



Estimación de parámetros de desempeño de cuatro algoritmos para diagnóstico de dengue después de adicionar los resultados de una prueba rápida.

Claudia Lorena Montes - Juan David Angel

Universidad del Valle

Facultad de Ingeniería, Escuela de Estadística

Santiago de Cali, Colombia

2018.

Estimación de parámetros de desempeño de cuatro algoritmos para diagnóstico de dengue después de adicionar los resultados de una prueba rápida.

Claudia Lorena Montes - Juan David Angel

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico

Director:
Ph.D., José Rafael Tovar

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia

2018.

Dedicatoria

'...al Rey de los siglos, inmortal, invisible, al único y sabio Dios, sea honor y gloria por los siglos de los siglos'

Con cariño a mis familiares y amigos, especialmente a mi madre Ana Mora, gracias por apoyarme siempre incondicionalmente, por el amor y por ser mi fuerza para culminar este proceso.

Claudia Lorena Montes Mora

A mi familia, por brindarme siempre la motivación y el apoyo para terminar con éxito mi carrera profesional.

Juan David Angel Rubio

Agradecimientos

Agradecemos a Dios inicialmente, quien nos dio salud y vida, para seguir adelante con esta carrera, permitirnos continuar con nuestros sueños y alcanzar nuestras metas. Gracias a nuestras familias por el apoyo total para culminar este logro como profesionales.

Gracias al profesor José Rafael Tovar Cuevas, docente de la escuela de estadística de la Universidad del Valle, por aceptar dirigir nuestro trabajo de grado y brindarnos la enorme colaboración para la realización exitosa del mismo.

Nos sentimos agradecidos con la escuela de Estadística y la Universidad del Valle, por la oportunidad de ingresar en esta gran carrera.

Agradecemos a nuestros compañeros de carrera, especialmente a Julián Tatalcha y Julián Quintero, por apoyarnos en todo momento y animarnos, les deseamos lo mejor en la vida y esperamos seguir juntos, conservando la amistad que forjamos durante este proceso académico.

Contenido

. Agradecimientos	vi
Lista de tablas	xi
1. Introducción	2
2. Planteamiento del Problema	4
2.1. Problema Clínico	4
2.2. Problema estadístico	5
3. Justificación del Proyecto	6
3.1. Justificación Clínica	6
3.2. Justificación Estadística	6
3.3. Pregunta de Investigación	7
4. Objetivos	8
4.1. Objetivo General	8
4.2. Objetivos específicos	8
5. Antecedentes	9
5.1. Antecedentes Clínicos	9
5.2. Antecedentes Estadísticos	10
6. Marco Teórico	12
6.1. Marco Teórico Clínico	12
6.1.1. Dengue	12

6.1.2.	Diagnóstico médico	12
6.1.3.	Algoritmos de Clasificación	13
6.1.4.	Hemograma	14
6.1.5.	Prueba Patrón de Oro para el Dengue	14
6.1.6.	Prueba rápida para identificar individuos con Dengue.	15
6.1.7.	Prueba diagnóstica	16
6.1.8.	Parámetros de desempeño de una prueba diagnóstica	16
6.1.9.	Diferentes procesos de aplicación de una prueba diagnóstica	17
6.2.	Marco Teórico Estadístico	17
6.2.1.	Distribución de probabilidad multinomial	17
6.2.2.	Método Bootstrap no paramétrico	19
6.2.3.	Procedimientos para simulación de datos binarios correlacionados	20
6.2.4.	Inferencia Bayesiana	21
6.2.5.	Formula de Bayes Caso Discreto	22
6.2.6.	Formula de Bayes Caso Continuo	22
6.2.7.	Distribución a priori	23
6.2.8.	Distribución Beta conjugada	24
6.2.9.	Elicitación de hiperparametros para familias de distribuciones Beta	24
7.	Metodología	26
7.1.	Identificación de Variables para la simulación	26
7.2.	Archivos de datos	29
7.2.1.	Archivo 1	29
7.2.2.	Archivo 2	30
7.2.3.	Algoritmo implementado para obtener los archivos de datos	30
7.3.	Experimento Estadístico	32
7.4.	Modelo Estadístico	32
7.5.	Escenarios para la obtención del resultado clasificador	35
7.5.1.	Pruebas para validación aplicadas en Paralelo	35
7.5.2.	Pruebas de validación aplicadas en Serie.	36

7.6.	Estimación de Parámetros de desempeño	36
7.6.1.	Estimación Clásica	37
7.6.2.	Estimación de parámetros de desempeño para pruebas aplicadas en paralelo	38
7.6.3.	Estimación de parámetros de desempeño para pruebas aplicadas en serie . .	38
7.6.4.	Estimación Bayesiana	39
7.6.5.	Estimación Bayesiana para escenarios de aplicación en Paralelo y Series . .	40
7.6.6.	Remuestreo Bootstrap: Calculo de parámetros de desempeño	41
7.7.	Comparación de las estimaciones mediante la estadística Clásica y Bayesiana . . .	41
8.	Resultados.	43
8.1.	Análisis Descriptivo	43
8.1.1.	Resultados prueba rápida	44
8.2.	Resultados obtenidos cuando las pruebas se aplican en Paralelo	45
8.2.1.	Estimación por máxima verosimilitud	46
8.2.2.	Estimación Bayesiana	48
8.2.3.	Estimación usando métodos de remuestreo	49
8.3.	Resultados obtenidos cuando las pruebas se aplican en Serie Negativo	50
8.3.1.	Estimación por máxima verosimilitud	50
8.4.	Resultados obtenidos cuando las pruebas se aplican en Serie Positivo	52
8.4.1.	Estimación por máxima verosimilitud	52
8.4.2.	Estimación Bayesiana	53
8.4.3.	Estimación usando métodos de Remuestreo.	55
8.5.	Estimación de los parámetros de desempeño bajo diferentes niveles de sensibilidad de la prueba rápida.	56
8.6.	Comparación de bases generadas desde las categorías y el resultados de los algo- ritmos.	58
9.	Conclusiones y Recomendaciones.	60
9.1.	Conclusiones	60
9.2.	Recomendaciones	62

A. Anexo: Código R	63
A.1. Código para simulación de los datos binarios correlacionados.	63
A.2. Código para estimación de los parámetros de desempeño a partir de las bases simuladas	68
A.3. Código para técnica de remuestreo bootstrap	84

Lista de Tablas

6-1.	Interpretación de Resultados Prueba Rápida SD BIOLINE NS1/IgM/IgG	15
6-2.	Clasificación de los resultados de una prueba diagnóstica	16
7-1.	Definiciones de signos y síntomas	27
7-2.	Categorías de signos y síntomas conformadas por los clínicos especialistas en dengue.	28
7-3.	Insumos para la simulación del segundo archivo de datos - estimaciones de los algoritmos de clasificación (Valores tomados de [Mendez, 2017]).	29
7-4.	Estructura general del primer archivo de datos	30
7-5.	Estructura general del segundo archivo de datos	30
7-6.	Tabla de clasificación	33
7-7.	Elicitación de hiperparámetros para la prueba rápida (Valores tomados de Osorio et al.,2010)	40
8-1.	Proporción de pacientes con resultado positivo y negativo para dengue en algorit- mos clasificadores	43
8-2.	Estimaciones de los parámetros de desempeño para la Prueba Rápida.	45
8-3.	Estimación clásica - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en paralelo	47
8-4.	Estimación bayesiana - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en Paralelo.	48
8-5.	Estimación Bootstrap - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en paralelo.	49
8-6.	Estimación clásica - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en serie negativo.	51

8-7.	Estimación clásica - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en serie positivo.	52
8-8.	Estimación bayesiana - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en serie positivo.	54
8-9.	Estimación Bootstrap - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en serie positivo.	55
8-10.	Medidas de desempeño bajo diferentes sensibilidades de la prueba rápida escenario de aplicación paralelo.	56
8-11.	Medidas de desempeño bajo diferentes sensibilidades de la prueba rápida escenario de aplicación serie positivo.	57
8-12.	Proporciones y estimaciones de las bases simuladas	58
8-13.	Medidas de desempeño escenarios en Serie Negativo y Serie Positivo de las bases simuladas a partir de las categorías y los resultados de los algoritmos.	59

1. Introducción

El dengue es la enfermedad viral transmitida por mosquito de más rápida propagación en el mundo, afectando principalmente a países tropicales y subtropicales, en los últimos 50 años su incidencia ha aumentado 30 veces y se ha expandido a nuevos países. Se estima que alrededor de 3 mil millones de personas viven en países con dengue endémico.

En el presente trabajo de grado, se realizó la estimación de los parámetros de desempeño sensibilidad y especificidad para evaluar el rendimiento de cuatro algoritmos de clasificación en el diagnóstico de individuos con dengue mediante un conjunto de signos y síntomas adicionando una prueba rápida que permita el pronto diagnóstico de esta enfermedad, a partir de la simulación de bases de datos con la información previa de las sensibilidades y especificidades de los algoritmos desarrollados y la prueba rápida encontrada.

Se requirió analizar la contribución que realiza esta prueba rápida a los algoritmos previamente desarrollados, mediante técnicas de la estadística clásica y la flexibilidad de la estadística bayesiana. Los principales insumos para la realización de este trabajo de grado, son los signos y síntomas del dengue, incluido el hemograma que arroja los individuos que presentan un cuadro febril como criterio de inclusión, obteniendo a su vez los resultados de la prueba rápida que se le aplica para estimar los parámetros de interés.

Desde el año 2017, el Departamento de Planeación Nacional (DPN), desarrolla un Megaproyecto en las regiones de Santander, Valle del Cauca y Casanare, que permita la contribución y realización de un modelo sensible y altamente efectivo para el diagnóstico de dengue disminuyendo así la carga

social que esta enfermedad representa para el país, dicho proyecto consta de doce subproyectos, el cuarto busca generar y validar pruebas comerciales que permitan el pronto diagnóstico de esta enfermedad.

Mediante simulación de datos binarios correlacionados [Park et al., 1996] bajo los niveles del 10%, 50% y 90% de prevalencia de la enfermedad de dengue, se combinó la prueba rápida con los algoritmos de clasificación propuestos por [Mendez, 2017]. Se evaluaron diferentes escenarios que incluían cambios en la forma de aplicar las pruebas de evaluación a los individuos en serie (secuencialmente) y paralelo (simultáneamente) y definían el resultado final del procedimiento clasificatorio.

Se encontró que el uso en serie de la prueba rápida cuando el diagnóstico clínico fue negativo (Serie Negativo) aumentó las sensibilidades siendo éstas aproximadamente del 90% y especificidades del 50% y, el uso en serie cuando el diagnóstico clínico fue positivo (Serie positivo) aumentó las especificidades al 90% y se alcanzaron sensibilidades de 60%, indicando que la prueba rápida SD-BIOLINE Dengue Duo bajo la simulación podría ser útil para aumentar la sensibilidad y/o especificidad del diagnóstico clínico de acuerdo con su interpretación.

En este documento, el capítulo 1 contiene la introducción de este trabajo; el capítulo 2 contiene el planteamiento de la investigación; el capítulo 3 y 4 contienen la justificación y los objetivos de esta investigación; el capítulo 5 contiene la revisión de la literatura (antecedentes); el capítulo 6 contiene el marco de referencia clínico y estadístico, en el capítulo 7 se presenta la metodología propuesta indicando los procesos realizados en esta investigación, en el capítulo 8 los resultados obtenidos y finalmente, en el capítulo 9 las conclusiones y recomendaciones del trabajo de investigación.

2. Planteamiento del Problema

2.1. Problema Clínico

El dengue es una amenaza significativa para la salud pública, con estimaciones de 50 a 100 millones de casos por año y alrededor de 3 mil millones de personas en riesgo de infección [Kyle and Harris, 2008], reportandose epidemias en más de 100 países.

Para el 2010, se estimaba que anualmente ocurrían entre 284 y 528 millones de casos de dengue clásico en el mundo, y la región de las Américas ocupaba el tercer lugar en cuanto al número de casos de dengue clínicamente evidentes [Fernández et al., 2016].

En Colombia en el año 2016 se presentaron 103.822 casos probables de dengue, de los cuales 45.809 fueron confirmados por laboratorio y 199 muertes por dengue grave, para el año 2017 hasta el mes de Octubre, se reportaron 22.476 casos probables, de los cuales 6.104 fueron confirmados por laboratorio, esto es el 27.1% de los casos probables reportados, 227 casos de dengue grave y 44 muertes por esta causa, información obtenida desde la plataforma virtual de la Organización Panamericana de la Salud (OPS) , se identifica a su vez la circulación de 4 serotipos de este virus [Padilla et al., 2012].

En Santiago de Cali, la Secretaria Departamental de Salud del Valle indica que para el año 2016 reporta el 19% de los casos del país con 19.632 casos de dengue, de los cuales el 14.677 con dengue sin signos de alarma, 4.843 casos de dengue con signos de alarma y 112 casos de dengue grave.

El principal problema que se aborda es la clasificación errónea en el sistema de vigilancia del dengue, ya que el cuadro clínico de signos y síntomas de la enfermedad es muy similar al de otras

enfermedades tales como Chicungunya, Zika, Fiebre amarilla, entre otras. Además el diagnóstico confirmatorio de esta enfermedad son pruebas identificadas como Gold Standard “Patrones de Oro” tales como la reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR), la inmunoglobulina G (IgG), la inmunoglobulina M (IgM), las cuales resultan muy costosas y a su vez el tiempo para la obtención de los resultados podría causar un retraso en el diagnóstico de la enfermedad y en casos desfavorables la muerte del paciente.

Por lo tanto, se ha observado que existen pruebas rápidas comerciales que permiten el diagnóstico de esta enfermedad basados en la glucoproteína de este virus llamada NS1, de fácil acceso pero con desempeños cuestionables [Osorio et al., 2010].

2.2. Problema estadístico

Debido a la dificultad que existe en salud pública para el diagnóstico de esta enfermedad, surge la necesidad de emplear una prueba rápida (**SD-BIOLINE Dengue Duo**) basada en anticuerpos y antígenos, que permita el pronto diagnóstico del Dengue [Osorio et al., 2010].

En América el dengue se ha propagado con brotes cíclicos que ocurren cada 3 a 5 años [OMS, 2009].

Actualmente se espera una alta presencia de dengue puesto que ya ha pasado el tiempo para su regreso, sin embargo en la ciudad de Cali se han reportado menos casos de dengue de los esperados, por tanto el desafío principal en este tipo de estudios es contar con bases de datos reales para estimar los parámetros de desempeño del procedimiento que incluye cuatro algoritmos de clasificación previamente desarrollados adicionando una prueba rápida, sin embargo una opción propia y útil es hacer uso de simulaciones, además de tener en cuenta que los resultados y conclusiones desarrolladas se ajusten a la realidad y brindar herramientas útiles que permitan observar el desempeño de los algoritmos y la prueba rápida para cuando haga presencia la enfermedad en el país.

3. Justificación del Proyecto

3.1. Justificación Clínica

El diagnóstico del Dengue actualmente resulta difícil de realizar, ya que en países endémicos en los cuales se encuentra este virus presenta otras enfermedades que poseen signos y síntomas muy similares a las observadas en el cuadro clínico de la enfermedad del Dengue, debido a esto, se deben realizar pruebas y procedimientos de laboratorio para un correcto diagnóstico, los cuales son costosos y por tal razón se convierte en una carga económica para estos países [Osorio et al., 2010]. Este trabajo de grado está orientado en ayudar a disminuir los procesos y pruebas de laboratorio que se deben realizar para el diagnóstico clínico del Dengue, evaluando el desempeño de un procedimiento que incluye una prueba rápida para tener un adecuado manejo de pacientes con esta enfermedad y por consiguiente reducir la carga económica que este virus representa a los sistemas de salud.

3.2. Justificación Estadística

En la persistente búsqueda de una solución factible al problema clínico expuesto, se implementan alternativas estadísticas que permitan la obtención de herramientas metodológicas para la toma de decisiones en el ámbito clínico [Podgorelec et al., 2002] para la confirmación y diagnósticos de Dengue.

Así, el presente trabajo permitirá estimar los parámetros de desempeño de los algoritmos propuestos adicionando los resultados de una prueba rápida mediante el uso de la estadística como principal disciplina para la evaluación de éstos. Además de ofrecer una comparación entre los dos

enfoques de la estadística mostrando un panorama más amplio para la toma de decisiones referentes a la enfermedad del Dengue y ofrecer una mirada integral sobre la contribución que realiza la prueba rápida **SD-BIOLINE Dengue Duo** a los algoritmos.

3.3. Pregunta de Investigación

¿ Adicionando los resultados de una prueba rápida, se pueden obtener valores considerablemente altos en las estimaciones tanto para la sensibilidad como especificidad en cada algoritmo con y sin hemograma ?

4. Objetivos

4.1. Objetivo General

Estimar los parámetros de desempeño (sensibilidad y especificidad) adicionando los resultados de una prueba rápida a cuatros algoritmos de clasificación para el diagnóstico del Dengue.

4.2. Objetivos específicos

- Estimar los parámetros de desempeño adicionando la prueba rápida, de cuatros algoritmos de clasificación cuando la aplicación del proceso clasificador se realiza en paralelo.
- Estimar los parámetros de desempeño adicionando la prueba rápida, de cuatros algoritmos de clasificación cuando la aplicación del proceso clasificador se realiza en serie.
- Realizar las estimaciones asumiendo procedimientos propios del enfoque clásico y del enfoque bayesiano y comparar los resultados obtenidos.

5. Antecedentes

5.1. Antecedentes Clínicos

[Low et al., 2011] Realizó un estudio para determinar los primeros síntomas y signos del dengue en adultos con un total de 2129 pacientes. En general, los casos de dengue presentaron altas tasas de síntomas enumerados en los esquemas de clasificación de la OMS 1997 y 2009 para un cuadro febril probable de dengue, se obtiene una alta sensibilidad de estos esquemas cuando se aplica para el diagnóstico precoz. Sin embargo, síntomas como la mialgia, artralgia, dolor retroorbital y hemorragia de la mucosa se observaron con menos frecuencia en adultos mayores. Esta tendencia dio como resultado una reducción de la sensibilidad de los esquemas de clasificación de la OMS en adultos mayores a pesar de que mostraron un mayor riesgo de hospitalización por dengue grave.

Sugiere que los adultos mayores que presentan fiebre aguda y leucopenia se les aplique un diagnóstico diferencial de dengue mediante la clasificación de la OMS 2009 para el dengue grave con una confirmación de laboratorio adicional, incluso en ausencia de otros síntomas. Esto podría ser útil para el diagnóstico clínico temprano en adultos mayores para que puedan ser monitoreados y tratados por dengue severo.

[Osorio et al., 2010] Realizó la Comparación de la precisión diagnóstica de las pruebas comerciales basadas en el antígeno de proteína no estructural del dengue (NS1) para la infección temprana por dengue, por tanto, compara 5 pruebas comerciales: Platelia Dengue NS1, Dengue NS1-ELISA, Dengue Early Elisa, Dengue NS1-STRIP, SD BIOLINE - Dengue Duo (NS1/IgM/IgG) en 310 muestras de sangre.

Se usó los programas de Stata 10, Mathematica 7 (Wolfram 2010) y realizó la comparación de los resultados de cada prueba mediante las pruebas de Cochran Q, McNemar y la reproducibilidad por medio del índice Kappa. La sensibilidad obtenida para cada una de las pruebas comparadas en el estudio tuvieron una variación entre 51% al 80.7%, la prueba SD BIOLINE (NS1/IgM/IgG) obtuvo la mayor sensibilidad de 80.7%, seguida por SD BIOLINE (NS1/IgM) con el 78.4%, la prueba de Platelia con una sensibilidad del 70.8%, las pruebas de ELISA no obtuvieron diferencias significativas, la prueba Dengue NS1-ELISA obtuvo una sensibilidad del 68.8%, Dengue Early Elisa 71.1% y la prueba Dengue NS1-STRIP una sensibilidad del 57.7%, sin embargo se observó que la prueba Dengue NS1-STRIP obtuvo una mayor precisión (sensibilidad) a los 30 minutos de realizada la prueba que a los 15 minutos, obteniendo una sensibilidad del 71.4% y 68.7% respectivamente, por tanto como gran conclusión la adición de los anticuerpos IgM e IgG aumentaron la sensibilidad de la prueba SD BIOLINE en todas las situaciones obteniendo así un efecto muy positivo para el pronto diagnóstico del dengue.

5.2. Antecedentes Estadísticos

[Mendez, 2017] Uso una metodología para construir reglas de clasificación de Dengue, desarrollando la propuesta de 4 algoritmos de clasificación, utilizando un archivo de datos de las ciudades de Palmira, Cali, Neiva y Bucaramanga. Se definen unos subconjuntos de signos y síntomas que ingresan en cada prototipo por medio de la prueba Chi-Cuadrado y ayuda de 5 expertos, el uso de la estadística Bayesiana permitió el cálculo de las probabilidades predictivas que cuantifican la incertidumbre de tener dengue y se realiza remuestreo por Bootstrap para la validación externa.

Los 4 prototipos de algoritmos clasificadores obtuvieron resultados satisfactorios, con unas sensibilidades del 64.2%, 76%, 85.8% y 81.3% respectivamente adicionando los resultados del hemograma, realizando el remuestreo de bootstrap con 1000 muestras se obtuvo las sensibilidades del 63.5%, 74.2%, 88.4% y 77.9% teniendo en cuenta los resultados del hemograma, por tanto, los prototipos tres y cuatro presentan los mejores índices de rendimientos para el diagnóstico del dengue.

[**Fernández et al., 2016**] Realizó un análisis de ondículas que permite descomponer una serie de tiempo muy larga en sus componentes de tiempo calendario y periodo, utilizando una técnica estadística para generar un modelo exploratorio del comportamiento del dengue y obtener los patrones de estacionalidad interanual del dengue en Colombia, en los cinco municipios más endémicos, para el periodo 2007 a 2012, y de los ciclos entre los años 1978 y 2013 a nivel nacional. Se utilizó un modelo autorregresivo de primer orden, fue evidente el efecto de la epidemia de 2010 sobre la serie de tiempo a nivel nacional y la de los cinco municipios. Se observaron diferencias en los patrones de estacionalidad interanual por municipio. Asimismo, a nivel nacional se hallaron ciclos de dos a cinco años desde el 2004.

[**Hess et al., 2012**] Indica pautas y recomendaciones con el objetivo de brindar a los investigadores un recurso accesible para la evaluación y realización de informes técnicos sobre el rendimiento de pruebas dicotómicas, es decir, que se clasifican en dos categorías. En este presenta la comparación de una nueva prueba con una ya existente por medio de los parámetros estimados como la sensibilidad y especificidad, como también aplicando los test de McNemar y Liddell para la comparación. A su vez recomienda el proceso de estimación de la sensibilidad y especificidad para lograr estimaciones precisas con sus respectivos intervalos de confianza, cuando existe ausencia de un estándar de oro. Indica que cada vez que se evalúa un nuevo diagnóstico dicotómico, los investigadores deben tratar de compararlo con una prueba existente y se debe informar la sensibilidad y especificidad de la nueva prueba en comparación con la anterior, al incluir intervalos de confianza con estas estimaciones permite evaluar la precisión de la prueba y de las estimaciones. Aunque no existe un estándar para el ancho de los intervalos de confianza en la literatura bioestadística, los intervalos más estrechos son mejores.

6. Marco Teórico

6.1. Marco Teórico Clínico

6.1.1. Dengue

El dengue es la enfermedad viral transmitida por mosquito de más rápida propagación en el mundo, afectando principalmente países tropicales y subtropicales [OMS, 2009], presente en 128 países, donde aproximadamente 3.970 millones de personas están en riesgo de contraer la infección, es un problema creciente tanto en su distribución geográfica como en su intensidad [Brady et al., 2012]. El dengue es un virus que pertenece a la familia de flavivirus, la cual posee un antígeno de proteína no estructural llamada NS1, que esta contenida en los virus de fiebre amarilla, encefalitis japonesa y encefalitis transmitida por garrapata [Osorio et al., 2010]. Actualmente en Colombia se han identificado 4 serotipos de dengue que son distintos sin embargo están estrechamente emparentados (DEN-1, DEN-2, DEN-3 y DEN-4) [Padilla et al., 2012]. Cuando una persona se recupera de la infección adquiere inmunidad de por vida contra el serotipo en particular. Sin embargo, la inmunidad cruzada a los otros serotipos es parcial y temporal. Las infecciones posteriores causadas por otros serotipos aumentan el riesgo de padecer el dengue grave.

6.1.2. Diagnóstico médico

El diagnóstico médico de esta enfermedad es realizado a partir de los signos y síntomas que presenta el paciente considerando el cuadro clínico del mismo, este diagnóstico resulta muy importante por que al realizar el correspondiente estudio del cuadro clínico del individuo hecho por el doctor o médico tratante, puede evitar que el paciente llegue a una etapa letal de la enfermedad con

implicaciones graves.

6.1.3. Algoritmos de Clasificación

El algoritmo de clasificación para la enfermedad del dengue es un criterio clínico basado en los signos y síntomas, los cuales deben ser identificados por el médico y por consiguiente realizar un buen diagnóstico de la enfermedad.

La OMS ha planteado algoritmos de clasificación los cuales permiten realizar un diagnóstico de la enfermedad basado en un conjunto de signos y síntomas que diferencian el dengue de otras enfermedades febriles en el entorno de atención primaria, el algoritmo propuesto en el 2009 genera una nueva clasificación del Dengue y Dengue severo para el manejo oportuno de este virus.

