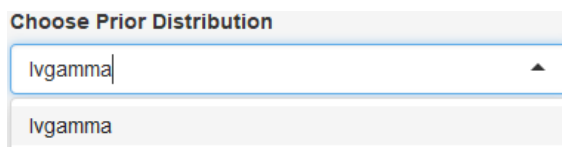
Expectation bayesian 95% Credible Interval: (-1.9261, 1.9177)
Expectation Frequentist 95% Confidence Interval:(-1.9644, 1.9556)
Estimated distance:0.0891
Sample size required:1

Figura 4–96: Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Normal en el caso de μ con σ_0 conocido

✓ $\sigma^2(\mu_0 \text{ known})$

En esta parte, se presenta una forma de seleccionar la distribución previa de σ^2 :

“Choose Prior Distribution:”

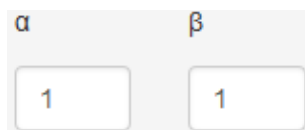


The image shows a web interface with a dropdown menu titled "Choose Prior Distribution". The dropdown is open, showing a list with the option "Ivgamma" selected. The text "Ivgamma" is also visible in the input field above the dropdown.

Figura 4–97: Selección de la distribución previa para el modelo Normal en el caso de σ^2 con μ_0 conocido

✓ Conjugate Invgamma

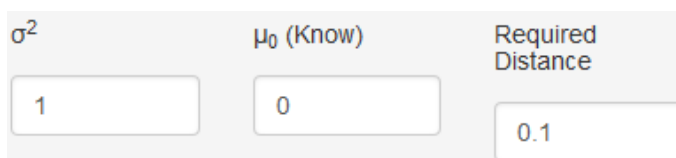
Aquí, se utiliza la distribución previa conjugate Ivgamma para el parámetro σ^2 del modelo Normal con μ conocido: “ $Ivgamma(\alpha, \beta)$ ”(ver Ejemplo 2.13)”, donde α y β pueden ser ingresados por el usuario:



The image shows two input fields for the parameters of the conjugate Invgamma distribution. The first field is labeled with the Greek letter alpha (α) and contains the value "1". The second field is labeled with the Greek letter beta (β) and also contains the value "1".

Figura 4–98: Parámetros de la previa Invgamma conjugada para el modelo Normal el caso de σ^2 con μ_0 conocido

Una vez seleccionada la distribución previa, se puede ingresar el valor para el parámetro σ^2 , $\mu_0(Know)$ del modelo Normal y el valor de la **distancia requerida** para la cual se quiere encontrar el tamaño de muestra.



The image shows three input fields arranged horizontally. The first field is labeled σ^2 and contains the value "1". The second field is labeled $\mu_0 (Know)$ and contains the value "0". The third field is labeled "Required Distance" and contains the value "0.1".

Figura 4–99: Parmetros de la distribucin Invgamma y distancia requerida para aproximar los intervalos en el caso de σ^2 con μ_0 conocido

Se establece el rango de valores para el tamaño de muestra n , donde el programa buscará el tamaño de muestra adecuado (utilizando el método de bisección), tal

que la distancia entre los intervalos sea menor o igual a la distancia requerida.

Search region for n:

min	max
1	1000

Figura 4–100: Región de búsqueda para el tamaño de muestra del modelo Normal en el caso de σ^2 con μ_0 conocido

min: valor mínimo del rango de búsqueda para n

max: valor máximo del rango de búsqueda para n

Se determina el número de simulaciones, el cual es el número de veces que se generan muestras Normales, para estimar la distancia entre los intervalos en cada iteración del n en el método de bisección.

Number of simulations

100

Figura 4–101: Número de simulaciones en el modelo Normal en el caso de σ^2 con μ_0 conocido

A continuación se determina la forma de estimar la distancia: “Way to estimate the simulated distance”

Way to estimate the simulated distance

Mean

Mean

Quantile

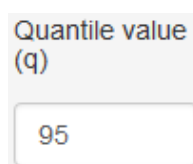
Figura 4–102: Forma de estimar la distancia simulada en el modelo Normal en el caso de σ^2 con μ_0 conocido

✓ Mean

Se estima la distancia calculando la media de las distancias simuladas.

