

Modelos Gráficos Probabilísticos

Tema 4. Problemas de inferencia en modelos gráficos: Cálculo de probabilidades condicionadas de forma aproximada

Andrés Cano Utrera



Curso 2024-2025

1 Introducción

2 Algoritmos de Monte Carlo

- Simulación de variables aleatorias
- Base intuitiva de los métodos de simulación
- Metodología general de los métodos de simulación en RBs
- Muestreo lógico probabilístico
- Método de ponderación por verosimilitud
- Método de muestreo de Markov

Introducción

Definición

Sea $\mathbf{E} \subset \mathbf{X}$ un conjunto de variables observadas y $e \in \Omega_{\mathbf{E}}$ el valor concreto que toman. Un algoritmo que calcula $p(X = x|e)$ para cada $x \in \Omega_X$, $X \in \mathbf{X} \setminus \mathbf{E}$ se llama **algoritmo de propagación exacto**.

Definición

En las mismas condiciones que en la definición anterior, un algoritmo que calcula $\hat{p}(X = x|e)$ para cada $x \in \Omega_X$, $X \in \mathbf{X} \setminus \mathbf{E}$ se llama **algoritmo de propagación aproximado**.

Clasificación de alg. de propagación para RBs

- **Algoritmos exactos:** Problema *NP-duro* (Cooper 90) :-)
- **Algoritmos aproximados:** También *NP-duros* (Dagum y Luby 93) :-)

Tipos de algoritmos aproximados

Métodos de Monte Carlo

Generan una muestra $\{\mathbf{x}^{(j)}\}$, $j = 1, \dots, m$ de configuraciones de las n variables a partir de la distribución conjunta de la red y luego aproximan las probabilidades de cada caso de la variable, $X_k = x_k$, como la frecuencia relativa de dicho caso en la muestra.

Métodos deterministas

Obtienen una **aproximación** para $p(X_k|e)$ que es siempre la **misma en cualquier ejecución** del algoritmo si se usan los **mismos parámetros** de entrada.

Precisión de la aproximación

Dada la distribución de probabilidad $p(\mathbf{X})$ y una aproximación $\hat{p}(\mathbf{X})$, ¿cómo de buena es la aproximación?

Definición

Dados $p(x_i|e)$ y su aproximación $\hat{p}(x_i|e)$, decimos que $\hat{p}(x_i|e)$ tiene un **error absoluto** ϵ para $p(x_i|e)$ si:

$$|p(x_i|e) - \hat{p}(x_i|e)| \leq \epsilon$$

Ejemplo: $p_1 = 0.5$, $p_2 = 0.0001$, $\hat{p}_1 = 0.49$, $\hat{p}_2 = 0.000001$. $\epsilon_1 = 0.01$ y $\epsilon_2 = 0.0001$

Precisión de la aproximación

Definición

Dados $p(x_i|e)$ y su aproximación $\hat{p}(x_i|e)$, decimos que $\hat{p}(x_i|e)$ tiene un error relativo ϵ para $p(x_i|e)$ si:

$$\frac{\hat{p}(x_i|e)}{p(x_i|e)} \in [1 - \epsilon, 1 + \epsilon]$$

Ejemplo: $\epsilon_1 = 0.02$ y $\epsilon_2 = 0.99$

Definición

El error de la distribución completa se suele medir con el root mean square error (RMSE):

$$\sqrt{\sum_{x_i \in \Omega_{X_i}} ((p(x_i|e) - \hat{p}(x_i|e))^2)}$$

Precisión de la aproximación

Dada la distribución de probabilidad $p(\mathbf{X})$ y una aproximación $\hat{p}(\mathbf{X})$, ¿cómo de buena es la aproximación?

Distancia de Kullback Leibler

Dados $p(x_i|e)$ y su aproximación $\hat{p}(x_i|e)$, decimos que $\hat{p}(x_i|e)$ tiene un **la distancia de Kullback-Leibler** para $p(x_i|e)$

$$\sum_{x_i \in \Omega_{X_i}} p(x_i|e) \log\left(\frac{p(x_i|e)}{\hat{p}(x_i|e)}\right)$$

Contenido del tema

- 1 Introducción
- 2 Algoritmos de Monte Carlo

- Simulación de variables aleatorias
 - Base intuitiva de los métodos de simulación
 - Metodología general de los métodos de simulación en RBs
 - Muestreo lógico probabilístico
 - Método de ponderación por verosimilitud
 - Método de muestreo de Markov

Simulación de variables aleatorias

- Para **generar una muestra** procedente de una determinada función de probabilidad $h(x)$, se calcula en primer lugar la **función de distribución** (probabilidad acumulada)

$$H(x) = P(X \leq x) = \int_{-\infty}^x h(x) dx$$

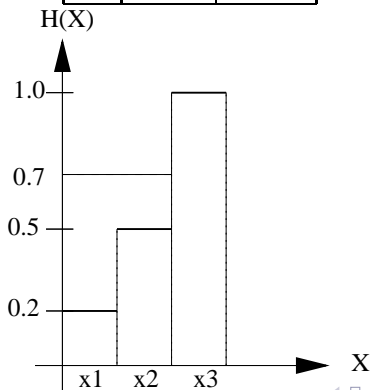
- Generamos ahora una sucesión de números aleatorios $\{u_1, \dots, u_N\}$ de la **distribución uniforme** $U(0, 1)$ obteniendo los valores correspondientes $\{x_1, \dots, x_N\}$ resolviendo la ecuación $H(x_i) = u_i$ con $i = 1, \dots, N$ que nos da

$$x_i = H^{-1}(u_i)$$

Ejemplo de simulación de variables aleatorias

Sea la distribución de probabilidad:

| | $p(X)$ | $H(X)$ |
|-------|--------|--------|
| x_1 | 0.2 | 0.2 |
| x_2 | 0.3 | 0.5 |
| x_3 | 0.5 | 1.0 |



Algoritmo básico de obtención de una muestra

Algorithm 1: Algoritmo básico de obtención de una muestra

Data: Variable X con n estados; Distribución $p(X = x_i)$ denotada por p_i

Result: Una muestra $\{x_1, \dots, x_N\}$ procedente de $p(X)$

```
1 for  $i = 1$  to  $N$  do
2   Generar un valor aleatorio  $u \in U(0, 1)$ ;
3    $P = p_1$ ;
4    $k = 1$ ;
5   while  $k \leq n$  y  $P < u$  do
6      $k = k + 1$ ;
7      $P = P + p_k$ ;
8   Añadir  $x_k$  a la muestra;
9 Devolver  $\{x_1, \dots, x_N\}$  ;
```

Obtención de probabilidades de un suceso mediante simulación de una muestra

Podemos calcular las probabilidades de ciertos sucesos de forma aproximada mediante simulación.

Obtención de probabilidades a partir de una muestra

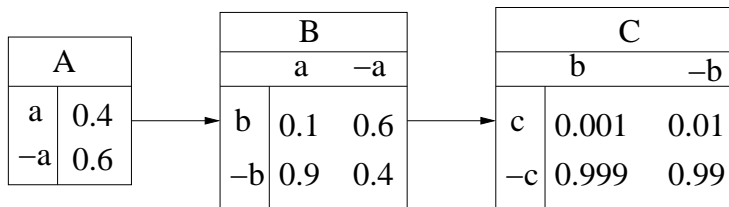
Repetir un experimento N veces obteniendo una *muestra* de tamaño N , obteniendo la probabilidad de un suceso como

$$\frac{\text{Nº veces que ocurre el suceso}}{\text{Nº total de simulaciones}}$$

Ejemplo: Aproximación de probabilidades mediante generación de una muestra

En una red bayesiana podríamos usar el algoritmo anterior para estimar $p(X_i = x_i)$

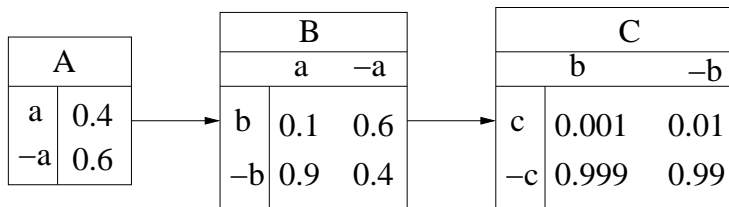
- Generamos una muestra a partir de la distribución conjunta de la red



Ejemplo: Aproximación de probabilidades mediante generación de una muestra

En una red bayesiana podríamos usar el algoritmo anterior para estimar $p(X_i = x_i)$

- Generamos una muestra a partir de la distribución conjunta de la red
- Estimamos las probabilidades $p(X_i = x_i)$ como la proporción de las realizaciones en las que $X_i = x_i$



Ejemplo: Aproximación de probabilidades mediante generación de una muestra

- La función de distribución conjunta y acumulada para cada configuración es:

| Configuración | prob. | prob. acumulada |
|-----------------------------|---------|-----------------|
| a, b, c | 0.00004 | 0.00004 |
| a, b, \bar{c} | 0.03996 | 0.04 |
| a, \bar{b}, c | 0.0036 | 0.0436 |
| a, \bar{b}, \bar{c} | 0.3564 | 0.4 |
| \bar{a}, b, c | 0.00036 | 0.40036 |
| \bar{a}, b, \bar{c} | 0.35964 | 0.76 |
| \bar{a}, \bar{b}, c | 0.0024 | 0.7624 |
| $\bar{a}, \bar{b}, \bar{c}$ | 0.2376 | 1.0 |

