Elección Bayesiana del Tamaño de Muestra Óptimo

Mª Jesús Bayarri García y Mª Asunción Martínez Mayoral
Universitat de València y Universidad Miguel Hernández

Resumen

El paradigma Bayesiano introduce información adicional a la proporcionada por los datos en forma de distribuciones de probabilidad para las cantidades desconocidas. Esta práctica no está universalmente aceptada en la resolución de problemas de inferencia y de toma de decisiones. No obstante, en los problemas de elección de tamanño de muestra en un experimento futuro, puesto que obviamente los futuors datos de este experimento no están disponibles, resulta crucial la introducción de información adicional. En este articulo argumentamos que la forma mas natural, intuitiva y eficiente de hacerlo es mediante distribuciones de probabilidad que incorporen la incertidumbre acerca de los valores desconocidos de los parámetros, es decir, utilizando la aproximación bayesiana. Esta aproximación puede utilizarse tanto si el análisis estadístico a realizar es frecuentista o bayesiano. Proporcionamos ajemplos de elección de tamaño de muestra óptimo para ambos tipos de análisis. Asimismo, planteamos la elección como un problema de decisión.

Keywords: Diseño de experimentos; Distribuciones a posteriori o posteriores; Distribuciones a priori o previas; Distribuciones predictivas; Funciones de pérdida; Teoría de la decisión.

1 Introducción

La forma habitual de determinar el tamaño muestral óptimo (TMO) de un estudio estadístico futuro consiste, bien en fijar un nivel de precisión a conseguir en las inferencias, bien en maximizar alguna función objetivo que refleja la utilidad de resolver satisfactoriamente el análisis. En cualquier caso, el TMO resultante suele venir expresado en términos de parámetros desconocidos.

Tradicionalmente, el diseñador utiliza su información previa (subjetiva, basada en experimentos previos, reflejando un deseo de actuar conservadoramente, etc.) para asignar valores específicos a estos parámetros. El TMO se calcula entonces como si estas asignaciones fueran los verdaderos valores de los parámetros desconocidos. La considerable incertidumbre que suele existir en la asignación de estos valores no suele incorporarse.

La Estadística Bayesiana, permite al planificador, si así lo desea, evitar comprometerse con una asignación tan precisa: en vez de utilizar la información inicial en la asignación de un solo valor para los parámetros, se propone cuantificar dicha información en términos de una distribución de probabilidad, la *distribución previa* o *a priori*. Esta distribución incorpora de forma natural la incertidumbre presente en el problema. A partir de la distribución previa, y simplemente utilizando herramientas simples del cálculo de probabilidad, es fácil obtener la *distribución predictiva* de los resultados que se espera obtener en el experimento futuro. Esta distribución refleja toda la información, y toda la incertidumbre, acerca de los resultados del experimento futuro, y es clave en la determinación del TMO en particular, y en el diseño de experimentos en general.

El TMO de un estudio en planificación dependerá siempre del objetivo del estudio y del tipo de análisis que se va a efectuar. En este trabajo presentamos diversas formas de obtener el TMO desde una perspectiva Bayesiana, y bajo dos enfoques alternativos: i) inferencial, en el que el TMO pretende resolver satisfactoriamente algun tipo de inferencia, realizada tanto utilizando técnicas frecuentistas, como bayesianas; ii) planteando la búsqueda del TMO como un problema de decisión, incorporando también una función de utilidad (o pérdida) adecuada. Es importante remarcar que el diseño puede ser bayesiano aun cuando el análisis a realizar en el experimento que se planifica vaya a ser frecuentista.

Como es habitual, suponemos que en el estudio en planificación se va a observar $\boldsymbol{X} = (X_1, \dots, X_n)$, cuya distribución conjunta, $f(\boldsymbol{x}|\theta)$, depende de un parámetro desconocido θ sobre el que se desea inferir. Frecuentemente la distribución de $\boldsymbol{X}$ depende tambiém de otros parámetros marginales, en cuyo caso θ denota al vector paramétrico que identifica f. Suponemos, además, que la información disponible sobre θ (información de expertos o de estudios anteriores) está expresada en términos de una distribución previa $\pi(\theta)$. La costumbre habitual en los análisis frecuentistas de realizar los cálculos en un valor específico de θ sería un caso particular en que $\pi(\theta)$ es degenerada.

Verdinelli (1992) y Chaloner y Verdinelli (1995) ofrecen una revisión de diseño bayesiano. Smeeton y Adcock (1997), y Adcock (1997) hacen una buena revisión del tema de tamaños muestrales óptimos inferenciales desde el punto de vista bayesiano. Spiegelhalter y Freedman (1986) plantean la incorporación de información previa para la determinación del TMO y la predicción de la decisión a tomar en un ensayo clínico respecto a la asignación de tratamientos a pacientes. Joseph y Bélisle (1997) obtienen el TMO bayesiano en el problema de inferencia sobre medias normales y diferencias de medias normales, y Joseph, Wolfson y du Berger (1995) lo hacen para proporciones, utilizando intervalos de densidad posterior máxima. Joseph, du Berger y Bélisle (1997) resumen varias aproximaciones bayesianas y mezclas de bayesianas y verosimilitudistas para la determinación del TMO.

Este trabajo se organiza en 6 secciones, esta Introducción es la primera. En las Secciones 2 y 3 hacemos una introducción muy breve a los métodos inferenciales bayesianos y su resolución mediante métodos Monte Carlo. La Sección 4 trata de la determinación del TMO como problema sólo inferencial, y la Sección 5 como problema de decisión. Por último, proporcionamos algunas conclusiones y generalizaciones en la Sección 6.

2 Proceso de Aprendizaje Bayesiano

El proceso de aprendizaje bayesiano es muy sencillo: basta utilizar el cálculo de probabilidades. La idea básica es que toda la incertidumbre se cuantifica mediante distribuciones de probabilidad, incluida la incertidumbre sobre θ. Como consecuencia tiene sentido hablar de $Pr(H_0)$, en un contraste de hipótesis, de $Pr(a < \mu < b)$ en una estimación por intervalos, ... , etc., convirtiéndose las inferencias en meras descripciones de la distribución sobre θ. Los pasos fundamentales en cualquier proceso inferencial bayesiano son *i)* determinar una distribución conjunta de observables, $\boldsymbol{X}$, y no observables, θ, y *ii)* condicionar a lo que se observa ($\boldsymbol{x}$) e integrar lo que no se observa y no es de interés. La herramienta básica para condicionar es el teorema de Bayes, y para integrar en modelos complejos, las técnicas Monte Carlo.

En general, a las variables que se observan en el experimento las denotaremos por $\boldsymbol{X}$ y las cantidades desconocidas por θ, que pueden ser parámetros de un modelo, datos faltantes, sucesos que no observamos directamente o exactamente, etc. Al igual que en cualquier tipo de inferencia, se relacionan los observables con los no observables mediante un modelo $f(\mathbf{x}|\,\theta)$. Generalmente, $\mathbf{X} = (X_1, X_2, \ldots , X_n)$, y las X_i se suponen (condicionalmente) i.i.d $X_i|\,\theta \sim f(\cdot|\theta)$, y, por tanto, la *función* *de verosimilitud* para los datos observados $\boldsymbol{x}$ es

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta).$$

En la inferencia bayesiana, es preciso además cuantificar la incertidumbre (información) sobre θ **antes** de observar $\mathbf{x}$. Esta es la distribución *a priori* o *previa*

$$\theta \sim \pi(\theta).$$

(Estas distribuciones también pueden elegirse de forma que no se incorpore información adicional. La inferencia resultante suele denominarse *'inferencia bayesiana objetiva'.*)

Una vez observados los datos $\mathbf{X} = \mathbf{x}$, $\mathbf{x}$ es *conocido* por lo que tiene sentido condicionar al valor observado, dando lugar a la distribución *a posteriori* (o *posterior*), $\pi(\theta\,|\mathbf{x})$. Esta distribución cuantifica la incertidumbre (información) sobre θ *despues* de observar $\mathbf{x}$, y se calcula a partir del Teorema de Bayes:

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

Generalmente no es preciso calcular el denominador, puesto que sólo es una constante normalizadora.

La *predicción* de observables futuros es asimismo directa en el paradigma bayesiano: antes de observar los datos $\mathbf{x}$,

$$m(\mathbf{x}) = \int p(\mathbf{x}, \theta)d\theta = \int f(\mathbf{x}|\theta)\; \pi(\theta)d\theta,$$

es la predictiva *a priori*, herramienta crucial en el diseño y crítica de modelos. Nótese que $m(\boldsymbol{x})$ es el *promedio* de $f(\mathbf{x}|\theta)$, ponderado según $\pi(\theta)$ (las regiones más probables de valores de θ reciben mayor peso). Contrástese con el procedimiento de utilizar un valor específico para θ, resultando en $f(\mathbf{x}|\theta_0)$.