✓ Quantile

Se estima la distancia calculando el cuantil q de las distancias simuladas, donde el valor cuantil puede ser ingresado por el usuario: “Quantile value (q)”



Quantile value (q)

95

Figura 4–103: Ingresar el valor del cuantil para estimar la distancia en el modelo Normal en el caso de σ^2 con μ_0 conocido

Con el botón **CALCULATE**, se calcula el tamaño de muestra n con el cual la estimación de la distancia (media o cuantil de las distancias simuladas) entre los intervalos de confianza y de credibilidad para el parámetro σ^2 del modelo Normal, sea menor o igual a la distancia requerida (ingresada por el usuario) y se muestran las medias de los intervalos simulados para este tamaño de muestra calculado.

Expectation bayesian 95% Credible Interval: (4.5381, 5.512)
Expectation Frequentist 95% Confidence Interval:(4.5423, 5.5178)
Estimated distance:0.01
Sample size required:813

Figura 4–104: Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Normal en el caso de σ^2 con μ_0 conocido

✓ $\mu(\sigma^2 \text{ Unknown})$

En esta parte, se presentan dos formas de seleccionar la distribución previa de μ :

“Choose Prior Distribution:”

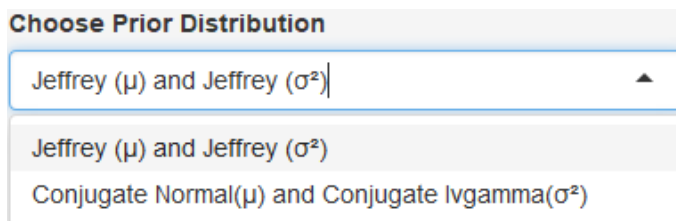


Figura 4–105: Selección de la distribución previa para el modelo Normal en el caso de μ con σ^2 desconocido

✓ Jeffrey (μ) and Jeffrey (σ^2)

Aquí, se utilizan las distribuciones previas de Jeffrey para los parámetros μ y σ^2 (donde μ y σ^2 son desconocidos y independientes) del modelo Normal (ver Ejemplo 2.7)

✓ Conjugate Normal (μ) and Conjugate Ivgamma (σ^2)

Aquí, se utilizan las distribuciones previas Conjugadas para los parámetros μ y σ^2 (donde μ y σ^2 son desconocidos y independientes) del modelo Normal: “ $Norm(\mu_1, \sigma_1)$ y $Ivgamma(\alpha, \beta)$ (ver Ejemplo 2.14)”, donde μ_1 , σ_1 , α y β pueden ser ingresados por el usuario.

μ_1 (Normal)	σ_1 (Normal)	α (Ivgamma)	β (Ivgamma)
0	1	0.5	0.5

Figura 4–106: Parámetros de la previa conjugada para el modelo Normal el caso de μ con σ^2 desconocido

Una vez seleccionada la distribución previa, se puede ingresar el valor para el parámetro μ , $\sigma^2(Unknown)$ del modelo Normal y el valor de la **distancia requerida** para la cual se quiere encontrar el tamaño de muestra.

μ	σ^2	Required Distance
<input type="text" value="0"/>	<input type="text" value="1"/>	<input type="text" value="0.1"/>

Figura 4–107: Parmetros de la distribucin Normal y distancia requerida para aproximar los intervalos en el caso de μ con σ^2 desconocido

Se establece el rango de valores para el tamaño de muestra n , donde el programa buscará el tamaño de muestra adecuado (utilizando el método de bisección), tal que la distancia entre los intervalos sean menor o igual a la distancia requerida.

Search region for n:	
min	max
<input type="text" value="1"/>	<input type="text" value="1000"/>

Figura 4–108: Región de búsqueda para el tamaño de muestra del modelo Normal en el caso de μ con σ^2 desconocido

min: valor mínimo del rango de búsqueda para n

max: valor máximo del rango de búsqueda para n

Se determina el número de simulaciones, el cual es el número de veces que se generan muestras Normales, para estimar la distancia entre los intervalos en cada iteración del n en el método de bisección.

Number of simulations
<input type="text" value="100"/>

Figura 4–109: Número de simulaciones en el modelo Normal en el caso de μ con σ^2 desconocido

A continuación, se determina la forma de estimar la distancia: “Way to estimate the simulated distance”

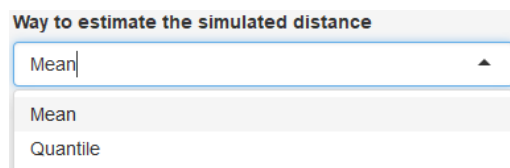


Figura 4–110: Forma de estimar la distancia simulada en el modelo Normal en el caso de μ con σ desconocido

✓ Mean

Se estima la distancia calculando la media de las distancias simuladas.

