

# 9

## Extrayendo valores de densidades

### 9.1 Introducción

La simulación de un estadístico consiste en extraer valores al azar de una densidad de probabilidad, calcular el estadístico para cada valor extraído y promediar los resultados. En todos los casos, el investigador quiere calcular un promedio de tipo  $\bar{t} = \int t(\varepsilon)f(\varepsilon)d\varepsilon$ , donde  $t(\cdot)$  es el estadístico de interés y  $f(\cdot)$  es una densidad de probabilidad. Para aproximar esta media a través de simulación, el investigador debe poder extraer valores de la densidad  $f(\cdot)$ . Para algunas densidades, esta tarea es simple. Sin embargo, en muchas situaciones, el cómo se extraen valores de la densidad correspondiente no es algo inmediato. Incluso con densidades simples, puede haber formas de extraer valores que proporcionen una mejor aproximación a la integral que una secuencia puramente aleatoria de extracciones.

En este capítulo exploramos estos temas. En las primeras secciones, se describen los métodos más importantes que han sido desarrollados para extraer valores puramente aleatorios de varias clases de densidades. Estos métodos se presentan de forma progresiva, empezando por procedimientos simples que funcionan para unas pocas densidades especialmente convenientes, pasando posteriormente a tratar métodos cada vez más complejos que trabajan con densidades menos convenientes. La exposición culmina con el algoritmo Metropolis-Hastings, que se puede utilizar con (prácticamente) cualquier densidad. El capítulo finaliza regresando a la cuestión de si es posible obtener - y cómo - una secuencia de valores que proporcionen una mejor aproximación a la integral correspondiente que una secuencia puramente aleatoria. Expondremos los métodos de antitéticos, el muestreo sistemático y las secuencias de Halton, y mostraremos el beneficio que este tipo de extracción de valores proporciona en la estimación de los parámetros de un modelo.

### 9.2 Extracción de valores aleatorios

#### 9.2.1 *Distribuciones normales y uniformes estándar*

Si el investigador quiere extraer un valor aleatorio de una densidad normal estándar (es decir, una normal con media cero y varianza unitaria) o de una densidad uniforme estándar (uniforme entre 0 y 1), el proceso es muy simple desde una perspectiva de programación. La mayoría de los paquetes de software estadístico contienen generadores de números aleatorios para estas densidades. El investigador simplemente debe llamar a estas rutinas para obtener una secuencia de valores al azar. En

las siguientes secciones nos referiremos a un valor extraído al azar de una densidad normal estándar como  $\eta$  y a una extracción de una densidad uniforme estándar como  $\mu$ .

Los valores generados por estas rutinas son en realidad números *pseudo-aleatorios*, porque nada de lo que una computadora puede hacer es verdaderamente aleatorio. Hay muchas cuestiones que intervienen en el diseño de estas rutinas. El objetivo de su diseño es producir números que exhiban las propiedades de extracciones al azar. En qué medida se logra este objetivo depende, por supuesto, de cómo se definen las propiedades de una extracción "al azar". Estas propiedades son difíciles de definir con precisión, ya que el azar es un concepto teórico que no tiene un reflejo en el mundo real. Desde una perspectiva práctica, mi consejo es el siguiente: a menos que uno esté dispuesto a pasar un tiempo considerable investigando y resolviendo (de hecho, diría re-resolviendo) estos temas, probablemente la mejor idea es usar las rutinas disponibles en lugar de programar una nueva.

### 9.2.2 Transformaciones de la normal estándar

Algunas variables aleatorias son transformaciones de una normal estándar. Por ejemplo, un valor al azar de una densidad normal con media  $b$  y varianza  $s^2$  se obtiene como  $\varepsilon = b + s\eta$ . Un valor al azar de una densidad log-normal se obtiene mediante una exponencial de una extracción de una densidad normal:  $\varepsilon = e^{b+s\eta}$ . Los momentos estadísticos de la log-normal son funciones de la media y la varianza de la normal exponenciada. En concreto, la media de  $\varepsilon$  es  $\exp(b + (s^2/2))$  y su varianza es  $\exp(2b + s^2)(\exp(s^2) - 1)$ . Dados los valores de la media y la varianza de la log-normal, es posible calcular los valores correctos de  $b$  y  $s$  a utilizar en la transformación. Sin embargo, es más común tratar  $b$  y  $s$  como los parámetros de la log-normal, y calcular su media y su varianza a partir de estos parámetros.

### 9.2.3 Densidades acumulativas inversas para densidades univariadas

Considere una variable aleatoria con densidad  $f(\varepsilon)$  y la correspondiente distribución acumulativa  $F(\varepsilon)$ . Si  $F$  es invertible (es decir, si se puede calcular  $F^{-1}$ ) entonces es posible extraer valores al azar de  $\varepsilon$  a partir de valores al azar de una distribución uniforme estándar. Por definición,  $F(\varepsilon) = k$  significa que la probabilidad de extraer un valor al azar igual o menor a  $\varepsilon$  es  $k$ , donde  $k$  está entre cero y uno. Un valor  $\mu$  extraído al azar de una distribución uniforme estándar proporciona un número entre cero y uno. Podemos establecer  $F(\varepsilon) = \mu$  y resolver para el correspondiente  $\varepsilon$ :  $\varepsilon = F^{-1}(\mu)$ . Cuando  $\varepsilon$  se obtiene de esta manera, la distribución acumulativa de los valores extraídos es igual a  $F$ , de tal manera que los valores son equivalentes a valores extraídos directamente de  $F$ . En la figura 9.1 puede observarse un ejemplo. Un valor  $\mu^1$  extraído de una distribución uniforme estándar se transforma en el valor de  $\varepsilon$  etiquetado como  $\varepsilon^1$ , valor en el que  $F(\varepsilon^1) = \mu^1$ .

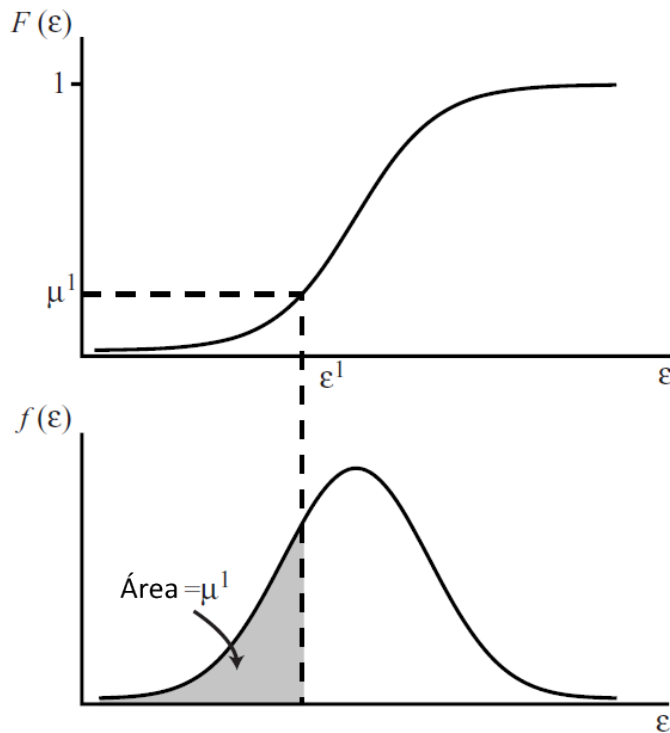


Figura 9.1 Extraemos un valor  $\mu^1$  de una uniforme y creamos  $\varepsilon^1 = F^{-1}(\mu)$

La distribución de valor extremo, que es la base de los modelos logit multinomiales, nos proporciona un ejemplo de este método. La densidad es  $f(\varepsilon) = \exp(-\varepsilon) \cdot \exp(-\exp(-\varepsilon))$  con una distribución acumulativa  $F(\varepsilon) = \exp(-\exp(-\varepsilon))$ . Podemos extraer un valor al azar de esta densidad de probabilidad como  $\varepsilon = -\ln(-\ln \mu)$ .

Tenga en cuenta que este procedimiento sólo funciona para distribuciones univariadas. Si  $\varepsilon$  tiene dos o más elementos, entonces  $F^{-1}(\mu)$  no es único, ya que diversas combinaciones de los elementos de  $\varepsilon$  tienen la misma probabilidad acumulativa.

#### 9.2.4 Densidades univariadas truncadas

Considere una variable aleatoria que va de  $a$  a  $b$  con densidad de probabilidad proporcional a  $f(\varepsilon)$  dentro de este rango. Es decir, la densidad es  $(1/k)f(\varepsilon)$  para  $a \leq \varepsilon \leq b$  y 0 en caso contrario, donde  $k$  es la constante de normalización que asegura que la integral de la densidad resulta 1:  $k = \int_a^b f(\varepsilon)d\varepsilon = F(b) - F(a)$ . Podemos extraer un valor de esta densidad mediante la aplicación del procedimiento descrito en la Sección 9.2.3, asegurando que el valor está dentro del rango apropiado.

Para ello, extraiga  $\mu$  de una densidad uniforme estándar. Calcule el promedio ponderado de  $F(a)$  y  $F(b)$  como  $\bar{\mu} = (1 - \mu)F(a) + \mu F(b)$ . Luego calcule  $\varepsilon = F^{-1}(\bar{\mu})$ . Dado que  $\bar{\mu}$  está entre  $F(a)$  y  $F(b)$ ,  $\varepsilon$  necesariamente está entre  $a$  y  $b$ . Básicamente, el valor extraído de  $\mu$  determina hasta dónde llegar entre  $a$  y  $b$ . Tenga en cuenta que la constante de normalización  $k$  no se utiliza en los cálculos y por lo tanto no necesita ser calculada. La figura 9.2 ilustra el proceso.

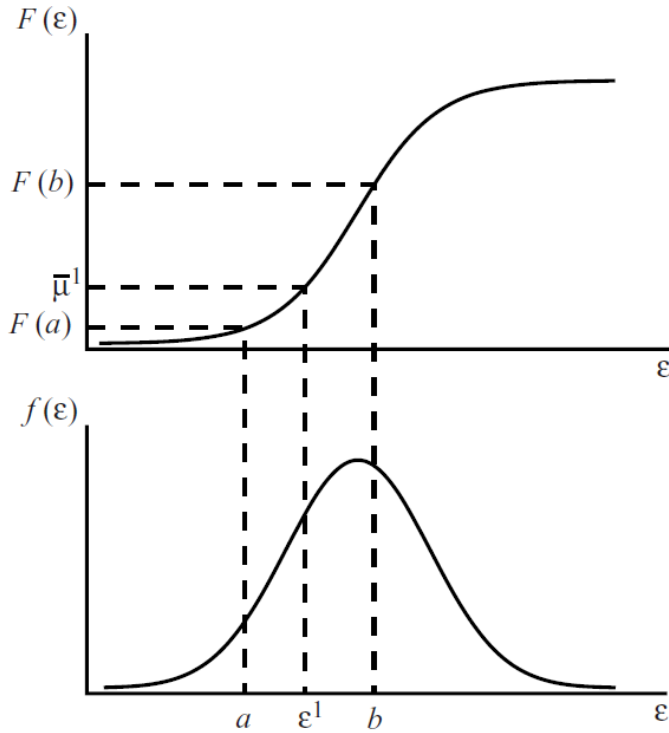


Figura 9.2. Un valor  $\bar{\mu}^1$  entre  $F(a)$  y  $F(b)$  genera un valor  $\varepsilon^1$  de  $f(\varepsilon)$  entre  $a$  y  $b$ .

### 9.2.5 Transformación Choleski de normales multivariadas

Como se describe en la Sección 9.2.2, una normal univariada con media  $b$  y varianza  $s^2$  se obtiene como  $\varepsilon = b + s\eta$ , donde  $\eta$  es una normal estándar. Es posible utilizar un procedimiento análogo para extraer valores al azar de una normal multivariada. Sea  $\varepsilon$  un vector con  $K$  elementos distribuidos  $N(b, \Omega)$ . Un factor Choleski de  $\Omega$  se define como una matriz triangular inferior  $L$  tal que  $LL' = \Omega$ . A menudo recibe el nombre de raíz cuadrada generalizada de  $\Omega$  o desviación estándar generalizada de  $\varepsilon$ . Con  $K = 1$  y varianza  $s^2$ , el factor Choleski es  $s$ , que simplemente es la desviación estándar de  $\varepsilon$ . La mayoría de los paquetes de software estadístico y de manipulación de matrices tienen rutinas para calcular un factor Choleski de cualquier matriz simétrica definida positiva.