1. Algoritmo 1: Este algoritmo se estableció a partir de doce signos y síntomas significativos que son presentes en individuos con dengue, realizando una secuencia de aplicación con Bayes discreto (Mendez 2018).
2. Algoritmo 2: Este algoritmo se planteó bajo ocho categorías que agrupan los signos y síntomas según sus características, realizando una secuencia de aplicación con Bayes discreto (Mendez 2018).
3. Algoritmo 3: Este algoritmo se estableció a partir de doce signos y síntomas significativos que son presentes en individuos con dengue, realizando una secuencia de aplicación con Bayes continuo (Mendez 2018).
4. Algoritmo 4: Este algoritmo se planteó bajo ocho categorías que agrupan los signos y síntomas según sus características, realizando una secuencia de aplicación con Bayes continuo (Mendez 2018).

Estos algoritmos son propuestos por Mendez 2018 a partir del algoritmo de la OMS del 2009.

6.1.4. Hemograma

El hemograma es un examen relativamente simple y en algunas situaciones ayuda en la evaluación diagnóstica. Este examen entrega datos sobre hematocrito (Hto), concentración de la hemoglobina (Hb), concentración de hemoglobina corpuscular media (CHCM), volumen corpuscular medio (VCM), recuento de eritrocitos, leucocitos y plaquetas.[Becker et al., 2001]

Para el diagnóstico del dengue se utilizan los recuentos de los leucocitos y plaquetas, a partir de los cuales se conocerá si el paciente padece de Leucopenia y Trombocitopenia.

- **Trombocitopenia:** Las plaquetas son pequeñas células que circulan en la sangre las cuales permiten la formación de coágulos de sangre y en la reparación de vasos sanguíneos dañados. El número y el tamaño de las plaquetas que tiene el paciente definen si este está expuesto a la enfermedad de trombocitopenia, en general se presenta la enfermedad cuando se tienen recuentos de plaquetas menor de 150.000/mm³ [Becker et al., 2001].
- **Leucopenia:** Los leucocitos son células sanguíneas llamadas glóbulos blancos de gran importancia para el funcionamiento idóneo del sistema inmunológico y defensa ante cualquier enfermedad. Se considera leucopenia la presencia al hemograma de menos de 4.000/mm³ [Becker et al., 2001].

6.1.5. Prueba Patrón de Oro para el Dengue

El rendimiento de todo test diagnóstico se basa en su comparación con un Gold Standard (Patrón de Oro). Es la técnica diagnóstica que define la presencia de la condición con la máxima certeza conocida [Saleh et al., 2008].

Para el diagnóstico de dengue existen pruebas Gold Standar las cuales tienen un 100% de confirmación de esta enfermedad, después del inicio de la enfermedad, el virus es fácilmente encontrado en las células sanguíneas, sueros o plasma y otros tejidos del paciente, las pruebas Gold Standar que se realizan son pruebas de laboratorio que duran entre 24 horas en dar resultados, la RT-PCR (**Reacción en Cadena de la Polimerasa con Transcriptasa Reversa**) realiza un diagnóstico molecular basado en la transcripción inversa, NS1 ELISA realiza un enzimoimmunoanálisis a partir

de la glucoproteína o antígeno que está altamente contenida por el virus del dengue y las pruebas IgM e IgG las cuales se realizan a partir de la detección cualitativa de los anticuerpos y la elevación de los mismo; en el proceso de laboratorio son pruebas costosas para los centros de salud, sin embargo el resultado es 100% confiable.

6.1.6. Prueba rápida para identificar individuos con Dengue.

A partir del estudio realizado por [Osorio et al., 2010] se obtuvo la prueba rápida DENGUE DUO SD-BIOLINE como la más sensible para el pronto diagnóstico del dengue, la cual, a partir de los anticuerpos IgM e IgG y el antígeno NS1 del virus, permite obtener resultados 15 minutos después de tomada la muestra de sangre o plasma e introducirla en la prueba.

Se espera una alta sensibilidad de esta prueba bajo un escenario epidémico del virus, ahora bien, otros estudios realizados demuestran que las pruebas basadas en el antígeno del dengue NS1 resultan altamente sensibles cuando se adicionan los anticuerpos que el paciente genera a partir de la primera infección permitiendo a su vez conocer la etapa de la infección y por lo tanto un pronto diagnóstico y buen manejo de los casos.

Los resultados de esta prueba son binarios, que indicarán si el individuo presenta positivo o negativo en los anticuerpos (IgM o IgG) y en NS1, por lo tanto la clasificación para obtener el resultado final de esta prueba es que al menos un resultado sea positivo, en la tabla **6-1** se especifica la interpretación de los resultados:

Tabla 6-1.: Interpretación de Resultados Prueba Rápida SD BIOLINE NS1/IgM/IgG

NS1	IgM	IgG	Resultado
+	-	-	+
-	+	-	+
-	-	+	+
+	+	+	+
-	-	-	-

6.1.7. Prueba diagnóstica

Una prueba diagnóstica es un proceso que pretende determinar en un paciente la presencia de cierta condición patológica, no susceptible de ser observable directamente con algunos de los cinco sentidos elementales [Morales and Zárate, 2004]. Son utilizadas para diferentes fines: tamizaje de una población, búsqueda de casos, descarte de un diagnóstico, confirmación de un diagnóstico o seguimiento de una patología [Morales and Zárate, 2004].

Tabla 6-2.: Clasificación de los resultados de una prueba diagnóstica

Prueba dicotómica clasificada como positiva o negativa en función de los resultados de la prueba		Enfermedad		
		Presente	Ausente	Total
	Positivo	Verdaderos Positivos (V+)	Falsos Positivos (F+)	Positivos
	Negativo	Falsos Negativos (F-)	Verdaderos Negativos (V-)	Negativos
	Total	Enfermos	Sanos	N

A partir de la tabla **6-2** se obtienen los indicadores importantes para el estudio de una prueba diagnóstica tales como: sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo y razón de verosimilitud positiva y negativa. [Donis, 2012].

6.1.8. Parámetros de desempeño de una prueba diagnóstica

[Bravo-Grau and Cruz, 2015] Define como prueba diagnóstica a cualquier método para obtener información adicional del estado de salud de un paciente, donde la utilidad de estas pruebas se cuantifican en términos de sensibilidad, especificidad y otros parámetros de interés que se determina a partir del test.

Estos parámetros son definidos como:

- **Sensibilidad $P(+|+)$:** Determina la capacidad de una prueba para identificar correctamente aquellos que tienen la enfermedad. Corresponde a la proporción de individuos con la enfer-

medad, Esta es igual al número de sujetos con un test positivo que tienen la enfermedad, dividido entre todos los sujetos que tienen la enfermedad [Donis, 2012, Bravo-Grau and Cruz, 2015].

- **Especificidad $P(-|-)$:** Es la capacidad de una prueba para identificar aquellos que no tienen la enfermedad. Corresponde a la proporción de individuos sin la enfermedad ,y es igual al número de sujetos que resultan negativos a la prueba y que no tienen la enfermedad, dividido entre el número de personas que no tienen la enfermedad o están sanos [Donis, 2012, Bravo-Grau and Cruz, 2015].

6.1.9. Diferentes procesos de aplicación de una prueba diagnóstica

Las pruebas diagnósticas no son perfectas, es por tal motivo se debe recurrir a múltiples pruebas que pueden ser realizadas en paralelo o en serie.

- **Pruebas en paralelo:** consiste en realizar simultáneamente dos o más pruebas en un individuo, y se realizan con el fin de aumentar la sensibilidad diagnóstica [Donis, 2012], además aumenta el valor predictivo negativo tras realizarse las dos pruebas y esto lleva a disminuir la especificidad pero se aumenta la sensibilidad, permitiendo detectar con mayor probabilidad los enfermos [Escrig-Sos et al., 2006].
- **Pruebas en serie:** Se aplican secuencialmente, ésto indica que se realiza una prueba en primer lugar, y después se indica la otra prueba solo sí el individuo resulta positivo en la primera prueba, para confirmar un primer resultado positivo, son aplicadas con el principal objetivo de aumentar la especificidad [Donis, 2012].

6.2. Marco Teórico Estadístico

6.2.1. Distribución de probabilidad multinomial

Teniendo en cuenta lo leído en [Wackerly and Scheaffer, 2010], con frecuencia encontramos situaciones similares en las que el número de posibles resultados por intento es más que dos. Por

ejemplo, experimentos que comprenden tipos de sangre por lo general tiene al menos cuatro posibles resultados por intento. Experimentos que comprenden el muestreo de artículos defectuosos donde se puede clasificar el tipo de defectos observados en más de dos clases, esta clase de problemas son asociados a modelo multinomial el cual es una generalización del experimento binomial.

La función de probabilidad de masa conjunta para las variables aleatorias de interés $y_1, y_2, \dots, y_k$ donde y_k es el resultado de cada intento que cae en una las k categorías o celdas, esta dado por

$$p(y_1, y_2, \dots, y_k | \theta_1, \theta_2, \dots, \theta_k) = \frac{n!}{y_1! y_2! \dots y_k!} \prod_{j=1}^k \theta_j^{y_j} \quad (6-1)$$

Donde $\theta_j \in \{0, 1\}$, $\sum_{j=1}^k \theta_j = 1$ y $\sum_{j=1}^k y_j = n$.

Se realiza la estimación considerando el método de multiplicadores de Lagrange a partir de las restricciones del modelo. El cual se plantea de la siguiente manera:

$$L(\theta_1, \theta_2, \dots, \theta_k | y_1, y_2, \dots, y_k) \propto \prod_{j=1}^k \theta_j^{y_j} \quad (6-2)$$

Aplicando logaritmo natural a la función de verosimilitud presentada en 6-2 y sujeto a la restricción $\sum_{j=1}^k \theta_j = 1, \theta_j \geq 1 (\forall j)$, se logra:

$$Z = Ln(L(\theta_1, \theta_2, \dots, \theta_k | y_1, y_2, \dots, y_k)) = \sum_{j=1}^k y_j Ln(\theta_j) - \lambda (\sum_{j=1}^k \theta_j - 1) \quad (6-3)$$

Al realizar los respectiva maximación a partir de la derivada y igualando a cero, se obtiene:

$$\frac{dZ}{d\theta_j} = \frac{y_j}{\theta_j} - \lambda = 0; \quad \theta_j = \frac{y_j}{\lambda}; \forall j \quad (6-4)$$

$$\frac{dZ}{d\lambda} = -(\sum_{j=1}^k \theta_j - 1) = 0; \quad \sum_{j=1}^k \theta_j = 1; \forall j \quad (6-5)$$

Desarrollando el sistema de ecuaciones para las expresiones obtenidas en 6-4 y 6-5, se llega al estimador máximo verosimil para una distribución multinomial (6-1) de la siguiente manera:

$$\hat{\theta}_j = \frac{y_j}{n} \quad (6-6)$$

6.2.2. Método Bootstrap no paramétrico

El bootstrap es un tipo de técnica de remuestreo de datos que permite resolver problemas relacionados con la estimación de intervalos de confianza [Ledesma, 2008]. El bootstrap puede incluirse dentro del enfoque mayor del remuestreo de datos o resampling, el cual comprende una variedad de técnicas que implementan procesos de simulación por computadora para estimar probabilidades empíricas a partir de experimentos con los propios datos [Rudner and Shafer, 1992].

[Behar, 2015] indica que para conocer o aproximar al valor de un parámetro desconocido θ se usa un estadístico $T_n = g(\underline{x})$, donde su varianza es $V(T_n)$. En pocas ocasiones la función de distribución de la población muestreada y/o de la estadística es conocida, al igual que la familia de distribuciones a la cual pertenece.

Dado el inconveniente de no conocer la distribución del estadístico, es de gran utilidad la técnica Bootstrap atribuida a Efrom, en la década del 70. Esta técnica consta de los siguientes pasos:

- De una muestra aleatoria $Y = (y_1, y_2, \dots, y_n)$
- Se generan B muestras con reemplazo de tamaño n, $Y_{1B}^*, Y_{2B}^*, \dots, Y_{nB}^*$.
- Para cada $Y_{1B}^*, Y_{2B}^*, \dots, Y_{nB}^*$ se hace el proceso de estimación del parámetro de interés $\hat{\theta}^{*B}$.
- Se obtiene una cantidad B de estimaciones del parámetro de interés, lo cual permitirá realizar inferencias a partir de éstas.

Los estimadores puntuales obtenidos a partir del remuestreo bootstrap se calculan mediante las siguientes formulas:

$$\hat{\theta}^* = \frac{\sum_{b=1}^B \hat{\theta}_b^*}{B}$$

$$Var(\hat{\theta}^*) = \frac{\sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta}^*)^2}{B - 1}$$

$$Sesgo = \hat{\theta}^* - \hat{\theta}$$

A partir de las ecuaciones anteriores se puede obtener el intervalo de confianza bootstrap del estimador, utilizando el intervalo de la distribución normal, puesto que la muestra original es lo

suficientemente grande. Se construye el intervalo de confianza con $(1 - \alpha)100\%$

$$\theta = [\hat{\theta}^* \pm Z_{\frac{\alpha}{2}} \sqrt{Var(\hat{\theta}^*)}]$$

Este intervalo funciona bien si la distribución del estimador por remuestreo Bootstrap tiene una distribución simétrica.

6.2.3. Procedimientos para simulación de datos binarios correlacionados

Existen diferentes métodos que permiten la simulación de datos correlacionados continuos, pero pocos son con datos binarios y aún más crece la dificultad al ser correlacionados, de tal manera que en este trabajo de grado existió la necesidad de realizar la simulación de una base de datos que contenga los algoritmos clasificatorios, la prueba rápida y el Gold Standard ; estos deben estar correlacionados ya que se aplican al mismo individuo y existirá por tanto una correlación entre los resultados de las variables anteriormente descritas, para esta simulación se utiliza la metodología propuesta por Park [Park et al., 1996], la cual indica:

Generar un vector aleatorio $(Z_1, \dots, Z_k)$ tales que $E(Z_i) = p_i$ y $corr(Z_i, Z_j) = \rho_{ij} > 0, i \neq j$. A partir de distribuciones poisson y su convolución para obtener datos binarios, primero se consideran $k = 2$ para presentar la idea y después se generaliza. Sea $X(\alpha)$ una variable aleatoria Poisson con media $\alpha > 0$. Se asume que las $X(\cdot)$ son mutuamente independientes si aparecen con diferentes subíndices. Se consideran dos variables aleatorias definidas por:

$$Y_1 = X_1(\alpha_{11} - \alpha_{12}) + X_3(\alpha_{12}) \quad (6-7)$$

$$Y_2 = X_2(\alpha_{22} - \alpha_{12}) + X_3(\alpha_{12}) \quad (6-8)$$

Donde α_{11} , α_{22} y α_{12} son constantes no negativas y hacen parte de los parámetros de la distribución poisson $X_i; i = 1, 2, 3$.

Obviamente, Y_1 y Y_2 siguen una distribución de Poisson. Debido a que comparten un término común $X_3(\alpha_{12})$, las dos variables de Poisson Y_1 y Y_2 están correlacionadas positivamente.

Las constantes α_{11} , α_{22} y α_{12} se pueden elegir de modo que $E(Z_1) = p_1$, $E(Z_2) = p_2$ y $\text{corr}(Z_1, Z_2) = \rho_{12}$.

$$E(Z_i) = E(Z_i^2) = P(X_i = X_3 = 0) = e^{-\alpha_{ii}} = p_i, i = 1, 2 \quad (6-9)$$

$$E(Z_1, Z_2) = P[X_1 = X_2 = X_3 = 0] = \exp[-(\alpha_{11} + \alpha_{22} - \alpha_{12})] = p_1 p_2 e^{\alpha_{12}} \quad (6-10)$$

Siendo la $\text{var}(Z_i) = p_i q_i$ y $\text{cov}(Z_1, Z_2) = p_1 p_2 (e^{-\alpha_{12}} - 1) = \rho_{12} (p_1 q_1 p_2 q_2)^{1/2}$.

Entonces, Z_1, Z_2 tienen coeficiente de correlación:

$$\rho_{12} = p_1 p_2 (e^{-\alpha_{12}} - 1) / (p_1 q_1 p_2 q_2)^{1/2} \quad (6-11)$$

Resolviendo la ecuación 6-11 dada la forma α'_{ij} s

$$\alpha_{ij} = \log \left[1 + \rho_{ij} \left(\frac{q_i q_j}{p_i p_j} \right)^{1/2} \right] \quad (6-12)$$

Por supuesto, el caso con $\alpha_{12} = 0$ corresponde a la independencia entre Z_1 y Z_2 . Tenga en cuenta que $E(Z_1, Z_2) = P(Z_1 = Z_2 = 1) \geq P(Z_i = 1) = p_i, i = 1, 2$, y tenemos $\text{cov}(Z_1, Z_2) \leq p_1 q_2$ y $\text{cov}(Z_1, Z_2) \leq p_2 q_1$. Por lo tanto, como se describe en [Emrich and Piedmonte, 1991], ρ_{12} no es libre sobre $[-1, 1]$. Es decir,

$$\rho_{12} \leq [p_2 q_1 / p_1 q_2] \quad y \quad \rho_{12} \leq [p_1 q_2 / p_2 q_1] \quad (6-13)$$

Por lo cual la fórmula 6-13 indica los valores máximos de la correlación para que exista solución entre Z_1 y Z_2 , se obtiene entonces que $Z_i \sim \text{Bernoulli}(p_i)$ donde:

$$Z_i = \begin{cases} 1 & \text{si } Y_i = 0 \\ 0 & \text{si } Y_i \neq 0 \end{cases} \quad i = 1, 2$$

Lo anterior puede generalizarse para más de dos variables.

6.2.4. Inferencia Bayesiana

El marco teórico en que se aplica la inferencia bayesiana es similar a la clásica: hay un parámetro poblacional respecto al cual se desea realizar inferencias y se tiene un modelo que determina la probabilidad de observar diferentes valores de X, bajo diferentes valores de los parámetros. Sin

embargo, la diferencia fundamental es que **la inferencia bayesiana considera al parámetro como una variable aleatoria**. Esto parecería que no tiene demasiada importancia, pero realmente si lo tiene pues conduce a una aproximación diferente para realizar el modelamiento del problema e inferencia propiamente dicha [Yumpaqui Pacheco, 2005].

En esencia, la inferencia bayesiana esta basada en la distribución de probabilidad del parámetro dado los datos **distribución a posteriori** $p(\theta|y)$, en lugar de la distribución de los datos dado el parámetro. Esta diferencia conduce a inferencias mucho más naturales, lo único que se requiere para el proceso de inferencia es la especificación previa de una **distribución a priori de probabilidad** $p(\theta)$, la cual representa el conocimiento acerca del parámetro antes de obtener cualquier informacion respecto a los datos [Yumpaqui Pacheco, 2005].

6.2.5. Formula de Bayes Caso Discreto

Según [Bernardo and Smith, 2001], sea $p(D|\theta)$, $\theta \in \{\theta_1, \dots, \theta_m\}$, el mecanismo de probabilidad que se supone ha generado los datos D , así que θ solo puede tomar un número finitos de valores. El uso finito de la forma del teorema de Bayes y omitiendo las condiciones predominantes de la notación, la probabilidad posterior de θ_i después de los datos D se ha observado es:

$$P_r(\theta_i|D) = \frac{p(D|\theta_i)P_r(\theta_i)}{\sum_{j=1}^m p(D|\theta_j)P_r(\theta_j)}; i = 1, \dots, m. \quad (6-14)$$

Para cualquier distribucion a priori $p(\theta) = \{P_r(\theta_1), \dots, P_r(\theta_m)\}$ describe el conocimiento disponible sobre el valor de θ , $P_r(\theta_i|D)$ mide que tan probable puede ser juzgado θ_i , dado tanto el conocimiento inicial descrito por la distribucion a priori, como la información por los datos D . [Bernardo and Smith, 2001]

6.2.6. Formula de Bayes Caso Continuo

En [Box and Tiao, 1992] indica que: Sea $y' = (y_1, \dots, y_n)$ un vector de n observaciones cuya distribución de probabilidad $p(y|\theta)$ depende de los valores de k parametros $\theta = (\theta_1, \dots, \theta_k)$, suponga tambien que θ cuenta con una distribución de probabilidad $p(\theta)$, entonces:

$$p(y|\theta) = p(y, \theta) = p(\theta|y)p(y) \quad (6-15)$$

Dado los datos observados de y , la distribución condicional de θ es:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d(\theta)} \quad (6-16)$$

Para el caso en la naturaleza de la variable aleatoria sea continua, donde la integral toma todos los valores admisibles dentro del rango de θ , y de donde $E[f(\theta)]$ es la esperanza matemática de $f(\theta)$ con respecto a la distribución $p(\theta)$. En esta expresión $p(\theta)$ es llamado la distribución apriori de θ , la cual condensa la información concerniente al parámetro θ antes de obtener los datos [Box and Tiao, 1992].

6.2.7. Distribución a priori

La distribución a priori cumple un papel importante en el análisis bayesiano ya que mide el grado de conocimiento inicial que se tiene de los parámetros en estudio [Yumpaqui Pacheco, 2005].

Si se tiene un conocimiento previo sobre los parámetros, esto se traducirá en una distribución a priori. Así, será posible plantear tantas distribuciones a priori como estados iniciales de conocimiento existan y los diferentes resultados obtenidos en la distribución bajo cada uno de los enfoques, adquiriran una importancia en relación con la convicción que tenga el investigador sobre cada estado inicial. Sin embargo, cuando nada es conocido sobre los parámetros, la selección de una distribución a priori adecuada adquiere una connotación especial pues será necesario elegir una distribución a priori que no influya sobre ninguno de los posibles valores de los parámetros en cuestión [Yumpaqui Pacheco, 2005]. Las distribuciones a priori pueden ser informativas o no informativas, las cuales son muy utiles según la necesidad del investigador.

- Distribución a priori no informativa

Una distribución de probabilidad a priori no informativa expresa la información de forma general acerca de una variable, los métodos utilizados son el método de LaPlace y el método de Jeffreys.

- Distribución a priori informativa

La distribución a priori informativa expresa información de forma más específica y definida acerca de una variable. Las distribuciones conjugadas es un método muy utilizado cuando la distribución $P(\theta)$ a priori de la variable pertenece a la familia de distribuciones exponenciales [Gelman et al., 2014].

6.2.8. Distribución Beta conjugada

La distribución Beta conjugada es considerada la distribución a priori de distribuciones tales como la Uniforme, Binomial, Bernoulli, entre otras. Por tanto, se denota la distribución Beta como la distribución a priori para el parametro θ , con función de densidad.

$$\pi(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \quad (6-17)$$

Ahora, el proceso para la obtención de la distribución posterior es el que sigue, teniendo presente la función de verosimilitud:

$$\begin{aligned} \pi(\theta|y_1, y_2, \dots, y_n) &\propto l(y_1, y_2, \dots, y_n|\theta)\pi(\theta) \\ &\propto \theta^y (1 - \theta)^{n-y} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \\ &\propto \theta^{\alpha+y-1} (1 - \theta)^{n-y+\beta-1} \end{aligned} \quad (6-18)$$

Por tanto, la distribución posterior es una $\text{Beta}(\alpha + y, \beta + n - y)$, donde y es el número de éxitos y $n - y$ los fracasos, por consiguiente es necesario realizar la elicitación de los hiperparametros por los métodos propuestos para tal fin [Zhu and Lu, 2004].

6.2.9. Elicitación de hiperparametros para familias de distribuciones Beta

El proceso para convertir el conocimiento que tiene un experto sobre los posibles valores que puede tomar un parámetro de interés, en una distribución conjunta de probabilidades, es llamado elicitación [Tovar, 2012].

En [Tovar, 2012] se describe el por que es posible modelar el comportamiento de un parámetro el cual esta definido en el intervalo $[0,1]$ mediante una distribución de la familia de distribuciones beta, el argumento principal se fundamenta en que esta familia de distribuciones en particular cuenta con una amplia variedad de formas y sus componentes son conjugadas a priori naturales de la distribución binomial.

Vamos a definir por θ la proporción de interes que es una cantidad aleatoria con una distribución de probabilidad desconocida $\pi(\theta)$. Una elección natural para $\pi(\theta)$ es la distribución Beta con hiperparámetros a y b . Usando la información disponible obtenida de alguna fuente, es posible determinar intervalos de probabilidad de $(1-\alpha)\%$ con límites θ_1 y θ_2 igual que [Joseph et al., 1995] hicieron dentro de un enfoque de intervalos fijos; por lo tanto, $(1 - \alpha)\%$ de los valores de θ cae dentro del intervalo $[\theta_1, \theta_2]$. En una segunda etapa, el centro del intervalo (θ_0) se corresponde con la media de la distribución beta(a,b); es decir:

$$\theta_0 = \frac{\theta_1 + \theta_2}{2} \cong \frac{a}{a + b} \quad (6-19)$$

Como la media $\theta_0 = E(\theta)$ y la varianza $\sigma^2 = V(\theta)$ se puede escribir como funciones de los hiperparámetros previos de Beta, entonces el problema se convierte en una cuestion de resolver un sistema de dos ecuaciones con dos incógnitas [Tovar, 2012].

Se tiene

$$\omega = \frac{\theta_0}{(1 - \theta_0)} a = \omega b \quad b = \frac{\omega - [(\omega + 1)^2 \sigma^2]}{(\omega^3 + 3\omega^2 + 3\omega + 1) \sigma^2} \quad (6-20)$$

a y b son los hiperparámetros que indexan la distribución Beta.