Ejemplo: Aproximación de probabilidades mediante generación de una muestra

- La función de distribución conjunta y acumulada para cada configuración es:

| Configuración | prob. | prob. acumulada |
|-----------------------------|---------|-----------------|
| a, b, c | 0.00004 | 0.00004 |
| a, b, \bar{c} | 0.03996 | 0.04 |
| a, \bar{b}, c | 0.0036 | 0.0436 |
| a, \bar{b}, \bar{c} | 0.3564 | 0.4 |
| \bar{a}, b, c | 0.00036 | 0.40036 |
| \bar{a}, b, \bar{c} | 0.35964 | 0.76 |
| \bar{a}, \bar{b}, c | 0.0024 | 0.7624 |
| $\bar{a}, \bar{b}, \bar{c}$ | 0.2376 | 1.0 |

- Consideremos la siguiente serie de números aleatorios uniformes $U(0, 1)$: 0.723, 0.070, 0.531, 0.559, 0.458

Ejemplo: Aproximación de probabilidades mediante generación de una muestra

- La función de distribución conjunta y acumulada para cada configuración es:

| Configuración | prob. | prob. acumulada |
|-----------------------------|---------|-----------------|
| a, b, c | 0.00004 | 0.00004 |
| a, b, \bar{c} | 0.03996 | 0.04 |
| a, \bar{b}, c | 0.0036 | 0.0436 |
| a, \bar{b}, \bar{c} | 0.3564 | 0.4 |
| \bar{a}, b, c | 0.00036 | 0.40036 |
| \bar{a}, b, \bar{c} | 0.35964 | 0.76 |
| \bar{a}, \bar{b}, c | 0.0024 | 0.7624 |
| $\bar{a}, \bar{b}, \bar{c}$ | 0.2376 | 1.0 |

- Consideremos la siguiente serie de números aleatorios uniformes $U(0, 1)$: 0.723, 0.070, 0.531, 0.559, 0.458
- Que nos dan las siguientes 5 realizaciones: (\bar{a}, b, \bar{c}) , (a, \bar{b}, \bar{c}) , (\bar{a}, b, \bar{c}) , (\bar{a}, b, \bar{c}) , (\bar{a}, b, \bar{c})

Ejemplo: Aproximación de probabilidades mediante generación de una muestra

- Las 5 realizaciones anteriores nos permiten obtener:

$$\hat{p}(a) = \frac{1}{5} = 0.2; \quad \hat{p}(b) = \frac{4}{5} = 0.8; \quad \hat{p}(c) = \frac{0}{5} = 0.0$$

Las verdaderas probabilidades son $p(a) = 0.4$, $p(b) = 0.4$ y $p(c) = 0.0064$

Ejemplo: Aproximación de probabilidades mediante generación de una muestra

- Las 5 realizaciones anteriores nos permiten obtener:

$$\hat{p}(a) = \frac{1}{5} = 0.2; \quad \hat{p}(b) = \frac{4}{5} = 0.8; \quad \hat{p}(c) = \frac{0}{5} = 0.0$$

Las verdaderas probabilidades son $p(a) = 0.4$, $p(b) = 0.4$ y $p(c) = 0.0064$

- Esta aproximación produce un error RMSE de 0.6325

Ejemplo: Aproximación de probabilidades mediante generación de una muestra

- Las 5 realizaciones anteriores nos permiten obtener:

$$\hat{p}(a) = \frac{1}{5} = 0.2; \quad \hat{p}(b) = \frac{4}{5} = 0.8; \quad \hat{p}(c) = \frac{0}{5} = 0.0$$

Las verdaderas probabilidades son $p(a) = 0.4$, $p(b) = 0.4$ y $p(c) = 0.0064$

- Esta aproximación produce un error RMSE de 0.6325
- Si generamos otros 5 números aleatorios: 0.817, 0.268, 0.733, 0.301, 0.348 obtenemos las realizaciones $(\bar{a}, \bar{b}, \bar{c})$, (a, \bar{b}, \bar{c}) , (\bar{a}, b, \bar{c}) , (a, \bar{b}, \bar{c}) , (a, \bar{b}, \bar{c})

Ejemplo: Aproximación de probabilidades mediante generación de una muestra

- Las 5 realizaciones anteriores nos permiten obtener:

$$\hat{p}(a) = \frac{1}{5} = 0.2; \quad \hat{p}(b) = \frac{4}{5} = 0.8; \quad \hat{p}(c) = \frac{0}{5} = 0.0$$

Las verdaderas probabilidades son $p(a) = 0.4$, $p(b) = 0.4$ y $p(c) = 0.0064$

- Esta aproximación produce un error RMSE de 0.6325
- Si generamos otros 5 números aleatorios: 0.817, 0.268, 0.733, 0.301, 0.348 obtenemos las realizaciones $(\bar{a}, \bar{b}, \bar{c})$, (a, \bar{b}, \bar{c}) , (\bar{a}, b, \bar{c}) , (a, \bar{b}, \bar{c}) , (a, \bar{b}, \bar{c})
- Con las 10 realizaciones obtenemos:

$$\hat{p}(a) = \frac{4}{10} = 0.4; \quad \hat{p}(b) = \frac{5}{10} = 0.5; \quad \hat{p}(c) = \frac{0}{10} = 0.0$$

Ejemplo: Aproximación de probabilidades mediante generación de una muestra

- Las 5 realizaciones anteriores nos permiten obtener:

$$\hat{p}(a) = \frac{1}{5} = 0.2; \quad \hat{p}(b) = \frac{4}{5} = 0.8; \quad \hat{p}(c) = \frac{0}{5} = 0.0$$

Las verdaderas probabilidades son $p(a) = 0.4$, $p(b) = 0.4$ y $p(c) = 0.0064$

- Esta aproximación produce un error RMSE de 0.6325
- Si generamos otros 5 números aleatorios: 0.817, 0.268, 0.733, 0.301, 0.348 obtenemos las realizaciones $(\bar{a}, \bar{b}, \bar{c})$, (a, \bar{b}, \bar{c}) , (\bar{a}, b, \bar{c}) , (a, \bar{b}, \bar{c}) , (a, \bar{b}, \bar{c})
- Con las 10 realizaciones obtenemos:

$$\hat{p}(a) = \frac{4}{10} = 0.4; \quad \hat{p}(b) = \frac{5}{10} = 0.5; \quad \hat{p}(c) = \frac{0}{10} = 0.0$$

- Y el nuevo error es 0.1417

Contenido del tema

1 Introducción

2 Algoritmos de Monte Carlo

- Simulación de variables aleatorias
- **Base intuitiva de los métodos de simulación**
- Metodología general de los métodos de simulación en RBs
- Muestreo lógico probabilístico
- Método de ponderación por verosimilitud
- Método de muestreo de Markov

Base intuitiva de los métodos de simulación

Sea una **Urna 1 con seis bolas numeradas** $\{1, \dots, 6\}$ y sea X_i el resultado de la extracción i -ésima de la urna haciendo muestreo con *reemplazamiento*. Entonces X_i es una variable *uniforme* con función de probabilidad $p(X_i = x_i) = 1/6$, $x_i = 1, \dots, 6$ y $i = 1, \dots, N$ (donde N es el número de extracciones o tamaño de la muestra).

Base intuitiva de los métodos de simulación

Sea una **Urna 1 con seis bolas numeradas** $\{1, \dots, 6\}$ y sea X_i el resultado de la extracción i -ésima de la urna haciendo muestreo con *reemplazamiento*. Entonces X_i es una variable *uniforme* con función de probabilidad $p(X_i = x_i) = 1/6$, $x_i = 1, \dots, 6$ y $i = 1, \dots, N$ (donde N es el número de extracciones o tamaño de la muestra).

- Si $X = \{X_1, \dots, X_N\}$, entonces la **función de probabilidad conjunta** $p(\mathbf{x})$ se obtiene con

$$p(\mathbf{x}) = \prod_{i=1}^N p(x_i)$$

Base intuitiva de los métodos de simulación

Sea una **Urna 1 con seis bolas numeradas** $\{1, \dots, 6\}$ y sea X_i el resultado de la extracción i -ésima de la urna haciendo muestreo con *reemplazamiento*. Entonces X_i es una variable *uniforme* con función de probabilidad $p(X_i = x_i) = 1/6$, $x_i = 1, \dots, 6$ y $i = 1, \dots, N$ (donde N es el número de extracciones o tamaño de la muestra).

- Si $X = \{X_1, \dots, X_N\}$, entonces la **función de probabilidad conjunta** $p(\mathbf{x})$ se obtiene con

$$p(\mathbf{x}) = \prod_{i=1}^N p(x_i)$$

- En el ejemplo anterior, es sencillo calcular probabilidades de ciertos sucesos tales como $p(X_1 = 1, \dots, X_N = 1)$ o $p(\text{número de pares} = \text{número de impares})$

Base intuitiva de los métodos de simulación

En situaciones más complicadas podemos calcular las probabilidades de ciertos sucesos de forma aproximada mediante simulación.

Simulación de una muestra

Repetir un experimento N veces obteniendo una *muestra* de tamaño N , obteniendo la probabilidad de un suceso como

$$\frac{\text{Nº veces que ocurre el suceso}}{\text{Nº total de simulaciones}}$$

Base intuitiva de los métodos de simulación

En situaciones más complicadas podemos calcular las probabilidades de ciertos sucesos de forma aproximada mediante simulación.

Simulación de una muestra

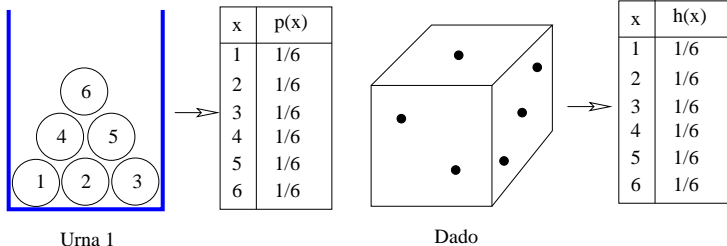
Repetir un experimento N veces obteniendo una *muestra* de tamaño N , obteniendo la probabilidad de un suceso como

$$\frac{\text{Nº veces que ocurre el suceso}}{\text{Nº total de simulaciones}}$$

Una forma equivalente de obtener una muestra con reemplazamiento de tamaño N de la Urna 1 es lanzando un dado N veces.

