

043/UB/VEG/opt

DEPARTAMENTO DE ESTADÍSTICA DE LA UNIVERSIDAD DE BARCELONA

PROGRAMA DE PROBABILITATS I ESTADÍSTICA (BIENNIO 87/89)

**Optimización
en estudios de Monte Carlo
en Estadística: Aplicaciones al
Contraste de Hipótesis.**

Esteban Vegas Lozano

Departamento de Estadística
Facultad de Biología
Universidad de Barcelona

DEPARTAMENTO DE ESTADÍSTICA DE LA UNIVERSIDAD DE BARCELONA

PROGRAMA DE PROBABILITATS I ESTADÍSTICA (BIENNIO 87/89)

Optimización en estudios de Monte Carlo en Estadística: Aplicaciones al Contraste de Hipótesis.

Memoria presentada para optar
al título de Doctor en Biología
por la Universidad de Barcelona,
por

Esteban Vegas Lozano

Visto bueno
El Director,

Dr. Jordi Ocaña Rebull,
Catedrático de Universidad.
Departamento de Estadística.
Universidad de Barcelona.

Barcelona, 20 de Junio de 1996

Prólogo

Esta memoria se enmarca dentro de lo que se denomina Simulación Estadística y, por consiguiente, incluye herramientas de Estadística Computacional.

Se divide en cinco bloques. Un primer bloque de introducción general a la estadística computacional y a la supercomputación (Capítulo 1), que se amplía con las técnicas de reducción de la varianza (Capítulo 2). Un segundo bloque en el que se presenta una técnica de optimización en estudios de Monte Carlo en estadística asociado a variables dicotómicas en el que se estudia sus propiedades teóricamente y se ilustra y valida por simulación utilizando técnicas vectoriales para su implementación (Capítulo 3). En el tercer bloque se desarrolla un nuevo test, geodésico, para el problema de Behrens-Fisher. Además se estudia, aplicando la anterior técnica de optimización, la potencia de este test en comparación con la solución más habitual, en la práctica, para el problema de Behrens-Fisher, el test de Cochran y Cox (Capítulo 4). A continuación se estudia por simulación la robustez del test geodésico y del test de Cochran y Cox para dos situaciones bastantes plausibles, bajo distribución lognormal y bajo distribución gamma (Capítulo 5). De nuevo, se utilizó la técnica de optimización. En el cuarto bloque se estima las verdaderas probabilidades de error, usando la técnica de optimización, de dos casos reales del problema de Behrens-Fisher en el ámbito de la experimentación médico-biológica en los cuales se ignoró la diferencia de varianzas e incluso en uno de ellos, se rechaza la hipótesis de normalidad (Capítulo 6). En el quinto y último bloque se presentan las conclusiones (Capítulo 7) y se incluye tres apéndices.

Cualquier trabajo intelectual no es solamente atribuible al esfuerzo del que lo presenta, hay muchas personas que han ayudado, de una manera u otra, a llevar a buen puerto, en este caso, la presente memoria. Espero no olvidar a ninguna.

He de agradecer, en primer término, al Dr. Jordi Ocaña que me dirigiera la tesis con todo lo que ello implica de dedicación y esfuerzo.

También he de agradecer la inestimable ayuda del Dr. José María Oller y del Dr. José Manuel Corcuera en el ámbito en el que ellos son unos especialistas, las distancias geodésicas y al Dr. Joan del Castillo por sus apreciaciones sobre la familia exponencial curvada.

No puede faltar el agradecimiento a todos mis compañeros/as del departa-

mento de Estadística por su colaboración y apoyo, en especial a mi compañero de caverna, Dr. Francesc Oliva.

Tampoco he de olvidar a dos compañeros de la carrera de Biología, Joan Saldaña y Ricard Vicente, que nos une una inquietud común, la biología teórica. Con ellos he disfrutado de unas tertulias nocturnas muy *reconfortantes*.

Finalmente, y de forma muy especial, he de poner en relieve la comprensión y la ayuda prestada por María José Segura gracias a la cual la carga del trabajo ha sido más liviana.

Índice General

1	ESTADÍSTICA COMPUTACIONAL Y SUPERCOMPUTACIÓN	1
1.1	Estadística computacional	3
1.1.1	Introducción	3
1.1.2	Apuntes históricos de la estadística computacional y algunas aportaciones del departamento de estadística de la Universidad de Barcelona	4
1.1.3	Relevancia de la estadística computacional	7
1.1.4	Posibles errores que se pueden cometer en estadística computacional	9
1.2	Supercomputación	12
1.2.1	Introducción	12
1.2.2	Evolución de las arquitecturas de los ordenadores	14
1.2.3	Procesador “pipeline”	16
1.2.4	Procesador vectorial	18
1.2.5	Procesador paralelo	20
1.2.6	Programación vectorial	21
1.3	Futuro de la estadística computacional y de la supercomputación	24
2	TÉCNICAS DE REDUCCIÓN DE LA VARIANZA: VARIABLES DE CONTROL Y MONTE CARLO CONDICIONAL	27

2.1	Introducción	29
2.2	Variables de Control	32
2.2.1	Variables de Control Simple	33
2.2.2	Variables de Control Ajustadas por Regresión	34
2.2.3	Ejemplo de aplicación de la técnica de variables de control	35
2.3	Muestreo condicional o Monte Carlo condicional	36
2.3.1	Ejemplo de aplicación de la técnica Monte Carlo condicional	38
3	REDUCCIÓN DE LA VARIANZA PARA VARIABLES DE RESPUESTA DICOTÓMICA EN SIMULACIÓN	41
3.1	Introducción	43
3.2	Relación del estimador máximo verosímil, $\tilde{p}_{1.}$, con otros métodos	45
3.2.1	Estimador de variable de control ajustado por regresión	45
3.2.2	Estimador basado en la esperanza de Y condicionada a C	46
3.2.3	Estimador basado en la curva de regresión de la media de Y versus C	47
3.2.4	Estimador basado en la esperanza de \bar{Y} condicionada a \bar{C}	47
3.2.5	Estimador basado en la curva de regresión de la media de \bar{Y} versus \bar{C}	49
3.2.6	Mínimos cuadrados ordinarios de Y_i/C_i sabiendo que $p_{.1}$ es conocido	49
3.3	Propiedades del nuevo estimador	50
3.3.1	Sesgo	51
3.3.2	Varianza	51
3.3.3	Eficiencia	53
3.3.4	Otras propiedades	57

3.4	Estimadores de la varianza del nuevo estimador	57
3.4.1	Estimador propuesto por Rothery de $var(\tilde{p}_1)$	58
3.4.2	Estimación insesgada de $var(\tilde{p}_1)$	60
3.4.3	Estimación aproximada al estimador insesgado de $var(\tilde{p}_1)$	60
3.4.4	Estimación bootstrap de $var(\tilde{p}_1)$	61
3.4.5	Estimación de la reducción de varianza y del margen de error	61
3.5	Estudio de simulación	62
3.5.1	Un caso de estudio: potencia de un nuevo test	62
3.5.2	Un estudio de simulación general	66
3.5.3	Conclusión	67
4	REDUCCIÓN DE LA VARIANZA EN EL ESTUDIO DE UN NUEVO TEST PARA EL PROBLEMA DE BEHRENS-FISHER	71
4.1	El problema de Behrens-Fisher	73
4.1.1	Introducción	73
4.1.2	Soluciones prácticas	76
4.2	Obtención del nuevo test	80
4.2.1	Planteamiento	80
4.2.2	Deducción del test	82
4.3	Estudio de simulación	84
4.3.1	Procedimiento	84
4.3.2	Implementación	88
4.3.3	Algoritmos vectoriales	88
4.3.4	Resultados	92
4.3.5	Gráficos	103

4.3.6	Conclusiones	110
5	ESTUDIO SOBRE LA ROBUSTEZ DEL TEST GEODÉSICO Y DEL TEST COCHRAN Y COX	117
5.1	Estudio de la robustez para muestras originadas bajo la distribu- ción lognormal	119
5.1.1	Introducción	119
5.1.2	Procedimiento	120
5.1.3	Resultados	121
5.1.4	Gráficos	126
5.1.5	Conclusiones	133
5.2	Diseño del estudio por simulación de la robustez para muestras originadas bajo la distribución gamma	139
5.2.1	Introducción	139
5.2.2	Procedimiento	139
5.2.3	Resultados	143
5.2.4	Gráficos	146
5.2.5	Conclusiones	150
6	ESTUDIOS DE SIMULACIÓN SOBRE ALGUNOS EJEM- PLOS DEL PROBLEMA DE BEHRENS-FISHER EN EL ÁMBITO DE LA EXPERIMENTACIÓN MÉDICO-BIOLÓGICA	153
6.1	Un estudio sobre la trombosis	155
6.1.1	Comparación de medias entre los dos tipos de pacientes . .	156
6.1.2	Estimación de la verdadera probabilidad de error de tipo I	157
6.2	Un estudio sobre la diabetes	165
6.2.1	Comparación de medias entre el grupo Normal y el diabetes Alloxan tratado con Insulina	166

6.2.2	Estudio de simulación para estimar la verdadera probabilidad de error alrededor de los valores muestrales	167
6.2.3	Resultados y conclusiones para la estimación de la verdadera probabilidad de error de tipo I y II cuando se ha realizado el logaritmo neperiano de los datos originales	180
7	CONCLUSIONES	189
7.1	Conclusiones y resumen de resultados	191
A	Obtención de μ_2 a partir del nomograma	199
B	Distancia de Rao	201
C	Bootstrap paramétrico	207
	Bibliografía	211

Capítulo 1

ESTADÍSTICA COMPUTACIONAL Y SUPERCOMPUTACIÓN

Se hace una breve exposición de la importancia de la simulación, haciendo hincapié en la estadística computacional. Esta es una herramienta básica, cada vez más empleada, en el avance de muchas áreas de la ciencia. Su crecimiento está en correspondencia con el crecimiento de la informática, de forma que no se puede entender la evolución de la primera sin considerar la de la segunda. Esta mejora de la informática a dado lugar a la aparición de nuevas arquitecturas en los ordenadores que han propiciado el desarrollo de las llamadas técnicas de supercomputación: paralelización y vectorización. Ambas, son cada día más necesarias en la realización de las simulaciones, sobre todo si son a tiempo real con mucha cantidad de información. Se mostrará como cambia la forma de entender los algoritmos en el caso de procesamiento vectorial. Por último se hará una reflexión sobre lo que nos depara el futuro de la estadística computacional y la supercomputación.

1.1 Estadística computacional

1.1.1 Introducción

La utilización de la simulación cada día es más frecuente en muchos y muy variados ámbitos. Se emplea, habitualmente, para imitar explícitamente el comportamiento de modelos, por ejemplo: poblaciones animales (Saunders y Tweedie (1976), Gibbs (1980), ...), epidemias (Bailey (1975), ...), etc. En todos estos casos la estadística está presente como pieza fundamental en la implementación de la simulación por ordenador y en el análisis de los resultados, es lo que se llama "estadística computacional": generación de variables aleatorias, método de Monte Carlo, técnicas de reducción de la varianza, etc.

Por otro lado, también se puede usar la simulación para evaluar el comportamiento de variables aleatorias complicadas cuya distribución exacta no se pueda encontrar matemáticamente, o más en general, el estudio del comportamiento de cualquier variable aleatoria. En este caso, lo que se simula es de interés estadístico y, por descontado, también se utilizan las técnicas de estadística computacional

antes mencionadas y otras técnicas más recientes llamadas de remuestreo (especialmente bootstrap y jackknife). A este tipo de simulación se la denomina “simulación estadística” y es donde se enmarca la presente memoria.

1.1.2 Apuntes históricos de la estadística computacional y algunas aportaciones del departamento de estadística de la Universidad de Barcelona

El uso del método de Monte Carlo se remonta a una de las primeras aplicaciones de los ordenadores. John v. Neumann, uno de los padres del ordenador, lo llamaba “el experimento estadístico” y sirvió para determinar el comportamiento de la bomba atómica. El posterior desarrollo en los métodos de Monte Carlo se hace patente en el número de conferencias en la década de los años 40 y en una monografía de Hammersley y Handscomb la cual se publica en Londres, en 1964.

Un importante paso en el crecimiento de la estadística computacional se produjo con el desarrollo de sistemas de simulación, un modelo más general de experimentación estadística. A los lenguajes de programación de uso más general les faltaban algunas herramientas para facilitar la programación de la simulación. Estas facilidades entran dentro de algunas de estas cinco categorías:

Manipulación de entidades. Creación de nuevas entidades, destrucción de las viejas, inclusión de unas entidades dentro de otras, ordenar o desordenar conjuntos (pilas, colas, etc).

Manipulación del tiempo y sucesos. Mantenimiento de un reloj y una lista de sucesos, y ejecución de las operaciones de planificación: Insertar sucesos, eliminarlos,

Números aleatorios. Implementación de algoritmos para generar distribuciones continuas y discretas, así como la capacidad para reproducir o/y cambiar de secuencia de números aleatorios entre diferentes ejecuciones del mismo programa.

Repertorio de estadísticos y gráficos. Inclusión de un grupo de estadísticos comunes (medias, varianzas, etc) y gráficos (histogramas, etc) para facilitar el análisis de los datos obtenidos de las simulaciones, tanto de una porción como de todo el proceso.

Computación numérica. Se requiere capacidad para implementar algoritmos numéricos. Por ejemplo, en el análisis estadístico de los resultados de las simulaciones y en la construcción de generadores de números aleatorios especiales.

Además, un buen lenguaje de simulación debe poseer todas las cualidades que son deseables en cualquier lenguaje de programación tales como potencia, flexibilidad, sencillez, modularidad, etc. Estos productos comienzan a presentarse en la década de los años sesenta. Como muestra solo se presentarán tres de los que tienen más renombre:

- Geoffrey Gordon desarrolló el GPSS (General Purpose Simulation System) (ver Gordon 1975, 1978) que estuvo disponible en 1961. Actualmente es un producto de IBM.
- Kristen Nygaard y Ole-Johan Dahl, en 1963, presentaron el lenguaje de simulación SIMULA que está basado en el Algol (ver Dahl and Nygaard, 1966; Birtwistle *et al.*, 1973). Al margen de su interés como lenguaje de simulación, tiene una importancia capital en la historia de la informática al tratarse del primer lenguaje en el que se implementó el paradigma de la programación orientada a objetos.
- En la corporación RAND se desarrolla el SIMSCRIPT que aparece a la luz sobre 1965 (Karr, Kleine and Markowitz, 1965; Kiviat, Villanueva and Markovitz, 1969, 1973). Se basa en el lenguaje FORTRAN.

Estos tres lenguajes han llegado a nuestros días con nuevas versiones más optimizadas. Sin embargo, la evolución de los sistemas de simulación ha ido tendiendo a crear nuevos productos muchos más específicos. Así, hay productos para simular redes de colas ("queuing networks"): RESQME, Q-PAS, ...; para simular aplicaciones de manufacturación: XCELL+, WITNESS ATT, ... y muchas cosas más. Dentro de esta línea, el departamento de Bioestadística de la Universidad de Barcelona creó en los años 80 un lenguaje de simulación específico para el entorno biológico, en concreto para genética y ecología, que se denominó ECOGEN (Ocaña, Alonso y Ruiz de Villa, 1986). Además, se creó un software llamado ISSP (Interactive Support Software Package) (Guigó y Ocaña, 1988; Guigó, 1988) con la misión de ayudar en el diseño y planificación de cualquier simulación, es decir,

automatizar la documentación, la definición interactiva de modelos, etc. pasos previos en el desarrollo de una simulación a partir de un paquete de simulación.

La mayoría de personas que se dedicaban a la simulación o tenían interés en analizar unos datos no eran estadísticos, igual que pasa actualmente, entonces las empresas de informática tenían un mercado potencial muy importante deseoso de unos programas que facilitaran el análisis estadístico como también la resolución de los problemas de cálculo numérico. Por esta razón, casi al mismo tiempo que surgieron los sistemas de simulación, aparecieron en Estados Unidos las primeras librerías de programas que incluían programas de cálculo numérico y estadística. Una de las primeras y más conocidas era las SSP (Scientific Subroutines Package) realizadas en FORTRAN (y posteriormente en PL/1) y desarrolladas por IBM. Más adelante se crearon empresas dedicadas exclusivamente a la producción de software dando origen a librerías tan comunes actualmente como las NAG y el IMSL. Con el paso del tiempo y con el desarrollo de la estadística y de los ordenadores, se produjo un gran incremento en número y variedad de software en diferentes temas entre los que la estadística no podía faltar (ver Koch and Haag(1993)):

- Se crearon programas de uso muy general para el análisis de datos estadísticos denominados genericamente paquetes estadísticos, por ejemplo, STATGRAPHICS, SAS, BMDP, SPSS, SYSTAT así como de uso más restringidos como GLIM, LISREL, MULTICUA. Este último ha sido desarrollado por miembros del departamento de Estadística de la Universidad de Barcelona (Arenas, Cuadras y Fortiana, 1991) y está pensado para estadística multivariante.
- Incluso los paquetes no estadísticos como Mathematica, Derive, Maple y otros sistemas algebraicos contienen algoritmos para la generación variables aleatorias y un gran número de funciones estadísticas.
- Con la creación de nuevos lenguajes de programación en los años 70 apareció el lenguaje S, de uso específicamente estadístico. Es decir, pensado para el análisis de datos, principalmente diseñado por estadísticos. Hoy es uno de los más conocidos y que tienen un futuro más prometedor (Venables and Ripley, 1994). Actualmente, un grupo del departamento de Estadística de la Universidad de Barcelona está desarrollando un sistema denominado EMSS (Environment for Montecarlo Simulation Studies) (Sánchez, Ocaña

y Ruíz de Villa, 1992) que es un entorno de desarrollo integrado que pretende ayudar al usuario a escribir, ejecutar y analizar programas para realizar estudios de simulación de Monte Carlo. Combina aspectos de un típico paquete estadístico (basado en menús) y de un lenguaje estadístico (potencia y eficiencia).

- Con el desarrollo de la inteligencia artificial y de los sistemas expertos se desencadenó un interés en introducir también la estadística (Hand, 1986). Al mismo tiempo, también se comenzaron a desarrollar bases de datos estadísticas (Michalewicz, 1990).
- Otro entorno que ha crecido especialmente en los últimos años son los gráficos estadísticos, principalmente a partir del libro *Exploratory Data Analysis*, de Tukey (1977). Ha abarcado a todos los campos de la estadística: modelos lineales, series temporales, multivariante, procesos de control, etc. Esto ha provocado que las compañías de software desarrollen en sus productos nuevos programas o módulos específicos para este propósito, por ejemplo SAS/INSIGHT.

En poco tiempo se llegó a un nivel tal de usuarios y de investigadores dentro del entorno de la estadística y los ordenadores que promovió la publicación de revistas especializadas, como por ejemplo: *Computational Statistics and Data Analysis*, *Mathematics and Computers in Simulation*, *Communications in Statistics, Part B: Simulation and Computation*, *Journal of Statistical Computational and Simulation* y *International Journal Computational Statistical*. Por otro lado, se organizaron también foros internacionales. En el Institute of Statistics de Viena se realizó en 1974 una conferencia de "Estadística Computacional" que dio origen a los actuales congresos bianuales denominados "COMPSTAT". Su marco de actuación abarca todo aquello que se pueda incluir dentro de la metodología de estadística computacional: probabilidad computacional, clasificación automática, aspectos numéricos de los métodos estadísticos, simulación y paquetes de software. Este año, 1996, es Barcelona el lugar elegido como sede del COMPSTAT.

1.1.3 Relevancia de la estadística computacional

Desde sus inicios, la década de los años sesenta, hasta hoy la estadística computacional ha ido creciendo y ampliando su ámbito de influencia.

En física de partículas el método de Monte Carlo se ha convertido en una pieza indispensable para el estudio del comportamiento de la materia. Los ordenadores más potentes son necesarios para simulaciones con un gran número de dimensiones. También sucede algo parecido en el campo de la química-física y la meteorología.

Este método también ha sido importante en el avance de la matemática pura: se usa como única técnica práctica para el cálculo de integrales múltiples de orden muy elevado (por ejemplo, por muestreo de Gibbs). Estadísticamente las secuencias calculadas convergen con probabilidad uno al valor correcto de esta integral.

Posteriormente se empezaron a emplear los métodos de remuestreo. Quenouille fué el primero que propuso el método pero Tukey lo introdujo dentro del ambiente computacional y le dio este nombre. Tuvo que pasar algún tiempo hasta que la utilización del ordenador se hizo más masiva para que tuviera más adeptos. Hoy, engloba a varias técnicas como el jackknife y el bootstrap (Efron, 1982). Esta última, se suele usar para evaluar el comportamiento de variables aleatorias mediante simulación estadística.

La filosofía de la simulación estadística, especialmente la inferencia, se basa en principios semejantes a los del análisis estadístico clásico. Así, por ejemplo, en la estimación de un parámetro sobre diferentes bases de datos suele dar valores típicamente diferentes. Un análisis estadístico clásico de una de tales estimaciones para una base de datos concreta comienza a partir de un modelo en el que se haga constar una medida de probabilidad para el conjunto de todos los estados concebibles de estas bases de datos. A partir de este modelo se deduce una distribución de probabilidad teórica que describe la variación de las estimaciones. Esto permite realizar afirmaciones acerca de la distancia de verdadero valor del parámetro al de la estimación con una cierta probabilidad.

La nueva contribución a la filosofía de la inferencia es que tales conclusiones se pueden efectuar como resultado de experimentos realizados dentro de la misma base de datos. Con muchos menos detalles acerca del modelo se pueden extraer subconjuntos de datos o muestras aleatorias a partir de un muestreo y estudiar sus estimaciones. Esto significa que la investigación teórica para una clase de medida de probabilidad sobre el espacio muestral completo ha sido reemplazado por variaciones sobre la propia base de datos. Esto sólo puede ser eficiente con

un ordenador (excepto que la base de datos sea muy pequeña) y puesto que los experimentos serán repetidos muchas veces, estos métodos necesitan computación intensiva. Se advierte que en ambos casos, clásica y nueva filosofía, el resultado final se basa enteramente en la base de datos dada. Esta es la única información disponible para realizar el análisis estadístico.

En 1964 Huber, en un artículo seminal sobre estimaciones robustas de los parámetros de localización, dio origen a lo que hoy se llama los “Métodos Robustos” en estadística. Son similares a las técnicas de remuestreo y necesitan de un gran esfuerzo computacional. Para más detalles (Huber, 1964; Stahel and Weisberg, 1991).

Otra aplicación de la computación intensiva es la clasificación automática. Hoy muchos de los más frecuentes métodos de clusters son disponibles en paquetes estándar (Whishart, 1978).

Hay también un número de nuevas teorías o campos, intrínsecamente orientadas al ordenador, las cuales usan los conceptos de estadística o conceptos relacionados con el razonamiento estadístico. Los conjuntos borrosos y la inferencia borrosa tienen connotaciones estadísticas y pueden ser usados en entornos donde el razonamiento estadístico es importante para los conceptos de borrosidad. De manera similar los algoritmos genéticos (Schaffer, 1989) están relacionados con una aproximación estocástica. Los modelos caóticos son especialmente interesantes al mostrar como aparentemente el comportamiento caótico puede ser gobernado por unas reglas y ecuaciones deterministas simples. Así, se generalizan las ideas que subyacen en la generación de números aleatorios y sirven para el estudio de la “planificación” de la aleatoriedad.

1.1.4 Posibles errores que se pueden cometer en estadística computacional

Al principio de la corta historia de la estadística y el ordenador muchos estadísticos estaban en contra de la utilización de la simulación como solución a los problemas. Insistían en el desarrollo de soluciones analíticas. Ahora, con la disponibilidad y potencia de los ordenadores, es muy común realizar simulaciones como forma de resolver algunos problemas aunque se debe de preferir, siempre que sea posible, una solución analítica que la obtenida por simulación aunque esta

última puede servir como corroboración y/o ilustración de la solución analítica. Debido a este auge y a la complejidad de los problemas a resolver, muchas veces, por ignorancia o descuido, se cometen errores que pueden invalidar los resultados obtenidos en la simulación.

Uno de los errores posibles reside en la generación de los números pseudoaleatorios, pieza fundamental para cualquier simulación. Una persona que intente realizar una simulación debe saber que estos números no son propiamente aleatorios, es decir, son unos números que se obtienen de un procedimiento totalmente determinista pero que sus valores tienen apariencia de ser aleatorios. Por esta razón, su principal ventaja es que se puede repetir la misma serie de valores aleatorios de simulación en simulación. Pero una desventaja importante es la existencia de una periodicidad, es decir, que se llega a un momento en que la serie de valores pseudoaleatorios generada se repite. En la mayoría de los casos es al cabo de un gran número de valores. También, según el algoritmo, puede existir un cierto patrón en la generación de los valores aleatorios. Para observar su presencia se suele pasar unos test estadísticos de aleatoriedad o hacer gráficos bidimensionales donde se marca los sucesivos pares de valores generados (o pares de valores a una distancia fija de K generaciones) y se observa si existe la presencia de alguna forma. Por tanto, es muy importante conocer los algoritmos de generación de números aleatorios que utiliza nuestro software y verificar, antes de utilizarlos, su aleatoriedad.

La pérdida de comprensión de los experimentos simulados se convierte particularmente preocupante para sistemas complejos. Estos son simulaciones de sistemas muy grandes caracterizados por interacciones locales como pasa en los modelos para el estudio de los gases. La clave de estas simulaciones está en el cambio de posición o estado de una partícula en el tiempo. Esta idea fue reconocida por Metropolis *et al.* en 1953 quien propuso simular cadenas de Markov que tengan la distribución deseada como su distribución de equilibrio. Se han desarrollado varios algoritmos (como por ejemplo el de Metropolis) para expresar estas ideas pero, a diferencia de los modelos más simples (generación de diferentes tipos de variables aleatorias bajo un esquema de muestreo aleatorio simple), no existe un buen conocimiento analítico de sus propiedades. Esto puede desencadenar ciertos problemas en los resultados o en el proceso de la simulación. Por ejemplo, ¿cómo saber cuando el algoritmo de Metropolis ha sido ejecutado el número suficiente de veces?

Otros problemas se dan en el análisis de resultados de las simulaciones. En los modelos simulados que van evolucionando a través del tiempo, la obvia autocorrelación de los resultados se ignora comunmente cuando se realiza su análisis en la mayoría de paquetes y lenguajes de simulación estándar. La esencia del problema se basa en tener una serie Z_1, \dots, Z_T y se desea estimar la esperanza $E(Z_T)$. Se supone que $\mu_i = E(Z_i) \rightarrow \mu$ y que después de un tránsito inicial la serie se convierte en estacionaria. Partiendo de la suposición de que el tránsito inicial ha sido descartado. Si se ignora la correlación de $\hat{\mu} = \bar{Z}$ su varianza es σ^2/T y se estima por s^2/T . Sin embargo, con autocorrelación positiva, $E(s_Z^2/T) < \sigma^2/T < var(\bar{Z})$, con lo cual el estimador anterior es bastante optimista en la precisión del resultado de μ .

Se han planteando diferentes procedimientos para poder estimar la $var(\bar{Z})$ más aceptablemente. Tal vez la más conocida es hacer lotes ("batching"). Se divide Z_1, \dots, Z_T en k segmentos de longitud m cada uno, y se calcula las medias por lotes B_1, \dots, B_k , los cuales tienen de nuevo una media μ . Se asume que todos los B_i son incorrelacionados (así es aproximadamente para m suficientemente grande) y se estima $var(\bar{B}) = var(\bar{Z})$ por s_B^2/k . Por tanto, se consigue un menor sesgo a costa de una menor precisión en la variabilidad del estimador.

También se puede analizar Z_i como una serie temporal (Ripley, 1987). Se sabe que $var(\bar{Z}) \approx 2\pi f(0)/T$ donde f es la densidad espectral del modelo de la serie temporal. Por tanto, solo se necesita estimar con cuidado el valor de $f(0)$, la densidad espectral en el origen. En manos de un analista experto es una buena técnica, pero es muy difícil de automatizar y fácil de abusar.

La idea más simple que se puede aplicar pero que ha sido repetidamente pasada por alto se basa en estimar la correlación. Se conoce que la varianza de la media muestral es

$$var(\bar{Z}) = \frac{\sigma^2}{T^2} \left\{ 1 + 2 \sum_1^{T-1} \left(1 - \frac{s}{T} \right) \rho_s \right\}$$

donde $\rho_s = corr(Z_i, Z_{i+s})$. Hannan (1957) propone utilizar el siguiente estimador

$$\widehat{var}(\bar{Z}) = \frac{T}{(T-L)(T-L-1)} \sum_{|s|<L} \left(1 - \frac{|s|}{T} \right) (c_s - \bar{Z}^2),$$

$$c_s = \sum_{i=1}^{T-s} \frac{Z_i Z_{i+s}}{T-s},$$

donde L se escoge de forma que $\rho_s \approx 0$ para $s \geq L$. Moran (1957) muestra que $\widehat{var}(\bar{Z})$ es un estimador insesgado de $var(\bar{Z})$ (en la práctica dependerá de la aproximación al valor de L).

1.2 Supercomputación

1.2.1 Introducción

No es fácil definir la palabra “supercomputador”. Generalmente hablando, los supercomputadores son los ordenadores más potentes en un momento dado del tiempo. Sin embargo, este término se viene asociando con ciertos tipos de arquitecturas que se adaptan, particularmente bien, hacia el cálculo (Cálculo Intensivo Numérico, Cálculo de altas prestaciones, etc); son los llamados **supercomputadores con arquitectura vectorial y paralela**. Se ha observado en el campo técnico y científico que la necesidad de potencia de cálculo se multiplica por 10 cada cinco o seis años. Como resultado, los fabricantes están constantemente produciendo máquinas más rápidas, y a lo que fue un supercomputador en 1990 no le corresponde este calificativo en 1996.

El mercado de los supercomputadores existe desde la década de los años 70 pero estaba muy limitado por las agencias de gobierno de Estados Unidos al ser tratados como material secreto. Preferentemente se utilizaban para la investigación nuclear y para proyectos espaciales. En los años 80 este mercado tuvo un fuerte crecimiento al incluir como clientes a la industria (industrias químicas, electrónicas, aeronáuticas, automóvil y petróleo) y universidades. En la actualidad, este mercado es mucho más amplio y el número de supercomputadores ha aumentado y se han extendido por todo el mundo. En España la supercomputación llegó al final de la década de los 80 pero ha ido aumentando paulatinamente. En particular, Cataluña inauguró al principio de los años 90 el Centre de Super-Computació de Catalunya (CESCA) con un IBM 3090 y posteriormente un Cray X-MP/EA. Ahora el IBM 3090 se ha cambiado por un IBM SP2 y el antiguo Cray por un Cray Y-MP. Además, en el mes de octubre de 1995 la Comissió Interdepartamental de Recerca i Innovació Tecnològica (CIRIT), la Fundació Catalana per a la Recerca (FCR) i la Universitat Politècnica de Catalunya (UPC) firmaron un convenio para crear el Centre de Computació i Comunicacions de Catalunya (C⁴). Este centro sirve para coordinar las actividades del CESCA y del CEPBA

(Centre Europeu de Paral.lisme de Barcelona) i para potenciar el acceso a sus recursos a través de la red de comunicaciones. De esta manera, se hace más sencillo y asequible el poder utilizar los recursos existentes de supercomputación.

La modelización y simulación numérica se ha convertido en herramientas básicas en la carrera para la productividad, calidad e innovación. Se presentan varios de los campos donde los supercomputadores se usan actualmente:

- Fabricación de automóviles, aeronáutica, etc

Tradicionalmente, para el diseño de automóviles y aviones se necesitaba del uso de prototipos que se empleaban para realizar gran cantidad de pruebas. Este método tiene dos grandes defectos: gasta mucho tiempo y es costoso. En cambio, ahora es posible realizar simulaciones numéricas consiguiendo un gran ahorro de tiempo y dinero. La mayor potencia de cálculo ha facilitado llegar a la fase de test con prototipos altamente optimizados. Además, la modelización matemática ha ayudado a comprender ciertos fenómenos (tales como las características del comportamiento por carretera de un automóvil) y calcular ciertos parámetros (tales como la temperatura que se tiene en la cámara de combustión) que antes no eran accesibles.

Los métodos de simulación usados se fundamentan en la descomposición de superficies y volúmenes en un gran número de puntos que son tratados como una malla. Cada punto tiene asociadas variables tales como la temperatura, la presión y el desplazamiento. Se requiere un alto número de iteraciones para simular un fenómeno físico, por tanto el tiempo necesario para cada simulación depende de la potencia de cálculo disponible.

Algunas de las aplicaciones específicas para la industria del automóvil son:

- La combustión y sus efectos en la generación de la polución.
- La acústica.
- Modelización de los accidentes para mejorar la seguridad de los pasajeros.
- Aerodinámica.

- Química y Biotecnología (por ejemplo, estudios de nuevos fármacos)
- Investigación petrolera y geofísica

- Meteorología en la cual la exactitud de las previsiones es una función directa de como es de fina la malla usada.
- Electrónica (por ejemplo, simulación de circuitos integrados)
- Física: mecánica, astrofísica, física de altas energías, etc.

El número de áreas en las cuales el supercomputador se emplea continua creciendo. Así, se están desarrollando aplicaciones en medicina, economía, estadística y ecología. Ver (Crowley, 1992; Martino *et al.*, 1994, ...)

1.2.2 Evolución de las arquitecturas de los ordenadores

Antes de hablar sobre las diferentes tipos de arquitecturas, se presenta una breve introducción de la historia de los ordenadores.

Las diferentes generaciones de ordenadores se definen a partir de la tecnología, arquitectura, sistemas operativos y lenguajes usados. En general, se puede clasificar en tres grandes periodos:

1. La primera generación (1945-1956).

La primera generación se caracterizó por ordenadores con funcionamiento electromecánico y, más tarde, con tubos de vacío. Un gran adelanto se hizo en el desarrollo de las memorias (con la aparición de la organización jerárquica) y en el medio de almacenamiento de los datos (con la invención del disco magnético y cintas).