7. Metodología

La simulación es el proceso de diseñar y desarrollar un modelo computarizado de un sistema o proceso y conducir experimentos con éste con el propósito de entender el comportamiento del sistema o evaluar varias estrategias con las cuales se puede operar, siendo este proceso lo más cercano a la realidad [Bú, 1996]. Por tanto, se realiza la simulación de archivos de datos a partir del algoritmo de [Park et al., 1996], donde el primer archivo cuenta con las categorías de signos y síntomas utilizadas como insumo para el algoritmo I y III, una vez creado, se realizó el proceso de clasificación planteado en el trabajo de grado de [Mendez, 2017].

El segundo archivo de datos se simuló a partir de las estimaciones obtenidas por [Mendez, 2017] ver tabla 7-3 y la prueba rápida bajo tres diferentes escenarios de prevalencia 10%, 50% y 90%, en consecuencia, se busca evaluar el desempeño de los cuatro algoritmos de clasificación para el diagnóstico de dengue adicionando los resultados de una prueba rápida.

7.1. Identificación de Variables para la simulación

Los métodos para diagnosticar Dengue indicados en [Low et al., 2011] y [Fernández et al., 2016] usan los signos y síntomas que esta enfermedad ocasiona en los pacientes que la padecen, donde se identificaron 25 signos y síntomas con la ayuda de médicos generales expertos en el diagnóstico del Dengue presentadas en la tabla 7-1.

Debido a la gran cantidad de signos y síntomas de dengue relacionados, se realizó la categorización de los mismos en ocho categorías (ver tabla 7-2), con la ayuda de médicos con amplia experiencia tratando individuos con dengue.

Se obtuvo el diagnóstico final del dengue realizado por la prueba Gold Standard o patrón de Oro, delimitada como variable binaria al igual que las otras relacionadas. Posteriormente, se hizo las

definiciones de los 25 signos y síntomas identificados para esta enfermedad (ver tabla 7-1), las ocho categorías obtenidas (ver tabla 7-2). Los signos y síntomas consignados en la tabla 7-1 fueron los principales insumos del trabajo de [Mendez, 2017] para la construcción de los algoritmos para el diagnóstico de Dengue.

Tabla 7-1.: Definiciones de signos y síntomas

Signos y Síntomas	Descripción
Artralgias	dolor en una o más articulaciones.
Cefalea	dolor de cabeza intenso y persistente, acompañado de sensación de pesadez.
Dolor Retroocular	Sensación urente pulsátil, dolorosa y lacerante en los alrededores del ojo.
Mialgias	Dolor muscular
Diarrea	Alteración intestinal caracterizada por mayor frecuencia y fluidez de las deposiciones.
Dolor abdominal	Dolor en la zona abdominal.
Hígado palpable o Hepatomegalia	Aumento tamaño del hígado sobre los límites estimados.
Vomito	expulsión violenta del contenido del estómago a través de la boca.
Edemas	hinchazón causada por la acumulación de líquido en los tejidos del cuerpo.
Escalofríos	Temblo involuntario y repentino bajo sensación de frío.
Hiporexia	Perdida o disminución parcial del apetito.
Exantema	Erupción de la piel, de color rojizo y más o menos extensa.
Erupción maculopapular	Enfermedad de la piel que se caracteriza por tener ciertas áreas de la piel roja.
Eritema	Enrojecimiento de la piel debido al aumento de la sangre contenida en los capilares.
Rash	Erupción que se presenta con cambios en el color o textura de la piel.
Hemorragias	Pérdida de sangre provocada por rupturas de vasos sanguíneos como venas, arterias y capilares.
Petequeias	Pequeños derrames vasculares cutáneos del tamaño de la cabeza de un alfiler.
Ictericia	Coloración amarillenta de la piel y las mucosas que se produce por un aumento de bilirrubina en la sangre como resultado de ciertos trastornos hepáticos..
Irritabilidad	Tendencia a irritarse por cualquier cosa.
Insomnio	Falta de sueño.
Somnolencia	Falta de sueño.
Odinofagia	Dolor de garganta.
Rinorrea	Flujo o emisión abundante de líquido por la nariz.
Inyección Conjuntival	Enrojecimiento del glóbulo ocular.
Tos	Reflejo del cuerpo que mantiene despejada la garganta y las vías respiratorias.
Leucopenia	Trastorno de la sangre, causado por la disminución de los glóbulos blancos en la sangre.
Trombocitopenia	Cualquier situación de disminución de la cantidad de plaquetas circulantes en el torrente sanguíneo por debajo de los niveles normales.

Tabla 7-2.: Categorías de signos y síntomas conformadas por los clínicos especialistas en dengue.

CATEGORÍA	SINTOMA	TIPO
DOLOR GENERAL	Artralgias	Binaria
	Cefalea	
	Dolor retroocular	
	Mialgias	
SIGNO DE ALARMA	Diarrea	Binaria
	Dolor abdominal	
	Hígado Palpable	
	Vómito	
	Edemas	
MALESTAR GENERAL	Escalofríos	Binaria
	Hiporexia	
ERUPCIÓN CUTÁNEA	Exantema	Binaria
	Erupción Maculopapular	
	Eritema	
	Rahs	
MANIFESTACIONES HEMORRÁGICAS	Hemorragias	Binaria
	Petequeias	
ICTERICIA (*)	Ictericia	Binaria
ALTERACIÓN NEUROLÓGICA	Irritabilidad	Binaria
	Insomnio	
	Somnolencia	
SÍNTOMA RESPIRATORIO(*)	Odinofagia	Binaria
	Rinorrea	
	Inyección Conjuntival	
	Tos	

Donde las categorías ictericia y síntomas respiratorios son diferenciables para dengue, ya que si el individuo presenta estas categorías tienen una alta probabilidad de no tener el virus.

A partir de las categorías relacionadas en la tabla 7-2 se produce la simulación del primer archivo de datos, una vez obtenido el archivo con los vectores binarios se realizó la clasificación con los algoritmos I y III bajo el proceso propuesto por Mendez,2017.

Para la simulación del segundo archivo de datos se tomaron como principales insumos las estimaciones de los algoritmos de clasificación propuestos con sus respectivos intervalos de confianza (Ver tabla 7-3) y el valor de la estimación de la prueba rápida obtenida por Osorio et al.,2010.

Tabla 7-3.: Insumos para la simulación del segundo archivo de datos - estimaciones de los algoritmos de clasificación (Valores tomados de [Mendez, 2017]).

Descripción	Medida	Sin Hemograma		Con Hemograma	
		Valor	IC 95%	Valor	IC 95%
Algoritmo 1	Sensibilidad	0,666	< 0, 446; 0, 885 >	0,642	< 0, 591; 0, 692 >
	Especificidad	0,590	< 0, 553; 0, 628 >	0,656	< 0, 573; 0, 738 >
Algoritmo 2	Sensibilidad	0,672	< 0, 643; 0, 700 >	0,76	< 0, 728; 0, 793 >
	Especificidad	0,677	< 0, 646; 0, 707 >	0,724	< 0, 688; 0, 758 >
Algoritmo 3	Sensibilidad	0,722	< 0, 687; 0, 756 >	0,858	< 0, 820; 0, 895 >
	Especificidad	0,603	< 0, 549; 0, 656 >	0,520	< 0, 432; 0, 607 >
Algoritmo 4	Sensibilidad	0,7572	< 0, 731; 0, 782 >	0,813	< 0, 783; 0, 842 >
	Especificidad	0,816	< 0, 790; 0, 841 >	0,882	< 0, 856; 0, 907 >

Ademas de los resultados desarrollados por los algoritmos para la simulación del segundo archivo de datos, se tuvo en cuenta las estimaciones de los parámetros de desempeño de la prueba rápida publicadas en el articulo de Osorio et al., 2010, la cual presento una sensibilidad del 80,7%, ($IC_{95\%}$: 75%; 85,7%) y una especificidad del 89,1%, ($IC_{95\%}$: 81%; 94,7%).

Utilizando la información de la tabla 7-3 y considerando los resultados reportados por Osorio et al., 2010 se obtuvo el segundo archivo de valores binarios.

7.2. Archivos de datos

7.2.1. Archivo 1

El primer archivo de datos tiene como variables principales las 8 categorías (ver tabla 7-2) para el diagnóstico de dengue, a partir de éstos se realizó la clasificación de cada individuo bajo el proceso realizado por [Mendez, 2017] para el algoritmo I Y III teniendo en cuenta a su vez los resultados del hemograma y la prueba patrón de oro. La estructura de este archivo de datos es la siguiente:

Tabla 7-4.: Estructura general del primer archivo de datos

$C_1, \dots, C_8$	A_1, A_3	GS	H
11...0	0 1	1	1
10...1	0 1	1	1
01...0	1 1	0	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
10...0	0 1	1	0

7.2.2. Archivo 2

El segundo archivo de datos dispone de los resultados del diagnóstico de dengue para los cuatro algoritmos sin hemograma y con hemograma, a su vez de la prueba rápida. A partir de estos, se procedió con la clasificación en caso de aplicar las pruebas en serie o paralelo.

La base tiene la siguiente estructura (A:Algoritmo sin hemograma-AH:Algoritmo con hemograma):

Tabla 7-5.: Estructura general del segundo archivo de datos

$A_1, \dots, A_4$	$AH_1, \dots, AH_4$	PR	GS
11...0	11...0	0	1
10...1	10...1	0	1
01...0	01...0	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
11...0	11...0	0	0

Los dos archivos de datos tienen 10000 individuos y son simuladas con el algoritmo propuesto por Park [Park et al., 1996].

7.2.3. Algoritmo implementado para obtener los archivos de datos

Para la simulación de los archivos de datos se implementa a partir del paquete estadístico R [R Development Core Team, 2007], aplicando la siguiente descripción del algoritmo:

Inicialmente, se deben plantear las proporciones iniciales p_i descritas en la ecuación 6-9. En el contexto de este trabajo de grado, se usó la proporción de individuos enfermos obtenidos de la

base de datos de [Mendez, 2017] para el primer archivo, por otra parte, en el segundo archivo de datos se utilizaron como proporciones iniciales los valores obtenidos de las estimaciones de los parámetros de desempeño para los algoritmos de clasificación mostrados en la tabla 7-3 y la prueba rápida.

Al tener en cuenta lo anterior, se procede a simular el primer archivo de datos posteriormente el segundo archivo siguiendo el paso a paso propuesto en el artículo de [Park et al., 1996], cuando se considera un conjunto de k variables de Poisson aleatorias $Y_1, Y_2, \dots, Y_k$.

Paso 0. Se ingresan las proporciones iniciales p_i .

Paso 1. A partir de las proporciones iniciales se calculan las constantes α_{ij} como se presenta en la ecuación 6-12 para $1 \leq i, j \leq k$. Usando un ciclo desde $l = 0$ y sea α_{ij} . Ya que estas constantes son los parámetros de las distribuciones Poisson para obtener columnas binarias en el proceso de convolución.

Paso 2. En este paso se obtienen los coeficientes de correlación ρ_{ij} , las cuales son simuladas teniendo en cuenta el intervalo presentado en la ecuación 6-12.

Paso 3. Con el uso del ciclo (repeticiones), se realizó lo siguiente: En cada ciclo se tiene un conjunto de $T_l = \alpha_{ij} : \alpha_{ij} > 0, 1 \leq i, j \leq k$, por tanto se toma $\beta_l = \alpha_{rs}$ como el elemento más pequeño en el conjunto T_l . A partir de esto debe cumplirse que si $\alpha_{rs} = 0$ o $\alpha_{ss} = 0$, entonces el ciclo para. De lo contrario, se repite el proceso hasta obtener los elementos que cumplen con las condiciones anteriores.

Paso 4. Una vez obtenido el valor de los β_l 's en el paso tres, se reemplaza α_{ij} con la diferencia de $\alpha_{ij} - \beta_l$. Entonces si se cumple que todo $\alpha_{ij} = 0$, se continua para obtener los vectores binarios deseados, en caso contrario se retrocede al paso uno.

Paso 5. Como último paso, una vez que se cumpla con todos los requisitos anteriores se obtiene el archivo final con los vectores binarios correlacionados como se muestra en la ecuación 6-8.

Es importante resaltar que en el segundo archivo se consideran tres valores de prevalencia los cuales son del 10%, 50% y 90% y tres valores de sensibilidad para la prueba rápida (75%-80.7%-85,7%). Como se dijo anteriormente, las proporciones iniciales para el archivo 2 son valores que se encontraban dentro del intervalo de confianza para cada parámetro de desempeño expuesto en la tabla 7-3 y la prueba rápida, esto con el propósito de obtener las estimaciones deseadas y así

realizar las pertinentes comparaciones al adicionar los resultados de la prueba rápida simulada. Las correlaciones se obtuvieron de forma aleatoria teniendo en cuenta los valores máximos mostrados en la ecuación 6-13.

7.3. Experimento Estadístico

Para efectos de la elaboración del modelo que representa la situación en estudio se hace necesario identificar un experimento estadístico el cual obedece a un diseño de cohorte con dos tratamientos donde a un individuo con cuadro febril como criterio de inclusión que realiza una consulta con el médico, se le aplica los algoritmos de clasificación de dengue propuestos, la prueba de sangre con la cual se obtiene el cuadro hemático del paciente (resultados del Hemograma) y la prueba rápida, por último es ejecutado la prueba confirmatoria llamada patrón de oro. Los archivos simulados se realizan asumiendo el proceso descrito aplicado en la realidad.

7.4. Modelo Estadístico

Sea $S_1, S_2, \dots, S_n$ los registros de signos y síntomas de los individuos que consultaron con el médico y presentaban cuadro febril, por tanto, se definen los siguientes eventos:

$D_1 = \{ \text{Individuo con resultado positivo en la prueba patrón de oro} \}$

$D_2 = \{ \text{Individuo con resultado positivo en el algoritmo} \}$

$D_3 = \{ \text{Individuo con resultado positivo en la prueba rápida} \}$

A cada evento se le asigna una variable aleatoria que cuantifique el respectivo resultado, tal que:

$$X_1 = \begin{cases} 1 & \text{si ocurre } D_1 \\ 0 & \text{si ocurre } D_1^c \end{cases} \quad X_2 = \begin{cases} 1 & \text{si ocurre } D_2 \\ 0 & \text{si ocurre } D_2^c \end{cases} \quad X_3 = \begin{cases} 1 & \text{si ocurre } D_3 \\ 0 & \text{si ocurre } D_3^c \end{cases}$$

Por lo tanto $X_i \sim \text{Bernoulli}(\pi_i)$ donde $\pi_i = P(X_i = 1)$

En este caso son asignados dos tratamientos y un patrón de oro, se tiene que las cantidades observadas son aquellas contenidas en las ocho celdas de la tabla 7-6 de clasificación generada a partir de los eventos D_i con $i = 1, 2, 3$ y la función de verosimilitud esta asociada a una distribución multinomial como en la ecuación 6-1.

Tabla 7-6.: Tabla de clasificación

	Patrón oro +		Total	Patrón oro -		Total
	Prueba Rápida +	Prueba Rápida -		Prueba Rápida +	Prueba Rápida -	
<i>Algoritmo +</i>	y_1	y_2	$y_1 + y_2$	y_5	y_6	$y_5 + y_6$
<i>Algoritmo -</i>	y_3	y_4	$y_3 + y_4$	y_7	y_8	$y_7 + y_8$
Total	$y_1 + y_3$	$y_2 + y_4$	n^+	$y_5 + y_7$	$y_6 + y_8$	n^-

De modo que se tienen Y_r repeticiones de un experimento bernoulli con probabilidad de éxito θ_i , de esta forma es posible decir que:

$$Y_r = \sum_{j=1}^N X_{rj}; \quad r = 1, 2, 3, \dots, 8$$

Que identifica los eventos formados a partir de las combinaciones de resultados entre los tratamientos y el patrón de oro aplicada a una población n de individuos, es decir:

$$y_1 = 1 \text{ si } X_1 = 1, X_2 = 1, X_3 = 1$$

$$y_2 = 1 \text{ si } X_1 = 1, X_2 = 1, X_3 = 0$$

$$y_3 = 1 \text{ si } X_1 = 1, X_2 = 0, X_3 = 1$$

$$y_4 = 1 \text{ si } X_1 = 1, X_2 = 0, X_3 = 0$$

$$y_5 = 1 \text{ si } X_1 = 0, X_2 = 1, X_3 = 1$$

$$y_6 = 1 \text{ si } X_1 = 0, X_2 = 1, X_3 = 0$$

$$y_7 = 1 \text{ si } X_1 = 0, X_2 = 0, X_3 = 1$$

$$y_8 = 1 \text{ si } X_1 = 0, X_2 = 0, X_3 = 0$$

Cada combinación de resultados define una categoría. El experimento estadístico asociado al problema se ajusta aquel presentado en la sección 7.3 y los componentes del vector de probabilidades $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_8)$ de la función conjunta de masa de la distribución multinomial (6-1), son las probabilidades de ocurrencia de los eventos que definen las categorías, de tal manera que:

$$\theta_j = P(Y_r = 1)$$

Por supuesto, cada categoría identifica una subpoblación compuesta por n_r individuos, de modo que $\sum_{r=1}^8 n_r = n$.

Dado que se requiere estimar los parámetros de desempeño de las pruebas de diagnóstico para dengue, es necesario reparametrizar sobre los componentes del vector $\vec{\theta}$ de la distribución multinomial de modo que se expresen en términos de la prevalencia, sensibilidad y especificidad, además se tiene en cuenta la dependencia que existe entre las pruebas, siendo éstas conocidas por medio de la simulación, en consecuencia se define:

$$\theta_1 = P(X_1 = 1, X_2 = 1, X_3 = 1) = P(X_1 = 1)P(X_2 = 1|X_1 = 1)P(X_3 = 1|X_1 = 1) = p[\alpha_1\alpha_2 + \psi_D]$$

$$\theta_2 = P(X_1 = 1, X_2 = 1, X_3 = 0) = P(X_1 = 1)P(X_2 = 1|X_1 = 1)P(X_3 = 0|X_1 = 1) = p[\alpha_1(1 - \alpha_2) - \psi_D]$$

$$\theta_3 = P(X_1 = 1, X_2 = 0, X_3 = 1) = P(X_1 = 1)P(X_2 = 0|X_1 = 1)P(X_3 = 1|X_1 = 1) = p[(1 - \alpha_1)\alpha_2 - \psi_D]$$

$$\theta_4 = P(X_1 = 1, X_2 = 0, X_3 = 0) = P(X_1 = 1)P(X_2 = 0|X_1 = 1)P(X_3 = 0|X_1 = 1) = p[(1 - \alpha_1)(1 - \alpha_2) + \psi_D]$$

$$\theta_5 = P(X_1 = 0, X_2 = 1, X_3 = 1) = P(X_1 = 0)P(X_2 = 1|X_1 = 0)P(X_3 = 1|X_1 = 0) = (1 - p)[(1 - \beta_1)(1 - \beta_2) + \psi_{ND}]$$

$$\theta_6 = P(X_1 = 0, X_2 = 1, X_3 = 0) = P(X_1 = 0)P(X_2 = 1|X_1 = 0)P(X_3 = 0|X_1 = 0) = (1 - p)[(1 - \beta_1)\beta_2 - \psi_{ND}]$$

$$\theta_7 = P(X_1 = 0, X_2 = 0, X_3 = 1) = P(X_1 = 0)P(X_2 = 0|X_1 = 0)P(X_3 = 1|X_1 = 0) = (1 - p)[\beta_1(1 - \beta_2) - \psi_{ND}]$$

$$\theta_8 = P(X_1 = 0, X_2 = 0, X_3 = 0) = P(X_1 = 0)P(X_2 = 0|X_1 = 0)P(X_3 = 0|X_1 = 0) = (1 - p)[\beta_1\beta_2 + \psi_{ND}]$$

Donde las letras α , β y p corresponden a las sensibilidades, especificidades y la prevalencia respectivamente y ψ_D y ψ_{ND} son las covarianzas de los individuos sanos (ND) y enfermos (D) determinadas por la dependencia inducida al obtener variables binarias correlacionadas por el proceso de simulación, es decir, que los algoritmos de clasificación y prueba rápida anteriormente expuestas se simulan considerando que a cada paciente se le aplican todas las pruebas a la vez.

Teniendo en cuenta las expresiones anteriores, se procede a reescribir la función de verosimilitud de la ecuación 6-1 de la siguiente manera:

$$L(p, \alpha_1, \alpha_2, \beta_1, \beta_2 | y_1, \dots, y_8, n) = \frac{n!}{\prod_{j=1}^8 y_j!} (p^{y_1+y_2+y_3+y_4} (1-p)^{y_5+y_6+y_7+y_8}) (\alpha_1 \alpha_2 + \psi_D)^{y_1} (\alpha_1(1-\alpha_2) - \psi_D)^{y_2} ((1-\alpha_1)\alpha_2 - \psi_D)^{y_3} ((1-\alpha_1)(1-\alpha_2) + \psi_D)^{y_4} ((1-\beta_1)(1-\beta_2) + \psi_{ND})^{y_5} ((1-\beta_1)\beta_2 - \psi_{ND})^{y_6} (\beta_1(1-\beta_2) - \psi_{ND})^{y_7} (\beta_1\beta_2 + \psi_{ND})^{y_8} \quad (7-1)$$

7.5. Escenarios para la obtención del resultado clasificador

Los siguientes escenarios se plantean con la ayuda de expertos en el diagnóstico de esta enfermedad, bajo el conocimiento clínico de los mismos.

7.5.1. Pruebas para validación aplicadas en Paralelo

En este escenario las pruebas se aplicaron en paralelo, esto indica que las tres pruebas, algoritmos, prueba rápida, hemograma y prueba confirmatoria patrón de oro fueron aplicadas a todos los individuos simultáneamente. Para este escenario se tiene que el resultado formal del proceso clasificador se obtiene de la siguiente manera:

$$P^+ = \begin{cases} \text{Si el individuo tiene : } & \text{Algoritmo} + \text{ y Prueba Rapida} + \\ & \text{o Algoritmo} - \text{ y Prueba Rapida} + \\ & \text{o Algoritmo} + \text{ y Prueba Rapida} - \end{cases} \quad (7-2)$$

$$P^- = \begin{cases} \text{Si el individuo tiene : } & \text{Algoritmo} - \text{ y Prueba Rapida} - \end{cases} \quad (7-3)$$

Sea U una variable aleatoria que cuantifica el respectivo resultado final del clasificador aplicado en Paralelo.

$$U = \begin{cases} 1 & \text{si ocurre } P^+ \\ 0 & \text{si ocurre } P^- \end{cases}$$

Por tanto $U \sim \text{Bernoulli}(\pi)$.

7.5.2. Pruebas de validación aplicadas en Serie.

En este escenario las pruebas se emplean en serie con el fin de aumentar sensibilidad (Serie Negativo) y especificidad (Serie Positivo), esto indica que la aplicación de las tres pruebas dependerá de lo que ocurra con el resultado de la prueba anteriormente realizada. En este marco es considerado los siguientes eventos:

Serie Negativo

$$Serie_{Negativo}^+ = \begin{cases} Si\ el\ individuo\ tiene: & Algoritmo + no\ se\ realiza\ la\ Prueba\ Rapida \\ & o\ Algoritmo - y\ Prueba\ Rapida + \end{cases} \quad (7-4)$$

$$Serie_{Negativo}^- = \begin{cases} Si\ el\ individuo\ tiene: & Algoritmo - y\ Prueba\ Rapida - \end{cases} \quad (7-5)$$

Serie Positivo

$$Serie_{Positivo}^+ = \begin{cases} Si\ el\ individuo\ tiene: & Algoritmo + y\ Prueba\ Rapida + \end{cases} \quad (7-6)$$

$$Serie_{Positivo}^- = \begin{cases} Si\ el\ individuo\ tiene: & Algoritmo + y\ Prueba\ Rapida - \\ & o\ Algoritmo - no\ se\ realiza\ la\ Prueba\ Rapida \end{cases} \quad (7-7)$$

Sea V_1 y V_2 variables aleatorias, tales que:

$$V_1 = \begin{cases} 1 & \text{si ocurre } Serie_{Negativo}^+ \\ 0 & \text{si ocurre } Serie_{Negativo}^- \end{cases} \quad V_2 = \begin{cases} 1 & \text{si ocurre } Serie_{Positivo}^+ \\ 0 & \text{si ocurre } Serie_{Positivo}^- \end{cases}$$

Por tanto $V_i \sim Bernoulli(\pi_i)$ $i = 1, 2$. Los escenarios en serie son necesarios puesto que son los procedimientos clínicos más ajustados a lo que ocurre en la realidad.

7.6. Estimación de Parámetros de desempeño

Aplicando el método de máxima verosimilitud a la función 7-1, se obtienen las respectivas estimaciones de los parámetros de desempeño para cada algoritmo clasificador a evaluar y prueba rápida

expuestas en las ecuaciones 7-8 al 7-11.

Tanto las sensibilidades como las especificidades y sus intervalos de confianza fueron estimados usando cada uno de los cuatro algoritmos de clasificación y los escenarios de aplicación propuestos en el presente trabajo de grado.