Es posible extraer un valor al azar  $\varepsilon$  de  $N(b, \Omega)$  de la siguiente manera. Extraiga  $K$  valores de una normal estándar y etiquete el vector que contiene estos valores como  $\eta = \langle \eta_1, \dots, \eta_K \rangle'$ . Calcule  $\varepsilon = b + L\eta$ . Podemos verificar las propiedades del  $\varepsilon$  resultante: se distribuye normalmente ya que la suma de normales es normal. Su media es  $b$ :  $E(\varepsilon) = b + LE(\eta) = b$ . Y su covarianza es  $\Omega$ :  $Var(\varepsilon) = E(L\eta(L\eta)') = LE(\eta\eta')L' = LVar(\eta)L' = LIL' = LL' = \Omega$ .

Para ser concretos, considere un caso de  $\varepsilon$  tridimensional con media cero. Un valor de  $\varepsilon$  puede calcularse como

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} = \begin{pmatrix} s_{11} & 0 & 0 \\ s_{21} & s_{22} & 0 \\ s_{31} & s_{32} & s_{33} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix}$$

o

$$\varepsilon_1 = s_{11}\eta_1,$$

$$\varepsilon_2 = s_{21}\eta_1 + s_{22}\eta_2,$$

$$\varepsilon_3 = s_{31}\eta_1 + s_{32}\eta_2 + s_{33}\eta_3,$$

A través de estas relaciones vemos que  $\text{Var}(\varepsilon_1) = s_{11}^2$ ,  $\text{Var}(\varepsilon_2) = s_{21}^2 + s_{22}^2$  y  $\text{Var}(\varepsilon_3) = s_{31}^2 + s_{32}^2 + s_{33}^2$ . También que  $\text{Cov}(\varepsilon_1, \varepsilon_2) = s_{11}s_{21}$  y así sucesivamente. Los elementos  $\varepsilon_1$  y  $\varepsilon_2$  están correlacionados debido a la influencia común de  $\eta_1$  en ambos. No están perfectamente correlacionados porque  $\eta_2$  entra en  $\varepsilon_2$  sin afectar  $\varepsilon_1$ . Un análisis similar aplica a  $\varepsilon_1$  y  $\varepsilon_3$ , y a  $\varepsilon_2$  y  $\varepsilon_3$ . En esencia, el factor Choleski expresa  $K$  términos correlacionados a partir de  $K$  componentes independientes, de manera que cada componente *carga* (afecta) de manera diferente cada término. Para cualquier patrón de covarianza, existe un conjunto de cargas de componentes independientes que reproduce esa covarianza.

### 9.2.6 Aceptación-rechazo para densidades multivariadas truncadas

El procedimiento descrito en la Sección 9.2.4 para la extracción de valores al azar de densidades truncadas sólo aplica a distribuciones univariadas. Con densidades multivariadas, extraer valores de distribuciones truncadas es más difícil. A continuación describimos un procedimiento de aceptación-rechazo que siempre es aplicable. Sin embargo, como veremos más adelante, este enfoque tiene desventajas que pueden hacer elegir al investigador otros métodos siempre que estén disponibles.

Supongamos que queremos extraer un valor al azar de una densidad multivariada  $g(\varepsilon)$  dentro del rango  $a \leq \varepsilon \leq b$ , donde  $a$  y  $b$  son vectores con la misma longitud de  $\varepsilon$ . Es decir, queremos extraer valores de  $f(\varepsilon) = \frac{1}{k}g(\varepsilon)$  si  $a \leq \varepsilon \leq b$ , e iguales a cero en caso contrario, donde  $k$  es la constante de normalización. Podemos extraer valores de  $f$  simplemente extrayendo valores de  $g$  y reteniendo ("aceptando") los valores que se encuentran dentro del área de la distribución pertinente y descartando ("rechazando") los valores que se encuentran fuera de ese área. La ventaja de este procedimiento es que se puede aplicar siempre que sea posible extraer valores de la densidad no truncada. Es importante destacar que no es necesario conocer la constante de normalización  $k$  para la densidad truncada. Este hecho es útil debido a que suele ser difícil calcular esta constante de normalización.

La desventaja de este procedimiento es que el número de valores extraídos que son aceptados (es decir, el número de valores de  $f$  que obtenemos) no es fijo, sino que es a su vez un número aleatorio. Si se extraen  $R$  valores al azar de  $g$ , el número esperado de aceptaciones es  $kR$ . Esta esperanza no puede conocerse sin conocer  $k$ , que, como ya se ha mencionado, es por lo general difícil de calcular. Por tanto, es difícil determinar un número apropiado de valores a extraer de  $g$ . Más importante aún, el número real de aceptaciones generalmente diferirá de la cantidad esperada. De hecho, hay una probabilidad positiva de no obtener ninguna aceptación a partir de un número fijo de valores extraídos. Cuando el espacio de truncamiento es pequeño (o, más precisamente, cuando  $k$  es pequeño) no obtener ninguna aceptación, y por lo tanto no obtener ningún valor de la densidad truncada, es un evento probable.

Esta dificultad se puede sortear extrayendo valores de  $g$  hasta obtener un cierto número de valores aceptados. Es decir, en lugar de establecer de antemano el número de valores a extraer de  $g$ , el investigador puede establecer el número de valores extraídos de  $f$  que quiere obtener. Por supuesto, el investigador no sabe cuánto tiempo tardará en alcanzar el número establecido. En la mayoría de situaciones, es posible aplicar más fácilmente otros procedimientos para extraer valores de una densidad truncada multivariada. Sin embargo, es importante recordar que cuando ningún otro método parece posible con una distribución truncada, el procedimiento de aceptación-rechazo puede ser aplicado.

### 9.2.7 Muestreo por importancia

Supongamos que  $\varepsilon$  tiene una densidad  $f(\varepsilon)$  de la cual no pueden extraerse valores fácilmente por otros procedimientos. Supongamos además que existe otra densidad  $g(\varepsilon)$  de la que sí pueden extraerse valores al azar fácilmente. Podemos extraer valores de  $f(\varepsilon)$  de la siguiente manera. Extraiga un valor de  $g(\varepsilon)$  y etiquételo como  $\varepsilon^1$ . Pondere el valor por el factor  $f(\varepsilon^1)/g(\varepsilon^1)$ . Repita este proceso muchas veces. El conjunto de valores ponderados es equivalente a un conjunto de valores extraídos de  $f$ .

Para verificar este hecho, mostraremos que la distribución acumulativa de los valores ponderados extraídos de  $g$  es la misma que la distribución acumulativa de valores extraídos de  $f$ . Considere la proporción de valores extraídos de  $g$  que están por debajo cierto valor  $m$ , con cada valor ponderado por el factor  $f/g$ . Esta proporción es

$$\int \frac{f(\varepsilon)}{g(\varepsilon)} I(\varepsilon < m) g(\varepsilon) d\varepsilon = \int_{-\infty}^m \frac{f(\varepsilon)}{g(\varepsilon)} g(\varepsilon) d\varepsilon = \int_{-\infty}^m f(\varepsilon) d\varepsilon = F(m).$$

En simulación, los valores extraídos de una densidad se usan para calcular el promedio de un estadístico sobre dicha densidad. El muestreo por importancia puede ser visto como un cambio en el estadístico acompañado del correspondiente cambio en la densidad que haga que extraer valores de esa densidad sea fácil. Supongamos que queremos calcular  $\int t(\varepsilon) f(\varepsilon) d\varepsilon$ , pero resulta difícil extraer valores de  $f$ . Podemos multiplicar el integrando por  $g \div g$  sin cambiar su valor, por lo que la integral es  $\int t(\varepsilon) [f(\varepsilon)/g(\varepsilon)] g(\varepsilon) d\varepsilon$ . Para simular la integral, extraemos valores de  $g$ , calculamos  $t(\varepsilon) [f(\varepsilon)/g(\varepsilon)]$  para cada valor y promediamos los resultados. Simplemente hemos transformado la integral de modo que resulte más fácil para la simulación.

La densidad  $f$  recibe el nombre de densidad objetivo y  $g$  de densidad propuesta. Los requisitos para poder aplicar el muestreo por importancia son que (1) el ámbito de  $g(\varepsilon)$  cubra el ámbito de  $f$ , de modo que cualquier  $\varepsilon$  que podría surgir con  $f$  también pueda surgir con  $g$ , y (2) el ratio  $f(\varepsilon)/g(\varepsilon)$  debe ser finito para todos los valores posibles de  $\varepsilon$ , de modo que este ratio siempre pueda ser calculado.

Un ejemplo útil de muestreo por importancia lo encontramos en densidades normales truncadas multivariadas. Supongamos que queremos extraer valores al azar de  $N(0, \Omega)$ , pero de forma que cada elemento sea positivo (es decir, truncando la densidad de probabilidad por debajo de cero). La densidad es

$$f(\varepsilon) = \frac{1}{k(2\pi)^{\frac{1}{2}K} |\Omega|^{1/2}} e^{-\frac{1}{2} \varepsilon' \Omega^{-1} \varepsilon},$$

para  $\varepsilon > 0$ , y 0 en caso contrario, donde  $K$  es la dimensión de  $\varepsilon$  y  $k$  es la constante de normalización. (Asumimos para los fines de este ejemplo que  $k$  es conocido. En realidad, el cálculo de  $k$  puede a su vez requerir simulación). Extraer valores al azar de esta densidad es difícil debido a que los elementos de  $\varepsilon$  están correlacionados así como truncados. Sin embargo, podemos utilizar el procedimiento de la sección 9.2.4 para extraer valores al azar de normales independientes truncadas y luego aplicar muestreo por importancia para crear la correlación. Para ello, extraiga valores de  $K$  normales univariadas truncadas por debajo de cero, utilizando el procedimiento de la Sección 9.2.4. Estos valores constituyen colectivamente una extracción de un vector  $k$ -dimensional de  $\varepsilon$  desde el cuadrante positivo con densidad

$$g(\varepsilon) = \frac{1}{m(2\pi)^{\frac{1}{2}K}} e^{-\frac{1}{2} \varepsilon' \varepsilon},$$

donde  $m = 1/2^K$ . Para cada valor extraído, asigne el peso

$$\frac{f(\varepsilon)}{g(\varepsilon)} = \frac{m}{k} |\Omega|^{-1/2} e^{\varepsilon'(\Omega^{-1} - I)\varepsilon}.$$

Los valores ponderados son equivalentes a valores extraídos a partir de  $N(0, \Omega)$  truncada por debajo de cero.

Una última reflexión: observe que el procedimiento de aceptación-rechazo descrito en la sección 9.2.6 es un tipo de muestreo por importancia. La distribución truncada es la densidad objetivo y la distribución no truncada es la densidad propuesta. Cada valor extraído de la densidad no truncada es ponderado por una constante si el valor está dentro del espacio de truncamiento y es ponderado por cero si el valor está fuera del espacio de truncamiento. Ponderar por una constante o por cero es equivalente a ponderar por uno (aceptar) o cero (rechazar).

### 9.2.8 Muestreo de Gibbs (Gibbs Sampling)

Para distribuciones multinomiales, a veces es difícil extraer valores al azar directamente de la densidad conjunta y sin embargo resulta sencillo hacerlo de la densidad condicionada de cada elemento dados los valores del resto de elementos. El muestreo de Gibbs (el término aparentemente fue introducido por Geman y Geman, 1984) se puede utilizar en estas situaciones. Una explicación general del método lo proporcionan Casella y George (1992), explicación que el lector puede utilizar como complemento a la descripción más concisa que facilito a continuación.

Considere dos variables aleatorias  $\varepsilon_1$  y  $\varepsilon_2$ . La generalización a una dimensión mayor es obvia. La densidad conjunta es  $f(\varepsilon_1, \varepsilon_2)$  y las densidades condicionadas son  $f(\varepsilon_1|\varepsilon_2)$  y  $f(\varepsilon_2|\varepsilon_1)$ . El muestreo de Gibbs actúa extrayendo valores al azar iterativamente de las densidades condicionadas: extrae  $\varepsilon_1$  condicionado a un valor de  $\varepsilon_2$ , extrae  $\varepsilon_2$  condicionado a ese valor de  $\varepsilon_1$ , extrae de nuevo  $\varepsilon_1$  condicionado al nuevo valor de  $\varepsilon_2$  y así sucesivamente. Este proceso converge a valores extraídos de la densidad conjunta.