2. La segunda generación (1954-1963).

La segunda generación se inicia con la aparición de los transistores. Estos producen un aumento considerable en la velocidad de cálculo. La creación de los circuitos complejos (llamados circuitos integrados) agrupó un gran número de componentes microscópicos directamente en una microplaqueta de silicona haciendo posible la reducción del volumen y del consumo de energía.

3. La tercera generación (1962-).

Rápidamente la tecnología avanzó por la senda de la miniaturización y se extendió el uso de los circuitos integrados basados en la tecnología MSI

(Medium Scale Integration), LSI (Large Scale Integration), y finalmente en VLSI (Very Large Scale Integration) trayendo consigo un incremento en la velocidad, fiabilidad y reducción de coste y consumo de energía.

Uno de los hechos más importantes es la evolución de la organización interna del ordenador: las operaciones de entrada/salida y de proceso se realizan simultáneamente y comienza a existir la multiprogramación, multiproceso y el almacenamiento virtual.

La tecnología aún se está desarrollando hoy, pero hay límites físicos para su progreso tales como la velocidad de la luz. Por esta razón en la última década, se han desarrollado nuevas ramas de investigación, especialmente en el campo de la arquitectura de los ordenadores.

La arquitectura de los ordenadores puede ser clasificada de manera general dentro de cuatro categorías usando la clasificación de Flynn (1972):

- “The single instruction stream single data stream organization” (SISD) que corresponde a la arquitectura de Von Neumann tradicional. La que existe en la mayoría de los ordenadores.
- “The single instruction stream multiple data stream organization” (SIMD). Incluye a la mayoría de los procesadores vectoriales.
- “The multiple instruction stream multiple data stream organization” (MIMD). Esta arquitectura aparece en la mayoría de los sistemas denominados como multiprocesadores.
- “The multiple instruction stream single data stream organization” (MISD).

En la arquitectura de Von Neumann tradicional se ejecuta el programa en serie. Los ordenadores diseñados con esta arquitectura tienen una memoria para almacenar datos e instrucciones, una unidad de control para interpretar instrucciones, una unidad aritmético-lógica para procesar los datos y una unidad de entrada/salida. Con esta arquitectura, las operaciones que constan de una instrucción se ejecutan secuencialmente. Por ejemplo, para ejecutar la multiplicación de punto flotante ($X = Y * Z$) se necesitan realizar los siguientes pasos:

1. Sumar las características

2. Normalizar las mantisas y ajustar las características
3. Multiplicar las mantisas
4. Normalizar la mantisa y ajustar la característica del resultado

La normalización y ajuste del exponente hace posible garantizar que el primer dígito de la mantisa sea significativo.

En la arquitectura de Von Neumann tradicional, estos pasos se realizan secuencialmente con el resultado de que una instrucción se ejecuta en cada paso para un flujo de datos (por ejemplo, un grupo de operandos). Además, no hay superposición de instrucciones y el tiempo necesario para ejecutar una instrucción es la suma de los tiempos de cada paso. Para reducir el tiempo de ejecución de un programa se puede 1) solapar instrucciones o 2) realizar una instrucción que opera con múltiples datos o 3) realizar múltiples instrucciones que se ejecutan a la vez. Estas tres variaciones corresponden, respectivamente, a los tres tipos de arquitectura más conocidas: procesador “pipeline”, procesador vectorial SIMD y procesador paralelo MIMD.

1.2.3 Procesador “pipeline”

Es una técnica implementada para que varias instrucciones sean simultáneamente superpuestas en la ejecución. Un “pipeline” se parece a una cadena de ensamblaje donde cada paso de la “pipeline” se responsabiliza de la ejecución de una parte de la instrucción (suboperación). Una vez que se completa una suboperación comienza una nueva iteración.

Para ilustrar su funcionamiento se muestra el siguiente ejemplo:

```
DO 10 I=1,N
    X(I) =Y(I)*Z(I)
10 CONTINUE
```

Como se ha visto anteriormente, la multiplicación de punto flotante suele necesitar de 4 suboperaciones. Por tanto, en la arquitectura de Von Neumann tradicional se necesitan $N \times 4$ ciclos de máquina y con respecto al tiempo requerido

se necesita $N \times T$, donde T es el tiempo empleado para ejecutar la instrucción. En cambio, mediante un "pipeline" con 4 suboperaciones se realiza así:

Ciclo	1	
subop. 1	Y(1),Z(1)	
Ciclo	2	
subop. 1	Y(2),Z(2)	
subop. 2	Y(1),Z(1)	
Ciclo	3	
subop. 1	Y(3),Z(3)	
subop. 2	Y(2),Z(2)	
subop. 3	Y(1),Z(1)	
Ciclo	4	
subop. 1	Y(4),Z(4)	
subop. 2	Y(3),Z(3)	
subop. 3	Y(2),Z(2)	
subop. 4	Y(1),Z(1)	
Ciclo	5	
subop. 1	Y(5),Z(5)	
subop. 2	Y(4),Z(4)	
subop. 3	Y(3),Z(3)	
subop. 4	Y(2),Z(2)	
salida del pipeline	Y(1),Z(1)	
....	
....	

El resultado de $Y(1) \times Z(1)$ está disponible después de 4 ciclos y el $Y(2) \times Z(2)$ después de 5 ciclos, mientras que en la arquitectura de Von Neumann tradicional se requieren de 8 ciclos para calcular $Y(2) \times Z(2)$. Por tanto, el tiempo requerido para obtener todos los resultados del bucle es de $3 + N$ ciclos, bastante menos que los $4 \times N$ ciclos normales.

En un caso general, si una instrucción requiere de K ciclos, el bucle necesitará $T + (N - 1) \times T/K$ de tiempo. En el caso de bucles cortos, el tiempo T domina. Cuanto mas grande es el bucle, el rendimiento se va mejorando asintóticamente.

1.2.4 Procesador vectorial

El siguiente paso para optimizar el tiempo de ejecución de un programa es trabajar con procesadores vectoriales. Estos son adecuados para aquellas aplicaciones en las cuales la mayor parte del programa se puede expresar como manipulación matemática de vectores.

Para obtener el máximo rendimiento de los “pipelines”, estos deben ser “alimentados” a un ritmo que corresponde a su ciclo de tiempo. El ciclo de tiempo de la memoria central (cantidad de tiempo entre una petición de CPU para datos y el momento cuando se recibe estos) es a menudo de un orden de 5 veces el tiempo de ciclo de CPU básico. Para evitar tener “pipelines” esperando datos se han añadido unos pequeños almacenamientos locales de datos llamados *registros vectoriales*. Los registros son pequeñas memorias localizadas en la CPU las cuales tienen el mismo ciclo de tiempo que la CPU. La mejora del tiempo de ejecución dependerá del número de registros y de su tamaño. Si su tamaño es, por ejemplo, de 64 elementos, en cada ciclo se puede almacenar 64 datos.

Así, el procesador vectorial consta de l registros vectoriales que “alimentan” a k “pipelines” que pueden trabajar en cadena. Por esta razón, los “pipelines” se han diseñado para manejar ciertas operaciones muy comunes para grupos de datos: suma de punto flotante, multiplicación de punto flotante, inversión de punto flotante, suma entera y operaciones lógicas.

El siguiente bucle puede servir como ejemplo del funcionamiento del procesador vectorial:

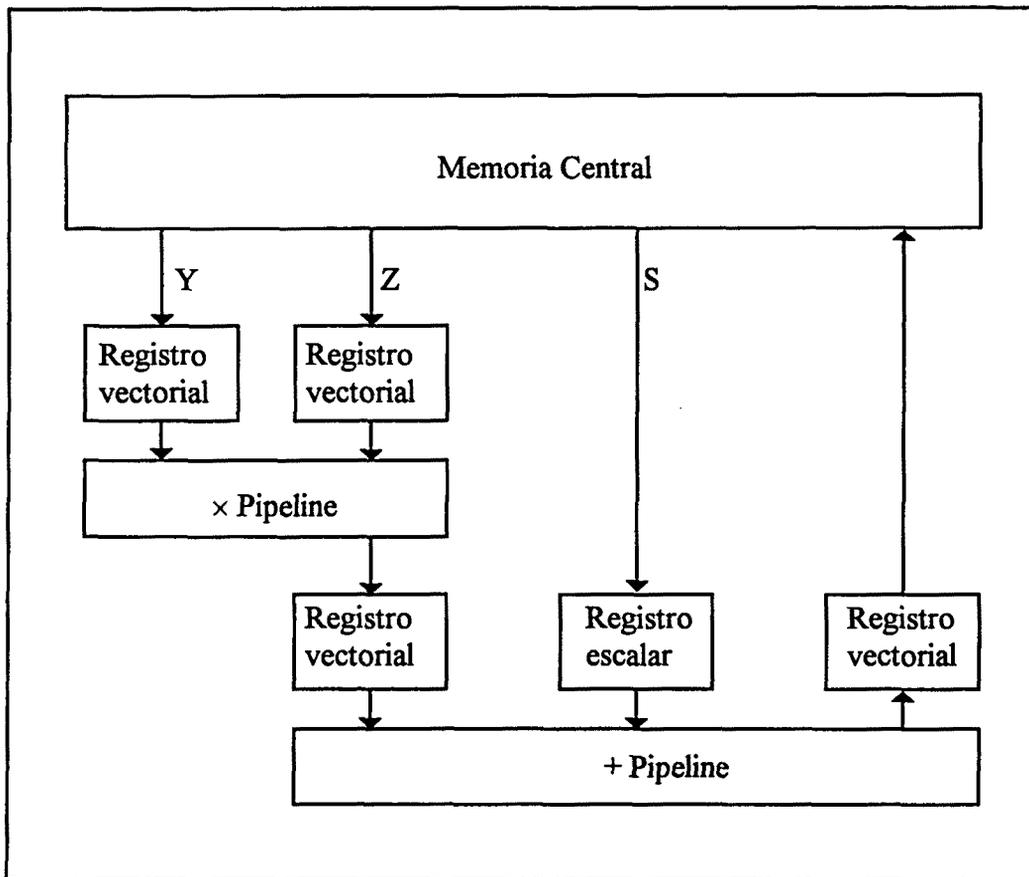
```

DO 10 I=1,N
      X(I) =S+Y(I)*Z(I)
10  CONTINUE

```

Se comienza cargando los primeros 64 elementos de Y en un registro y los otros primeros 64 elementos de Z en otro registro. Se realiza el “pipeline” de la multiplicación de punto flotante. Tan pronto se obtenga el resultado de $Y(1) \times Z(1)$ se almacena en un registro intermedio y se activa el “pipeline” de la suma de punto flotante. Cuando se tenga resultados se van acumulando en otro registro que finalmente se almacenará en la memoria central. Cuando la primera pareja

de vectores de 64 elementos ha sido procesada el sistema carga la próxima pareja de vectores de 64 elementos.



Así, cuanto mayor sea el tamaño del registro mayor rendimiento se tiene. Por otro lado, dado un tamaño de registro concreto el rendimiento de un bucle es mejor cuanto mayor sea el número de iteraciones y más próximo esté de un múltiplo del tamaño del registro.

No hay que olvidar que un ordenador con procesador vectorial también puede trabajar de forma escalar ya que, en muchos casos, no es posible realizar las operaciones que se indica en el programa de forma vectorial.

1.2.5 Procesador paralelo

El concepto de paralización se refiere a un grupo de técnicas de hardware y software que hacen posible ejecutar secuencias independientes de instrucciones simultáneamente en distintos procesadores.

El siguiente programa muestra que se entiende por paralización en el estricto sentido del término:

```

DO 10 J=1,2
  DO 20 I=1,10000
    A(I,J) = .....
20  CONTINUE
10  CONTINUE
DO 30 J=1,2
  .....
  ... = ..... A(N,J) ...
  .....
10  CONTINUE

```

En el caso de paralelización con dos procesadores, los dos bucles

```

DO 20 I=1,10000
  A(I,1) = .....
20  CONTINUE

```

```
DO 20 I=1,10000
    A(I,2) = .....
20 CONTINUE
```

se ejecutarían separadamente en los dos diferentes procesadores.

El bucle etiquetado DO 30 solo será ejecutado una vez que ambos procesadores hayan acabado.

Bajo este punto de vista se engloba tanto la arquitectura paralela SIMD como MIMD. Los procesadores deben de comunicarse para compartir datos o para sincronizar la ejecución de los programas. Hay dos mecanismos para facilitar la comunicación entre procesadores: (1) la memoria compartida o (2) la red de interconexión. Las arquitecturas paralelas que usan el primer esquema son llamadas arquitecturas paralelas de memoria compartida y aquellos que usan el segundo esquema son llamadas arquitecturas paralelas de memoria distribuida.

1.2.6 Programación vectorial

La arquitectura de Von Neumann tradicional hacía que la forma de programar fuera secuencial y escalar. A consecuencia de las nuevas arquitecturas, vectorial y paralela, el diseño de programación ha cambiado para aprovechar al máximo su potencia.

La programación vectorial se basa en realizar el programa con el mayor número de bucles que sean vectorizables. Un ejemplo sencillo de bucle vectorizable, expresado en FORTRAN, puede ser el siguiente:

```
DO 10 I=1,20
    C(I) =A(I)+B(I)
10 CONTINUE
```

Los valores de los elementos de C calculados en cada iteración son *independientes*. Además, es posible calcular todos ellos “al mismo tiempo”, por tanto el bucle se puede vectorizar. En modo vectorial, con un tamaño de registro z , el proceso de calculo de C sería:

- Primero

$$C(1) = A(1) + B(1)$$

$$C(2) = A(2) + B(2)$$

...

$$C(z) = A(z) + B(z)$$

- Después

$$C(z + 1) = A(z + 1) + B(z + 1)$$

$$C(z + 2) = A(z + 2) + B(z + 2)$$

y así sucesivamente

Los elementos de C se calculan en grupos de z elementos. Esto es como si se calculasen los elementos de un grupo casi todos al mismo tiempo.

Por contra, en el siguiente ejemplo

```
DO 99 J=1,20
  A(J+1) =A(J)+B(J)
99  CONTINUE
```

el bucle no se puede vectorizar. En modo escalar, A se calcula como:

$$A(2) = A(1) + B(1)$$

$$A(3) = A(2) + B(2)$$

...

$$A(21) = A(20) + B(20)$$

Los valores de la segunda iteración de A dependen de los valores de la primera y así sucesivamente. Los valores son *dependientes* y en tal caso, el bucle no se puede vectorizar porque los resultados de la iteración precedente aún no se conoce.

En general, las condiciones que inhiben la vectorización de un bucle son debidas a la presencia de:

- Instrucciones IF
- Recurrencias
- Dependencias
- Interrupciones
- Instrucciones de Entrada/Salida
- Funciones no vectorizables
- Procesos CALL
- Inicialización de variables

Además, existen pseudo-inhibidores que reducen el rendimiento de la vectorización. Estos son principalmente:

- Tamaño del bucle DO
- Ordenación de bucles anidados DO

Por tanto, al intentar expresar algoritmos desde un punto de vista vectorial hay que evitar, en la medida de lo posible, estas construcciones.

Un ejemplo de como se puede resolver un caso de bucle que puede tener un rendimiento vectorial bajo como consecuencia del tamaño del bucle es el siguiente:

Se presenta una subrutina que multiplica una matriz tridimensional, A , por un valor escalar S

```
SUBROUTINE MULTI(A,NX,NY,NZ,S)
  DIMENSION A(NX,NY,NZ)
  DO 10 IZ=1,NZ
    DO 20 IY=1,NY
      DO 10 IX=1,NX
        A(IX,IY,IZ) =S*A(IX,IY,IZ)
      30    CONTINUE
    20    CONTINUE
  10  CONTINUE
```

```
RETURN
END
```

Esta operación sería más efectiva hacerla en un solo DO, ya que se aumentaría el número de iteraciones del bucle y, a la vez, se reduciría el número de índices. Esto es,

```
SUBROUTINE MULTI(A,NX,NY,NZ,S)
DIMENSION A(NX*NY*NZ)
DO 10 I=1,NZ*NY*NX
    A(IX*IY*IZ) =S*A(I)
10 CONTINUE
RETURN
END
```

En el capítulo 4 se dan algunas indicaciones sobre vectorización de algoritmos en estadística y simulación.

1.3 Futuro de la estadística computacional y de la supercomputación

La evolución de los superordenadores ha ido pasando de la vectorización a la paralización, que actualmente está más de moda. Pero el futuro puede deparar una combinación entre un procesador vectorial y un sistema paralelo para aprovechar ambas arquitecturas, como apunta Keiichiro Uchida, director del grupo High Performance Computing de Fujitsu (Teraflop, 1995). La bajada de precio de los superordenadores y el incremento y mejora de las comunicaciones facilitarán el aumento del número de usuarios de los superordenadores. Estos nuevos usuarios suelen ser de la industria. En Japón, más de la mitad de los superordenadores lo tienen las industrias que hacen investigación y el resto las universidades y los centros de investigación pública. Por tanto, serán las grandes industrias las que, a partir de sus necesidades, irán empujando, cada vez más, el mercado de superordenadores, desplazando a los habituales usuarios (universidades).

Por otro lado, la estadística computacional ha ido creciendo tanto en usuarios como en prestaciones. Ahora se puede plantear hacer grandes estudios de ecología, epidemiología, economía, bolsa, de inferencia estadística, etc. que necesitan de la estadística computacional para poder realizar las simulaciones pero además, necesitan de un superordenador por la gran cantidad de pasos que requieren. En el futuro la estadística computacional pasa por que haya cada vez más facilidades para que los posibles usuarios sean capaces de aplicarla de una manera cómoda y sencilla. Esta labor corresponde tanto a los especialistas en este campo como a las empresas de software que deben de ayudar a crear unos programas para facilitar la estadística computacional.

Por tanto, el futuro de la estadística computacional como de la supercomputación, o más general de la informática, están muy unidos ya que existe un efecto de retroalimentación positiva entre ellas.

Capítulo 2

TÉCNICAS DE REDUCCIÓN DE LA VARIANZA: VARIABLES DE CONTROL Y MONTE CARLO CONDICIONAL

En las simulaciones de Monte Carlo tan importante es la estimación puntual de la característica de interés como obtener un valor con el menor error posible. En este capítulo se presenta algunas técnicas que intentan reducir la varianza sin incrementar el número de réplicas realizadas. Se suelen basar en la teoría de la correlación o en informaciones que se tienen del proceso simulado. Se expondrán más ampliamente la técnica de variables de control y la de Monte Carlo condicionado.

2.1 Introducción

Las simulaciones de Monte Carlo se realizan para estudiar alguna característica del proceso simulado (por ejemplo, en la simulación de un proceso de colas puede ser el tiempo medio de espera). En general, se estima alguna medida de centralización de la característica de interés, casi siempre se usa la media muestral, y además, se da una estimación de su bondad, habitualmente expresado como una estimación de la varianza de la media muestral. La varianza de la característica estimada (por ejemplo, media muestral) suele disminuir como c/m , donde c es una constante, a menudo la varianza de la característica (por ejemplo, tiempo de espera), y m el número de réplicas independientes de la característica estudiada en el proceso de Monte Carlo. Así, para doblar la precisión absoluta de la estimación, medida por su desviación estándar (es decir, reducir a la mitad ésta), o su dispersión relativa, medida por su desviación estándar partida por su esperanza, se debe de generar una muestra cuatro veces más grande ($4m$). Por tanto, para reducir la desviación estándar de la característica estimada por un factor de $1/h$, se necesita, en general, una muestra que sea h^2 más grande que antes. Como se puede observar, esta forma de reducir la dispersión provoca que fácilmente se tenga un número de réplicas en el proceso de Monte Carlo exorbitante, que puede ser inviable a la hora de la realización práctica, a partir del ordenador, del proceso simulado.

Las técnicas denominadas de reducción de la varianza se basan en la idea de que la simulación es un experimento estadístico controlado. Así, se puede reducir la varianza obtenida para un número m de réplicas por el uso de ideas

probabilísticas.

Si en la simulación se induce correlación, positiva o negativa, deliberadamente entre observaciones se puede reducir la varianza por aplicación de la teoría de la correlación. Sean X e Y los resultados aleatorios de dos simulaciones obtenidas a partir del vector de números aleatorios U y V , respectivamente. Adviértase que $var(X \pm Y) = var(X) + var(Y) \pm 2cov(X, Y)$. Por tanto, para unas marginales dadas, maximizar la covarianza es equivalente a maximizar la correlación; igualmente para minimizar. Si X e Y son funciones crecientes de U y V , respectivamente, entonces

$$1. U = V \Rightarrow corr(X, Y) \geq 0$$

$$2. \left. \begin{array}{l} U = (U_1, U_2, \dots, U_m) \\ V = (V_1, V_2, \dots, V_m) \\ V_i = 1 - U_i \end{array} \right\} \Rightarrow corr(X, Y) \leq 0$$

También, si $m = 1$ y X e Y tienen, respectivamente, las distribuciones F y G , entonces:

$$3. \text{ para maximizar } corr(X, Y), \text{ se genera } (X, Y) \text{ como } (F^{-1}(U), G^{-1}(U))$$

$$4. \text{ para minimizar } corr(X, Y), \text{ se genera } (X, Y) \text{ como } (F^{-1}(U), G^{-1}(1 - U))$$

A menos que Y sea función lineal de X , los puntos 3 y 4 son las únicas formas de conseguir una correlación extrema.

Entonces, con las cuatro propiedades anteriores y la relación básica

$$var(X \pm kY) = var(X) + k^2 var(Y) \pm 2kcov(X, Y)$$

se tienen los pilares fundamentales de las siguientes técnicas de reducción de la varianza:

- **Números aleatorios comunes:** Se suele utilizar para estudiar la diferencia entre dos variables o características (X, Y) . Por ejemplo, se tienen dos tipos de procesos de colas, interesa saber la esperanza de la diferencia de los tiempos de espera de las dos colas con la mayor precisión. En este caso se utilizará el promedio de $X - Y$ como un estimador de la $E(X - Y)$
- **Variables de control:** Sea Y la variables principal de la cual se desea obtener una estimación de su esperanza, $E(Y)$. A la vez, se tiene otra

variable denominada de control, X , con $E(X)$ conocida (suponemos igual a cero). Se usará como un estimador con mayor precisión la media de $Y - kX$ para estimar la $E(Y)$.

- **Variabes antitéticas:** Sea Y los resultados de la ejecución uno, X los resultados de la ejecución dos, se empleará el promedio de $(Y + X)/2$ como un estimador de $E(Y)$

Las dos primeras técnicas se basan en la aplicación de las propiedades de correlación 1 y 3 mientras que la última técnica se basan en las propiedades 2 y 4. Además, es habitual emplear la técnica de números aleatorios comunes cuando se usa cualquiera de las otras dos técnicas, tal como se hará en la presente memoria.

Un detalle importante es que para maximizar (o minimizar) la $corr(X, Y)$, los valores introducidos de U y V deben estar sincronizados. Entre otras cosas, esto significa diseñar la simulación para que X e Y sean funciones crecientes de U y V , respectivamente. A grandes rasgos, esta monotonicidad asegura que la correlación inducida tiene el signo adecuado para la reducción de la varianza. En la sección 2 del capítulo 5 se hará referencia a este hecho.

Otras técnicas de reducción de la varianza no dependen de la correlación inducida. La estratificación y el muestreo por importancia emplean el muestreo ponderado basado en información cualitativa o cuantitativa "a priori" para reducir la varianza. El muestreo condicional o Monte Carlo condicional se basa en poder calcular la esperanza de la variable de interés a partir de esperanzas condicionadas, es decir, emplear esta propiedad: $E(Y) = E_X(E_Y(Y/X))$.

Estas son las técnicas más comunes dentro del marco de las técnicas de reducción de la varianza pero existen algunas más (véase Kleinjen, 1974 Cap III; Wilson, 1984; Nelson and Schmeiser, 1986). Gaver and Thompson (1973, Cap. 12) muestra un sumario de técnicas de reducción de la varianza con aplicaciones a la investigación operativa.

Por último, hay que destacar las limitaciones de las técnicas de reducción de la varianza. En simulaciones complejas, su implementación debe de hacerse con mucho cuidado ya que puede producir distorsiones no controladas del proceso o en el mejor de los casos no hacer el efecto esperado. La implementación de estas técnicas puede requerir un incremento del tiempo de computación. Este nunca

debe ser superior al que se necesitaría para obtener la misma precisión por la simple generación de más réplicas. Además, tampoco debe ser muy costoso el tiempo de programación. En resumen, la reducción de la varianza es un arte que requiere destreza, experiencia y tiempo para su implementación. De hecho, Handscomb (1969) llama a una técnica “reducción de la varianza” si reduce la varianza proporcionalmente más que el incremento de trabajo que precisa.

2.2 Variables de Control

Es una de las técnicas más útiles porque puede ser aplicada en las simulaciones de todos los aspectos de la distribución (media, varianza, cuantiles, etc) de la característica de interés y porque las condiciones de aplicabilidad son más fáciles de cumplir que algunas de las premisas requeridas para otras técnicas de reducción de la varianza. Además, simular una variable de control, de la cual se debe de conocer o aproximadamente conocer la esperanza, es una ayuda importante para la validación del programa de ordenador que genera la simulación.

Supongamos que se tiene la variable aleatoria bivalente (Y, C) y una muestra de tamaño m de esta población, $(Y_1, C_1), \dots, (Y_m, C_m)$, y se desea estimar algún atributo de la distribución marginal de Y : por ejemplo, la esperanza, $E(Y)$. Además, se conoce alguna propiedad acerca de C (por ejemplo, $E(C)$). La idea de esta técnica de reducción de la varianza es utilizar esta información y la muestra de C para mejorar la estimación del atributo de Y basado solamente en la muestra de Y . Para ello es necesario que exista correlación entre la variable Y y C , para poder aplicar la teoría de la correlación. Por tanto, las condiciones que debe de cumplir una buena variable de control es:

- Conocer el valor de la $E(C)$
- Existencia de correlación entre Y y C .

Si $E(C)$ solo se conoce aproximadamente, este valor aproximado debe de ser obtenido independientemente de la simulación en la que está siendo usada. Además, se induce sesgo en el estimador de variable de control.

2.2.1 Variables de Control Simple

Es el caso más sencillo de emplear y, también, el menos eficaz en relación al aumento de precisión. La nueva variable \bar{Y}^s es función lineal simple de la media muestral de Y , \bar{Y} , de la media muestral de C , \bar{C} , y de la $E(C)$ que se supone conocida e igual a μ_c :

$$\bar{Y}^s = \bar{Y} - (\bar{C} - \mu_c) = \frac{1}{m} \sum_{i=1}^m (Y_i - C_i + \mu_c) \quad (2.1)$$

Las propiedades de \bar{Y}^s son su ausencia de sesgo,

$$E(\bar{Y}^s) = E(\bar{Y}) - E(\bar{C} - \mu_c) = E(Y) - 0 = E(Y)$$

y una posible varianza menor que la varianza de \bar{Y} . Así, el cálculo de la varianza es:

$$var(\bar{Y}^s) = var(\bar{Y}) + var(\bar{C} - \mu_c) - 2cov(\bar{Y}, (\bar{C} - \mu_c))$$

Por tanto, para que se reduzca la varianza en comparación a la $var(\bar{Y})$ se necesita que la correlación entre Y y C , y por tanto, también, entre \bar{Y} y \bar{C} , sea positiva. Si fuera negativa, se debería cambiar el signo negativo delante de $(\bar{C} - \mu_c)$ en la ecuación (2.1) por el signo positivo para poder obtener de nuevo un valor negativo en el tercer sumando de la $var(\bar{Y}^s)$. Además, debe de cumplirse la siguiente desigualdad

$$var(\bar{C} - \mu_c) - 2cov(\bar{Y}, (\bar{C} - \mu_c)) < 0 \quad (2.2)$$

ya que en caso contrario no se compensa el incremento de varianza, $var(\bar{C} - \mu_c)$, con la correlación entre las variables Y y C y se produciría una inflación de la varianza.

Para simplificar, si consideramos que la $var(C) = var(Y) = \sigma^2$, entonces, la varianza de \bar{Y}^s es:

$$var(\bar{Y}^s) = \frac{2\sigma^2(1 - corr(Y, C))}{m} = var(\bar{Y})[2(1 - corr(Y, C))]$$

Por consiguiente, si $2(1 - corr(Y, C)) < 1$, o equivalentemente $corr(Y, C) > 1/2$, el uso de esta técnica de variables de control reduce la varianza de la estimación de la media de Y en comparación a la que se obtendría directamente de $var(\bar{Y})$.

2.2.2 Variables de Control Ajustadas por Regresión

Es una generalización del método anterior que se asienta en el uso de un estimador ajustado por regresión:

$$\bar{Y}^* = \bar{Y} + \beta(\bar{C} - \mu_c) \quad (2.3)$$

Este nuevo estimador, \bar{Y}^* , sigue siendo un estimador insesgado de $E(Y)$ si β es una constante. Además, se puede obtener \bar{Y}^* como un caso especial de β cuando valga -1 . Este refinamiento sirve para maximizar la reducción de la varianza que se puede obtener con una variable de control, C . Así, si se expresa la desigualdad (2.2) como

$$\text{corr}(Y, C) \sqrt{\frac{\text{var}(Y)}{\text{var}(C)}} > 1/2 \quad (2.4)$$

se puede ver que un factor importante para que funcione correctamente la técnica de variable de control simple es el valor del cociente de desviaciones típicas entre la variable de interés, Y , y la variable de control, C . Si este cociente es mayor que uno, una $\text{corr}(Y, C) < 1/2$ puede servir para lograr una reducción de la varianza mientras que al contrario, cociente menor que uno, se dificulta la obtención de la reducción de la varianza ya que se necesita una $\text{corr}(Y, C) > 1/2$.

Este efecto queda aplacado con la utilización de la constante β . Se busca β que haga mínima la varianza de \bar{Y}^* ,

$$\text{var}(\bar{Y}^*) = \text{var}(\bar{Y}) + \beta^2 \text{var}(\bar{C}) - 2\beta \text{cov}(\bar{Y}, \bar{C}) \quad (2.5)$$

por tanto, se deriva respecto a β y se iguala a cero encontrando que la $\text{var}(\bar{Y}^*)$ es mínima si

$$\beta = \frac{-\text{cov}(\bar{Y}, \bar{C})}{\text{var}(\bar{C})} = \frac{-\text{cov}(Y, C)}{\text{var}(C)} = -\text{corr}(Y, C) \sqrt{\frac{\text{var}(Y)}{\text{var}(C)}} \quad (2.6)$$

Como se observa este valor óptimo de β tiene en consideración las diferencias entre las varianzas de Y y C . Si se sustituye (2.6) en la ecuación (2.5) se encuentra que la varianza mínima es

$$\text{var}(\bar{Y}^*) = \text{var}(\bar{Y})(1 - \text{corr}^2(Y, C)) \quad (2.7)$$

que sólo depende de la correlación entre Y y C . Por consiguiente, si se conocen los valores de la varianzas de Y , de C y la correlación de Y y C la variable de control ajustada por regresión, \bar{Y}^* , siempre se reduce la varianza excepto cuando la $\text{corr}(Y, C) = 0$. También, la varianza de \bar{Y}^* siempre es menor que la varianza de la variable de control simple \bar{Y}^s , la cual si es menor de $1/2$, puede ser mayor que la varianza de \bar{Y} . En el caso concreto de $\text{var}(Y) = \text{var}(C)$, se puede calcular que

$$\begin{aligned} \text{var}(\bar{Y}^*) &= \frac{1}{m} \text{var}(Y)(1 - \text{corr}^2(Y, C)) \\ &= \frac{1}{m} \text{var}(Y)(1 - \text{corr}(Y, C))(1 + \text{corr}(Y, C)) \\ &\leq \frac{2}{m} \text{var}(Y)(1 - \text{corr}(Y, C)) = \text{var}(\bar{Y}^s) \end{aligned}$$

Además, se puede advertir que el cociente de desviaciones estándar, $\sqrt{\frac{\text{var}(\bar{Y}^*)}{\text{var}(\bar{Y})}}$ que es igual a $(1 - \text{corr}^2(Y, C))^{1/2}$ disminuye más lentamente que como se incrementa el valor absoluto de la $\text{corr}(Y, C)$. Esto produce como efecto que para dividir la desviación estándar de \bar{Y} a la mitad con el mismo número de réplicas se necesita una correlación entre Y y C alta, alrededor de 0.87.

Un problema bastante común es el desconocimiento de $\text{var}(C)$, $\text{var}(Y)$ y $\text{corr}(Y, C)$. Esto trae consigo que se deban de estimar a partir de la muestra $(Y_1, C_1), \dots, (Y_m, C_m)$. Además, al no emplear una constante β sino una variable aleatoria $\hat{\beta}$, se puede inducir un sesgo en \bar{Y}^* y reducir las predicciones de reducción de la varianza de la ecuación (2.7).

2.2.3 Ejemplo de aplicación de la técnica de variables de control

Este ejemplo fue presentado por Barnett (1965, pag. 27). Un servicio de reparación se encarga del mantenimiento de n máquinas las cuales se averían de vez en cuando. Se supone que las máquinas fallan independientemente unas de otras, y que para cualquier máquina, las averías son sucesos en un proceso de Poisson de parámetro λ . Además, se supone que el tiempo necesario para realizar la reparación de cualquier máquina es constante y se denominará r . Los entorpecimientos en las máquinas surgen si se forman colas de máquinas estropeadas, es decir,

cuando falla al menos una máquina en el momento que se está reparando otra. Simular este proceso es muy sencillo, así como, estimar la “disponibilidad de las máquinas”, S , sobre un período de tiempo t , por,

$$\hat{S} = \frac{\sum_{i=1}^n f_i}{nt}$$

donde f_i es la cantidad de tiempo en que la máquina i , $\forall i = 1, \dots, n$, ha funcionado correctamente durante el período de tiempo t .

Como variable de control, Barnett usó la estimación, \hat{L} , de $1/\lambda$, expresada por

$$\hat{L} = r \times \frac{\sum_{i=1}^n f_i}{nt - \sum_{i=1}^n f_i}$$

donde el denominador representa el total de tiempo gastado en la reparación de todas las máquinas.