Base intuitiva de los métodos de simulación

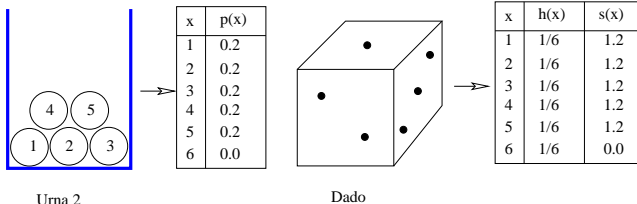
- Sea Y_i el número obtenido al lanzar el dado, entonces Y_i tiene la misma función de probabilidad que X_i
- Llamemos $p(x)$ a la función de probabilidad de X y $h(y)$ a la función de probabilidad de Y ; entonces $p(x) = h(x)$
- Por tanto, extraer N bolas con reemplazamiento de la Urna 1 puede simularse con el lanzamiento de un dado N veces.



- La distribución $p(x)$ se denomina **distribución de la población** y $h(x)$ **distribución de simulación**.

Base intuitiva de los métodos de simulación

- La distribución de simulación se usa porque es más fácil obtener muestras de ésta que de la distribución de la población.
- En el ejemplo anterior la **distribución de la población** $p(x)$ y la **distribución de simulación** $h(x)$ coinciden.
- Pero normalmente $p(x) \neq h(x)$



- Aunque $p(x) \neq h(x)$, sí que podemos usar el dado para simular una muestra de extracción de bolas de Urna 2. Cuando en el dado sale 6, se ignora la tirada y se repite de nuevo: **método de aceptación-rechazo** (ver por ejemplo Rubin-stein (1981)).

Método de aceptación-rechazo

Teorema: Método de aceptación-rechazo

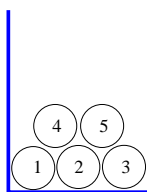
Sea X una variable aleatoria con función de probabilidad $p(x)$.
Supongamos que $p(x)$ puede expresarse como

$$p(x) = c \cdot g(x) \cdot h(x)$$

donde $c \geq 1$, $0 \leq g(x) \leq 1$ y $h(x)$ es una función de probabilidad.
Sea U una variable aleatoria uniforme $U(0, 1)$ y sea Y una variable aleatoria con función de probabilidad $h(y)$ independiente de U . Entonces, la función de probabilidad condicional de Y dado que $u \leq g(y)$ coincide con la función de probabilidad de X . Por otra parte, la probabilidad de aceptar la muestra (eficiencia) es $1/c$.

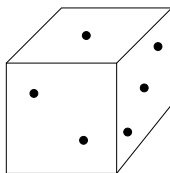
La probabilidad $1/c$ de aceptar la muestra es alta cuando $h(x)$ es próxima a $p(x)$

Ejemplo: Obtención de una muestra de la Urna 2



Urna 2

| x | p(x) |
|---|------|
| 1 | 0.2 |
| 2 | 0.2 |
| 3 | 0.2 |
| 4 | 0.2 |
| 5 | 0.2 |
| 6 | 0.0 |



Dado

| x | h(x) | s(x) |
|---|------|------|
| 1 | 1/6 | 1.2 |
| 2 | 1/6 | 1.2 |
| 3 | 1/6 | 1.2 |
| 4 | 1/6 | 1.2 |
| 5 | 1/6 | 1.2 |
| 6 | 1/6 | 0.0 |

En el caso de la Urna 2 se puede escribir $p(x) = c \cdot g(x) \cdot h(x)$, con $c = 6/5$ y

$$g(x) = \begin{cases} 0, & \text{si } x = 6 \\ 1, & \text{en otro caso} \end{cases}$$

Se puede obtener una muestra de $p(x)$ (Urna 2), usando $h(x)$ (el dado) y comprobando la condición $u \leq g(x)$ para todo valor x que se simule de $h(x)$, donde u es un número obtenido de $U(0, 1)$.

El suceso $x = 6$ siempre se rechaza, ya que $g(6) = 0$

Ejemplo del método de aceptación-rechazo (caso continuo)

- Supongamos que queremos obtener una muestra de tamaño N de una distribución con función de densidad $p(x) = 3x(2 - x)/4$, $0 \leq x \leq 2$:
- Si $h(x) = 1/2$, $0 \leq x \leq 2$, $g(x) = x(2 - x)$ y $c = 3/2$

$$p(x) = c \cdot g(x) \cdot h(x)$$

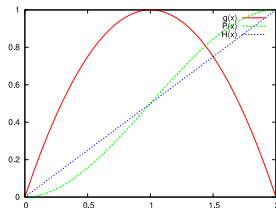
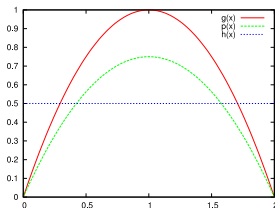
- Las funciones de distribución de $h(x)$ y $p(x)$ son

$$H(x) = \int_0^x h(x)dx = \int_0^x (1/2)dx = x/2$$

$$P(x) = 3/4x^2 - x^3/4$$

- Es más fácil simular con $H(x)$ que con $P(X)$

Ejemplo del método de aceptación-rechazo (caso continuo)



- Generamos un número aleatorio y de $h(y)$ y un u de $U(0,1)$.
- Si $u \leq g(y)$ se acepta y como número procedente de $p(x)$.
En caso contrario se rechazan u e y

Por ejemplo,

- Si $y = 1.5$: $g(y) = 0.75$
- Si $y = 1$: $g(y) = 1$

En este caso la probabilidad $1/3$ de aceptar un número generado con $h(y)$ es $2/3$.

Algoritmo del método de aceptación-rechazo

Algorithm 2: El Método de rechazo

Data: Funciones $p(x)$, $h(x)$ y tamaño muestra N

Result: Una muestra $\{x_1, \dots, x_N\}$ procedente de $p(x)$

```
1 for  $i = 1$  to  $N$  do
2   Generar un valor aleatorio  $u$  de la distribución  $U(0, 1)$ ;
3   Generar un valor aleatorio  $y$  de  $h(x)$  ;
4   if  $u \leq g(y)$  then
5     | hacer  $x_i = y$ ;
6   else
7     | Ir a la etapa 2;
8 Devolver  $\{x_1, \dots, x_N\}$  ;
```

Modificación del método: Muestreo por importancia

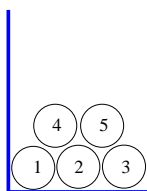
- Cuando c es grande, la eficiencia del algoritmo es baja.
- El algoritmo puede hacerse más eficiente haciendo:

$$p(x) = \frac{p(x)}{h(x)} h(x) = s(x) h(x)$$

donde $s(x) = p(x)/h(x)$ es una **función peso**.

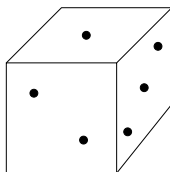
- Puede deducirse que $s(x) = c \cdot g(x)$
- Ahora, en vez de rechazar un número x generado de $h(x)$, se le asigna una probabilidad proporcional a $s(x)$ o $g(x)$, normalizando los pesos al final de las simulaciones para obtener la probabilidad de cualquier suceso de interés.

Ejemplos usando muestreo por importancia



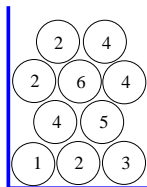
Urna 2

| x | p(x) |
|---|------|
| 1 | 0.2 |
| 2 | 0.2 |
| 3 | 0.2 |
| 4 | 0.2 |
| 5 | 0.2 |
| 6 | 0.0 |



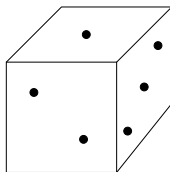
Dado

| x | h(x) | s(x) |
|---|------|------|
| 1 | 1/6 | 1.2 |
| 2 | 1/6 | 1.2 |
| 3 | 1/6 | 1.2 |
| 4 | 1/6 | 1.2 |
| 5 | 1/6 | 1.2 |
| 6 | 1/6 | 0.0 |



Urna 3

| x | p(x) |
|---|------|
| 1 | 0.1 |
| 2 | 0.3 |
| 3 | 0.1 |
| 4 | 0.3 |
| 5 | 0.1 |
| 6 | 0.1 |



Dado

| x | h(x) | s(x) |
|---|------|------|
| 1 | 1/6 | 0.6 |
| 2 | 1/6 | 1.8 |
| 3 | 1/6 | 0.6 |
| 4 | 1/6 | 1.8 |
| 5 | 1/6 | 0.6 |
| 6 | 1/6 | 0.6 |

Contenido del tema

- 1 Introducción
- 2 Algoritmos de Monte Carlo
 - Simulación de variables aleatorias
 - Base intuitiva de los métodos de simulación
 - Metodología general de los métodos de simulación en RBs
 - Muestreo lógico probabilístico
 - Método de ponderación por verosimilitud
 - Método de muestreo de Markov

El problema en el caso de Redes Bayesianas

Como hemos visto, en **redes bayesianas** el problema es el siguiente:

- Sea $\mathbf{X} = \{X_1, \dots, X_n\}$ un conjunto de variables con distribución de probabilidad conjunta $p(\mathbf{x})$.
- Supongamos que el conjunto de variables evidenciales \mathbf{E} toman el valor e
- Nuestro objetivo es calcular $p(X = x|e)$ para cada $x \in \Omega_X$
- En general, dado $\mathbf{Y} \subset \mathbf{X}$, podemos calcular $p(\mathbf{Y} = y|e)$ para cada $y \in \Omega_Y$
- $p(\mathbf{Y} = y|e)$ puede escribirse como

$$p(\mathbf{Y} = y|e) = \frac{p_e(y)}{p(e)} \propto p_e(y)$$

donde

$$p_e(y) = \begin{cases} p(y \cup e), & \text{si } y \text{ es consistente con } e \\ 0, & \text{en otro caso} \end{cases}$$