Para ser más precisos, el método consiste en: (1) Elija un valor inicial para  $\varepsilon_1$ , llamado  $\varepsilon_1^0$ . Cualquier valor con densidad distinta de cero puede ser elegido. (2) Extraiga un valor de  $\varepsilon_2$ , llamado  $\varepsilon_2^0$ , de la densidad  $f(\varepsilon_2|\varepsilon_1^0)$ . (3) Extraiga un valor de  $\varepsilon_1$ , llamado  $\varepsilon_1^1$ , a partir de  $f(\varepsilon_1|\varepsilon_2^0)$ . (4) Extraiga  $\varepsilon_2^1$  de  $f(\varepsilon_2|\varepsilon_1^1)$  y así sucesivamente. Los valores  $\varepsilon_1^t$  de  $f(\varepsilon_1|\varepsilon_2^{t-1})$  y los valores  $\varepsilon_2^t$  de  $f(\varepsilon_2|\varepsilon_1^t)$  constituyen una secuencia en  $t$ . Para una  $t$  suficientemente grande (es decir, para un número suficientemente grande de iteraciones) la secuencia converge a valores extraídos de la densidad conjunta  $f(\varepsilon_1, \varepsilon_2)$ .

Como ejemplo, considere dos normales con desviación estándar, independientes salvo por el hecho de que se truncan en base a su suma:  $\varepsilon_1 + \varepsilon_2 \leq m$ . La figura 9.3 representa la densidad truncada. Los círculos son contornos de la densidad no truncada y el área sombreada representa la densidad truncada. Para obtener las densidades condicionadas, considere primero las normales no truncadas. Puesto que las dos variables son independientes, la densidad condicionada de cada una es igual a su densidad no condicionada. Es decir, haciendo caso omiso del truncamiento,  $\varepsilon_1|\varepsilon_2 \sim N(0,1)$ . La regla de truncamiento que hemos definido es  $\varepsilon_1 + \varepsilon_2 \leq m$  que puede ser reexpresada como  $\varepsilon_1 \leq m - \varepsilon_2$ . Por lo tanto,  $\varepsilon_1|\varepsilon_2$  se distribuye como una normal estándar univariada, truncada por encima de  $m - \varepsilon_2$ . Dado  $\varepsilon_2$ , un valor de  $\varepsilon_1$  se obtiene usando el procedimiento de la sección 9.2.4:  $\varepsilon_1 = \Phi^{-1}(\mu\Phi(m - \varepsilon_2))$ , donde  $\mu$  es un valor extraído de una uniforme estándar y  $\Phi(\cdot)$  es la distribución normal estándar acumulativa. Valores de  $\varepsilon_2$  condicionados a  $\varepsilon_1$  se obtienen de forma análoga. Obtener valores secuencialmente de estas densidades condicionadas finalmente proporciona valores de la densidad truncada conjunta.

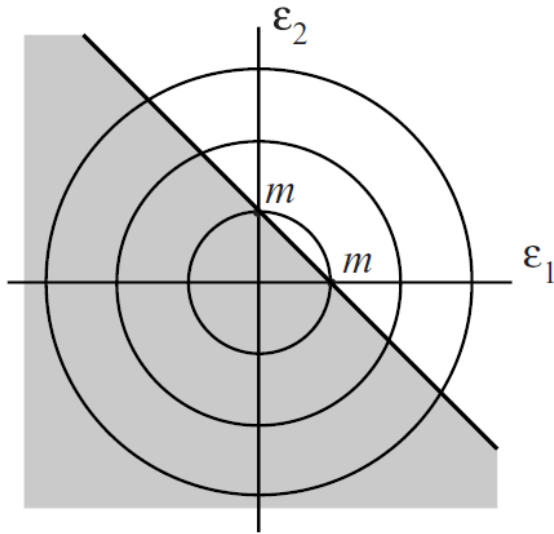


Figura 9.3 Densidad normal truncada

### 9.2.9 Algoritmo Metropolis-Hastings

Si todo lo demás falla, puede usarse el algoritmo Metropolis-Hastings (MH) para extraer valores de una densidad de probabilidad. El algoritmo fue inicialmente desarrollado por Metropolis et al. (1953) y generalizado posteriormente por Hastings (1970). Funciona como se indica a continuación. El objetivo es extraer valores al azar de  $f(\varepsilon)$ :

1. Empezar con un valor del vector  $\varepsilon$ , etiquetado  $\varepsilon^0$ .
2. Elija un valor de prueba de  $\varepsilon^1$  como  $\tilde{\varepsilon}^1 = \varepsilon^0 + \eta$ , donde  $\eta$  se extrae de una distribución  $g(\eta)$  que tiene media cero. Por lo general, se suele especificar una distribución normal para  $g(\eta)$ .
3. Calcule la densidad en el valor de prueba  $\tilde{\varepsilon}^1$  y compárelo con la densidad en el valor original  $\varepsilon^0$ . Es decir, compare  $f(\tilde{\varepsilon}^1)$  con  $f(\varepsilon^0)$ . Si  $f(\tilde{\varepsilon}^1) > f(\varepsilon^0)$ , entonces acepte  $\tilde{\varepsilon}^1$ , etiquételo como  $\varepsilon^1$  y vaya al paso 4. Si  $f(\tilde{\varepsilon}^1) \leq f(\varepsilon^0)$ , entonces acepte  $\tilde{\varepsilon}^1$  con probabilidad  $f(\tilde{\varepsilon}^1)/f(\varepsilon^0)$  y recházelo con probabilidad  $1 - f(\tilde{\varepsilon}^1)/f(\varepsilon^0)$ . Para determinar si aceptamos o rechazamos  $\tilde{\varepsilon}^1$ , en este caso, extraiga un valor al azar de una distribución uniforme estándar  $\mu$ . Si  $\mu \leq f(\tilde{\varepsilon}^1)/f(\varepsilon^0)$ , entonces mantenga  $\tilde{\varepsilon}^1$ . De lo contrario, rechace  $\tilde{\varepsilon}^1$ . Si se acepta  $\tilde{\varepsilon}^1$ , etiquételo como  $\varepsilon^1$ . Si  $\tilde{\varepsilon}^1$  es rechazado, entonces utilice  $\varepsilon^0$  como  $\varepsilon^1$ .
4. Elija un valor de prueba de  $\varepsilon^2$  como  $\tilde{\varepsilon}^2 = \varepsilon^1 + \eta$ , donde  $\eta$  es un nuevo valor extraído al azar de  $g(\eta)$ .
5. Aplique la regla descrita en el paso 3 para aceptar  $\tilde{\varepsilon}^2$  como  $\varepsilon^2$  o rechace  $\tilde{\varepsilon}^2$  y utilice  $\varepsilon^1$  como valor de  $\varepsilon^2$ .
6. Continúe este proceso durante múltiples iteraciones. La secuencia  $\varepsilon^t$  acaba siendo equivalente a extraer valores al azar de  $f(\varepsilon)$  para una  $t$  suficientemente grande.

Los valores generados se correlacionan en serie, ya que cada valor depende del valor anterior. De hecho, cuando un valor de prueba es rechazado, el valor empleado en la serie es igual al valor previo. Esta correlación en serie debe tenerse en cuenta cuando se utilizan los valores generados.

El algoritmo MH se puede aplicar a cualquier densidad que pueda ser calculada. Este algoritmo es particularmente útil cuando la constante de normalización de una densidad no se conoce o no se puede calcular fácilmente. Supongamos que sabemos que  $\varepsilon$  se distribuye proporcional a  $f^*(\varepsilon)$ . Esto significa



que la densidad de  $\varepsilon$  es  $f(\varepsilon) = \frac{1}{k} f^*(\varepsilon)$ , donde la constante de normalización  $k = \int f^*(\varepsilon) d\varepsilon$  asegura que la integral de  $f$  es 1. Por lo general,  $k$  no se puede calcular analíticamente, por la misma razón por la que tenemos que simular integrales en otras circunstancias. Por suerte, el algoritmo MH no utiliza  $k$ . Un valor de prueba de  $\varepsilon^t$  se prueba en primer lugar determinando si  $f(\varepsilon^t) > f(\varepsilon^{t-1})$ . Esta comparación no se ve afectada por la constante de normalización, ya que la constante entra en el denominador en ambos lados. Por lo tanto, si  $f(\varepsilon^t) \leq f(\varepsilon^{t-1})$ , aceptamos el valor de prueba con probabilidad  $f(\varepsilon^t)/f(\varepsilon^{t-1})$ . La constante de normalización desaparece de este ratio.

El algoritmo MH es en realidad más general de lo que describo en esta explicación, aunque en la práctica suele aplicarse como lo describo. Chib y Greenberg (1995) proporcionan una excelente descripción del algoritmo más general, así como una explicación de por qué funciona. Bajo la definición más general, el muestreo de Gibbs es un caso particular del algoritmo MH, tal y como Gelman (1992) señaló. El algoritmo MH y el muestreo de Gibbs a menudo se llaman métodos de Monte Carlo – Cadena de Markov (*Markov chain Monte Carlo*, MCMC o MC al cuadrado); Chib y Greenberg (1996) proporcionan una descripción de su uso en econometría. Los valores extraídos son cadenas de Markov porque cada valor depende sólo del valor extraído inmediatamente antes, y los métodos son Monte Carlo porque los valores se extraen de forma aleatoria. Exploraremos más cuestiones sobre el algoritmo MH, tales como la forma de elegir  $g(\varepsilon)$ , en el contexto de su uso con los procedimientos bayesianos jerárquicos (en el capítulo 12).

### 9.3 Reducción de la varianza

El uso de valores independientes extraídos al azar para la simulación es atractivo por ser conceptualmente simple y porque las propiedades estadísticas del simulador resultante son fáciles de obtener. Sin embargo, hay otras formas de extraer valores de una densidad que pueden proporcionar mayor precisión para un número dado de extracciones. Examinaremos estos métodos alternativos en las siguientes secciones.

Recordemos que el objetivo es aproximar una integral de la forma  $\int t(\varepsilon) f(\varepsilon) d\varepsilon$ . Al extraer una secuencia de valores de la densidad  $f(\varepsilon)$ , dos cuestiones están en juego: la cobertura y la covarianza. Considere en primer lugar la cobertura. La integral es sobre toda la densidad  $f$ . Parece razonable pensar que obtendríamos una aproximación más precisa si evaluásemos  $t(\varepsilon)$  en valores de  $\varepsilon$  que se extendiesen en todo el dominio de  $f$ . Con valores independientes extraídos al azar, es posible que los valores se agrupen, sin que obtengamos valores en grandes áreas del dominio. Es de esperar que procedimientos que garanticen una mejor cobertura del dominio proporcionen una mejor aproximación.

La covarianza es otro tema relevante. Con valores independientes, la covarianza entre valores es cero.

Por tanto, la varianza de un simulador basado en  $R$  valores independientes es la varianza basada en un valor dividida por  $R$ . Si los valores extraídos se correlacionan negativamente en lugar de ser independientes, entonces la varianza del simulador es menor. Considere  $R = 2$ . La varianza de  $\bar{t} = [t(\varepsilon_1) + t(\varepsilon_2)]/2$  es  $[V(t(\varepsilon_1)) + V(t(\varepsilon_2)) + 2Cov(t(\varepsilon_1), t(\varepsilon_2))]/4$ . Si los valores son independientes, la varianza es  $V(t(\varepsilon_r))/2$ . Si los dos valores se correlacionan negativamente entre sí, el término de covarianza es negativo y la varianza es inferior a  $V(t(\varepsilon_r))/2$ . Básicamente, cuando los valores están correlacionados negativamente dentro de un simulador no sesgado, un valor extraído que esté por encima de  $\bar{t} = E_r(t(\varepsilon))$  tiende a estar asociado a un valor para la siguiente extracción que está por debajo de  $\bar{t} = E_r(t(\varepsilon))$ , de tal manera que su media está más cerca del verdadero valor  $\bar{t}$ .