7.6.1. Estimación Clásica

Asumiendo el procedimiento propio del paradigma clásico de la estadística se realizó la estimación por máxima verosimilitud para obtener las sensibilidades y especificidades, las cuales estarán identificadas con las letras α y β respectivamente, por tanto, las estimaciones serán:

$$\hat{\alpha}_1 = \frac{y_1 + y_2}{\sum_{j=1}^4 y_j} = \frac{y_1 + y_2}{n^+} \quad (7-8)$$

$$\hat{\beta}_1 = \frac{y_7 + y_8}{\sum_{j=5}^8 y_j} = \frac{y_7 + y_8}{n^-} \quad (7-9)$$

$$\hat{\alpha}_2 = \frac{y_1 + y_3}{\sum_{j=1}^4 y_j} = \frac{y_1 + y_3}{n^+} \quad (7-10)$$

$$\hat{\beta}_2 = \frac{y_6 + y_8}{\sum_{j=5}^8 y_j} = \frac{y_6 + y_8}{n^-} \quad (7-11)$$

Donde α_1 y β_1 , es la sensibilidad y especificidad correspondiente al algoritmo, en contraste, α_2 y β_2 es la sensibilidad y especificidad respecto a la prueba rápida. A continuación se presentan sus intervalos de confianza: :

$$IC\langle\alpha_k\rangle_{1-\alpha} = \left\{ \hat{\alpha}_k \pm Z_{1-\alpha} \sqrt{\frac{\hat{\alpha}_k(1 - \hat{\alpha}_k)}{n^+}} \right\} \quad k = 1, 2. \quad (7-12)$$

$$IC\langle\beta_k\rangle_{1-\alpha} = \left\{ \hat{\beta}_k \pm Z_{1-\alpha} \sqrt{\frac{\hat{\beta}_k(1 - \hat{\beta}_k)}{n^-}} \right\} \quad k = 1, 2. \quad (7-13)$$

7.6.2. Estimación de parámetros de desempeño para pruebas aplicadas en paralelo

Teniendo en cuenta los eventos planteados en 7-2 y 7-3; se realizó la estimación de los parámetros de desempeño los cuales serán:

Paralelo

$$\hat{\alpha}_3 = \frac{y_1 + y_2 + y_3}{\sum_{j=1}^4 y_j} = \frac{y_1 + y_2 + y_3}{n^+} \quad (7-14)$$

$$\hat{\beta}_3 = \frac{y_8}{\sum_{j=5}^8 y_j} = \frac{y_8}{n^-} \quad (7-15)$$

En este escenario, α_3 y β_3 es la sensibilidad y especificidad en el marco que se consideren la aplicación de las pruebas en paralelo. Los intervalos de confianza en cada caso se hacen a partir de las ecuaciones 7-12 y 7-13.

7.6.3. Estimación de parámetros de desempeño para pruebas aplicadas en serie

A partir de los eventos planteados en 7-4 al 7-7 se realizó la estimación de los parámetros:

Serie Negativo

$$\hat{\alpha}_4 = \frac{y_1 + y_2 + y_3}{\sum_{j=1}^4 y_j} = \frac{y_1 + y_2 + y_3}{n^+} \quad \hat{\beta}_4 = \frac{y_8}{\sum_{j=5}^8 y_j} = \frac{y_8}{n^-}$$

Serie Positivo

$$\hat{\alpha}_5 = \frac{y_1}{\sum_{j=1}^4 y_j} = \frac{y_1}{n^+} \quad \hat{\beta}_5 = \frac{y_6 + y_7 + y_8}{\sum_{j=5}^8 y_j} = \frac{y_6 + y_7 + y_8}{n^-}$$

Donde α representa la sensibilidad y β la especificidad. De forma similar para obtener los intervalos de confianza se obtuvo como en las ecuaciones 7-12 y 7-13.

7.6.4. Estimación Bayesiana

Dado que los parámetros de interés tienen un espacio paramétrico acotado en $(0,1)$ se podría asumir una distribución a priori $Beta(\alpha, \beta)$, cuya forma se expresa en la ecuación 6-17.

Por lo que se asumen las siguientes distribuciones a priori para la estimación.

Distribución a priori no informativa:

Asumiendo el método de Laplace se obtuvo una distribución no informativa para realizar la estimación bayesiana, tomando una distribución Beta con $\alpha = \beta = 1$ para cada parámetro de desempeño.

De manera que las distribuciones posteriores obtenidas son:

$$\alpha_1|y \sim Beta(y_1 + y_2 + 1, y_3 + y_4 + 1); \alpha_2|y \sim Beta(y_1 + y_3 + 1, y_2 + y_4 + 1)$$

$$\beta_1|y \sim Beta(y_7 + y_8 + 1, y_5 + y_6 + 1); \beta_2|y \sim Beta(y_6 + y_8 + 1, y_5 + y_7 + 1)$$

Con sus respectivos intervalos de credibilidad:

$$IC_{\alpha_1} = [B(y_1 + y_2 + 1, \delta/2), B(y_3 + y_4 + 1, 1 - \delta/2)]$$

$$IC_{\alpha_2} = [B(y_1 + y_3 + 1, \delta/2), B(y_2 + y_4 + 1, 1 - \delta/2)]$$

$$IC_{\beta_1} = [B(y_7 + y_8 + 1, \delta/2), B(y_5 + y_6 + 1, 1 - \delta/2)]$$

$$IC_{\beta_2} = [B(y_6 + y_8 + 1, \delta/2), B(y_5 + y_7 + 1, 1 - \delta/2)]$$

Donde $1 - \delta$ es el nivel de credibilidad y δ es un cuantil de la distribución posterior beta asociada a cada parámetro de desempeño .

Distribución a priori informativa:

A partir de la información obtenida por Mendez,2017 y Osorio et al., 2010, se tomó como información previa para la estimación de los parámetros de cada algoritmo y la prueba rápida, por tanto, se escoge una distribución Beta con parámetros α y β , se obtuvo las siguientes distribuciones posteriores bajo el proceso descrito en 6-18:

$$\alpha_1|y \sim Beta(y_1 + y_2 + \alpha, y_3 + y_4 + \beta); \beta_1|y \sim Beta(y_7 + y_8 + \alpha, y_5 + y_6 + \beta)$$

$$\alpha_2|y \sim Beta(y_1 + y_3 + \alpha, y_2 + y_4 + \beta); \beta_2|y \sim Beta(y_6 + y_8 + \alpha, y_5 + y_7 + \beta)$$

Para obtener los valores de α y β se realiza la elicitación de los hiperparametros por medio del método de [Tovar, 2012] (Ver tabla 7-7) descrito en la ecuación 6-20.

Tabla 7-7.: Elicitación de hiperparametros para la prueba rápida (Valores tomados de Osorio et al.,2010)

	Limite inferior	Limite superior	θ_0	a	b
Sensibilidad	0.75	0.857	0.8035	885.65	216.59
Especificidad	0.81	0.947	0.8785	398.79	55.15

Donde los límites inferiores y superiores para la sensibilidad y especificidad son tomados de Osorio et al.,2010. En este caso, los intervalos de credibilidad se obtienen de la siguiente manera:

$$IC_{\alpha_1} = [B(y_1 + y_2 + \alpha, \delta/2), B(y_3 + y_4 + \beta, 1 - \delta/2)]$$

$$IC_{\alpha_2} = [B(y_1 + y_3 + \alpha, \delta/2), B(y_2 + y_4 + \beta, 1 - \delta/2)]$$

$$IC_{\beta_1} = [B(y_7 + y_8 + \alpha, \delta/2), B(y_5 + y_6 + \beta, 1 - \delta/2)]$$

$$IC_{\beta_2} = [B(y_6 + y_8 + \alpha, \delta/2), B(y_5 + y_7 + \beta, 1 - \delta/2)]$$

Al igual que lo descrito en la sección anterior, $1 - \delta$ es el nivel de credibilidad y δ es un cuantil de la distribución posterior asociada a cada parámetro de desempeño .

7.6.5. Estimación Bayesiana para escenarios de aplicación en Paralelo y Series

Para la estimación bayesiana de los escenarios propuestos, se opta por usar una distribución Beta no informativa con parámetros $\alpha = \beta = 1$, ya que no se cuenta con información externa para estos escenarios, por tanto la distribución posterior será:

- Paralelo

$$\alpha_3|y \sim Beta(y_1 + y_2 + y_3 + 1, y_4 + 1); \beta_3|y \sim Beta(y_8 + 1, y_5 + y_6 + y_7 + 1).$$

- Serie Negativo

$$\alpha_4|y \sim \text{Beta}(y_1 + y_2 + y_3 + 1, y_4 + 1); \beta_4|y \sim \text{Beta}(y_8 + 1, y_5 + y_6 + y_7 + 1).$$

- Serie Positivo

$$\alpha_5|y \sim \text{Beta}(y_1 + 1, y_2 + y_3 + y_4 + 1); \beta_5|y \sim \text{Beta}(y_6 + y_7 + y_8 + 1, y_5 + 1)$$

Se obtienen las regiones de credibilidad bajo el mismo proceso descrito anteriormente.

7.6.6. Remuestreo Bootstrap: Calculo de parámetros de desempeño

Es desarrollado el remuestreo Bootstrap a la tabla de clasificación presentada en 7-6 con el objetivo de encontrar la convergencia de los parámetros de desempeño con B=2000, teniendo en cuenta los siguientes pasos:

1. Inicialmente, se procedió a remuestrear la matriz de datos obtenida mediante simulación .
2. Se realizó la estimación de los parámetros de desempeño utilizando las formulas 7-8 y 7-11 para cada una de las matrices de datos remuestreadas.
3. Se repitió de forma independiente los pasos 1 y 2 un total de B veces.
4. Se calculó las estimaciones e intervalos de confianza Bootstrap para la sensibilidad y Especificidad.

En el caso de los escenarios para las pruebas aplicadas en serie y paralelo se procedió a realizar la respectiva clasificación de los individuos teniendo en cuenta los eventos del 7-2 al 7-7.

7.7. Comparación de las estimaciones mediante la estadística Clásica y Bayesiana

Uno de los objetivos de este trabajo de grado, es realizar la comparación de las estimaciones de los parámetros de desempeño obtenidas por medio de estadística clásica y bayesiana.

Esta comparación puede realizarse siempre y cuando los parámetros de desempeño de cada proceso, sean calculados usando el mismo conjunto de datos o en este caso la misma base de datos de individuos. Se comparó los intervalos de confianza y las regiones de credibilidad para así dar respuesta a este objetivo.

8. Resultados.

En esta sección se presenta un análisis para los diferentes niveles de prevalencia utilizados, analizando así la cantidad de individuos que presentaron dengue y no dengue, la estimación de los parámetros de desempeño aplicados a los escenarios en serie y paralelo como también la técnica de Bootstrap para observar en que valor se estabilizan las estimaciones realizadas, en los resultados consignados en las diferentes tablas se presentan dos valores, los valores que se encuentran en negrilla corresponden a los cálculos realizados para los algoritmos con hemograma.

8.1. Análisis Descriptivo

El análisis descriptivo para los archivos de datos simulados con el algoritmo Park [Park et al., 1996], es llevado a cabo con la intención de conocer la proporción de pacientes positivos y negativos para diagnóstico de dengue teniendo en cuenta los distintos escenarios en la prevalencia para los algoritmos solos y adicionando el hemograma considerando los resultados de una prueba rápida.

Tabla 8-1.: Proporción de pacientes con resultado positivo y negativo para dengue en algoritmos clasificadores

Prevalencia	Resultados	Algoritmo I	Algoritmo II	Algoritmo III	Algoritmo IV	Prueba rápida
0.1	Positivo	0.4678 (0.4908)	0.3375 (0.3796)	0.5391 (0.4193)	0.2727 (0.3226)	0.3705
	Negativo	0.5322 (0.5092)	0.6625 (0.6204)	0.4609 (0.5807)	0.7273 (0.6774)	0.6295
0.5	Positivo	0.5554 (0.5350)	0.4459 (0.4895)	0.5003 (0.7361)	0.5698 (0.4515)	0.5169
	Negativo	0.4446 (0.4650)	0.5541 (0.5105)	0.4997 (0.2639)	0.4302 (0.5485)	0.4831
0.9	Positivo	0.6093 (0.7368)	0.6420 (0.5839)	0.7230 (0.6656)	0.6282 (0.6188)	0.7530
	Negativo	0.3907 (0.2632)	0.3580 (0.4161)	0.2770 (0.3344)	0, 3718 (0,3812)	0.2470

Los valores en **negrilla** corresponden a proporciones para los algoritmos con hemograma.

Los resultados que muestra la tabla **8-1** se observó que a medida que aumenta el nivel de prevalencia se tiene una mayor cantidad de resultados positivos para cada prototipo clasificador solo y cuando se tiene en cuenta el resultado del hemograma, la prueba rápida aumenta la proporción de positivos cuando aumenta la prevalencia, este primer análisis es relevante por que se esta evidenciando el comportamiento ante una situación de tener pocos casos de dengue, es decir una prevalencia del 0.1 (10%) y pasar a un estado de epidemia al obtener una prevalencia del 0.9 (90%).

El anterior resultado explica lo que realmente se presenta sobre el comportamiento del diagnóstico de una enfermedad como el dengue a medida que el nivel de prevalencia aumenta, es decir, cuando en una población se van generando muchos casos de la enfermedad los 4 prototipos analizados detectarán en mayor medida a pacientes con la enfermedad, lo cual es lógico en el contexto del problema en estudio.

Para ser más específicos, en el caso de una prevalencia del 10% o 90% de la enfermedad del Dengue, los algoritmos que generan la mayor proporción de individuos clasificados como positivo para dengue son el prototipo I y III cuando se tiene en cuenta los resultados del hemograma, en el marco en el que se tenga una cantidad moderada de pacientes con Dengue, es decir, una prevalencia del 0.5(50%), el prototipo III con hemograma y el IV sin hemograma determinaron la mayor proporción de individuos diagnosticados con dengue.

8.1.1. Resultados prueba rápida

Es realizado el proceso de clasificación cuando se tiene la prueba rápida únicamente versus la prueba confirmatoria patrón de oro.

Los resultados determinados en la tabla **8-2** representan las medidas de desempeño de la prueba rápida en diferentes niveles de prevalencia, en la cual se puede observar que la sensibilidad bajo los tres niveles no presenta cambios significativos a diferencia de la especificidad, Se obtienen valores de especificidad de 67.87%, 78.06% y 79.39% para los niveles del 10%, 50% y 90% de

prevalencia respectivamente. Por estimación clásica las medidas de desempeño son más altos, sin embargo la diferencia no es relevante.

Tabla 8-2.: Estimaciones de los parámetros de desempeño para la Prueba Rápida.

Prevalencia	Medida	Métodos de estimación			
		Clásico	LaPlace	Conjugada	Bootstrap
0.1	Sensibilidad 95%	0.8183 <i>IC</i> : $\langle 0.7943; 0.8423 \rangle$	0.8175 <i>RC</i> : $\langle 0.7931; 0.8411 \rangle$	0.8106 <i>RC</i> : $\langle 0.7934; 0.8270 \rangle$	0.8130 <i>IC</i> : $\langle 0.8127; 0.8131 \rangle$
	Especificidad 95%	0.6787 <i>IC</i> : $\langle 0.6691; 0.6884 \rangle$	0.6786 <i>RC</i> : $\langle 0.669; 0.6883 \rangle$	0.6882 <i>RC</i> : $\langle 0.6789; 0.6976 \rangle$	0.7939 <i>IC</i> : $\langle 0.7934; 0.7946 \rangle$
0.5	Sensibilidad 95%	0.8108 <i>IC</i> : $\langle 0.7999; 0.8216 \rangle$	0.8106 <i>RC</i> : $\langle 0.7997; 0.8214 \rangle$	0.8094 <i>RC</i> : $\langle 0.7995; 0.8191 \rangle$	0.8106 <i>IC</i> : $\langle 0.8105; 0.8113 \rangle$
	Especificidad 95%	0.7806 <i>IC</i> : $\langle 0.7691; 0.7921 \rangle$	0.7805 <i>RC</i> : $\langle 0.7689; 0.7919 \rangle$	0.7888 <i>RC</i> : $\langle 0.77790.7996 \rangle$	0.7809 <i>IC</i> : $\langle 0.7807; 0.7812 \rangle$
0.9	Sensibilidad 95%	0.8131 <i>IC</i> : $\langle 0.805; 0.8211 \rangle$	0.813 <i>RC</i> : $\langle 0.8049; 0.82099 \rangle$	0.8121 <i>RC</i> : $\langle 0.80440.8196 \rangle$	0.8130 <i>IC</i> : $\langle 0.8127; 0.8131 \rangle$
	Especificidad 95%	0.7931 <i>IC</i> : $\langle 0.7679; 0.8183 \rangle$	0.7925 <i>RC</i> : $\langle 0.7668; 0.8172 \rangle$	0.8201 <i>RC</i> : $\langle 0.7997; 0.8393 \rangle$	0.7939 <i>IC</i> : $\langle 0.7934; 0.7946 \rangle$

Lo anterior demuestra que el proceso de estimación de los parámetros de desempeño de la prueba rápida condicionada a un patrón de oro desde el enfoque bayesiano y el clásico, no representan diferencias significativas por el hecho de que se obtiene 10000 individuos en la base de datos, es decir, que la verosimilitud impuesta de una distribución binomial por la prueba rápida con respecto a los datos extrae la mayoría de la información comparada con cualquier apriori no informativa o informativa que se considere.

8.2. Resultados obtenidos cuando las pruebas se aplican en Paralelo

Inicialmente para este escenario, es ejecutada la nueva clasificación teniendo en cuenta los eventos planteados anteriormente, examinando la contribución que realiza la prueba rápida a los algoritmos

cuando se aplican en paralelo (simultáneamente) como clasificador final.

8.2.1. Estimación por máxima verosimilitud

El procedimiento bajo la estadística clásica se realizó para cada uno de los algoritmos de clasificación propuestos donde los valores presentados en negrilla son las estimaciones realizadas para los algoritmos con hemograma.

En la tabla **8-3** son presentados los resultados de las estimaciones de los parámetros de desempeño sin hemograma y con hemograma. En el caso de contar con condiciones bajo una prevalencia del 0.1(10%), las sensibilidades obtenidas por los algoritmos de clasificación sin hemograma no varía entre ellos, alcanzando sensibilidades mayores al 90%, las especificidades más altas alcanzadas bajo la misma prevalencia, son notables en los algoritmos II y IV, con estimaciones del 48.62% y 54.95% respectivamente.

En una prevalencia del 50%, las sensibilidades más altas alcanzadas son del 95.66% por el algoritmo I sin hemograma y del 94.83% para el algoritmo IV sin hemograma, las especificidades obtenidas bajo este nivel de prevalencia no presenta gran variación, sin embargo los algoritmos con mayores especificidades fueron el I y II sin hemograma, siendo estas medidas de 54.98% y 57.53% respectivamente.

Para una prevalencia del 90%, las sensibilidades para los cuatro algoritmos sin hemograma, presentan valores por encima del 90% de clasificación, siendo el 94.9% la estimación más alta por el algoritmo III seguido por el algoritmo II con una sensibilidad del 93.5%, las especificidades obtenidas en este grado de prevalencia cambian entre ellas, sin embargo los algoritmos con mayores medidas alcanzadas fueron 68.41% por el algoritmo IV y 54.39% correspondiente al algoritmo II. Se percibe que la especificidad se ve afectada cuando el nivel de prevalencia aumenta, al notar que el proceso de clasificación para diagnosticar dengue es más específico en el momento que el estado de la prevalencia de la enfermedad es más alta.

Tabla 8-3.: Estimación clásica - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en paralelo

Algoritmo	Medida	Prevalencia		
		0.1	0.5	0.9
I	Sensibilidad	0.9344 (0.9203)	0.9566 (0.9292)	0.9304 (0.9549)
	Intervalo (95%)	⟨0.9189; 0.0.9498⟩ (0.9034;0.9371)	⟨0.951; 0.9623⟩ (0.9222;0.9363)	⟨0.9251; 0.9356⟩ (0.9506;0.9592)
	Especificidad	0.3897 (0.3639)	0.5498 (0.4560)	0.3996 (0.6155)
	Intervalo (95%)	⟨0.3796; 0.3997⟩ (0.3540;0.3739)	⟨0.5359; 0.5636⟩ (0.4422;0.4697)	⟨0.3691; 0.4301⟩ (0.5853;0.6458)
II	Sensibilidad	0.9293 (0.9344)	0.9199 (0.9079)	0.935 (0.9139)
	Intervalo (95%)	⟨0.9134; 0.9453⟩ (0.9189;0.9498)	⟨0.9124; 0.9274⟩ (0.899;0.9159)	⟨0.9299; 0.9401⟩ (0.9082;0.9197)
	Especificidad	0.4862(0.4684)	0.5753 (0.5286)	0.5439 (0.671)
	Intervalo (95%)	⟨0.4758; 0.4965⟩ (0.4581;0.4787)	⟨0.5616; 0.5891⟩ (0.5148;0.5425)	⟨0.5129; 0.5749⟩ (0.6417;0.7003)
III	Sensibilidad	0.9243 (0.9606)	0.9272 (0.9161)	0.949 (0.9167)
	Intervalo (95%)	⟨0.9078; 0.9408⟩ (0.9485;0.9727)	⟨0.9201; 0.9344⟩ (0.9084;0.9238)	⟨0.9445; 0.9536⟩ (0.911;0.9225)
	Especificidad	0.3283 (0.4457)	0.5055 (0.2092)	0.5378 (0.5943)
	Intervalo (95%)	⟨0.3186; 0.3380⟩ (0.4355;0.4560)	⟨0.4916; 0.5194⟩ (0.1979;0.2206)	⟨0.5068; 0.5689⟩ (0.5637;0.6249)
IV	Sensibilidad	0.9273 (0.9243)	0.9483 (0.9083)	0.9149 (0.9091)
	Intervalo (95%)	⟨0.9112; 0.9435⟩ (0.9078;0.9408)	⟨0.9422; 0.9544⟩ (0.9004;0.9163)	⟨0.9092; 0.9207⟩ (0.9032;0.915)
	Especificidad	0.5495 (0.4872)	0.5105 (0.5806)	0.6841 (0.5368)
	Intervalo (95%)	⟨0.5393; 0.5598⟩ (0.4769;0.4976)	⟨0.4966; 0.5244⟩ (0.5668;0.5943)	⟨0.6552; 0.7131⟩ (0.5057;0.5678)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

De los algoritmos con el hemograma (valores con **negrilla**) se observa que bajo los tres niveles de prevalencia se alcanzan sensibilidades mayores al 90% y especificidades mayores al 50%, sin embargo bajo una prevalencia del 50%, el prototipo III alcanza una especificidad del 20.92% de individuos que verdaderamente están sanos, se podría pensar que en este algoritmo son reflejados los mayores casos de falsos positivos. Es importante resaltar que las mejores estimaciones alcanzadas fueron con los algoritmos sin hemograma.

En comparación con la tabla de los algoritmos 7-3, es importante mencionar que las sensibilidades estimadas al aplicar las pruebas en paralelo aumenta tanto en los prototipos con hemograma como sin el, siendo el valor de las estimaciones mayores al 90%, es decir, la prueba rápida esta realizando una alta contribución en las sensibilidades de este escenario clasificatorio, sin embargo las especificidades presentan valores entre 50% y 60%, sin presentarse gran diferencia con las estimaciones obtenidas en los algoritmos (ver tabla 7-3).

8.2.2. Estimación Bayesiana

Considerando una distribución previa no informativa, teniendo en cuenta las pruebas aplicadas en paralelo como clasificador final, se obtiene las estimaciones para los parámetros de desempeño después de asumir tres valores distintos de prevalencia que aparecen en la tabla 8-4.

Tabla 8-4.: Estimación bayesiana - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en Paralelo.

Algoritmo	Medida	Prevalencia		
		0.1	0.5	0.9
Algoritmo I	Sensibilidad	0.9335 (0.9194)	0.9565 (0.9291)	0.9303 (0.9548)
	RC (95%)	⟨0.9172; 0.9482⟩ (0.9017 ; 0.9355)	⟨0.9507; 0.9619⟩(0.9218 ; 0.9360)	⟨0.9250; 0.9355⟩ (0.9505 ; 0.9590)
	Especificidad	0.3898 (0.3640)	0.5498 (0.4561)	0.3997 (0.6151)
	RC (95%)	⟨0.3797; 0.3998⟩ (0.3541 ; 0.3740)	⟨0.5359; 0.5636⟩ (0.4422 ; 0.4699)	⟨0.3695; 0.4304⟩ (0.5848 ; 0.6453)
Algoritmo II	Sensibilidad	0.9285 (0.9336)	0.9197 (0.9078)	0.9350 (0.9139)
	RC (95%)	⟨0.9117; 0.9437⟩ (0.9172 ; 0.9482)	⟨0.9121; 0.9271⟩(0.8997 ; 0.9156)	⟨0.9298; 0.9400⟩ (0.9080 ; 0.9196)
	Especificidad	0.4863 (0.4684)	0.5753 (0.5285)	0.5437 (0.6707)
	RC (95%)	⟨0.4759; 0.4965⟩ (0.4581 ; 0.4787)	⟨0.5616; 0.5890⟩ (0.5148 ; 0.5425)	⟨0.5128; 0.5747⟩ (0.6412 ; 0.6996)
Algoritmo III	Sensibilidad	0.9235 (0.9598)	0.9271(0.9159)	0.9489(0.9167)
	RC (95%)	⟨0.9061; 0.9392⟩ (0.9466 ; 0.9710)	⟨0.9197; 0.9341⟩ (0.9081 ; 0.9235)	⟨0.9443; 0.9534⟩ (0.9109 ; 0.9223)
	Especificidad	0.3283 (0.4458)	0.5055 (0.2095)	0.5378(0.5941)
	RC (95%)	⟨0.3187; 0.3381⟩ (0.4355 ; 0.4561)	⟨0.4916; 0.5194⟩ (0.1982 ; 0.2208)	⟨0.5067; 0.5687⟩ (0.5635 ; 0.6245)
Algoritmo IV	Sensibilidad	0.9265(0.9233)	0.9482 (0.9082)	0.9149 (0.9090)
	RC (95%)	⟨0.9095; 0.9419⟩ (0.9061 ; 0.9392)	⟨0.9419; 0.9541⟩ (0.9001 ; 0.9160)	⟨0.9090; 0.9206⟩ (0.9030 ; 0.9149)
	Especificidad	0.5496 (0.4873)	0.5105 (0.5805)	0.6838 (0.5367)
	RC (95%)	⟨0.5393; 0.5598⟩ (0.4769 ; 0.4976)	⟨0.4967; 0.5245⟩ (0.5668 ; 0.5942)	⟨0.6545; 0.7123⟩ (0.5057 ; 0.5677)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

Con base en los valores reportados en la tabla 8-4, los resultados de las estimaciones no varían mucho en comparación a los presentados en la tabla 8-3. Se obtienen resultados mayores al 90% en las sensibilidades y especificidades bajas considerando todos los niveles de prevalencia, al igual que en el escenario clásico.