Metodología general de los métodos de simulación en RBs

El cálculo aproximado de $p(x_i|e)$ lo haremos mediante:

- Generar una muestra $x^j = \{x_1^j, \dots, x_n^j\}$ de tamaño N de $p(x)$ pero usando la función $h(x)$
- Calcular y normalizar los pesos
- Aproximar $p(x_i|e)$ mediante la suma de todos los pesos de las realizaciones consistentes con los sucesos x_i y e .

$$p(x) \approx \frac{\sum_{x \in x^j} s(x^j)}{\sum_{j=1}^N s(x^j)}$$

Metodología general de los métodos de simulación en RBs

El **cálculo aproximado** de $p(x_i|e)$ lo haremos mediante:

- Generar una muestra $x^j = \{x_1^j, \dots, x_n^j\}$ de tamaño N de $p(x)$ pero usando la función $h(x)$
- Calcular y normalizar los pesos
- Aproximar $p(x_i|e)$ mediante la suma de todos los pesos de las realizaciones consistentes con los sucesos x_i y e .

$$p(x) \approx \frac{\sum_{x \in x^j} s(x^j)}{\sum_{j=1}^N s(x^j)}$$

La **calidad de la aproximación** depende de:

- La función $h(x)$ elegida para obtener la muestra
- Método usado para obtener realizaciones de $h(x)$ (pesos similares)
- Tamaño de la muestra N

Metodología general de los métodos de simulación en RBs

Algorithm 3: Algoritmo general de simulación

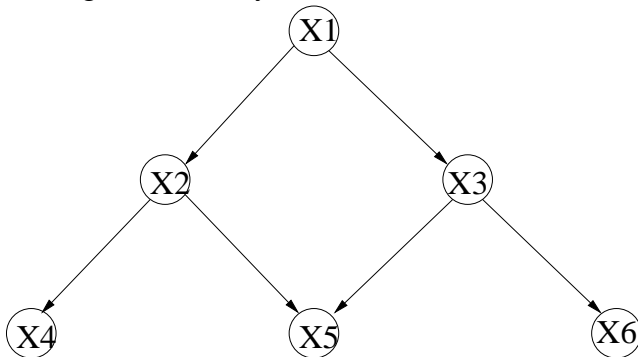
Data: Funciones de probabilidad real $p(x)$ y de simulación $h(x)$, tamaño de la muestra N , y un subconjunto $\mathbf{Y} \subset \mathbf{X}$

Result: Una aproximación de $p(y)$ para todo $y \in \Omega_Y$

- 1 **for** $j = 1$ to N **do**
 - 2 Generar $x^j = (x_1^j, \dots, x_n^j)$ usando $h(x)$;
 - 3 Calcular $s(x^j) = \frac{p(x^j)}{h(x^j)}$;
 - 4 Para cada $y \in \Omega_Y$, aproximar $p(y)$ usando $p(y) \approx \frac{\sum_{y \in x^j} s(x^j)}{\sum_{j=1}^N s(x^j)}$
-

Ejemplo de aproximación en red bayesiana

Supongamos la siguiente red bayesiana con todas sus variables binarias



Ejemplo de aproximación en red bayesiana

Distribuciones de probabilidad condicional

| x_1 | $p(x_1)$ | x_1 | x_2 | $p(x_2 x_1)$ | x_1 | x_3 | $p(x_3 x_1)$ | x_2 | x_4 | $p(x_4 x_2)$ |
|-------|----------|-------|-------|--------------|-------|-------|--------------|-------|-------|--------------|
| 0 | 0.3 | 0 | 0 | 0.4 | 0 | 0 | 0.2 | 0 | 0 | 0.3 |
| 1 | 0.7 | 0 | 1 | 0.6 | 0 | 1 | 0.8 | 0 | 1 | 0.7 |
| | | 1 | 0 | 0.1 | 1 | 0 | 0.5 | 1 | 0 | 0.2 |
| | | 1 | 1 | 0.9 | 1 | 1 | 0.5 | 1 | 1 | 0.8 |

| x_3 | x_6 | $p(x_6 x_3)$ | x_2 | x_3 | x_5 | $p(x_5 x_2, x_3)$ |
|-------|-------|--------------|-------|-------|-------|-------------------|
| 0 | 0 | 0.1 | 0 | 0 | 0 | 0.4 |
| 0 | 1 | 0.9 | 0 | 0 | 1 | 0.6 |
| 1 | 0 | 0.4 | 0 | 1 | 0 | 0.5 |
| 1 | 1 | 0.6 | 0 | 1 | 1 | 0.5 |
| | | | 1 | 0 | 0 | 0.7 |
| | | | 1 | 0 | 1 | 0.3 |
| | | | 1 | 1 | 0 | 0.2 |
| | | | 1 | 1 | 1 | 0.8 |

Ejemplo de aproximación en red bayesiana

Generaremos una muestra de N realizaciones $x^j = \{x_1^j, \dots, x_6^j\}$ (2^6 posibles realizaciones)

Supongamos que seleccionamos cinco de las 64 realizaciones al azar y con reemplazamiento.

- Distribución conjunta:

$$p(\mathbf{x}) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2)p(x_5|x_2, x_3)p(x_6|x_3)$$

- Distribución de simulación:

$$h(x^j) = 1/64; j = 1, \dots, 5$$

Ejemplo de aproximación en red bayesiana

Tabla con distribuciones real, de simulación y pesos

| Realización x^j | $p(x^j)$ | $h(x^j)$ | $s(x^j)$ |
|----------------------------|----------|----------|----------|
| $x^1 = (0, 1, 1, 1, 0, 0)$ | 0.0092 | 1/64 | 0.5898 |
| $x^2 = (1, 1, 0, 1, 1, 0)$ | 0.0076 | 1/64 | 0.4838 |
| $x^3 = (0, 0, 1, 0, 0, 1)$ | 0.0086 | 1/64 | 0.5529 |
| $x^4 = (1, 0, 0, 1, 1, 0)$ | 0.0015 | 1/64 | 0.0941 |
| $x^5 = (1, 0, 0, 0, 1, 1)$ | 0.0057 | 1/64 | 0.3629 |

Ejemplo de aproximación en red bayesiana

Tabla con distribuciones real, de simulación y pesos

| Realización x^j | $p(x^j)$ | $h(x^j)$ | $s(x^j)$ |
|----------------------------|----------|----------|----------|
| $x^1 = (0, 1, 1, 1, 0, 0)$ | 0.0092 | 1/64 | 0.5898 |
| $x^2 = (1, 1, 0, 1, 1, 0)$ | 0.0076 | 1/64 | 0.4838 |
| $x^3 = (0, 0, 1, 0, 0, 1)$ | 0.0086 | 1/64 | 0.5529 |
| $x^4 = (1, 0, 0, 1, 1, 0)$ | 0.0015 | 1/64 | 0.0941 |
| $x^5 = (1, 0, 0, 0, 1, 1)$ | 0.0057 | 1/64 | 0.3629 |

Ejemplos suponiendo que no hay evidencia

- $$p(X_1 = 0) \approx \frac{s(x^1) + s(x^3)}{\sum_{i=1}^5 s(x^i)} = \frac{0.5898 + 0.5529}{2.0835} = 0.5485$$

Ejemplo de aproximación en red bayesiana

Tabla con distribuciones real, de simulación y pesos

| Realización x^j | $p(x^j)$ | $h(x^j)$ | $s(x^j)$ |
|----------------------------|----------|----------|----------|
| $x^1 = (0, 1, 1, 1, 0, 0)$ | 0.0092 | 1/64 | 0.5898 |
| $x^2 = (1, 1, 0, 1, 1, 0)$ | 0.0076 | 1/64 | 0.4838 |
| $x^3 = (0, 0, 1, 0, 0, 1)$ | 0.0086 | 1/64 | 0.5529 |
| $x^4 = (1, 0, 0, 1, 1, 0)$ | 0.0015 | 1/64 | 0.0941 |
| $x^5 = (1, 0, 0, 0, 1, 1)$ | 0.0057 | 1/64 | 0.3629 |

Ejemplos suponiendo que no hay evidencia

- $$p(X_1 = 0) \approx \frac{s(x^1) + s(x^3)}{\sum_{i=1}^5 s(x^i)} = \frac{0.5898 + 0.5529}{2.0835} = 0.5485$$
- $$p(X_2 = 0) \approx \frac{s(x^3) + s(x^4) + s(x^5)}{\sum_{i=1}^5 s(x^i)} = \frac{0.5529 + 0.0941 + 0.3629}{2.0835} = 0.4847$$

Ejemplo de aproximación en red bayesiana

Ejemplos suponiendo $e = \{X_3 = 1, X_4 = 1\}$

- Ahora $h(x^j) = 1/16$
- Seleccionemos al azar cinco de las 16 realizaciones

Tabla con distribuciones real, de simulación y pesos

| Realización x^j | $p(x^j)$ | $h(x^j)$ | $s(x^j)$ |
|----------------------------|----------|----------|----------|
| $x^1 = (0, 1, 1, 1, 0, 1)$ | 0.0138 | 1/16 | 0.2212 |
| $x^2 = (1, 0, 1, 1, 0, 0)$ | 0.0049 | 1/16 | 0.0784 |
| $x^3 = (1, 0, 1, 1, 1, 1)$ | 0.0073 | 1/16 | 0.1176 |
| $x^4 = (0, 1, 1, 1, 1, 1)$ | 0.0369 | 1/16 | 0.5898 |
| $x^5 = (1, 1, 1, 1, 0, 0)$ | 0.0202 | 1/16 | 0.3226 |