El mismo concepto aplica cuando los simuladores se suman entre observaciones. Por ejemplo, la función log-verosimilitud simulada es una suma sobre observaciones del logaritmo de probabilidades simuladas. Si los valores extraídos para la simulación de cada observación son independientes de los valores

extraídos para otras observaciones, la varianza de la suma es simplemente la suma de las varianzas. Si los valores extraídos se toman de manera que se cree una correlación negativa entre observaciones, la varianza de la suma es inferior.

Para una observación dada, la cuestión de la covarianza está relacionada con la de la cobertura. Al inducir una correlación negativa entre valores, usualmente estamos asegurando una mejor cobertura. Con  $R = 2$ , si los dos valores se extraen de forma independiente, entonces ambos pueden estar en la parte baja de la distribución. Por el contrario, si se induce correlación negativa, el segundo valor tenderá a ser alto si el primer valor fue bajo, lo que proporciona una mejor cobertura.

A continuación se describen métodos para lograr una mejor cobertura para la integral de cada observación y para inducir correlación negativa entre valores extraídos para cada observación, así como entre observaciones. Para facilitar la explicación asumimos que la integral es una probabilidad de elección y que la suma sobre las observaciones es la función log-verosimilitud simulada. Sin embargo, los conceptos aplican a otras integrales, tales como puntuaciones (*scores*), así como para otras sumas, como las condiciones de momentos (*moment conditions*) y cuotas de mercado (*market shares*). Además, a menos que se indique lo contrario, ilustraremos los métodos con sólo dos términos aleatorios de manera que los valores extraídos puedan ser representados gráficamente. Los términos aleatorios se etiquetan como  $\varepsilon^a$  y  $\varepsilon^b$ , y colectivamente como  $\varepsilon = \langle \varepsilon^a, \varepsilon^b \rangle'$ . Un valor de  $\varepsilon$  extraído de su densidad  $f(\varepsilon)$  se denota como  $\varepsilon_r = \langle \varepsilon_r^a, \varepsilon_r^b \rangle'$  para  $r = 1, \dots, R$ . Por lo tanto,  $\varepsilon_3^a$ , por ejemplo, es el tercer valor extraído del primer término aleatorio.

### 9.3.1 Antitéticos (antithetics)

La extracción de valores antitéticos, sugerida por Hammersley y Morton (1956), se obtiene mediante la creación de diferentes tipos de imágenes espejo de un mismo valor. Para una densidad simétrica centrada en cero, la variable aleatoria antitética más simple se crea invirtiendo el signo de todos los elementos de un valor extraído. La figura 9.4 ilustra este método. Supongamos que extraemos de  $f(\varepsilon)$  un valor al azar  $\varepsilon_1 = \langle \varepsilon_1^a, \varepsilon_1^b \rangle'$ . El segundo "valor extraído", al que llamamos antítesis del primer valor, se crea como  $\varepsilon_2 = \langle -\varepsilon_1^a, -\varepsilon_1^b \rangle'$ . Cada valor extraído de  $f$  crea un par de "extracciones", el valor original y su imagen espejo (reflejado en relación al origen). Para obtener un total de  $R$  valores, extraemos de  $f$  un total de  $R/2$  valores independientemente y los otros  $R/2$  restantes se crean como los negativos de los valores originales.

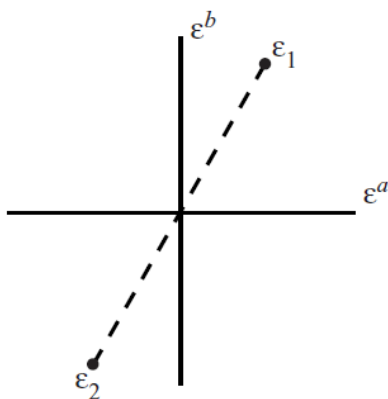


Figura 9.4 Signo invertido de los dos elementos.

Cuando la densidad no está centrada en cero, es posible aplicar el mismo concepto pero a través de un proceso diferente. Por ejemplo, la densidad uniforme estándar se extiende entre 0 y 1, con 0.5 como valor central. Si extraemos de esta densidad un valor  $\mu_1$ , podemos crear su valor antitético como  $\mu_2 = 1 - \mu_1$ .

Este valor dista igual de 0.5 que la primera extracción realizada, pero en el otro lado de 0.5. En general, para cualquier densidad univariada con función acumulativa  $F(\varepsilon)$ , la antítesis de un valor se crea como  $F^{-1}(1 - F(\varepsilon))$ . En el caso de una densidad simétrica centrada en cero, esta fórmula general es equivalente a invertir directamente el signo. En lo que queda de exposición supondremos que la densidad es simétrica y está centrada en cero, lo que hace los conceptos más fáciles de expresar y visualizar.

La correlación entre un valor extraído y su antítesis es exactamente -1, de manera que la varianza de su suma es igual a cero:  $V(\varepsilon_1 + \varepsilon_2) = V(\varepsilon_1) + V(\varepsilon_2) + 2Cov(\varepsilon_1, \varepsilon_2) = 0$ . Este hecho no significa que no haya varianza en la probabilidad simulada que se basa en estos valores. La probabilidad simulada es una función no lineal de los términos aleatorios y, por lo tanto, la correlación entre  $P(\varepsilon_1)$  y  $P(\varepsilon_2)$  es menor que uno. La varianza de la probabilidad simulada  $\check{P} = \frac{1}{2}[P(\varepsilon_1) + P(\varepsilon_2)]$  es mayor que cero. Sin embargo, la varianza de las probabilidades simuladas es menor que  $\frac{1}{2}V_r(P(\varepsilon_r))$ , que es la varianza que tendríamos con dos valores extraídos de forma independiente.

Como se muestra en la figura 9.4, invertir el signo de un valor extraído da puntos de evaluación en cuadrantes opuestos. El concepto se puede ampliar para obtener valores en cada cuadrante. Para ello, se extrae un valor y posteriormente se crean valores antitéticos invirtiendo cada elemento por separado (dejando el signo de los otros elementos sin alterar), invirtiendo el signo de cada pareja de elementos, de cada triplete de elementos, y así sucesivamente. Para un  $\varepsilon$  con dos elementos, este proceso crea tres valores antitéticos por cada valor extraído independiente. Para  $\varepsilon_1 = \langle \varepsilon_1^a, \varepsilon_1^b \rangle'$ , los valores antitéticos son

$$\varepsilon_2 = \langle -\varepsilon_1^a, \varepsilon_1^b \rangle',$$

$$\varepsilon_3 = \langle \varepsilon_1^a, -\varepsilon_1^b \rangle',$$

$$\varepsilon_4 = \langle -\varepsilon_1^a, -\varepsilon_1^b \rangle'.$$

Estos valores se muestran en la figura 9.5. Cada cuadrante contiene un valor.

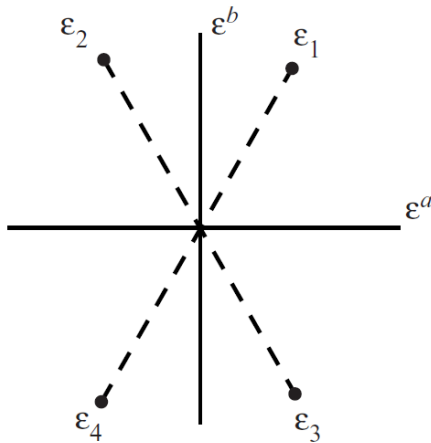


Figura 9.5. Invierta el signo de cada elemento, luego el de ambos.

Moviendo la posición de cada elemento e invirtiendo sus signos, se puede obtener mejor cobertura y mayor correlación negativa. En la figura 9.5,  $\varepsilon_1$  y  $\varepsilon_2$  están bastante próximos entre sí, al igual que  $\varepsilon_3$  y  $\varepsilon_4$ . Esta disposición deja grandes áreas descubiertas entre  $\varepsilon_1$  y  $\varepsilon_3$  y entre  $\varepsilon_2$  y  $\varepsilon_4$ . Es posible obtener valores ortogonales con disposición equilibrada intercambiando el elemento  $\varepsilon_1^a$  con el elemento  $\varepsilon_1^b$ , e invirtiendo los signos al mismo tiempo. Los valores antitéticos resultantes son

$$\varepsilon_2 = \langle -\varepsilon_1^b, \varepsilon_1^a \rangle',$$

$$\varepsilon_3 = \langle \varepsilon_1^b, -\varepsilon_1^a \rangle',$$

los cuales se ilustran en la figura 9.6. Por supuesto, estos conceptos pueden extenderse a cualquier número de dimensiones. Para un  $\varepsilon$   $M$ -dimensional, cada valor extraído crea  $2^M$  valores antitéticos (incluyendo el original), con un valor en cada cuadrante.

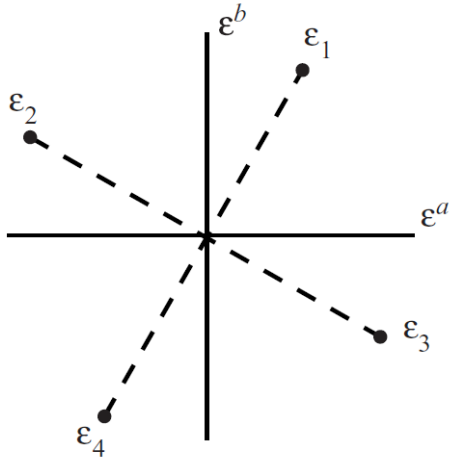


Figura 9.6. Intercambie posiciones e invierta signos.

Las comparaciones realizadas por Vijverberg (1997) y Sándor y András (2001) muestran que los antitéticos mejoran sustancialmente la estimación de modelos probit. Del mismo modo, Geweke (1988) ha demostrado el valor de su uso al calcular estadísticos basados en distribuciones posteriores bayesianas.

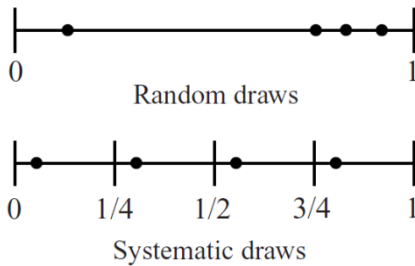


Figura 9.7. Extracción de valores de una uniforme estándar.

### 9.3.2 Muestreo sistemático

La cobertura también puede mejorarse mediante muestreo sistemático (McGrath, 1970), un método que crea una malla de puntos sobre el dominio de la densidad y que aleatoriamente desplaza la malla de puntos al completo. Considere la extracción de valores de una distribución uniforme entre 0 y 1. Si se extraen cuatro valores de forma independiente, los puntos podrían distribuirse como se muestra en la parte superior de la figura 9.7, obteniendo una cobertura bastante pobre. En lugar de extraerlos de forma totalmente independiente, podemos dividir el intervalo unitario en cuatro segmentos y extraer valores al azar de forma que aseguremos que hay un valor en cada segmento, con idéntica distancia entre valores. Para ello, extraiga un valor de una uniforme entre 0 y 0.25 (mediante una extracción de

un valor de una densidad uniforme estándar, dividiendo el resultado por 4). Etiquete el valor como  $\varepsilon_1$ . Los otros tres valores se crean como

$$\varepsilon_2 = 0.25 + \varepsilon_1,$$

$$\varepsilon_3 = 0.50 + \varepsilon_1,$$

$$\varepsilon_4 = 0.75 + \varepsilon_1.$$

Estos valores se distribuyen tal y como se muestra en la parte inferior de la figura 9.7, de forma que proporcionan una mejora cobertura que la extracción de valores independientes.

La cuestión es, ¿qué nivel de segmentación debemos aplicar sobre el intervalo? Por ejemplo, para extraer un total de 100 valores, el intervalo unitario puede dividirse en 100 segmentos. Se extrae un valor entre 0 y 0.01, y posteriormente definimos los restantes 99 valores a partir de este primer valor. En lugar de hacer esto, el intervalo unitario puede dividirse en menos de 100 valores y extraer más valores independientes. Si el intervalo se divide en cuatro segmentos, podemos extraer 25 valores independientes entre 0 y 0.25, y posteriormente calcular tres valores en los otros segmentos para cada uno de los valores extraídos de forma independiente. El investigador debe buscar un término medio al decidir el nivel de segmentación de la malla empleada en el muestro sistemático. Un número mayor de segmentos proporcionan una cobertura más uniforme para un número total dado de valores. Sin embargo, un número menor de segmentos proporcionan más aleatoriedad al proceso. En nuestro ejemplo con  $R = 100$ , sólo hay un valor aleatorio si utilizamos 100 segmentos, mientras que hay 25 valores aleatorios si usamos cuatro segmentos.