Se estimó empíricamente la correlación entre \hat{S} y \hat{L} logrando un valor alrededor de 0.95 para diferentes valores de n , t y del producto λr . Entonces, S se estimó mediante variables de control ajustadas por regresión como

$$\hat{S}^* = \hat{S} - \beta \left(\hat{L} - \frac{1}{\lambda} \right)$$

donde β viene dado por la ecuación (2.6), usando los estimadores de los momentos de segundo orden obtenidos en un estudio piloto. Además, la estimación del porcentaje de reducción de la varianza, $\left(1 - \frac{\text{var}(\hat{S}^*)}{\text{var}(\hat{S})} \right) \times 100$, dio un valor alrededor del 90%, corroborando así, la eficiencia de la técnica de variables de control para este proceso.

2.3 Muestreo condicional o Monte Carlo condicional

Nuestro objetivo es usar la simulación para encontrar propiedades de una variable aleatoria Y que es función de otras variables. El caso más sencillo sería condicionar respecto de una variable aleatoria univariante. La suposición básica es que cuando la variable aleatoria que condiciona, X , está fijada, se puede conocer la distribución o los momentos de Y . Esto es, se conoce $E_Y(Y|X = x)$, la

esperanza condicionada de Y con respecto a X cuando X es un valor dado. Más generalmente, se asume que $E_Y(g(Y)|X = x)$ se conoce para incluir otras cuantificaciones de la variable aleatoria Y . Por ejemplo, para percentiles, la función $g(Y)$ corresponde a una función dicotómica o indicador con parámetro y que se puede expresar como :

$$I(Y; y) = \begin{cases} 1 & \text{si } Y \leq y \\ 0 & \text{si } Y > y \end{cases}$$

Entonces, $E_Y(I(Y; y)|X = x) = P(Y \leq y|X = x)$.

Una vez que se conoce esta esperanza condicionada se puede aplicar la siguiente propiedad para obtener la esperanza de Y .

$$E(Y) = E_X[E_Y(Y|X)] \quad (2.8)$$

La forma de aplicar estas ideas en la simulación se basa en tener una muestra de X , (X_1, \dots, X_m) , y estimar $E(Y)$ como la media de X usando la expresión conocida de $E_Y(Y|X)$,

$$Y^c = \frac{1}{m} \sum_{i=1}^m E_Y(Y|X_i)$$

El estimador Y^c es un estimador insesgado de $E(Y)$, con una varianza teórica

$$\text{var}(Y^c) = \frac{1}{m} \text{var}_X\{E_Y(Y|X)\}$$

Para poder comparar este resultado con el que se obtendría a partir de una simulación sencilla, con $\text{var}(\bar{Y}) = (1/m) \text{var}(Y)$, se calcula la varianza de Y en términos condicionados. En efecto,

$$\begin{aligned} \text{var}(Y) &= E(Y^2) - E^2(Y) \\ &= E_X\{E_Y(Y^2|X)\} - [E_X\{E_Y(Y|X)\}]^2 \\ &= E_X\{E_Y(Y^2|X)\} - E_X\{E_Y^2(Y|X)\} + E_X\{E_Y^2(Y|X)\} - [E_X\{E_Y(Y|X)\}]^2 \end{aligned}$$

donde en la última línea se suma y se resta el mismo término para poder llegar a expresar:

$$\text{var}(Y) = E_X\{\text{var}_Y(Y|X)\} + \text{var}_X\{E_Y(Y|X)\} \quad (2.9)$$

Por tanto, la $\text{var}_X\{E_Y(Y|X)\} = \text{var}(Y) - E_X\{\text{var}_Y(Y|X)\}$ y como la $\text{var}_Y(Y|X)$ es una variable aleatoria positiva tendrá una esperanza positiva resultando que

$$\text{var}_X\{E_Y(Y|X)\} \leq \text{var}(Y)$$

Volviendo al estimador Y^c se tiene que

$$\begin{aligned} \text{var}(Y^c) &= \frac{1}{m} \text{var}_X \{E_Y(Y|X)\} = \frac{1}{m} [\text{var}(Y) - E_X \{\text{var}_Y(Y|X)\}] \\ &\leq \frac{1}{m} \text{var}(Y) = \text{var}(\bar{Y}) \end{aligned}$$

Esto es, $\text{var}(Y^c) \leq \text{var}(\bar{Y})$.

Así, aplicando el muestreo condicionado o Monte Carlo condicionado siempre habrá reducción de la varianza excepto en el peor de los casos que será nula. Además, se seguirá la regla que cuanto mayor sea $E_X \{\text{var}_Y(Y|X)\}$, mayor será la reducción de la varianza.

2.3.1 Ejemplo de aplicación de la técnica Monte Carlo condicional

Fue presentado por Simon (1976). Se desea estimar el valor medio, θ , de una variable aleatoria Y , la cual tiene una distribución beta $B(X, X^2 + 1)$, donde X es a la vez una variable aleatoria, con una distribución de Poisson de media conocida, λ . La simulación de Monte Carlo más sencilla sería simular n valores de Y y simplemente realizar su media. Sin embargo, esto incluye simular una distribución beta y una Poisson. Un punto de vista diferente sería aplicar la esperanza condicionada.

Se conoce que una variable aleatoria con distribución beta(α, β) tiene una esperanza igual a $\frac{\alpha}{\alpha + \beta}$, por tanto

$$E(Y|X = x) = \frac{x}{x^2 + x + 1}$$

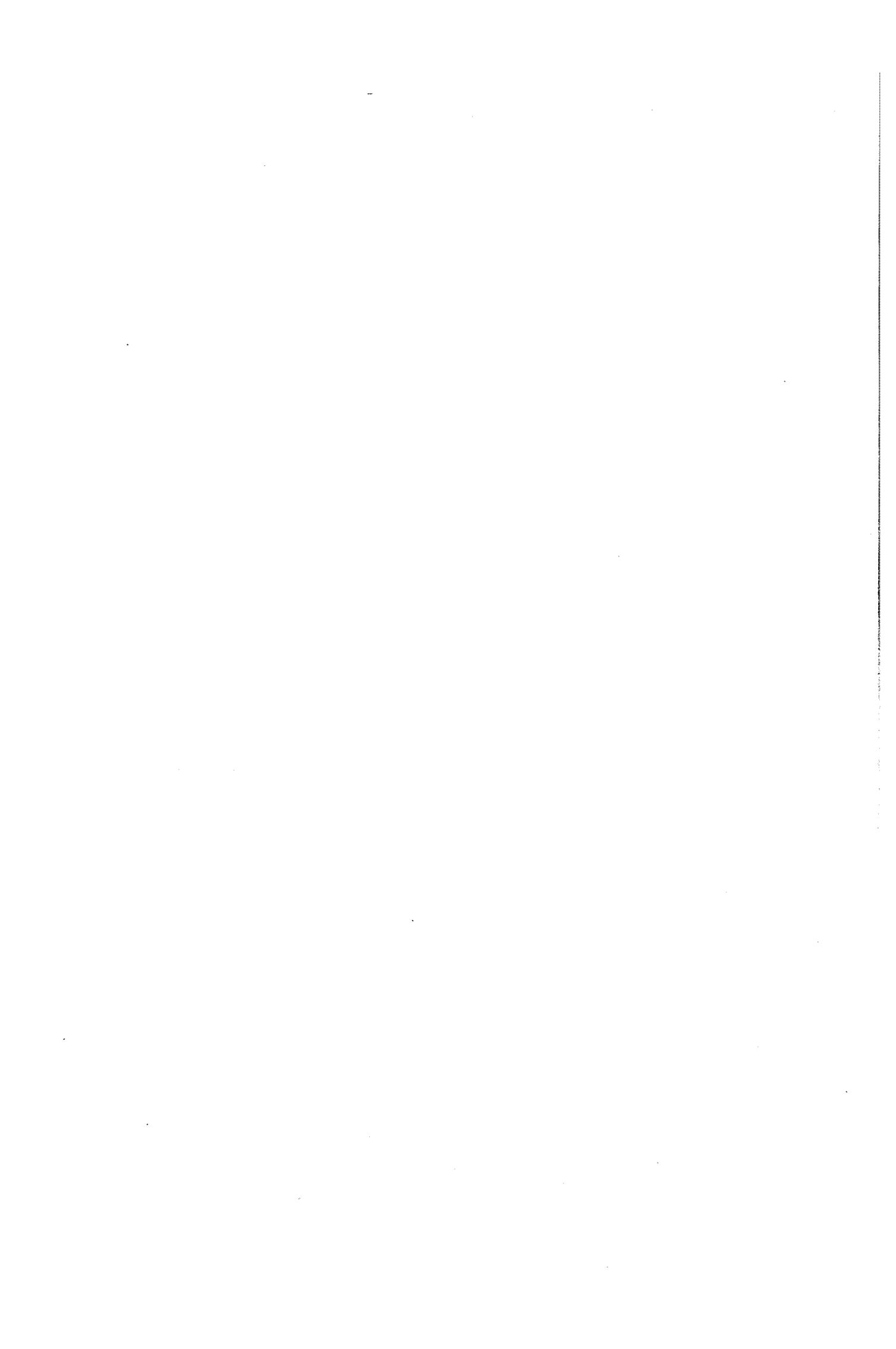
y aplicando la esperanza condicionada (2.8) se consigue

$$\sum_{x=0}^{\infty} E(Y|X = x) \frac{e^{-\lambda} \lambda^x}{x!} = \theta$$

Así, se puede estimar θ por

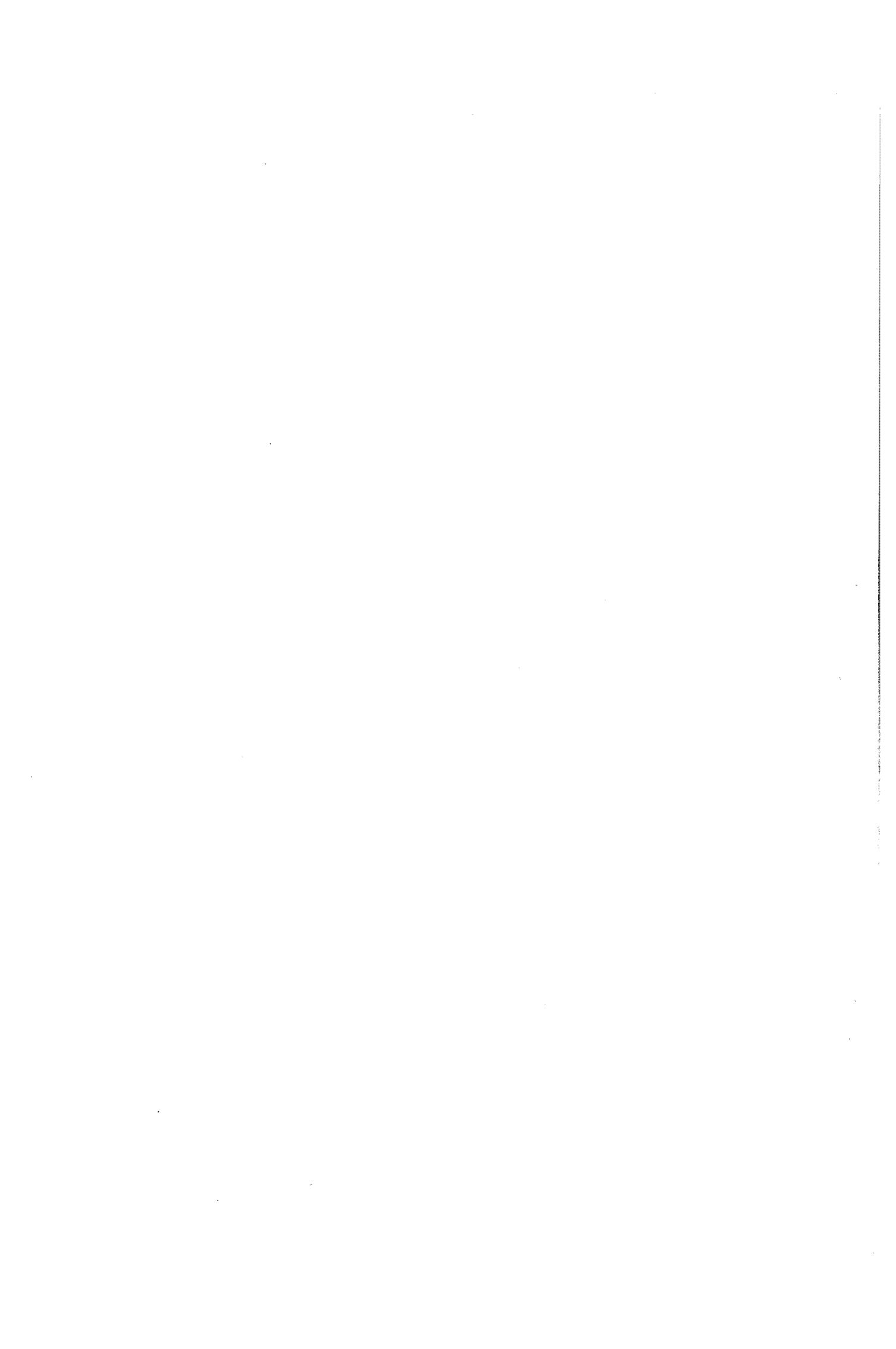
$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{x_i^2 + x_i + 1}$$

donde x_1, x_2, \dots, x_n es una muestra aleatoria con distribución de Poisson de parámetro λ que se puede realizar en un proceso de Monte Carlo. Adviértase que en este caso no hace falta simular la variable Y haciendo más sencilla su programación y además se obtiene una mayor precisión en la estimación de θ .



Capítulo 3

REDUCCIÓN DE LA VARIANZA PARA VARIABLES DE RESPUESTA DICOTÓMICA EN SIMULACIÓN



Se presentan algunas propiedades estadísticas del estimador asociado a una técnica de reducción de la varianza. Esta técnica es especialmente adecuada en simulaciones de Monte Carlo para estudios en Estadística, como por ejemplo en el estudio de la potencia de un nuevo test que será presentado en capítulos posteriores. Además, se mostrarán dos estudios de simulación para ilustrar la validez y uso de esta técnica de reducción de la varianza.

3.1 Introducción

Supongamos que en cada réplica de una simulación de Monte Carlo se obtiene un valor de una variable de respuesta Y con distribución de Bernoulli. Por ejemplo, Y puede ser el resultado final de un test que acepta o rechaza una hipótesis, o un intervalo de confianza el cual incluye o no el verdadero valor del parámetro. Normalmente se realizará la simulación para estimar la esperanza $p_1 = E(Y)$, por ejemplo para estimar la verdadera probabilidad de rechazo de la hipótesis nula o la verdadera probabilidad de recubrimiento. El estimador insesgado más obvio es la frecuencia relativa \hat{p}_1 del suceso $\{Y = 1\}$. Asumamos que, juntamente con un valor de Y , en cada réplica de simulación se produce un valor de otra variable dicotómica C , correlacionada con Y , cuya esperanza $p_1 = E(C) = P(C = 1)$ sea conocida. Por ejemplo C puede ser un test paramétrico cuya curva de potencia sea conocida, relacionado con el nuevo test Y bajo estudio de Monte Carlo o un intervalo de confianza relacionado cuyo verdadero recubrimiento sea conocido.

Al realizar el proceso de simulación se obtienen n pares de datos (Y_i, C_i) que se pueden resumir en forma de una tabla de contingencia 2×2 :

Tabla 1

	$C = 0$	$C = 1$	
$Y = 0$	n_{00}	n_{01}	$n_{0.}$
$Y = 1$	n_{10}	n_{11}	$n_{1.}$
	$n_{.0}$	$n_{.1}$	n

generada de acuerdo con las siguientes probabilidades p_{ij}

Tabla 2

	$C = 0$	$C = 1$	
$Y = 0$	p_{00}	p_{01}	$p_{0.}$
$Y = 1$	p_{10}	p_{11}	$p_{1.}$
	$p_{.0}$	$p_{.1}$	1

La función de masa de probabilidad conjunta del vector aleatorio (Y, C) con ambas variables aleatorias del tipo Bernoulli y bajo la condición de conocer el valor de $p_{.1}$ se puede expresar como:

$$p((y, c); p_{10}, p_{11}) = \begin{cases} p_{.0} - p_{10} & \text{si } (y = 0, c = 0) \\ p_{10} & \text{si } (y = 1, c = 0) \\ p_{.1} - p_{11} & \text{si } (y = 0, c = 1) \\ p_{11} & \text{si } (y = 1, c = 1) \end{cases} \quad (3.1)$$

$$= (p_{.0} - p_{10})^{(1-y)(1-c)} p_{10}^{y(1-c)} (p_{.1} - p_{11})^{(1-y)c} p_{11}^{yc} \quad y, c = 0, 1$$

Por lo tanto, la función de verosimilitud asociada a la muestra aleatoria simple presentada en la tabla de contingencia anterior quedará como:

$$L([(y_1, c_1), \dots, (y_n, c_n)]; p_{10}, p_{11}) = (p_{.0} - p_{10})^{n_{00}} p_{10}^{n_{10}} (p_{.1} - p_{11})^{n_{01}} p_{11}^{n_{11}}.$$

Las ecuaciones para los estimadores máximo verosímiles (EMV) de p_{10} y p_{11} son

$$\frac{\partial \ln L}{\partial p_{10}} = \frac{n_{10}}{p_{10}} - \frac{n_{00}}{p_{.0} - p_{10}} = 0$$

y

$$\frac{\partial \ln L}{\partial p_{11}} = \frac{n_{11}}{p_{11}} - \frac{n_{01}}{p_{.1} - p_{11}} = 0$$

de donde, simplificando, se obtiene

$$\tilde{p}_{10} = p_{.0} \frac{n_{10}}{n_{00} + n_{10}} \quad \text{y} \quad \tilde{p}_{11} = p_{.1} \frac{n_{11}}{n_{01} + n_{11}},$$

(obviamente, $\tilde{p}_{00} = p_{.0} - \tilde{p}_{10}$ y $\tilde{p}_{01} = p_{.1} - \tilde{p}_{11}$) y el EMV de $p_{.1}$ será

$$\tilde{p}_{.1} = p_{.0} \frac{n_{10}}{n_{00} + n_{10}} + p_{.1} \frac{n_{11}}{n_{01} + n_{11}}. \quad (3.2)$$

Este estimador fue introducido por Rothery (1982). Además, en ese mismo artículo, el autor compara este estimador con el estimador obtenido a partir de la técnica de reducción de varianza denominada "variables de control" utilizando como estimador el más sencillo posible (ver capítulo 2):

$$p_{1.}^s = \bar{Y} - (\bar{C} - p_{.1}) = \frac{n_{1.}}{n} - \left(\frac{n_{.1}}{n} - p_{.1} \right) \quad (3.3)$$

Rothery, concluye su artículo realizando unas pruebas de simulación donde observa que el estimador máximo verosímil reduce más la varianza que el estimador, $p_{1.}^s$, basado en la técnica de variable de control.

Nótese que este estimador, $\tilde{p}_{1.}$, es válido sólo para valores distintos de cero de $n_{.j} = n_{0j} + n_{1j}$. Por tanto, todas las propiedades de (3.2) y los estimadores relacionados que aparecerán más adelante son sólo válidos en este caso truncado, condicionado por el suceso $\{0 < n_{.0} < n\}$. Así, cuando se realicen simulaciones en los capítulos posteriores, los muy pocos casos con valor nulo de $n_{.0}$ o $n_{.1}$ serán eliminados.

3.2 Relación del estimador máximo verosímil, $\tilde{p}_{1.}$, con otros métodos

A continuación se presentan diferentes alternativas para obtener el estimador $\tilde{p}_{1.}$ basadas en la técnica de variables de control, esperanza condicionada y curva de regresión de la media. Para cerrar el círculo de las diferentes técnicas se comenta en el último apartado de esta sección la relación existente entre el estimador máximo verosímil y el estimador por mínimos cuadrados.

3.2.1 Estimador de variable de control ajustado por regresión

En el artículo de Rothery (1982) citado anteriormente, el estimador utilizado al usar la técnica de variable de control, $p_{1.}^s$, tiene una varianza mayor que el estimador máximo verosímil. Sin embargo, si seguimos usando un estimador de variable de control pero ahora ajustado por regresión (ver capítulo 2):

$$p_{1.}^* = \bar{Y} + \beta (\bar{C} - p_{.1}) = \frac{n_{1.}}{n} + \beta \left(\frac{n_{.1}}{n} - p_{.1} \right) \quad (3.4)$$

se obtiene, al estimar β a partir de los datos muestrales, el estimador máximo verosímil, \tilde{p}_1 .¹

En efecto, al desconocer la covarianza $cov(\bar{Y}, \bar{C}) = (p_{11} - p_{1.}p_{.1})/n$ que aparece en el cálculo del parámetro β , se puede estimar como $(\tilde{p}_{11} - \tilde{p}_{1.}p_{.1})/n$. Así, el estimador de β se convierte en

$$\hat{\beta} = -\frac{\tilde{p}_{11} - \tilde{p}_{1.}p_{.1}}{p_{.1}(1 - p_{.1})} = \frac{n_{10}}{n_{.0}} - \frac{n_{11}}{n_{.1}},$$

sustituyéndolo en (3.4), teniendo en cuenta que β se aproxima por $\hat{\beta}$, y simplificando se obtiene (3.2).

3.2.2 Estimador basado en la esperanza de Y condicionada a C

Alternativamente, se puede argumentar que, si Y y C son variables Bernouilli, la estimación de la esperanza de Y obtenida a partir de la esperanza condicionada de C producirá la misma estimación \tilde{p}_1 . Además, coincide con la técnica de reducción de la varianza denominada Monte Carlo condicional (ver Capítulo 2).

En efecto, a partir del teorema fundamental de la esperanza condicionada

$$E(Y) = E_C[E(Y|C)]$$

donde

$$\phi(C) = E(Y|C) = (1 - C) p_{1.|.0} + C p_{1.|.1} \quad (3.5)$$

con

$$\begin{aligned} p_{1.|.0} &= P\{Y = 1|C = 0\} \\ p_{1.|.1} &= P\{Y = 1|C = 1\}. \end{aligned} \quad (3.6)$$

Por tanto $E(Y|C)$ es una función lineal de C y coincide con la función de regresión lineal simple mínimo cuadrática:

$$\phi(C) = \alpha + \beta C = p_{1.|.0} + (p_{1.|.1} - p_{1.|.0}) C. \quad (3.7)$$

¹Adviértase que el estimador p_1^* es un caso particular de p_1^* si $\beta = 1$, de esta manera, se evita su estimación pero, al mismo tiempo, se tiene un estimador que reduce menos la varianza.

A partir de la propiedad de linealidad de la esperanza

$$E(Y) = E_C[\phi(C)] = \phi(E(C)) = \phi(p_{.1}) = p_{1.|.0} + (p_{1.|.1} - p_{1.|.0}) p_{.1}, \quad (3.8)$$

pero a causa del desconocimiento de las probabilidades condicionadas se reemplaza por sus estimaciones quedando como

$$\widehat{E}(Y) = E_C[\widehat{\phi}(C)] = \widehat{\phi}(E(C)) = \widehat{\phi}(p_{.1}) = \hat{p}_{1.|.0} + (\hat{p}_{1.|.1} - \hat{p}_{1.|.0}) p_{.1} \quad (3.9)$$

donde

$$\hat{p}_{1.|.0} = \frac{n_{10}}{n_{.0}}$$

$$\hat{p}_{1.|.1} = \frac{n_{11}}{n_{.1}}$$

que sustituyendo en (3.9) y simplificando se convierte en la igualdad (3.2).

3.2.3 Estimador basado en la curva de regresión de la media de Y versus C

Otra manera de llegar a la misma conclusión, utilizando los resultados de la esperanza condicionada, es a partir de la teoría de la regresión. Se puede demostrar que de todas las curvas posibles función de C , $f(C)$, la que cumple que:

$$E((Y - f(C))^2) = \text{mínimo}$$

es $E(Y|C)$, que es denominada "curva de regresión de la media". En este caso, es una función lineal de C (3.5), obteniéndose la función de regresión simple mínimo cuadrática (3.7). Además, a partir de la linealidad de la esperanza se obtiene que $\phi(p_{.1})$ es la "curva de regresión de la media" de $E(Y)$ en función de $p_{.1}$. Con lo cual, (3.9) es la estimación de curva óptima en el sentido de mínimos cuadrados.

3.2.4 Estimador basado en la esperanza de \bar{Y} condicionada a \bar{C}

Ahora se plantearán las mismas ideas que en la sección 3.2.2 pero para la media aritmética de Y , \bar{Y} , y para la media aritmética de C , \bar{C} . Entonces, teniendo en cuenta que la $E(Y) = E(\bar{Y})$ y que en la técnica de variables de control se conoce

$E(C) = E(\bar{C})$ se puede estimar la esperanza de \bar{Y} en función de la esperanza de \bar{C} dando el mismo resultado que (3.9).

En efecto, utilizando el teorema fundamental de la esperanza condicionada

$$E(Y) = E(\bar{Y}) = E_{\bar{C}}[E(\bar{Y}|\bar{C})]$$

donde

$$\begin{aligned} \phi(\bar{C}) &= E(\bar{Y}|\bar{C}) = \frac{1}{n} \sum_{j=1}^n E(Y_j|\bar{C}) = P(Y_j = 1|\bar{C}) \\ &= P(Y_j = 1 | \sum_{i=1}^n C_i = nk). \end{aligned} \quad (3.10)$$

Si se aplica la fórmula de Bayes y se sustituye nk por s queda como

$$\frac{P(Y_j = 1) P(\sum_{i=1}^n C_i = s | Y_j = 1)}{P(\sum_{i=1}^n C_i = s)} \quad (3.11)$$

donde la probabilidad condicionada de la expresión anterior se calcula utilizando la fórmula de probabilidades totales en función del valor, 0 o 1, de C_j .

$$\begin{aligned} P(\sum_{i=1}^n C_i = s | Y_j = 1) &= \sum_{h=0}^1 P_{[Y_j=1]}(C_j = h) P_{[Y_j=1]}(\sum_{i=1}^n C_i = s | C_j = h) \\ &= \sum_{h=0}^1 P_{[Y_j=1]}(C_j = h) P_{[Y_j=1]}(\sum_{i \neq j}^n C_i = s - h) \end{aligned} \quad (3.12)$$

La variable aleatoria $(C_j | Y_j = 1)$ sigue una distribución Bernouilli de probabilidad $\frac{p_{11}}{p_{1.}}$ y la variable aleatoria $(\sum_{i \neq j}^n C_i | Y_j = 1)$ sigue una distribución binomial $b(n-1, p_{1.})$. Por tanto, el resultado anterior es

$$= \sum_{h=0}^1 \frac{p_{1h}}{p_{1.}} \binom{n-1}{s-h} p_{1.}^{(s-h)} p_{0.}^{(n-1-(s-h))}.$$

Sabiendo que Y_j es una variable Bernouilli de probabilidad $p_{1.}$ y que $\sum_{i=1}^n C_i$ es una variable binomial $b(n, p_{1.})$ la expresión de la fórmula de Bayes (3.11) se reduce a

$$\frac{\sum_{h=0}^1 p_{1h} \binom{n-1}{s-h} p_{1.}^{(s-h)} p_{0.}^{(n-1-(s-h))}}{\binom{n}{s} p_{1.}^s p_{0.}^{(n-s)}} \quad (3.13)$$

que simplificando se obtiene finalmente

$$\phi(\bar{C}) = E(\bar{Y}|\bar{C}) = p_{1.,0} + (p_{1.,1} - p_{1.,0})\bar{C} \quad (3.14)$$

donde $p_{1.,0}$ y $p_{1.,1}$ son probabilidades condicionadas definidas en (3.6).

Aplicando la propiedad de linealidad de la esperanza

$$E(\bar{Y}) = E_{\bar{C}}[\phi(\bar{C})] = \phi(E(\bar{C})) = \phi(p_{.1}) = p_{1..0} + (p_{1..1} - p_{1..0}) p_{.1}$$

que es el mismo resultado obtenido en (3.8). Y a partir de aquí, como ya se vio, se puede obtener (3.9) y, posteriormente (3.2).

3.2.5 Estimador basado en la curva de regresión de la media de \bar{Y} versus \bar{C}

De manera análoga al punto 3.2.3 se puede observar que la curva de regresión de la media de \bar{Y} versus \bar{C} , $E(\bar{Y}|\bar{C})$, es una función lineal de \bar{C} (3.14), con lo cual, es una función de regresión simple mínimo cuadrática que coincide con el estimador, $p_{.1}^*$, de variable de control ajustado por regresión (3.4).

3.2.6 Mínimos cuadrados ordinarios de Y_i/C_i sabiendo que $p_{.1}$ es conocido

A partir de aplicar el criterio de mínimos cuadrados ordinarios para una muestra de n parejas de datos (Y_i, C_i) donde cada pareja es el resultado obtenido de un par de variables aleatorias dicotómicas Y y C con la condición de conocer $p_{.1}$ se logra el estimador máximo verosímil $\tilde{p}_{.1}$. De esta manera, se cierra el círculo de obtención del estimador $\tilde{p}_{.1}$ por máxima verosimilitud y por mínimos cuadrados.

En efecto, la función lineal entre dos variables aleatorias se puede expresar como:

$$\hat{Y}_i = \alpha + \beta C_i \quad (3.15)$$

donde $\alpha = E(Y) - \beta E(C)$ que al ser, (Y, C) , variables aleatorias Bernouilli la expresión anterior se convierte en $\alpha = p_{.1} - \beta p_{.1}$ que sólo dependerá de $p_{.1}$ y β al ser $p_{.1}$ conocida. Por consiguiente, la función a minimizar es:

$$h(p_{.1}, \beta) = \frac{1}{n} \sum_{i=1}^n (Y_i - \beta(C_i - p_{.1}) - p_{.1})^2 \quad (3.16)$$

con respecto a $p_{.1}$ y a β .

El resultado de las derivadas parciales es:

$$\begin{aligned} \frac{\partial h(p_{1.}, \beta)}{\partial p_{1.}} &= - \sum_{i=1}^n (Y_i - \beta(C_i - p_{1.}) - p_{1.}) = 0 \\ \frac{\partial h(p_{1.}, \beta)}{\partial \beta} &= - \sum_{i=1}^n (Y_i - \beta(C_i - p_{1.}) - p_{1.})(C_i - p_{1.}) = 0 \end{aligned} \quad (3.17)$$

De la primera ecuación, simplificando y despejando $p_{1.}$, se obtiene el estimador mínimo cuadrático ordinario, $p_{1.}^{mc}$

$$p_{1.}^{mc} = \bar{Y} - \beta (\bar{C} - p_{1.}) \quad (3.18)$$

Aplicando el resultado anterior en la segunda ecuación y simplificando aparece el estimador mínimo cuadrático ordinario de β

$$\beta^{mc} = \frac{S_{YC}}{S_C^2} \quad (3.19)$$

donde S_{YC} es la covarianza muestral de Y y C , es decir, $S_{YC} = 1/n \sum_{i=1}^n Y_i C_i - \bar{Y} \bar{C}$ y $S_C^2 = 1/n \sum_{i=1}^n C_i^2 - \bar{C}^2$ es la varianza muestral de Y . Pero aún se puede simplificar más. Por ser C variable dicotómica se tiene que $\sum_{i=1}^n C_i^2 = \sum_{i=1}^n C_i$ y si además, se expresa β^{mc} en función del número de veces que sale el valor (i, j) $i, j = 0, 1, n_{ij}$, indicado en la Tabla 2, queda como

$$\beta^{mc} = \frac{n_{11}}{n_{.1}} - \frac{n_{10}}{n_{.0}}$$

Si ahora, se sustituye β por β^{mc} en la expresión de $p_{1.}^{mc}$ se obtiene la expresión que se utiliza en la técnica de variables de control ajustada por regresión (3.4) o lo que es lo mismo, el estimador máximo verosímil $\tilde{p}_{1.}$.

$$p_{1.}^{mc} = \bar{Y} - \frac{S_{YC}}{S_C^2} (\bar{C} - p_{1.}) = \bar{Y} + \left(\frac{n_{10}}{n_{.0}} - \frac{n_{11}}{n_{.1}} \right) (\bar{C} - p_{1.}) = \tilde{p}_{1.}$$

3.3 Propiedades del nuevo estimador

EMV

Se estudiarán las propiedades típicas de cualquier estimador: sesgo, eficiencia, consistencia, normalidad, etc. Para ello, se tendrá en consideración los métodos de obtención del nuevo estimador, $\tilde{p}_{1.}$.

3.3.1 Sesgo

A partir del cálculo directo de la esperanza del estimador \tilde{p}_1 . (recuérdese que condicionado a $\{0 < n_{.0} < n\}$) se puede deducir que es un estimador insesgado, es decir que $E(\tilde{p}_1) = p_1$. En efecto, por una parte

$$E(\tilde{p}_1) = p_{.0}E\left(\frac{n_{10}}{n_{.0}}\right) + p_{.1}E\left(\frac{n_{11}}{n_{.1}}\right),$$

y calculando las esperanzas de la parte derecha de la igualdad anterior:

$$\begin{aligned} E\left(\frac{n_{1i}}{n_{.i}}\right) &= \sum_{n_{.i} > 0; n_{1i}} \frac{n_{1i}}{n_{.i}} P\{n_{.i}, n_{1i}\} \\ &= \sum_{n_{.i} > 0} \frac{1}{n_{.i}} P\{n_{.i}\} \sum_{n_{1i}} n_{1i} P\{n_{1i}|n_{.i}\}, \end{aligned}$$

donde el último sumatorio corresponde a la esperanza condicionada de n_{1i} . Debido a que n_{1i} condicionado a $n_{.i}$ sigue una distribución binomial, $b\left(n_{.i}, \frac{p_{1i}}{p_{.i}}\right)$, el último sumatorio se convierte en $n_{.i} \frac{p_{1i}}{p_{.i}}$ y el resultado se deduce inmediatamente.