- $p(X_1 = 0 | X_3 = 1, X_4 = 1) \approx \frac{s(x^1) + s(x^4)}{\sum_{i=1}^5 s(x^i)} = \frac{0.2212 + 0.5898}{1.3296} = 0.6099$
- $p(X_2 = 0 | X_3 = 1, X_4 = 1) \approx \frac{s(x^2) + s(x^3)}{\sum_{i=1}^5 s(x^i)} = \frac{0.0784 + 0.1176}{1.3296} = 0.1474$

Ejemplo de aproximación en red bayesiana

Ejemplos suponiendo $e = \{X_3 = 1, X_4 = 1\}$

- Ahora $h(x^j) = 1/16$
- Seleccionemos al azar cinco de las 16 realizaciones

Tabla con distribuciones real, de simulación y pesos

| Realización x^j | $p(x^j)$ | $h(x^j)$ | $s(x^j)$ |
|----------------------------|----------|----------|----------|
| $x^1 = (0, 1, 1, 1, 0, 1)$ | 0.0138 | 1/16 | 0.2212 |
| $x^2 = (1, 0, 1, 1, 0, 0)$ | 0.0049 | 1/16 | 0.0784 |
| $x^3 = (1, 0, 1, 1, 1, 1)$ | 0.0073 | 1/16 | 0.1176 |
| $x^4 = (0, 1, 1, 1, 1, 1)$ | 0.0369 | 1/16 | 0.5898 |
| $x^5 = (1, 1, 1, 1, 0, 0)$ | 0.0202 | 1/16 | 0.3226 |

- $p(X_1 = 0 | X_3 = 1, X_4 = 1) \approx \frac{s(x^1) + s(x^4)}{\sum_{i=1}^5 s(x^i)} = \frac{0.2212 + 0.5898}{1.3296} = 0.6099$
- $p(X_2 = 0 | X_3 = 1, X_4 = 1) \approx \frac{s(x^2) + s(x^3)}{\sum_{i=1}^5 s(x^i)} = \frac{0.0784 + 0.1176}{1.3296} = 0.1474$

Ejemplo de aproximación en red bayesiana

Ejemplo para aproximar distribuciones multivariadas

Tabla con distribuciones real, de simulación y pesos

| Realización x^j | $p(x^j)$ | $h(x^j)$ | $s(x^j)$ |
|----------------------------|----------|----------|----------|
| $x^1 = (0, 1, 1, 1, 0, 0)$ | 0.0092 | 1/64 | 0.5898 |
| $x^2 = (1, 1, 0, 1, 1, 0)$ | 0.0076 | 1/64 | 0.4838 |
| $x^3 = (0, 0, 1, 0, 0, 1)$ | 0.0086 | 1/64 | 0.5529 |
| $x^4 = (1, 0, 0, 1, 1, 0)$ | 0.0015 | 1/64 | 0.0941 |
| $x^5 = (1, 0, 0, 0, 1, 1)$ | 0.0057 | 1/64 | 0.3629 |

- $p(X_5 = 0, X_6 = 0) \approx \frac{s(x^1)}{\sum_{i=1}^5 s(x^i)} = \frac{0.5898}{2.0835} = 0.2831$
- $p(X_5 = 0, X_6 = 1) \approx \frac{s(x^3)}{\sum_{i=1}^5 s(x^i)} = \frac{0.5529}{2.0835} = 0.2654$
- $p(X_5 = 1, X_6 = 0) \approx \frac{s(x^2)+s(x^4)}{\sum_{i=1}^5 s(x^i)} = \frac{0.4838+0.0941}{2.0835} = 0.2773$
- $p(X_5 = 1, X_6 = 1) \approx \frac{s(x^5)}{\sum_{i=1}^5 s(x^i)} = \frac{0.3629}{2.0835} = 0.1742$

Factorización de la función peso con el algoritmo 3

- Especialmente útil es el caso de que la distribución real y la de simulación puedan factorizarse (como en redes bayesianas)

$$p(x) = \prod_{i=1}^n p(x_i | \pi_i)$$

$$h(x) = \prod_{i=1}^n h(x_i | \pi_i)$$

- De esta forma el peso puede también factorizarse:

$$s(x) = \frac{p(x)}{h(x)} = \prod \frac{p(x_i | \pi_i)}{h(x_i | \pi_i)} = \prod s(x_i | \pi_i)$$

Factorización de la función peso con el algoritmo 3

- Cuando hay evidencia usaremos $p_e(x)$ en lugar de $p(x)$

$$p_e(x) = \begin{cases} p(x \cup e), & \text{si } x \text{ es consistente con } e \\ 0, & \text{en otro caso} \end{cases}$$

Que también puede factorizarse teniendo en cuenta:

$$p_e(x) \propto \prod_{i=1}^n p_e(x_i | \pi_i)$$

donde

$$p_e(x_i | \pi_i) = \begin{cases} p(x_i | \pi_i), & \text{si } x_i \text{ y } \pi_i \text{ son consistentes con } e \\ 0, & \text{en otro caso} \end{cases}$$

Métodos basados en muestreo por importancia

Todo método de simulación consta de **tres componentes**:

- Una distribución de simulación $h(x)$ usada para generar la muestra
- Un método para obtener realizaciones de $h(x)$
- Una fórmula para calcular los pesos

Ejemplos de estos métodos:

- Muestreo lógico probabilístico (Henrion, 1988)
- Ponderación por verosimilitud (Fung y Chang, 1990; Shachter y Peot, 1990)
- Muestreo de Markov (Pearl, 1987)

Contenido del tema

1 Introducción

2 Algoritmos de Monte Carlo

- Simulación de variables aleatorias
- Base intuitiva de los métodos de simulación
- Metodología general de los métodos de simulación en RBs
- **Muestreo lógico probabilístico**
- Método de ponderación por verosimilitud
- Método de muestreo de Markov

Muestreo lógico probabilístico (Henrion,1988)

- Está basado en el teorema de aceptación-rechazo.
- Es uno de los métodos de **propagación hacia adelante**.
*Se muestrea una variable solo cuando ya han sido muestreados todos sus padres (**ordenación ancestral de los nodos**).*
- Se simulan todas las variables, incluso las evidenciales.
- La función de simulación usada para X_i es su función de probabilidad condicional:

$$h(x_i|\pi_i) = p(x_i|\pi_i), \quad i \in \{1, \dots, n\}$$

Muestreo lógico probabilístico (Henrion,1988)

- Los **pesos** se obtienen con:

$$\begin{aligned}
 s(x) &= \frac{p_e(x)}{h(x)} = \frac{\prod_{X_i \notin \mathbf{E}} p_e(x_i | \pi_i) \prod_{X_i \in \mathbf{E}} p_e(x_i | \pi_i)}{\prod_{X_i \notin \mathbf{E}} p(x_i | \pi_i) \prod_{X_i \in \mathbf{E}} p(x_i | \pi_i)} = \\
 &= \begin{cases} 1, & \text{si } x_i = e_i \ \forall X_i \in \mathbf{E} \\ 0, & \text{en otro caso} \end{cases} \quad (1)
 \end{aligned}$$

- Puede verse que si $x_i \neq e_i$ para algún $X_i \in \mathbf{E}$, entonces el peso es cero: **se rechaza la muestra**.
- Las probabilidades condicionales $p(x_i | e)$ se aproximan como el cociente entre los casos no rechazados consistentes con x_i y el número de casos totales N (no rechazados).
- Cuando $p(e)$ es pequeño el método es **muy ineficiente** porque hay un gran porcentaje de rechazo.

Muestreo lógico probabilístico (Henrion,1988)

Algorithm 4: Algoritmo de muestreo lógico

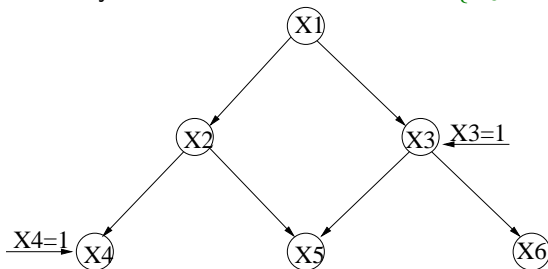
```

1 Ordenar los nodos ancestralmente;
2 for  $j = 1$  to  $N$  do
3   for  $i = 1$  to  $n$  do
4      $x_i =$  valor generado a partir de  $p(x_i|\pi_i)$ ;
5     if  $X_i \in \mathbf{E}$  y  $x_i \neq e_i$  then
6       Repetir el ciclo  $i$ ;

```

Ejemplo de muestreo lógico probabilístico

Sea la siguiente red bayesiana con la evidencia $\mathbf{E} = \{X_3 = 1, X_4 = 1\}$:



Ejemplo de muestreo lógico probabilístico

Distribuciones de probabilidad condicional

| x_1 | $p(x_1)$ | x_1 | x_2 | $p(x_2 x_1)$ | x_1 | x_3 | $p(x_3 x_1)$ | x_2 | x_4 | $p(x_4 x_2)$ |
|-------|----------|-------|-------|--------------|-------|-------|--------------|-------|-------|--------------|
| 0 | 0.3 | 0 | 0 | 0.4 | 0 | 0 | 0.2 | 0 | 0 | 0.3 |
| 1 | 0.7 | 0 | 1 | 0.6 | 0 | 1 | 0.8 | 0 | 1 | 0.7 |
| | | 1 | 0 | 0.1 | 1 | 0 | 0.5 | 1 | 0 | 0.2 |
| | | 1 | 1 | 0.9 | 1 | 1 | 0.5 | 1 | 1 | 0.8 |

| x_3 | x_6 | $p(x_6 x_3)$ | x_2 | x_3 | x_5 | $p(x_5 x_2, x_3)$ |
|-------|-------|--------------|-------|-------|-------|-------------------|
| 0 | 0 | 0.1 | 0 | 0 | 0 | 0.4 |
| 0 | 1 | 0.9 | 0 | 0 | 1 | 0.6 |
| 1 | 0 | 0.4 | 0 | 1 | 0 | 0.5 |
| 1 | 1 | 0.6 | 0 | 1 | 1 | 0.5 |
| | | | 1 | 0 | 0 | 0.7 |
| | | | 1 | 0 | 1 | 0.3 |
| | | | 1 | 1 | 0 | 0.2 |
| | | | 1 | 1 | 1 | 0.8 |