La aleatoriedad de los valores extraídos para una simulación es una propiedad necesaria en la obtención de las propiedades asintóticas de los estimadores basados en simulación, como se describe en el capítulo 10. Muchas de las propiedades asintóticas se basan en el concepto de que el número de valores extraídos al azar aumenta sin límites con el tamaño de la muestra. Las distribuciones asintóticas se vuelven relativamente exactas sólo cuando se han tomado suficientes valores al azar. Por lo tanto, para un número total dado de valores, el objetivo de lograr una mejor cobertura, que se logra con una segmentación definida con más precisión, debe ser equilibrado con el objetivo de tener suficiente aleatoriedad en las fórmulas asintóticas como para poder aplicarlas, lo que se logra con un menor nivel de segmentación (segmentos mayores). El mismo problema aplica a los antitéticos vistos anteriormente.

El muestreo sistemático puede realizarse en múltiples dimensiones. Considere una distribución uniforme bidimensional en una región cuadrada unitaria. Podemos crear una malla dividiendo cada dimensión en segmentos. Tal y como se muestra en la figura 9.8, cuando cada dimensión se divide en cuatro segmentos, el cuadrado unitario queda dividido en 16 zonas. Se extrae al azar un valor entre 0 y 0.25 para cada elemento de  $\varepsilon$ , resultando  $\varepsilon_1 = \langle \varepsilon_1^a, \varepsilon_1^b \rangle'$ , donde  $0 < \varepsilon_1^a < 0.25$  y  $0 < \varepsilon_1^b < 0.25$ . Este valor extraído cae en algún lugar de la zona inferior izquierda en la figura 9.8. Para obtener otros quince valores, sumamos al "origen" de cada zona el valor  $\varepsilon_1 = \langle \varepsilon_1^a, \varepsilon_1^b \rangle'$ . Por ejemplo, el punto que se crea para la zona inferior derecha es  $\varepsilon_4 = \langle (0.75 + \varepsilon_1^a), (0 + \varepsilon_1^b) \rangle'$ .

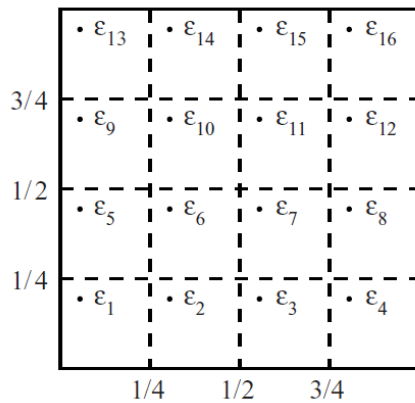


Figura 9.8. Valores extraídos sistemáticamente en dos dimensiones.

Estos valores se definen para una distribución uniforme. Cuando  $f$  representa otra densidad, los puntos se transforman utilizando el método descrito en la sección 9.2.3. En particular, sea  $F$  la distribución acumulativa asociada a la densidad  $f$  univariada. Es posible hacer un muestreo sistemático de  $f$  transformando cada extracción sistemática de una uniforme mediante  $F^{-1}$ . Por ejemplo, para una normal estándar, se divide la densidad en cuatro segmentos de igual tamaño con puntos de corte:  $\Phi^{-1}(0.25) = -0.67$ ,  $\Phi^{-1}(0.5) = 0$  y  $\Phi^{-1}(0.75) = 0.67$ . Tal y como se muestra en la figura 9.9, estos segmentos son de igual tamaño en el sentido de que cada uno contiene la misma masa de densidad. Las extracciones de la normal estándar se obtienen mediante la extracción de un valor de una uniforme entre 0 y 0.25, etiquetado  $\mu_1$ . El punto correspondiente en la normal es  $\varepsilon_1 = \Phi^{-1}(\mu_1)$ , que se ubica en el primer segmento. Los puntos para los otros tres segmentos restantes se crean como  $\varepsilon_2 = \Phi^{-1}(0.25 + \mu_1)$ ,  $\varepsilon_3 = \Phi^{-1}(0.5 + \mu_1)$  y  $\varepsilon_4 = \Phi^{-1}(0.75 + \mu_1)$ .

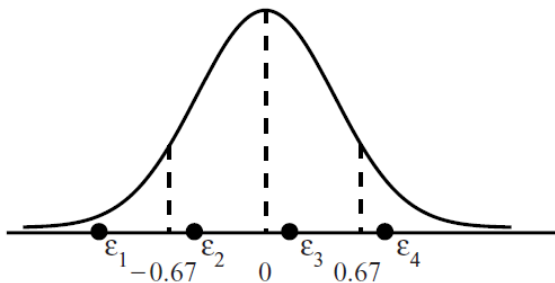


Figura 9.9 Valores sistemáticos para una normal univariada.

Se pueden extraer valores de términos aleatorios multidimensionales de forma similar, con la condición de que los elementos sean independientes. Por ejemplo, si  $\varepsilon$  se compone de dos elementos cada uno de los cuales se distribuye de forma normal estándar, entonces se pueden extraer valores análogos a los de la figura 9.8 de la siguiente manera: Extraiga los valores  $\mu_1^a$  y  $\mu_1^b$  de una uniforme entre 0 y 0.25. Calcule  $\varepsilon_1$  como  $\langle \Phi^{-1}(\mu_1^a), \Phi^{-1}(\mu_1^b) \rangle'$ . Calcule los otros 15 puntos restantes  $\varepsilon_r$  como  $\langle \Phi^{-1}(x_r + \mu_1^a), \Phi^{-1}(y_r + \mu_1^b) \rangle'$ , donde  $\langle x_r, y_r \rangle'$  es el origen del área  $r$  en el cuadrado unitario.

El requisito de que los elementos de  $\varepsilon$  sean independientes no es restrictivo. Se pueden crear elementos aleatorios correlacionados a través de transformaciones de elementos independientes, como la transformación de Choleski. Los elementos independientes se extraen de su densidad y posteriormente se crea la correlación dentro del modelo.

Obviamente, se pueden obtener varios conjuntos de valores muestreados sistemáticamente para lograr más aleatorización. En dos dimensiones con cuatro segmentos en cada dimensión, se pueden obtener

64 valores extrayendo 4 valores independientes en el cuadrado que va de 0 a  $1/4$  y creando 15 valores adicionales para cada valor independiente. Este procedimiento proporciona mayor aleatorización pero una cobertura menos uniforme que la obtenida al generar los valores usando 8 segmentos en cada dimensión, de forma que cada extracción aleatoria en el cuadrante que va de 0 a  $1/8$  se traduce en 64 valores sistemáticos.

Para una distribución normal, los valores que se extraen como se acaba de describir no son simétricos alrededor de cero. Se puede utilizar una aproximación alternativa a este problema para asegurar dicha simetría. Para una normal unidimensional, se pueden obtener 4 valores que son simétricos alrededor de cero como sigue. Extraiga un valor al azar de una uniforme entre 0 y 0.25, y etiquételo como  $\mu_1$ . Cree el valor correspondiente para la normal como  $\varepsilon_1 = \Phi^{-1}(\mu_1)$ . Calcule el valor para el segundo segmento como  $\varepsilon_2 = \Phi^{-1}(0.25 + \mu_1)$ . A continuación, cree los valores para el tercer y cuarto segmento como los negativos de estos valores:  $\varepsilon_3 = -\varepsilon_2$  y  $\varepsilon_4 = -\varepsilon_1$ . La figura 9.10 ilustra los valores generados usando el mismo  $\mu_1$  que en la figura 9.9. Este procedimiento combina el muestreo sistemático con los antitéticos. Se puede extender a múltiples dimensiones mediante la creación de valores extraídos sistemáticamente para el cuadrante positivo y la posterior creación de valores antitéticos para los otros cuadrantes.

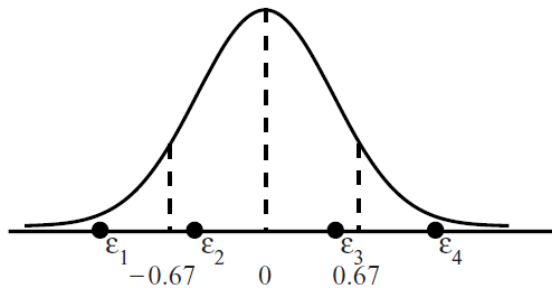


Figura 9.10 Valores simétricos extraídos sistemáticamente.

### 9.3.3 Secuencias de Halton

Las secuencias de Halton (Halton, 1960) proporcionan cobertura y, a diferencia de los otros métodos que hemos expuesto aquí, inducen correlación negativa entre observaciones. Una secuencia de Halton se define en función de un número dado, por lo general un número primo. El concepto detrás de la secuencia se entiende más fácilmente a través de un ejemplo. Considere el número primo 3. La secuencia de Halton para el 3 se crea dividiendo el intervalo unitario en tres partes con puntos de corte en  $\frac{1}{3}$  y  $\frac{2}{3}$ , como se muestra en la parte superior de la figura 9.11. Los primeros términos de la secuencia son estos puntos de corte:  $\frac{1}{3}, \frac{2}{3}$ . A continuación, cada uno de los tres segmentos se divide en tres partes, y los puntos de corte de estos nuevos segmentos se añaden a la secuencia pero de una forma particular. La secuencia se convierte en  $\frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}$ . Observe que los puntos de corte más bajos de los tres segmentos ( $\frac{1}{9}, \frac{4}{9}, \frac{7}{9}$ ) se introducen en la secuencia antes que los puntos de corte más altos ( $\frac{2}{9}, \frac{5}{9}, \frac{8}{9}$ ). Posteriormente, cada uno de los nueve segmentos se divide en tres partes y se añaden nuevamente los puntos de corte a la secuencia. La secuencia se convierte así en  $\frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \frac{10}{27}, \frac{19}{27}, \frac{4}{27}, \frac{13}{27}$ , y así sucesivamente. Este proceso continúa hasta obtener tantos puntos como el investigador desee.

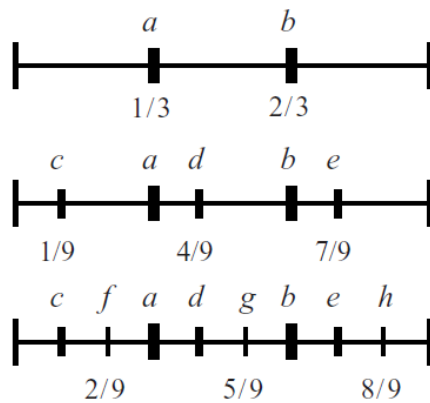


Figura 9.11. Secuencia de Halton para el número primo 3.