3.3.2 Varianza

Otra propiedad, tan importante como el sesgo del nuevo estimador, \tilde{p}_1 , es su varianza. A partir de cálculos directos se ha podido obtener la siguiente fórmula para la verdadera varianza de \tilde{p}_1 , teniendo en cuenta que $\{0 < n_{.0} < n\}$:

$$var(\tilde{p}_1) = p_{00}p_{10}E\{n_{.0}^{-1}\} + p_{01}p_{11}E\{n_{.1}^{-1}\} \tag{3.20}$$

donde $E\{n_{.i}^{-1}\}$ es la esperanza de la inversa de la variable binomial $n_{.i}$, truncada a $\{0 < n_{.0} < n\}$.

En efecto, como \tilde{p}_1 es un estimador insesgado, su varianza se puede expresar como:

$$var(\tilde{p}_1) = E(\tilde{p}_1^2) - p_1^2,$$

donde

$$E(\tilde{p}_1^2) = \sum_{n_{.1}, n_{11}, n_{10}} \left(p_{.0} \frac{n_{10}}{n_{.0}} + p_{.1} \frac{n_{11}}{n_{.1}}\right)^2 P(n_{.1}, n_{11}, n_{10}),$$

y, como n_{10} y n_{11} son condicionalmente independientes bajo $n_{.1}$,

$$P(n_{.1}, n_{11}, n_{10}) = P(n_{.1})P(n_{10}|n_{.0})P(n_{11}|n_{.1}).$$

A partir de la anterior descomposición y usando, de nuevo, la propiedad que n_{1i} condicionado a $n_{.i}$ sigue una distribución binomial, $b\left(n_{.i}, \frac{p_{1i}}{p_{.i}}\right)$, nos conduce a

$$E(\tilde{p}_{1.}^2) = \sum_{n_{.1}} P(n_{.1})\{A_0 + A_1 + B\},$$

donde

$$A_i = \left(\frac{p_{.i}}{n_{.i}}\right)^2 (\text{var}(n_{1i}|n_{.i}) + E^2(n_{1i}|n_{.i})) = \frac{p_{1i}p_{0i}}{n_{.i}} + p_{1i}^2 \quad \text{para } i = 0, 1$$

y

$$B = \frac{2p_{.0}p_{.1}}{n_{.0}n_{.1}} E(n_{10}|n_{.0})E(n_{11}|n_{.1}) = 2p_{10}p_{11}.$$

Por tanto,

$$E(\tilde{p}_{1.}^2) = \sum_{n_{.1}} \left\{ \frac{p_{10}p_{00}}{n_{.0}} + \frac{p_{11}p_{01}}{n_{.1}} \right\} P(n_{.1}) + p_{1.}^2. \quad (3.21)$$

y el resultado (3.20) se obtiene directamente.

Otra forma de conseguir (3.20) es a partir de

$$\text{var}(\tilde{p}_{1.}) = \text{var}(E\{\tilde{p}_{1.}|n_{.1}\}) + E(\text{var}\{\tilde{p}_{1.}|n_{.1}\}) \quad (3.22)$$

teniendo en cuenta que, tanto las varianzas como las esperanzas, están condicionadas por $\{0 < n_{.0} < n\}$. La deducción de la ecuación (3.22) se presenta en la sección Monte Carlo condicional del capítulo 2.

En efecto, como $n_{.0}$ es lo que condiciona, tanto $n_{.0}$ como $n_{.1}$, se tratarán como constantes si aparecen en la esperanza o en la varianza, así la esperanza del primer sumando puede quedar como

$$E\{\tilde{p}_{1.}|n_{.1}\} = \frac{p_{.0}}{n_{.0}} E(n_{10}|n_{.0}) + \frac{p_{.1}}{n_{.1}} E(n_{11}|n_{.1})$$

y de nuevo aplicando la propiedad que n_{1i} condicionado a $n_{.i}$ sigue una distribución binomial, $b\left(n_{.i}, \frac{p_{1i}}{p_{.i}}\right)$, obtenemos que es igual $p_{1.}$. Por tanto el primer sumando valdrá cero,

$$\text{var}(E\{\tilde{p}_{1.}|n_{.1}\}) = \text{var}(p_{1.}) = 0$$

La varianza del segundo sumando es

$$\text{var}\{\tilde{p}_{1.}|n_{.1}\} = \frac{p_{.0}^2}{n_{.0}^2} \text{var}(n_{10}|n_{.0}) + \frac{p_{.1}^2}{n_{.1}^2} \text{var}(n_{11}|n_{.1}) + \frac{2p_{.1}p_{.0}}{n_{.1}n_{.0}} \text{cov}(n_{10}, n_{11}|n_{.1})$$

Al usar la propiedad anterior, $\frac{n_{1i}}{n_{.i}} \sim b\left(n_{.i}, \frac{p_{1i}}{p_{.i}}\right)$, y saber que la covarianza es cero ya que n_{10} y n_{11} son condicionalmente independientes bajo $n_{.1}$ resulta que

$$\text{var}\{\tilde{p}_{1.}|n_{.1}\} = \frac{p_{10}p_{00}}{n_{.0}} + \frac{p_{11}p_{01}}{n_{.1}}$$

con lo cual su esperanza se convierte en la igualdad (3.20).

Esta varianza (3.20) nunca es mayor que la varianza de la frecuencia relativa de $\{Y = 1\}$, $\hat{p}_{1.}$:

$$\text{var}(\hat{p}_{1.}) = \frac{p_{1.}(1 - p_{1.})}{n}.$$

3.3.3 Eficiencia

Conocemos la varianza del nuevo estimador pero ¿este valor corresponde al valor de la cota de Cramer-Rao? En esta sección se plantea buscar la cota de Cramer-Rao en la estimación de $p_{1.}$ y ver si el estimador propuesto, $\tilde{p}_{1.}$, llega a esta cota. En caso contrario, se buscará la existencia de un nuevo estimador que se aproxime más a ese valor.

Para ello se tiene en consideración que $p_{1.}$ se obtiene como suma de dos parámetros, p_{10} y p_{11} , y que, por otra parte, se conoce la función de masa de probabilidad conjunta del vector aleatorio (Y, C) indicado en (3.1).

Como la función de probabilidad depende de dos parámetros, p_{10} y p_{11} , la generalización de la cota de Cramer-Rao será la elipse fija (ver Cramer, 1953, pp 562-567).

$$\begin{aligned} n \left[E \left(\frac{\partial \ln p}{\partial p_{10}} \right)^2 (u - p_{10})^2 + 2E \left(\frac{\partial \ln p}{\partial p_{10}} \frac{\partial \ln p}{\partial p_{11}} \right) (u - p_{10})(v - p_{11}) + \right. \\ \left. E \left(\frac{\partial \ln p}{\partial p_{11}} \right)^2 (v - p_{11})^2 \right] = 4 \end{aligned} \quad (3.23)$$

que está por completo contenida dentro de la elipse de concentración de la distribución conjunta de todo par de estimaciones regulares insesgadas de p_{10} y p_{11} que tiene por ecuación

$$\frac{1}{1 - \rho^2} \left(\frac{(u - p_{10})^2}{\sigma_1^2} - \frac{2\rho(u - p_{10})(v - p_{11})}{\sigma_1\sigma_2} + \frac{(v - p_{11})^2}{\sigma_2^2} \right) = 4 \quad (3.24)$$

donde σ_1 , σ_2 y ρ representan las desviaciones estándar y el coeficiente de correlación de las estimaciones regulares insesgadas de p_{10} y p_{11} .

Entonces, el primer paso es calcular las derivadas parciales

$$\frac{\partial \ln p}{\partial p_{10}} = \frac{-(1-y)(1-c)}{p_{.0} - p_{10}} + \frac{y(1-c)}{p_{10}}$$

$$\frac{\partial \ln p}{\partial p_{11}} = \frac{-(1-y)c}{p_{.1} - p_{11}} + \frac{yc}{p_{11}}$$

A continuación, se calcula las esperanzas

$$E\left(\frac{\partial \ln p}{\partial p_{10}}\right)^2 = \frac{1}{p_{.0} - p_{10}} + \frac{1}{p_{10}}$$

$$E\left(\frac{\partial \ln p}{\partial p_{11}}\right)^2 = \frac{1}{p_{.1} - p_{11}} + \frac{1}{p_{11}}$$

$$E\left(\frac{\partial \ln p}{\partial p_{10}} \frac{\partial \ln p}{\partial p_{11}}\right) = 0$$

de modo que la elipse óptima (3.23) es igual a:

$$\frac{p_{.0}}{(p_{.0} - p_{10})p_{10}}(u - p_{10})^2 + \frac{p_{.1}}{(p_{.1} - p_{11})p_{11}}(v - p_{11})^2 = \frac{4}{n} \quad (3.25)$$

Si p_{10}^e y p_{11}^e son un par de estimaciones eficientes conjuntas, las varianzas de p_{10}^e y de p_{11}^e , y el coeficiente de correlación entre estas dos estimaciones, se obtienen a partir de la recíproca de la forma cuadrática del primer miembro de (3.23)

$$\text{var}(p_{10}^e) = \frac{1}{n\Delta} E\left(\frac{\partial \ln p}{\partial p_{11}}\right)^2$$

$$\text{var}(p_{11}^e) = \frac{1}{n\Delta} E\left(\frac{\partial \ln p}{\partial p_{10}}\right)^2$$

$$\rho(p_{10}^e, p_{11}^e) = \frac{E\left(\frac{\partial \ln p}{\partial p_{10}} \frac{\partial \ln p}{\partial p_{11}}\right)}{\sqrt{E\left(\frac{\partial \ln p}{\partial p_{10}}\right)^2 E\left(\frac{\partial \ln p}{\partial p_{11}}\right)^2}}$$

en donde

$$\Delta = E \left(\frac{\partial \ln p}{\partial p_{11}} \right)^2 E \left(\frac{\partial \ln p}{\partial p_{10}} \right)^2 - E^2 \left(\frac{\partial \ln p}{\partial p_{10}} \frac{\partial \ln p}{\partial p_{11}} \right).$$

Por tanto, obtenemos los siguientes resultados

$$\text{var}(p_{10}^e) = \frac{(p_{.0} - p_{10})p_{10}}{np_{.0}}, \quad \text{var}(p_{11}^e) = \frac{(p_{.1} - p_{11})p_{11}}{np_{.1}}, \quad \rho(p_{10}^e, p_{11}^e) = 0 \quad (3.26)$$

Pero como lo que nos interesa es la varianza mínima para la estimación de p_1 , esta será la suma de las dos varianzas ya que la correlación es cero.

$$\text{var}(p_{1.}^e) = \text{var}(p_{10}^e) + \text{var}(p_{11}^e) = \frac{(p_{.0} - p_{10})p_{10}}{np_{.0}} + \frac{(p_{.1} - p_{11})p_{11}}{np_{.1}} \quad (3.27)$$

Entonces, (3.27) será la cota de Cramer-Rao, varianza del estimador eficiente de $p_{1.}$, si éste existe.

Para el par de estimaciones, \tilde{p}_{10} y \tilde{p}_{11} , obtenidas por máxima verosimilitud la elipse de concentración (3.24) queda como

$$\frac{(u - p_{10})^2}{p_{00}p_{10}E(n_{.0}^{-1})} + \frac{(v - p_{11})^2}{p_{01}p_{11}E(n_{.1}^{-1})} = 4 \quad (3.28)$$

que no es igual a la elipse óptima (3.25).

La razón entre la varianza bidimensional de un par de estimaciones conjuntamente eficientes p_{10}^e, p_{11}^e y la cantidad correspondiente a cualquier par de estimaciones insesgadas regulares p_{10}^r, p_{11}^r es la eficiencia conjunta de p_{10}^r, p_{11}^r , que se representará por $e(p_{10}^r, p_{11}^r)$. Esta es idéntica al cuadrado de la razón entre las áreas de las elipses (3.23) y (3.24), la cual se puede expresar como:

$$\frac{1}{n^2 \Delta \sigma_1^2 \sigma_2^2 (1 - \rho)^2}$$

que para $\tilde{p}_{10}, \tilde{p}_{11}$ quedará como

$$e(\tilde{p}_{10}, \tilde{p}_{11}) = \frac{1}{n^2 p_{.0} p_{.1} E(n_{.0}^{-1}) E(n_{.1}^{-1})}$$

que tiende a uno y, por tanto, el estimador es asintóticamente eficiente.

Por consiguiente, la $\text{var}(\tilde{p}_{1.})$ (3.20) no llegará a la cota de Cramer-Rao (3.27). No obstante, si calculamos la eficiencia de $\tilde{p}_{1.}$ como el cociente entre el valor

mínimo de varianza de $p_{1.}$ y el valor de la varianza de $\tilde{p}_{1.}$ será menor que uno pero asintóticamente eficiente.

Una justificación de que la EMV no sea eficiente y que sólo sea asintóticamente eficiente se basa en el hecho que la función de probabilidad conjunta (3.1) pertenece a una familia exponencial curvada, es decir, es una variedad no plana de una familia exponencial, puesto que existe, al menos, una restricción sobre los parámetros naturales que no es función lineal de los mismos.

En efecto, la familia exponencial se caracteriza por poder expresar la función de densidad o de probabilidad de la forma:

$$p(x; \omega) = a(\omega)b(x) \exp \left\{ \sum_{i=1}^k \theta_i(\omega)t_i(x) \right\} \quad (3.29)$$

donde ω es el parámetro (ambos, x y ω , pueden ser multidimensionales), $\theta = \theta(\omega) = (\theta_1(\omega), \dots, \theta_k(\omega))$ y $t = t(x) = (t_1(x), \dots, t_k(x))$ son vectores de dimensión común, k , que se denominan parámetro natural o canónico y estadístico privilegiado o canónico, respectivamente.

En nuestro caso, la función de probabilidad conjunta del vector aleatorio (Y, C) asociado a la tabla de contingencia 2×2 (Tabla 1) generadas de acuerdo a las probabilidades p_{ij} (Tabla 2) se asociaría a una multinomial, $M(n, p_{00}, p_{10}, p_{01}, p_{11})$.

$$p((n_{00}, n_{01}, n_{10}, n_{11}); p_{00}, p_{10}, p_{01}, p_{11}) = \frac{n!}{n_{00}! n_{01}! n_{10}! n_{11}!} p_{00}^{n_{00}} p_{10}^{n_{10}} p_{01}^{n_{01}} p_{11}^{n_{11}} \quad (3.30)$$

que se puede reescribir según la familia exponencial (3.29) como

$$p((n_{00}, n_{01}, n_{10}, n_{11}); p_{00}, p_{10}, p_{01}, p_{11}) = (1 + e^{\theta_1} + e^{\theta_2} + e^{\theta_3})^{-n} \frac{n!}{n_{00}! n_{01}! n_{10}! n_{11}!} \exp\{n_{01}\theta_1 + n_{10}\theta_2 + n_{11}\theta_3\} \quad (3.31)$$

donde $\theta = (\ln(p_{01}/p_{00}), \ln(p_{10}/p_{00}), \ln(p_{11}/p_{11}))$ es el vector paramétrico natural minimal, y estos parámetros varían sobre \mathfrak{R}^3 .

Además, sabemos que $p_{.1}$ es conocida, así que también, conocemos $p_{.0} = 1 - p_{.1}$. Si expresamos estas condiciones en función de los parámetros naturales

$$p_{.1}(\theta_1, \theta_2, \theta_3) = p_{01}(\theta_1, \theta_2, \theta_3) + p_{11}(\theta_1, \theta_2, \theta_3) = (1 + e^{\theta_1} + e^{\theta_2} + e^{\theta_3})^{-1} \{e^{\theta_2} + e^{\theta_3}\} \quad (3.32)$$

y

$$p_{.0}(\theta_1, \theta_2, \theta_3) = 1 - (1 + e^{\theta_1} + e^{\theta_2} + e^{\theta_3})^{-1} \{e^{\theta_2} + e^{\theta_3}\} \quad (3.33)$$

se observa que no son función lineal de éstos.

Por tanto, los parámetros, bajo la anterior condición, varían sobre una variedad no plana \mathfrak{R}^2 de \mathfrak{R}^3 , con lo cual nos indica que es una familia exponencial curvada².

Debido a que estamos en una variedad \mathfrak{R}^2 de \mathfrak{R}^3 el estimador máximo verosímil obtenido en esa variedad es una proyección del estimador máximo verosímil de la familia exponencial base, siendo ésta la que puede alcanzar la cota de Cramer-Rao en la familia completa. Por consiguiente, los estimadores \tilde{p}_{10} y \tilde{p}_{11} , así como, la del estimador suma de ambos, \tilde{p}_1 , no alcanzan la cota de Cramer-Rao.

3.3.4 Otras propiedades

A partir de las diferentes formas de obtener el nuevo estimador se puede indicar las siguientes características:

- Como es un EMV tiene las propiedades de este tipo de estimadores (Cramer, 1953, pp. 570-573). Si existe una estimación eficiente p_1^e de p_1 , esta será \tilde{p}_1 , y, a la vez, si existe una estimación suficiente esta será función de \tilde{p}_1 . Además, es asintóticamente normal y asintóticamente eficiente.
- Tanto para la $E(Y|C)$ como para $E(\bar{Y}|\bar{C})$ se observa que es función lineal de la variable que condiciona, C o \bar{C} , respectivamente. Con lo cual, esta función es la mejor predicción de Y en función de C o de \bar{Y} en función de \bar{C} en el sentido de mínimos cuadrados, respectivamente. En ambos casos, coincide la estimación de $E(Y)$ con el resultado (3.2) obtenido directamente por EMV.

3.4 Estimadores de la varianza del nuevo estimador

Como no es posible conocer el verdadero valor de la varianza del nuevo estimador \tilde{p}_1 , nos vemos obligados a buscar su mejor estimación. Por tanto, se han

²En el próximo capítulo se insistirá más en el concepto de familia exponencial curvada.

ido proponiendo y probando diferentes estimadores y viendo sus características, escogiendo finalmente uno de ellos como el más óptimo.

3.4.1 Estimador propuesto por Rothery de $var(\tilde{p}_1)$

Del método de máxima verosimilitud se sabe que sus estimadores tienen la propiedad de ser asintóticamente eficientes, así que la cota de Cramer-Rao obtenida en (3.27) se aproxima al valor de la $var(\tilde{p}_1)$ cuanto mayor sea el tamaño muestral. Así pues, un estimador congruente de la varianza anterior se consigue al sustituir los parámetros desconocidos por sus EMV: \tilde{p}_{10} , \tilde{p}_{00} , \tilde{p}_{01} y \tilde{p}_{11} . Esto es lo que hizo Rothery (1982) al proponer como estimador de la varianza de \tilde{p}_1 a

$$\hat{\sigma}_R^2 = \frac{\tilde{p}_{00}\tilde{p}_{10}}{np_{.0}} + \frac{\tilde{p}_{01}\tilde{p}_{11}}{np_{.1}}. \quad (3.34)$$

Alternativamente, se puede llegar a este mismo resultado a partir de otros dos caminos diferentes.

El primero es obtener el estimador de la varianza de \tilde{p}_1 a partir de la técnica de variables de control (p_1^*).

En efecto,

$$p_1^* = \bar{Y} + \beta(\bar{C} - p_{.1}) \simeq \tilde{p}_1 = \bar{Y} + \hat{\beta}(\bar{C} - p_{.1}) \quad (3.35)$$

así que, las varianzas teóricas serán semejantes

$$\begin{aligned} var(\tilde{p}_1) &\simeq var(p_1^*) \\ &\simeq var(\hat{p}_1)(1 - \rho^2) \end{aligned} \quad (3.36)$$

por lo tanto, tiene sentido estimar la $var(\tilde{p}_1)$ a partir de la estimación de la $var(p_1^*)$ quedando como

$$\hat{\sigma}_R^2 = \frac{1}{n} \left\{ \tilde{p}_1(1 - \tilde{p}_1) - \frac{(\tilde{p}_{1.p.1} - \tilde{p}_{11})^2}{(1 - p_{.1})p_{.1}} \right\} \quad (3.37)$$

donde el último término indica una forma directa de estimar la cantidad de reducción de la varianza.

El segundo es el estimador de la varianza de \tilde{p}_1 por el método delta de primer orden. Este se puede obtener como sigue (ver por ejemplo, Bishop *et al.*(1975), pp. 486-489).

Las derivadas parciales de la función

$$g(n_{11}, n_{01}, n_{10}) = \tilde{p}_1 = p_{.0} \frac{n_{10}}{n_{00} + n_{10}} + p_{.1} \frac{n_{11}}{n_{01} + n_{11}}$$

son:

$$\begin{aligned} \frac{\partial \tilde{p}_1}{\partial n_{11}} &= p_{.0} \frac{n_{10}}{(n - n_{.1})^2} + p_{.1} \frac{n_{01}}{n_{.1}^2} \\ \frac{\partial \tilde{p}_1}{\partial n_{01}} &= p_{.0} \frac{n_{10}}{(n - n_{.1})^2} - p_{.1} \frac{n_{11}}{n_{.1}^2} \\ \frac{\partial \tilde{p}_1}{\partial n_{10}} &= p_{.0} \frac{1}{(n - n_{.1})} \end{aligned}$$

Entonces, el desarrollo de Taylor de \tilde{p}_1 con respecto al punto $(\nu_{11}, \nu_{01}, \nu_{10})$, donde $\nu_{ij} = np_{ij}$, es

$$\begin{aligned} \tilde{p}_1 \simeq & \left(p_{.0} \frac{\nu_{10}}{\nu_{.0}} + p_{.1} \frac{\nu_{11}}{\nu_{.1}} \right) + \\ & \frac{1}{n} \left\{ \left(\frac{p_{10}}{p_{.0}} + \frac{p_{01}}{p_{.1}} \right) (n_{11} - \nu_{11}) + \left(\frac{p_{10}}{p_{.0}} - \frac{p_{11}}{p_{.1}} \right) (n_{01} - \nu_{01}) + (n_{10} - \nu_{10}) \right\}, \end{aligned} \tag{3.38}$$

el cual en término de la constante $\beta_\nu = \frac{p_{10}}{p_{.0}} - \frac{p_{11}}{p_{.1}}$ se convierte en

$$\tilde{p}_1 \simeq \frac{1}{n} \{ n_{.1} + \beta_\nu (n_{.1} - \nu_{.1}) \} = p_{1.}^*$$

que es igual a la expresión (3.4) utilizada en el estimador de variables de control ajustado por regresión $p_{1.}^*$. Por lo tanto, la aproximación obtenida por el método delta de primer orden coincide con la $var(p_{1.}^*)$ (3.36), que como vimos es una buena aproximación a la $var(\tilde{p}_1)$ dando

$$var(\tilde{p}_1) \simeq \frac{1}{n} \left\{ p_{1.} (1 - p_{1.}) - \frac{(p_{1.} p_{.1} - p_{11})^2}{(1 - p_{.1}) p_{.1}} \right\}.$$

Si se reemplaza los parámetros p por sus estimadores, se obtiene la varianza del estimador (3.37) en Vegas y Ocaña (1992) que, como es lógico, coincide con $\hat{\sigma}_R^2$.

El estimador de la varianza de Rothery, $\hat{\sigma}_R^2$, tiene un sesgo negativo bajo las condiciones usuales en el estudio de potencia por Monte Carlo. Esto se puede deducir de los siguientes argumentos:

La verdadera varianza de \tilde{p}_1 es (3.20) y la esperanza de $\hat{\sigma}_R^2$ es:

$$E(\hat{\sigma}_R^2) = p_{00} p_{10} \frac{1 - E\{n_{.0}^{-1}\}}{np_{.0}} + p_{01} p_{11} \frac{1 - E\{n_{.1}^{-1}\}}{np_{.1}}, \tag{3.39}$$

donde $E\{n_{.i}^{-1}\}$ es la esperanza de la inversa de la variable binomial $n_{.i}$, truncada a $\{0 < n_{.0} < n\}$.

Este último resultado (3.39) se deduce inmediatamente de la igualdad

$$E(\tilde{p}_{0i}\tilde{p}_{1i}) = p_{0i}p_{1i}(1 - E\{n_{.i}^{-1}\}).$$

que es directamente consecuencia de (3.21)

$$E(\tilde{p}_{0i}\tilde{p}_{1i}) = E((p_{.i} - \tilde{p}_{1i})\tilde{p}_{1i}) = p_{.i}E(\tilde{p}_{1i}) - E(\tilde{p}_{1i}^2) = p_{.i}p_{1i} - E(\tilde{p}_{1i}^2).$$

Finkner (1950) sugiere que

$$\frac{1}{np_{.i}} < E\{n_{.i}^{-1}\}$$

para "valores grandes de $np_{.i}$ ". Nuestros resultados empíricos confirman que esta afirmación es cierta si $np_{.i} \geq 2$. Por tanto, el sesgo negativo de $\hat{\sigma}_R^2$ viene de

$$\frac{1 - E\{n_{.i}^{-1}\}}{np_{.i}} < \frac{1}{np_{.i}} < E\{n_{.i}^{-1}\}.$$

3.4.2 Estimación insesgada de $var(\tilde{p}_{1.})$

Los resultados (3.20) y (3.39) sugieren una forma para corregir el sesgo de $\hat{\sigma}_R^2$ y, a la vez, de obtener un nuevo estimador de la varianza

$$\hat{\sigma}_U^2 = \tilde{p}_{00}\tilde{p}_{10} \frac{E\{n_{.0}^{-1}\}}{1 - E\{n_{.0}^{-1}\}} + \tilde{p}_{01}\tilde{p}_{11} \frac{E\{n_{.1}^{-1}\}}{1 - E\{n_{.1}^{-1}\}} \quad (3.40)$$

que es obviamente insesgado. Las esperanzas $E\{n_{.i}^{-1}\}$, para $i = 0, 1$, se conocen si n y $p_{.i}$ son conocidas, como ocurre en la presente situación (recuérdese que $p_{.i}$ corresponde a la esperanza, conocida, de la "variable de control"). Sin embargo, no existe una forma cerrada para esta esperanza, así pues, se debe calcular directamente sumando todos los posibles valores de $k^{-1}P\{n_{.i} = k\}$ para $k = 1, \dots, n - 1$.

3.4.3 Estimación aproximada al estimador insesgado de $var(\tilde{p}_{1.})$

Un estimador alternativo a (3.40) es:

$$\hat{\sigma}_{GS}^2 = \frac{\tilde{p}_{00}\tilde{p}_{10}}{np_{.0} - 2 + p_{.0}} + \frac{\tilde{p}_{01}\tilde{p}_{11}}{np_{.1} - 2 + p_{.1}} \quad (3.41)$$

donde $\hat{\sigma}_{GS}^2$ es un estimador casi insesgado de (3.20).

GS significa que (3.41) se basa en la aproximación $E\{n_{.i}^{-1}\} \simeq (np_{.i} - (1 - p_{.i}))^{-1}$. Esta aproximación se propuso en (Grab & Savage, 1954) a partir de observaciones empíricas y posteriormente se estudió analíticamente en (Wooff, 1985). Es válida para binomiales truncadas positivamente, $n_{.0} > 0$ y, separadamente, $n_{.1} > 0$, con tal que $(n + 1)p_{.i} - \frac{0.3}{n} > 2.94$ (Wooff, 1985), pero por nuestra parte se realizaron pruebas numéricas que muestran que su comportamiento es semejante bajo la condición $\{0 < n_{.0} < n\}$. Además, las diferencias aparecen cuando $np_{.i} < 10$, lo cual no es común en simulaciones de Monte Carlo en estudios de Estadística. Luego el estimador $\hat{\sigma}_{GS}^2$ es más sencillo y muy próximo a ser insesgado y, por tanto, preferible a $\hat{\sigma}_U^2$.

3.4.4 Estimación bootstrap de $var(\tilde{p}_1)$

Los métodos de cálculo intensivo, tal como el bootstrap (véase Efron, 1979 y 1982) pueden producir mejores estimadores de la varianza. Su uso se puede justificar (en términos puramente computacionales) cuando el tiempo adicional de ordenador requerido para obtener las remuestras a partir de los resultados de simulación, $((Y_1, C_1), \dots, (Y_n, C_n))$ (o, más exactamente, realizar el remuestreo de las nuevas tablas de contingencia a partir de las probabilidades conjuntas estimadas $\frac{n_{ij}}{n}$), es pequeño comparado con el tiempo de ordenador requerido para generar los propios resultados de simulación. A la estimación de la varianza de \tilde{p}_1 bootstrap se ha designado como $\hat{\sigma}_B^2$. En nuestro caso se calculó y estudió las propiedades de $\hat{\sigma}_B^2$ a partir de dos estudios de simulación presentados en la próxima sección.

3.4.5 Estimación de la reducción de varianza y del margen de error

No hemos obtenido de momento resultados teóricos concernientes a la estimación de la verdadera cantidad de reducción de varianza. Se ha de recordar que la falta de sesgo de los estimadores de la varianza no garantiza la falta de sesgo del estimador de la verdadera cantidad de reducción de la varianza basado en ellos. El verdadero porcentaje de reducción de varianza es la fracción

$$VR = \left(1 - \frac{var(\tilde{p}_1)}{var(\hat{p}_1)}\right) 100, \quad (3.42)$$

donde \hat{p}_1 es la frecuencia de el suceso $\{Y = 1\}$.

Su más obvio estimador es

$$\widehat{VR} = \left(1 - \frac{(n-1)\hat{\sigma}^2}{\hat{p}_1(1-\hat{p}_1)}\right) 100,$$

donde $\hat{\sigma}^2$ es uno de los estimadores de la varianza de \tilde{p}_1 precedentes, $\hat{\sigma}_U^2$, $\hat{\sigma}_{GS}^2$, $\hat{\sigma}_R^2$ o $\hat{\sigma}_B^2$. Sus propiedades se estudian por simulación en la siguiente sección.

Si $\hat{\sigma}$ es una de las estimaciones de la desviación estándar ($\hat{\sigma}_U$, $\hat{\sigma}_{GS}$, $\hat{\sigma}_R$ or $\hat{\sigma}_B$) entonces

$$\tilde{p}_1 \pm 1.959964\hat{\sigma} \quad (3.43)$$

define un intervalo de confianza del 95% para el verdadero valor de p_1 . (asumiendo aproximación a la normalidad para $(\tilde{p}_1 - p_1)\hat{\sigma}^{-1}$). El verdadero recubrimiento de estos intervalos ha sido también estimado por simulación.

3.5 Estudio de simulación

Las propiedades teóricas \tilde{p}_1 y sus estimadores de la varianza se ilustran en dos series de estudios de Monte Carlo presentados en Ocaña y Vegas (1995). Los resultados de estas simulaciones complementan unos resultados anteriores en Vegas y Ocaña (1992).

3.5.1 Un caso de estudio: potencia de un nuevo test

En Burbea y Oller (1982) se plantea un test que compara simultáneamente las medias μ_1, μ_2 y las varianzas σ_1^2, σ_2^2 de dos variables aleatorias normales independientes. Se basa en el criterio de distancia geodésica entre distribuciones³(Rao, 1945) que se expone, sucintamente, en un apéndice denominado "distancia de Rao". La hipótesis nula $\mu_1 = \mu_2, \sigma_1 = \sigma_2$ será rechazada si $U \geq A_\epsilon$, donde

$$U = \frac{n_1 n_2}{n_1 + n_2} \lambda^2,$$

λ es la distancia geodésica

$$\lambda = \sqrt{2} \log \frac{1 + \Delta}{1 - \Delta} = 2\sqrt{2} \tanh^{-1}(\Delta)$$

³En el capítulo 4 se insistirá sobre la distancia geodésica o de Rao

$$\Delta = \left\{ \frac{(\bar{x}_1 - \bar{x}_2)^2 + 2(s_1 - s_2)^2}{(\bar{x}_1 - \bar{x}_2)^2 + 2(s_1 + s_2)^2} \right\}^{1/2}$$

y (\bar{x}_1, s_1) , (\bar{x}_2, s_2) son los estimadores máximo verosímiles de los parámetros poblacionales, (μ_i, σ_i) $i = 1, 2$. A_ϵ debe satisfacer que la $P(U > A_\epsilon | H_0) = \epsilon$. Para muestras de gran tamaño, A_ϵ se puede obtener de acuerdo al teorema 2 en Burbea y Oller (1982) en el cual se afirma que U sigue asintóticamente una distribución ji-cuadrado con dos grados de libertad,

$$U \xrightarrow{\epsilon} \chi_2^2.$$

(Por cierto, los resultados de simulación indican que la mencionada aproximación por una distribución ji-cuadrado es muy pobre, y que un test basado en los mismos estadísticos pero usando una aproximación por bootstrap paramétrico funciona mucho mejor)

Una cualidad a tener en cuenta en cualquier test es su potencia. Por tanto, se debe conocer, al menos, una estimación de ésta que será mejor cuanto menor sea su varianza, así pues, un procedimiento adecuado es aplicar la metodología indicada en este capítulo. Para estimar la potencia del test propuesto por Burbea y Oller (1982), $p_{1.}$, se necesita otro test correlacionado con éste con potencia conocida, $p_{.1}$, que será empleado como variable de control. En este caso se opta por el test t de Student de igualdad de medias para dos variables normales independientes con igual varianzas.

Recordando la nomenclatura utilizada al principio del capítulo, si Y representa la variable dicotómica de interés y C la variable dicotómica de control entonces, en nuestro caso, se corresponde al test bajo estudio y al test t de Student, respectivamente. El proceso⁴ consiste en generar muestras $\mathbf{x} = (x_{11}, \dots, x_{1n_1}; x_{21}, \dots, x_{2n_2})$ donde \mathbf{x} está compuesta de dos submuestras de tamaños n_1 y n_2 generadas según la distribución normal de parámetros μ_1, σ_1^2 y μ_2, σ_2^2 , respectivamente. A continuación, se evalúa el test bajo estudio, si se acepta la hipótesis nula, $Y(\mathbf{x}) = 0$, en caso contrario, $Y(\mathbf{x}) = 1$. De manera análoga se hace para el test t de Student, pero el vector que se utiliza es $\mathbf{v} = (v_{11}, \dots, v_{1n_1}; v_{21}, \dots, v_{2n_2})$ que sirve para evaluar el test y obtener $C(\mathbf{v})$. El vector \mathbf{v} se genera bajo la condición de tener igual varianza (ya que en caso contrario no se conocería la potencia de control $p_{.1}$).