Una vez observado $\mathbf{x}$, si se desea predecir un futuro observable Z (observación, muestra, estadístico de una muestra futura, ... etc.), la distribución relevante es la predictiva *a posteriori*:

$$m(z|\,\mathbf{x}) = \int p(z, \theta|\,\mathbf{x})d\theta = \int f(z|\,\mathbf{x}, \theta)\; \pi(\theta|\,\mathbf{x})d\theta \overset{\text{ind}}{=} \int f(z|\theta)\; \pi(\theta|\mathbf{x})d\theta.$$

De nuevo debe remarcarse que las predicciones a posteriori bayesianas están basadas en el promedio ponderado de $f(\mathbf{x}|\theta)$, con los valores de θ con mayor densidad a posteriori recibiendo un peso mayor; La predicción frecuentista está basada a menudo simplemente en $f(\mathbf{x}|\hat{\theta})$, como si $\hat{\theta}$, la estimación de θ, fuera con toda certidumbre el valor desconocido de θ.

El tratamiento de los parámetros marginales es tambien muy simple: basta, de nuevo, con utilizar el cálculo de probabilidades e integrarlos de la distribución conjunta. Este tratamiento incorpora de forma natural la incertidumbre existente en la estimación de dichos parámetros.

En general, si $\boldsymbol{\theta = (\theta_1, \theta_2)}$ con $\boldsymbol{\theta_1}$ de interés ($\boldsymbol{\theta_2}$ marginal), el procedimiento bayesiano se limita a *i)* asignar una distribución previa conjunta, $\pi(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \pi(\boldsymbol{\theta_1}|\boldsymbol{\theta_2})\ \pi(\boldsymbol{\theta_2})$; *ii)* calcular la posterior conjunta correspondiente : $\pi(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}|\mathbf{x}) \propto f(\mathbf{x}|\ \boldsymbol{\theta_1}, \boldsymbol{\theta_2})\pi(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}\ )$, y *iii)* obtener la correspondiente distribución posterior marginal:

$$\pi(\boldsymbol{\theta_1}|\mathbf{x}) = \int \pi(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}|\mathbf{x}) = \int \pi(\boldsymbol{\theta_1}|\boldsymbol{\theta_2}, \mathbf{x})\pi(\boldsymbol{\theta_2}|\mathbf{x})\ d\boldsymbol{\theta_2}. \tag{1}$$

Nótese que, de nuevo, $\pi(\boldsymbol{\theta_1}|\mathbf{x})$ es el promedio ponderado de las distribuciones condicionales $\pi(\boldsymbol{\theta_1}|\boldsymbol{\theta_2}, \mathbf{x})$, recibiendo mas peso aquellas con mayor $\pi(\boldsymbol{\theta_2}|\mathbf{x})$; Contrástese con la simple utilización de $\pi(\boldsymbol{\theta_1}|\hat{\boldsymbol{\theta_2}}, \mathbf{x}))$, para alguna estimación $\hat{\boldsymbol{\theta_2}}$ de $\boldsymbol{\theta_2}$ que ignoraría la incertidumbre en la estimación de $\boldsymbol{\theta_2}$

La expresión (1) proporciona además una forma muy conveniente para la obtención de la marginal posterior mediante simulación Monte Carlo que tratamos a continuación.

3 Métodos MC y MCMC

Como hemos visto en la Sección anterior, la inferencia Bayesiana es metodológicamente muy sencilla. Los problemas aparecen en la implementación práctica, en particular, en la resolución de las numerosas integrales involucradas. En efecto, si la distribución a posteriori es

$$\theta|\ \mathbf{x} \sim \pi(\theta|\ \mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta) \propto p^*(\theta),$$

a partir de $p^*(\theta)$ (sin normalizar) hay que obtener:

Momentos a posteriori:

$$E[\theta^k] = \int \theta^k \pi(\theta|\mathbf{x}) d\theta = \frac{\int \theta^k f(\mathbf{x}|\theta)\pi(\theta) d\theta}{\int f(\mathbf{x}|\theta)\pi(\theta) d\theta} = \frac{\int \theta^k p^*(\theta) d\theta}{\int p^*(\theta) d\theta}$$

Probabilidades de intervalos:

$$Pr(a < \theta < b) = \int_a^b \pi(\theta|\mathbf{x}) d\theta = \int I_{(a,b)}(\theta)\pi(\theta|\mathbf{x}) d\theta = E[I_{(a,b)}(\theta)],$$

donde $I_{(a,b)}(\theta) = \begin{cases} 1 & \text{si} \quad \theta \in (a,b) \\ 0 & \text{si} \quad \theta \notin (a,b) \end{cases}$

Distribuciones predictivas:

$$m(z|\mathbf{x}) = \int f(z|\mathbf{x},\theta)\pi(\theta|\mathbf{x}) d\theta = E[f(z|\mathbf{x},\theta)]$$

y sus momentos, probabilidades, etc. asociadas.

Distribuciones marginales: si $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ necesitamos obtener, además de las cantidades previas, las distribuciones marginales

$$\pi(\theta_1) = \int \ldots \int \pi(\boldsymbol{\theta}|\mathbf{x}) d\theta_2, \ldots, d\theta_k = \int I_{(\theta_1,\cdot)}(\boldsymbol{\theta}) d\boldsymbol{\theta} = E[I_{(\theta_1,\cdot)}(\boldsymbol{\theta})]$$

y sus correspondientes momentos, probabilidades, etc.

Todas las integrales que necesitamos para llevar a cabo las inferencias deseadas pueden, pues, expresarse como el cálculo de esperanzas. Estas esperanzas no son generalmente analíticas (salvo en casos sencillos de modelos simples). Además, en la práctica la dimensión de θ puede ser muy grande, por lo que integración por métodos numéricos es también inviable. Acudimos entonces a resolver estas esperanzas mediante simulación, lo que genericamente se denominan *Métodos Monte-Carlo* (o MC).

La idea fundamental consiste en generar una muestra aleatoria $\theta^{(l)}$, $l = 1, \ldots, L$ de la distribución a posteriori $\pi(\theta|\mathbf{x})$ de interés, y utilizar dicha muestra como sustitutivo de $\pi(\theta|\mathbf{x})$. Por tanto, aproximamos las esperanzas como:

$$\int g(\theta)\pi(\theta|\mathbf{x}) \approx \frac{1}{L}\sum_{l=1}^{L} g(\theta^{(l)}).$$

Asimismo, aproximamos la densidad mediante el histograma de los valores generados (suavizado, si se desea), los momentos mediante los correspondientes momentos muestrales de los $\theta^{(l)}$'s, las distribuciones y momentos de funciones $g(\theta)$ a partir de $g(\theta^{(l)}),\ l = 1, \ldots, L$ que es m.a. de la distribución de $g(\theta)$, etc.

Si tenemos un vector paramétrico de dimensión n, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \ldots, \theta_n)$, y se desea generar de $\pi(\boldsymbol{\theta}|\mathbf{x})$ una posible estragia consiste en generar

$$\begin{array}{rcl} \theta_n & \sim & \pi(\theta_n|\mathbf{x}) \\ \theta_{n-1} & \sim & \pi(\theta_{n-1}|\theta_n, \mathbf{x}) \\ & \vdots & \\ \theta_2 & \sim & \pi(\theta_2|\theta_3, \ldots, \theta_n, \mathbf{x}) \\ \theta_1 & \sim & \pi(\theta_1|\theta_2, \theta_3, \ldots, \theta_n, \mathbf{x}) \end{array}$$

o bien utilizar el mismo procedimiento con 'bloques' de θ_i's. A este procedimiento se le conoce como *método de sustitución*, y una vez realizado L veces:

– $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(L)}$ es una m.a. de $\pi(\boldsymbol{\theta}|\,\mathbf{x})$

– $\theta_i^{(1)}, \theta_i^{(2)}, \ldots, \theta_i^{(L)}$ es una m.a. de $\pi(\theta_i|\,\mathbf{x})$

– $g(\boldsymbol{\theta}^{(1)}),\ g(\boldsymbol{\theta}^{(2)}), \ldots,\ g(\boldsymbol{\theta}^{(L)})$ es una m.a. de $\pi(g(\boldsymbol{\theta})|\,\mathbf{x})$

Si en vez de inferencias sobre θ lo que queremos es predecir un futuro $\mathbf{z}$, el procedimiento es virualmente idéntico: generar de $p(\boldsymbol{\theta}, \mathbf{z}|\mathbf{x})$, y retener para las predicciones las generaciones de $\mathbf{z}$. Para ello, basta generar *i)* $\boldsymbol{\theta}^{(l)}$ de $\pi(\boldsymbol{\theta}|\mathbf{x})$ y *ii)* $\mathbf{z}$ de $f(\mathbf{z}|\boldsymbol{\theta}^{(l)})$. Además, la generación de $f(\mathbf{z}|\boldsymbol{\theta}\ )$ es muy fácil cuando, como es habitual, los z_i son condicionalmente independientes dado $\boldsymbol{\theta}$, en cuyo caso basta generar $z_i \sim f_i(z|\ \boldsymbol{\theta})$.