Desde el punto de vista de programación, es fácil crear una secuencia de Halton. La secuencia se crea de forma iterativa. En cada iteración  $t$ , la secuencia se denota  $s_t$ , que es una serie de números. La secuencia se amplía en cada iteración, siendo la nueva secuencia  $s_{t+1} = \{s_t, s_t + 1/3^t, s_t + 2/3^t\}$ . El proceso sería el siguiente: comience con 0 como secuencia inicial:  $s_0 = \{0\}$ . El número cero en realidad no es parte de una secuencia de Halton, pero considerarlo como primer elemento facilita la creación de la secuencia, como veremos más adelante. Se puede suprimir una vez se ha creado la secuencia completa. En la primera iteración, agregue  $1/3^1 = \frac{1}{3}$  y luego  $2/3^1 = \frac{2}{3}$  al elemento inicial, obteniendo  $\{0, \frac{1}{3}, \frac{2}{3}\}$ . La secuencia tiene tres elementos. En la segunda iteración, añada  $1/3^2 = \frac{1}{9}$  y luego  $2/3^2 = \frac{2}{9}$  a cada elemento de la secuencia, y añada los resultados:

$$0 = 0,$$

$$1/3 = 1/3,$$

$$2/3 = 2/3,$$

$$0 + 1/9 = 1/9,$$

$$1/3 + 1/9 = 4/9,$$

$$2/3 + 1/9 = 7/9,$$

$$0 + 2/9 = 2/9,$$

$$1/3 + 2/9 = 5/9,$$

$$2/3 + 2/9 = 8/9.$$

La nueva secuencia consta de nueve elementos. En la tercera iteración, añada  $1/3^3 = \frac{1}{27}$  y luego  $2/3^3 = \frac{2}{27}$  a cada elemento de esta secuencia y añada los resultados:

$$0 = 0,$$

$$1/3 = 1/3,$$

$$2/3 = 2/3,$$

$$1/9 = 1/9,$$

$$4/9 = 4/9,$$

$$7/9 = 7/9,$$

$$2/9 = 2/9,$$

$$5/9 = 5/9,$$

$$8/9 = 8/9,$$



$$\begin{aligned}
0 + 1/27 &= 1/27, \\
1/3 + 1/27 &= 10/27, \\
2/3 + 1/27 &= 19/27, \\
1/9 + 1/27 &= 4/27, \\
4/9 + 1/27 &= 13/27, \\
7/9 + 1/27 &= 22/27, \\
2/9 + 1/27 &= 7/27, \\
5/9 + 1/27 &= 16/27, \\
8/9 + 1/27 &= 25/27, \\
0 + 2/27 &= 2/27, \\
1/3 + 2/27 &= 11/27, \\
2/3 + 2/27 &= 20/27, \\
1/9 + 2/27 &= 5/27, \\
4/9 + 2/27 &= 14/27, \\
7/9 + 2/27 &= 23/27, \\
2/9 + 2/27 &= 8/27, \\
5/9 + 2/27 &= 17/27, \\
8/9 + 2/27 &= 26/27.
\end{aligned}$$

La secuencia se compone ahora de 27 elementos. En la cuarta iteración, añada  $1/3^4 = \frac{1}{81}$  y luego  $2/3^4 = \frac{2}{81}$  a cada elemento de la secuencia y añada los resultados, y así sucesivamente.

Observe que la secuencia se compone así de ciclos sobre el intervalo unitario cada tres números:

0	1/3	2/3
1/9	4/9	7/9
2/9	5/9	8/9
1/27	10/27	19/27
4/27	13/27	22/27
7/27	16/27	25/27
2/27	11/27	20/27
5/27	14/27	23/27
8/27	17/27	26/27

Dentro de cada ciclo, los números son ascendentes.

Las secuencias de Halton para otros números primos se crean de manera similar. La secuencia para 2 es  $\left\{\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \frac{9}{16}, \dots\right\}$ . En general, la secuencia para el número primo  $k$  se crea iterativamente, siendo la secuencia en la iteración  $t + 1$   $s_{t+1} = \left\{s_t, s_t + \frac{1}{k^t}, s_t + \frac{2}{k^t}, \dots, s_t + \frac{k-1}{k^t}\right\}$ . La secuencia contiene ciclos de longitud  $k$ , donde cada ciclo consiste en  $k$  puntos ascendentes, equidistantes entre sí, dentro del intervalo unitario.

Dado que una secuencia de Halton está definida en el intervalo unitario, sus elementos pueden ser considerados como una extracción de valores “bien colocados” de una densidad uniforme estándar. La extracción de valores de Halton proporciona, en promedio, una mejor cobertura que la extracción de

valores puramente al azar, debido a que se crean para llenar progresivamente el intervalo unitario de forma uniforme y cada vez más densa. Los elementos de cada ciclo son equidistantes entre sí, y cada ciclo abarca las zonas del intervalo unitario no cubiertas por los ciclos anteriores.

Cuando se utilizan secuencias de Halton para generar una muestra de observaciones, normalmente se crea una secuencia larga y posteriormente se utiliza una parte de la secuencia para cada observación. Los elementos iniciales de la secuencia se descartan por razones que detallaremos posteriormente. Los elementos restantes se utilizan en grupos, con cada grupo de elementos constituyendo los valores extraídos para una observación. Por ejemplo, supongamos que hay dos observaciones, y el investigador quiere  $R = 5$  valores extraídos para cada una. Si se utiliza el número primo 3, y el investigador decide descartar los primeros 10 elementos, se debe crear una secuencia de longitud 20. Esta secuencia es

0	1/3	2/3
1/9	4/9	7/9
2/9	5/9	8/9
1/27	10/27	19/27
4/27	13/27	22/27
7/27	16/27	25/27
2/27	11/27	

Después de eliminar los 10 primeros elementos, los valores de Halton para la primera observación son  $\{\frac{10}{27}, \frac{19}{27}, \frac{4}{27}, \frac{13}{27}, \frac{22}{27}\}$  y los valores de Halton para la segunda observación son  $\{\frac{7}{27}, \frac{16}{27}, \frac{25}{27}, \frac{2}{27}, \frac{11}{27}\}$ . Estos valores se ilustran en la figura 9.12. Observe que las deficiencias en la cobertura de la primera observación se compensan por los valores de la segunda observación. Por ejemplo, la gran brecha que existe entre  $\frac{4}{27}$  y  $\frac{10}{27}$  para la primera observación es rellenada por el punto medio de este vacío,  $\frac{7}{27}$ , en la segunda observación. La brecha entre  $\frac{13}{27}$  y  $\frac{19}{27}$  es rellenada por su punto medio,  $\frac{16}{27}$ , en la segunda observación, y así sucesivamente. El patrón por el cual se crean secuencias de Halton las hace de tal manera que cada sub-secuencia llena los vacíos de las sub-secuencias anteriores.

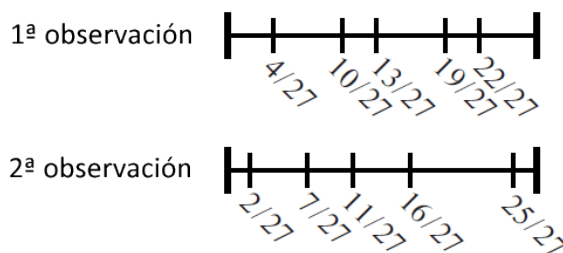


Figura 9.12. Valores de Halton para dos observaciones.

Debido a esta propiedad de llenado de espacios, las probabilidades simuladas basadas en la extracción de valores de Halton tienden a ser auto-correctoras entre observaciones. Los valores de una observación tienden a estar correlacionados negativamente con los de la observación anterior. En nuestro ejemplo, el promedio de los valores correspondiente a la primera observación está por encima de 0.5, mientras que el promedio de los valores para la segunda observación es inferior a 0.5. Esta correlación negativa reduce el error en la función log-verosimilitud simulada.

Cuando el número de valores utilizados para cada observación se eleva, la cobertura para cada observación mejora. La covarianza negativa entre observaciones disminuye, ya que hay menos espacios en la cobertura de cada observación para ser rellenados por la siguiente observación. Esta característica

de auto-corrección entre observaciones de las secuencias de Halton es mayor cuando se extraen pocos valores para cada observación, de manera que la corrección es más necesaria. Sin embargo, la precisión mejora cuantos más valores de Halton se utilizan, ya que la cobertura es mejor para cada observación.

Como se ha descrito hasta ahora, la extracción de valores de Halton se define para una densidad uniforme. Para obtener una secuencia de puntos para otras densidades univariadas, la distribución acumulativa inversa se evalúa en cada elemento de la secuencia de Halton. Por ejemplo, supongamos que el investigador quiere extraer valores de una densidad normal estándar. Se crea una secuencia de Halton para, por ejemplo, el número primo 3, y se evalúa la inversa de la normal acumulativa para cada elemento. La secuencia resultante es

$$\Phi^{-1}\left(\frac{1}{3}\right) = -0.43,$$

$$\Phi^{-1}\left(\frac{2}{3}\right) = 0.43,$$

$$\Phi^{-1}\left(\frac{1}{9}\right) = -1.2,$$

$$\Phi^{-1}\left(\frac{4}{9}\right) = -0.14,$$

$$\Phi^{-1}\left(\frac{7}{9}\right) = 0.76,$$

$\vdots$

Esta secuencia se representa en la figura 9.13. Se puede considerar igual a la del intervalo unitario, en el sentido de que divide la densidad en tres segmentos de igual masa, con puntos de corte en  $-0.43$  y  $0.43$ , y luego divide cada segmento en tres sub-segmentos de igual masa, y así sucesivamente.

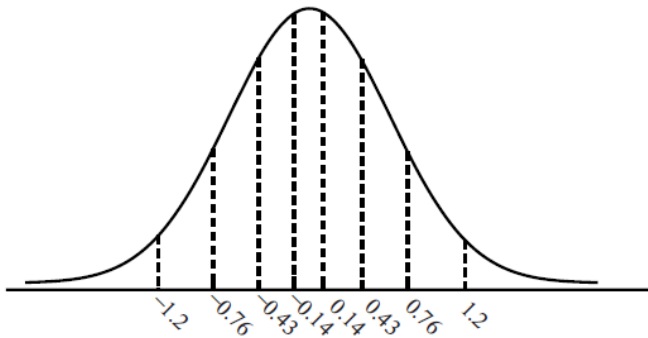


Figura 9.13. Valores de Halton para una normal estándar.

Para obtener secuencias de Halton en múltiples dimensiones se crea una secuencia de Halton para cada dimensión, usando un número primo diferente para cada una. Por ejemplo, una secuencia en dos dimensiones se obtiene mediante la creación de pares de valores de las secuencias de Halton para los números primos 2 y 3. Los puntos son

$$\varepsilon_1 = \left\langle \frac{1}{2}, \frac{1}{3} \right\rangle,$$

$$\varepsilon_2 = \left\langle \frac{1}{4}, \frac{2}{3} \right\rangle,$$

$$\varepsilon_3 = \left\langle \frac{3}{4}, \frac{1}{9} \right\rangle,$$

$$\varepsilon_4 = \left\langle \frac{1}{8}, \frac{4}{9} \right\rangle,$$

$$\varepsilon_5 = \left\langle \frac{5}{8}, \frac{7}{9} \right\rangle,$$

$$\varepsilon_6 = \left\langle \frac{3}{8}, \frac{2}{9} \right\rangle,$$

⋮

Esta secuencia se representa en la figura 9.14. Para extraer valores de una normal estándar de dos dimensiones independientes, se calcula la normal acumulativa inversa de cada elemento de estos pares. Los valores son

$$\varepsilon_1 = \langle 0, -0.43 \rangle,$$

$$\varepsilon_2 = \langle -0.67, 0.43 \rangle,$$

$$\varepsilon_3 = \langle 0.67, -1.2 \rangle,$$

$$\varepsilon_4 = \langle -1.15, -1.2 \rangle,$$

$$\varepsilon_5 = \langle 0.32, 0.76 \rangle,$$

$$\varepsilon_6 = \langle -0.32, -0.76 \rangle,$$

⋮

que se muestran en la figura 9.15.

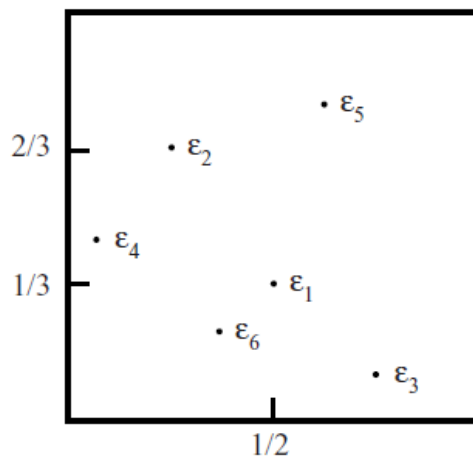


Figura 9.14. Secuencia de Halton en dos dimensiones para los números primeros 2 y 3.

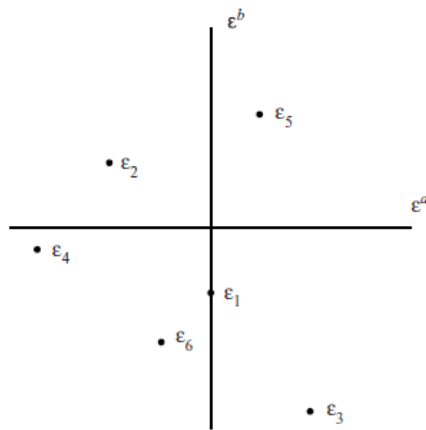


Figura 9.15. Secuencia de Halton para una normal estándar de dos dimensiones.