✓ Quantile

Se estima la distancia calculando el cuantil q de las distancias simuladas, donde el valor cuantil puede ser ingresado por el usuario: “Quantile value (q)”

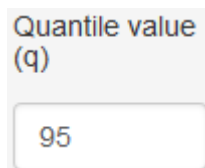


Figura 4–111: Ingresar el valor del cuantil para estimar la distancia en el modelo Normal en el caso de μ con σ^2 desconocido

Con el botón **CALCULATE**, se calcula el tamaño de muestra n con el cual la estimación de la distancia (media o cuantil de las distancias simuladas) entre los intervalos de confianza y de credibilidad para el parámetro μ del modelo Normal, sea menor o igual a la distancia requerida (ingresada por el usuario) y se muestran las medias de los intervalos simulados para este tamaño de muestra calculado.

Expectation bayesian 95% Credible Interval: (-0.5396, 0.5812)
Expectation Frequentist 95% Confidence Interval:(-0.5787, 0.6169)
Estimated distance:0.0909
Sample size required:13

Figura 4–112: Tamaño de muestra encontrado para aproximar los intervalos de confianza y de credibilidad en el modelo Normal en el caso de μ con σ^2 desconocido

Capítulo 5

CONCLUSIONES Y TRABAJOS FUTUROS

En los casos estudiados se confirmó que, cuando el tamaño de muestra es grande, la distancia calculada para los intervalos de confianza y credibilidad convergen a cero. Es decir, que los dos enfoques de la estadística coinciden en la estimación por intervalos. Ahora, en este trabajo se realizó un estudio detallado en el cual se encontraron valores numéricos del tamaño de muestra que se necesitan para asegurar que los intervalos de confianza y de credibilidad sean iguales, excepto por una distancia requerida (error predeterminado) que un analista o experimentador puede seleccionar.

Los tamaños de muestra necesarios para que la distancia entre los intervalos de confianza y los intervalos de credibilidad sea menor o igual a una distancia requerida, es mucho más grandes cuando se utiliza la previa Conjugada que cuando se utiliza la previa de Jeffrey.

En la mayoría de los casos estudiados con la previa de Jeffrey, los tamaños de muestra n_{50} y n_{99} son iguales o casi iguales, lo que quiere decir que los límites inferiores y los límites superiores para los intervalos de confianza y de credibilidad tiene comportamientos muy similares o la distancia entre estos, es constantes para cualquier observación obtenida.

Para trabajos futuros sobre este tema, se podrían explorar más distribuciones previas de forma subjetiva y otros intervalos de confianza para los diferentes modelos que hemos estudiado. También, realizar la misma metodología para el coeficiente de

correlación, diferencias de proporciones, diferencias de medias y cociente de varianzas, que por tiempo no fueron posibles estudiar.

Por otro lado, se podría realizar este mismo análisis estudiando otras distancias, como por ejemplo la distancia de Mahalanobis.

Bibliografía

- Agresti, A. (2002). *Categorical Data Analysis*, 2nd edn. New York: Wiley
- Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. New York: Wiley.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. (2nd edition). Springer-Verlag, New York.
- Casella, G. and Berger, R (1990) *Statistical Inference*. Wadsworth, Belmont, CA.
- Cepeda, E., Aguilar, W., Cervantes, V., Corrales, M., Diaz, I. y Rodríguez, D.: *Intervalos de confianza e intervalos de credibilidad para una proporción*, Revista Colombiana de Estadística, Diciembre 2008, volumen 31, no. 2, pp. 211 a 228.
- Clopper C. J., Pearson E. S. (1934), *The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial*, Biometrika 26, 404 – 413.
- Garwood, F., 1936: *Fiducial limits for the Poisson distribution*. Biometrika 28, 437 – 442.
- S. Jackman. Bayesian Analysis for the Social Sciences. John Wiley & Sons, Ltd, 2009.
- Jeffreys, H. (1961) *Theory of Probability* (3rd edition). Oxford University Press, Oxford.
- Lawrence, M., Kishor, S.: *A Comparison of Approximate Interval Estimators for the Bernoulli Parameter*, The American Statistician, February 2012, 50 : 1, pp. 63 – 68.
- Pacheco, L., Tesis de Maestría: *Comparación de Intervalos de Confianza para el Coeficiente de Correlación*, Facultad de Ciencias, Escuela de Estadística, Universidad Nacional de Colombia, Medellín, 2013.