Ejemplo de muestreo lógico probabilístico

- Elegimos un **orden ancestral** de los nodos: $\{X_1, X_2, X_3, X_4, X_5, X_6\}$

Ejemplo de muestreo lógico probabilístico

- Elegimos un **orden ancestral** de los nodos: $\{X_1, X_2, X_3, X_4, X_5, X_6\}$
- Muestreamos X_1 con $p(X_1)$

| x_1 | $p(x_1)$ |
|-------|----------|
| 0 | 0.3 |
| 1 | 0.7 |

Suponer que obtenemos que $X_1 = 1$

Ejemplo de muestreo lógico probabilístico

- Elegimos un **orden ancestral** de los nodos: $\{X_1, X_2, X_3, X_4, X_5, X_6\}$
- Muestreamos X_1 con $p(X_1)$

| x_1 | $p(x_1)$ |
|-------|----------|
| 0 | 0.3 |
| 1 | 0.7 |

Suponer que obtenemos que $X_1 = 1$

- Muestreamos X_2 usando $p(X_2|X_1 = 1)$

| x_1 | x_2 | $p(x_2 x_1)$ |
|-------|-------|--------------|
| 0 | 0 | 0.4 |
| 0 | 1 | 0.6 |
| 1 | 0 | 0.1 |
| 1 | 1 | 0.9 |

Suponer que obtenemos $X_2 = 0$

Ejemplo de muestreo lógico probabilístico

- Muestreamos X_3 usando $p(X_3|X_1 = 1)$

| x_1 | x_3 | $p(x_3 x_1)$ |
|-------|-------|--------------|
| 0 | 0 | 0.2 |
| 0 | 1 | 0.8 |
| 1 | 0 | 0.5 |
| 1 | 1 | 0.5 |

Ejemplo de muestreo lógico probabilístico

- Muestreamos X_3 usando $p(X_3|X_1 = 1)$

| x_1 | x_3 | $p(x_3 x_1)$ |
|-------|-------|--------------|
| 0 | 0 | 0.2 |
| 0 | 1 | 0.8 |
| 1 | 0 | 0.5 |
| 1 | 1 | 0.5 |

- Si obtuviesemos $X_3 = 0$ se rechazaría esta realización completa y se comenzaría desde el principio.

Ejemplo de muestreo lógico probabilístico

- Muestreamos X_3 usando $p(X_3|X_1 = 1)$

| x_1 | x_3 | $p(x_3 x_1)$ |
|-------|-------|--------------|
| 0 | 0 | 0.2 |
| 0 | 1 | 0.8 |
| 1 | 0 | 0.5 |
| 1 | 1 | 0.5 |

- Si obtuviesemos $X_3 = 0$ se rechazaría esta realización completa y se comenzaría desde el principio.
- Supongamos que obtenemos $X_3 = 1$

Ejemplo de muestreo lógico probabilístico

- Muestreamos X_4 , X_5 y X_6 de forma similar

| x_2 | x_4 | $p(x_4 x_2)$ |
|-------|-------|----------------|
| 0 | 0 | 0.3 |
| 0 | 1 | 0.7 |
| 1 | 0 | 0.2 |
| 1 | 1 | 0.8 |

| x_3 | x_6 | $p(x_6 x_3)$ |
|-------|-------|----------------|
| 0 | 0 | 0.1 |
| 0 | 1 | 0.9 |
| 1 | 0 | 0.4 |
| 1 | 1 | 0.6 |

| x_2 | x_3 | x_5 | $p(x_5 x_2, x_3)$ |
|-------|-------|-------|---------------------|
| 0 | 0 | 0 | 0.4 |
| 0 | 0 | 1 | 0.6 |
| 0 | 1 | 0 | 0.5 |
| 0 | 1 | 1 | 0.5 |
| 1 | 0 | 0 | 0.7 |
| 1 | 0 | 1 | 0.3 |
| 1 | 1 | 0 | 0.2 |
| 1 | 1 | 1 | 0.8 |

Ejemplo de muestreo lógico probabilístico

- Muestreamos X_4 , X_5 y X_6 de forma similar

| x_2 | x_4 | $p(x_4 x_2)$ | x_3 | x_6 | $p(x_6 x_3)$ | x_5 | $p(x_5 x_2, x_3)$ |
|-------|-------|--------------|-------|-------|--------------|-------|-------------------|
| 0 | 0 | 0.3 | 0 | 0 | 0.1 | 0 | 0.4 |
| 0 | 1 | 0.7 | 0 | 1 | 0.9 | 0 | 0.6 |
| 1 | 0 | 0.2 | 1 | 0 | 0.4 | 0 | 0.5 |
| 1 | 1 | 0.8 | 1 | 1 | 0.6 | 1 | 0.5 |
| | | | | | | 0 | 0.7 |
| | | | | | | 1 | 0.3 |
| | | | | | | 0 | 0.2 |
| | | | | | | 1 | 0.8 |

- Supongamos que la primera realización es $x^1 = (1, 0, 1, 1, 1, 0)$

Ejemplo de muestreo lógico probabilístico

- Muestreamos X_4 , X_5 y X_6 de forma similar

| x_2 | x_4 | $p(x_4 x_2)$ | x_3 | x_6 | $p(x_6 x_3)$ | x_2 | x_3 | x_5 | $p(x_5 x_2, x_3)$ |
|-------|-------|--------------|-------|-------|--------------|-------|-------|-------|-------------------|
| 0 | 0 | 0.3 | 0 | 0 | 0.1 | 0 | 0 | 0 | 0.4 |
| 0 | 1 | 0.7 | 0 | 1 | 0.9 | 0 | 0 | 1 | 0.6 |
| 1 | 0 | 0.2 | 1 | 0 | 0.4 | 0 | 1 | 0 | 0.5 |
| 1 | 1 | 0.8 | 1 | 1 | 0.6 | 0 | 1 | 1 | 0.5 |
| | | | | | | 1 | 0 | 0 | 0.7 |
| | | | | | | 1 | 0 | 1 | 0.3 |
| | | | | | | 1 | 1 | 0 | 0.2 |
| | | | | | | 1 | 1 | 1 | 0.8 |

- Supongamos que la primera realización es $x^1 = (1, 0, 1, 1, 1, 0)$
- El proceso se repite hasta obtener N realizaciones.

Ejemplo de muestreo lógico probabilístico

- Muestreamos X_4 , X_5 y X_6 de forma similar

| x_2 | x_4 | $p(x_4 x_2)$ | x_3 | x_6 | $p(x_6 x_3)$ | x_2 | x_3 | x_5 | $p(x_5 x_2, x_3)$ |
|-------|-------|--------------|-------|-------|--------------|-------|-------|-------|-------------------|
| 0 | 0 | 0.3 | 0 | 0 | 0.1 | 0 | 0 | 0 | 0.4 |
| 0 | 1 | 0.7 | 0 | 1 | 0.9 | 0 | 0 | 1 | 0.6 |
| 1 | 0 | 0.2 | 1 | 0 | 0.4 | 0 | 1 | 0 | 0.5 |
| 1 | 1 | 0.8 | 1 | 1 | 0.6 | 0 | 1 | 1 | 0.5 |
| | | | | | | 1 | 0 | 0 | 0.7 |
| | | | | | | 1 | 0 | 1 | 0.3 |
| | | | | | | 1 | 1 | 0 | 0.2 |
| | | | | | | 1 | 1 | 1 | 0.8 |

- Supongamos que la primera realización es $x^1 = (1, 0, 1, 1, 1, 0)$
- El proceso se repite hasta obtener N realizaciones.
- La distribución de probabilidad de cualquier variable se aproxima por el **porcentaje de realizaciones** en las que ocurre el suceso de interés.

Contenido del tema

1 Introducción

2 Algoritmos de Monte Carlo

- Simulación de variables aleatorias
- Base intuitiva de los métodos de simulación
- Metodología general de los métodos de simulación en RBs
- Muestreo lógico probabilístico
- **Método de ponderación por verosimilitud**
- Método de muestreo de Markov

Método de ponderación por verosimilitud

- Fue desarrollado independientemente por Fung y Chang (1990) y Shachter y Peot (1990)
- Funciona de forma parecida al algoritmo de muestreo lógico, pero cuando llega a un nodo observado $X_i = e_i$, lo hace de forma distinta:
 - En vez de simular un valor para X_i , **fija** el valor de $X_i = e_i$
- Con esto intenta resolver el problema del alto rechazo del muestreo lógico.
- Los nodos también se ordenan según una **ordenación ancestral**

Método de ponderación por verosimilitud

- La **distribución de la población** es de nuevo:

$$p_e(x) \propto \prod_{i=1}^n p_e(x_i | \pi_i)$$

donde

$$p_e(x_i | \pi_i) = \begin{cases} p(x_i | \pi_i), & \text{si } x_i \text{ y } \pi_i \text{ son consistentes con } e \\ 0, & \text{en otro caso} \end{cases}$$

- La **distribución de simulación** es:

$$h(x_i) = \begin{cases} p(x_i | \pi_i), & \text{si } X_i \notin \mathbf{E} \\ 1, & \text{si } X_i \in \mathbf{E} \text{ y } x_i = e_i \\ 0, & \text{en otro caso} \end{cases}$$