Al crear secuencias en varias dimensiones, es costumbre eliminar la parte inicial de las series. Los términos iniciales de dos secuencias de Halton están altamente correlacionados, por lo menos durante el primer ciclo de cada secuencia. Por ejemplo, las secuencias para 7 y 11 comienzan con  $\left\{\frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7}\right\}$  y  $\left\{\frac{1}{11}, \frac{2}{11}, \frac{3}{11}, \frac{4}{11}, \frac{5}{11}, \frac{6}{11}\right\}$ . Estos primeros elementos caen en una línea en dos dimensiones, como se muestra en la figura 9.16. La correlación se disipa después de que cada secuencia haya completado un ciclo a través del intervalo unitario, puesto que secuencias con diferentes números primos completan un ciclo a diferentes velocidades. El descarte de la parte inicial de la secuencia elimina la correlación. El número de elementos iniciales a descartar necesita ser al menos tan grande como el mayor primo empleado en la creación de las secuencias.

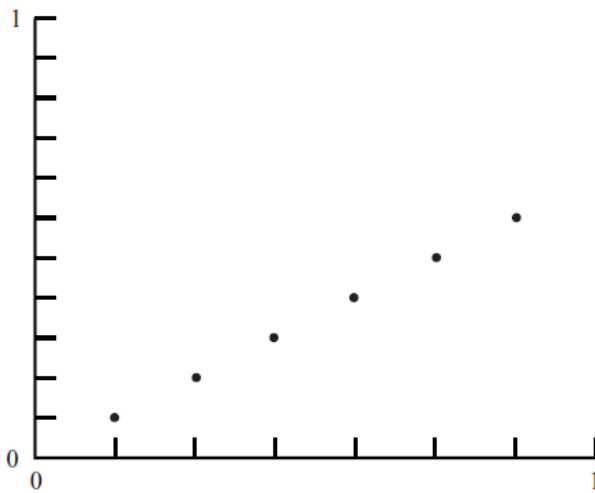


Figura 9.16. Primeros 6 elementos de una secuencia de Halton para los primos 7 y 11.

La potencial correlación es la razón por la que se utilizan números primos para crear las secuencias de Halton en lugar de no primos. Si se utiliza un número compuesto (no primo), entonces hay una posibilidad de que los ciclos coincidan a lo largo de toda la secuencia, en lugar de sólo para los elementos iniciales. Por ejemplo, si se crean secuencias de Halton con los números 3 y 6, la secuencia de 3 completa un ciclo dos veces por cada ciclo completado de la secuencia de 6. Puesto que los elementos dentro de un ciclo son ascendentes, los elementos en cada ciclo de la secuencia de 3

estarán correlacionados con los elementos en el ciclo de la secuencia de 6. Usando sólo números primos evitamos esta superposición de ciclos.

La mayor cobertura y la correlación negativa entre observaciones que se obtiene al usar secuencias de Halton hacen que este método sea mucho más eficaz que la extracción de valores al azar a efectos de simulación. Spanier y Maize (1991) han demostrado que un pequeño número de valores de Halton proporciona una integración relativamente buena. En el contexto de los modelos de elección discreta, Bhat (2001) encontró que 100 valores de Halton proporcionaban resultados más precisos para su logit mixto que 1.000 valores extraídos al azar. De hecho, el error de simulación con 125 valores de Halton era la mitad del error obtenido con 1.000 valores extraídos de forma aleatoria y algo inferior al de 2.000 valores al azar. Train (2000), Munizaga y Alvarez-Daziano (2001), y Hensher (2001) confirman estos resultados en otros conjuntos de datos.

Como ilustración, considere el modelo logit mixto que se describe ampliamente en el capítulo 11. De forma resumida, el modelo describe la elección de proveedor de electricidad de los hogares. En una encuesta de preferencias declaradas, a los encuestados se les presentó una serie de situaciones de elección hipotéticas. En cada situación, se describieron cuatro proveedores de energía y se preguntó al entrevistado qué empresa elegiría. Los proveedores se diferenciaban en función de su precio, de si la empresa requería que el cliente firmase un contrato a largo plazo, de si el proveedor era el suministrador local de energía, de si el proveedor era una empresa bien conocida y de si el proveedor ofrecía tarifas diferenciadas por horario (*time-of-day*, *TOD*) o tarifas estacionales. Se estimó un modelo logit mixto con estas seis características como variables explicativas. Se supuso que el coeficiente de cada variable seguía una distribución normal, excepto para el coeficiente de precio, que se supuso fijo. Por tanto el modelo contenía cinco términos aleatorios para la simulación. Una descripción completa de los datos, del modelo estimado y de sus implicaciones, se facilitan en el capítulo 11, en el cual el contenido del modelo es relevante para el tema tratado en el capítulo. Por ahora, sólo nos interesa la comparación entre el uso de secuencias de Halton y el uso de valores al azar.

Para investigar esta cuestión, el modelo se estimó con 1.000 valores extraídos al azar y luego con 100 valores de Halton. Más concretamente, el modelo se estimó en cinco ocasiones utilizando cinco grupos diferentes de 1.000 valores al azar. Se calculó la media y la desviación estándar de los parámetros estimados en estas cinco simulaciones. Posteriormente, el modelo se estimó en cinco ocasiones usando secuencias de Halton. El primer modelo utilizó los números primos 2, 3, 5, 7 y 11 para las cinco dimensiones de la simulación. Para el resto de modelos se cambió el orden de los números primos, de modo que la dimensión para la que se utilizó cada número primo cambió en cada una de las cinco estimaciones. A continuación, se calculó el promedio y la desviación estándar de los cinco conjuntos de estimaciones.

Las medias de las estimaciones de los parámetros de las cinco simulaciones se pueden ver en la tabla 9.1. La media de las simulaciones basadas en valores extraídos al azar se dan en la primera columna, y las medias para las simulaciones basadas en secuencias de Halton se dan en la segunda columna. Los dos conjuntos de medias son muy similares. Este resultado indica que las secuencias de Halton proporcionan las mismas estimaciones, *en promedio*, que los valores extraídos al azar.

Tabla 9.1. Medias de las estimaciones de parámetros

	1.000 valores al azar	100 valores de Halton
Precio:	-0.8607	-0.8588
Duración del contrato:		
Media	-0.1955	-0.1965

Desv. Estándar	0.3092	0.3158
Proveedor local:		
Media	2.0967	2.1142
Desv. Estándar	1.0535	1.0236
Compañía conocida:		
Media	1.431	1.4419
Desv. Estándar	0.8208	0.6894
Tarifas por franjas horarias:		
Media	-8.3760	-8.4149
Desv. Estándar	2.4647	2.5466
Tarifas estacionales:		
Media	-8.6286	-8.6381
Desv. Estándar	1.8492	1.8977

Las desviaciones estándar de las estimaciones de los parámetros se pueden ver en la tabla 9.2. Para los 11 parámetros excepto uno, las desviaciones estándar son inferiores al emplear 100 valores de Halton respecto a 1.000 valores extraídos al azar. Para ocho de los parámetros, las desviaciones estándar son la mitad de grandes. Teniendo en cuenta que ambos conjuntos de valores dan esencialmente las mismas medias, las desviaciones estándar más bajas obtenidas con los valores de Halton indican que el investigador puede esperar estar más cerca de los valores esperados de las estimaciones si usa 100 valores de Halton en lugar de usar 1.000 valores extraídos al azar.

*Tabla 9.2. Desviaciones estándar de las estimaciones de parámetros*

	1.000 valores al azar	100 valores de Halton
Precio:	0.0310	0.0169
Duración del contrato:		
Media	0.0093	0.0045
Desv. Estándar	0.0222	0.0108
Proveedor local:		
Media	0.0844	0.0361
Desv. Estándar	0.1584	0.1180
Compañía conocida:		
Media	0.0580	0.0242
Desv. Estándar	0.0738	0.1753
Tarifas por franjas horarias:		
Media	0.3372	0.1650
Desv. Estándar	0.1578	0.0696
Tarifas estacionales:		
Media	0.4134	0.1789
Desv. Estándar	0.2418	0.0679

Estos resultados demuestran el beneficio proporcionado por las secuencias de Halton. El tiempo de computación requerido en simulación se puede reducir en un factor diez mediante el uso de secuencias de Halton en lugar de valores extraídos al azar, sin reducir, y de hecho incrementando, la precisión.

Sin embargo, estos resultados deben ser tomados con precaución. El uso de secuencias de Halton y otros números cuasi-aleatorios en la estimación basada en simulación es algo bastante nuevo y no completamente comprendido. Por ejemplo, durante el análisis surgió una anomalía que sirve como advertencia. El modelo se volvió a estimar con 125 valores de Halton en lugar de 100. Se estimó cinco veces con cada una de las cinco posibles ordenaciones de los números primos, como se ha descrito anteriormente. Cuatro de las cinco simulaciones proporcionaron estimaciones muy similares. Sin embargo, la quinta simulación dio estimaciones que eran notablemente diferentes de las demás. Por ejemplo, el coeficiente de precio estimado para las primeras cuatro simulaciones fue -0.862, -0.865, -0.863 y -0.864 respectivamente, mientras que la quinta resultó -0.911. Las desviaciones estándar entre los cinco conjuntos de estimaciones fueron menores a las obtenidas con 1.000 valores al azar, lo que confirma el valor de las secuencias de Halton. Sin embargo, las desviaciones estándar fueron mayores con 125 valores de Halton que con 100 valores de Halton, debido a que la última simulación con 125 valores proporcionaba resultados muy diferentes. La razón de esta anomalía no ha sido determinada aún. Su presencia indica la necesidad de una mayor investigación de las propiedades de las secuencias de Halton en la estimación basada en simulación.

#### 9.3.4 Secuencias de Halton aleatorizadas

Las secuencias Halton son sistemáticas y no aleatorias. Sin embargo, las propiedades asintóticas de los estimadores basados en simulación se obtienen bajo el supuesto de que los valores empleados son aleatorios. Hay dos maneras de abordar esta cuestión. En primer lugar, uno puede darse cuenta de que cuando emplea un generador de números aleatorios, estos tampoco son realmente al azar. Son sistemáticos, como cualquier cosa hecha por una computadora. Un generador de números aleatorios crea valores que tienen muchas de las características de los valores realmente generados al azar, pero en realidad son sólo pseudo-aleatorios. En este sentido, por lo tanto, los valores de Halton pueden ser vistos como una forma sistemática de aproximar la integración que es más precisa que el uso de valores pseudo-aleatorios, que también son sistemáticos. Ninguno de los dos métodos coincide con el concepto teórico de aleatoriedad y, de hecho, no está claro que el concepto teórico en realidad tenga una representación en el mundo real. Ambos métodos cumplen con el objetivo básico subyacente de aproximar una integral sobre una densidad.

En segundo lugar, las secuencias de Halton pueden ser transformadas de una manera que pasan a ser aleatorias, al menos en la misma medida en que los números pseudo-aleatorios son aleatorios. Bhat (2003) describe el proceso, basado en los procedimientos introducidos por Tuffin (1996):

1. Extraiga un valor de una densidad uniforme estándar. Etiquete este valor como  $\mu$ .
2. Añada  $\mu$  a cada elemento de la secuencia de Halton. Si el elemento resultante es superior a 1, se resta 1 del mismo. De lo contrario, mantenga el elemento resultante como está (sin restar 1).

La fórmula que describe esta transformación es  $s_n = \text{mod}(s_0 + \mu)$ , donde  $s_0$  es el elemento original de la secuencia de Halton,  $s_n$  es el elemento transformado y  $\text{mod}$  es la operación consistente en tomar la parte decimal del argumento entre paréntesis.

La transformación se representa en la figura 9.17. Supongamos que el valor  $\mu$  extraído de la densidad uniforme es 0.40. El número 0.33 es el primer elemento de la secuencia de Halton para el número primo 3. Transformamos este elemento, como se muestra en la parte superior de la figura 9.17, a  $0.33 + 0.40 = 0.73$ , que es sólo un desplazamiento de 0.40 en la línea. El número 0.67 es el segundo elemento de la



secuencia. Se transforma mediante la adición de 0.40 y, a continuación, ya que el resultado es superior a 1, restamos 1 para obtener 0.07 ( $0.67 + 0.40 - 1 = 0.07$ ). Como se muestra en la parte inferior de la figura 9.17, esta transformación se visualiza como un movimiento del punto original hacia arriba a una distancia de 0.40, pero volviendo al inicio cuando se alcanza el final del intervalo unitario. El punto se mueve hacia arriba 0.33 hasta donde termina la línea y luego se lleva al inicio de la línea y continúa moviéndose hacia arriba otros 0.07, para completar así un movimiento total de 0.40.