Existen multitud de de técnicas llevar a cabo todas estas generaciones: CDF inversa, aceptación-rechazo, cociente de uniformes, muestreo 'importante', bootstrap bayesiano (SIR) (ver, por ejemplo, Ripley, 1987, y Tanner, 1993). En la mayoría de ellas, basta con disponer de la expresión de densidad sin normalizar (es decir, proporcional a verosimilitud $\times$ previa) para conseguir una muestra de la densidad (normalizada). Asimismo, si no se pueden obtener generaciones directamente de la distribución objetivo, puede muestrearse de *otra* distribución más sencilla y 'reconvertirla' en una muestra de la distribución objetivo

Los métodos MC son muy potentes, pero a veces insuficientes para la complejidad de algunos modelos que se utilizan en las aplicaciones prácticas. Se recurre entonces a los *métodos MCMC*, que pueden aplicarse a virtualmente *cualquier* problema, de *cualquier* complejidad. De hecho, los métodos MCMC pueden considerarse los 'responsables' del increible 'boom' que los métodos bayesianos han experimentado en los últimos años.

Cuando generar m.a. de $\pi(\theta|\mathbf{x})$ es 'imposible' o muy dificultoso, los *métodos* **M***onte*-**C***arlo con* **C***adenas de* **M***arkov (MCMC)* proponen generar una cadena de Markov $\theta^{(l)},\ l = 1, \ldots, L$ cuya distribución estacionaria sea $\pi(\theta|\mathbf{x})$. Entonces, despues de un 'quemado inicial', se considera $\theta^{(l)},\ l = 1, \ldots, L$ como una muestra, *correlada* en general, (a veces con correlación muy alta) de $\pi(\theta|\mathbf{x})$, pero el Teorema Central del Límite sigue siendo válido en esta muestra correlada, por lo que podemos utilizar las mismas aproximaciones que utilizabamos en los métodos MC para aproximar las esperanzas requeridas.

Existen varios métodos para obtener cadenas de Markov convergentes a $\pi(\theta|\mathbf{x})$. Los más populares son el método de *Metrópolis-Hastings* y el *Muestreo de Gibbs*. Este último es especialmente sencillo de implementar, puesto que basta con generar de las condicionales (univariantes) completas $p(\theta_i \mid \text{todo lo demás})$. Además, estas condicionales completas son *muy* sencillas de obtener, puesto que son simplemente proporcionales a la distribución conjunta, es decir, al producto de la verosimilitud por la previa.

Tanto las generaciones MC como las MCMC puede simplificarse a veces utilizando variables auxiliares (latentes), lo que suele denominarse *data augmentation*.

Una buena introducción a los métodos MCMC para el análisis bayesiano, junto con interesantes aplicaciones puede encontrarse en Gilks et al. (1995). Además, existe software de uso público para llevarlo a cabo (BUGS: **B**ayesian analysis **U**sing **G**ibbs **S**ampling).

4 Tamaño Muestral Óptimo como Problema Inferencial.

Presentamos aquí diversas aproximaciones para el cálculo del TMO cuando el objetivo del estudio en planificación es el de resolver las inferencias de interés con un nivel satisfactorio de precisión. Proponemos el diseño bayesiano para obtener el TMO ante dos posibilidades básicas respecto al tipo de análisis a realizar:

1. Un *análisis frecuentista* de estimación o contraste de hipótesis;

2. Un *análisis bayesiano* de estimación, contraste de hipótesis o de predicción de resultados.

4.1 Diseño Bayesiano, Análisis Frecuentista

Estimación.- Si el objetivo es estimar θ mediante el estimador $\hat{\theta}$ con suficiente precisión, la solución clásica pasa por fijar un error tolerable, d, para la estimación, d. Se busca n tal que:

$$q_\alpha \; s.e.(\hat{\theta}) \leq d, \tag{2}$$

donde q_α es el cuantil adecuado, basado en la distribución del estimador $\hat{\theta}$ (cuando $2d$ es la longitud de un intervalo de confianza aproximado para θ), o bien $q_\alpha = 1$ (cuando d es una cota superior del error standard)

Contraste de Hipótesis.- Si se pretende resolver un contraste de hipótesis sobre θ a un nivel de confianza $1-\alpha$ y garantizando una potencia superior a $1-\beta$, el tamaño muestral se consigue resolviendo:

$$Potencia(\theta) \;=\; Pr(\text{ rechazar } H_0 \text{ a nivel } \alpha \mid \theta, \theta \in H_1) \geq 1-\beta. \tag{3}$$

Cualquiera de los requisitos previos sobre n suelen expresarse finalmente de la forma:

$$n \geq g(\theta), \tag{4}$$

donde $g(\cdot)$ es una función que con frecuencia depende del valor desconocido del parámetro (y a veces incluso de otros parámetros del modelo). Así, para determinar el n óptimo, el planificador frecuentista tiene que asignar un valor especiífico de θ, a menudo con grandes dosis de arbitrariedad. Generalmente, utiliza:

– un valor esperado o "conjetura", que suele ocasionar TMO a veces demasiado pequeños, puesto que ignora la incertidumbre sobre dichos valores, o bien

– un valor conservador, que suele dar lugar a TMO demasiado grandes debido a que prescinde de todo tipo de información.

La solución Bayesiana consiste en evitar el compromiso con un valor concreto de θ, incorporando la información disponible y la incertidumbre que existe sobre ella, todo ello contenido en la distribución de probabilidad previa para θ, $\pi(\theta)$. Algunas posibilidades son:

Bayes 1 $\rightsquigarrow$ promediar $g(\theta)$ y tomar el menor n tal que

$$n \geq E^{\theta}[g(\theta)]. \tag{5}$$

Bayes 2 $\rightsquigarrow$ calcular la probabilidad de (4) y tomar el menor n tal que garantiza que dicha probabilidad es suficientemente alta, esto es,

$$Pr^{\theta}[n \geq g(\theta)] \geq \rho, \tag{6}$$

con $\rho \in (0, 1)$ suficientemente grande.

Bayes 3 $\rightsquigarrow$ en el caso específico del contraste, tomar el menor n que proporcione una potencia "esperada" superior a $1 - \beta$

$$E^{\theta}[potencia] \geq 1 - \beta. \tag{7}$$

Cuando estamos planteando un contraste sobre θ, puesto que el n óptimo lo estamos planteando para obtener una potencia específica, los valores esperados y probabilidades se calculan respecto de la distribución de θ bajo H_1. En el caso de estimación, las probabilidades y esperanzas se calculan respecto de la (única) distribución previa $\pi(\theta)$ asumida.

Con frecuencia estos cálculos no resultan analíticos o son excesivamente complejos, de modo que es aconsejable utilizar simulaciones $\theta^{(i)} \sim \pi(\theta),\ i = 1, \ldots, M$, y aproximaciones Monte Carlo de las cantidades de interés (promedios y probabilidades).

Lee y Zelen (2000) dan una formulación *global* alternativa para análisis frecuentistas, con un diseño *aún* más bayesiano. Básicamente defienden que la asignación de α y $1 - \beta$ en un contraste no resulta muy intuitivo, y que es mejor asignar $Pr(H_0|\text{rech. } H_0)$, $Pr(H_1|\text{acep. } H_0)$, y $Pr(H_0)$, incluso para investigadores frecuentistas. De dichas asignaciones *deducen* entonces los valores apropiados de α y $1-\beta$. Finalmente, concluyen su *'rejoinder'* con:

> *"There are some statistical applications that can only be carried out using Bayesian ideas. In our opinion, this is one of them. It is not an issue of faith or adherence to Bayesian ideas. It is simply doing the right thing"*

Ejemplo 1 (Inferencia sobre una proporción. Datos Bernoulli). *Supongamos que va a observarse una m.a. Bernoulli de parámetro θ, esto es,*

$$X_1, \ldots, X_n \overset{iid}{\sim} Br(\theta).$$

Supongamos además que la información disponible sobre θ se cuantifica mediante una distribución (conjugada) Beta,

$$\theta \sim Be(a,b), \tag{8}$$

donde a y b se eligen para representar adecuadamente la información de que se dispone. En este caso, también es posible llevar a cabo un análisis 'conservador', para lo que basta utilizar $Be(1,1)$ (es decir, la uniforme) como distribución a priori "mínimo–informativa" para θ.

Planteamos como posibles objetivos de un análisis frecuentista sobre θ:

Estimación.- *Si queremos que el intervalo al nivel $1-\alpha$ tenga longitud inferior a $2d$ y utilizando como estimador la proporción muestral $\hat{\pi}$:*

$$\hat{\pi} = \sum_i X_i/n \;\dot{\sim}\; N\left(\theta, \frac{\theta(1-\theta)}{n}\right),$$

el TMO frecuentista se obtiene resolviendo:

$$n \geq \left(\frac{z_{1-\alpha/2}}{d}\right)^2 \theta(1-\theta) = g_e(\theta). \tag{9}$$

Contraste de hipótesis.- *Se plantea un contraste de hipótesis unilateral sobre θ a nivel $1-\alpha$ y potencia $1-\beta$,*

$$\begin{aligned} H_0 &: \theta \leq \theta_0 \\ H_1 &: \theta > \theta_0. \end{aligned} \tag{10}$$

La función de potencia para un n fijo es:

$$\begin{aligned} \Phi_n(\theta;\alpha) &\equiv Pr(rechazar\ H_0|\ \theta, \theta \in H_1) \\ &= Pr(\hat{\pi} - \theta_0 > z_{1-\alpha}\ \sqrt{\theta(1-\theta)/n}|\ \theta, \theta \in H_1) \\ &= \Phi\left(\frac{z_\alpha\sqrt{\theta_0(1-\theta_0)} + (\theta-\theta_0)\sqrt{n}}{\sqrt{\theta(1-\theta)}}\right), \end{aligned} \tag{11}$$

ya que $\hat{\theta} \sim N(\theta, \theta(1-\theta)/n)$, con $\theta > \theta_0$ bajo H_1.