En lo que corresponde a las regiones de credibilidad asociadas a las estimaciones de los parámetros de desempeño, en algunos casos su longitud es mas pequeña en comparación a los intervalos de confianza presentados en la tabla 8-3

8.2.3. Estimación usando métodos de remuestreo

El remuestreo realizado para este escenario de aplicación del procedimiento en paralelo como clasificador final, contó con un total de $B=2000$ remuestras, dado que es en 2000 que se alcanza la convergencia de las estimaciones, en cada uno de ellos se realiza el calculo de sensibilidad y especificidad, con el objetivo de observar la convergencia de las estimaciones.

Tabla 8-5.: Estimación Bootstrap - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en paralelo.

Algoritmo	Medida	Prevalencia		
		0.1	0.5	0.9
I	Sensibilidad	0.9342 (0.9199)	0.9567 (0.9291)	0.9303 (0.9549)
	Intervalo (95%)	⟨0.9337; 0.9344⟩ (0.9193 ; 0.9202)	⟨0.9566; 0.9570⟩ (0.9291; 0.9293)	⟨0.9302; 0.9305⟩ (0.9549 ; 0.9550)
	Especificidad	0.3897 (0.3640)	0.5499 (0.4562)	0.3999 (0.6157)
	Intervalo (95%)	⟨0.3894; 0.3898⟩ (0.3638 ; 0.3642)	⟨0.5492; 0.5502⟩ (0.4559; 0.4567)	⟨0.3991; 0.4035⟩ (0.6151 ; 0.6172)
II	Sensibilidad	0.9294(0.9340)	0.9198 (0.9079)	0.9350 (0.9139)
	Intervalo (95%)	⟨0.9289; 0.9296⟩ (0.9331 ; 0.9342)	⟨0.9196; 0.9199⟩ (0.9078 ; 0.9081)	⟨0.9346; 0.9351⟩ (0.9135 ; 0.9140)
	Especificidad	0.4863 (0.4685)	0.5753 (0.5287)	0.5454 (0.6718)
	Intervalo (95%)	⟨0.4861; 0.4867⟩ (0.4683 ; 0.4689)	⟨0.5747; 0.5755⟩ (0.5280 ; 0.5288)	⟨0.5448; 0.5482⟩ (0.6713 ; 0.6728)
III	Sensibilidad	0.9241 (0.9604)	0.9271 (0.9160)	0.9490 (0.9168)
	Intervalo (95%)	⟨0.9237; 0.9253⟩ (0.9601 ; 0.9607)	⟨0.9269; 0.9273⟩ (0.9159 ; 0.9163)	⟨0.9487; 0.9491⟩ (0.9167 ; 0.9170)
	Especificidad	0.3282 (0.4454)	0.5062 (0.2092)	0.5376 (0.5955)
	Intervalo (95%)	⟨0.3276; 0.3284⟩ (0.4450 ; 0.4456)	⟨0.5058; 0.5067⟩ (0.2088 ; 0.2093)	⟨0.5359; 0.5381⟩ (0.5947 ; 0.5976)
IV	Sensibilidad	0.9273 (0.9240)	0.9484 (0.9083)	0.9150 (0.9081)
	Intervalo (95%)	⟨0.9268; 0.9280⟩ (0.9225 ; 0.9242)	⟨0.9483; 0.9488⟩ (0.9082 ; 0.9090)	⟨0.9147; 0.9152⟩ (0.9090 ; 0.9092)
	Especificidad	0.5496 (0.4872)	0.5104(5809)	0.6844 (0.5375)
	Intervalo (95%)	⟨0.5495; 0.5501⟩ (0.4869 ; 0.4875)	⟨0.5095; 0.5107⟩ (0.5807 ; 0.5813)	⟨0.6840; 0.6854⟩ (0.5368 ; 0.5401)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

Las estimaciones obtenidas en la tabla 8-5 difieren de una manera leve a comparación de las estimaciones presentadas en las tablas 8-3 y 8-4, si bien en algunos prototipos los valores son muy parecidos y se diferencian por muy poco, tal es el caso del prototipo II, con un nivel de prevalencia del 90% presentan los mismos valores sin y con hemograma tanto en el remuestreo como en la estimación clásica. Es importante resaltar que los intervalos de confianza calculados por la metodología bootstrap determinan una longitud más alta comparados a los intervalos realizados por estimación clásica, esto podría indicar que al desarrollar esta herramienta estadística aumenta

la variabilidad de las estimaciones por el remuestreo. El prototipo I sin hemograma alcanza una mayor sensibilidad promedio dada la prevalencia del 10% y 50% en contraste con los otros algoritmos, siendo estas del 93.42% y 95.67%, mientras en el marco en el que se considere el resultado con hemograma, el prototipo I logra una mayor sensibilidad promedio en los niveles de prevalencia del 50% y del 90%. La especificidad mayor promedio fue conseguida por el prototipo II sin hemograma, sin embargo en la condición de tener una prevalencia del 90% es más alta en el prototipo IV sin hemograma con 68.4%, por otra parte al analizar los resultados con hemograma no presentan cambios en las estimaciones de la especificidad. Por último, cabe mencionar que la estabilidad de los parámetros estimados se alcanza con un valor de 2000 repeticiones.

8.3. Resultados obtenidos cuando las pruebas se aplican en Serie Negativo

Este escenario propuesto para el análisis, se lleva a cabo teniendo en cuenta los eventos ya planteados en 7-4 y 7-5, los cuales se le realiza las estimaciones bajo los tres niveles de prevalencia simulados.

8.3.1. Estimación por máxima verosimilitud

El procedimiento bajo la estadística clásica se realiza para cada uno de los algoritmos de clasificación propuestos donde los valores presentados en negrilla son las estimaciones realizadas para los algoritmos con hemograma.

En la tabla **8-6** se presentan las medidas estimadas para el escenario de aplicación serie negativo, se percibe que los valores contenidos en la tabla **8-3** son exactamente los mismos, debido a lo cual el proceso de aplicación en serie negativo es igual al realizar un escenario de aplicación en paralelo, teniendo en cuenta que los resultados son equivalentes, de tal modo, las estimaciones en las técnicas bootstrap y bayesiana presentan el mismo efecto y por esta razón se omiten los resultados para este contexto.

Tabla 8-6.: Estimación clásica - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en serie negativo.

Algoritmo	Medida	Prevalencia		
		0.1	0.5	0.9
I	Sensibilidad	0.9344 (0.9203)	0.9566 (0.9292)	0.9304 (0.9549)
	Intervalo (95%)	⟨0.9189; 0.0.9498⟩ (0.9034;0.9371)	⟨0.951; 0.9623⟩ (0.9222;0.9363)	⟨0.9251; 0.9356⟩ (0.9506;0.9592)
	Especificidad	0.3897 (0.3639)	0.5498 (0.4560)	0.3996 (0.6155)
	Intervalo (95%)	⟨0.3796; 0.3997⟩ (0.3540;0.3739)	⟨0.5359; 0.5636⟩ (0.4422;0.4697)	⟨0.3691; 0.4301⟩ (0.5853;0.6458)
II	Sensibilidad	0.9293 (0.9344)	0.9199 (0.9079)	0.935 (0.9139)
	Intervalo (95%)	⟨0.9134; 0.9453⟩ (0.9189;0.9498)	⟨0.9124; 0.9274⟩ (0.899;0.9159)	⟨0.9299; 0.9401⟩ (0.9082;0.9197)
	Especificidad	0.4862(0.4684)	0.5753 (0.5286)	0.5439 (0.671)
	Intervalo (95%)	⟨0.4758; 0.4965⟩ (0.4581;0.4787)	⟨0.5616; 0.5891⟩ (0.5148;0.5425)	⟨0.5129; 0.5749⟩ (0.6417;0.7003)
III	Sensibilidad	0.9243 (0.9606)	0.9272 (0.9161)	0.949 (0.9167)
	Intervalo (95%)	⟨0.9078; 0.9408⟩ (0.9485;0.9727)	⟨0.9201; 0.9344⟩ (0.9084;0.9238)	⟨0.9445; 0.9536⟩ (0.911;0.9225)
	Especificidad	0.3283 (0.4457)	0.5055 (0.2092)	0.5378 (0.5943)
	Intervalo (95%)	⟨0.3186; 0.3380⟩ (0.4355;0.4560)	⟨0.4916; 0.5194⟩ (0.1979;0.2206)	⟨0.5068; 0.5689⟩ (0.5637;0.6249)
IV	Sensibilidad	0.9273 (0.9243)	0.9483 (0.9083)	0.9149 (0.9091)
	Intervalo (95%)	⟨0.9112; 0.9435⟩ (0.9078;0.9408)	⟨0.9422; 0.9544⟩ (0.9004;0.9163)	⟨0.9092; 0.9207⟩ (0.9032;0.915)
	Especificidad	0.5495 (0.4872)	0.5105 (0.5806)	0.6841 (0.5368)
	Intervalo (95%)	⟨0.5393; 0.5598⟩ (0.4769;0.4976)	⟨0.4966; 0.5244⟩ (0.5668;0.5943)	⟨0.6552; 0.7131⟩ (0.5057;0.5678)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

El algoritmo IV con hemograma y sin hemograma presenta valores de sensibilidades y especificidades que pueden considerarse buenos en la aplicación, ya que bajo los tres niveles de prevalencia alcanza sensibilidades mayores al 90% y especificidades mayores al 50% con y sin hemograma, indicando así que este realiza una mejor clasificación de individuos bajo este escenario clínico.

En este escenario se observa una situación peculiar al tener valores altos de sensibilidades pero bajos en especificidades, esto se debe a que el objetivo principal en este marco es aumentar la sensibilidad pero obteniendo mayores resultados de falsos positivos, esto conlleva en sí a un aumento en costos en el trabajo de campo si se aplica el estudio ante una situación real, si bien el gasto en personas que no tienen dengue pero han sido clasificadas como portadoras de la enfermedad es alto, por lo tanto al aplicar este escenario es necesario sacrificar el costo y aumentar la inversión desde el punto de vista de la salud.

8.4. Resultados obtenidos cuando las pruebas se aplican en Serie Positivo

El escenario en serie positivo busca obtener una mayor especificidad a la hora de hacer la clasificación de los algoritmos sin hemograma y con hemograma donde el objetivo primordial es analizar la contribución de la prueba rápida a cada prototipo, se realizó el mismo procedimiento con los escenarios anteriores, obteniendo las estimaciones por estadística clásica, bayesiana y la metodología de bootstrap.

8.4.1. Estimación por máxima verosimilitud

Inicialmente al igual que en los escenarios anteriores, se procede a la estimación de los parámetros de desempeño bajo la estadística clásica donde sus resultados están consignados en la tabla 8-7.

Tabla 8-7.: Estimación clásica - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en serie positivo.

Algoritmo	Medida	Prevalencia		
		0.1	0.5	0.9
Algoritmo I	Sensibilidad	0.5913 (0.559)	0.6501 (0.5237)	0.5047 (0.649)
	Intervalo (95%)	⟨0.5607; 0.6219⟩ (0.5281;0.5899)	⟨0.6369; 0.6633⟩ (0.5099;0.5375)	⟨0.4944; 0.515⟩ (0.6391;0.6588)
	Especificidad	0.8476 (0.8423)	0.9191 (0.8982)	0.9001 (0.9324)
	Intervalo (95%)	⟨0.8402; 0.855⟩ (0.8352;0.8502)	⟨0.9115; 0.9267⟩ (0.8897;0.9066)	⟨0.8814; 0.9188⟩ (0.9167;0.948)
Algoritmo II	Sensibilidad	0.5469 (0.6196)	0.5106 (0.5333)	0.553 (0.5292)
	Intervalo (95%)	⟨0.5159; 0.5779⟩ (0.5893;0.6498)	⟨0.4968; 0.5244⟩ (0.5195;0.5471)	⟨0.542; 0.5632⟩ (0.5189;0.5395)
	Especificidad	0.8903 (0.8699)	0.9354 (0.9052)	0.9072 (0.9586)
	Intervalo (95%)	⟨0.8839; 0.8968⟩ (0.8629;0.8768)	⟨0.9286; 0.9422⟩ (0.897;0.913)	⟨0.8891; 0.9252⟩ (0.9462;0.971)
Algoritmo III	Sensibilidad	0.6004 (0.6196)	0.5199 (0.6338)	0.6294 (0.6059)
	Intervalo (95%)	⟨0.5699; 0.6309⟩ (0.5893;0.6498)	⟨0.5062; 0.5338⟩ (0.6205;0.6471)	⟨0.6194; 0.6393⟩ (0.5958;0.616)
	Especificidad	0.8297 (0.8514)	0.9126 (0.8383)	0.9173 (0.9334)
	Intervalo (95%)	⟨0.8219; 0.8375⟩ (0.8440;0.8587)	⟨0.9048; 0.9205⟩ (0.8282;0.8486)	⟨0.9001; 0.9344⟩ (0.9178;0.9489)
Algoritmo IV	Sensibilidad	0.5247 (0.5318)	0.6086 (0.5152)	0.5767 (0.5557)
	Intervalo (95%)	⟨0.4936; 0.5558⟩ (0.5007;0.5628)	⟨0.5951; 0.6221⟩ (0.5014;0.529)	⟨0.5665; 0.5869⟩ (0.5455;0.566)
	Especificidad	0.8962 (0.9035)	0.8788 (0.9118)	0.9395 (0.9374)
	Intervalo (95%)	⟨0.8899; 0.9025⟩ (0.8974;0.9096)	⟨0.8697; 0.8879⟩ (0.9039;0.9197)	⟨0.9246; 0.9543⟩ (0.9224;0.9525)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

A partir de las estimaciones alcanzadas en la tabla **8-7**, las sensibilidades obtenidas teniendo en cuenta los tres valores de prevalencia son mayores al 50%, es decir, considerando este procedimiento se realiza una clasificación correcta mayor del 50% de los individuos que realmente están enfermos, mientras que las especificidades en los niveles de prevalencia del 10% y 50% son mayores del 80% indicando que en este contexto se clasifican bien los individuos que están realmente sanos, mientras que con una prevalencia del 90% se logran especificidades mayores al 90%, por esta razón, cuando hay un escenario endémico de la enfermedad se presenta alta capacidad de clasificar a los individuos verdaderamente sanos.

Teniendo en cuenta los prototipos con hemograma, algunas estimaciones bajan en comparación a cuando no se tiene presente los resultados del hemograma, sin embargo a un nivel de prevalencia del 90% la mayoría de las especificidades aumentan cuando se realiza el proceso de clasificación con lo resultados del hemograma.

Es evidente que en este escenario se mejora las especificidades de los algoritmos comparado con la tabla **7-3**, donde los algoritmos presentan una especificidad de entre 60% y 80%, a pesar de ello las sensibilidades obtenidas realizando el proceso de aplicación en serie positivo están entre 50% y 60% en los algoritmos sin y con hemograma, por ello la prueba rápida en este escenario aporta más a la especificidad y no con gran impacto a la sensibilidad, dado que no varía mucho respecto a los valores de los algoritmos en la tabla **7-3**.

8.4.2. Estimación Bayesiana

Se realiza la estimación de los parámetros de desempeño con base la estadística bayesiana, estos resultados se encuentran en la tabla **8-8**.

A partir de la tabla **8-8** se obtiene el mismo comportamiento que en el paradigma clásico representado en la tabla **8-7**, donde a una prevalencia del 90% se logra un mayor nivel de especificidad comparado con los escenarios anteriormente estudiados.

Tabla 8-8.: Estimación bayesiana - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en serie positivo.

Algoritmo	Medida	Prevalencia		
		0.1	0.5	0.9
Algoritmo I	Sensibilidad	0.5910 (0.5589)	0.6500 (0.5237)	0.5047 (0.6490)
	RC (95%)	⟨0.5604; 0.6215⟩ (0.5279; 0.5897)	⟨0.6369; 0.6632⟩ (0.5099 ; 0.5375)	⟨0.4944; 0.5150⟩ (0.6391 ; 0.6588)
	Especificidad	0.8475 (0.8426)	0.9189 (0.8980)	0.8993 (0.9315)
	RC (95%)	⟨0.8400; 0.8549⟩ (0.8350 ; 0.8501)	⟨0.9112; 0.9264⟩ (0.8894 ; 0.9063)	⟨0.8798; 0.9172⟩ (0.9150 ; 0.9464)
Algoritmo II	Sensibilidad	0.5468 (0.6196)	0.5106 (0.5333)	0.5529 (0.5293)
	RC (95%)	⟨0.5158; 0.5777⟩ (0.5889 ; 0.6493)	⟨0.4968; 0.5244⟩ (0.5195 ; 0.5470)	⟨0.5427; 0.5632⟩ (0.5189)
	Especificidad	0.8902 (0.8698)	0.9352 (0.9050)	0.9063 (0.9578)
	RC (95%)	⟨0.8837; 0.8966⟩ (0.8628 ; 0.8767)	⟨0.9282; 0.9419⟩ (0.8967 ; 0.9130)	⟨0.8984; 0.9328⟩ (0.9161 ; 0.9473)
Algoritmo III	Sensibilidad	0.6000 (0.6194)	0.5200 (0.6338)	0.6293 (0.6058)
	RC (95%)	⟨0.5696; 0.6305⟩ (0.5889 ; 0.6493)	⟨0.4968; 0.5244⟩ (0.5195 ; 0.5470)	⟨0.6193; 0.6393⟩ (0.5958 ; 0.6160)
	Especificidad	0.8297 (0.8513)	0.9124 (0.8383)	0.9164 (0.9326)
	RC (95%)	⟨0.8218; 0.8373⟩ (0.8439 ; 0.8586)	⟨0.9045; 0.9202⟩ (0.8279 ; 0.8484)	⟨0.8984; 0.9328⟩ (0.9161 ; 0.9473)
Algoritmo IV	Sensibilidad	0.5244 (0.5321)	0.6086 (0.5152)	0.5767 (0.5558)
	RC (95%)	⟨0.4936; 0.5557⟩ (0.5006; 0.5627)	⟨0.5951; 0.6220⟩ (0.5014 ; 0.5290)	⟨0.5665; 0.5869⟩ (0.5455 ; 0.5660)
	Especificidad	0.8961 (0.9034)	0.8787 (0.9117)	0.9386 (0.9365)
	RC (95%)	⟨0.8897; 0.9023⟩ (0.8973 ; 0.9094)	⟨0.8695; 0.8876⟩ (0.9036 ; 0.9194)	⟨0.9228; 0.9526⟩ (0.9205 ; 0.9508)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

Es relevante notar que este escenario muestra una gran especificidad pero una sensibilidad aproximadamente de 50%, de tal forma que en este contexto se obtiene una cantidad de falsos negativos alta, lo cual indica que puede existir un mal procedimiento en el diagnóstico aunque se esten disminuyendo los costos, dado que si la sensibilidad es baja afecta a la población por los casos que verdaderamente son enfermos de dengue.

Además, es importante observar que las diferencias entre el paradigma clásico y bayesiano no son significativas, ya que las bases simuladas fueron de 10.000 individuos indicando así que los tres niveles de prevalencia proporcionan una cantidad significativa de casos de dengue, ya que la estadística bayesiana resulta potente y útil cuando se obtienen pocos casos.

8.4.3. Estimación usando métodos de Remuestreo.

Al igual que los escenarios anteriormente presentados, es realizado el remuestro bootstrap 2000 veces, valor en el cual se logra la convergencia de los parámetros estimados (ver tabla 8-9).

Tabla 8-9.: Estimación Bootstrap - Estimación de los parámetros de desempeño obtenidos cuando las pruebas se aplican en serie positivo.

Algoritmo	Medida	Prevalencia		
		0.1	0.5	0.9
I	Sensibilidad	0.5913 (0.5588)	0.6501 (0.5237)	0.5048 (0.6491)
	Intervalo (95%)	⟨0.5907; 0.5923⟩ (0.5580 ; 0.5592)	⟨0.6499; 0.6507⟩ (0.5234 ; 0.5240)	⟨0.5046; 0.5052⟩ (0.6489 ; 0.6494)
	Especificidad	0.8475 (0.8427)	0.9193 (0.8984)	0.9004 (0.9328)
	Intervalo (95%)	⟨0.8473; 0.8476⟩ (0.8424 ; 0.8430)	⟨0.9192; 0.9295⟩ (0.8982 ; 0.8986)	⟨0.9192; 0.9295⟩ (0.8982 ; 0.8986)
II	Sensibilidad	0.5460(0.6190)	0.5102 (0.5332)	0.5527 (0.5294)
	Intervalo (95%)	⟨0.5447; 0.5467⟩ (0.6175 ; 0.6194)	⟨0.5100; 0.5106⟩ (0.5330 ; 0.5335)	⟨0.5517; 0.5529⟩ (0.5291 ; 0.5297)
	Especificidad	0.8904 (0.8699)	0.9355 (0.9052)	0.9075 (0.9588)
	Intervalo (95%)	⟨0.8903; 0.8906⟩ (0.8698 ; 0.8701)	⟨0.9353; 0.9356⟩ (0.9051 ; 0.9054)	⟨0.9072; 0.9084⟩ (0.9587 ; 0.9593)
III	Sensibilidad	0.60 (0.6192)	0.5199 (0.6340)	0.6293 (0.6061)
	Intervalo (95%)	⟨0.599; 0.6004⟩ (0.6187 ; 0.6203)	⟨0.5197; 0.5207⟩ (0.6339 ; 0.6350)	⟨0.6292; 0.6296⟩ (0.6055 ; 0.6062)
	Especificidad	0.8296 (0.8512)	0.9129 (0.8386)	0.9174 (0.9340)
	Intervalo (95%)	⟨0.8291; 0.8298⟩ (0.8511 ; 0.8514)	⟨0.9126; 0.9132⟩ (0.8385 ; 0.8389)	⟨0.9172; 0.9183⟩ (0.9335 ; 0.9349)
IV	Sensibilidad	0.5246 (0.5315)	0.6087 (0.5153)	0.5765 (0.5555)
	Intervalo (95%)	⟨0.5238; 0.5251⟩ (0.5308;0.5327)	⟨0.6085; 0.6098⟩ (0.5150;0.5158)	⟨0.5755; 0.5767⟩ (0.5458;0.5557)
	Especificidad	0.8960 (0.9036)	0.8788 (0.9119)	0.9400 (0.9375)
	Intervalo (95%)	⟨0.8960; 0.8963⟩ (0.9034;0.9036)	⟨0.8788; 0.8793⟩ (0.9117;0.9121)	⟨0.9397; 0.9414⟩ (0.9373;0.9381)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

A partir de las estimaciones obtenidas en la tabla 8-9 por la técnica de bootstrap, donde su función es estudiar la convergencia de los parámetros de desempeño en cada proceso de aplicación cuando se tiene un archivo de datos diferente al simulado. Específicamente en los tres valores de prevalencia, se puede observar que las sensibilidades promedios se encuentran entre el 50% y 65%, siendo parecidas a las alcanzadas por el paradigma clásico y bayesiano (Ver tablas 8-7 y 8-8), a su vez, las especificidades obtenidas bajo esta técnica logran estimaciones promedios entre el 80% y 94%, en general el prototipo que alcanza sensibilidades promedios mayores bajo las 2000 repeticiones es el prototipo III sin hemograma y con hemograma en los tres escenarios de prevalencia, por otro lado, el prototipo IV sin hemograma y con hemograma alcanza especificidades promedios mayores en

comparación con los otros prototipos bajo los tres niveles de prevalencia realizados.

8.5. Estimación de los parámetros de desempeño bajo diferentes niveles de sensibilidad de la prueba rápida.

Puesto que los objetivos de este trabajo de grado es estudiar la contribución que realiza la prueba rápida a los algoritmos de clasificación propuestos, por lo cual, se realiza el proceso bajo tres niveles de sensibilidad de la prueba rápida, respectivamente del 75%, 80.7% y 85.7%, que corresponden a los valores mínimo y máximo del intervalo de esta prueba, dado que los resultados anteriores indican que no hay cambios significativos entre los valores de prevalencia aplicados es realizado el siguiente análisis con un nivel de prevalencia del 10%.

Tabla 8-10.: Medidas de desempeño bajo diferentes sensibilidades de la prueba rápida escenario de aplicación paralelo.