⁴En la sección Estudio de Simulación del próximo capítulo se aportan más detalles sobre todo el proceso de simulación e implementación

Para inducir la correlación necesaria entre $Y(\mathbf{x})$ y $C(\mathbf{v})$ se pueden usar variables aleatorias comunes en la generación de los distintos vectores \mathbf{x} y \mathbf{v} . Así, mientras las muestras usadas para evaluar el test t de Student se generan de acuerdo a una diferencia de medias $\Delta\mu = \mu_2 - \mu_1 = 0$ (potencia de control = 0.05) y $\Delta\mu = \mu_2 - \mu_1 = 0.5388159061$ (potencia de control = 0.55) para igual desviación estándar $\sigma_1 = \sigma_2 = 1$; las muestras usadas para evaluar el test bajo estudio se generan con los mismos valores de $\Delta\mu$ pero posiblemente con desviaciones estándar diferentes.

Para calcular las “verdaderas” propiedades del nuevo estimador, en una amplia variedad de condiciones (medias, varianzas y tamaños muestrales para cada una de las poblaciones bajo comparación), se realizaron 5 000 000 o 10 000 000 de pares (Y_i, C_i) de réplicas de simulación que fueron divididas en 10000 submuestras de tamaño 500 o 1000. Esto se hizo con la intención de producir resultados aplicables a simulaciones “reales” de tamaño $n = 500$ o 1000. Cada una de estas 10000 submuestras da un valor estimado de $\tilde{p}_1, \hat{\sigma}_R^2, \hat{\sigma}_{GS}^2$, etc (ver figura 1). Estos valores se reúnen para estimar la “verdadera” varianza de \tilde{p}_1 y la “verdadera” esperanza de los diversos estimadores de la varianza. Además, también se usan para estimar otras cantidades como el “verdadero” porcentaje de reducción de varianza comparada con la media de porcentaje de reducción de varianza pronosticado para cada estimador de la varianza. Nótese que cada una de las anteriores réplicas de simulación fue a la vez remuestreada 200 veces para calcular el estimador bootstrap, $\hat{\sigma}_B^2$.

A causa del gran número de réplicas de simulación requerido se tuvo que plantear técnicas intensivas de cálculo o también llamadas de supercomputación. Así, se utilizaron técnicas vectoriales de computación en un IBM 3090 600J corriendo bajo VM/XA en el CESCA (Centre de Supercomputació de Catalunya). Los programas se escribieron en VS FORTRAN v.2.0 r.3.0 usando el paquete de subrutinas vectoriales ESSL v.1.0 r.4.0. Para la generación de números aleatorios se empleó la subrutina DURAND del paquete ESSL que implementaba un generador congruencial multiplicativo siendo el multiplicador 16807 y la base $2^{31} - 1$. Para generar la distribución normal se usó la subrutina DNRAND del paquete ESSL. Se fundamenta en el método polar de Box y Muller (1958) en la implementación propuesta por Knuth (1981, pp 117-118).

Los resultados de las simulaciones aseveran los resultados teóricos que hacen referencia a la validez de la técnica de reducción de la varianza. Se confirman los

resultados respecto al sesgo: $\hat{\sigma}_U^2$ y $\hat{\sigma}_{GS}^2$ son casi iguales y, además, preferibles al estimador sesgado negativamente $\hat{\sigma}_R^2$. El estimador bootstrap, $\hat{\sigma}_B^2$, da resultados semejantes a los dos primeros estimadores anteriores.

La verdadera distribución de todos estos estimadores (\tilde{p}_1 y sus estimadores de la varianza) es aproximadamente normal, como se puede esperar de la teoría (máxima verosimilitud y método delta). Esto forma la base empírica para los métodos de inferencia basado en estos estimadores.

Tabla 3.

“Verdadero” porcentaje de reducción de la varianza comparado con la media, sobre todas las simulaciones, de los porcentajes estimados de reducción de la varianza con respecto a los diversos estimadores de la varianza $\hat{\sigma}_R^2$, $\hat{\sigma}_{GS}^2$, $\hat{\sigma}_U^2$ y $\hat{\sigma}_B^2$.

potencia (t Stud.)	$\sigma_1 = 1.05 \sigma_2 = 0.95$					$\sigma_1 = 1.10 \sigma_2 = 0.90$				
	Verd.	$\hat{\sigma}_R^2$	$\hat{\sigma}_{GS}^2$	$\hat{\sigma}_U^2$	$\hat{\sigma}_B^2$	Verd.	$\hat{\sigma}_R^2$	$\hat{\sigma}_{GS}^2$	$\hat{\sigma}_U^2$	$\hat{\sigma}_B^2$
0.05	19.09	19.40	18.78	18.75	18.77	12.07	12.06	11.77	11.76	11.80
0.55	53.92	53.49	53.40	53.40	53.41	45.92	45.99	45.87	45.87	45.88

La tabla 3 muestra la “verdadera” reducción de la varianza comparada con la media, sobre un total de 10000 submuestras, de las estimaciones de reducción de la varianza que corresponden al test t de Student con probabilidades de rechazo de 0.05 y 0.55, cuando las muestras usadas para evaluar los tests son ambas de igual tamaño $n_1 = n_2 = 31$ y las “simulaciones” (submuestra de la simulación completa) son de tamaño $n = 500$.

Estos resultados sugieren que todos los estimadores de la verdadera cantidad de reducción de la varianza son sesgados. Como era previsible, los estimadores basados en $\hat{\sigma}_{GS}^2$ y $\hat{\sigma}_U^2$ tienden a dar valores menores que el estimador basado en $\hat{\sigma}_R^2$ y, por otro lado, no muy diferentes de $\hat{\sigma}_B^2$. Tanto $\hat{\sigma}_{GS}^2$, $\hat{\sigma}_U^2$ y $\hat{\sigma}_B^2$ dan, de forma sistemática, estimaciones más bajas que la “verdadera” cantidad de reducción de la varianza lo cual es preferible a estimaciones superiores a esta cantidad, como pasa en algunas ocasiones con $\hat{\sigma}_R^2$.

Cuando se usa $\hat{\sigma}_{GS}^2$, $\hat{\sigma}_U^2$ y $\hat{\sigma}_B^2$, las estimaciones del porcentaje de recubrimiento del intervalo de confianza definido por (3.43) está muy próximo al 95%. Esto también se cumple para el usual intervalo de confianza basado en la frecuencia

relativa \hat{p}_1 , es decir,

$$\hat{p}_1 \pm 1.959964 \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n - 1}}.$$

Por otro lado, el intervalo de confianza basado en $\hat{\sigma}_R$ es muy cercano al 95% aunque sistemáticamente está por debajo.

Por tanto, estos resultados indican que las inferencias basadas en \tilde{p}_1 son adecuadas, especialmente cuando se usan $\hat{\sigma}_{GS}^2$, $\hat{\sigma}_U^2$ o $\hat{\sigma}_B^2$.

3.5.2 Un estudio de simulación general

La segunda serie de simulaciones se ejecutaron en ordenadores personales 386 y 486 usando el software en Borland Pascal 7.0 (EMSS) descrito en Sánchez *et al.*(1992). La generación de números aleatorios se basa en un generador muy rápido desarrollado por Zaman y Marsaglia, que combina un generador “substract-with-borrow” con base 2^{32} y retrasos (“lags”) de 37 y 24 (Marsaglia and Zaman, 1991) con un generador congruencial siendo el multiplicador 69069 y la base 2^{32} . El resultado final es la combinación de ambos generadores por medio de la operación *xor*. Debido al diseño orientado a objetos del software EMSS, se sustituyó el antiguo generador incorporado de números aleatorios, en el programa escrito para este estudio, por el más rápido generador de Marsaglia y Zaman sin recompilar las librerías de EMSS.

Las serie de simulaciones que se presentan no corresponden a ningún estudio de un test específico, intervalo de confianza, etc. Sirven para ilustrar la validez de la técnica de reducción de la varianza y los estimadores asociados a la varianza en un amplio rango de condiciones posibles y, además, validar los resultados de la anterior simulación.

Se generaron 10 000 (y algunas veces 1 000 000) tablas de contingencia semejantes a la tabla 1, cada una de ellas representa el resultado final de una “verdadera” simulación de un tamaño $n = 250, 500$ o 1000 . Estas tablas se generaron en correspondencia a los valores prefijados de p_{ij} . En algunos casos, se escogieron como valores de p_{ij} los obtenidos de \tilde{p}_{ij} en el estudio de simulación precedente, y el resultado fue totalmente compatible. En otros casos, se ejecutaron series de simulaciones para valores fijos de p_1 (en correspondencia a la “potencia del test bajo estudio”) y valores fijos de $p_{.1}$ (la “potencia del test de control”) con diferentes valores de p_{ij} , para simular diferentes grados de correlación entre la variable

respuesta y la de control. Todos los resultados estuvieron en concordancia con la primera serie de simulaciones.

Como ejemplo de resultados, en conformidad con el presentado en el estudio de la potencia, se muestra la tabla 4 donde $n = 250$ y las probabilidades de “rechazo” en la variable de control son de 0.05 y 0.54 con diferentes probabilidades p_{ij} para simular diferentes valores de ρ^2 entre Y y C . Se realizaron 200 remuestras de las tablas de frecuencias relativas en la aplicación del bootstrap.

Table 4.

“Verdadero” porcentaje de reducción de la varianza comparado con la media, sobre todas las simulaciones, de los porcentajes estimados de reducción de la varianza con respecto a los diversos estimadores de la varianza $\hat{\sigma}_R^2$, $\hat{\sigma}_{GS}^2$, $\hat{\sigma}_U^2$ y $\hat{\sigma}_B^2$. Las tablas se generaron con las probabilidades $p_{00} = 0.94, p_{01} = 0.0, p_{10} = 0.01, p_{11} = 0.05$ ($p_{1.} = 0.06, \rho^2 = 0.8246$); $p_{00} = 0.93, p_{01} = 0.01, p_{10} = 0.02, p_{11} = 0.04$ ($p_{1.} = 0.06, \rho^2 = 0.5110$); $p_{00} = 0.45, p_{01} = 0.04, p_{10} = 0.01, p_{11} = 0.50$ ($p_{1.} = 0.51, \rho^2 = 0.8126$); $p_{00} = 0.40, p_{01} = 0.09, p_{10} = 0.06, p_{11} = 0.45$ ($p_{1.} = 0.51, \rho^2 = 0.4911$), respectivamente.

potencia (v. respuesta)	$\rho^2 = 0.8$ (aprox.)					$\rho^2 = 0.5$ (aprox.)				
	Verd.	$\hat{\sigma}_R^2$	$\hat{\sigma}_{GS}^2$	$\hat{\sigma}_U^2$	$\hat{\sigma}_B^2$	Verd.	$\hat{\sigma}_R^2$	$\hat{\sigma}_{GS}^2$	$\hat{\sigma}_U^2$	$\hat{\sigma}_B^2$
0.06	82.82	82.41	82.40	82.40	82.39	49.75	51.32	50.14	50.15	50.41
0.51	81.14	81.37	81.30	81.25	81.26	48.79	49.20	49.00	48.97	48.96

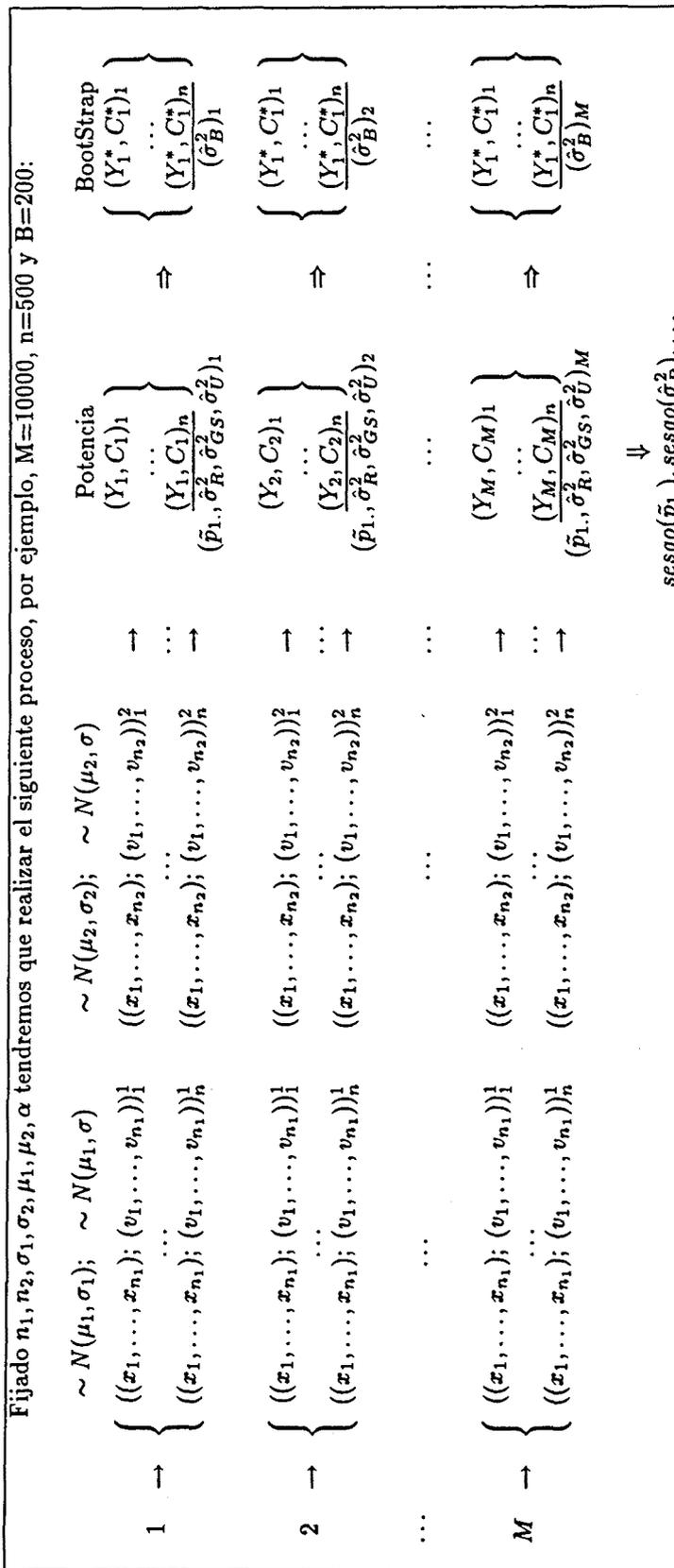
3.5.3 Conclusión

En resumen, todas las simulaciones confirman los resultados teóricos presentados al principio del capítulo. El estimador \tilde{p}_1 , usado en lugar de la habitual frecuencia relativa, obtiene una reducción de la varianza aceptable (incluso cuando la correlación entre Y y C no es muy alta, lo cual ocurre, por ejemplo, cuando el parámetro de escala diverge mucho en la primera serie de simulaciones) con un esfuerzo extra de cálculo y complejidad en el diseño bajo. Todos los estimadores de la varianza de \tilde{p}_1 dan valores razonables de la verdadera varianza aunque para $\hat{\sigma}_R^2$ se obtiene, sistemáticamente, estimaciones inferiores. Bajo las usuales condiciones de simulación en estadística (gran número, n , de réplicas de simulación, y sin tener que estimar valores extremos de p_{ij}) no hay diferencias apreciables entre

$\hat{\sigma}_{GS}^2$, $\hat{\sigma}_U^2$ y $\hat{\sigma}_B^2$. Así, se escoge el primer estimador por su simplicidad y menor coste (en comparación al estimador bootstrap, $\hat{\sigma}_B^2$).

Por tanto, en los próximos capítulos se utilizará $\hat{\sigma}_{GS}^2$ como un estimador adecuado de la varianza de $p_{1.}$, obtenido a partir de una aproximación de la esperanza de la inversa de una binomial positiva propuesta originalmente por Grab y Savage.

FIG.1: ESQUEMA DEL PROCESO REALIZADO EN EL CASO DEL ESTUDIO DE LA POTENCIA DE UN NUEVO TEST PARA EVALUAR LA BONDAD DE LA TÉCNICA PROPUESTA



Capítulo 4

REDUCCIÓN DE LA VARIANZA EN EL ESTUDIO DE UN NUEVO TEST PARA EL PROBLEMA DE BEHRENS-FISHER

En este capítulo deducimos un nuevo test para el problema de Behrens-Fisher usando la distancia de Rao. A continuación, estudiamos la potencia del nuevo test tal como se plantea en el capítulo anterior. Para ello se precisa realizar unos estudios de simulación que requieren una gran cantidad de tiempo de ordenador a pesar de usar técnicas de reducción de varianza y otras optimizaciones. Por lo tanto, nuestro objetivo es doble: estudiar, por simulación, las propiedades de un nuevo procedimiento estadístico, una nueva solución al problema de Behrens-Fisher, y proporcionar un caso de estudio ilustrativo de cómo usar la técnica de reducción de la varianza propuesta en el capítulo anterior y las ventajas asociadas con su uso.

4.1 El problema de Behrens-Fisher

4.1.1 Introducción

El problema de Behrens-Fisher es el de contrastar la igualdad de medias de dos poblaciones normales, sin hacer ninguna suposición acerca de las varianzas. De una manera más formal, el problema puede ser expuesto como sigue:

Sean X_1, X_2 dos variables aleatorias normales independientes con sus respectivas medias $\mu_i, i = 1, 2$, y con varianzas desconocidas y arbitrarias $\sigma_i^2, i = 1, 2$. Bajo estas premisas se considera el contraste de hipótesis:

$$\begin{aligned} H_0 : \mu_1 = \mu_2 = \mu & \quad \text{arbitrarios} \\ H_1 : \mu_1 \neq \mu_2 & \quad \sigma_1 > 0 \quad \sigma_2 > 0 \end{aligned}$$

Se supone que el contraste está basado en una muestra de $n_1 + n_2$ valores

$$\mathbf{x} = (x_{11}, \dots, x_{1n_1}; x_{21}, \dots, x_{2n_2}),$$

donde cada submuestra viene de n_i realizaciones *iid* de X_i , para $i = 1, 2$.

Por lo tanto, la diferencia con respecto al conocido test t de Student para la comparación de medias de dos poblaciones normales con varianzas desconocidas pero iguales es no hacer ninguna suposición respecto de las varianzas. Este detalle produce un cambio cualitativo muy importante, reflejado a un nivel más abstracto en la geometría de ciertas familias paramétricas de densidad. Así, tenemos que tanto el modelo paramétrico correspondiente al caso de varianzas iguales como el del caso de varianzas desiguales se pueden expresar como una familia exponencial de 3 y 4 parámetros, respectivamente. Pero, es en el submodelo paramétrico asociado a la hipótesis nula donde se produce la diferencia a remarcar. El espacio de distribuciones de probabilidad asociado a la hipótesis nula en el caso de varianzas iguales es un “subespacio plano” del espacio de la familia exponencial de dimensión 3, mientras que en el caso de varianzas desiguales es un “subespacio curvado” de la familia exponencial de dimensión 4. Como consecuencia, se produce una ruptura de las buenas propiedades para los problemas de inferencia que tiene la familia exponencial, que se acrecienta cuanto mayor es la curvatura. Así, por ejemplo, la varianza del estimador máximo verosímil excede del valor de la cota de Cramer-Rao en proporción aproximada al valor de su curvatura al cuadrado. Tampoco existe un estadístico suficiente con la misma dimensión que la familia exponencial curvada (Efron, 1975).

El test t de Student de comparación de medias se basa en obtener dos submuestras independientes x_{11}, \dots, x_{1n_1} y x_{21}, \dots, x_{2n_2} , la primera de una distribución normal $N(\mu_1, \sigma)$ y la segunda de una $N(\mu_2, \sigma)$. El modelo paramétrico subyacente para el conjunto de todas las observaciones si los tres parámetros son desconocidos es una familia exponencial de dimensión 3 que podemos expresar como:

$$\{f(x_1; \mu_1, \sigma) \cdot f(x_2; \mu_2, \sigma); \mu_1, \mu_2 \in \mathfrak{R} \ \sigma > 0\}$$

donde $f(x_i; \mu_i, \sigma)$ es la función de densidad de la variable aleatoria X_i que sigue una distribución $N(\mu_i, \sigma)$ para $i = 1, 2$. El submodelo determinado bajo la hipótesis nula, igualdad de medias, es una familia paramétrica bidimensional. Estos posibles valores de los parámetros varían sobre una variedad plana $\mathfrak{R} \times \mathfrak{R}^+$ de $\mathfrak{R}^2 \times \mathfrak{R}^+$. Por consiguiente, se forma una familia exponencial bidimensional dentro de la anterior.

En el caso del problema de Behrens-Fisher el modelo paramétrico para el conjunto de todas las observaciones es

$$\{f(x_1; \mu_1, \sigma_1) \cdot f(x_2; \mu_2, \sigma_2); \mu_1, \mu_2 \in \mathfrak{R} \ \sigma_1, \sigma_2 > 0\}$$

donde $f(x_i; \mu_i, \sigma_i)$ es la función de densidad de la variable aleatoria X_i que sigue una distribución $N(\mu_i, \sigma_i)$ para $i = 1, 2$. Por lo tanto, es una familia exponencial de dimensión 4. Bajo la hipótesis nula de igualdad de medias el submodelo es tridimensional. Los posibles valores de estos parámetros forman una variedad no plana $\mathfrak{R} \times \mathfrak{R}^{+2}$ de $\mathfrak{R}^2 \times \mathfrak{R}^{+2}$. Por consiguiente, se tiene una familia exponencial curvada.

Las familias exponenciales curvadas suelen aparecer bastante a menudo en la práctica y tiene gran importancia en la discusión de varios conceptos y métodos claves de inferencia estadística (ver Efron, 1975, 1978 y Efron and Hinkley, 1978).

Veamos ahora como los submodelos bajo la hipótesis nula de igualdad de medias constituyen una familia exponencial y una familia exponencial curvada, respectivamente.

En efecto, la función de densidad en el caso de igualdad de varianzas es una familia exponencial caracterizada por:

$$p(\mathbf{x}; \mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{1}{(\sigma_1 \sqrt{2\pi})^{n_1}} \frac{1}{(\sigma_2 \sqrt{2\pi})^{n_2}} \exp \left\{ -\frac{n_1 \mu_1^2}{2\sigma_1^2} - \frac{n_2 \mu_2^2}{2\sigma_2^2} \right\} \cdot \exp \left\{ -\frac{\sum_{i=1}^{n_1} x_{1i}^2}{2\sigma_1^2} - \frac{\sum_{i=1}^{n_2} x_{2i}^2}{2\sigma_2^2} + \frac{\mu_1 \sum_{i=1}^{n_1} x_{1i}}{\sigma_1^2} + \frac{\mu_2 \sum_{i=1}^{n_2} x_{2i}}{\sigma_2^2} \right\},$$

donde $\mu_1, \mu_2 \in \mathfrak{R}$ $\sigma_1, \sigma_2 \in \mathfrak{R}^+$ (4.1)

que tiene como vector de parámetros naturales y estadístico privilegiado a

$$(\theta_1, \theta_2, \theta_3) = \left(\frac{\mu_1}{\sigma^2}, \frac{\mu_2}{\sigma^2}, -\frac{1}{2\sigma^2} \right)$$

$$t(\mathbf{x}) = \left(\sum_{i=1}^{n_1} x_{1i}, \sum_{i=1}^{n_2} x_{2i}, \sum_{i=1}^{n_1} x_{1i}^2 + \sum_{i=1}^{n_2} x_{2i}^2 \right)$$

(4.2)

respectivamente siendo ambos vectores de dimensión 3. Si ahora expresamos la condición de igualdad de medias en función de los parámetros naturales:

$$\mu_1 = \mu_2 \iff \theta_1 = \theta_2$$

(4.3)

entonces, el submodelo bajo la hipótesis nula queda como una función lineal de estos. Por tanto, es un plano $(\mathfrak{R} \times \mathfrak{R}^+)$ de $\mathfrak{R}^2 \times \mathfrak{R}^+$.

En cambio, la función de densidad de las observaciones en el modelo para el problema de Behrens-Fisher también es una familia exponencial pero caracteri-

zada por

$$p(\mathbf{x}; \mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{1}{(\sigma_1 \sqrt{2\pi})^{n_1}} \frac{1}{(\sigma_2 \sqrt{2\pi})^{n_2}} \exp \left\{ -\frac{n_1 \mu_1^2}{2\sigma_1^2} - \frac{n_2 \mu_2^2}{2\sigma_2^2} \right\} \cdot \exp \left\{ -\frac{\sum_{i=1}^{n_1} x_{1i}^2}{2\sigma_1^2} - \frac{\sum_{i=1}^{n_2} x_{2i}^2}{2\sigma_2^2} + \frac{\mu_1 \sum_{i=1}^{n_1} x_{1i}}{\sigma_1^2} + \frac{\mu_2 \sum_{i=1}^{n_2} x_{2i}}{\sigma_2^2} \right\},$$

(4.4)

donde $\mu_1, \mu_2 \in \mathfrak{R}$ $\sigma_1, \sigma_2 \in \mathfrak{R}^+$

siendo

$$(\theta_1, \theta_2, \theta_3, \theta_4) = \left(\frac{\mu_1}{\sigma_1^2}, \frac{\mu_2}{\sigma_2^2}, -\frac{1}{2\sigma_1^2}, -\frac{1}{2\sigma_2^2} \right) \tag{4.5}$$

$$t(\mathbf{x}) = \left(\sum_{i=1}^{n_1} x_{1i}, \sum_{i=1}^{n_2} x_{2i}, \sum_{i=1}^{n_1} x_{1i}^2, \sum_{i=1}^{n_2} x_{2i}^2 \right)$$

el vector de parámetros naturales y el estadístico privilegiado respectivamente, ambos de dimensión 4. Expresando la condición de igualdad de medias en función de los parámetros naturales:

$$\mu_1 = \mu_2 \iff \theta_1 \theta_4 = \theta_2 \theta_3 \tag{4.6}$$

entonces, el submodelo bajo la hipótesis nula queda como una función no lineal de estos. Por tanto, el modelo bajo la hipótesis nula es una familia exponencial curvada.

Antes de continuar, se indicará la notación referente al problema de Behrens-Fisher que será utilizada durante todo este capítulo. Para cada submuestra, $(x_{i1}, \dots, x_{in_i})$, \bar{x}_1, \bar{x}_2 representa la media muestral: $\bar{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$; s_1^2, s_2^2 corresponde a la varianza muestral con denominador n_i : $s_i^2 = n_i^{-1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$; mientras \hat{s}_1^2, \hat{s}_2^2 designa a las varianzas con denominador $n_i - 1$: $\hat{s}_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$ y $s_i = \sqrt{s_i^2}$, $\hat{s}_i = \sqrt{\hat{s}_i^2}$ las correspondientes desviaciones estándar muestrales.

4.1.2 Soluciones prácticas

El problema de Behrens-Fisher es una cuestión controvertida sin una solución óptima conocida y donde diferentes escuelas de pensamiento estadístico dan diferentes soluciones. Así, como breve resumen, se indicarán algunas de las soluciones prácticas más habituales utilizadas. Para una revisión más exhaustiva ver Scheffé (1970) y Lee and Gurland (1975).

Para tamaños muestrales mayores que 10, las diferencias entre las diversas soluciones propuestas son generalmente mucho menores que sus diferencias con el test t de Student para la igualdad de medias. Por tanto, el uso de cualquiera de ellos es mejor que el uso del test t de Student, a menos que la suposición de igualdad de varianzas sea garantizada.

De todas estas soluciones, las más comunmente utilizadas son:

- El test de Cochran y Cox.
- El test de Welch o el test de Welch-Aspin.

La primera puede ser considerada, con el test de McCullough-Barnerjee, como aproximaciones a la solución de Behrens-Fisher, mientras que la mayoría de otros test discutidos por Lee and Gurland (1975) pueden ser considerados como aproximaciones test de Welch-Aspin.

En todos los casos, el estadístico de contraste es:

$$T' = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(\hat{s}_1^2/n_1) + (\hat{s}_2^2/n_2)}} \quad (4.7)$$

El problema es que el estadístico T' no se distribuye, necesariamente, según el modelo de probabilidad t de Student con $n_1 + n_2 - 2$ grados de libertad. Por tanto, todo el esfuerzo se centra en conocer de forma aproximada la distribución muestral de T' .

Cochran y Cox (1950) diseñaron un método de aproximación a los puntos críticos de la distribución muestral T' . El método consiste en obtener estos puntos mediante:

$$t'_p = \frac{t(p, n_1 - 1)(\hat{s}_1^2/n_1) + t(p, n_2 - 1)(\hat{s}_2^2/n_2)}{(\hat{s}_1^2/n_1) + (\hat{s}_2^2/n_2)} \quad (4.8)$$

donde $t(p, \nu)$ es el p -percentil superior de una distribución t de Student con ν grados de libertad. El proceso de decisión es rechazar H_0 si $|T'| \geq t'_{\alpha/2}$ siendo α el nivel de significación del test.

La principal ventaja de esta solución es su simplicidad. Tal vez es por esta razón lo que hace al test más extensamente utilizado en la práctica, incluso sabiendo que la extensión del test puede diferir substancialmente del nivel de significación nominal, como fue mostrado por Cochran (1964). Afortunadamente,

en la mayoría de ocasiones, la extensión del test está por debajo del nivel de significación nominal, es decir, se trata de un test conservador.

Welch (1938) propuso una aproximación alternativa. En esta aproximación T' se concibe como una variable aleatoria distribuida según la t de Student, pero con un número desconocido de grados de libertad. La solución pasa por determinar los grados de libertad (g') que corresponden a la distribución de T' mediante la ecuación:

$$g' = \frac{\left(\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}\right)^2}{\frac{(\hat{s}_1^2/n_1)^2}{n_1 - 1} + \frac{(\hat{s}_2^2/n_2)^2}{n_2 - 1}} \quad (4.9)$$

El resultado obtenido para g' se redondea al entero más próximo¹. Se obtienen así unos grados de libertad comprendidos entre un mínimo y un máximo conocidos: el mínimo es el valor más pequeño de $n_1 - 1$ y $n_2 - 1$; el máximo es $n_1 + n_2 - 2$. El criterio de decisión es rechazar H_0 si $|T'| \geq t(\alpha/2, g')$ donde $t(\alpha/2, g')$ es el $(\alpha/2)$ -percentil superior de una distribución t de Student con g' grados de libertad.

El test de Welch-Aspin se basa en cálculos asintóticos sobre T' . Welch (1947) obtiene los puntos porcentuales de T' como una serie de potencias en $1/f_i = 1/(n_i - 1)$ para $i = 1, 2$. La fracción P ($0 < P < 1$) de T' es explícitamente desarrollada hasta orden 2 como

$$\alpha \left[1 + \frac{(1 + \alpha)^2 \sum_{i=1}^2 (\hat{s}_i^4/n_i^2 f_i)}{4 (\sum_{i=1}^2 \hat{s}_i^2/n_i)^2} + \frac{(3 + 5\alpha^2 + \alpha^4) \sum_{i=1}^2 (\hat{s}_i^6/n_i^3 f_i^2)}{3 (\sum_{i=1}^2 \hat{s}_i^2/n_i)^3} - \frac{(15 + 32\alpha^2 + 9\alpha^4) \sum_{i=1}^2 (\hat{s}_i^4/n_i^2 f_i)^2}{32 (\sum_{i=1}^2 \hat{s}_i^2/n_i)^4} \right], \quad (4.10)$$

donde $\alpha = \Phi^{-1}(P)$ y Φ es la función de distribución de la normal estándar. Al año siguiente, Aspin (1948) extendió estos resultados hasta el orden 4. Por

¹El propio Welch (1947) sugirió posteriormente que hacer:

$$g' = \left[\frac{\left(\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}\right)^2}{\frac{(\hat{s}_1^2/n_1)^2}{n_1 - 1} + \frac{(\hat{s}_2^2/n_2)^2}{n_2 - 1}} \right] - 2$$

puede ofrecer una solución más exacta para g' . No obstante, la diferencia entre ambas soluciones es, en la mayor parte de los casos, insignificante.

tanto, para poder utilizar este test es necesario tabular los puntos críticos. Así, se pueden encontrar en la tabla 11 de *Biometrika Tables* (1976) los valores tabulados para cuatro niveles de probabilidad ($\alpha/2 = 0.05, 0.025, 0.01$ y 0.005) para la cola superior. El criterio de decisión es rechazar H_0 si $|T'| \geq V(c; f_1, f_2, \alpha/2)$ donde $c = \frac{\hat{s}_1^2/n_1}{\hat{s}_1^2/n_1 + \hat{s}_2^2/n_2}$ y el valor $V(c; f_1, f_2, \alpha/2)$ se obtiene en las tablas anteriores.

Las diferencias entre los dos últimos test son mínimas siendo más común utilizar el test de Welch ya que no requiere de unas tablas específicas y en sólo un caso hace falta determinar los grados de libertad gl' . Dado que los valores de la distribución t de Student van disminuyendo a medida que van aumentando los grados de libertad, antes de calcular gl' podemos evaluar T' utilizando el gl' mínimo (es decir, el menor de $n_1 - 1$ y $n_2 - 1$); si rechazamos $H_0 : \mu_1 = \mu_2$, también la rechazaremos con el valor proporcionado por (4.9) para gl' ; si no rechazamos H_0 , podemos evaluar T' con el gl' máximo ($n_1 + n_2 - 2$); si seguimos sin rechazar H_0 , tampoco la rechazaremos calculando el valor exacto de gl' . De modo que el único caso en el que necesitaremos hacer uso de (4.9) para calcular el valor exacto de gl' será aquel en el que manteniendo H_0 con el gl' mínimo, la rechazamos con el gl' máximo.