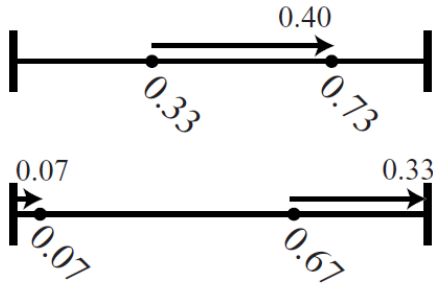


Figura 9.17. Transformación aleatoria de secuencias de Halton con  $\mu=0.40$ .

La figura 9.18 representa la transformación para los cinco primeros elementos de la secuencia. La posición relativa entre los puntos y el grado de cobertura es el mismo antes y después de la transformación. Sin embargo, dado que la transformación se basa en el valor aleatorio  $\mu$ , los valores numéricos de la secuencia transformada son aleatorios. La secuencia resultante se denomina secuencia de Halton aleatorizada. Tiene las mismas propiedades de cobertura y correlación negativa entre observaciones de la secuencia de Halton original, ya que la colocación relativa de los elementos es la misma; sin embargo, ahora la secuencia es aleatoria.

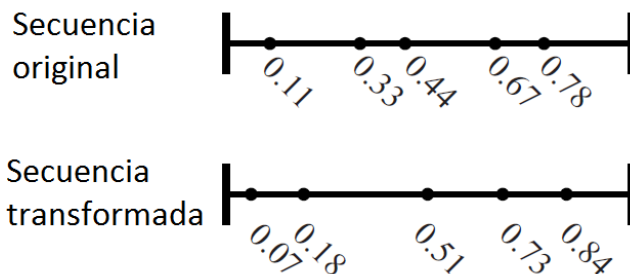


Figura 9.18. Aleatorización de una secuencia de Halton en una dimensión.

Con múltiples dimensiones, la secuencia utilizada para cada dimensión se transforma por separado en función de su propio valor extraído al azar de la densidad uniforme estándar. La figura 9.19 representa una transformación de una secuencia de dos dimensiones de longitud 3 definida para números primos 2 y 3. La secuencia para el número primo 3 está representada en el eje x y obtiene un valor aleatorio de 0.40. La secuencia para el número primo 2 obtiene un valor aleatorio de 0.35. Cada punto en la secuencia bidimensional original se mueve a la derecha 0.40 y hacia arriba 0.35, volviendo al origen cuando es necesario. La posición relativa entre los puntos en cada dimensión se mantiene, y sin embargo la secuencia es ahora aleatoria.

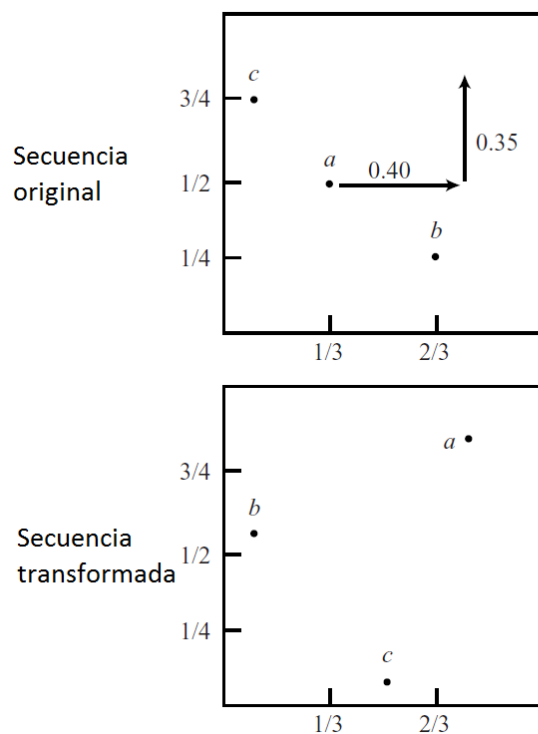


Figura 9.19. Aleatorización de una secuencia de Halton en dos dimensiones.

### 9.3.5 Secuencias de Halton mezcladas

Otro problema con las secuencias de Halton surge cuando se utilizan con muchas dimensiones. Para la simulación de integrales de alta dimensionalidad, se necesitan secuencias de Halton basadas en números primos grandes. Por ejemplo, con 15 dimensiones, se necesitan los números primos existentes hasta 47. Sin embargo, las secuencias de Halton definidas por números primos grandes pueden estar altamente correlacionadas entre sí a través de extensas porciones de la secuencia. La correlación no se limita a los elementos iniciales como se ha descrito anteriormente, por lo que no podemos eliminar la correlación descartando estos elementos. Dos secuencias definidas por números primos grandes y similares se sincronizan periódicamente entre sí y permanecen así durante muchos ciclos.

Bhat (2003) describe el problema y proporciona una solución eficaz. La figura 9.20 reproduce un gráfico incluido en su artículo que representa la secuencia de Halton para los números primos 43 y 47. Claramente, estas secuencias están altamente correlacionadas.

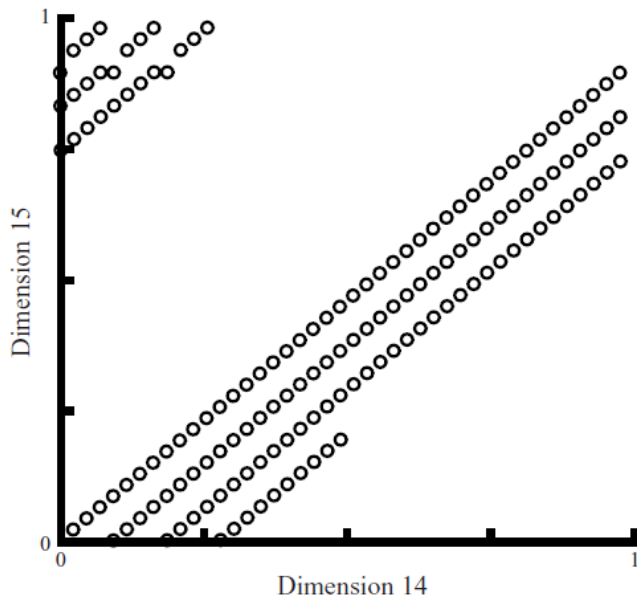


Figura 9.20. Secuencia de Halton estándar.

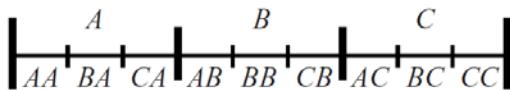


Figura 9.21. Segmentos definidos para mezclar secuencias de Halton.

Esta correlación se puede suprimir, preservando al mismo tiempo la cobertura deseable de las secuencias de Halton, mediante la mezcla de los dígitos de cada elemento de la secuencia. La mezcla se puede hacer de varias maneras. Braatan y Weller (1979) describen un procedimiento que se explica más fácilmente a través de un ejemplo. Considere la secuencia de Halton para el número primo 3:

$$\frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \dots$$

Recordemos que la secuencia se crea dividiendo el intervalo unitario en tres segmentos, que denominamos A, B y C en la figura 9.21. Cada segmento se divide en tres sub-segmentos, etiquetados como AA (indicando sub-segmento A del segmento A), BA (sub-segmento B del segmento A), CA, AB, BB, CB, AC, BC y CC. La secuencia de Halton no mezclada es el punto de inicio de cada segmento en orden alfabético e ignorando A (es decir, ignoramos A,  $\frac{1}{3}$  para B,  $\frac{2}{3}$  para C), seguido por el punto de inicio de cada sub-segmento en orden alfabético e ignorando A (es decir, ignoramos AA, AB y AC,  $\frac{1}{9}$  para BA,  $\frac{4}{9}$  para BB,  $\frac{7}{9}$  para BC,  $\frac{2}{9}$  para CA,  $\frac{5}{9}$  para CB y  $\frac{8}{9}$  para CC). Tenga en cuenta que los segmentos y sub-segmentos que empiezan con A se ignoran debido a que sus puntos de inicio o bien son 0 (para el segmento A) o ya están incluidos en la secuencia (por ejemplo, el punto de inicio del sub-segmento AB es el mismo que el punto de inicio del segmento B).

La secuencia mezclada se obtiene mediante la inversión de B y C, es decir, considerando que C está antes que B en el alfabeto. La lista alfabética pasa a ser: segmentos A C B, sub-segmentos AA AC AB CA

CC CB BA BC BB. La secuencia se crea así de la misma manera que antes pero con este nuevo orden alfabético: ignoramos A,  $\frac{2}{3}$  para C,  $\frac{1}{3}$  para B; ignoramos AA, AC y AB,  $\frac{2}{9}$  para CA,  $\frac{8}{9}$  para CC,  $\frac{5}{9}$  para CB,  $\frac{1}{9}$  para BA,  $\frac{7}{9}$  para BC,  $\frac{4}{9}$  para BB. Las secuencia original y la mezclada son:

original	mezclado
1/3	2/3
2/3	1/3
1/9	2/9
4/9	8/9
7/9	5/9
2/9	1/9
5/9	7/9
8/9	4/9

Diferentes permutaciones de las letras se utilizan para diferentes números primos. La figura 9.22, hecha por Bhat (2003), muestra la secuencia mezclada para los números primos 43 y 47. Los puntos no están correlacionados como sí están en la secuencia original. Bhat demuestra que las secuencias mezcladas funcionan bien para integrales de alta dimensionalidad de la misma manera que las secuencias no mezcladas lo hacen para integrales de baja dimensionalidad.

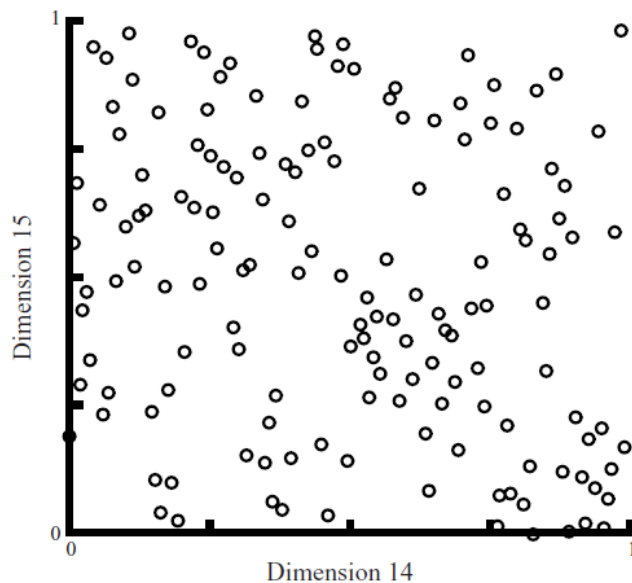


Figura 9.22. Secuencia de Halton mezclada.

### 9.3.6 Otros procedimientos

Hemos descrito sólo unos pocos de los procedimientos antitéticos y cuasi-aleatorios más importantes y directos. Procedimientos más complejos, con propiedades teóricas deseables, son descritos por Niederreiter (1978, 1988), Morokoff y Caflisch (1995), Joe y Sloan (1993), y Sloan y Wozniakowski (1998), por nombrar sólo unos pocos en esta creciente área de investigación. Como hemos visto con las secuencias de Halton, procedimientos bastante simples pueden proporcionar grandes mejoras respecto a la extracción de valores al azar. Las comparaciones realizadas por Sándor y András (2001) sobre probit y Sándor y Train (2004) sobre logit mixtos, indican que la precisión de la estimación basada en la

simulación de modelos de elección discreta se puede mejorar aún más con procedimientos más complejos. Es importante recordar, sin embargo, pese al entusiasmo que despiertan los resultados de estos métodos, que la precisión siempre se puede mejorar mediante el simple uso de más valores. El investigador debe decidir si le conviene aprender y programar nuevos métodos de extracción de valores, dadas sus limitaciones de tiempo, en lugar de simplemente estimar su modelo con más valores.