- El **peso** asociado a una realización $x^j = (x_1, \dots, x_n)$ resulta:

$$s(x) = \frac{p_e(x)}{h(x)} = \prod_{X_i \notin \mathbf{E}} \frac{p_e(x_i | \pi_i)}{p_e(x_i | \pi_i)} \prod_{X_i \in \mathbf{E}} \frac{p_e(x_i | \pi_i)}{1} = \prod_{X_i \in \mathbf{E}} p_e(x_i | \pi_i) = \prod_{X_i \in \mathbf{E}} p(e_i | \pi_i)$$

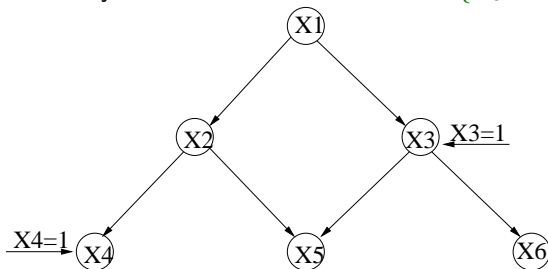
Método de ponderación por verosimilitud

Algorithm 5: Algoritmo de ponderación por verosimilitud

```
1 Ordenar los nodos ancestralmente;
2 foreach  $X_i \in \mathbf{E}$  do
3    $x_i = e_i$ 
4 for  $j = 1$  to  $N$  do
5   foreach  $X_i \notin \mathbf{E}$  do
6      $x_i = \text{valor generado a partir de } h(x_i);$ 
7    $s_j = \prod_{X_i \in \mathbf{E}} p(e_i | \pi_i);$ 
8 Normalizar los pesos;
```

Ejemplo de ponderación por verosimilitud

Sea de nuevo la red bayesiana con la evidencia $\mathbf{E} = \{X_3 = 1, X_4 = 1\}$:



Ejemplo de ponderación por verosimilitud

Distribuciones de probabilidad condicional

| x_1 | $p(x_1)$ | x_1 | x_2 | $p(x_2 x_1)$ | x_1 | x_3 | $p(x_3 x_1)$ | x_2 | x_4 | $p(x_4 x_2)$ |
|-------|----------|-------|-------|--------------|-------|-------|--------------|-------|-------|--------------|
| 0 | 0.3 | 0 | 0 | 0.4 | 0 | 0 | 0.2 | 0 | 0 | 0.3 |
| 1 | 0.7 | 0 | 1 | 0.6 | 0 | 1 | 0.8 | 0 | 1 | 0.7 |
| | | 1 | 0 | 0.1 | 1 | 0 | 0.5 | 1 | 0 | 0.2 |
| | | 1 | 1 | 0.9 | 1 | 1 | 0.5 | 1 | 1 | 0.8 |

| x_3 | x_6 | $p(x_6 x_3)$ | x_2 | x_3 | x_5 | $p(x_5 x_2, x_3)$ |
|-------|-------|--------------|-------|-------|-------|-------------------|
| 0 | 0 | 0.1 | 0 | 0 | 0 | 0.4 |
| 0 | 1 | 0.9 | 0 | 0 | 1 | 0.6 |
| 1 | 0 | 0.4 | 0 | 1 | 0 | 0.5 |
| 1 | 1 | 0.6 | 0 | 1 | 1 | 0.5 |
| | | | 1 | 0 | 0 | 0.7 |
| | | | 1 | 0 | 1 | 0.3 |
| | | | 1 | 1 | 0 | 0.2 |
| | | | 1 | 1 | 1 | 0.8 |

Ejemplo de ponderación por verosimilitud

- Elegimos un **orden ancestral** de los nodos no evidenciales:
 $\{X_1, X_2, X_5, X_6\}$

Ejemplo de ponderación por verosimilitud

- Elegimos un **orden ancestral** de los nodos no evidenciales:
 $\{X_1, X_2, X_5, X_6\}$
- Asignamos a los nodos evidenciales sus correspondientes valores observados $X_3 = 1$ y $X_4 = 1$.

Ejemplo de ponderación por verosimilitud

- Elegimos un **orden ancestral** de los nodos no evidenciales:
 $\{X_1, X_2, X_5, X_6\}$
- Asignamos a los nodos evidenciales sus correspondientes valores observados $X_3 = 1$ y $X_4 = 1$.
- Muestreamos X_1 con $p(X_1)$

| x_1 | $p(x_1)$ |
|-------|----------|
| 0 | 0.3 |
| 1 | 0.7 |

Suponer que obtenemos que $X_1 = 0$

Ejemplo de ponderación por verosimilitud

- Elegimos un **orden ancestral** de los nodos no evidenciales:
 $\{X_1, X_2, X_5, X_6\}$
- Asignamos a los nodos evidenciales sus correspondientes valores observados $X_3 = 1$ y $X_4 = 1$.
- Muestreamos X_1 con $p(X_1)$

| x_1 | $p(x_1)$ |
|-------|----------|
| 0 | 0.3 |
| 1 | 0.7 |

Suponer que obtenemos que $X_1 = 0$

- Muestreamos X_2 usando $p(X_2|X_1 = 0)$

| x_1 | x_2 | $p(x_2 x_1)$ |
|-------|-------|--------------|
| 0 | 0 | 0.4 |
| 0 | 1 | 0.6 |
| 1 | 0 | 0.1 |
| 1 | 1 | 0.9 |

Suponer que obtenemos $X_2 = 1$

Ejemplo de ponderación por verosimilitud

- Muestreamos X_5 con $p(X_5|X_2 = 1, X_3 = 1)$

| x_2 | x_3 | x_5 | $p(x_5 x_2, x_3)$ |
|-------|-------|-------|-------------------|
| 0 | 0 | 0 | 0.4 |
| 0 | 0 | 1 | 0.6 |
| 0 | 1 | 0 | 0.5 |
| 0 | 1 | 1 | 0.5 |
| 1 | 0 | 0 | 0.7 |
| 1 | 0 | 1 | 0.3 |
| 1 | 1 | 0 | 0.2 |
| 1 | 1 | 1 | 0.8 |

Suponer que obtenemos $X_5 = 0$

Ejemplo de ponderación por verosimilitud

- Muestreamos X_5 con $p(X_5|X_2 = 1, X_3 = 1)$

| x_2 | x_3 | x_5 | $p(x_5 x_2, x_3)$ |
|-------|-------|-------|-------------------|
| 0 | 0 | 0 | 0.4 |
| 0 | 0 | 1 | 0.6 |
| 0 | 1 | 0 | 0.5 |
| 0 | 1 | 1 | 0.5 |
| 1 | 0 | 0 | 0.7 |
| 1 | 0 | 1 | 0.3 |
| 1 | 1 | 0 | 0.2 |
| 1 | 1 | 1 | 0.8 |

Suponer que obtenemos $X_5 = 0$

- Muestreamos X_6 usando $p(x_6|X_3 = 1)$

| x_3 | x_6 | $p(x_6 x_3)$ |
|-------|-------|--------------|
| 0 | 0 | 0.1 |
| 0 | 1 | 0.9 |
| 1 | 0 | 0.4 |
| 1 | 1 | 0.6 |

Suponer que obtenemos $X_6 = 1$

Ejemplo de ponderación por verosimilitud

- Por tanto la primera realización es $x^1 = (0, 1, 1, 1, 0, 1)$ y su peso asociado es:

$$s(x^1) = p(X_3 = 1|X_1 = 0) \cdot p(X_4 = 1|X_2 = 1) = 0.8 \times 0.8 = 0.64$$

- El proceso se repite hasta obtener N realizaciones.
- La distribución de probabilidad de la variable de interés $p(x_i|e)$ se aproxima por el cociente entre la suma de pesos de las realizaciones en las que $X_i = x_i$ y la suma total de pesos.

Contenido del tema

1 Introducción

2 Algoritmos de Monte Carlo

- Simulación de variables aleatorias
- Base intuitiva de los métodos de simulación
- Metodología general de los métodos de simulación en RBs
- Muestreo lógico probabilístico
- Método de ponderación por verosimilitud
- Método de muestreo de Markov

Método de muestreo de Markov (Pearl, 1987)

- Comienza asignando a las variables evidencia su correspondiente valor
- Se simula luego la red de **forma estocástica**:
 - Inicialmente, se genera una realización aleatoriamente, eligiendo una al azar o bien aplicando uno de los métodos previos.
 - Simular las variables no evidenciales, una a una, siguiendo un orden arbitrario, mediante su función de probabilidad condicionada a todas las demás $p(x_i | \mathbf{x} \setminus x_i)$.
- El peso asociado a cada realización es siempre igual a 1.
- La función de probabilidad condicional $p(x_i | e)$ se estima por la proporción de realizaciones en las que ocurre x_i

Método de muestreo de Markov (Pearl, 1987)

Teorema: Función de probabilidad de una variable condicionada a todas las demás

La función de probabilidad de una variable X_i condicionada a todas las demás, se puede obtener con

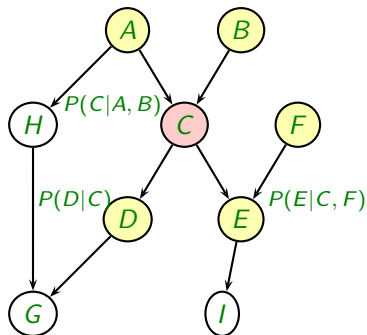
$$h(x_i) = p(x_i | \mathbf{x} \setminus x_i) \propto p(x_i | \pi_i) \prod_{X_j \in C_i} p(x_j | \pi_j)$$

donde C_i es el conjunto de hijos de X_i y $\mathbf{x} \setminus X_i$ denota todas las variables de \mathbf{x} que no están en X_i .