El TMO frecuentista se obtiene resolviedno

$$n \geq \left(\frac{z_{1-\alpha}\sqrt{\theta_0(1-\theta_0)} - z_\beta\sqrt{\theta(1-\theta)}}{\theta-\theta_0}\right)^2 = g_c(\theta). \tag{12}$$

para algún valor específico de θ.

Utilizando la información previa sobre θ, en (8)*, determinamos el TMO desde una perspectiva Bayesiana según* (5), (6) *y* (7)*. Para ello hay que evaluar valores esperados y probabilidades de las funciones* (9), (11) *y* (12)*. Una forma rápida de resolverlo es a través de simulaciones y $\theta^{(j)}, j = 1, \dots, M$ de la distribución previa* (8) *y aproximaciones Monte Carlo:*

Bayes 1 $\rightsquigarrow n \geq \sum_{j=1}^{M} g(\theta^{(j)})/M \approx E[g(\theta)]$

Bayes 2 $\rightsquigarrow$ *para un grid de valores de n, calcular*

$$Pr[g(\theta) \leq n] \approx \widehat{P}_n = \frac{\#\{g(\theta^{(j)}) \leq n\}}{M}$$

y tomar el menor n tal que la probabilidad $\widehat{P}_n$ es igual o mayor que un ρ fijado suficientemente alto.

Bayes 3 $\rightsquigarrow$ *para un grid de valores de n, aproximar $E[potencia]$ y considerar el menor n que satisfaga:*

$$E[\Phi_n(\theta; \alpha)] \approx \sum_{j=1}^{M} \Phi_n(\theta^{(j)}; \alpha)/M \geq 1 - \beta.$$

Ejemplo 2 (Contraste sobre una media. Datos Normales). *Consideremos datos normales i.i.d., con varianza conocida σ^2,*

$$X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

con el propósito de contrastar

$$\begin{aligned} H_0 &: \mu \leq 0 \\ H_1 &: \mu > 0. \end{aligned}$$

(Ver Carlin y Louis (1985), Louis y Bailey (1990) y Shih (1995), para modelos más complejos.)

Suponemos que la información previa sobre μ se cuantifica mediante una distribución (conjugada) normal

$$\mu \sim N(m_0, v_0^2), \tag{13}$$

con media y varianza, m_0 y v_0^2, asignadas para reflejar la información sobre μ.

Vamos a obtener el TM de forma que $Pr(\textit{error tipo I}) = \alpha$ *(para* $\mu = 0$*) y potencia* $= 1 - \beta$*. Para el valor* μ *es fácil ver que* n *debe cumplir:*

$$\Phi_n(\mu;\alpha) \equiv \Phi\left(\frac{\sqrt{n}\,\mu}{\sigma} + z_\alpha\right) \geq 1 - \beta, \tag{14}$$

es decir,

$$n \geq \sigma^2\,\frac{(z_{1-\beta} - z_\alpha)^2}{\mu^2} = g(\mu). \tag{15}$$

La solución frecuentista sustituye μ *en la encuación (14) por un valor concreto, presumiblemente la mejor 'conjetura' sobre el valor de* μ*, es decir,* m_0 *en (13). Las soluciones bayesianas, por el contrario, tienen en cuenta la incertidumbre acerca del valor* m_0 *tal y como se cuantifica en (13). Para ello basta obtener simulaciones* $\mu^{(j)} \sim N(m_0, v_0^2)$*,* $j = 1, \ldots, M$*, en las que evaluar las funciones* $g(\mu)$ *en* (15) *y la de potencia en* (14) *y proponer:*

Bayes 1 $\rightsquigarrow n \geq \sum_{j=1}^{M} g(\mu^{(j)})/M \approx E[g(\mu)]$.

Bayes 2 $\rightsquigarrow$ *a partir de un grid de valores para* n*, elegir el menor tal que para un* ρ *dado, suficientemente alto, se consigue:*

$$Pr[g(\mu) \leq n] \approx \frac{\#\{g(\mu^{(j)}) \leq n\}}{M} \geq \rho.$$

Bayes 3 $\rightsquigarrow$ *para un grid de valores para* n*, elegir el menor tal que:*

$$E[\Phi_n(\mu;\alpha)] \approx \sum_{j=1}^{M} \Phi_n(\mu^{(j)};\alpha)/M \geq 1 - \beta.$$

4.2 Diseño bayesiano, análisis bayesiano

Supongamos ahora que se va a realizar un análisis bayesiano del estudio que se está planificando. Dicho anaálisis estará basado en la distribución posterior $\pi(\theta|\boldsymbol{x})$, obtenida tras observar el resultado $\boldsymbol{X} = \boldsymbol{x}$. Al elegir el TMO, obviamente el valor $\boldsymbol{X} = \boldsymbol{x}$ que se observará es desconocido, pero la información (e incertidumbre) sobre el mismo vienen cuantificada mediante distribución predictiva:

$$\text{distribución predictiva} \quad \rightsquigarrow \quad m(\boldsymbol{x}) = \int m(\boldsymbol{x}|\theta)\pi(\theta)d\theta. \tag{16}$$

Un objetivo natural de cualquier tipo de análisis consiste en obtener inferencias suficientemente precisas. Dependiendo del tipo de inferencias específicas a realizar, esto se expresa según:

4.2.1 Estimación.

Obtener una estimación de θ suficientemente precisa se puede expresar en términos de conseguir una varianza posterior de θ, suficientemente pequeña,

$$Var(\theta|\boldsymbol{X}) \leq C^2,$$

donde $\boldsymbol{X}$ es un vector aleatorio con distribución $m(\boldsymbol{x})$, lo que nos permite calcular el n tal que:

BB.Est1 $\rightsquigarrow$ proporciona un valor esperado de dicha varianza suficientemente pequeño,

$$E^{\boldsymbol{X}}[Var(\theta|\boldsymbol{X})] \leq d^*, \tag{17}$$

donde la esperanza es con respecto a $m(\boldsymbol{x})$.

BB.Est2 $\rightsquigarrow$ garantiza, con probabilidad alta, que dicha varianza resultará suficientemente pequeña,

$$Pr^{\boldsymbol{X}}[Var(\theta|\boldsymbol{X}) \leq d^*] \geq \rho, \tag{18}$$

donde de nuevo la probabilidad se calcula con la distribución $m(\boldsymbol{x})$.

Alternativamente, la precisión a posteriori se podría concretar en términos de la longitud de un intervalo creíble para θ, $L(\boldsymbol{X}) \leq C^2$, etc., y se plantearían soluciones análogas.

Ejemplo 3 (Estimación Bayesiana de una media. Datos normales). *Consideramos los datos normales del Ejemplo 2. Queremos elegir n que garantice una estimación bayesiana suficientemente precisa de la media μ.*

Suponemos que la información previa se cuantifica de nuevo mediante la distribución normal (13). *El estimador bayes habitual es la media de la distribución posterior $\pi(\mu|\boldsymbol{x})$ obtenida cuando se observa $\boldsymbol{X} = \boldsymbol{x}$,*

$$\pi(\mu|\boldsymbol{X}) = N(\mu|m(\boldsymbol{X}),\ v^2(\boldsymbol{X})) \tag{19}$$

donde

$$v^2(\boldsymbol{X}) = \frac{\sigma^2\ v_0^2}{\sigma^2 + n v_0^2},$$

depende de los datos a observar.

Requerir un valor de la varianza a posteriori suficientemente pequeño da lugar a un tamaño muestral n *de la forma:*

$$v^2 \leq C^2 \Leftrightarrow n \approx \left(\frac{\sigma}{C}\right)^2 \left[1 - \left(\frac{C}{v_0}\right)^2\right],$$

que en este caso no depende de μ *sino sólo de la incertidumbre a priori* (v_0^2) *y de la varianza del modelo* (σ^2)*. Puede notarse que, de acuerdo a la intuición, cuanto menor es la precisión que exigimos a la estimación (mayor* C^2*), menor tamaño muestral necesitamos, y cuanto más imprecisa es la información previa* ($v_0^2 \uparrow$)*, mayor* n *es necesario. De hecho, si* $v_0^2 \to \infty$*, entonces* $n \to (\frac{\sigma}{C})^2$*, que es el resultado que proporciona un diseño frecuentista.*

4.2.2 Contraste de Hipótesis.

Hay varias alternativas sobre qué exigir para resolver "satisfactoriamente" un contraste de hipótesis. Aquí nos centramos en resolver el contraste con suficiente precisión, esto es, obtener suficiente evidencia a favor de una u otra hipótesis. Si el contraste se resuelve en términos de las probabilidades a posteriori de cada hipótesis, exigiendo un determinado nivel de "evidencias" $1/B$ a favor de una hipótesis y en contra de la otra (con $B < 1$), tendremos resuelto el contraste, en un sentido u otro, si

BB.Cont $\rightsquigarrow$

$$Pr^{\theta|\boldsymbol{X}}(H_0) < B \cdot Pr^{\theta|\boldsymbol{X}}(H_1), \quad o \quad Pr^{\theta|\boldsymbol{X}}(H_1) < B \cdot Pr^{\theta|\boldsymbol{X}}(H_0),$$

es decir,

$$Pr^{\theta|\boldsymbol{X}}(H_0) < \frac{B}{1+B}, \quad o \quad Pr^{\theta|\boldsymbol{X}}(H_0) > \frac{1}{1+B},$$

donde las probabilidades están calculadas respecto de la distribución a posteriori de θ. Utilizando la distribución predictiva de $\boldsymbol{X}$, (16), podemos calcular el n que con probabilidad alta garantice la resolución del contraste a un nivel de evidencias $1/B$ suficientemente alto,

$$Pr^{\boldsymbol{X}}[\{rechazar\ H_0\ a\ nivel\ B\}\ o\ \{rechazar\ H_1\ a\ nivel\ B\}] \geq \rho$$

es decir,

$$Pr^{\boldsymbol{X}}[Pr^{\boldsymbol{\theta}|\boldsymbol{X}}(H_0) < B/(1+B)] + Pr^{\boldsymbol{X}}[Pr^{\boldsymbol{\theta}|\boldsymbol{X}}(H_0) > 1/(1+B)] \geq \rho. \qquad (20)$$

Weiss (1997) presenta la determinación del tamaño muestral para contrastes resueltos en función de factores Bayes.