A diferencia del test de Cochran y Cox, tanto el test de Welch como el test de Wech-Aspin, la probabilidad de error es muy cercana al valor nominal en todo el espacio paramétrico, aunque Fisher (1956) los critica por mostrar en un subconjunto relevante la existencia de un sesgo negativo en el sentido de Buehler (1959). Así, independientemente de \hat{s}_1/\hat{s}_2 , $U = (n_1 - 1)(\hat{s}_1^2/\sigma_1^2) + (n_2 - 1)(\hat{s}_2^2/\sigma_2^2)$ sigue una distribución de ji-cuadrado con $n_1 + n_2 - 2$ grados de libertad y $(\bar{x}_1 - \mu_1) + (\bar{x}_2 - \mu_2)$ está distribuido normalmente con media cero y varianza $\sigma_1/n_1 + \sigma_2/n_2$. Por consiguiente,

$$T' \sqrt{\frac{\hat{s}_1^2/n_1 + \hat{s}_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} \frac{n_1 + n_2 - 2}{U}} \quad (4.11)$$

tiene una distribución t de Student con $n_1 + n_2 - 2$ grados de libertad dados $\hat{s}_1/\hat{s}_2, \sigma_1^2, \sigma_2^2$. Cuando $\hat{s}_1/\hat{s}_2 = 1, n_1 = n_2$ y $\sigma_1^2/\sigma_2^2 = w$, la expresión (4.11) se simplifica a

$$T' \sqrt{\frac{4}{2 + 1/w + w}}$$

Ahora $1/w + w \geq 2$, entonces

$$\sqrt{\frac{4}{2 + 1/w + w}} \leq 1$$

y por tanto

$$Pr(|T'| > a | \hat{s}_1 / \hat{s}_2 = 1) \geq Pr(|t_{(n_1+n_2-2)}| > a)$$

para todo $a > 0$, donde $t_{(n_1+n_2-2)}$ indica la distribución t de Student con n_1+n_2-2 grados de libertad. En particular, tomando $n_1 = n_2 = 7$ y $a = 1.74$ (de la tabla 11 de *Biometrika Tables*),

$$Pr(|T'| > 1.74 | \hat{s}_1 / \hat{s}_2 = 1) \geq Pr(|t_{12}| > 1.74) > 0.1.$$

Entonces el conjunto donde $\hat{s}_1 / \hat{s}_2 = 1$ es un subconjunto relevante donde el intervalo de confianza basado en el test de Welch-Aspin cubre el verdadero valor de $\mu_1 - \mu_2 = 0$ menos a menudo que el nivel de confianza sugerido.

4.2 Obtención del nuevo test

4.2.1 Planteamiento

En el artículo de C.R. Rao de 1945, entre otros importantes resultados, introduce una distancia entre poblaciones y la base de su uso en inferencia. Esta distancia esta basada en conceptos de geometría diferencial, por tanto es apropiado llamarla distancia "geodésica", como se esboza en el apéndice distancia de Rao. Así, el proceso de decisión del nuevo test se basa en la distancia geodésica² para resolver el problema de Behrens-Fisher.

Asumimos que los datos muestrales han sido resumidos por $\bar{x}_1, \bar{x}_2, s_1, s_2$, esto es, por las estimaciones máximo verosímiles de las medias y desviaciones estándar poblacionales, respectivamente, calculadas a partir de los datos muestrales x . Como se muestra en el apéndice distancia de Rao, bajo la suposición de normalidad, la distancia de Rao entre el punto muestral (submuestral) (\bar{x}_i, s_i) y un punto paramétrico dado (μ_i, σ_i) es:

$$\lambda_i = \lambda[(\bar{x}_i, s_i); (\mu_i, \sigma_i)] = 2\sqrt{2} \tanh^{-1} \Delta_i = \sqrt{2} \left| \log \frac{1 + \Delta_i}{1 - \Delta_i} \right|, \quad (4.12)$$

²Comunmente llamada distancia de Rao en honor a su introductor en el ámbito estadístico

con:

$$\Delta_i = \left\{ \frac{(\bar{x}_i - \mu_i)^2 + 2(s_i - \sigma_i)^2}{(\bar{x}_i - \mu_i)^2 + 2(s_i + \sigma_i)^2} \right\}^{1/2}. \tag{4.13}$$

Sobre esta base, una medida natural de discrepancia entre una muestra compuesta de dos submuestras *independientes*, caracterizadas por los puntos (\bar{x}_1, s_1) y (\bar{x}_2, s_2) , y un punto paramétrico compuesto por (μ_1, σ_1) y (μ_2, σ_2) es la *distancia de Rao al cuadrado ponderada* por los tamaños muestrales n_1 y n_2 (Oller, comunicación personal).

$$d^2[(\bar{x}_1, \bar{x}_2, s_1, s_2), (\mu_1, \mu_2, \sigma_1, \sigma_2)] = 2 \sum_{i=1}^2 n_i \left\{ \log \frac{1 + \Delta_i}{1 - \Delta_i} \right\}^2. \tag{4.14}$$

Definimos:

$$D^2 = \min_{\mu, \sigma_1, \sigma_2} \{ d^2[(\bar{x}_1, \bar{x}_2, s_1, s_2), (\mu, \mu, \sigma_1, \sigma_2)] \}, \tag{4.15}$$

intuitivamente, (4.15) representa el grado de discrepancia, en términos de distancia geodésica, entre los datos muestrales (resumidos por $\bar{x}_1, \bar{x}_2, s_1, s_2$) y el conjunto de todos los puntos paramétricos $(\mu, \mu, \sigma_1, \sigma_2)$ que satisfacen la hipótesis nula de igualdad de medias. Valores grandes de D^2 proporcionan evidencias contra esta hipótesis. Por tanto, el siguiente procedimiento de rechazo: “rechazar H_0 si $D^2 \geq c_\alpha$ ”, donde c_α debe satisfacer

$$P\{D^2 \geq c_\alpha | H_0\} = \alpha, \tag{4.16}$$

producirá un test “geodésico” (en el sentido de Burbea y Oller, (Burbea y Oller, 1982)) con un nivel de significación α para la hipótesis nula de igualdad de medias.

El test estadístico D^2 no tiene una forma cerrada, quedando en función de un sistema de ecuaciones (ver la siguiente sección para más detalles)

$$\begin{cases} \sigma_1^2 = s_1^2 + \frac{(\bar{x}_1 - \mu)^2}{2} \\ \sigma_2^2 = s_2^2 + \frac{(\bar{x}_2 - \mu)^2}{2} \\ \frac{\sigma_1}{\sigma_2} = \frac{n_1 \cosh^{-1}(\sigma_1/s_1)}{n_2 \cosh^{-1}(\sigma_2/s_2)} \end{cases} \tag{4.17}$$

que debe ser resuelto, numéricamente, para obtener los valores $\hat{\mu}, \hat{\sigma}_1$ y $\hat{\sigma}_2$ que minimizan d^2 en (4.15), requerido para calcular $D^2 = d^2[(\bar{x}_1, \bar{x}_2, s_1, s_2); (\hat{\mu}, \hat{\mu}, \hat{\sigma}_1, \hat{\sigma}_2)]$. Esto hace que sea muy difícil de resolver el problema distribucional de calcular c_α . Por consiguiente, optamos por aproximar su valor mediante bootstrap paramétrico (ver el apéndice de bootstrap paramétrico)

4.2.2 Deducción del test

Para obtener la distancia geodésica mínima al cuadrado del punto muestral $(\bar{x}_1, \bar{x}_2, s_1, s_2)$ a la región paramétrica definida por la hipótesis nula, D^2 , debemos resolver las siguientes ecuaciones:

$$\begin{cases} \frac{\partial d^2[(\bar{x}_1, \bar{x}_2, s_1, s_2), (\mu, \mu, \sigma_1, \sigma_2)]}{\partial \sigma_i} = 0 & i = 1, 2 \\ \frac{\partial d^2[(\bar{x}_1, \bar{x}_2, s_1, s_2), (\mu, \mu, \sigma_1, \sigma_2)]}{\partial \mu} = 0. \end{cases} \quad (4.18)$$

Como

$$\frac{\partial d^2}{\partial \sigma_i} = 8n_i \log \left(\frac{1 + \Delta_i}{1 - \Delta_i} \right) \frac{1}{1 - \Delta_i^2} \frac{\partial \Delta_i}{\partial \sigma_i} = 0 \quad i = 1, 2,$$

las dos primeras ecuaciones de (4.18) son equivalentes a $\frac{\partial \Delta_i}{\partial \sigma_i} = 0$ ya que sólo la derivada parcial de Δ_i con respecto a σ_i puede eliminarse.

Continuando con ellas, obtenemos:

$$\frac{\partial \Delta_i}{\partial \sigma_i} = -(\bar{x}_1 - \mu)^2 - 2(s_i - \sigma_i^2) = 0 \quad i = 1, 2, \quad (4.19)$$

donde

$$\Delta_i = \left\{ \frac{(\bar{x}_i - \mu)^2 + 2(s_i - \sigma_i)^2}{(\bar{x}_i - \mu)^2 + 2(s_i + \sigma_i)^2} \right\}^{1/2} = \left\{ 1 - \frac{8s_i\sigma_i}{(\bar{x}_i - \mu)^2 + 2(s_i + \sigma_i)^2} \right\}^{1/2} \quad i = 1, 2.$$

Entonces (4.19) da las dos primeras ecuaciones de (4.17):

$$\sigma_i^2 = s_i^2 + \frac{(\bar{x}_i - \mu)^2}{2} \quad i = 1, 2.$$

y además podemos observar la siguiente desigualdad:

$$\sigma_i^2 \geq s_i^2 \quad i = 1, 2. \quad (4.20)$$

Desarrollando la derivada parcial de la tercera ecuación de (4.18):

$$\frac{\partial d^2}{\partial \mu} = 2 \sum_{i=1}^2 n_i \log \left(\frac{1 + \Delta_i}{1 - \Delta_i} \right) \frac{1}{1 - \Delta_i^2} \frac{\partial \Delta_i}{\partial \mu} = 0, \quad (4.21)$$

donde

$$\frac{\partial \Delta_i}{\partial \mu} = \frac{1}{\Delta_i} (1 - \Delta_i^2) \frac{-(\bar{x}_i - \mu)}{(\bar{x}_i - \mu)^2 + 2(s_i + \sigma_i)^2},$$

obtenemos como resultado,

$$\begin{aligned} \frac{\partial d^2}{\partial \mu} &= \sum_{i=1}^2 n_i \log \left(\frac{1 + \Delta_i}{1 - \Delta_i} \right) \frac{1}{\Delta_i} \frac{-(\bar{x}_i - \mu)}{(\bar{x}_i - \mu)^2 + 2(s_i + \sigma_i)^2} = 0 \\ &= \sum_{i=1}^2 n_i \log \left(\frac{1 + \Delta_i}{1 - \Delta_i} \right) \frac{-(\bar{x}_i - \mu)}{\sqrt{(\bar{x}_i - \mu)^4 + 4[(\bar{x}_i - \mu)^2(s_i^2 + \sigma_i^2) + (s_i^2 - \sigma_i^2)^2]}} = 0. \end{aligned} \tag{4.22}$$

Sustituyendo

$$\sigma_i^2 = s_i^2 + \frac{(\bar{x}_i - \mu)^2}{2} \quad i = 1, 2,$$

en (4.22) se producen los cambios siguientes:

$$\Delta_i = \left(\frac{1 - (s_i/\sigma_i)}{1 + (s_i/\sigma_i)} \right)^{1/2}$$

$$\log \left(\frac{1 + \Delta_i}{1 - \Delta_i} \right) = \log \left(\frac{\sigma_i}{s_i} + \sqrt{\frac{\sigma_i^2}{s_i^2} - 1} \right)$$

$$\frac{-(\bar{x}_i - \mu)}{\sqrt{(\bar{x}_i - \mu)^4 + 4[(\bar{x}_i - \mu)^2(s_i^2 + \sigma_i^2) + (s_i^2 - \sigma_i^2)^2]}} = \frac{-sgn(\bar{x}_i - \mu)}{2\sqrt{2}\sigma_i},$$

así que, la igualdad (4.21) finalmente se reduce a:

$$\sum_{i=1}^2 n_i \log \left(\frac{\sigma_i}{s_i} + \sqrt{\frac{\sigma_i^2}{s_i^2} - 1} \right) \frac{-sgn(\bar{x}_i - \mu)}{2\sqrt{2}\sigma_i} = 0 \tag{4.23}$$

Por consiguiente, ambos sumandos deben tener igual valor pero con signos opuestos o ser iguales a cero (solución trivial). Además, si consideramos la desigualdad (4.20) obtenemos una nueva desigualdad:

$$\min(\bar{x}_1, \bar{x}_2) \leq \mu \leq \max(\bar{x}_1, \bar{x}_2) \tag{4.24}$$

Aplicando la anterior desigualdad (4.24) y recordando que

$$\cosh^{-1}(x) = \log(x + \sqrt{x^2 - 1})$$

la ecuación (4.23) podemos expresarla como

$$n_1 \frac{1}{\sigma_1} \cosh^{-1}(\sigma_1/s_1) = n_2 \frac{1}{\sigma_2} \cosh^{-1}(\sigma_2/s_2)$$

y por tanto, obtener la tercera ecuación de (4.17):

$$\frac{\sigma_1}{\sigma_2} = \frac{n_1 \cosh^{-1}(\sigma_1/s_1)}{n_2 \cosh^{-1}(\sigma_2/s_2)}.$$

4.3 Estudio de simulación

Es obvio que se requiere de un estudio de Monte Carlo del test, que por poco escueto que sea necesitará de un gran número de réplicas de simulación, por consiguiente, de una gran cantidad de tiempo de ordenador. Una forma adecuada de resolver este inconveniente es usar técnicas de reducción de la varianza.

El estudio que se plantea es, de nuevo, sobre la potencia³ del nuevo test utilizando la técnica de reducción de la varianza desarrollada en el capítulo anterior y las optimizaciones de los algoritmos adecuadas para la programación vectorial que serán descritas más adelante.

4.3.1 Procedimiento

Se han hecho dos estudios análogos de simulación.

En el primero estudiamos la potencia del nuevo test basado en la distancia geodésica al cuadrado, D^2 , utilizando el test t de Student de comparación de medias (para varianzas iguales pero desconocidas) como variable de control C para aplicar la técnica de reducción de la varianza (ver Fig. 1). Esta es una elección razonable ya que la curva de potencia del test t de Student es bien conocida e, intuitivamente, ambos tests deben proporcionar resultados similares tanto mejor cuanto más cercano a uno sea el cociente de varianzas.

El segundo tipo de simulación también tiene como finalidad estudiar la potencia, en este caso, del test de Cochran y Cox (para más detalles ver la sección Soluciones prácticas). Este test es uno de los que habitualmente se utilizan para resolver el problema de Behrens-Fisher. El proceso de simulación (ver Fig. 2) que se seguirá es el mismo que en el primer estudio pero sustituyendo el nuevo test geodésico por el de Cochran y Cox.

De esta manera los resultados obtenidos nos indicarán la bondad de la potencia del nuevo test con respecto a uno de los más ampliamente utilizados, además sabremos si el nuevo test cumple la probabilidad de error de tipo I nominal y podremos observar, en ambas simulaciones, si la técnica de reducción de la

³Bajo este término también incluiremos el concepto de extensión del test, es decir, la probabilidad de rechazar la hipótesis nula sabiendo que es cierta, no solo el concepto puro de potencia de un test, probabilidad de rechazar la hipótesis nula sabiendo que es falsa.

varianza funciona bien.

En ambos casos el algoritmo de la simulación se divide en tres módulos (ver Fig 1 y 2):

1. Entrada de parámetros.
2. Obtención de los dos vectores de resultados (Y_i, C_i) $i = 1, \dots, n$.
3. Aplicación de la técnica de reducción de la varianza.

El segundo apartado es el fundamental para el estudio de la potencia de los tests. Se divide en:

- Generación de las muestras.
- Calcular el test de estudio y el test t de Student.
- Obtención de Y y C .

Por tanto, es en el ítem "Calcular el test de estudio" donde aparece la diferencia entre una simulación y otra. Así, en el estudio de la potencia del test geodésico se debe de realizar dos pasos:

- Calcular D^2 .
- Bootstrap paramétrico.

Con el primero obtenemos el valor del nuevo estadístico pero, como se desconoce la distribución muestral de c_α que debe satisfacer

$$P\{D^2 \geq c_\alpha | H_0\} = \alpha,$$

se realiza una estimación con la técnica de bootstrap paramétrico, mientras que en el estudio de la potencia del test de Cochran y Cox es mucho más sencillo.

4.3.1.1 Entrada de parámetros

Para aplicar el algoritmo se necesitan introducir 8 valores:

n_1 = Tamaño muestral para la población 1.

n_2 = Tamaño muestral para la población 2.

σ_1 = Desviación típica de la población 1.

σ_2 = Desviación típica de la población 2.

β_c = Potencia de la variable de control (C). En este caso es el test t de Student de comparación de medias para varianzas desconocidas pero iguales.

α = Nivel de significación de los tests.

n = Número de simulaciones en el proceso de Monte Carlo.

B = Número de remuestras en el proceso de Bootstrap.

Los valores de las medias, μ_1 y μ_2 , se obtienen empleando la siguiente regla: El valor de la media de la primera población, μ_1 , es siempre 0 con lo cual la diferencia de medias entre las dos poblaciones afecta sólo a la segunda media, μ_2 , que se obtiene utilizando la potencia teórica del test de control, t de Student (Ver Apéndice 1)

4.3.1.2 Obtención de los dos vectores de resultados

La correlación requerida entre Y y C se induce usando variables aleatorias comunes para evaluar cada uno de los tests, como se describe a continuación.

Se generan n_1 valores normales estándar iid , $\mathbf{z}_1 = (z_{11}, \dots, z_{1n_1})$, independientes de otra serie de n_2 valores normales estándar iid , $\mathbf{z}_2 = (z_{21}, \dots, z_{2n_2})$. Utilizando la transformación $x_{ij} = \sigma_i z_{ij} + \mu_i$, $i = 1, 2$, se obtiene los $n_1 + n_2$ valores

$$\mathbf{x} = (x_{11}, \dots, x_{1n_1}; x_{21}, \dots, x_{2n_2})$$

para diferentes configuraciones de medias y varianzas. De igual manera, se aplica la misma transformación para obtener también $n_1 + n_2$ valores

$$\mathbf{v} = (v_{11}, \dots, v_{1n_1}; v_{21}, \dots, v_{2n_2})$$

pero con la diferencia de usar como varianza,

$$\sigma^2 = \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n_1 + n_2}, \quad (4.25)$$

un valor común a ambas submuestras e intermedio, entre σ_1^2 y σ_2^2 . Se debe tener en cuenta que si la generación del vector de valores \mathbf{v} no fuera con una varianza común, (4.25), el test t de Student no serviría como una variable de control al ser desconocida su potencia (es decir, la esperanza de la variable de control) ya que se violaría una de sus condiciones: la de igualdad de varianzas.

Una vez obtenido la muestra \mathbf{v} ya se puede calcular el test t de Student $t(\mathbf{v})$ y evaluar el valor de la variable de control C :

$$\begin{aligned} |t(\mathbf{v})| &\geq t_{(n_1+n_2-2, \alpha/2)} \longrightarrow C = 1 \\ &< t_{(n_1+n_2-2, \alpha/2)} \longrightarrow C = 0 \end{aligned}$$

donde $t_{(n_1+n_2-2, \alpha/2)}$ es el percentil $(\alpha/2)$ superior de una distribución t de Student con $n_1 + n_2 - 2$ grados de libertad.

En el caso del estudio de la potencia del test geodésico se debe de calcular $D^2(\mathbf{x})$ resolviendo el sistema de ecuaciones (4.17) y sustituyendo los valores resultantes $(\hat{\mu}, \hat{\mu}, \hat{\sigma}_1, \hat{\sigma}_2)$ en (4.14). Pero aún es necesario estimar la distribución de muestreo de D^2 bajo la hipótesis nula, para lo cual se emplea la técnica de bootstrap paramétrico (ver apéndice Bootstrap paramétrico).

Si estamos en el test de Cochran y Cox se obtiene el valor experimental T' aplicando (4.7) y el punto de crítico $t'_{\alpha/2}$ (4.8). Con todo ello, podemos emplear la siguiente regla de decisión para dar el valor adecuado a Y

$$\begin{aligned} |t_{cc}(\mathbf{x})| &\geq t'_{\alpha/2} \longrightarrow Y = 1 \\ &< t'_{\alpha/2} \longrightarrow Y = 0 \end{aligned}$$

Este proceso se repite n veces, generándose $2 \times n \times (n_1 + n_2)$ muestras y $n \times B \times (n_1 + n_2)$ remuestras.

4.3.1.3 Aplicación de la técnica de reducción de la varianza

Con los valores (Y_i, C_i) $i = 1, \dots, n$ obtenidos en el paso anterior se calcula la potencia del nuevo test (\tilde{p}_1), la estimación de la varianza de \tilde{p}_1 , de Grab & Savage, el tanto por ciento de reducción de varianza, etc. tal como se indicó en el capítulo 3.

4.3.2 Implementación

Todas las simulaciones fueron realizadas en un IBM 3090 600J corriendo bajo VM/XA y un IBM POWERparallel SP2 corriendo bajo Unix/AIX, ambos en el CESCA (Centre de Supercomputació de Catalunya).

Los programas que ejecutan estas simulaciones en el 3090 fueron escritos en VS FORTRAN v.2.0 r.3.0 vectorial usando el paquete de subrutinas ESSL v.1.0 r.4.0 y el paquete NAG (Mark 15 version). Estos programas fueron más tarde adaptados al nuevo ordenador SP2 y recompilados con FORTRAN v.3.2 usando las ESSL v.2.0 r.2.0 y el paquete NAG (Mark 16 version). Los resultados en ambos ordenadores fueron completamente coincidentes.

Para generar los valores normales se usó, de nuevo, la subrutina DNRAND de ESSL.

Para resolver el sistema de ecuaciones (4.17) se sustituyó la primera y segunda ecuación en la tercera con lo cual queda una función dependiente de μ igualada a cero

$$\sum_{i=1}^2 n_i \log \left(\frac{(\bar{x}_i - \mu)}{s_i \sqrt{2}} + \sqrt{\frac{(\bar{x}_i - \mu)^2}{2s_i^2} + 1} \right) \frac{1}{\sqrt{2s_i^2 + (\bar{x}_i - \mu)^2}} = 0. \quad (4.26)$$

Se resolvió numéricamente con la subrutina C05AZF del paquete NAG que busca los ceros de una función continua en un intervalo dado. Esta se basa en una combinación de métodos de interpolación lineal, extrapolación lineal y bisección (Dahlquist and Bjorck, 1974 y Bus and Dekker, 1975). Una vez conocida μ se sustituye en la primera y segunda ecuación para encontrar los valores de σ_1 y σ_2 .

4.3.3 Algoritmos vectoriales

Todos los programas que se han empleado en esta memoria se han realizado de la manera más adecuada para aprovechar las características de la vectorización (Capítulo 1).

En general, si se piensa de manera escalar y secuencial, casi todos los programas presentados, y en particular el del estudio de la potencia de un nuevo test para el problema de Behrens-Fisher, estarían formados por un bucle externo, con muchas instrucciones, donde en cada iteración se obtiene el par de valores

$(Y(i), C(i))$ hasta llegar a los n pares (número total de réplicas para una simulación dada – parámetros μ_1, μ_2 , etc. fijos). Una vez obtenidos todos los pares, se calculan diferentes estimaciones ($\tilde{p}_1, \hat{\sigma}^2, \%VR$). Esquemáticamente se puede expresar como (ver fig. 1):

Esquema 1

- Lectura de los parámetros iniciales.
- Repite desde I= 1 hasta n
 - Generación de las dos parejas de muestras normales:
 $((x_{11}, \dots, x_{1n_1}), (x_{21}, \dots, x_{2n_2}))$ y $((v_{11}, \dots, v_{1n_1}), (v_{21}, \dots, v_{2n_2}))$.
 - Evaluar la distancia geodésica en el primer par de vectores, $D^2(\mathbf{x})$.
 - Evaluar el t de Student en el segundo par de vectores, $t(\mathbf{v})$.
 - Asignar un 0 o 1, dependiendo de aceptación o rechazo de $t(\mathbf{v})$, a la variable de control, C .
 - Obtener una estimación del p-valor $D^2(\mathbf{x})$ mediante bootstrap paramétrico.
 - Asignar un 0 o 1, dependiendo de aceptación o rechazo de $D^2(\mathbf{x})$, a la variable de respuesta, Y .
- Fin Repite
- Con todos los n pares de valores de (Y_i, C_i) se aplica la técnica de reducción de la varianza y se obtiene estimaciones de $p_1, \sigma^2, \%VR, \dots$

Al pensar en forma vectorial esta manera de programar no es la más adecuada. Se tiene que intentar que los bucles sean sencillos y con muchas iteraciones, por tanto la idea escalar de un bucle exterior se elimina y es en cada paso (generación de normales, evaluación de los tests, etc) donde se realiza el bucle (es una internalización del bucle), de esta forma se obtienen todos los valores a la vez, siendo ubicados en un vector enorme que se utilizará en el siguiente paso. Por tanto, con una sola iteración se ejecuta todo el programa. Por ejemplo, el proceso de obtención de los n pares de vectores muestrales, $((x_{11}, \dots, x_{1n_1}), (v_{11}, \dots, v_{1n_1}))$, que siguen cada uno de ellos, respectivamente, la distribución normal $N(\mu_1, \sigma_1)$ y $N(\mu_1, \sigma)$

se realizan todos a la vez. Se necesita un vector de tamaño $n \times (n_1 + n_1)$ (en la simulación especificada puede llegar a tener un tamaño de $5000 \times (121 + 121)$) e igualmente para los datos de la segunda "población", que requieren un vector de $n \times (n_2 + n_2)$ elementos. Como se puede observar una limitación importante es el espacio disponible en memoria, ya que, fácilmente, los tamaños de los vectores se disparan. Una circunstancia añadida es que en FORTRAN no se puede dimensionar los vectores o matrices en tiempo de ejecución y, para curarse en salud, se pone los tamaños más grandes, acrecentando el problema anterior. Por tanto, las soluciones prácticas pasan por aprovechar los grandes vectores en el mayor número de procesos, utilizar COMMONs y nunca dejar la memoria disponible a cero.

En el proceso de simulación del estudio de la potencia del test propuesto por Burbea y Oller (1982) que sirvió para evaluar la técnica de reducción de la varianza presentada en el capítulo anterior (ver Fig 1 del capítulo 3), se acrecienta la necesidad de utilizar técnicas de vectorización para optimizar el proceso ya que requiere de un gran número de iteraciones al tenerse que repetir el esquema 1, a su vez, un número elevado de veces, M (10000) de veces, para poder obtener los "verdaderos" valores de p_1 , σ^2 , $\%VR$, ... y así, poder estudiar las propiedades de sus estimadores. Por tanto, igual que pasaba antes, la forma de proceder vectorial implica la eliminación de este bucle y, como consecuencia realizar todo el proceso en un solo paso. Si se considera el ejemplo anterior, de un vector de tamaño $n \times (n_1 + n_1)$ que contiene apilados los n pares de vectores muestrales, $((x_{11}, \dots, x_{1n_1}), (v_{11}, \dots, v_{1n_1}))$, ahora se convierte en un vector aún más grande de tamaño $M \times n \times (n_1 + n_1)$ que puede llegar a ser de $10000 \times 1000 \times (121 + 121)$.

Uno de los muchos problemas, más específicos, que se tuvieron que solucionar era cómo vectorizar el cálculo de los n promedios de n muestras de cualquier tamaño. Si se piensa escalarmente, una posible solución sería

```

DO 340 J=1,N
      DO 350 I=1,N1
            XM(J) = XM(J) + AUX((J-1)*N+I)
350      CONTINUE
            XM(J) = XM(J) / N
340 CONTINUE

```

El vector AUX tiene todos los vectores muestrales (n) de tamaño n_1 (por ejemplo 121) de manera secuencial. Por tanto, el tamaño del vector AUX es $n \times n_1$. En $XM(J)$, puesto a cero inicialmente, se van sumando los distintos valores de una muestra, cuando se acaba se calcula su promedio. El principal problema de este algoritmo es que existe dependencia entre una iteración y la siguiente.

$$XM(1) = XM(1) + AUX(1)$$

$$XM(1) = XM(1) + AUX(2)$$

...

$$XM(1) = XM(1) + AUX(N1)$$

$$XM(1) = XM(1)/N$$

y así sucesivamente.

Para evitar esta iteración se dividió el proceso en dos bucles y se cambió la forma de ir sumando los valores a $XM(J)$ para que no existiera dependencia entre una iteración y la siguiente.

```

DO 340 I=1,N1
  DO 350 J=1,N
    XM(J) = XM(J) + AUX((J-1)*N+I)
350  CONTINUE
340 CONTINUE
DO 360 I=1,N
  XM(I) = XM(I) / N
360 CONTINUE

```

De esta manera, se puede vectorizar el bucle DO 350

$$XM(1) = XM(1) + AUX(1)$$

$$XM(2) = XM(2) + AUX(N + 1)$$

...

$$XM(N) = XM(N) + AUX((N - 1) \times N + 1)$$

y así sucesivamente.

ya que son instrucciones independientes y, también, el bucle DO 360.

Como norma general, los bucles que desde un punto de vista secuencial corresponderían a las iteraciones de nivel más anidado, deben de corresponder a los bucles más externos desde un punto de vista vectorial, y viceversa.

Todos los programas realizados en esta memoria son un buen ejemplo de la implementación de este tipo de ideas y están a la disposición de cualquiera que lo desee.

4.3.4 Resultados

Para tener una visión lo más amplia posible realizando el menor número de simulaciones se ha planteado la siguiente estrategia. Todas las muestras (cualquier x o v) han sido generadas de acuerdo a tres valores dados $\Delta\mu = \mu_2 - \mu_1$ (dependiendo de los valores específicos de n_1 y n_2) proporcionados por tres potencias conocidas ($p_{.1}$) de 0.05 (hipótesis nula), 0.55 y 0.92 para el test de control, t de Student. Para cada uno de los anteriores valores $\Delta\mu$, se ejecutaron dos series de simulaciones. La primera era para varianzas iguales o semejantes: ($\sigma_1^2 = \sigma_2^2 = 1$), ($\sigma_1^2 = 1.05, \sigma_2^2 = 0.95$) y ($\sigma_1^2 = 1.10, \sigma_2^2 = 0.90$). En tanto que la segunda, era para unas varianzas mucho más diferentes: ($\sigma_1^2 = 1.333333333, \sigma_2^2 = 0.666666667$) y ($\sigma_1^2 = 1.6, \sigma_2^2 = 0.4$). Bajo estas condiciones⁴, se generaban los valores del vector x , mientras que se producían los valores del vector v para un valor σ^2 común e intermedio entre σ_1^2 y σ_2^2 que se obtenía al aplicar la expresión (4.25). Al utilizar tamaños muestrales iguales en ambas muestras, $n_1 = n_2$, se consiguió que en todas las configuraciones (σ_1^2, σ_2^2) estudiadas el valor de σ^2 siempre fuera 1. Los valores de los tamaños muestrales introducidos fueron 11, 31, 61 y 121.

El total de simulaciones realizadas se calcula al multiplicar el número de valores $\Delta\mu$ por el número de patrones de varianza (σ_1^2, σ_2^2) y el número de tamaños muestrales ($n_1 = n_2$) estudiados. Por consiguiente, en la primera serie se realizan

⁴La pareja de valores (σ_1^2, σ_2^2) se escogió con la condición de que (4.25) fuera igual a uno y que el cociente entre σ_1^2/σ_2^2 sea cada vez mayor. Así, pues, en la primera serie de simulaciones los valores del cociente son próximos a uno: 1, 1.105, 1.22; mientras que en la segunda serie, son altos: 2, 4.

$3 \times 3 \times 4 = 36$ simulaciones y en la segunda serie $3 \times 2 \times 4 = 24$, un total de 60 simulaciones para cada uno de los dos estudios, el test geodésico y el test de Cochran y Cox.

Las tablas 1 hasta 8 muestran algunos resultados de las simulaciones consistentes en la generación de $n = 5000$ parejas de valores de (Y, C) . Las cuatro primeras tablas muestran los resultados del estudio de la potencia del nuevo test, donde la variable respuesta, Y , se basa en la distancia geodésica al cuadrado, D^2 , mientras que en las cuatro restantes Y se obtiene de evaluar el test de Cochran y Cox.

En ambos casos, cada tabla, se divide en cinco subtablas según el valor de la pareja de (σ_1^2, σ_2^2) utilizada. Las tres primeras son para las varianzas semejantes y las dos siguientes para las varianzas más diferentes.

En cada subtabla se hace mención a la siguiente información:

\hat{p}_1 : Potencia del test de estudio estimada sólo a partir de la frecuencia relativa de Y .

$\pm 1.959964\hat{\sigma}$: Margen de error asociado a un nivel de confianza del 95% para la verdadera potencia p_1 , a partir de la estimación del error estándar de la frecuencia relativa.

\tilde{p}_1 : Potencia del test de estudio estimada usando el nuevo estimador.

$\pm 1.959964\hat{\sigma}_{GS}$: Margen de error asociado a un nivel de confianza del 95% para la verdadera potencia p_1 , utilizando la aproximación de Grab & Savage (Grab & Savage, 1954) en la estimación del error estándar del nuevo estimador.