Robert, C. and Casella, G. (1998). Monte Carlo Statistical Methods. New York: Springer-Verlag.

Robert, C. y Casella, G.: *Monte Carlo Statistical Methods* Springer, 2da edition, 2005.

APENDICES

Apéndice A

PROPOSICIONES Y CÓDIGOS EN R

Teorema 3. *(Inferencia en μ con σ_0 conocido) Los intervalos de confianza son equivalentes con los intervalos de credibilidad. Si la distribución previa es la distribución de Jeffrey (ver Ejemplo 2.5).*

Demostración. Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria iid de una distribución Normal, en el caso de σ_0 conocido, entonces la distribución a posteriori es:

$$\begin{aligned}\pi(\mu|x_1, x_2, \dots, x_n) &\propto \exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2\right) \pi(\mu) \\ &\propto \exp\left(-\frac{1}{2} \frac{(\mu - \bar{x})^2}{\sigma_0^2/n}\right)\end{aligned}\tag{A.1}$$

por lo tanto, $\mu|x_1, \dots, x_n \sim \text{norm}(\bar{x}, \frac{\sigma_0^2}{n})$ y el intervalo bayesiano es de la forma:
 $\left(\bar{x} - z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}}\right)$ □

Teorema 4. *(inferencia en λ para un modelo exponencial) Los intervalos de confianza son equivalentes con los intervalos de credibilidad. Si la distribución previa es la distribución de Jeffrey (ver Ejemplo 2.2).*

Demostración. Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria iid de una distribución Exponencial, entonces la distribución a posteriori es:

$$\pi(\mu|x_1, x_2, \dots, x_n) \propto \lambda^{n-1} \exp(-\lambda \sum x_i) \quad (\text{A.2})$$

Así,

$$\lambda|x_1, \dots, x_n \sim \text{gamma}(n, \sum x_i) \quad (\text{A.3})$$

$$\bar{x}\lambda|x_1, \dots, x_n \sim \text{gamma}(n, n) \quad (\text{A.4})$$

Así, tenemos el intervalo bayesiano para λ :

$$(q\text{gamma}(\alpha/2, n, n)/\bar{x}, q\text{gamma}(1 - \alpha/2, n, n)/\bar{x})$$

□

Teorema 5. (*Inferencia en σ^2 con μ_0 conocido*) *Los intervalos de confianza son equivalentes con los intervalos de credibilidad. Si la distribución previa es la distribución de Jeffrey (ver Ejemplo 2.6).*

Demostración. Sea $\mathbf{x} = (x_1, \dots, x_n)$ una muestra aleatoria iid de una distribución Normal, en el caso de μ_0 conocido, entonces la distribución a posteriori es:

$$\pi(\sigma^2|\mathbf{x}) \propto (\sigma^2)^{-\frac{n}{2}-1} \exp\left(-\frac{ns^2}{2\sigma^2}\right) \quad (\text{A.5})$$

luego $\sigma^2|\mathbf{x} \sim \text{Invgamma}\left(\frac{n}{2}, \frac{ns^2}{2}\right)$, donde $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$.

De forma equivalente la distribución a posteriori se puede escribir $\sigma^2|\mathbf{x} \sim \chi^{-2}(n, ns^2)$, es decir una distribución $\chi^2 - \text{Inversa}$ con n grados de libertad y parámetro de escala ns^2 . Luego se tiene $\frac{ns^2}{\sigma^2}|\mathbf{x} \sim \chi_n^2$, es la misma cantidad pivotal, con la cual se construye el intervalo de confianza. □

metodología del muestreador de Gibbs, en el caso donde μ y σ^2 son desconocidos en el modelo Normal.

El proceso en general es el siguiente:

1. Obtener una expresión para la distribución posterior conjunta $\pi(\mu, \sigma^2)$. Igualmente obtener las expresiones para las distribuciones marginales posteriores.
2. Determinar un valor inicial para los parámetros μ, σ^2
3. Generar μ de la distribución condicional para μ teniendo en cuenta los valores iniciales.
4. Reemplazar μ obtenido del paso anterior en la expresión para la distribución condicional de σ^2 y generar un valor para σ^2 .
5. Finalmente se obtiene una muestra con valores de cada uno de los parámetros de la distribución a posteriori conjunta.