4.2.3 Predicción.

Si el objetivo del análisis es, una vez observada $\boldsymbol{X}$ la predicción de una futura observación, X^*, el cálculo del TMO estará basado en tratar de minimizar el "error" de predicción, dado en términos de la varianza de la distribución predictiva $m(X^*|\boldsymbol{X})$ (o análogamente en función de la longitud de un intervalo creíble),

$$Var(X^*|\boldsymbol{X}) \leq C^2,$$

para lo que podemos calcular el n tal que garantiza:

BB.Pred1 $\rightsquigarrow$ un valor esperado a la varianza predictiva de X^* suficientemente pequeño,

$$E^{\boldsymbol{X}}[Var(X^*|\boldsymbol{X})] \leq C^2. \qquad (21)$$

BB.Pred2 $\rightsquigarrow$ con probabilidad alta, una varianza suficientemente pequeña,

$$Pr^{\boldsymbol{X}}[Var(X^*|\boldsymbol{X}) \leq C^2] \geq \rho. \qquad (22)$$

De nuevo, en la determinación del TMO resultarán útiles la simulación y las aproximaciones numéricas Monte Carlo. Para simular posibles resultados $\boldsymbol{X}$ del estudio, basta con simular $\theta^{(i)} \sim \pi(\theta)$ y $\boldsymbol{X}^{(i)} \sim f(x|\theta^{(i)})$, $i = 1, \ldots, M$. Cada uno de esos posibles resultados $\boldsymbol{X}^{(i)}$ origina una distribución posterior distinta $\pi(\theta|\boldsymbol{X}^{(i)})$.

Ejemplo 4 (Análisis Bayesiano sobre una proporción. Datos Bernoulli). *El estudio en planificación consiste en n pruebas independientes Bernoulli de parámetro θ, o equivalentemente, en la realización de una prueba binomial de parámetros n y θ,*

$$Y|\theta \sim Bi(n, \theta).$$

Utilizando la distribución previa (8)*, una vez realizado el estudio en planificación, tendremos como distribución posterior:*

$$\theta|Y \sim Be(a + Y, b + n - Y). \tag{23}$$

La predicción de posibles resultados del experimento en planificación se obtiene con simulaciones:

1. $\theta^{(j)} \sim \pi(\theta)$ *de la distribución previa* (8)*,* $j = 1, \ldots, M$
2. $y^{(j)} \sim Bi(n, \theta^{(j)})$*,* $j = 1, \ldots, M$.

Estimación de θ. *Si se observa un resultado Y en el estudio, la varianza posterior de θ es*

$$Var(\theta|Y) = \frac{(a + Y)(b + n - Y)}{(a + b + n)(a + b + n + 1)}. \tag{24}$$

La predicción de $Var(\theta|Y)$ para un n fijo se consigue evaluando (24) *en las predicciones $y^{(j)}$, $j = 1, \ldots, M$. El TMO lo obtenemos aproximando, para un grid de valores de n, el valor esperado de $Var(\theta|Y)$ y la probabilidad de que resulte suficientemente pequeña; se toma finalmente el menor n para el que se satisface* (17) *o* (18)*:*

$$\text{BB.E}\textsc{st}1 \rightsquigarrow \sum_{j=1}^{M} Var(\theta|\, y^{(j)})/M \leq C^2$$

$$\text{BB.E}\textsc{st}2 \rightsquigarrow \frac{\#\{Var(\theta|\, y^{(j)}) \leq C^2\}}{M} \geq \rho.$$

Resolución del Contraste (10) *sobre θ, a un nivel de evidencias $1/B$. La probabilidad de H_0 (resp. H_1) se obtiene, para cada predicción $y^{(j)}$ del experimento, con:*

$$Pr(H_0|\, y^{(j)}) \approx \widehat{P_0}^{(j)} = \frac{\#\{\theta^{i_j} \leq \theta_0\}}{m_j},$$

donde θ^{i_j}, $i_j = 1, \ldots, m_j$, *son simulaciones de la distribución posterior* $\pi(\theta|\, y^{(j)})$, *en* (23).

La probabilidad de concluir en el contraste para cada n *a un nivel de evidencias* $1/B$ *se aproxima con:*

$$\widehat{Pr}(concluir) = \frac{\#\{\widehat{P_{0n}}^{(j)} < B/(1+B)\}}{M} + \frac{\#\{\widehat{P_{0n}}^{(j)} > 1/(1+B)\}}{M}$$

El TMO será aquel n *menor para el que*

$$\text{BB.C}{\small\text{ONT}} \rightsquigarrow \widehat{Pr}(concluir) \geq \rho.$$

Predicción de otro experimento Binomial. *La distribución de una nueva observación binomial de tamaño* r, Y_r, *es una Beta-Binomial de parámetros* (r, a^*, b^*). *Los parámetros* $a^*, b*$ *son los de la distribución posterior de* θ *en* (23), *una vez realizado el estudio en planificación y obtenidos* Y_n *éxitos en* n *pruebas.*

La varianza de Y_r *es entonces:*

$$Var(Y_r|\, Y_n) = \frac{a^*\, b^*}{(a^* + b^*)^2} = \frac{(a + Y_n)(b + n - Y_n)}{(a + b + n)^2}$$

Predecir dicha varianza consiste simplemente en evaluarla, para cada n, *sobre las predicciones* $y^{(j)}$, $j = 1, \ldots, M$. *De nuevo, se considera el menor* n *que satisface* (21) *o* (22)*:*

$$\text{BB.P}{\small\text{RED}}1 \rightsquigarrow \sum_j Var(Y_r|y^{(j)})/M \leq C^2$$

$$\text{BB.P}{\small\text{RED}}2 \rightsquigarrow \frac{\#\{Var(Y_{n+1}|y^{(j)}) \leq C^2\}}{M} \geq \rho.$$

5 Tamaño Muestral Óptimo como Problema de Decisión

Otra forma de tratar la obtención del TMO es planteándolo como un problema de decisión. El propósito es entonces, lograr ciertos objetivos inferenciales (estimar, resolver un contraste, etc.) minimizando las pérdidas esperadas asociadas a los errores de decisión y a los costes del muestreo.

Una buena presentación de Teoría de la Decisión se hace en Berger (1985), Carlin and Louis (2000) y Müller and Parmigiani (1995). Lindley (1997) plantea la maximización de la utilidad esperada para la determinación bayesiana del tamaño de muestra óptimo, y lo compara con otros procedimientos.

5.1 Elementos básicos del problema

Enfocados hacia la obtención del TMO, los elementos básicos del problema de decisión son:

- El espacio parámetrico Θ que contiene los posibles valores del parámetro desconocido θ del estudio; estos valores se denominan en Teoría de la Decisión, *'estados de la naturaleza'*.
- Datos experimentales $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ que proporcionan información sobre θ, esto es, cuya distribución de probabilidad depende del valor de θ, $\boldsymbol{X}|\,\theta \sim f(\boldsymbol{x}|\,\theta)$. $\mathcal{X}$ denota el conjunto de todos los posibles resultados, esto es, el *espacio muestral*.
- Un conjunto $\mathcal{A}$ de decisiones (o acciones) a. En estimación, una acción es un estimador; en un problema de contraste de hip'otesis, las acciones posibles son aceptar o rechazar H_0. En un problema de decisión estadístico, $a = a(\boldsymbol{X})$ es una función de los datos.
- Una *función de pérdida (o utilidad) global* $L(\theta, a, n)$ que cuantifica el coste de decidir a cuando el verdadero estado de la naturaleza es θ y se han tomado n unidades muestrales en el experimento. Es equivalente hablar en términos de pérdidas (L) o utilidades (U) ya que $L(\theta, a, n) = -U(\theta, a, n)$.

  En el problema de determinación del TMO, podemos suponer que las pérdidas ocasionadas en un estudio estadístico son debidas, por un lado a la gravedad de cometer un error al decidir (error estadístico), y por otro debido al coste del muestreo. Suponemos además una relación lineal, esto es,

  $$L(\theta, a, n) = L(\theta, a) + C(n).$$

Una vez presentados los elementos, planteamos cómo evaluar los costes asociados a las acciones en un problema inferencial para acabar determinando el tamaño muestral óptimo.