$\widehat{\%VR}$: Estimación (conservadora) del porcentaje de reducción de varianza del nuevo estimador con respecto al estimador habitual, \hat{p}_1 .

$$\widehat{\%VR} = \left(1 - \frac{(n-1)\hat{\sigma}_{GS}^2}{\hat{p}_1(1-\hat{p}_1)} \right) 100$$

$\widehat{\%MER}$: Estimación del porcentaje de reducción del margen de error del nuevo estimador en comparación con el estimador usual, \hat{p}_1 .

$$\widehat{\%MER} = \left(1 - \hat{\sigma}_{GS} \sqrt{\frac{(n-1)}{\hat{p}_1(1-\hat{p}_1)}} \right) 100$$

Además, pueden verse algunos signos, † o ‡, en los valores de las columnas etiquetadas como \hat{p}_1 y \tilde{p}_1 . El primero, †, saldrá cuando la extensión del test, tanto en el test Cochran y Cox como en el test geodésico, para \hat{p}_1 (o \tilde{p}_1), no incluya el nivel de significación nominal (0.05), por esta razón solo puede aparecer sobre la primera fila ($\Delta\mu = 0$). El signo ‡ manifiesta la existencia de diferencias significativas entre el valor obtenido en el test de Cochran y Cox de \hat{p}_1 (o \tilde{p}_1) y su correspondiente valor en el test geodésico. Más concretamente, estos casos significativos corresponde a rechazar la hipótesis nula de igualdad de valores de p_1 versus una hipótesis alternativa afirmando que el test geodésico es más potente que el test de Cochran y Cox. El estadístico del test se basa en la aplicación del teorema central del límite, de tal manera que se supone que

$$\frac{\hat{p}_1^g - \hat{p}_1^{cc}}{\sqrt{(\hat{\sigma}^g)^2 + (\hat{\sigma}^{cc})^2}} \sim N(0, 1)$$

y

$$\frac{\tilde{p}_1^g - \tilde{p}_1^{cc}}{\sqrt{(\hat{\sigma}_{GS}^g)^2 + (\hat{\sigma}_{GS}^{cc})^2}} \sim N(0, 1)$$

donde el superíndice (g) indica que ha sido obtenido sobre los datos del test geodésico y (cc) sobre los datos del test de Cochran y Cox. Por ejemplo, \hat{p}_1^g expresa el valor de \hat{p}_1 en el test geodésico y \hat{p}_1^{cc} su homólogo en el test de Cochran y Cox.

4.3.4.1 Resultados del estudio de simulación del test geodésico

Tabla 1. Nivel de significación, potencia y grado de reducción de la varianza en el estudio del test geodésico definido mediante (4.15).

† =significativamente diferente del nivel de significación nominal.

‡ =significativamente diferente del valor correspondiente en el test Cochran y Cox.

$n = 5000, n_1 = n_2 = 11, B = 1000$

$\Delta\mu$	$\sigma_1^2 = 1.0$		$\sigma_2^2 = 1.0$		$\widehat{\%VR}$	$\widehat{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0488†	0.005973	0.04993†	0.001985	88.95	66.76
0.927147567	0.5516†	0.013786	0.5472†	0.004048	91.38	70.64
1.50755596	0.922 ‡	0.007434	0.9184 ‡	0.002624	87.54	64.70

$\Delta\mu$	$\sigma_1^2 = 1.05$		$\sigma_2^2 = 0.95$		$\widehat{\%VR}$	$\widehat{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0478†	0.005914	0.04892†	0.002014	88.40	65.94
0.927147567	0.5516†	0.013786	0.54721†	0.004120	91.07	70.11
1.50755596	0.9228†	0.007399	0.91923†	0.002678	86.90	63.80

$\Delta\mu$	$\sigma_1^2 = 1.10$		$\sigma_2^2 = 0.90$		$\widehat{\%VR}$	$\widehat{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0472†	0.005879	0.0483 ‡	0.002113	87.08	64.06
0.927147567	0.5532†	0.013782	0.54884†	0.004430	89.67	67.86
1.50755596	0.923 ‡	0.007390	0.91949†	0.002963	83.92	59.91

$\Delta\mu$	$\sigma_1^2 = 1.333333333$		$\sigma_2^2 = 0.666666667$		$\widehat{\%VR}$	$\widehat{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.051 ‡	0.006099	0.052 ‡	0.003540	66.30	41.95
0.927147567	0.5514†	0.013787	0.54736†	0.006606	77.04	52.09
1.50755596	0.9212†	0.007469	0.91807†	0.004410	65.13	40.95

Δ	$\sigma_1^2 = 1.6$		$\sigma_2^2 = 0.4$		$\widehat{\%VR}$	$\widehat{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0522†	0.006166	0.05302†	0.004632	43.56	24.87
0.927147567	0.5426†	0.013810	0.53899†	0.008606	61.17	37.69
1.50755596	0.9098†	0.007941	0.90706†	0.005955	43.76	25.01

Tabla 2. Nivel de significación, potencia y grado de reducción de la varianza en el estudio del test geodésico definido mediante (4.15).

† =significativamente diferente del nivel de significación nominal.

‡ =significativamente diferente del valor correspondiente en el test Cochran y Cox.

$n = 5000$, $n_1 = n_2 = 31$, $B = 1000$

$\Delta\mu$	$\sigma_1^2 = 1.0$		$\sigma_2^2 = 1.0$		$\overline{\%VR}$	$\overline{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0486	0.005961	0.04879†	0.001890	89.95	68.29
0.538815856	0.5622	0.013753	0.54521‡	0.004166	90.83	69.71
0.866595685	0.9228	0.007399	0.91876‡	0.002257	90.70	69.50

$\Delta\mu$	$\sigma_1^2 = 1.05$		$\sigma_2^2 = 0.95$		$\overline{\%VR}$	$\overline{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.049	0.005984	0.04919‡	0.001894	89.98	68.35
0.538815856	0.5618	0.013754	0.54486‡	0.004269	90.37	68.96
0.866595685	0.9228	0.007399	0.91874‡	0.002117	91.81	71.38

$\Delta\mu$	$\sigma_1^2 = 1.10$		$\sigma_2^2 = 0.90$		$\overline{\%VR}$	$\overline{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0488	0.005973	0.04898‡	0.002142	87.14	64.14
0.538815856	0.563	0.013750	0.54622‡	0.004634	88.64	66.30
0.866595685	0.9224	0.007417	0.91841‡	0.002569	88.00	65.36

$\Delta\mu$	$\sigma_1^2 = 1.333333333$		$\sigma_2^2 = 0.666666667$		$\overline{\%VR}$	$\overline{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0464	0.005831	0.04656†	0.003280	68.36	43.76
0.538815856	0.5564	0.013772	0.54096‡	0.006908	74.84	49.84
0.866595685	0.9244	0.007328	0.92085‡	0.003945	71.02	46.17

$\Delta\mu$	$\sigma_1^2 = 1.6$		$\sigma_2^2 = 0.4$		$\overline{\%VR}$	$\overline{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0442†	0.005698	0.04433†	0.004190	45.92	26.46
0.538815856	0.5538	0.013780	0.53979	0.008550	61.50	37.95
0.866595685	0.9214	0.007460	0.91822	0.005033	54.49	32.53

Tabla 3. Nivel de significación, potencia y grado de reducción de la varianza en el estudio del test geodésico definido mediante (4.15).

† =significativamente diferente del nivel de significación nominal.

‡ =significativamente diferente del valor correspondiente en el test Cochran y Cox.

$n = 5000, n_1 = n_2 = 61, B = 1000$

$\Delta\mu$	$\sigma_1^2 = 1.0 \quad \sigma_2^2 = 1.0$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0486	0.005961	0.05032	0.001838	90.50	69.17
0.377708852	0.5458	0.013802	0.54676†	0.004067	91.32	70.53
0.608175159	0.916	0.007689	0.91866	0.002557	88.94	66.75

$\Delta\mu$	$\sigma_1^2 = 1.05 \quad \sigma_2^2 = 0.95$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0478	0.005914	0.0495	0.001913	89.54	67.65
0.377708852	0.547	0.013799	0.54795†	0.004174	90.85	69.75
0.608175159	0.9156	0.007706	0.91827	0.002503	89.45	67.52

$\Delta\mu$	$\sigma_1^2 = 1.10 \quad \sigma_2^2 = 0.90$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0478	0.005914	0.04948	0.002131	87.02	63.97
0.377708852	0.5482	0.013796	0.54914‡	0.004699	88.40	65.94
0.608175159	0.9148	0.007739	0.91745	0.002730	87.56	64.73

$\Delta\mu$	$\sigma_1^2 = 1.333333333 \quad \sigma_2^2 = 0.666666667$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0482	0.005937	0.04967	0.003458	66.08	41.76
0.377708852	0.5476	0.013798	0.54848‡	0.006593	77.16	52.21
0.608175159	0.916	0.007689	0.91838	0.004112	71.40	46.52

$\Delta\mu$	$\sigma_1^2 = 1.6 \quad \sigma_2^2 = 0.4$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0512	0.006110	0.05247	0.004487	46.07	26.56
0.377708852	0.5408	0.013814	0.54158	0.008654	60.75	37.35
0.608175159	0.9128	0.007821	0.91491	0.005274	54.53	32.57

Tabla 4. Nivel de significación, potencia y grado de reducción de la varianza en el estudio del test geodésico definido mediante (4.15).

† =significativamente diferente del nivel de significación nominal.

‡ =significativamente diferente del valor correspondiente en el test Cochran y Cox.

$n = 5000, n_1 = n_2 = 121, B = 1000$

$\Delta\mu$	$\sigma_1^2 = 1.0 \quad \sigma_2^2 = 1.0$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\overline{\%VR}$	$\overline{\%MER}$
0	0.0526	0.006188	0.04941†	0.002006	89.48	67.58
0.268181860	0.545	0.013804	0.5475†	0.003844	92.25	72.16
0.4318118128	0.9108	0.007901	0.91603	0.002482	90.14	68.59

$\Delta\mu$	$\sigma_1^2 = 1.05 \quad \sigma_2^2 = 0.95$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\overline{\%VR}$	$\overline{\%MER}$
0	0.052	0.006155	0.04886	0.002164	87.64	64.84
0.268181860	0.5432	0.013809	0.54569	0.003976	91.71	71.21
0.4318118128	0.9098	0.007941	0.91499	0.002738	88.11	65.52

$\Delta\mu$	$\sigma_1^2 = 1.10 \quad \sigma_2^2 = 0.90$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\overline{\%VR}$	$\overline{\%MER}$
0	0.0528	0.006199	0.04967†	0.002302	86.22	62.87
0.268181860	0.5418	0.013812	0.54427	0.004380	89.95	68.29
0.4318118128	0.91	0.007933	0.91511	0.003010	85.61	62.06

$\Delta\mu$	$\sigma_1^2 = 1.333333333 \quad \sigma_2^2 = 0.666666667$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\overline{\%VR}$	$\overline{\%MER}$
0	0.0506	0.006076	0.04786	0.003377	69.11	44.43
0.268181860	0.5422	0.013811	0.5445	0.006460	78.12	53.23
0.4318118128	0.9122	0.007845	0.91681	0.004171	71.74	46.84

$\Delta\mu$	$\sigma_1^2 = 1.6 \quad \sigma_2^2 = 0.4$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\overline{\%VR}$	$\overline{\%MER}$
0	0.0524	0.006177	0.05009	0.004441	48.33	28.11
0.268181860	0.5408	0.013814	0.54285	0.008527	61.90	38.27
0.4318118128	0.91	0.007933	0.9142	0.005114	58.45	35.54

4.3.4.2 Resultados del estudio de simulación del test de Cochran y Cox

Tabla 5. Nivel de significación, potencia y grado de reducción de la varianza en el estudio del test de Cochran y Cox.

† =significativamente diferente del nivel de significación nominal.

‡ =significativamente diferente del valor correspondiente en el test geodésico.

$n = 5000, n_1 = n_2 = 11$

$\Delta\mu$	$\sigma_1^2 = 1.0$		$\sigma_2^2 = 1.0$		$\overline{\%VR}$	$\overline{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0386†‡	0.005340	0.03955†‡	0.002530	77.55	52.63
0.927147567	0.5046†	0.013860	0.50041†	0.005889	81.95	57.51
1.50755596	0.9026†	0.008219	0.89889†	0.003982	76.53	51.56

$\Delta\mu$	$\sigma_1^2 = 1.05$		$\sigma_2^2 = 0.95$		$\overline{\%VR}$	$\overline{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.038 †‡	0.005300	0.03893†‡	0.002583	76.25	51.26
0.927147567	0.5052†	0.013860	0.50101†	0.005857	82.14	57.74
1.50755596	0.9022†	0.008234	0.89849†	0.004018	76.19	51.20

$\Delta\mu$	$\sigma_1^2 = 1.10$		$\sigma_2^2 = 0.90$		$\overline{\%VR}$	$\overline{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0378†‡	0.005287	0.03873†‡	0.002600	75.81	50.82
0.927147567	0.5044†	0.013860	0.50022†	0.005923	81.74	57.26
1.50755596	0.9016†	0.008257	0.89789†	0.004072	75.68	50.69

$\Delta\mu$	$\sigma_1^2 = 1.333333333$		$\sigma_2^2 = 0.666666667$		$\overline{\%VR}$	$\overline{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0426†‡	0.005598	0.04352†‡	0.003229	66.74	42.33
0.927147567	0.5106†	0.013857	0.50656†	0.006730	76.41	51.43
1.50755596	0.8982†	0.008382	0.89472†	0.005000	64.42	40.35

$\Delta\mu$	$\sigma_1^2 = 1.6$		$\sigma_2^2 = 0.4$		$\overline{\%VR}$	$\overline{\%MER}$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0406†‡	0.005471	0.04132†‡	0.004171	41.87	23.76
0.927147567	0.5122†	0.013856	0.50857†	0.008581	61.64	38.07
1.50755596	0.8986†	0.008368	0.89569†	0.006222	44.71	25.64

Tabla 6. Nivel de significación, potencia y grado de reducción de la varianza en el estudio del test de Cochran y Cox.

† =significativamente diferente del nivel de significación nominal.

‡ =significativamente diferente del valor correspondiente en el test geodésico.

$n = 5000$, $n_1 = n_2 = 31$

$\Delta\mu$	$\sigma_1^2 = 1.0$		$\sigma_2^2 = 1.0$		$\%VR$	$\%MER$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0452	0.005759	0.04538†‡	0.001802	90.21	68.71
0.538815856	0.5494	0.013793	0.53218‡	0.003641	93.03	73.60
0.866595685	0.9182	0.007597	0.91403‡	0.002135	92.10	71.89

$\Delta\mu$	$\sigma_1^2 = 1.05$		$\sigma_2^2 = 0.95$		$\%VR$	$\%MER$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0452	0.005759	0.04538†‡	0.001802	90.21	68.71
0.538815856	0.5488	0.013794	0.5316 ‡	0.003698	92.81	73.19
0.866595685	0.9182	0.007597	0.91403‡	0.002135	92.10	71.89

$\Delta\mu$	$\sigma_1^2 = 1.10$		$\sigma_2^2 = 0.90$		$\%VR$	$\%MER$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0454	0.005771	0.04558†‡	0.001989	88.12	65.53
0.538815856	0.5494	0.013793	0.53238‡	0.004174	90.84	69.74
0.866595685	0.9182	0.007597	0.91405‡	0.002277	91.01	70.02

$\Delta\mu$	$\sigma_1^2 = 1.333333333$		$\sigma_2^2 = 0.666666667$		$\%VR$	$\%MER$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0446	0.005722	0.04476†	0.003024	72.08	47.16
0.538815856	0.5464	0.013801	0.53084‡	0.006784	75.84	50.84
0.866595685	0.9192	0.007555	0.91556‡	0.004124	70.20	45.41

$\Delta\mu$	$\sigma_1^2 = 1.6$		$\sigma_2^2 = 0.4$		$\%VR$	$\%MER$
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$		
0	0.0424†	0.005586	0.04253†	0.004050	47.43	27.49
0.538815856	0.5474	0.013798	0.53334	0.008526	61.82	38.21
0.866595685	0.9186	0.007580	0.91536	0.005093	54.86	32.81

Tabla 7. Nivel de significación, potencia y grado de reducción de la varianza en el estudio del test de Cochran y Cox.

† =significativamente diferente del nivel de significación nominal.

‡ =significativamente diferente del valor correspondiente en el test geodésico.

$n = 5000, n_1 = n_2 = 61$

$\Delta\mu$	$\sigma_1^2 = 1.0 \quad \sigma_2^2 = 1.0$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.047	0.005867	0.04876†	0.000970	97.27	83.46
0.377708852	0.5408	0.013814	0.54179‡	0.002494	96.74	81.95
0.608175159	0.9144	0.007756	0.91719	0.001467	96.42	81.09

$\Delta\mu$	$\sigma_1^2 = 1.05 \quad \sigma_2^2 = 0.95$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0468	0.005855	0.04855†	0.001044	96.81	82.16
0.377708852	0.5384	0.013819	0.53938‡	0.003132	94.86	77.33
0.608175159	0.9134	0.007796	0.91619	0.001708	95.20	78.10

$\Delta\mu$	$\sigma_1^2 = 1.10 \quad \sigma_2^2 = 0.90$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0468	0.005855	0.04852†	0.001415	94.17	75.84
0.377708852	0.5402	0.013816	0.54116‡	0.003954	91.81	71.38
0.608175159	0.9138	0.007780	0.91655	0.002100	92.71	73.01

$\Delta\mu$	$\sigma_1^2 = 1.333333333 \quad \sigma_2^2 = 0.666666667$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0466	0.005843	0.04811	0.003075	72.31	47.38
0.377708852	0.5392	0.013818	0.54008‡	0.006574	77.37	52.42
0.608175159	0.9146	0.007747	0.91708	0.003778	76.21	51.23

$\Delta\mu$	$\sigma_1^2 = 1.6 \quad \sigma_2^2 = 0.4$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.049	0.005984	0.05026	0.004352	47.11	27.28
0.377708852	0.5364	0.013824	0.53718	0.008655	60.80	37.39
0.608175159	0.913	0.007813	0.91515	0.005149	56.57	34.10

Tabla 8. Nivel de significación, potencia y grado de reducción de la varianza en el estudio del test de Cochran y Cox.

† =significativamente diferente del nivel de significación nominal.

‡ =significativamente diferente del valor correspondiente en el test geodésico.

$n = 5000$, $n_1 = n_2 = 121$

$\Delta\mu$	$\sigma_1^2 = 1.0 \quad \sigma_2^2 = 1.0$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0504	0.006064	0.04719†‡	0.001432	94.42	76.38
0.268181860	0.5406	0.013815	0.54317‡	0.002277	97.28	83.52
0.4318118128	0.911	0.007893	0.91638	0.001665	95.55	78.90

$\Delta\mu$	$\sigma_1^2 = 1.05 \quad \sigma_2^2 = 0.95$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.05	0.006042	0.04682†	0.001519	93.68	74.85
0.268181860	0.5396	0.013817	0.54215	0.002838	95.78	79.46
0.4318118128	0.91	0.007933	0.91535	0.002033	93.43	74.37

$\Delta\mu$	$\sigma_1^2 = 1.10 \quad \sigma_2^2 = 0.90$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.05	0.006042	0.04686†‡	0.001774	91.38	70.64
0.268181860	0.5382	0.013820	0.54071	0.003724	92.74	73.05
0.4318118128	0.9108	0.007901	0.91605	0.002362	91.07	70.11

$\Delta\mu$	$\sigma_1^2 = 1.333333333 \quad \sigma_2^2 = 0.666666667$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0504	0.006064	0.04758	0.003126	73.42	48.45
0.268181860	0.5402	0.013816	0.54252	0.006267	79.42	54.64
0.4318118128	0.9116	0.007869	0.91633	0.003938	74.96	49.96

$\Delta\mu$	$\sigma_1^2 = 1.6 \quad \sigma_2^2 = 0.4$					
	\hat{p}_1	$\pm 1.959964\hat{\sigma}$	\tilde{p}_1	$\pm 1.959964\hat{\sigma}_{GS}$	$\widehat{\%VR}$	$\widehat{\%MER}$
0	0.0506	0.006076	0.04829	0.004310	49.68	29.06
0.268181860	0.5394	0.013817	0.54148	0.008303	63.89	39.91
0.4318118128	0.9112	0.007885	0.91536	0.005096	58.24	35.38

4.3.5 Gráficos

Se presenta un conjunto de gráficos asociados a las anteriores tablas de resultados.

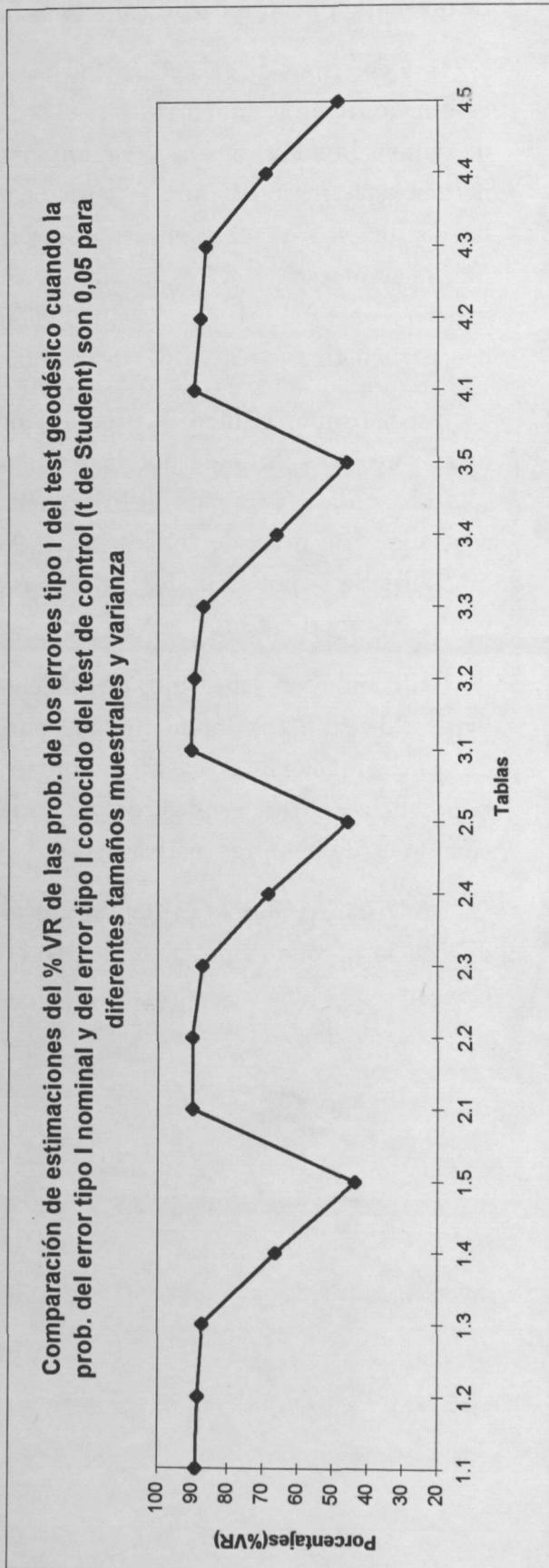
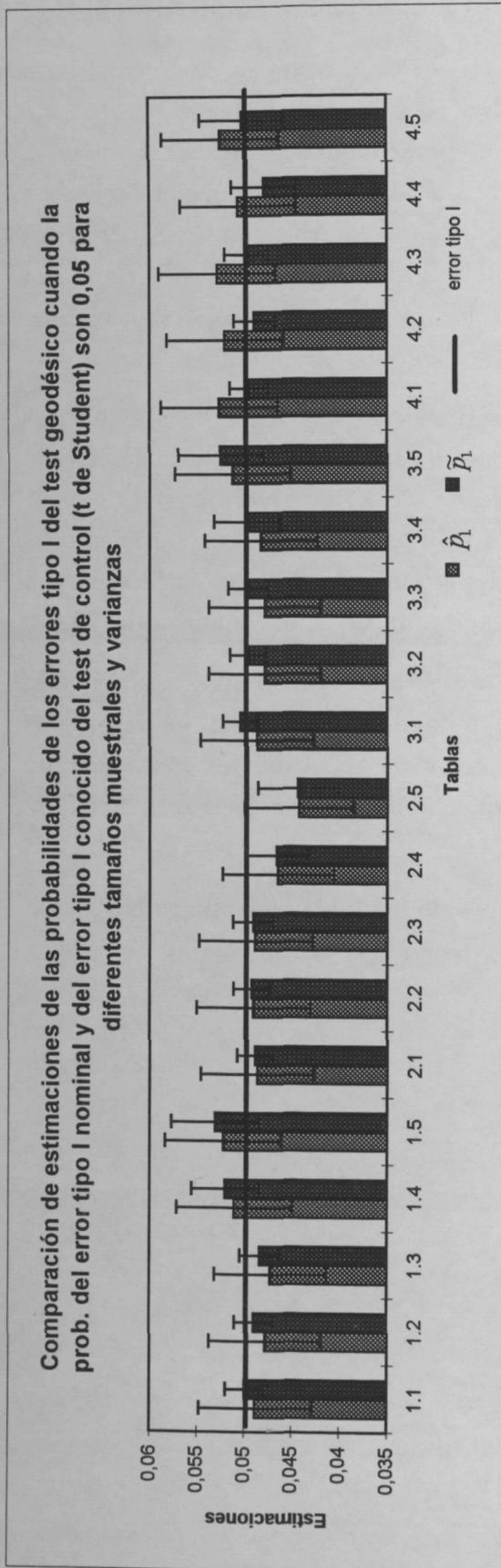
Hay dos tipos de gráficos. El primero es el que tiene como eje de ordenadas las estimaciones de p_1 , mientras que el segundo tipo son estimaciones del porcentaje de reducción de la varianza. En ambos, la numeración del eje de abscisas sigue la notación "número de tabla". "posición de la subtabla", de esta manera se indica las condiciones experimentales que corresponden a esa subtabla. Así, por ejemplo, la numeración 3.2 es la tabla 3 subtabla 2 que corresponde a la condición experimental del estudio de simulación de la potencia del nuevo test geodésico con $n = 5000$, $n_1 = n_2 = 61$, $\sigma_1^2 = 1.05$ $\sigma_2^2 = 0.95$.

En el primer gráfico se presenta dos series de estimaciones de p_1 . La primera serie está formada por las estimaciones de la potencia del test de estudio sólo obtenida a partir de la frecuencia relativa de Y , \hat{p}_1 , y el margen de error asociado a un nivel de confianza del 95%. Y la segunda serie son también estimaciones de la potencia del test de estudio pero usando el nuevo estimador, \tilde{p}_1 , y el correspondiente margen de error asociado a un nivel de confianza del 95%.

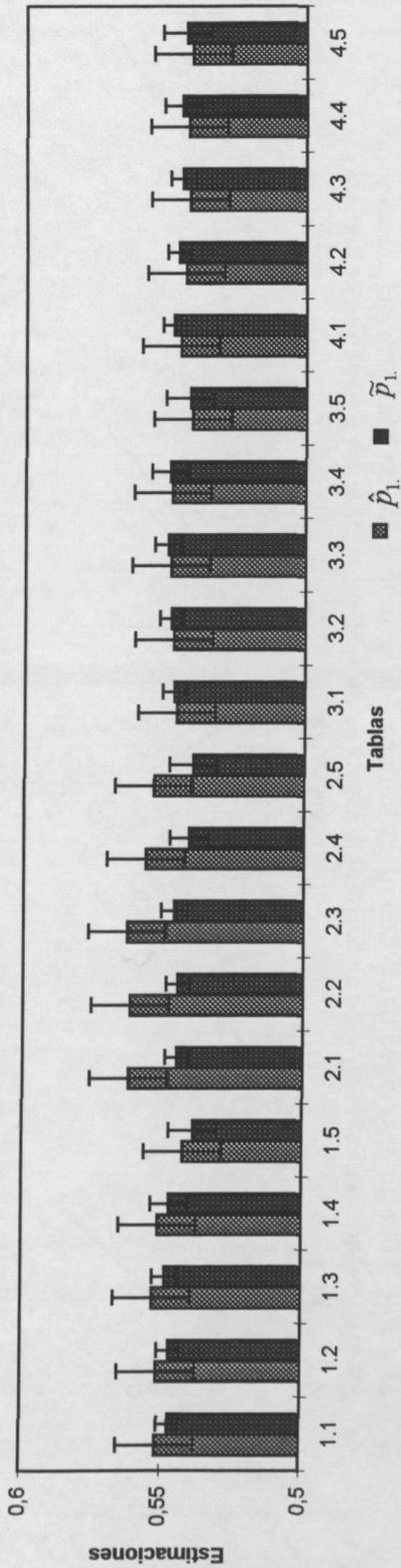
Utilizando estos dos tipos de gráficos se presentan las tablas 1 hasta la 4 en tres parejas de gráficos, donde en cada pareja (o página) se muestran los resultados de las estimaciones de la probabilidad de error de tipo I del test geodésico, de la potencia del test geodésico cuando la potencia del test de control es de 0.55 y cuando es de 0.92, respectivamente.

De igual manera se presentan los gráficos de las tablas 4 hasta la 8. Con la diferencia de que la potencia que se está evaluando es la del test de Cochran y Cox.

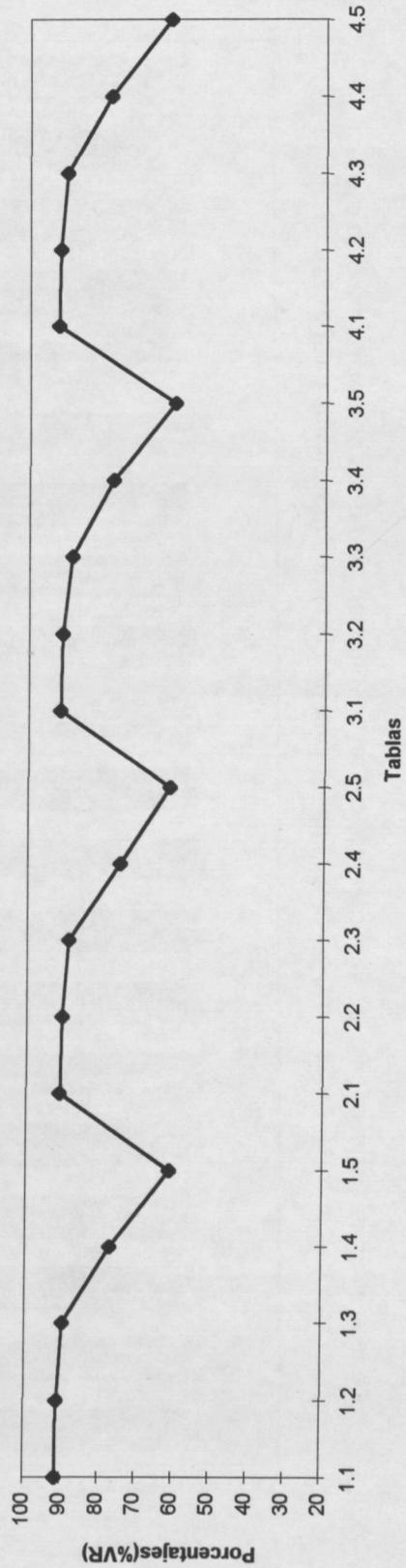
4.3.5.1 Gráficos de los resultados del estudio simulación del test geodésico



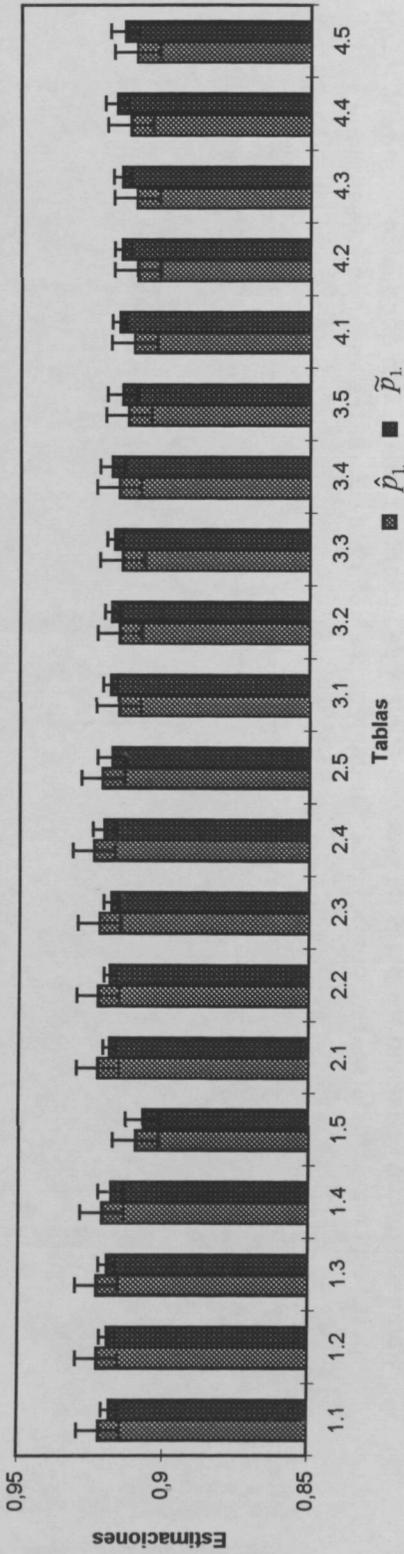
Comparación de estimaciones de las potencias del test geodésico cuando la potencia conocida del test de control (t de Student) es de 0,55 para diferentes tamaños muestrales y varianzas



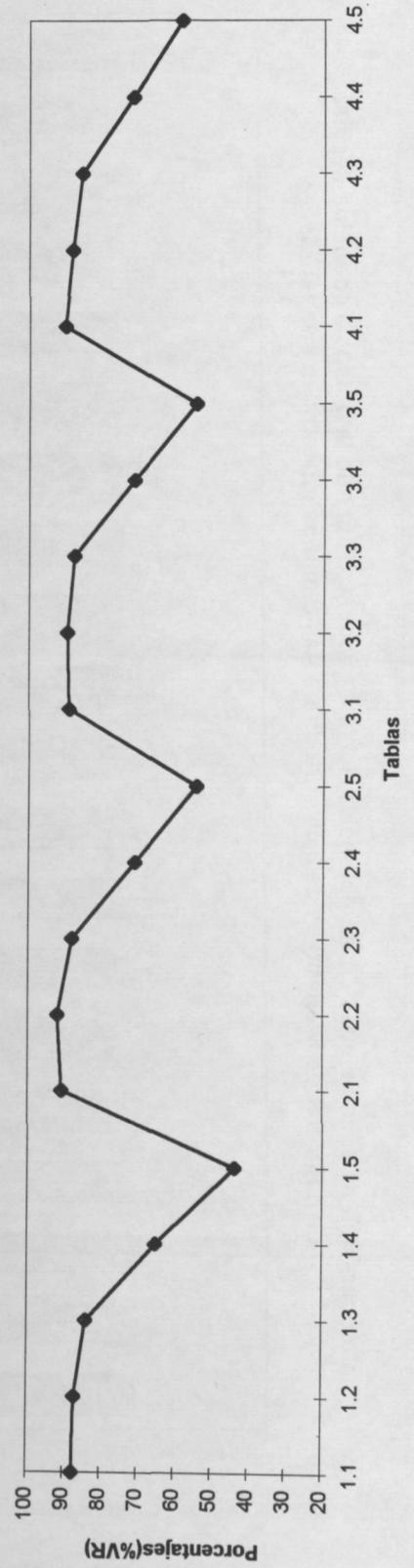
Comparación de estimaciones del % de red. de varianza de las potencias del test geodésico cuando la potencia conocida del test de control (t de Student) es de 0,55 para diferentes tamaños muestrales y varianzas