Medida	Valores			
	Algoritmo I	Algoritmo II	Algoritmo III	Algoritmo IV
Sensibilidad del 75%				
Sensibilidad	0.936 (0.928)	0.921 (0.947)	0.935 (0.964)	0.911 (0.919)
Intervalo (95%)	⟨0.919; 0.951⟩ (0.91;0.944)	⟨0.903; 0.937⟩ (0.931;0.96)	⟨0.918; 0.95⟩ (0.95;0.974)	⟨0.892; 0.928⟩ (0.901;0.935)
Especificidad	0.369 (0.346)	0.464 (0.44)	0.318 (0.413)	0.511 (0.473)
Intervalo (95%)	⟨0.359; 0.379⟩ (0.336;0.356)	⟨0.453; 0.474⟩ (0.43;0.45)	⟨0.309; 0.328⟩ (0.403;0.424)	⟨0.50; 0.521⟩ (0.462;0.483)
Sensibilidad del 85.7%				
Sensibilidad	0.963 (0.95)	0.957 (0.968)	0.963 (0.969)	0.941 (0.957)
Intervalo (95%)	⟨0.949; 0.974⟩ (0.934;0.962)	⟨0.942; 0.968⟩ (0.955;0.978)	⟨0.949; 0.974⟩ (0.956;0.979)	⟨0.925; 0.955⟩ (0.942;0.968)
Especificidad	0.404 (0.381)	0.506 (0.477)	0.344 (0.446)	0.553 (0.517)
Intervalo (95%)	⟨0.393; 0.414⟩ (0.371;0.391)	⟨0.495; 0.516⟩ (0.467;0.488)	⟨0.334; 0.354⟩ (0.436;0.456)	⟨0.543; 0.563⟩ (0.507;0.528)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

A partir de la tabla **8-10** se encuentran los valores estimados en diferentes niveles de sensibilidad de la prueba rápida, de la cual se puede observar que las sensibilidades aumentan en la medida que cambia la sensibilidad de la prueba rápida, además se mejora cuando se aplican los algoritmos con el hemograma, sin embargo la especificidad disminuye en comparación con los resultados

obtenidos en **8-3** correspondiente a un nivel de sensibilidad del 80.7% de la prueba rápida, por lo tanto, bajo los diferentes niveles de sensibilidad de esta prueba los valores en el escenario de aplicación en paralelo tienen un buen desempeño en la sensibilidad y la estimación de la especificidad aproximadamente es del 30% a 50%, en el caso de presentarse una sensibilidad del 85.7% los valores estimados aumentan sin embargo no es muy considerable el cambio.

Tabla 8-11.: Medidas de desempeño bajo diferentes sensibilidades de la prueba rápida escenario de aplicación serie positivo.

Medida	Valores			
	Algoritmo I	Algoritmo II	Algoritmo III	Algoritmo IV
Sensibilidad del 75%				
Sensibilidad	0.549 (0.511)	0.515(0.567)	0.549(0.576)	0.501 (0.496)
Intervalo (95%)	⟨0.517; 0.58⟩ (0.479;0.542)	⟨0.483; 0.546⟩ (0.536;0.598)	⟨0.517; 0.58⟩ (0.545;0.607)	⟨0.469; 0.532⟩ (0.465;0.528)
Especificidad	0.855 (0.848)	0.90 (0.885)	0.827 (0.87)	0.922 (0.905)
Intervalo (95%)	⟨0.847; 0.862⟩ (0.84;0.855)	⟨0.893; 0.906⟩ (0.878;0.892)	⟨0.819; 0.834⟩ (0.863;0.877)	⟨0.393; 0.414⟩ (0.899;0.911)
Sensibilidad del 85.7%				
Sensibilidad	0.599(0.566)	0.556(0.623)	0.598(0.648)	0.547(0.536)
Intervalo (95%)	⟨0.568; 0.63⟩ (0.535;0.597)	⟨0.524; 0.587⟩ (0.592;0.653)	⟨0.567; 0.629⟩ (0.617;0.678)	⟨0.515; 0.578⟩ (0.504;0.567)
Especificidad	0.877(0.869)	0.914(0.905)	0.857(0.894)	0.936(0.917)
Intervalo (95%)	⟨0.87; 0.884⟩ (0.535;0.597)	⟨0.908; 0.92⟩ (0.898;0.911)	⟨0.85; 0.864⟩ (0.888;0.901)	⟨0.931; 0.041⟩ (0.911;0.923)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

Ahora bien, para el escenario de aplicación serie positivo los resultados se encuentran en la tabla **8-11**, teniendo en cuenta los niveles de sensibilidad de la prueba rápida, las especificidades aumenta equivalente al comportamiento de la tabla **8-7**, en contexto, al presentar una sensibilidad de la prueba del 85.7% se logra alcanzar sensibilidad mayor al 60% sea el caso del prototipo III con hemograma con un valor calculado del 64.8%, desde un análisis global en este caso aplicado, al aumentar la sensibilidad de la prueba rápida incrementa los valores estimados tanto en los algoritmos sin y con hemograma.

8.6. Comparación de bases generadas desde las categorías y el resultados de los algoritmos.

Parte de la metodología del trabajo de grado, es analizar las diferencias que se puedan presentar entre las bases simuladas a partir de las categorías de signos y síntomas, en la cual se aplica la probabilidad que tiene cada categoría para clasificar a un individuo con dengue dado el punto de corte establecido por [Mendez, 2017] y la base de datos simulada a partir de los resultados de los algoritmos de clasificación, por esta razón analizamos como influye el proceso de simulación en los escenarios de aplicación en serie Positivo y paralelo.

Las bases de datos están simuladas aplicando un nivel de prevalencia del 38.83%, se obtiene las proporciones iniciales de la clasificación realizada por cada base en la tabla **8-12**. Posteriormente se hace la comparación para los algoritmos I y III puesto que las categorías de signos y síntomas son el insumo para diagnosticar dengue en estos prototipos.

Tabla 8-12.: Proporciones y estimaciones de las bases simuladas

Medida	Valores			
	Proporciones		Medidas de desempeño	
	Positivo	Negativo	Sensibilidad	Especificidad
Base de datos simulada a partir de las categorías de signos y síntomas				
Algoritmo I	0.6421 (0.5759)	0.3579 (0.4241)	0.793 (0.7053)	0.4538 (0.5062)
Algoritmo III	0.4642 (0.7366)	0.5358 (0.2634)	0.6443 (0.8266)	0.6502 (0.3206)
Base de datos simulada a partir de los resultados de los algoritmos				
Algoritmo I	0.4934 (0.5094)	0.5066 (0.54906)	0.62863(0.6201)	0.5924 (0.5608)
Algoritmo III	0.5467 (0.5395)	0.4533 (0.4605)	0.6368 (0.7476)	0.5105 (0.5926)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

En los resultados de las bases simuladas a partir de las categorías y los resultados de los algoritmos, indica que cuando se realiza la simulación con las categorías se obtiene una estimación menor en las especificidades, sin embargo son valores que se encuentran dentro del intervalo planteado en los prototipos, cuando se realiza la simulación bajo los resultados directos de los algoritmos se observa que cuando se agrega los resultados del hemograma aumentan sus medidas de desempeño, el algo-

8.6 Comparación de bases generadas desde las categorías y el resultados de los algoritmos. 59

ritmo I sin hemograma bajo la primera simulación tiene medidas de desempeño de 79.3% y 45.38% de sensibilidad y especificidad respectivamente, mientras que cuando se realiza la simulación directamente de los resultados de los algoritmos este mismo prototipo I alcanza una sensibilidad del 62.863% y especificidad del 59.24%. Se procede a realizar la estimación de los parámetros de desempeño de los escenarios propuestos, estos valores se resumen en la tabla **8-13**, sin embargo se observa que no hay diferencias significativas entre las estimaciones de las bases.

Tabla 8-13.: Medidas de desempeño escenarios en Serie Negativo y Serie Positivo de las bases simuladas a partir de las categorías y los resultados de los algoritmos.

Medida	Valores			
	Serie Clínico Negativo		Serie Clínico Positivo	
	Algoritmo I	Algoritmo III	Algoritmo I	Algoritmo III
Base de datos simulada a partir de las categorías de signos y síntomas				
Sensibilidad	0.873 (0.818)	0.776 (0.891)	0.289 (0.256)	0.237 (0.304)
Intervalo (95%)	⟨0.862; 0.884⟩ (0.806;0.83)	⟨0.763; 0.789⟩ (0.881;0.901)	⟨0.274; 0.303⟩ (0.242;0.27)	⟨0.224; 0.251⟩ (0.29;0.319)
Especificidad	0.286 (0.319)	0.408 (0.197)	0.796 (0.816)	0.871 (0.752)
Intervalo (95%)	⟨0.275; 0.298⟩ (0.307;0.331)	⟨0.396; 0.42⟩ (0.187;0.207)	⟨0.786; 0.806⟩ (0.806;0.825)	⟨0.862; 0.879⟩ (0.741;0.763)
Base de datos simulada a partir de los resultados de los algoritmos				
Sensibilidad	0.742 (0.742)	0.751 (0.827)	0.204 (0.196)	0.204 (0.239)
Intervalo (95%)	⟨0.728; 0.756⟩ (0.728;0.756)	⟨0.737; 0.764⟩ (0.536;0.598)	⟨0.192; 0.218⟩ (0.184;0.209)	⟨0.192; 0.218⟩ (0.225;0.252)
Especificidad	0.356 (0.334)	0.304 (0.352)	0.833 (0.823)	0.803 (0.837)
Intervalo (95%)	⟨0.344; 0.368⟩ (0.323;0.346)	⟨0.292; 0.315⟩ (0.34;0.364)	⟨0.823; 0.842⟩ (0.813;0.832)	⟨0.793; 0.813⟩ (0.828;0.846)
*Los valores en negrilla son los obtenidos para los algoritmos con hemograma.				

A partir de los resultados en la tabla **8-13** se observa que los escenarios presentan el mismo comportamiento evidenciado en el análisis realizado en este trabajo, sin embargo cuando se realiza el proceso clínico en serie Negativo se logran mejores sensibilidades en la base simulada a partir de las categorías sin embargo se logra mejores especificidades en la base simulada a partir de los resultados de los algoritmos, al contrario sucede cuando se aplica el caso en serie Positivo, ya que las sensibilidades mayores se alcanzan bajo la base simulada a partir de los algoritmos y las especificidades son mayores en la base simulada a partir de las categorías, podría decirse que ambas simulaciones son validas, esto indica que bajo el nivel del aproximadamente del 38% se logra sensibilidades mayores al 70% en el escenario clínico en serie Negativo y especificidades mayores al 70% en el escenario de aplicación Serie Positivo.

9. Conclusiones y Recomendaciones.

9.1. Conclusiones

El uso de los algoritmos con y sin hemograma en conjunto con la prueba rápida tienen en la mayoría de los casos valores por encima del 90% en las sensibilidades en un escenario de estudio clínico en paralelo y serie negativo, es decir, al realizar los tratamientos para el diagnóstico de dengue contrastado con una prueba de referencia patrón de oro (gold standard) en distintos niveles de prevalencia se obtiene una muy buena clasificación de los pacientes que están verdaderamente enfermos. La especificidad de los prototipos en general se va aumentando a medida que la prevalencia de la enfermedad es mayor, es decir, la importancia de los resultados no solo corresponde con tener valores altos de sensibilidad, el funcionamiento de los algoritmos con o sin hemograma adicionando los resultados de la prueba rápida mejora la clasificación de los individuos realmente sanos a medida que el nivel de prevalencia es más alto.

Cabe mencionar que al obtener estimaciones altas de sensibilidades comparado con las especificidades en el escenario cuando las pruebas son aplicadas simultáneamente (paralelo) se puede evidenciar los altos valores que toma la tasa de falsos positivos a la hora de hacer el proceso de clasificación de las pruebas, lo que llevaría a una subestimación de la proporción de individuos que están realmente enfermos. Es importante aclarar la descripción anterior, dado que en este proyecto de grado los archivos de datos son generados a partir de simulación, por tanto los errores de mala clasificación por parte del diagnóstico no son influyentes en los recursos económicos que puede imponer la metodología de investigación sobre un estudio de una enfermedad; sin embargo, es pertinente tener en cuenta estas descripciones con la finalidad de disminuir el posible sesgo de selección

en el caso de adaptación en un contexto real.

Desde la teoría médica, en el escenario clínico de serie positivo se espera encontrar en los resultados una mayor especificidad en el rendimiento de las pruebas diagnósticas lo cual se presenta en la tabla **8-7**, obteniendo estimaciones entre 80% y más del 90% en cada algoritmo con y sin hemograma teniendo en cuenta los resultados de la prueba rápida.

Se sugiere usar el procedimiento clínico aplicado en serie negativo, ya que se alcanzó sensibilidades mayores al 90% y especificidades moderadas de aproximadamente entre el 40% y 50%, por ello, es recomendable aplicar este escenario puesto que la tasa de falsos negativos disminuye y a pesar que la tasa de falsos positivos es alta, lo cual puede verse reflejada en los costos del estudio en un marco de aplicación real, la salud y el buen diagnóstico de los pacientes que tengan dengue va a predominar.

Al realizar la comparación de las bases obtenidas mediante las categorías de los signos y síntomas y directamente de los resultados binarios para los algoritmos I y III se observó que no existe diferencia significativa entre sus estimaciones, por tanto realizarla la simulación de cualquiera de las dos formas propuestas brindarán resultados útiles.

En lo que corresponde a la validez externa al proceso de estimación, se hizo a partir de la técnica de remuestreo de bootstrap con la finalidad de evaluar el desempeño de los tratamientos y se encontró la convergencia de las estimaciones de los parámetros de desempeño a distintos niveles de prevalencia, en donde en cada escenario de clasificación no se obtuvo una diferencia notable con respecto a la estimación en las aproximaciones clásica y bayesiana.

Es muy importante tener en cuenta los resultados expuestos en este trabajo de grado en el panorama el cual se tenga una baja, moderada o alta prevalencia del dengue con bases de datos suministradas por alguna entidad de salud interesada en el estudio del dengue y en el mejor de los casos con la ayuda del conocimiento que tenemos como estadísticos estar desde el primer momento de la planificación de los objetivos de la posible investigación para así asegurar archivos de datos ordenados y dignos para el correspondiente análisis, además de aplicar otras herramientas estadísticas como

por ejemplo realizar comparaciones con otros métodos de simulación para obtener datos binarios correlacionados y demás.

9.2. Recomendaciones

Las prevalencias utilizadas en este trabajo resultan de utilidad, sin embargo para mejorar esta investigación y brindar mayores herramientas a los médicos y mecanismos encargados del diagnóstico del Dengue se recomienda ampliar el rango de los valores de prevalencia para la simulación.

El modelo aplicado tuvo en cuenta una correlación binaria tanto para individuos sanos como enfermos, dado que en la simulación realizada se plantean estas correlaciones y son conocidas, la estimación de los parámetros de desempeño se obtuvo directamente del arreglo de datos en la tabla 7-6 se propone realizar la comparación de las estimaciones por medio de métodos numéricos tales como Newton-Raphson.

A. Anexo: Código R

A.1. Código para simulación de los datos binarios correlacionados.

```
#----- SIMULACIÓN BASE -----#
#      nrep = Numero de repeticiones
#      d = numero de variables
# mean.vec = vector de medias (proporciones iniciales)
# corr.mat = Matriz de correlaciones

draw.correlated.binary<-function(nrep,d,mean.vec,corr.mat){
  if((nrep<1)|(floor(nrep)!=nrep)){
    # Si el numero de repeticiones es menor que uno o
    # floor devuelve el numero entero más grande no mayor
    #que el numero que da
    # Por ejemplo floor(-3.5) = -4 o floor(2.5)=2
    stop("El numero de repeticiones debe ser un n\úmero entro
    #cuyo valor
        sea al menos 1!\n")
  }
  if((d<2)|(floor(d)!=d)){
    stop("La dimension debe ser un numero entero
    cuyo valor sea al menos de 2!\n")
  }
  if((max(mean.vec)>=1)|(min(mean.vec)<=0)){
```

```

    Vector de medias debe estar compuesto por proporciones
    stop("Las medias deben estar entre 0 y 1!\n")
  }
  if(length(mean.vec)!=d){
    stop("El vector de medias esta mal especificado, la dimension
    es incorrecta!\n")
  }
  if((ncol(corr.mat)!=d)|(nrow(corr.mat)!=d)){
    stop("La matriz de correlacion esta mal especificada, la
    dimension es incorrecta!\n")
  }
  if(sum(corr.mat!=t(corr.mat))>0){
    stop("La matriz de correlacion no es simetrica!\n")
  }
  if(sum(diag(corr.mat)!=rep(1,d))>0){
    stop("No todos los elementos de la diagonal de la matriz de
    correlacion son 1!\n")
  }
  if((max(corr.mat)>1)|(min(corr.mat)<0)){
    stop("Las correlaciones deben estar entre 0 y 1!\n")
  }

  alpha<-matrix(0,d,d) ; cor.limit<-matrix(0,d,d)
  for (i in 1:d){ # Calcula el valor maximo que puede tomar
#las correlaciones entre las variables
    for (j in 1:d){ # Formula (2.4)
      cor.limit[i,j]<-min(sqrt(mean.vec[j]*(1-mean.vec[i]))/
(mean.vec[i]*(1-mean.vec[j])),
      sqrt((mean.vec[i]*(1-mean.vec[j]))/(mean.vec[j]*
(1-mean.vec[i]))))
    }
  }
  if(sum(cor.limit>=corr.mat)<d^2){
    stop("Las correlaciones estan mas alla de sus limites
    superiores impuestos por las expectativas")
  }
  for (i in 1:d){for (j in 1:d){ #Formula (2.3)
    alpha[i,j]<-log(1+corr.mat[i,j]*sqrt((1-mean.vec[i])*

```

```

    (1-mean.vec[j])/(mean.vec[i]*mean.vec[j]))))}}
#Valores de alphaij para las variables X poisson
beta<-matrix(0,d,d*d)
summ<-1 ; counter<-0
while (summ>0){
    counter<-counter+1 # Contador
    minloc<-min.loc.finder(alpha,d) # Localizacion del minimo
    #global,
    #indices para S1
    w<-matrix(1,d,d)
    my.min<-apply(matrix(alpha[,-minloc],d,
    d-length(unique(minloc))),2,min)

    # Unique quita los valores duplicados
    # Calcula el valor minimo para cada columna sin tener en
    #cuenta las columnas correspondiente a las
    # coordenadas de localizacion del minimo calculado
    #anteriormente
    if (length(my.min)==1){
        w[,-minloc][my.min==0]<-0
        w[-minloc,][my.min==0]<-0}
    if (length(my.min)>1){
        w[,-minloc][,my.min==0]<-0
        w[-minloc,][my.min==0,]<-0
        w[alpha==0]<-0}
    for (i in 1:d){
        beta[i,counter]<- alpha[minloc[1],minloc[2]]*1*
        ((minloc[1]==i)|(minloc[2]==i)|(sum(w[,i])==d)})
    }
    alpha<-alpha-alpha[minloc[1],minloc[2]]*w
}

```

```

summ<-sum(alpha) }

#hacer el while hasta que la diferencia entre los alpha
#sea cero
tbeta<-t(beta)
w<-(tbeta!=0) # muestra si los valores de los betas son
#diferentes de cero
x<-matrix(0,nrep,d); y<-matrix(0,nrep,d) # Matriz que
#almacenara los valores Poisson correlacionados
pois<-numeric(nrow(tbeta)) # Vector numerico del tamano de
#las filas de beta traspuesta
sump<-numeric(d) #Vector del tamano de las variables
for (k in 1:nrep){
  for (j in 1:nrow(tbeta)){
    pois[j]<-rpois(1,max(tbeta[j,]))} #Genero un valor Poisson
    #con el parametro beta seleccionado
    for (i in 1:d){
      sump[i]<-sum(pois*w[,i])} # Almacena las suma solo para
      #los Poisson que son verdaderos
    x[k,]<-sump} # Calculo de las variables Y convolucionadas
y[x==0]<-1 # Donde sea verdadero ponga un 1
y[x!=0]<-0 # Donde sea vardadero ponga un 0
y} # Para lo que en las copias es Z

min.loc.finder<-function(my.mat,d){ # Funcion que localiza el
#minimo de los alpha
w<-is.matrix(my.mat) # Verfica si la matriz corresponda a la
#clase que se necesita

```

```

if (w==F){stop("Esto no es una matriz!\n")} # Si la matriz no
#es de la clase "matrix" arroja error

if (nrow(my.mat)!=ncol(my.mat)){ # Verifica si la matriz es
#cuadrada
  stop("Esta no es una matriz cuadrada!\n")}
n<-nrow(my.mat) # Numero de filas o observaciones
my.vec<-as.vector(t(my.mat)) # Vuelve vector la matrix agrada
my.vec[my.vec==0]<-999
my.index<-min((1:length(my.vec))[my.vec==min(my.vec)])
#Celda donde esta el minimo valor
row.index<-floor((my.index-1)/n)+1 # Numero de la fila donde
#se encuentra el valor minimo
col.index<-my.index-d*floor((my.index-1)/n) #Numero de la
#columna donde sen encuentra el valor minimo
c(row.index,col.index)} # Coordenada dentro de la tabla,
#donde esta el valor minimo

### Para generar matrices aleatorias de correlaci'on
matriz_cor<-function(nvar,p){
  rho=matrix(0,nvar,nvar)
  for(i in 1:nvar){
    for(j in 1:nvar){
      if(i==j){ rho[i,j]=1 }
      if(i<j){
        a=sqrt((p[j]*(1-p[i]))/(p[i]*(1-p[j])))
        b=sqrt((p[i]*(1-p[j]))/(p[j]*(1-p[i])))
        rho[i,j]=rho[j,i]=round(runif(1,0,min(a,b)),4)
      } } }
} } }

```

```

return(rho) }

#----- GERANDO BASE -----#
rep=1 ### Numero de individuos
nvar=9 ## Numero de variables
mean= read.delim("algoritmos1.txt", sep = " " ) #llamando propociones inicial
rho=matriz_cor(nvar = nrow(mean),p=t(mean) )
#rho=read.delim("algoritmos.txt", sep = " " )
#rho=read.delim("correlaciones1.txt", sep = " " )
variables=draw.correlated.binary(nrep=rep, d=nvar,
mean.vec = t(mean), corr.mat = rho)
colnames(variables)<-c("A1", "A2", "A3", "A4",
"AH1", "AH2", "AH3", "AH4", "PR")
#View(variables)
#variables1<-as.matrix(variables)
#getwd() ### escritorio
write.table(variables, "variablesENF.txt", sep = "\t",
quote = F, row.names = F)
write.table(rho, "correlacionesENF.txt", sep = "\t",
quote = F, row.names = F)

```

A.2. Código para estimación de los parámetros de desempeño a partir de las bases simuladas

```

#-----ANALISIS DESCRIPTIVO -----#
#Base de datos
base<-read.delim("basefinal.txt", sep="")
head(base)

```

A.2 Código para estimación de los parámetros de desempeño a partir de las bases simuladas 69

```
attach(base)
```

```
sapply(base,class)
```

```
#Se transforman en variables de categoricas
```

```
#Algoritmos sin hemograma
```

```
Algorit1<-factor(base$A1)
```

```
Algorit2<-factor(base$A2)
```

```
Algorit3<-factor(base$A3)
```

```
Algorit4<-factor(base$A4)
```

```
#Algoritmos con hemograma
```

```
Algorit1H<-factor(base$AH1)
```

```
Algorit2H<-factor(base$AH2)
```

```
Algorit3H<-factor(base$AH3)
```

```
Algorit4H<-factor(base$AH4)
```

```
#Prueba rapida
```

```
Pr<-factor(base$PR)
```

```
levels (Algorit1); levels(Algorit2); levels(Algorit3)
```

```
levels(Algorit4); levels(Algorit1H) ; levels(Algorit2H); levels(Algorit3H)
```

```
levels(Algorit4H); levels(Pr) ;
```

```
#base de datos con todas las variables categoricas
```

```
datos1<-data.frame(Algorit1,Algorit2,Algorit3,
```

```
Algorit4,Algorit1H,Algorit3H,Algorit4H,Pr)
```

```
attach(datos1)

#resumen del archivo de datos donde se muestran
# la cantidad de positivos (1) y negativos (0) para dengue
# y asi proceder a realizar las respectivas proporciones.
summary(datos1)

#----- CLASIFICACION -----#
library(xtable)
#----- ESTIMACION -----#
base<-as.matrix(basefinal)
View(base)
n<-10000
# CONTEOS SIN HEMOGRAMA
a1<-0;a2<-0;a3<-0;a4<-0;a5<-0;a6<-0;a7<-0;a8<-0 # A1-PR-GS
b1<-0;b2<-0;b3<-0;b4<-0;b5<-0;b6<-0;b7<-0;b8<-0 # A2-PR-GS
c1<-0;c2<-0;c3<-0;c4<-0;c5<-0;c6<-0;c7<-0;c8<-0 # A3-PR-GS
d1<-0;d2<-0;d3<-0;d4<-0;d5<-0;d6<-0;d7<-0;d8<-0 # A4-PR-GS

# CONTEOS CON HEMOGRAMA
e1<-0;e2<-0;e3<-0;e4<-0;e5<-0;e6<-0;e7<-0;e8<-0 # AH1-PR-GS
f1<-0;f2<-0;f3<-0;f4<-0;f5<-0;f6<-0;f7<-0;f8<-0 # AH2-PR-GS
g1<-0;g2<-0;g3<-0;g4<-0;g5<-0;g6<-0;g7<-0;g8<-0 # AH3-PR-GS
h1<-0;h2<-0;h3<-0;h4<-0;h5<-0;h6<-0;h7<-0;h8<-0 # AH4-PR-GS

for(i in 1:n){
  # resultado de combinacion para patron de oro Positivo
  if(base[i,1]==1 && base[i,9]==1 && base[i,10]==1){a1=1+a1}
  if(base[i,1]==1 && base[i,9]==0 && base[i,10]==1){a2=1+a2}
```