Cuando se ha observado una muestra de tamaño n, la información disponible sobre θ viene actualizada por los datos observados $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ a través de la distribución posterior $\pi(\theta|\,\boldsymbol{x})$. Podemos calcular entonces cuál es la *pérdida esperada* por tomar una decisión cualquiera a, calculando:

$$\rho(\pi(\theta|\boldsymbol{x}), a, n) = \int L(\theta, a, n)\pi(\theta|\boldsymbol{x})d\theta. \tag{25}$$

Si hemos decidido tomar una muestra de tamaño n y hemos observado $\boldsymbol{x}$, la acción óptima $a_n^*(\boldsymbol{x})$ es aquella que ocasiona una pérdida esperada mínima, esto es,

$$a_n^*(\boldsymbol{x}) \quad \text{tal que} \quad min_a \rho(\pi(\theta|\boldsymbol{x}), a, n). \tag{26}$$

Además, el *riesgo* asociado a considerar una muestra de tamaño n y observar $\boldsymbol{x}$ viene dado por la pérdida esperada de la decisión óptima, esto es:

$$r(\boldsymbol{x}, n) = min_a \rho(\pi(\theta|\boldsymbol{x}), a, n) = \rho(\pi(\theta|\boldsymbol{x}), a_n^*(\boldsymbol{x}), n). \tag{27}$$

Puesto que estamos tratando de determinar el TMO para un experimento futuro, su resultado lo desconocemos. Sin embargo lo podemos predecir en función de la información disponible, contenida en su distribución predictiva $m(\boldsymbol{x})$, en (16). Podemos así elegir el tamaño de muestra necesario para minimizar el *riesgo* que se espera de una planificación con tamaño n, $r(n)$, esto es,

$$\text{elegir } n^* \text{ tal que } min_n\ r(n)$$

donde

$$r(n) = E^{\boldsymbol{X}}[r(\boldsymbol{X}, n)] = \int r(\boldsymbol{x}, n)m(\boldsymbol{x})d\boldsymbol{x}. \tag{28}$$

Planificar en función del n que minimiza $r(n)$ es equivalente a hacerlo en términos de aquel que minimiza el riesgo Bayes de la *regla de decisión* óptima $a_n^*(\boldsymbol{x})$, $RB(n)$, definido como el valor esperado, respecto de $\pi(\theta)$, de la función de riesgo frecuentista $R(\theta, a_n^*(\boldsymbol{x}), n)$:

$$RB(n) = \int \underbrace{\int L(\theta, a_n^*(\boldsymbol{x}), n) f(\boldsymbol{x}|\theta)d\boldsymbol{x}}_{R(\theta, a_n^*, n)}\,\pi(\theta)d\theta.$$

5.2 Funciones de Pérdida

Existen diversas funciones de pérdida $L(\theta, a)$ estándar para los distintos problemas de inferencia estadística. En principio, la opción por una determinada función de pérdida depende del tipo de problema a resolver (estimación, contraste), y de la valoración que se hace de la gravedad de los errores que se pueden cometer al resolver estadísticamente el problema. Para completar la función de pérdida global en un problema de planificación, $L(\theta, a, n)$, será necesario además, establecer los costes que ocasiona el muestreo, $C(n)$.

Funciones de pérdida $L(\theta, a)$ habituales son:

5.2.1 Pérdida Cuadrática.

Se suele utilizar en estimación y es de la forma:

$$L(\theta, a) = (\theta - a)^2 \tag{29}$$

El riesgo frecuentista $R(\theta, a) = \int [\theta - a(x)]^2 f(\boldsymbol{x}|\theta) d\boldsymbol{x}$ representa la varianza del estimador cuando es insesgado; minimizar dicho riesgo es pues, conseguir el estimador insesgado de mínima varianza.

Hay diversas variaciones de esta pérdida basadas en ponderaciones y traslaciones, por ejemplo:

$$L_b(\theta, a) = (\theta - a - b)^2; \quad L_w(\theta, a) = w(\theta)(\theta - a)^2; \; \ldots$$

Ejemplo 5 (Estimación con datos normales. Pérdida cuadrática). *Supongamos de nuevo, datos normales i.i.d. que se van a observar en el estudio a planificar,*

$$X_1, X_2, \ldots, X_n; \quad X_i \sim N(\mu, \sigma^2) \;\; \textit{para } i = 1, \ldots, n,$$

y el objetivo de estimar μ. El espacio de acciones $\mathcal{A}$ está constituido por todas las posibles estimaciones para μ, esto es, $\mathcal{A} = \mathcal{R}$.

Considerando como distribución previa la normal (13), *las inferencias estarán basadas en la distribución posterior normal* (19). *Así, estimar μ minimizando la pérdida cuadrática esperada,*

$$min_{a(\boldsymbol{X})} L(\mu, a(\boldsymbol{X}))$$

resulta en elegir como estimador la media de la distribución posterior $a_n^*(\boldsymbol{X})$, con un riesgo asociado igual a la varianza posterior de $\pi(\mu|\boldsymbol{X})$, $v^2 = v_0^2\sigma^2/(\sigma^2 + nv_0^2)$. Esto es, el riesgo global asociado a la decisión de minimizar el error cuadrático y muestrear n unidades viene dado por

$$r(\boldsymbol{X}, n) \equiv r(n) = v^2 + C(n),$$

donde $C(n)$ representa la función de costes asociados al muestreo. Consideremos dos posibilidades para $C(n)$:

CASO 1: *cada observación cuesta lo mismo, c, esto es,*

$$C(n) = c\ n.$$

El riesgo global asociado al estimador Bayes y a un tamaño n viene dado por:

$$r(n) = \frac{\sigma^2 v_0^2}{\sigma^2 + nv_0^2} + cn.$$

Entonces el n^* óptimo se obtiene minimizando el riesgo anterior, que puede obtenerse aproximadamente derivando respecto de n e igualando a cero. Así se obtiene:

$$n^* \approx \sqrt{\frac{\sigma}{c}} - \frac{\sigma^2}{v_0^2}.$$

CASO 2: *cuanto mayor es el tamaño de la muestra, menor es el precio de cada unidad. Una posible función de costes es entonces,*

$$C(n) = \log(1 + n).$$

El riesgo global asociado al estimador Bayes y a un tamaño n viene dado ahora por:

$$r(n) = \frac{\sigma^2 v_0^2}{\sigma^2 + nv_0^2} + \log(1 + n),$$

y el TMO por:

$$n^* \approx \frac{\sigma^2}{2} - \frac{\sigma^2}{v_0^2} + \frac{\sigma}{2v_0}\sqrt{4(v_0^2 - \sigma^2) + \sigma^2 v_0^2}.$$

5.2.2 Pérdida Lineal.

La pérdida es una función lineal del error cometido. Funciones de pérdida lineal habituales son:

$$L(\theta, a) = \begin{cases} k_0(\theta - a), & si\ \theta - a \geq 0 \\ k_1(a - \theta), & si\ \theta - a < 0, \end{cases}$$

en la que se ponderan de forma diferente la sobrestimación y la infraestimación. A veces, $k_0 = k_0(\theta)$ y $k_1 = k_1(\theta)$.

Otra habitual es la pérdida absoluta, que valora igualmente un error por la derecha que uno por la izquierda, siendo de la misma magnitud, y está definida en general por:

$$L(\theta, a) = |\theta - a|^k. \tag{30}$$

Se pueden definir igualmente, versiones ponderadas.

Ejemplo 6 (Estimación con datos normales. Pérdida absoluta). *Consideremos igual que en el ejemplo anterior, observaciones normales i.i.d. para el estudio en planificación, y la misma información previa sobre la media μ, (13), que queremos estimar. Si valoramos el error decisorio por estimar μ con $a(\boldsymbol{X})$ con la función de pérdida del error absoluto,*

$$L(\mu, a(\boldsymbol{X}) = |\mu - a(\boldsymbol{X})|^k,$$

tenemos como función de pérdida global:

$$L(\mu, a(\boldsymbol{X}), n) = |\mu - a(\boldsymbol{X})|^k + c\, n.$$

El estimador que minimiza la pérdida esperada con una muestra de tamaño n también es en este caso la media de la distribución posterior, $a_n^*(\boldsymbol{X})$. El riesgo asociado a utilizar la estimación óptima en una muestra de tamaño n es:

$$r(\boldsymbol{X}, n) \equiv r(n) = v(k) \left(\frac{\sigma^2 v_0^2}{\sigma^2 + n v_0^2} \right)^{k/2} + C(n),$$

con $v(k)$ el momento absoluto de orden k para una distribución $N(0, 1)$. Es decir, $r(\boldsymbol{X}, n) = r(n)$ no depende del resultado $\boldsymbol{X}$ observado, y el TMO se obtiene minimizando la expresión anterior, es decir, aproximadamente,

$$n^* \approx \sigma^2 \left(\frac{v(k) k}{2\sigma^2 c} \right)^{\frac{2}{k+2}} - \frac{\sigma^2}{v_0^2}.$$

5.2.3 Pérdida $0 - k_i$.

Es habitual en contraste de hipótesis $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$, con $\Theta = \Theta_0 \cup \Theta_1$, donde las acciones son "$a_i \equiv$aceptar H_i", $i = 0, 1$. Se define la pérdida $0 - k_i$ como aquella que ocasiona pérdidas k_i sólo cuando nos equivocamos aceptando H_i, esto es,

$$L(\theta, a_i) = \begin{cases} 0, & si\ \theta \in \Theta_i \\ k_i, & si\ \theta \in \Theta_j,\ j \neq i. \end{cases} \tag{31}$$

Un caso particular es la pérdida $0-1$, cuando $k_i = 1$. Otra opción es considerar $k_i = k_i(\theta)$, una función creciente con la distancia entre θ y Θ_i. De hecho, la función de pérdida 0-1 tiene como función de riesgo frecuentista las probabilidades de error de tipo I y II (dependiendo del valor real de θ). También la pérdida esperada Bayes tiene sentido pleno para resolver el contraste, ya que evalúa la probabilidad de cada hipótesis cuando se opta por ella, $E^\theta[L(\theta, a_i)] = 1 - Pr(\theta \in \Theta_i)$.