- Una vez muestreadas todas las variables no evidenciales, se obtiene una realización.
- Los valores de las variables obtenidos se utilizan para generar la siguiente realización.

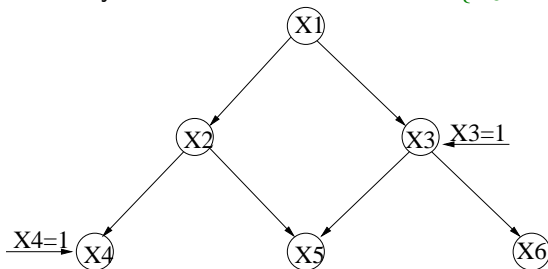
Método de muestreo de Markov (Pearl, 1987)

La frontera de Markov de C (nodos con fondo amarillo).



Ejemplo

Sea de nuevo la red bayesiana con la evidencia $\mathbf{E} = \{X_3 = 1, X_4 = 1\}$:



Ejemplo

Se procede de la siguiente forma:

- Inicialmente, asignamos un valor arbitrario inicial a cada variable no evidencial. Las variables observadas nunca cambian de valor:
 $\{X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 1\}$
- Supongamos que elegimos el siguiente orden arbitrario para simular las variables (no evidenciales): $\{X_1, X_2, X_5, X_6\}$
- Cada vez se cambia solo un valor condicionado a los anteriores y a **E**

| <i>N.iter.</i> | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 |
|----------------|-------|-------|-------|-------|-------|-------|
| 0 | 0 | 1 | 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 2 | 1 | 0 | 1 | 1 | 0 | 1 |
| 3 | 1 | 0 | 1 | 1 | 1 | 1 |
| 4 | 1 | 0 | 1 | 1 | 1 | 1 |
| 5 | 1 | 0 | 1 | 1 | 1 | 1 |
| 6 | 1 | 0 | 1 | 1 | 1 | 1 |
| 7 | 1 | 0 | 1 | 1 | 0 | 1 |
| 8 | 1 | 0 | 1 | 1 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... |

Método de muestreo de Markov (Pearl, 1987)

Algorithm 6: Algoritmo de muestreo de Markov

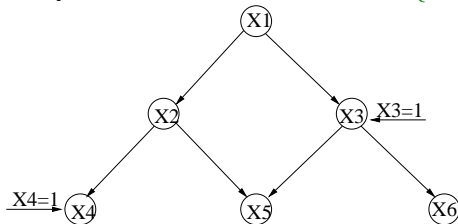
```

1  foreach  $X_i \in \mathbf{E}$  do
2     $x_i = e_i$ 
3  foreach  $X_i \notin \mathbf{E}$  do
4     $x_i = \text{valor generado con } U(0, 1);$ 
5  for  $j = 1$  to  $N$  do
6    foreach  $X_i \notin \mathbf{E}$  do
7      foreach  $x_i \in X_i$  do
8         $q(x_i) = p(x_i | \pi_i) \prod_{X_j \in C_i} p(x_j | \pi_j);$ 
9        Normalizar  $q(x_i);$ 
10        $x_i = \text{valor generado a partir de } q(x_i) \text{ normalizada;}$ 

```

Ejemplo método de muestreo de Markov

Sea de nuevo la red bayesiana con la evidencia $\mathbf{E} = \{X_3 = 1, X_4 = 1\}$:



Ejemplo

Distribuciones de probabilidad condicional

| x_1 | $p(x_1)$ | x_1 | x_2 | $p(x_2 x_1)$ | x_1 | x_3 | $p(x_3 x_1)$ | x_2 | x_4 | $p(x_4 x_2)$ |
|-------|----------|-------|-------|--------------|-------|-------|--------------|-------|-------|--------------|
| 0 | 0.3 | 0 | 0 | 0.4 | 0 | 0 | 0.2 | 0 | 0 | 0.3 |
| 1 | 0.7 | 0 | 1 | 0.6 | 0 | 1 | 0.8 | 0 | 1 | 0.7 |
| | | 1 | 0 | 0.1 | 1 | 0 | 0.5 | 1 | 0 | 0.2 |
| | | 1 | 1 | 0.9 | 1 | 1 | 0.5 | 1 | 1 | 0.8 |

| x_3 | x_6 | $p(x_6 x_3)$ | x_2 | x_3 | x_5 | $p(x_5 x_2, x_3)$ |
|-------|-------|--------------|-------|-------|-------|-------------------|
| 0 | 0 | 0.1 | 0 | 0 | 0 | 0.4 |
| 0 | 1 | 0.9 | 0 | 0 | 1 | 0.6 |
| 1 | 0 | 0.4 | 0 | 1 | 0 | 0.5 |
| 1 | 1 | 0.6 | 0 | 1 | 1 | 0.5 |
| | | | 1 | 0 | 0 | 0.7 |
| | | | 1 | 0 | 1 | 0.3 |
| | | | 1 | 1 | 0 | 0.2 |
| | | | 1 | 1 | 1 | 0.8 |

Ejemplo método de muestreo de Markov

- Asignamos valores a variables evidenciales: $X_3 = 1, X_4 = 1$

Ejemplo método de muestreo de Markov

- Asignamos valores a variables evidenciales: $X_3 = 1, X_4 = 1$
- Asignar un valor arbitrario inicial a cada variable no evidencial:
 $X_1 = 0, X_2 = 1, X_5 = 0, X_6 = 1$. Obtenemos $x^0 = (0, 1, 1, 1, 0, 1)$

Ejemplo método de muestreo de Markov

- Asignamos valores a variables evidenciales: $X_3 = 1, X_4 = 1$
- Asignar un valor arbitrario inicial a cada variable no evidencial:
 $X_1 = 0, X_2 = 1, X_5 = 0, X_6 = 1$. Obtenemos $x^0 = (0, 1, 1, 1, 0, 1)$
- Elegir ordenación arbitraria de nodos no evidenciales: $\{X_1, X_2, X_5, X_6\}$

Ejemplo método de muestreo de Markov

- Asignamos valores a variables evidenciales: $X_3 = 1, X_4 = 1$
- Asignar un valor arbitrario inicial a cada variable no evidencial: $X_1 = 0, X_2 = 1, X_5 = 0, X_6 = 1$. Obtenemos $x^0 = (0, 1, 1, 1, 0, 1)$
- Elegir ordenación arbitraria de nodos no evidenciales: $\{X_1, X_2, X_5, X_6\}$
- Para cada variable de $\{X_1, X_2, X_5, X_6\}$ generar un valor aleatorio mediante las distribuciones de simulación $h(x_i)$

Ejemplo método de muestreo de Markov

- Las distribuciones usadas para simulación de variables no evidenciales son:

$$\begin{aligned}
 h(x_1) &= p(x_1 | \mathbf{x} \setminus x_1) \propto p(x_1) p(x_2 | x_1) p(x_3 | x_1) \\
 h(x_2) &= p(x_2 | \mathbf{x} \setminus x_2) \propto p(x_2 | x_1) p(x_4 | x_2) p(x_5 | x_2, x_3) \\
 h(x_5) &= p(x_5 | \mathbf{x} \setminus x_5) \propto p(x_5 | x_2, x_3) \\
 h(x_6) &= p(x_6 | \mathbf{x} \setminus x_6) \propto p(x_6 | x_3)
 \end{aligned} \tag{2}$$

Ejemplo método de muestreo de Markov

- Variable X_1 :

$$\begin{aligned} p(X_1 = 0 | \mathbf{x} \setminus x_1) &\propto p(X_1 = 0)p(X_2 = 1 | X_1 = 0)p(X_3 = 1 | X_1 = 0) \\ &= 0.3 \times 0.6 \times 0.8 = 0.144 \end{aligned}$$

$$\begin{aligned} p(X_1 = 1 | \mathbf{x} \setminus x_1) &\propto p(X_1 = 1)p(X_2 = 1 | X_1 = 1)p(X_3 = 1 | X_1 = 1) \\ &= 0.7 \times 0.9 \times 0.5 = 0.315 \end{aligned}$$

- Normalizamos las anteriores probabilidades dividiendo por la suma 0.459. Obtenemos $p(X_1 = 0 | \mathbf{x} \setminus x_1) = 0.144/0.459 = 0.314$ y $p(X_1 = 1 | \mathbf{x} \setminus x_1) = 0.315/0.459 = 0.686$
- Generamos un valor aleatorio para X_1 con la anterior distribución:
 $X_1 = 0$

Ejemplo método de muestreo de Markov

- Variable X_2 :

- De la misma forma que con X_1 , usando el estado actual de las variables obtenemos:

$$p(X_2 = 0 | \mathbf{x} \setminus x_2) \propto 0.4 \times 0.3 \times 0.5 = 0.06$$

$$p(X_2 = 1 | \mathbf{x} \setminus x_2) \propto 0.6 \times 0.2 \times 0.2 = 0.024$$

(3)

- Normalizamos las probabilidades anteriores.
Obtenemos $p(X_2 = 0 | \mathbf{x} \setminus x_2) = 0.714$ y $p(X_2 = 1 | \mathbf{x} \setminus x_2) = 0.286$
- Generamos un valor aleatorio para X_2 con la anterior distribución:
 $X_2 = 1$

- Las variables X_5 y X_6 se simulan de forma similar: $X_5 = 0$ y $X_6 = 1$
- La primera realización obtenida es $\mathbf{x}^1 = (0, 1, 1, 1, 0, 1)$
- Repetimos hasta obtener N extracciones

Ejemplo método de muestreo de Markov

- La **distribución de probabilidad condicional** $p(x_i|e)$ (por ejemplo $p(X_2 = 1|e)$) puede obtenerse con:
 - Porcentaje de realizaciones en las que $X_2 = 1$
 - Calcular la media de las $p(X_2 = 1|\mathbf{x} \setminus x_2)$ de todas las realizaciones.