A.2 Código para estimación de los parámetros de desempeño a partir de las bases simuladas 71

```
if (base[i,1]==0 && base[i,9]==1 && base[i,10]==1) {a3=1+a3}
if (base[i,1]==0 && base[i,9]==0 && base[i,10]==1) {a4=1+a4}
if (base[i,2]==1 && base[i,9]==1 && base[i,10]==1) {b1=1+b1}
if (base[i,2]==1 && base[i,9]==0 && base[i,10]==1) {b2=1+b2}
if (base[i,2]==0 && base[i,9]==1 && base[i,10]==1) {b3=1+b3}
if (base[i,2]==0 && base[i,9]==0 && base[i,10]==1) {b4=1+b4}
if (base[i,3]==1 && base[i,9]==1 && base[i,10]==1) {c1=1+c1}
if (base[i,3]==1 && base[i,9]==0 && base[i,10]==1) {c2=1+c2}
if (base[i,3]==0 && base[i,9]==1 && base[i,10]==1) {c3=1+c3}
if (base[i,3]==0 && base[i,9]==0 && base[i,10]==1) {c4=1+c4}
if (base[i,4]==1 && base[i,9]==1 && base[i,10]==1) {d1=1+d1}
if (base[i,4]==1 && base[i,9]==0 && base[i,10]==1) {d2=1+d2}
if (base[i,4]==0 && base[i,9]==1 && base[i,10]==1) {d3=1+d3}
if (base[i,4]==0 && base[i,9]==0 && base[i,10]==1) {d4=1+d4}
if (base[i,5]==1 && base[i,9]==1 && base[i,10]==1) {e1=1+e1}
if (base[i,5]==1 && base[i,9]==0 && base[i,10]==1) {e2=1+e2}
if (base[i,5]==0 && base[i,9]==1 && base[i,10]==1) {e3=1+e3}
if (base[i,5]==0 && base[i,9]==0 && base[i,10]==1) {e4=1+e4}
if (base[i,6]==1 && base[i,9]==1 && base[i,10]==1) {f1=1+f1}
if (base[i,6]==1 && base[i,9]==0 && base[i,10]==1) {f2=1+f2}
if (base[i,6]==0 && base[i,9]==1 && base[i,10]==1) {f3=1+f3}
if (base[i,6]==0 && base[i,9]==0 && base[i,10]==1) {f4=1+f4}
if (base[i,7]==1 && base[i,9]==1 && base[i,10]==1) {g1=1+g1}
if (base[i,7]==1 && base[i,9]==0 && base[i,10]==1) {g2=1+g2}
if (base[i,7]==0 && base[i,9]==1 && base[i,10]==1) {g3=1+g3}
if (base[i,7]==0 && base[i,9]==0 && base[i,10]==1) {g4=1+g4}
if (base[i,8]==1 && base[i,9]==1 && base[i,10]==1) {h1=1+h1}
if (base[i,8]==1 && base[i,9]==0 && base[i,10]==1) {h2=1+h2}
if (base[i,8]==0 && base[i,9]==1 && base[i,10]==1) {h3=1+h3}
```

```

if(base[i,8]==0 && base[i,9]==0 && base[i,10]==1){h4=1+h4}
# resultado combinacion para patron de oro negativo
if(base[i,1]==1 && base[i,9]==1 && base[i,10]==0){a5=1+a5}
if(base[i,1]==1 && base[i,9]==0 && base[i,10]==0){a6=1+a6}
if(base[i,1]==0 && base[i,9]==1 && base[i,10]==0){a7=1+a7}
if(base[i,1]==0 && base[i,9]==0 && base[i,10]==0){a8=1+a8}
if(base[i,2]==1 && base[i,9]==1 && base[i,10]==0){b5=1+b5}
if(base[i,2]==1 && base[i,9]==0 && base[i,10]==0){b6=1+b6}
if(base[i,2]==0 && base[i,9]==1 && base[i,10]==0){b7=1+b7}
if(base[i,2]==0 && base[i,9]==0 && base[i,10]==0){b8=1+b8}
if(base[i,3]==1 && base[i,9]==1 && base[i,10]==0){c5=1+c5}
if(base[i,3]==1 && base[i,9]==0 && base[i,10]==0){c6=1+c6}
if(base[i,3]==0 && base[i,9]==1 && base[i,10]==0){c7=1+c7}
if(base[i,3]==0 && base[i,9]==0 && base[i,10]==0){c8=1+c8}
if(base[i,4]==1 && base[i,9]==1 && base[i,10]==0){d5=1+d5}
if(base[i,4]==1 && base[i,9]==0 && base[i,10]==0){d6=1+d6}
if(base[i,4]==0 && base[i,9]==1 && base[i,10]==0){d7=1+d7}
if(base[i,4]==0 && base[i,9]==0 && base[i,10]==0){d8=1+d8}
if(base[i,5]==1 && base[i,9]==1 && base[i,10]==0){e5=1+e5}
if(base[i,5]==1 && base[i,9]==0 && base[i,10]==0){e6=1+e6}
if(base[i,5]==0 && base[i,9]==1 && base[i,10]==0){e7=1+e7}
if(base[i,5]==0 && base[i,9]==0 && base[i,10]==0){e8=1+e8}
if(base[i,6]==1 && base[i,9]==1 && base[i,10]==0){f5=1+f5}
if(base[i,6]==1 && base[i,9]==0 && base[i,10]==0){f6=1+f6}
if(base[i,6]==0 && base[i,9]==1 && base[i,10]==0){f7=1+f7}
if(base[i,6]==0 && base[i,9]==0 && base[i,10]==0){f8=1+f8}
if(base[i,7]==1 && base[i,9]==1 && base[i,10]==0){g5=1+g5}
if(base[i,7]==1 && base[i,9]==0 && base[i,10]==0){g6=1+g6}
if(base[i,7]==0 && base[i,9]==1 && base[i,10]==0){g7=1+g7}

```

A.2 Código para estimación de los parámetros de desempeño a partir de las bases simuladas 73

```
if(base[i,7]==0 && base[i,9]==0 && base[i,10]==0){g8=1+g8}
if(base[i,8]==1 && base[i,9]==1 && base[i,10]==0){h5=1+g5}
if(base[i,8]==1 && base[i,9]==0 && base[i,10]==0){h6=1+g6}
if(base[i,8]==0 && base[i,9]==1 && base[i,10]==0){h7=1+g7}
if(base[i,8]==0 && base[i,9]==0 && base[i,10]==0){h8=1+g8}
}

##tabla clasificacion para cada algoritmo sin hemograma
A1<-c(a1,a2,a3,a4,a5,a6,a7,a8);tabla1<-matrix(A1,2,4,byrow=T)
A2<-c(b1,b2,b3,b4,b5,b6,b7,b8);tabla2<-matrix(A2,2,4,byrow=T)
A3<-c(c1,c2,c3,c4,c5,c6,c7,c8);tabla2<-matrix(A3,2,4,byrow=T)
A4<-c(d1,d2,d3,d4,d5,d6,d7,d8);tabla2<-matrix(A3,2,4,byrow=T)

##tabla clasificacion para cada algoritmo con hemograma
AH1<-c(e1,e2,e3,e4,e5,e6,e7,e8);tabla2<-matrix(AH1,2,4,byrow=T)
AH2<-c(f1,f2,f3,f4,f5,f6,f7,f8);tabla2<-matrix(AH2,2,4,byrow=T)
AH3<-c(g1,g2,g3,g4,g5,g6,g7,f8);tabla2<-matrix(AH3,2,4,byrow=T)
AH4<-c(h1,h2,h3,h4,h5,h6,h7,h8);tabla2<-matrix(AH4,2,4,byrow=T)

##estimacion de los parametros de desempe~no
## por estadistica clasica
# ALGORITMOS SIN HEMOGRAMA
A1s1<-(a1+a2)/(a1+a2+a3+a4);A1e1<-(a7+a8)/(a5+a6+a7+a8)
A2s1<-(b1+b2)/(b1+b2+b3+b4);A2e1<-(b7+b8)/(b5+b6+b7+b8)
A3s1<-(c1+c2)/(c1+c2+c3+c4);A3e1<-(c7+c8)/(c5+c6+c7+c8)
A4s1<-(d1+d2)/(d1+d2+d3+d4);A4e1<-(d7+d8)/(d5+d6+d7+d8)
# ALGORITMOS CON HEMOGRAMA
AH1s1<-(e1+e2)/(e1+e2+e3+e4);AH1e1<-(e7+e8)/(e5+e6+e7+e8)
```

```

AH2s1<-(f1+f2)/(f1+f2+f3+f4);AH2e1<-(f7+f8)/(f5+f6+f7+f8)
AH3s1<-(g1+g2)/(g1+g2+g3+g4);AH3e1<-(g7+g8)/(g5+g6+g7+g8)
AH4s1<-(h1+h2)/(h1+h2+h3+h4);AH4e1<-(h7+h8)/(h5+h6+h7+h8)
#PRUEBA RAPIDA
A1s2<-(a1+a3)/(a1+a2+a3+a4);A1e2<-(a6+a8)/(a5+a6+a7+a8)

## estimacion de los parametros de desempe~no para
## escenario clinico en paralelo

# ALGORITMOS SIN HEMOGRAMA
A1s3<-(a1+a3)/(a1+a2+a3+a4); A1e3<-(a6+a8)/(a5+a6+a7+a8)
A2s3<-(b1+b3)/(b1+b2+b3+b4); A2e3<-(b6+b8)/(b5+b6+b7+b8)
A3s3<-(c1+c3)/(c1+c2+c3+c4); A3e3<-(c6+c8)/(c5+c6+c7+c8)
A4s3<-(d1+d3)/(d1+d2+d3+d4); A4e3<-(d6+d8)/(d5+d6+d7+d8)

#ALGORITMOS CON HEMOGRAMA
AH1s3<-(e1+e3)/(e1+e2+e3+e4); AH1e3<-(e6+e8)/(e5+e6+e7+e8)
AH2s3<-(f1+f3)/(f1+f2+f3+f4); AH2e3<-(f6+f8)/(f5+f6+f7+f8)
AH3s3<-(g1+g3)/(g1+g2+g3+g4); AH3e3<-(g6+g8)/(g5+g6+g7+g8)
AH4s3<-(h1+h3)/(h1+h2+h3+h4); AH4e3<-(h6+h8)/(h5+h6+h7+h8)

## estimacion de los parametros de desempe~no para
## escenario clinico en serie negativo

#ALGORITMOS SIN HEMOGRAMA
A1s5<-(a1+a2+a3)/(a1+a2+a3+a4); A1e5<-a8/(a5+a6+a7+a8)
A2s5<-(b1+b2+b3)/(b1+b2+b3+b4); A2e5<-b8/(b5+b6+b7+b8)
A3s5<-(c1+c2+c3)/(c1+c2+c3+c4); A3e5<-c8/(c5+c6+c7+c8)
A4s5<-(d1+d2+d3)/(d1+d2+d3+d4); A4e5<-d8/(d5+d6+d7+d8)

```

A.2 Código para estimación de los parámetros de desempeño a partir de las bases simuladas 75

```
#ALGORITMOS CON HEMOGRAMA
```

```
AH1s5<-(e1+e2+e3)/(e1+e2+e3+e4); AH1e5<-e8/(e5+e6+e7+e8)
```

```
AH2s5<-(f1+f2+f3)/(f1+f2+f3+f4); AH2e5<-f8/(f5+f6+f7+f8)
```

```
AH3s5<-(g1+g2+g3)/(g1+g2+g3+g4); AH3e5<-g8/(g5+g6+g7+g8)
```

```
AH4s5<-(h1+h2+h3)/(h1+h2+h3+h4); AH4e5<-h8/(h5+h6+h7+h8)
```

```
## estimacion de los parametros de desempe~no para
```

```
## escenario clinico en serie positivo
```

```
#ALGORITMOS SIN HEMOGRAMA
```

```
A1s6<-(a1)/(a1+a2+a3+a4); A1e6<-(a6+a7+a8)/(a5+a6+a7+a8)
```

```
A2s6<-(b1)/(b1+b2+b3+b4); A2e6<-(b6+b7+b8)/(b5+b6+b7+b8)
```

```
A3s6<-(c1)/(c1+c2+c3+c4); A3e6<-(c6+c7+c8)/(c5+c6+c7+c8)
```

```
A4s6<-(d1)/(d1+d2+d3+d4); A4e6<-(d6+d7+d8)/(d5+d6+d7+d8)
```

```
#ALGORITMOS CON HEMOGRAMA
```

```
AH1s6<-(e1)/(e1+e2+e3+e4); AH1e6<-(e6+e7+e8)/(e5+e6+e7+e8)
```

```
AH2s6<-(f1)/(f1+f2+f3+f4); AH2e6<-(f6+f7+f8)/(f5+f6+f7+f8)
```

```
AH3s6<-(g1)/(g1+g2+g3+g4); AH3e6<-(g6+g7+g8)/(g5+g6+g7+g8)
```

```
AH4s6<-(h1)/(h1+h2+h3+h4); AH4e6<-(h6+h7+h8)/(h5+h6+h7+h8)
```

```
#-----TABLAS DE RESULTADOS-----#
```

```
#- estimaciones
```

```
es<-c(A1s2,A1e2,A1s3,A1e3,A2s3,A2e3,A3s3,A3e3,  
      A4s3,A4e3,AH1s3,AH1e3,AH2s3,AH2e3,AH3s3,AH3e3,  
      AH4s3,AH4e3,A1s5,A1e5,A2s5,A2e5,A3s5,A3e5,A4s5,  
      A4e5,AH1s5,AH1e5,AH2s5,AH2e5,AH3s5,AH3e5,AH4s5,  
      AH4e5,A1s6,A1e6,A2s6,A2e6,A3s6,A3e6,A4s6,A4e6,  
      AH1s6,AH1e6,AH2s6,AH2e6,AH3s6,AH3e6,AH4s6,AH4e6)
```

```

estim<-matrix(es,2,25,FALSE)
colnames(estim)<-c("PR","P1","P2","P3","P4","PH1","PH2",
                  "PH3","PH4","S1A1","S1A2","S1A3","S1A4",
                  "S1AH1","S1AH2","S1AH3","S1AH4",
                  "S2A1","S2A2","S2A3","S2A4",
                  "S2AH1","S2AH2","S2AH3","S2AH4")
rownames(estim)<-c("sensibilidad","especificidad")

#- Intervalos
est1<-c(A1s2,A1s3,A2s3,A3s3,A4s3,AH1s3,AH2s3,AH3s3,
        AH4s3,A1s5,A2s5,A3s5,A4s5,AH1s5,AH2s5,AH3s5,
        AH4s5,A1s6,A2s6,A3s6,A4s6,AH1s6,AH2s6,AH3s6,
        AH4s6)

est2<-c(A1e2,A1e3,A2e3,A3e3,A4e3,AH1e3,AH2e3,AH3e3,
        AH4e3,A1e5,A2e5,A3e5,A4e5,AH1e5,AH2e5,AH3e5,
        AH4e5,A1e6,A2e6,A3e6,A4e6,AH1e6,AH2e6,AH3e6,
        AH4e6)

npos<-c(a1+a2+a3+a4,a1+a2+a3+a4,b1+b2+b3+b4,c1+c2+c3+c4,
        d1+d2+d3+d4,e1+e2+e3+e4,f1+f2+f3+f4,g1+g2+g3+g4,
        h1+h2+h3+h4,a1+a2+a3+a4,b1+b2+b3+b4,c1+c2+c3+c4,
        d1+d2+d3+d4,e1+e2+e3+e4,f1+f2+f3+f4,g1+g2+g3+g4,
        h1+h2+h3+h4,a1+a2+a3+a4,b1+b2+b3+b4,c1+c2+c3+c4,
        d1+d2+d3+d4,e1+e2+e3+e4,f1+f2+f3+f4,g1+g2+g3+g4,
        h1+h2+h3+h4)

liminf_sen<-0 ; liminf_esp<-0
limsup_sen<-0 ; limsup_esp<-0

```

A.2 Código para estimación de los parámetros de desempeño a partir de las bases simuladas 77

```
for(i in 1:length(est1)){
  liminf_sen[i]<-(est1[i] + qnorm(0.025,0,1)* sqrt(est1[i]*
    (1-est1[i])/(npos[i]))))
  limsup_sen[i]<-(est1[i] + qnorm(0.975,0,1)* sqrt(est1[i]*
    (1-est1[i])/(npos[i]))))
  liminf_esp[i]<-(est2[i] + qnorm(0.025,0,1)* sqrt(est2[i]*
    (1-est2[i])/(10000-npos[i]))))
  limsup_esp[i]<-(est2[i] + qnorm(0.975,0,1)* sqrt(est2[i]*
    (1-est2[i])/(10000-npos[i]))))}

sensi<-data.frame(liminf_sen,limsup_sen)
especi<-data.frame(liminf_esp,limsup_esp)

#-----INFERENCIA BAYESIANA -----#
#a priori no informativa beta(1,1)
# APOSTERIORI PARA SENSIBILIDAD METODO LAPLACE

# A1
alfa1<-rbeta(10000,(a1+a2 + 1), (a3+a4 + 1))
media<-mean(alfa1)
delta<-c(0.025,0.975)
qbeta(delta,(a1+a2 + 1),(a3+a4 + 1)) #IC

#PR LAPLACE
alfa2<-rbeta(10000,(a1+a3 + 1), (a2+a4 + 1))
mediana<-median(alfa2)
media<-mean(alfa2)
delta<-c(0.025,0.975)
qbeta(delta,(a1+a3 + 1),(a2+a4 + 1)) #IC
```

```

##### POSTERIOR CONJUGADA PR
alfa2<-rbeta(10000, (a1+a3 + 885.65), (a2+a4 + 216.59))
mediana<-median(alfa2)
media<-mean(alfa2)
delta<-c(0.025,0.975)
qbeta(delta, (a1+a3 + 885.65), (a2+a4 + 216.59)) #IC

# APOSTERIOR PARA LA ESPECIFICIDAD
#A1
beta1<-rbeta(10000, (a7+a8 + 1), (a5+a6 + 1))
media<-mean(beta1)
delta<-c(0.025,0.975)
qbeta(delta, (a7+a8 + 1), (a5+a6 + 1)) #IC

#PR LAPLACE
beta2<-rbeta(10000, (a6+a8 + 1), (a5+a7 + 1))
media<-mean(beta2)
delta<-c(0.025,0.975)
qbeta(delta, (a6+a8 + 1), (a5+a7 + 1)) #IC

##### POSTERIOR CONJUGADA PR
beta2<-rbeta(10000, (a6+a8 + 398.79), (a5+a7 + 55.15))
media<-mean(beta2)
delta<-c(0.025,0.975)
qbeta(delta, (a6+a8 + 398.79), (a5+a7 + 55.15)) #IC

#Como analisis opcional, observamos el grafico de densidad
#de la posteriori para la sensibilidad y especificidad
#con sus respectiva media y intervalos de credibilidad.

x11();plot(density(alfa2), main="Posterior Sensibilidad Algoritmo 1")

```

A.2 Código para estimación de los parámetros de desempeño a partir de las bases simuladas 79

```
abline(v=c(qbeta(delta, (a1+a3 + 1), (a2+a4 + 1))), col="red")
abline(v=mean(alfa2))

x11();plot(density(beta2), main="Posterior especificidad Algoritmo 1")
abline(v=c(qbeta(delta, (a6 +a8 + 1), (a5+a7 + 1))), col="red")
abline(v=mean(beta2))

# POSTERIORES EN PARALELO

#PARALELO 1
#SENSIBILIDAD
paralel.1<-rbeta(10000, (a1+a3 + 1), (a2+a4 + 1))
mediaps<-mean(paralel.1)
delta<-c(0.025,0.975)
mediaps
qbeta(delta, (a1+a3 + 1), (a2+a4 + 1)) #IC

#ESPECIFICIDAD
paralel.2<-rbeta(10000, (a6+a8 + 1), (a5+a7 + 1))
mediape<-mean(paralel.2)
delta<-c(0.025,0.975)
mediape
qbeta(delta, (a6+a8 + 1), (a5+a7 + 1)) #IC

#POSTERIORES EN SERIE
#SERIE1
#SENSIBILIDAD
serie1.1<-rbeta(10000, (a1+a2+a3+1), (a4+1))
```

```

serie1.2<-rbeta(10000, (b1+b2+b3+1), (b4+1))
serie1.3<-rbeta(10000, (c1+c2+c3+1), (c4+1))
serie1.4<-rbeta(10000, (d1+d2+d3+1), (d4+1))

serie1h.1<-rbeta(10000, (e1+e2+e3+1), (e4+1))
serie1h.2<-rbeta(10000, (f1+f2+f3+1), (f4+1))
serie1h.3<-rbeta(10000, (g1+g2+g3+1), (g4+1))
serie1h.4<-rbeta(10000, (h1+h2+h3+1), (h4+1))
media_s<-c(mean(serie1.1),mean(serie1.2),mean(serie1.3),
            mean(serie1.4),mean(serie1h.1),mean(serie1h.2),
            mean(serie1h.3),mean(serie1h.4))
delta<-c(0.025,0.975)
IC_s<-c(qbeta(delta, (a1+a2+a3+1), (a4+1)),
        qbeta(delta, (b1+b2+b3+1), (b4+1)),
        qbeta(delta, (c1+c2+c3+1), (c4+1)),
        qbeta(delta, (d1+d2+d3+1), (d4+1)),
        qbeta(delta, (e1+e2+e3+1), (e4+1)),
        qbeta(delta, (f1+f2+f3+1), (f4+1)),
        qbeta(delta, (g1+g2+g3+1), (g4+1)),
        qbeta(delta, (h1+h2+h3+1), (h4+1)))
IC_s<-matrix(IC_s,8,2,T)#intervalo de credibilidad
rownames(IC_s)<-c("A1", "A2", "A3", "A4", "AH1", "AH2",
                 "AH3", "AH4")
colnames(IC_s)<-c("LimInf", "LimSup")
media_s
IC_s

#ESPECIFICIDAD
serie1.1.1<-rbeta(10000, (a8+1), (a5+a6+a7+1))

```

A.2 Código para estimación de los parámetros de desempeño a partir de las bases simuladas 81

```
serie1.1.2<-rbeta(10000,(b8+1),(b5+b6+b7+1))
serie1.1.3<-rbeta(10000,(c8+1),(c5+c6+c7+1))
serie1.1.4<-rbeta(10000,(d8+1),(d5+d6+d7+1))

serie1.1h.1<-rbeta(10000,(e8+1),(e5+e6+e7+1))
serie1.1h.2<-rbeta(10000,(f8+1),(f5+f6+f7+1))
serie1.1h.3<-rbeta(10000,(g8+1),(g5+g6+g7+1))
serie1.1h.4<-rbeta(10000,(h8+1),(h5+h6+h7+1))
media_e<-c(mean(serie1.1.1),mean(serie1.1.2),
            mean(serie1.1.3),mean(serie1.1.4),
            mean(serie1.1h.1),mean(serie1.1h.2),
            mean(serie1.1h.3),mean(serie1.1h.4))
delta<-c(0.025,0.975)
IC_e<-c(qbeta(delta,(a8+1),(a5+a6+a7+1)),
        qbeta(delta,(b8+1),(b5+b6+b7+1)),
        qbeta(delta,(c8+1),(c5+c6+c7+1)),
        qbeta(delta,(d8+1),(d5+d6+d7+1)),
        qbeta(delta,(e8+1),(e5+e6+e7+1)),
        qbeta(delta,(f8+1),(f5+f6+f7+1)),
        qbeta(delta,(g8+1),(g5+g6+g7+1)),
        qbeta(delta,(h8+1),(h5+h6+h7+1)))
IC_e<-matrix(IC_e,8,2,T)#intervalo de credibilidad
rownames(IC_e)<-c("A1","A2","A3","A4","AH1","AH2",
                 "AH3","AH4")
colnames(IC_e)<-c("LimInf","LimSup")
media_e
IC_e

#----- RESUMEN DE LOS VALORES BAYESIANOS SERIE1
```

```

bayesiana<-rbind(media_s , media_e)
colnames(bayesiana)<-c("A1", "A2", "A3", "A4", "AH1", "AH2",
                      "AH3", "AH4")
rownames(bayesiana)<-c("Sensibilidad" , "Especificidad")
IC_seriel<-cbind(IC_s, IC_e)
bayesiana
IC_seriel

#SERIE2
#SENSIBILIDAD
serie2.1<-rbeta(10000, (a1+1), (a2+a3+a4+1))
serie2.2<-rbeta(10000, (b1+1), (b2+b3+b4+1))
serie2.3<-rbeta(10000, (c1+1), (c2+c3+c4+1))
serie2.4<-rbeta(10000, (d1+1), (d2+d3+d4+1))

serie2h.1<-rbeta(10000, (e1+1), (e2+e3+e4+1))
serie2h.2<-rbeta(10000, (f1+1), (f2+f3+f4+1))
serie2h.3<-rbeta(10000, (g1+1), (g2+g3+g4+1))
serie2h.4<-rbeta(10000, (h1+1), (h2+h3+h4+1))
media_s2<-c(mean(serie2.1), mean(serie2.2), mean(serie2.3),
            mean(serie2.4), mean(serie2h.1), mean(serie2h.2),
            mean(serie2h.3), mean(serie2h.4))
delta<-c(0.025, 0.975)
IC_s2<-c(qbeta(delta, (a1+1), (a2+a3+a4+1)),
        qbeta(delta, (b1+1), (b2+b3+b4+1)),
        qbeta(delta, (c1+1), (c2+c3+c4+1)),
        qbeta(delta, (d1+1), (d2+d3+d4+1)),
        qbeta(delta, (e1+1), (e2+e3+e4+1)),
        qbeta(delta, (f1+1), (f2+f3+f4+1)),