Comparación de estimaciones de las potencias del test geodésico cuando la potencia conocida del test de control (t de Student) es de 0,92 para diferentes tamaños muestrales y varianzas

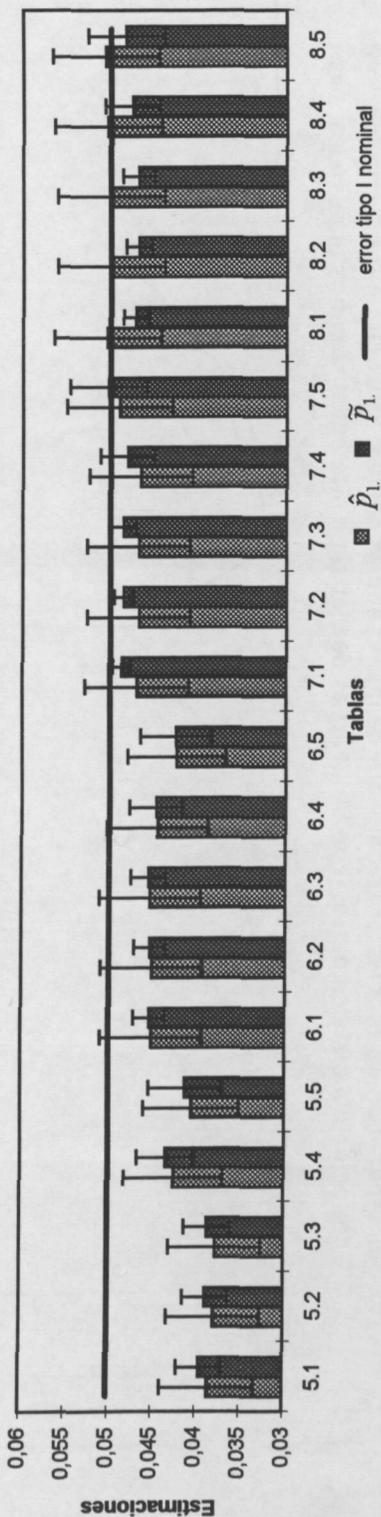


Comparación de estimaciones del % de red. de varianza de las potencias del test geodésico cuando la potencia conocida del test de control (t de Student) es de 0,92 para diferentes tamaños muestrales y varianzas

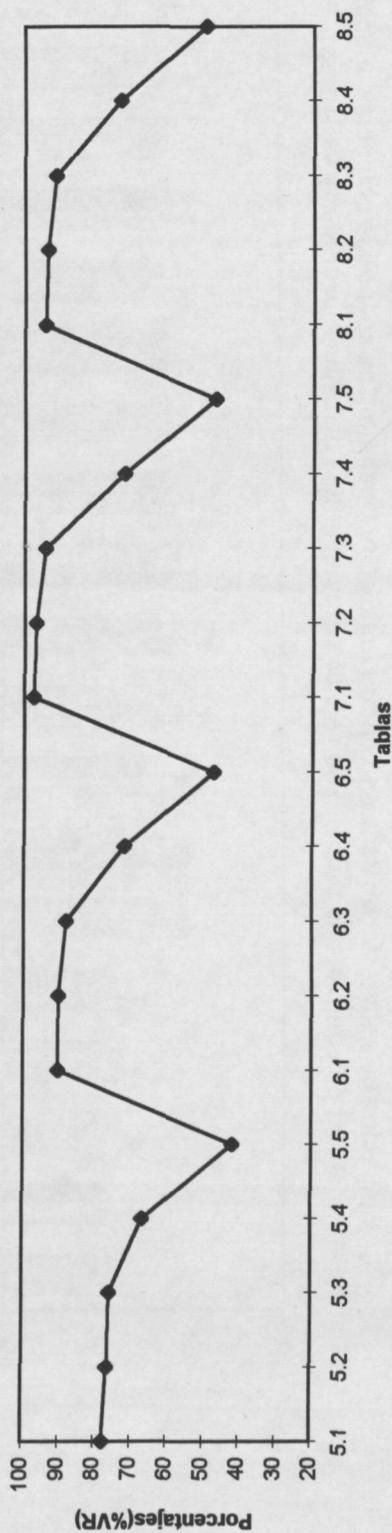


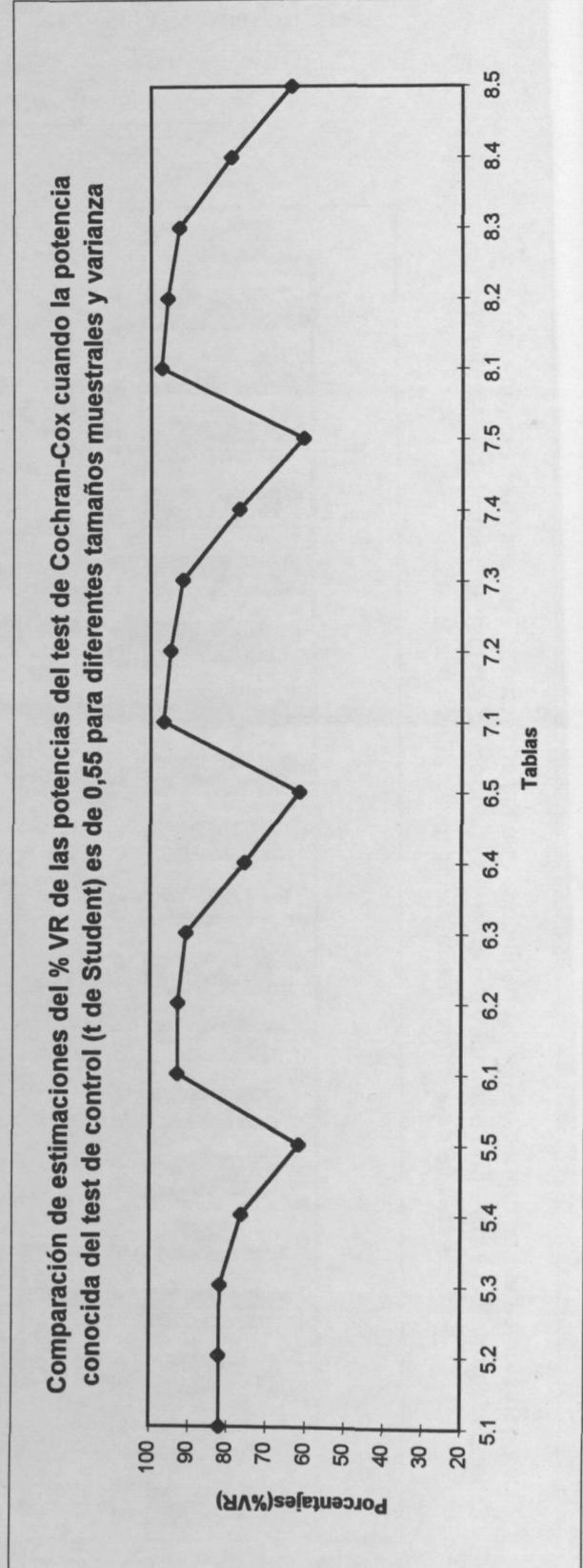
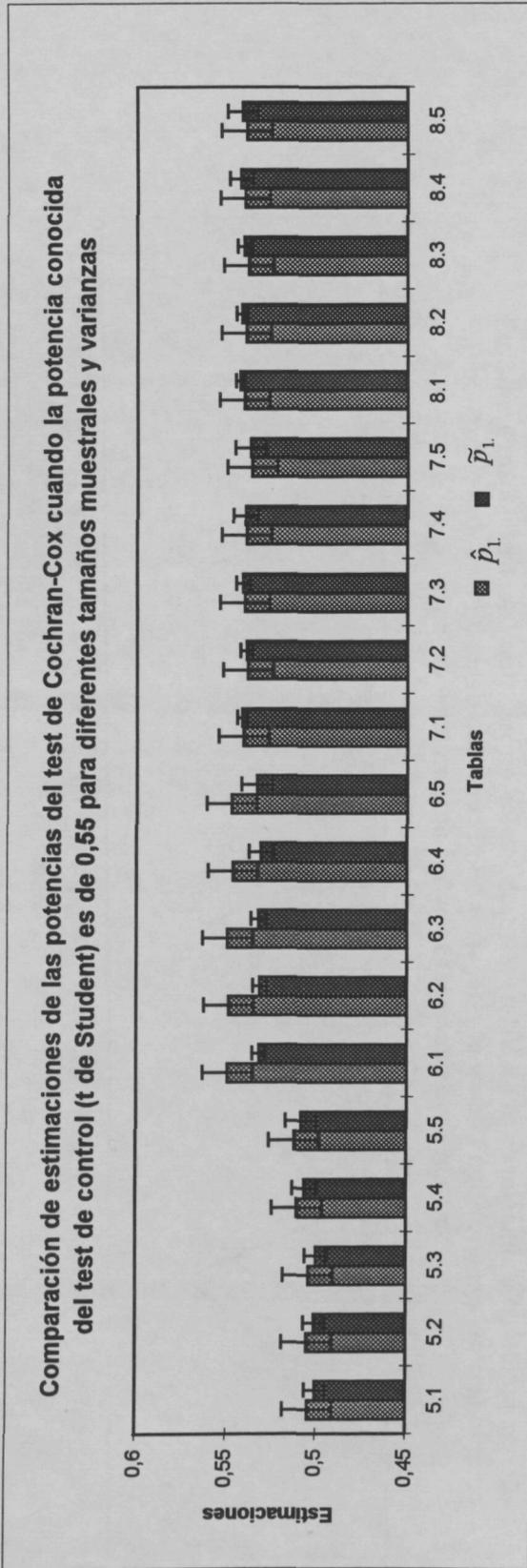
4.3.5.2 Gráficos de los resultados del estudio de simulación del test de Cochran y Cox

Comparación de estimaciones de las probabilidades de los errores tipo I del test de Cochran-Cox cuando la prob. del error tipo I nominal y del error tipo I conocido del test de control (t de Student) son 0,05 para diferentes tamaños muestrales y varianzas

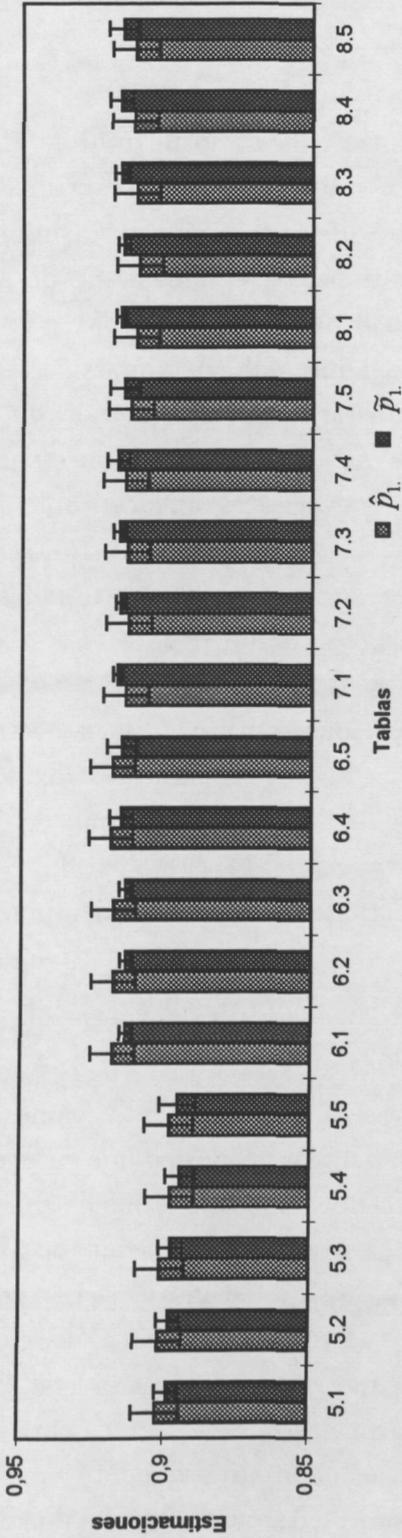


Comparación de estimaciones del % VR de las prob. de los errores tipo I del test de Cochran-Cox cuando la prob. del error tipo I nominal y del error tipo I conocido del test de control (t de Student) son 0,05 para diferentes tamaños muestrales y varianzas

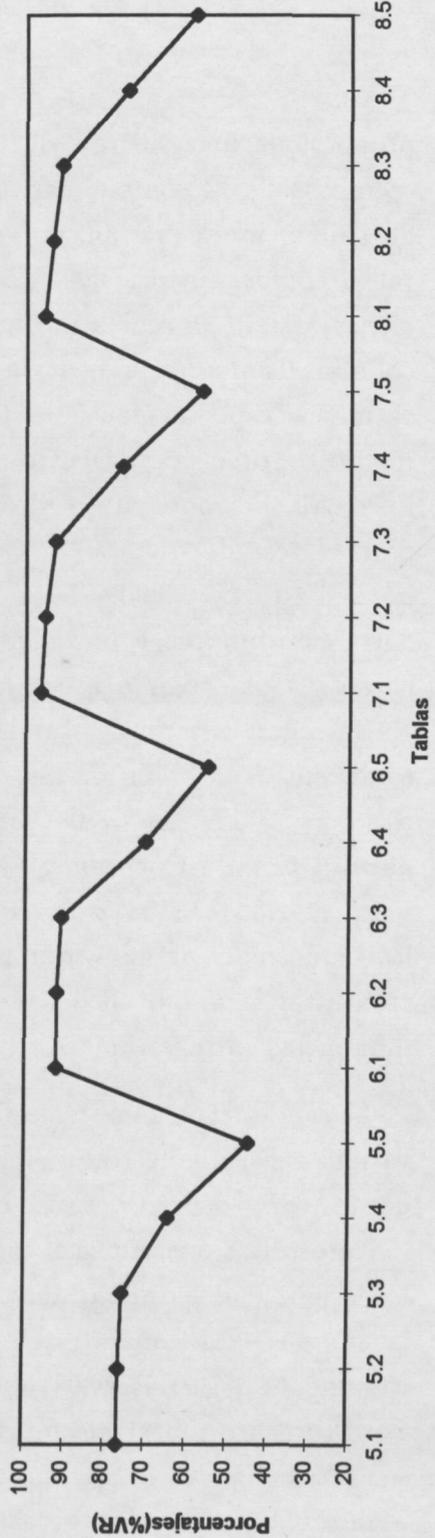




Comparación de estimaciones de las potencias del test de Cochran-Cox cuando la potencia conocida del test de control (t de Student) es de 0,92 para diferentes tamaños muestrales y varianzas



Comparación de estimaciones del % VR de las potencias del test de Cochran-Cox cuando la potencia conocida del test de control (t de Student) es de 0,92 para diferentes tamaños muestrales y varianzas



4.3.6 Conclusiones

El uso del estimador definido en el capítulo 3 como:

$$\tilde{p}_1 = p_0 \frac{n_{10}}{n_{00} + n_{10}} + p_1 \frac{n_{11}}{n_{01} + n_{11}} \quad (4.27)$$

proporciona una buena reducción de la varianza con un coste mínimo en el incremento de la complejidad del diseño de la simulación para la primera serie de simulaciones (varianzas iguales o no muy diferentes) que va disminuyendo tanto como aumenta las diferencias entre varianzas (segunda serie de simulaciones). Esto se refleja en las tablas 1 hasta 8 por la columna $\widehat{\%VR}$ y por la columna $\widehat{\%MER}$ (porcentaje de reducción del intervalo de confianza), el cual es menos espectacular pero tal vez más realista como una medida del incremento de precisión. Dentro de la primera serie de simulaciones la estimación más baja del porcentaje de reducción de la varianza es un aceptable 75.68% para el test de Cochran y Cox, cuando $n_1 = n_2 = 11$, $\Delta\mu = 1.50755596$ y $\sigma_1^2 = 1.10$ $\sigma_2^2 = 0.90$. Todas las reducciones de las varianzas estimadas están entre este mínimo y un valor máximo de 97.28%, también para el test de Cochran y Cox, cuando $n_1 = n_2 = 121$, $\Delta\mu = 0.268181860$ y $\sigma_1^2 = \sigma_2^2 = 1.0$. En la segunda serie de simulaciones las reducciones de las varianzas estimadas están entre un valor mínimo de 41.87% para el test Cochran y Cox, cuando $n_1 = n_2 = 11$, $\Delta\mu = 0$ y $\sigma_1^2 = 1.6$ $\sigma_2^2 = 0.4$, y un valor máximo de 79.42% para el test de Cochran y Cox, cuando $n_1 = n_2 = 121$, $\Delta\mu = 0.268181860$ y $\sigma_1^2 = 1.333333333$ $\sigma_2^2 = 0.666666667$. Esta tendencia a la disminución de las reducciones de las varianzas estimadas queda plasmada en las gráficas al incrementar la amplitud del intervalo de confianza de \tilde{p}_1 , cuando se aumenta las diferencias entre varianzas para un mismo tamaño muestral.

Para tamaños muestrales pequeños ($n_i = 11$), todas las reducciones de las varianzas estimadas son mejores en el estudio del test geodésico que en el estudio del test de Cochran y Cox. Esta situación se invierte para tamaños muestrales mayores. Tanto en un caso como otro, las diferencias se reducen cuando aumenta la desigualdad entre varianzas (segunda serie de simulaciones). Así se refleja el hecho que para tamaños muestrales pequeños el test de Cochran y Cox es marcadamente más conservativo y menos potente que el test de control t de Student. Para tamaños muestrales más grandes, por otra parte, el test de Cochran y Cox se acerca al test t de Student y ambos se convierten progresivamente equivalentes a un test basado en las tablas normal, y por tanto altamente correlacionados. Sin

embargo, en cualquier caso, la técnica de reducción de la varianza se comporta de forma similar y aceptablemente bien en el estudio de ambos tests.

El test de Cochran y Cox es conocido por ser conservativo. Su verdadera extensión tiende a ser menor que el nivel de significación nominal (Lee and Gurland, 1975) y esta tendencia es más fuerte en el caso de igualdad de varianzas (o, más exactamente, cuando $n_1\sigma_1^2 = n_2\sigma_2^2$). Esta propiedad puede ayudarnos a contrastar la sensibilidad de las pruebas de simulación y a señalar como se mejora por la técnica de reducción de la varianza. Al hacer 5000 réplicas de simulación, usando la frecuencia relativa para estimar la extensión del test (\hat{p}_1), es muy difícil detectar esta tendencia conservativa. Como así se muestra en las tablas 5 hasta 8 cuando $\Delta\mu = 0$, esto es, cuando las simulaciones son realizadas bajo la hipótesis de igualdad de medias. En estas situaciones, la mayoría de los valores estimados \hat{p}_1 no difieren significativamente del valor nominal de 0.05 ya que el intervalo de confianza aproximado del 95% alrededor de \hat{p}_1 , que aparece en la columna etiquetada por $\pm 1.959964\hat{\sigma}$ en las tablas 5 hasta 8, cubre erróneamente el valor 0.05 en muchos casos. Esta circunstancia se visualiza por el cruce de la línea $p = 0.05$ por muchos de estos intervalos de confianza en la gráfica (pág. 107) que corresponde al test de Cochran y Cox para $\Delta\mu = 0$.

Las únicas excepciones al principio precedente corresponde a casos muy claros de tamaños muestrales pequeños $n_1 = n_2 = 11$ y a un caso aislado con tamaños muestrales $n_1 = n_2 = 31$. Estos casos significativos se etiquetan con un signo † en las tablas. Incluso la hipótesis nula más fácil de rechazar, $p_1 = 0.05$ versus la alternativa $p_1 < 0.05$, se acepta en las mismas situaciones cuando el test se basa en \hat{p}_1 . Por otro lado, cuando \tilde{p}_1 se usó en el estudio del test de Cochran y Cox, la hipótesis $p_1 = 0.05$ se rechaza en todos los casos en que las varianzas son iguales o semejantes y se acepta cuando las varianzas son muy diferentes sólo en aquellos casos de tamaños muestrales grandes (61 y 121). Así queda expuesto en las tablas 5 hasta 8 (si hay diferencias significativas de nuevo se marca con †) y en la gráfica que corresponde al test de Cochran y Cox para $\Delta\mu = 0$, donde el intervalo de confianza basado en \tilde{p}_1 no incluye el valor 0.05 en la primera serie de simulaciones mientras que esto no sucede en la segunda serie (para $n_1 = n_2 = 61$ y 121) ya que el intervalo de confianza basado en \tilde{p}_1 se hace cada vez más amplio. La falta de rechazo en la segunda serie de simulaciones está en concordancia con el hecho que el test de Cochran y Cox pierde progresivamente su carácter conservador cuando las varianzas se convierten gradualmente en desiguales. Por otra parte, se puede

observar como el carácter conservador de este test también se pone de manifiesto para las potencias de 0.55 y 0.92 en muestras pequeñas, $n_1 = n_2 = 11$.

Por tanto, se presenta un caso claro en el cual un rasgo cualitativo se puede detectar gracias al uso de la reducción de la varianza.

En contraste con la tendencia observada para el test de Cochran y Cox, ninguna de las extensiones estimadas para el test geodésico (tanto si las medimos usando \hat{p}_1 como \tilde{p}_1 .) difiere de manera significativa del valor nominal del 0.05 incluso para las varianzas más desiguales (excepto para un extraño caso de $n_1 = n_2 = 31, \sigma_1^2 = 1.6, \sigma_2^2 = 0.4$). Así queda de manifiesto en la pérdida del signo † en las tablas 1 hasta 4 y en la gráfica que corresponde al test geodésico para $\Delta\mu = 0$, donde la línea $p = 0.05$ corta el intervalo de confianza alrededor de la estimación de la extensión. En vista de lo anterior, se puede concluir que la verdadera extensión del test geodésico es igual al nivel de significación establecido o muy próximo a él.

Ambos tests bajo estudio, tienen potencias comparables pero significativamente diferentes para tamaños muestrales pequeños de n_1 y n_2 , a un nivel del 0.05, pero paulatinamente van disminuyendo sus diferencias al aumentar la desigualdad entre varianzas llegando al caso extremo de $\sigma_1^2 = 1.6, \sigma_2^2 = 0.4$ donde solamente para $\Delta\mu = 0$ existen diferencias significativas entre los tests. En estos casos las potencias estimadas del test de Cochran y Cox son siempre menores que las potencias estimadas del test geodésico. Estas diferencias también se atenúan con el incremento progresivo del tamaño muestral. Esta tendencia queda reflejada en las tablas por la dispersión gradual del signo ‡. De nuevo, estas diferencias son más fáciles de detectar bajo el uso de la reducción de la varianza, esto es, cuando \tilde{p}_1 se usa en lugar de la frecuencia relativa \hat{p}_1 , aunque al aumentar la desigualdad entre varianzas estas diferencias tienden a desaparecer. Se puede observar que cuando el tamaño muestral es $n_1 = n_2 = 31$ ninguna de las diferencias basadas en \hat{p}_1 es significativa, mientras que todas aquellas diferencias basadas en \tilde{p}_1 de la primera serie de simulaciones son significativas. Este fenómeno no se repite para la segunda serie de simulaciones que tienen unas varianzas más desiguales.

En la segunda serie de simulaciones se observa, tanto en el test geodésico como en el test de Cochran y Cox y tanto si utilizamos \hat{p}_1 como \tilde{p}_1 , que no afecta, por lo menos de manera ostensible, el incremento de las diferencias entre las varianzas en la variación de las potencias. Tal como se expone en las tablas y en

sus gráficas asociadas.

Un efecto remarcable que produce la utilización de la técnica de reducción de la varianza es que si el valor medio de la variable de control C , en este caso la potencia del test t de Student para las 5000 réplicas de simulación, es inferior al valor esperado (0.05, 0.55 o 0.92) entonces, se producirá un incremento del valor obtenido en la variable respuesta Y , en nuestro caso el test geodésico o el test de Cochran y Cox, que será tanto mayor como sea la diferencia entre el valor medio y esperado de la variable de control C . Y de manera semejante ocurrirá en el caso contrario. Por tanto, los errores de muestreo para la obtención del valor medio de la variable de control servirán para corregir los correspondientes errores de muestreo en la variable de respuesta, influyendo en los cambios de valores de \hat{p}_1 a \tilde{p}_1 en los dos test de estudio.

En nuestras simulaciones, los valores medios de la potencia del test t de Student, \hat{p}_1 , y del margen de error asociado a un nivel de confianza del 95%, $\pm 1.96\hat{\sigma}_{ST}$, fueron los siguientes:

$n = 5000$		Valores de potencia teórica					
		0.05		0.55		0.92	
$n_1 = n_2$		\hat{p}_1	$\pm 1.96\hat{\sigma}_{ST}$	\hat{p}_1	$\pm 1.96\hat{\sigma}_{ST}$	\hat{p}_1	$\pm 1.96\hat{\sigma}_{ST}$
11		0.0488	0.00597	0.5546	0.01378	0.9238	0.00735
31		0.0498	0.00603	0.5678	0.01373	0.9242	0.00734
61		0.0482	0.00594	0.5492	0.01379	0.9172	0.00764
121		0.0534	0.00623	0.5474	0.01379	0.9146	0.00775

En todos ellos, se observa que el intervalo de confianza del 95% para la potencia de la t de Student incluye el valor teórico esperado, como era de suponer. Si nos fijamos en la potencia teórica 0.05 de la t de Student, vemos que todos los valores medios son menores de 0.05 excepto para el tamaño muestral 121. Por consiguiente, los valores de la potencia del test de estudio, tanto el geodésico como el de Cochran y Cox, usando el nuevo estimador (\tilde{p}_1) son mayores que sus respectivos valores en la potencia del test de estudio estimada como frecuencia relativa (\hat{p}_1) excepto para $n_1 = n_2 = 121$ que es menor. De la misma manera se puede razonar para las otras potencias teniendo en consideración el efecto corrector de la utilización de la técnica de reducción de la varianza.

De los párrafos anteriores se puede inferir que el nuevo (difícil de implementar) test geodésico es mejor que el test de Cochran y Cox, especialmente para muestras

pequeñas. Las diferencias se reducen tanto al aumentar la desigualdad entre varianzas, como al aumentar el tamaño muestral. La extensión del test geodésico es siempre igual o muy próxima al nivel de significación nominal y es más potente que el test de Cochran y Cox. Estos buenos resultados en el test geodésico son, de hecho, de limitado interés práctico. Nuestro estudio de simulación no es una comparación completa de las diferentes soluciones prácticas al problema de Behrens-Fisher. Muchas de estas soluciones dan mejor resultado que el test de Cochran y Cox (y se comportan de forma similar al test geodésico) y su uso está parcialmente limitado por su comparativamente mayor dificultad, la cual es extrema para el test geodésico.

Por otro lado, los resultados concernientes a la técnica de reducción de varianza son buenos cuando el grado de correlación con el test de control, en este caso, test *t* de Student, es alto. Así, cuando las diferencias entre las varianzas no es grande (primera serie de simulaciones) en todos los casos es de un valor superior al 75% para $\widehat{\%VR}$ y de un valor superior al 50% para $\widehat{\%MER}$. Además, la utilización de esta técnica supone un mínimo esfuerzo en la complejidad del diseño de la simulación.

FIG.1: ALGORITMO PARA EL ESTUDIO DE LA POTENCIA DEL TEST GEODÉSICO

Introducción de parámetros

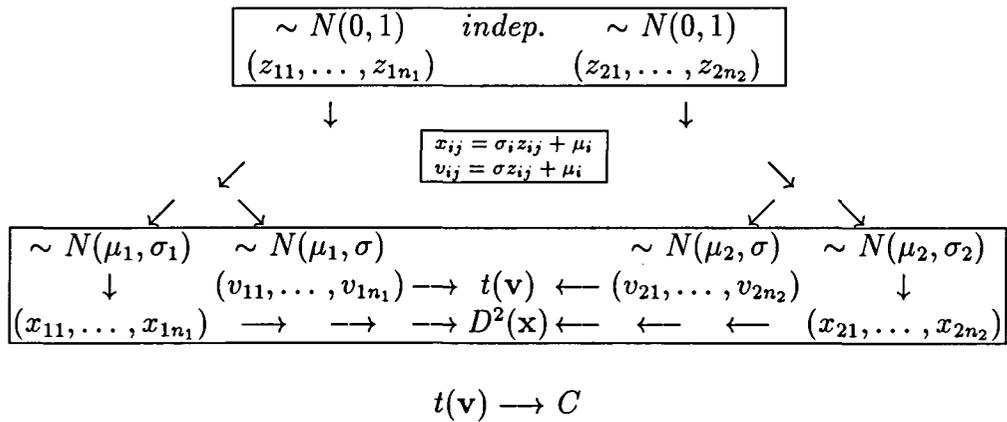
$$n_1, n_2, \sigma_1, \sigma_2, \beta_c, \alpha, n, B$$

↓

$$\begin{matrix} \mu_1 = 0 \\ \mu_2 = \phi_{\beta_c} \sigma \sqrt{\frac{2(n_1 + n_2)}{n_1 n_2}} \end{matrix} \quad \text{donde } \sigma^2 = \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n_1 + n_2} \text{ y } \phi_{\beta_c} \text{ es obtenido en el nomograma} \\ \text{que determina la función potencia de la t de Student}$$

Repite desde j=1 hasta n:

Generación de las muestras



Bootstrap paramétrico sobre x

$$\left. \begin{matrix} \sim N(\bar{x}, s_1) & \sim N(\bar{x}, s_2) \\ (x_{11}^*, \dots, x_{1n_1}^*)_1 & (x_{21}^*, \dots, x_{2n_2}^*)_1 \\ \dots & \dots \\ (x_{11}^*, \dots, x_{1n_1}^*)_B & (x_{21}^*, \dots, x_{2n_2}^*)_B \end{matrix} \right\} \rightarrow \begin{matrix} D_{*(1)}^2 \\ \dots \\ D_{*(B)}^2 \end{matrix}$$

$$\frac{\#[D_*^2 \geq D^2(\mathbf{x})] + 1}{B + 1} \geq \alpha \longrightarrow Y = 0$$

$$< \alpha \longrightarrow Y = 1$$

Fin Repite

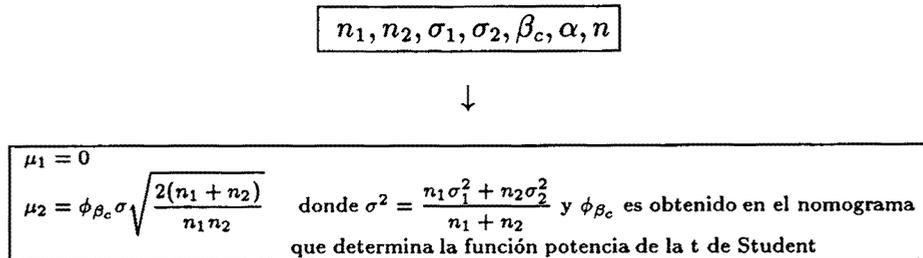
Del proceso anterior:

Reducción de la Varianza

$$\left. \begin{matrix} (Y_1, C_1) \\ \dots \\ (Y_n, C_n) \end{matrix} \right\} \longrightarrow (\tilde{p}_1, \hat{\sigma}^2, \%VR, \dots)$$

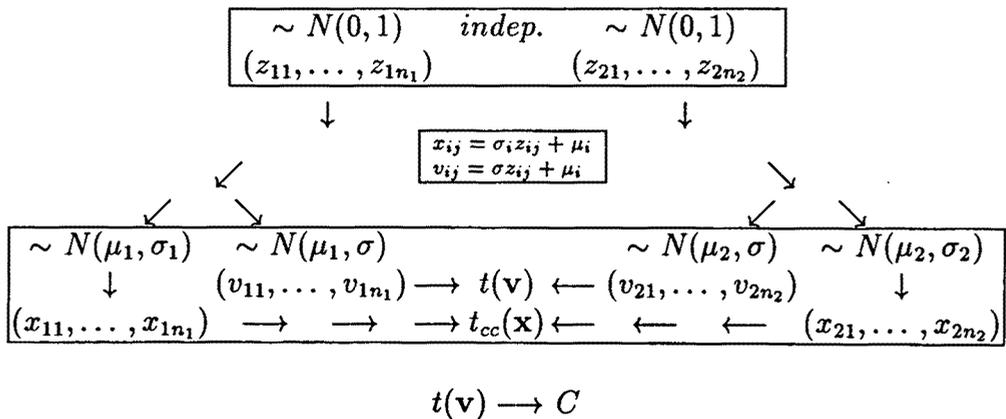
FIG.2: ALGORITMO PARA EL ESTUDIO DE LA POTENCIA DEL TEST DE COCHRAN Y COX

Introducción de parámetros



Repite desde j=1 hasta n:

Generación de las muestras



Test de Cochran y Cox

$$t_{cc}(X) = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(\hat{s}_1^2/n_1) + (\hat{s}_2^2/n_2)}}$$

$$t'_{\alpha/2} = \frac{t(\alpha/2, n_1 - 1)(\hat{s}_1^2/n_1) + t(\alpha/2, n_2 - 1)(\hat{s}_2^2/n_2)}{(\hat{s}_1^2/n_1) + (\hat{s}_2^2/n_2)}$$

$$\begin{aligned} |t_{cc}(X)| &\geq t'_{\alpha/2} \rightarrow Y = 1 \\ &< t'_{\alpha/2} \rightarrow Y = 0 \end{aligned}$$

Fin repite

Del proceso anterior:

Reducción de la Varianza

$$\left. \begin{matrix} (Y_1, C_1) \\ \dots \\ (Y_n, C_n) \end{matrix} \right\} \rightarrow (\tilde{p}_1, \hat{\sigma}^2, \%VR, \dots)$$