Ejemplo 7 (Contraste con datos normales. Pérdida $0 - k$ y pérdida absoluta). *Consideremos que el experimento a planificar consiste en observar datos normales i.i.d. $N(\mu, \sigma^2)$, con σ^2 conocida, con el propósito de resolver un contraste sobre μ. Consideramos que los costes muestrales por unidad son constantes, de forma que $C(n) = cn$. Respecto a los costes decisorios, distinguimos dos situaciones:*

Contraste puntual *Se plantea el contraste:*

$$H_0 : \mu = \mu_0,\ vs.\ H_1 : \mu = \mu_1,\ con\ \mu_0 < \mu_1.$$

Respecto a la pérdida decisoria, supongamos que acertar no ocasiona pérdidas y equivocarse sí, y que consideramos igual de gravoso equivocarse aceptando una u otra hipótesis. Es decir, tenemos una función de pérdida del tipo (31), con $k_0 = k_1 = k$. Si además ambas hipótesis son igualmente verosiímiles a priori, $\pi(\mu_0) = \pi(\mu_1) = 0.5$, tenemos

$$L(\mu_0, a_1)\, \pi(\mu_0) = L(\mu_1, a_0)\, \pi(\mu_1) = b,$$

La función de pérdida global resulta entonces:

$$L(\mu, a, n) = \begin{cases} c\,n, & si\ acertamos \\ b/2 + c\,n, & si\ nos\ equivocamos. \end{cases}$$

La pérdida esperada (25) *una vez hayamos optado por un tamaño muestral* n *y observado* $\boldsymbol{X} = \boldsymbol{x}$*, será:*

$$\begin{aligned} \rho(\pi(\mu|\boldsymbol{x}), a, n) &= E^{\mu|\boldsymbol{x}}[L(\mu, a, n)] \\ &= b/2\ \pi(\mu_i|\boldsymbol{x}) + c\,n, \quad si\ a = a_j,\ j \neq i. \end{aligned}$$

La decisión óptima consiste en aceptar aquella hipótesis con menor pérdida esperada, esto es, mayor probabilidad a posteriori. Puesto que

$$\pi(\mu_0|\boldsymbol{x}) = \left[1 + \frac{f(\boldsymbol{x}|\mu_1)}{f(\boldsymbol{x}|\mu_0)}\right]^{-1},$$

denotando, como es habitual la media por $\bar{x} = \sum_i x_i/n$*, tenemos:*

$$a_n^*(\boldsymbol{x}) = \begin{cases} a_0, & si\ \pi(\mu_0|\boldsymbol{x}) \geq 0.5 \;\;\Leftrightarrow\;\; \bar{x} \leq (\mu_0 + \mu_1)/2 \\ a_1, & si\ \pi(\mu_0|\boldsymbol{x}) < 0.5 \;\;\Leftrightarrow\;\; \bar{x} > (\mu_0 + \mu_1)/2. \end{cases}$$

noindent El riesgo esperado de un estudio con tamaño de muestra n *es el valor esperado de* $\rho(\pi(\mu|\boldsymbol{x}), a, n)$ *respecto de la distribución predictiva de* $\boldsymbol{X}$*:*

$$m(\boldsymbol{x}) = f(\boldsymbol{x}|\mu_0)\,\pi(\mu_0) + f(\boldsymbol{x}|\mu_1)\,\pi(\mu_1) \propto \frac{1}{2}\,[N(\bar{x}|\mu_0, \sigma^2/n) + N(\bar{x}|\mu_1, \sigma^2/n)],$$

resultando:

$$\begin{aligned} r(n) &= \int_{-\infty}^{\frac{\mu_0+\mu_1}{2}} [b/2\,\pi(\mu_1|\boldsymbol{x}) + c\,n]\ m(\boldsymbol{x})d\boldsymbol{x} + \int_{\frac{\mu_0+\mu_1}{2}}^{+\infty} [b/2\,\pi(\mu_0|\boldsymbol{x}) + c\,n]\ m(\boldsymbol{x})d\boldsymbol{x} \\ &= c\,n + \frac{b}{2}\left[1 - \Phi\left(\frac{\mu_1 - \mu_0}{2\sqrt{\sigma^2/n}}\right)\right] \end{aligned}$$

Es decir, el coste muestral domina al riesgo decisorio en la función de riesgo global si $c \approx b$.

El TMO es el valor de n *que minimiza el riesgo esperado* $r(n)$*. Derivando en la expresión anterior e igualando a cero, obtenemos aproximadamente el* n^* *óptimo como solución de la ecuación:*

$$\frac{\mu_1 - \mu_0}{\sqrt{2\theta\sigma^2 n^*}}\ exp\left\{-\frac{(\mu_1 - \mu_0)^2\, n^*}{4\sigma^2}\right\} = \frac{8\,c}{b}$$

Contraste Unilateral *El objetivo del estudio es resolver el contraste unilateral*

$$H_0 : \mu \leq \mu_0 \; vs. \; H_1 : \mu > \mu_0.$$

Suponemos, como es habitual, que la distribución a priori sobre μ *es una* $N(m_0, v_0^2)$ *con media y varianza conocidas. Si las consecuencias de equivocarse no dependen de la dirección en que nos equivoquemos, una función de pérdida razonable es:*

$$L(\mu, a, n) = \begin{cases} c\,n, & \text{si acertamos} \\ |\mu - \mu_0|^k + c\,n, & \text{si nos equivocamos .} \end{cases}$$

La regla de decisión óptima en $\boldsymbol{X} = \boldsymbol{x}$ *es:*

$$\text{aceptar } H_0 \text{ si } \int_{\mu > \mu_0} (\mu - \mu_0)^k \pi(\mu|\boldsymbol{x}) d\boldsymbol{x} < \int_{\mu \leq \mu_0} (\mu_0 - \mu)^k \pi(\mu|\boldsymbol{x}) d\boldsymbol{x}.$$

Tras promediar el riesgo asociado a la decisión óptima respecto de la distribución predictiva para la media $\bar{x}$,

$$m(\bar{x}) = \int N(\bar{x}|\mu, \sigma^2/n) \; N(\mu|m_0, v_0^2) \; d\mu = N(\bar{x}|m_0, \sigma^2/n + v_0^2),$$

y minimizarlo respecto de n*, obtenemos como TMO (aproximado),*

$$n^* \approx \left[\frac{v(k+1)\sigma^{(k+1)} \exp\{-(1/2v_0^2)(\theta_0 - m_0)^2\}}{2(2\theta)^{1/2} v_0 c} \right]^{2/(k+3)}.$$

Otros casos más generales pueden encontrarse en Berger (1985).

5.3 Recapitulación y aproximaciones

Si bien hemos presentado unas cuantas funciones de pérdida habituales, las opciones de elección son ilimitadas: funciones que consideran (o equilibran) objetivos distintos (por ejemplo, en investigaciones farmacéuticas, cabe valorar los costes de experimentación, los beneficios al paciente, el precio de mercado, ...); las que permiten una asignación óptima de recursos (Tan y Smith, 1998), etc. En todo caso, la elección de la función de pérdida ha de depender del problema específico que tratamos y de sus objetivos.

Básicamente, plantear y resolver un problema estadístico como un *problema de decisión* consistirá en:

– especificar las creencias previas sobre los parámetros desconocidos a través de una distribución a priori;

– valorar la gravedad de un error decisorio y de los costes muestrales, y expresarlos mediante una función de pérdida (utilidad)

– determinar y optar por aquellas decisiones (estimador, forma de resolver el contraste, etc.) y tamaños óptimos que minimizan la pérdida esperada (respecto de $(\mathbf{X}, \theta)$)

– utilizar aproximaciones MC o MCMC cuando resulte complejo calcular las 'esperanzas' involucradas en la determinación de las decisiones y tamaños óptimos.