Esto se repite las veces que se consideren necesarias para obtener una muestra de tamaño adecuado. Se realiza un quemado de muestras iniciales y finalmente se calculan los cuantiles de las muestras obtenidas para μ . Para hacer uso del Muestreador de Gibbs en el estudio de simulación, es necesario obtener las distribuciones condicionales de cada parámetro.

Distribuciones condicionales utilizando la previa de Jeffrey (ver Ejemplo 2.7)

Distribución a posteriori para μ condicionada σ^2 :

$$\pi(\mu|\sigma^2) \propto (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right) \quad (\text{A.6})$$

Por lo tanto $\mu|\sigma^2 \sim N(\bar{x}, \frac{\sigma^2}{n})$.

Distribución a posteriori para σ^2 condicionada μ :

$$\pi(\sigma^2|\mu) \propto (\sigma^2)^{-\frac{n}{2}-1} \exp \left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right) \right) \quad (\text{A.7})$$

Por lo tanto $\sigma^2|\mu \sim \text{Invgamma} \left(\frac{n}{2}, \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2} \right)$.

Distribuciones condicionales utilizando la previa de conjugada (ver Ejemplo 2.14)

Distribución a posteriori para μ condicionada σ^2 :

$$\pi(\mu|\sigma^2) \propto \exp \left(-\frac{1}{2} \left(\frac{n(\bar{x} - \mu)^2}{\sigma^2} + \frac{(\mu - \mu_1)^2}{\sigma_1^2} \right) \right) \quad (\text{A.8})$$

Por lo tanto $\mu|\sigma^2 \sim N \left(\frac{n\sigma_1^2\bar{x} + \sigma^2\mu_1}{n\sigma_1^2 + \sigma^2}, \frac{\sigma^2\sigma_1^2}{n\sigma_1^2 + \sigma^2} \right)$.

Distribución a posteriori para σ^2 condicionada μ :

$$\pi(\sigma^2|\mu) \propto (\sigma^2)^{-\frac{n}{2}-\alpha-1} \exp \left(-\frac{1}{\sigma^2} \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2} + \frac{n(\bar{x} - \mu)^2}{2} + \beta \right) \right) \quad (\text{A.9})$$

Por lo tanto $\sigma^2|\mu \sim \text{Invgamma} \left(\frac{n}{2} + \alpha, \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2} + \frac{n(\bar{x} - \mu)^2}{2} + \beta \right)$.

Código de él intervalo de Clopper-pearson y intervalo de credibilidad con distribución previa Beta(3,3)

```
int.prop <- function(x, n, conf.level = 95) {
  alpha <- 1-conf.level/100
  Lim_B <- qbeta(c(alpha/2,1-alpha/2),3+x,3+n-x)
  ci <- binom.test(x, n, conf.level = 1-alpha)$conf.int
  round(c(Lim_B, ci),4)
}
IC <- int.prop(3,10)
```

```
names(IC) <- c("LB","HB","LF","HF")
```

```
IC
```

```
LB      HB      LF      HF
```

```
0.0927 0.6058 0.0667 0.6525
```

Código de las distancias D_{50} y D_{99} en el modelo Binomial para $n = 10$ y $p = 0.3$

```
calc.D <- function(n,p) {
  x <- rbinom(1000, n, p)
  Fx_L <- c();Fx_H <- c();Bx_L <- c();Bx_H <- c();D <- c()
  for (i in 1:length(x)) {
    Fx_L[i] <- int.prop(x[i],n)[3]
    Fx_H[i] <- int.prop(x[i],n)[4]
    Bx_L[i] <- int.prop(x[i],n)[1]
    Bx_H[i] <- int.prop(x[i],n)[2]
    D[i] <- abs(Fx_L[i]- Bx_L[i])+abs(Fx_H[i]-Bx_H[i])
  }
  T <- data.frame(x,D)
  D.estimado <- c(mean(D),quantile(D,0.99))
  print(qplot(D, geom="histogram",fill=I("gray"), col=I("red"),
    xlab="D's",binwidth = diff(range(D))/3.5)+
    geom_point(aes(x=D.estimado[1],y=0),col = c("blue"))+
    geom_point(aes(x=D.estimado[2],y=0),col = c("red"))+
    annotate("text",x=D.estimado[2],y=-16,label=expression(D[99])
    ,color="red")+annotate("text",x=D.estimado[1],y=-16,
    label=expression(D[50]),color="blue")
```

```
)  
print(head(T, n = 15,... = 2))  
print( c("D50","D99"))  
print( D.estimado )  
}
```