```

A.2 Código para estimación de los parámetros de desempeño a partir de las bases simuladas 83

```
qbeta(delta, (g1+1), (g2+g3+g4+1)),
qbeta(delta, (h1+1), (h2+h3+h4+1)))
IC_s2<-matrix(IC_s2,8,2,T)#intervalo de credibilidad
rownames(IC_s2)<-c("A1", "A2", "A3", "A4", "AH1", "AH2",
                  "AH3", "AH4")
colnames(IC_s2)<-c("LimInf", "LimSup")
media_s2
IC_s2

#ESPECIFICIDAD
serie21.1<-rbeta(10000, (a6+a7+a8+1), (a5+1))
serie21.2<-rbeta(10000, (b6+b7+b8+1), (b5+1))
serie21.3<-rbeta(10000, (c6+c7+c8+1), (c5+1))
serie21.4<-rbeta(10000, (d6+d7+d8+1), (d5+1))

serie21h.1<-rbeta(10000, (e6+e7+e8+1), (e5+1))
serie21h.2<-rbeta(10000, (f6+f7+f8+1), (f5+1))
serie21h.3<-rbeta(10000, (g6+g7+g8+1), (g5+1))
serie21h.4<-rbeta(10000, (h6+h7+h8+1), (h5+1))
media_e2<-c(mean(serie21.1), mean(serie21.2), mean(serie21.3),
             mean(serie21.4), mean(serie21h.1), mean(serie21h.2),
             mean(serie21h.3), mean(serie21h.4))
delta<-c(0.025, 0.975)
IC_e2<-c(qbeta(delta, (a6+a7+a8+1), (a5+1)),
         qbeta(delta, (b6+b7+b8+1), (b5+1)),
         qbeta(delta, (c6+c7+c8+1), (c5+1)),
         qbeta(delta, (d6+d7+d8+1), (d5+1)),
         qbeta(delta, (e6+e7+e8+1), (e5+1)),
         qbeta(delta, (f6+f7+f8+1), (f5+1)),
```

```

      qbeta(delta, (g6+g7+g8+1), (g5+1)),
      qbeta(delta, (h6+h7+h8+1), (h5+1)))
IC_e2<-matrix(IC_e2,8,2,T)#intervalo de credibilidad
rownames(IC_e2)<-c("A1","A2","A3","A4","AH1","AH2",
                  "AH3","AH4")
colnames(IC_e2)<-c("LimInf","LimSup")
media_e2
IC_e2

#--- RESUMEN DE LOS VALORES BAYESIANOS P/ SERIE2
bayesiana2<-rbind(media_s2 , media_e2)
colnames(bayesiana2)<-c("A1","A2","A3","A4","AH1","AH2",
                      "AH3","AH4")
rownames(bayesiana2)<-c("Sensibilidad" , "Especificidad")

IC_serie2<-cbind(IC_s2,IC_e2)
bayesiana2
IC_serie2
###TABLAS TIPO LATEX
print(xtable(bayesiana,digits = 4), include.rownames = T)
print(xtable(bayesiana2,digits = 4), include.rownames =T)
print(xtable(IC_serie1,digits = 4), include.rownames = T)
print(xtable(IC_serie2,digits = 4), include.rownames = T)

```

A.3. Código para técnica de remuestreo bootstrap

```

#--- BOOTSTRAP ---#
bootstrap<-function(y,bo){

```

```

y<-as.matrix(y) # convierte el archivo a matriz
N<-10000 # cantidad de individuos en el archivo de datos
ind<-seq(1:N); n<-length(ind) # genera individuos aleatorios.
x <- vector('list', bo) ## donde se guardaran las bases remuestreadas.
muestra<-matrix(0,n,bo) # Matriz con los individuos simulados.
for(j in 1:bo){
  muestra[,j]=sample(ind,n,replace = T) }
column<-10 # numero de columnas de la matriz
for(j in 1:bo){
  x[[j]]<-(matrix(y[muestra[,j]],n,column))
} # Obteniendo las matrices remuestreadas.

#--- CLASIFICACION ---#
# ALGORITMOS SIN HEMOGRAMA
a1<-0; a2<-0; a3<-0; a4<-0; a5<-0; a6<-0; a7<-0; a8<-0 # A1-PR-GS
b1<-0; b2<-0; b3<-0; b4<-0; b5<-0; b6<-0 ; b7<-0 ; b8<-0 # A2-PR-GS
c1<-0; c2<-0; c3<-0; c4<-0; c5<-0; c6<-0 ; c7<-0 ; c8<-0 # A3-PR-GS
d1<-0; d2<-0; d3<-0; d4<-0; d5<-0; d6<-0 ; d7<-0 ; d8<-0 # A4-PR-GS
# ALGORITMOS CON HEMOGRAMA
e1<-0; e2<-0; e3<-0; e4<-0; e5<-0; e6<-0 ; e7<-0 ; e8<-0 # AH1-PR-GS
f1<-0; f2<-0; f3<-0; f4<-0; f5<-0; f6<-0 ; f7<-0 ; f8<-0 # AH2-PR-GS
g1<-0; g2<-0; g3<-0; g4<-0; g5<-0; g6<-0 ; g7<-0 ; g8<-0 # AH3-PR-GS
h1<-0; h2<-0; h3<-0; h4<-0; h5<-0; h6<-0 ; h7<-0 ; h8<-0 # AH4-PR-GS
##### ESTIMACION ALGORITMOS SOLOS
A1s1<-0;A1s2<-0;A1e1<-0;A1e2<-0
A2s1<-0;A2s2<-0;A2e1<-0;A2e2<-0
A3s1<-0;A3s2<-0;A3e1<-0;A3e2<-0
A4s1<-0;A4s2<-0;A4e1<-0;A4e2<-0

```

```
##### ESTIMACION ALGORITMOS CON HEMOGRAMA
```

```
AH1s1<-0;AH1s2<-0;AH1e1<-0;AH1e2<-0
```

```
AH2s1<-0;AH2s2<-0;AH2e1<-0;AH2e2<-0
```

```
AH3s1<-0;AH3s2<-0;AH3e1<-0;AH3e2<-0
```

```
AH4s1<-0;AH4s2<-0;AH4e1<-0;AH4e2<-0
```

```
##### PARALELO
```

```
A1s3<-0; A1e3<-0; A2s3<-0; A2e3<-0; A3s3<-0; A3e3<-0; A4s3<-0; A4e3<-0
```

```
AH1s3<-0; AH1e3<-0; AH2s3<-0; AH2e3<-0; AH3s3<-0; AH3e3<-0; AH4s3<-0; AH4e3<-0
```

```
A1s4<-0; A1e4<-0; A2s4<-0; A2e4<-0; A3s4<-0; A3e4<-0; A4s4<-0; A4e4<-0
```

```
AH1s4<-0; AH1e4<-0; AH2s4<-0; AH2e4<-0; AH3s4<-0; AH3e4<-0; AH4s4<-0; AH4e4<-0
```

```
##### SERIE
```

```
A1s5<-0; A1e5<-0; A2s5<-0; A2e5<-0; A3s5<-0; A3e5<-0; A4s5<-0; A4e5<-0
```

```
AH1s5<-0; AH1e5<-0; AH2s5<-0; AH2e5<-0; AH3s5<-0; AH3e5<-0; AH4s5<-0; AH4e5<-0
```

```
A1s6<-0; A1e6<-0; A2s6<-0; A2e6<-0; A3s6<-0; A3e6<-0; A4s6<-0; A4e6<-0
```

```
AH1s6<-0; AH1e6<-0; AH2s6<-0; AH2e6<-0; AH3s6<-0; AH3e6<-0; AH4s6<-0; AH4e6<-0
```

```
n<-N
```

```
for(j in 1:bo){
```

```
  for(i in 1:n){
```

```
    ### Gold Standard Positivo
```

```
    if(x[[j]][i,1]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==1){a1=1+a1}
```

```
    if(x[[j]][i,1]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==1){a2=1+a2}
```

```
    if(x[[j]][i,1]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==1){a3=1+a3}
```

```
    if(x[[j]][i,1]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==1){a4=1+a4}
```

```
    if(x[[j]][i,2]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==1){b1=1+b1}
```

```

if(x[[j]][i,2]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {b2=1+b2}
if(x[[j]][i,2]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {b3=1+b3}
if(x[[j]][i,2]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {b4=1+b4}
if(x[[j]][i,3]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {c1=1+c1}
if(x[[j]][i,3]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {c2=1+c2}
if(x[[j]][i,3]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {c3=1+c3}
if(x[[j]][i,3]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {c4=1+c4}
if(x[[j]][i,4]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {d1=1+d1}
if(x[[j]][i,4]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {d2=1+d2}
if(x[[j]][i,4]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {d3=1+d3}
if(x[[j]][i,4]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {d4=1+d4}
if(x[[j]][i,5]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {e1=1+e1}
if(x[[j]][i,5]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {e2=1+e2}
if(x[[j]][i,5]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {e3=1+e3}
if(x[[j]][i,5]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {e4=1+e4}
if(x[[j]][i,6]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {f1=1+f1}
if(x[[j]][i,6]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {f2=1+f2}
if(x[[j]][i,6]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {f3=1+f3}
if(x[[j]][i,6]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {f4=1+f4}
if(x[[j]][i,7]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {g1=1+g1}
if(x[[j]][i,7]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {g2=1+g2}
if(x[[j]][i,7]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {g3=1+g3}
if(x[[j]][i,7]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {g4=1+g4}
if(x[[j]][i,8]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {h1=1+h1}
if(x[[j]][i,8]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {h2=1+h2}
if(x[[j]][i,8]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==1) {h3=1+h3}
if(x[[j]][i,8]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==1) {h4=1+h4}

```

```
### Gold Standard Negativo
```

```

if(x[[j]][i,1]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==0){a5=1+a5}
if(x[[j]][i,1]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==0){a6=1+a6}
if(x[[j]][i,1]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==0){a7=1+a7}
if(x[[j]][i,1]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==0){a8=1+a8}
if(x[[j]][i,2]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==0){b5=1+b5}
if(x[[j]][i,2]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==0){b6=1+b6}
if(x[[j]][i,2]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==0){b7=1+b7}
if(x[[j]][i,2]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==0){b8=1+b8}
if(x[[j]][i,3]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==0){c5=1+c5}
if(x[[j]][i,3]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==0){c6=1+c6}
if(x[[j]][i,3]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==0){c7=1+c7}
if(x[[j]][i,3]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==0){c8=1+c8}
if(x[[j]][i,4]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==0){d5=1+d5}
if(x[[j]][i,4]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==0){d6=1+d6}
if(x[[j]][i,4]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==0){d7=1+d7}
if(x[[j]][i,4]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==0){d8=1+d8}
if(x[[j]][i,5]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==0){e5=1+e5}
if(x[[j]][i,5]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==0){e6=1+e6}
if(x[[j]][i,5]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==0){e7=1+e7}
if(x[[j]][i,5]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==0){e8=1+e8}
if(x[[j]][i,6]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==0){f5=1+f5}
if(x[[j]][i,6]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==0){f6=1+f6}
if(x[[j]][i,6]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==0){f7=1+f7}
if(x[[j]][i,6]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==0){f8=1+f8}
if(x[[j]][i,7]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==0){g5=1+g5}
if(x[[j]][i,7]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==0){g6=1+g6}
if(x[[j]][i,7]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==0){g7=1+g7}
if(x[[j]][i,7]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==0){g8=1+g8}
if(x[[j]][i,8]==1 && x[[j]][i,9]==1 && x[[j]][i,10]==0){h5=1+g5}

```

```

    if(x[[j]][i,8]==1 && x[[j]][i,9]==0 && x[[j]][i,10]==0){h6=1+g6}
    if(x[[j]][i,8]==0 && x[[j]][i,9]==1 && x[[j]][i,10]==0){h7=1+g7}
    if(x[[j]][i,8]==0 && x[[j]][i,9]==0 && x[[j]][i,10]==0){h8=1+g8}
  }

#--- ESTIMACION CLASICA ---#
#ALGORITMOS SIN HEMOGRAMA
A1s1[j]<-(a1+a2)/(a1+a2+a3+a4);A1e1[j]<-(a7+a8)/(a5+a6+a7+a8)
A2s1[j]<-(b1+b2)/(b1+b2+b3+b4);A2e1[j]<-(b7+b8)/(b5+b6+b7+b8)
A3s1[j]<-(c1+c2)/(c1+c2+c3+c4);A3e1[j]<-(c7+c8)/(c5+c6+c7+c8)
A4s1[j]<-(d1+d2)/(d1+d2+d3+d4);A4e1[j]<-(d7+d8)/(d5+d6+d7+d8)
##ALGORITMOS CON HEMOGRAMA
AH1s1[j]<-(e1+e2)/(e1+e2+e3+e4);AH1e1[j]<-(e7+e8)/(e5+e6+e7+e8)
AH2s1[j]<-(f1+f2)/(f1+f2+f3+f4);AH2e1[j]<-(f7+f8)/(f5+f6+f7+f8)
AH3s1[j]<-(g1+g2)/(g1+g2+g3+g4);AH3e1[j]<-(g7+g8)/(g5+g6+g7+g8)
AH4s1[j]<-(h1+h2)/(h1+h2+h3+h4);AH4e1[j]<-(h7+h8)/(h5+h6+h7+h8)
#PRUEBA RAPIDA
A1s2[j]<-(a1+a3)/(a1+a2+a3+a4);A1e2[j]<-(a6+a8)/(a5+a6+a7+a8)

# PRUEBAS APLICADAS EN PARALELO
#ALGORITMOS SIN HEMOGRAMA
A1s3[j]<-(a1+a3)/(a1+a2+a3+a4); A1e3[j]<-(a6+a8)/(a5+a6+a7+a8)
A2s3[j]<-(b1+b3)/(b1+b2+b3+b4); A2e3[j]<-(b6+b8)/(b5+b6+b7+b8)
A3s3[j]<-(c1+c3)/(c1+c2+c3+c4); A3e3[j]<-(c6+c8)/(c5+c6+c7+c8)
A4s3[j]<-(d1+d3)/(d1+d2+d3+d4); A4e3[j]<-(d6+d8)/(d5+d6+d7+d8)
#ALGORITMOS CON HEMOGRAMA
AH1s3[j]<-(e1+e3)/(e1+e2+e3+e4); AH1e3[j]<-(e6+e8)/(e5+e6+e7+e8)
AH2s3[j]<-(f1+f3)/(f1+f2+f3+f4); AH2e3[j]<-(f6+f8)/(f5+f6+f7+f8)
AH3s3[j]<-(g1+g3)/(g1+g2+g3+g4); AH3e3[j]<-(g6+g8)/(g5+g6+g7+g8)

```

```
AH4s3[j]<-(h1+h3)/(h1+h2+h3+h4); AH4e3[j]<-(h6+h8)/(h5+h6+h7+h8)
```

```
### SERIE NEGATIVO
```

```
#ALGORITMOS SIN HEMOGRAMA
```

```
A1s5[j]<-(a1+a2+a3)/(a1+a2+a3+a4); A1e5[j]<-a8/(a5+a6+a7+a8)
```

```
A2s5[j]<-(b1+b2+b3)/(b1+b2+b3+b4); A2e5[j]<-b8/(b5+b6+b7+b8)
```

```
A3s5[j]<-(c1+c2+c3)/(c1+c2+c3+c4); A3e5[j]<-c8/(c5+c6+c7+c8)
```

```
A4s5[j]<-(d1+d2+d3)/(d1+d2+d3+d4); A4e5[j]<-d8/(d5+d6+d7+d8)
```

```
#ALGORITMOS CON HEMOGRAMA
```

```
AH1s5[j]<-(e1+e2+e3)/(e1+e2+e3+e4); AH1e5[j]<-e8/(e5+e6+e7+e8)
```

```
AH2s5[j]<-(f1+f2+f3)/(f1+f2+f3+f4); AH2e5[j]<-f8/(f5+f6+f7+f8)
```

```
AH3s5[j]<-(g1+g2+g3)/(g1+g2+g3+g4); AH3e5[j]<-g8/(g5+g6+g7+g8)
```

```
AH4s5[j]<-(h1+h2+h3)/(h1+h2+h3+h4); AH4e5[j]<-h8/(h5+h6+h7+h8)
```

```
### SERIE POSITIVO
```

```
#ALGORITMOS SIN HEMOGRAMA
```

```
A1s6[j]<-(a1)/(a1+a2+a3+a4); A1e6[j]<-(a6+a7+a8)/(a5+a6+a7+a8)
```

```
A2s6[j]<-(b1)/(b1+b2+b3+b4); A2e6[j]<-(b6+b7+b8)/(b5+b6+b7+b8)
```

```
A3s6[j]<-(c1)/(c1+c2+c3+c4); A3e6[j]<-(c6+c7+c8)/(c5+c6+c7+c8)
```

```
A4s6[j]<-(d1)/(d1+d2+d3+d4); A4e6[j]<-(d6+d7+d8)/(d5+d6+d7+d8)
```

```
#ALGORITMOS CON HEMOGRAMA
```

```
AH1s6[j]<-(e1)/(e1+e2+e3+e4); AH1e6[j]<-(e6+e7+e8)/(e5+e6+e7+e8)
```

```
AH2s6[j]<-(f1)/(f1+f2+f3+f4); AH2e6[j]<-(f6+f7+f8)/(f5+f6+f7+f8)
```

```
AH3s6[j]<-(g1)/(g1+g2+g3+g4); AH3e6[j]<-(g6+g7+g8)/(g5+g6+g7+g8)
```

```
AH4s6[j]<-(h1)/(h1+h2+h3+h4); AH4e6[j]<-(h6+h7+h8)/(h5+h6+h7+h8)
```

```
print(c(j))
```

```
}
```

```
estim<-rbind(A1s1,A1s2,A1e1,A1e2,A2s1,A2s2,A2e1,A2e2,
```

```

      A3s1,A3s2,A3e1,A3e2,A4s1,A4s2,A4e1,A4e2,
      AH1s1,AH1s2,AH1e1,AH1e2,AH2s1,AH2s2,AH2e1,
      AH2e2,AH3s1,AH3s2,AH3e1,AH3e2,AH4s1,AH4s2,
      AH4e1,AH4e2,A1s3,A1e3,A2s3,A2e3,A3s3,A3e3,A4s3,
      A4e3,AH1s3,AH1e3,AH2s3,AH2e3,AH3s3,AH3e3,
      AH4s3,AH4e3,A1s5,A1e5,A2s5,A2e5,A3s5,A3e5,A4s5,A4e5,
      AH1s5,AH1e5,AH2s5,AH2e5,AH3s5,AH3e5,AH4s5,AH4e5,
      A1s6,A1e6,A2s6,A2e6,A3s6,A3e6,A4s6,A4e6,AH1s6,AH1e6,
      AH2s6,AH2e6,AH3s6, AH3e6,AH4s6,AH4e6)

return(estim)
}

##### datos a la funcion bootstrap #####
setwd("C:/Users/Usuario/Dropbox/TESIS/CODIGOS R/CODIGO TESIS/GS 0.9/PR - 0")
getwd()

s<-2000 ## Veces del bootstrap
x<-read.delim("basefinal.txt", sep = " " ) ## base a remuestrear
BOOS2000<-bootstrap(y=x,bo=2000) # funcion resultado.

```

Bibliography

- [Becker et al., 2001] Becker, K. et al. (2001). Interpretación del hemograma. *Revista chilena de pediatría*, 72(5):460–465.
- [Behar, 2015] Behar, R. (2015). *Notas sobre el remuestreo(Bootstrap)*. Universidad del valle, escuela de estadística, Santiago de Cali.
- [Bernardo and Smith, 2001] Bernardo, J. M. and Smith, A. F. (2001). Bayesian theory.
- [Box and Tiao, 1992] Box, G. E. and Tiao, G. C. (1992). *Bayesian inference in statistical analysis*, volume 40. John Wiley & Sons.
- [Brady et al., 2012] Brady, O., Gething, P., Bhatt, S., Messina, J., Brownstein, J., Hoen, A., Moyes, C., Farlow, A., Scott, T., and Hay, S. (2012). Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS neglected tropical diseases*, 6(8):e1760.
- [Bravo-Grau and Cruz, 2015] Bravo-Grau, S. and Cruz, J. P. (2015). Estudios de exactitud diagnóstica: Herramientas para su interpretación. *Revista chilena de radiología*, 21(4):158–164.
- [Bú, 1996] Bú, R. C. (1996). *Simulación: un enfoque práctico*. Editorial Limusa.
- [Donis, 2012] Donis, J. H. (2012). Evaluación de la validez y confiabilidad de una prueba diagnóstica (assessment of the validity and reliability of a diagnostic test).
- [Emrich and Piedmonte, 1991] Emrich, L. J. and Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *The American Statistician*, 45(4):302–304.

- [Escrig-Sos et al., 2006] Escrig-Sos, J., Martínez-Ramos, D., and Miralles-Tena, J. M. (2006). Pruebas diagnósticas: nociones básicas para su correcta interpretación y uso. *Cirugía Española*, 79(5):267–273.
- [Fernández et al., 2016] Fernández, J., Cárdenas, L., Hernández, J., Palacio, L., and Castañeda, C. (2016). Exploratory wavelet analysis of dengue seasonal patterns in colombia. *Biomédica*, 36:44–55.
- [Gelman et al., 2014] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. CRC press Boca Raton, FL.
- [Hess et al., 2012] Hess, A., Shardell, M., Johnson, J., Thom, K., Strassle, P., Netzer, G., and Harris, A. (2012). Methods and recommendations for evaluating and reporting a new diagnostic test. *European journal of clinical microbiology & infectious diseases*, 31(9):2111–2116.
- [Joseph et al., 1995] Joseph, L., Gyorkos, T. W., and Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American journal of epidemiology*, 141(3):263–272.
- [Kyle and Harris, 2008] Kyle, J. and Harris, E. (2008). Global spread and persistence of dengue. *Annu. Rev. Microbiol.*, 62:71–92.
- [Ledesma, 2008] Ledesma, R. (2008). Introduccción al bootstrap. desarrollo de un ejemplo acompañado de software de aplicación. *Tutorials in quantitative methods for psychology*, 4(2):51–60.
- [Low et al., 2011] Low, J., Ong, A., Tan, L., Chaterji, S., Chow, A., Lim, W., Lee, K., Chua, R., Chua, C., Tan, S., et al. (2011). The early clinical features of dengue in adults: challenges for early clinical diagnosis. *PLoS neglected tropical diseases*, 5(5):e1191.
- [Mendez, 2017] Mendez, A. (2017). *Propuesta metodológica para construir reglas de clasificación - caso de aplicación Dengue*. Tesis de pregrado, Universidad del Valle.
- [Morales and Zárate, 2004] Morales, A. R. and Zárate, L. E. M. (2004). *Epidemiología clínica: investigación clínica aplicada*. Ed. Médica Panamericana.

- [OMS, 2009] OMS, T. (2009). Dengue: guías para el diagnóstico, tratamiento, prevención y control. 2009.1. *Bolivia: OMS*.
- [OPS and OMS,] OPS and OMS. Dengue: mapas, datos y estadísticas. *OPS*, URL: https://www.paho.org/hq/index.php?option=com_topicsview = rdmored = 3274lte = 4073lang = eslang = es.
- [Osorio et al., 2010] Osorio, L., Ramirez, M., Bonelo, A., Villar, L. A., and Parra, B. (2010). Comparison of the diagnostic accuracy of commercial ns1-based diagnostic tests for early dengue infection. *Virology journal*, 7(1):361.
- [Padilla et al., 2012] Padilla, J. C., Rojas, D. P., and Gómez, R. S. (2012). *Dengue en Colombia: epidemiología de la reemergencia a la hiperendemia*. Guías de Impresión Ltda.
- [Park et al., 1996] Park, C. G., Park, T., and Shin, D. W. (1996). A simple method for generating correlated binary variates. *The American Statistician*, 50(4):306–310.
- [Podgorelec et al., 2002] Podgorelec, V., Kokol, P., Stiglic, B., and Rozman, I. (2002). Decision trees: an overview and their use in medicine. *Journal of medical systems*, 26(5):445–463.
- [R Development Core Team, 2007] R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- [Rudner and Shafer, 1992] Rudner, L. M. and Shafer, M. M. (1992). Resampling: a marriage of computers and statistics. *Practical Assessment, Research & Evaluation*, 3(5):3.
- [Salech et al., 2008] Salech, F., Mery, V., Larrondo, F., and Rada, G. (2008). Estudios que evalúan un test diagnóstico: interpretando sus resultados. *Revista médica de Chile*, 136(9):1208–1208.
- [Tovar, 2012] Tovar, J. R. (2012). Eliciting beta prior distributions for binomial sampling. *Rev. Bras. Biom*, 30(1):159–172.
- [Wackerly and Scheaffer, 2010] Wackerly, D. D. M. and Scheaffer, W. (2010). *Estadística matemática con aplicaciones*. Number 519.5 W3.

- [Yumpaqui Pacheco, 2005] Yumpaqui Pacheco, R. M. (2005). *Introducción a la estadística bayesiana (caso comparación de medias)*. Trabajo monográfico, Universidad Nacional Mayor de San Marcos.
- [Zhu and Lu, 2004] Zhu, M. and Lu, A. Y. (2004). The counter-intuitive non-informative prior for the bernoulli family. *Journal of Statistics Education*, 12(2).