De hecho, para resolver la búsqueda del TMO utilizando la aproximación de minimizar el riesgo esperado $r(n)$, es preciso resolver varias integrales $(d\boldsymbol{X}, d\theta)$, la mayoría de las veces muy complejas, de dimensiones altas y con frecuencia no analíticas. Una opción sencilla es utilizar aproximaciones Monte Carlo. Para un grid de valores de n $(n_i, i = 1, \ldots, I)$, se puede aproximar $r(n)$ puntualmente con:

$$\hat{r}(n) = \frac{1}{L}\sum_{l=1}^{L} L(\theta^{(l)}, a_n^*(\mathbf{x}^{(l)}), n), \tag{32}$$

donde $\theta^{(l)} \sim \pi(\theta)$, y $\mathbf{x}^{(l)} \sim f(\boldsymbol{x}|\theta^{(l)}),\ l = 1, 2, \ldots, L$. Finalmente se determina el n de dicho grid para el que se obtiene el mínimo valor de $\hat{r}(n)$.

Si este procedimiento es demasiado costoso, bien por ser un grid demasiado grande o porque la dimensión de las integrales a resolver es demasiado alta, una buena aproximación práctica es la propuesta de Müller y Parmigiani (1995), que proponen:

– elegir aleatoriamente algunos n_i,

– generar *sólo un* $(\theta^{(i)}, \mathbf{x}^{(i)})$ para cada uno de los correspondientes i's y evaluar (32) con una sola simulación para cada n_i considerado; es decir, evaluar $L(\theta^{(i)}, a_n^*(\mathbf{x}^{(i)}), n_i)$,

– considerar el grafo de $L(\theta^{(i)}, a_n^*(\mathbf{x}^{(i)}), n_i)$ vs. n_i,

– ajustar una curva 'suave´ $\tilde{r}(n)$ a los puntos anteriores,

– minimizar $\tilde{r}(n)$.

Ejemplo 8 (Estimación binomial, pérdida absoluta). *Consideramos la planificación de un estudio en el que se van a observar* n *pruebas Bernoulli de parámetro*

θ, o lo que es lo mismo, una prueba binomial $Y|\theta \sim Bi(n,\theta)$. En ausencia de información inicial sobre θ utilizamos una distribución previa uniforme,

$$\theta \sim U(0,1).$$

Como objetivo es estimar θ, las acciones son las estimaciones. Si consideramos como función de pérdida decisoria la del error absoluto cometido al estimar θ con a, y como función de costes muestrales $C(n) = c\,n$, tenemos como función de pérdida global:

$$L(\theta, a(Y), n) = |\theta - a(Y)| + c\,n.$$

Para un n fijo y un resultado $Y = y$, la acción óptima (estimación) es la mediana de la distribución posterior para θ,

$$a_n^*(y) = m_y = \text{ mediana de } Be(y+1, n-y+1).$$

Si bien es posible resolver analíticamente las integrales involucradas, se puede conseguir de un modo más sencillo una aproximación MC al riesgo esperado para un grid de valores de n, $(n_i; i = 1, 2, \ldots, L)$, sin más que simular $\theta^{(l)} \sim Un(0,1)$ y calcular:

$$\hat{r}(n_i) = \frac{1}{L}\sum_{l=0}^{L}\sum_{x=0}^{n_i}[|m_x - \theta^{(l)}| - c\,n]\left(\begin{array}{c} n_i \\ x \end{array}\right)(\theta^{(l)})^x(1-\theta^{(l)})^{n_i-x}.$$

Müller y Parmigiani (1995) llevan a cabo esta aproximación comparándola con la solución exacta encontrando que la aproximación funciona muy bien.

6 Conclusiones

El uso de información adicional a la proporcionada por los datos es un tema aun controvertido en la inferencia: Los métodos bayesianos subjetivos la utilizan explícitamente en forma de una distribución previa propia, los métodos frecuentistas, los verosimilitudistas y los bayesianos objetivos no la utilizan. Sin embargo, dicha información es la única disponible en el diseño y determinación del tamaño de muestra óptimo, y se utiliza, explícitamente o no, tanto en el diseño frecuentista como en el bayesiano.

Puestos a utilizar información adicional, creemos que es más razonable (aparte de proporcionar mejores resultadoa) hacerlo en términos de una distribución para θ que refleje la incertidumbre sobre su valor desconocido, en vez de utilizar un valor concreto como si fuera el verdadero valor de θ, con certidumbre plena.

Volvemos a remarcar que puede utilizarse el diseño bayesiano, mas natural, aun cuando el analisis a realizar en el experimento que se diseña vaya a ser frecuentista. De forma similar, si el analisis posterior es bayesiano, el hecho de utilizar una distribución subjetiva en el diseño no significa que tenga que utilizarse la misma distribución en el análisis. Esta práctica es, obviamente, la más natural, pero en ocasiones por conveniencia se desea realizar un análisis Bayesiano objetivo en cuyo caso no se incorporaría la información adicional.

En este trabajo hemos dado una pequeña muestra de la filosofía subyacente a la determinación bayesiana del tamño de muestra óptimo, así como un pequeño muestrario de aplicaciones concretas para ejemplificar distintos objetivos, métodos de análisis, ... , etc. Técnicas muy similares se utilizan para problemas más complejos. Por ejemplo, puede utilizarse la información de un experimento previo para determinar el TMO de *réplicas* del mismo, tanto réplicas cercanas, como réplicas en las que se vá a variar algunas de las condiciones de la experimentación. En Bayarri y Mayoral (1997, 1999, 2002, 2003) se obtienen TMO para conseguir 'éxito' al replicar, con distintas acepciones de 'éxito', análisis bayesianos y frecuentistas, réplicas parecidas al experimento original o moderadamente distintas, y tanto si la réplica se vá analizar separadamente o vá a incorporarse en un *meta-análisis* con otros experimentos. El potencial de los métodos bayesianos es muy grande.

Agradecimientos

Este trabajo ha sido desarrollado en parte con ayuda del Ministerio de Ciencia y Tecnología, bajo el Proyecto de Investigación SAF2001-2931. Las autoras quieren agradecer al profesor Jaoquín Muñoz García su apoyo y estímulo.

Bibliografía

Adcock, C.J. (1997). Sample size determination: a review. *The Statistician*, **46**, N.2, 261-283.

Bayarri, M.J. y Mayoral, A.M. (1997). Diseño y análisis bayesianos de réplicas en la experimentación científica. *Qüestiió* **21**, 59–97.

Bayarri, M.J. y Mayoral, A.M. (1999). Graphical models for hierarchical computations in the analysis and design of replications. *Revista de la Real Academia de las Ciencias, Special Issue on Bayesian Methods in the Sciences* **93**, 281–293.

Bayarri, M.J. y Mayoral, A.M. (2002). Bayesian analysis and design for comparison of effect-sizes. *Journal of Statistical Planning and Inference* **103**, 225–243.

Bayarri, M.J. y Mayoral, A.M. (2003). Designing "successful" replications. *The American Statistician* (en prensa).

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis, 2nd Edition.* New York: Springer-Verlag.

Carlin, J.B. y Louis, T.A. (1985). Controlling error rates by using conditional expected power to select tumor sites. *Proc. Biopharmaceutical Section of the Amer. Statist. Assoc.*, 11–18.

Carlin, J.B. y Louis, T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis, 2nd Edition* Boca Raton: Chapman&Hall.

Chaloner, K. y Verdinelli, I. (1995). Bayesian experimental design: a review. *Statistical Science* **10**, 273–304.

Gilks, W.R., Richardson, S., and Speigelhalter, D. (1995). *Practical Markov Chain Monte Carlo.* Chapman and Hall.

Joseph, L. and Bélisle, P. (1997). Bayesian sample size determination for normal means and differences between normal means. *Statistician*, **46**, 209-226.

Joseph, L., du Berger, R. and Bélisle,P. (1997). Bayesian and mixed bayesian/likelihood criteria for sample size determination. *Statistics in Medicine*, **16**, 769-781.

Joseph, L., Wolfson, D.B. and du Berger, R. (1995). Sample size calculations for binomial proportions via highest posterior density intervals. *Statistician*, **44**, 143-154.

Lee, S.J., y Zelen, M. (2000). Clinical trials and sample size considerations: another perspective. *Statistical Science* **15**, 95–110 (with discussion).

Lindley, D.V. (1997). The choice of sample size. *The Statistician*, **46**, N.2, 129-138.

Louis, T.A. y Bailey, J.K. (1990). Controlling error rates using prior information and marginal totals to select tumor sites. *J. Statisticial Planning and Inference*, **24**, 297–316.

Müller. P. y Parmigiani, G. (1995). Optimal design via curve fitting of Monte Carlo Experiments. *J. Amer. Statist. Assoc.*, **90**, 1322–1330.

Ripley, B.D. (1987). *Stochastic Simulation*. Wiley.

Shih, J. (1995). Sample size calculation for complex clinical trials with survival endpoints. *Controlled Clinical Trials*, **16**, 395–407.

Smeeton, C. y Adcock, C.J., editors (1997). *Special Issue on Sample Size Determination. The Statistician*, **46**, 129–283.

Spiegelhalter, D.J. and Freedman, L.S. (1986). A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. *Statistics in Medicine*, **5**, 1-13.

Tanner, M.A. (1993). *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions, Second ed.*. Springer-Verlag.

Verdinelli, I. (1992). Advances in Bayesian experimental design (with discussion). In *Bayesian Statistics 4* (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds.). Oxford: Oxford University Press. pp 467–481.

Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *Statistician*, **46**, 145-191.