Tema 3

Métodos específicos de generación de diversas distribuciones discretas

3.1. Distribución de Bernoulli

Sea $X \sim \mathcal{B}(p)$. La función de probabilidad puntual de X es:

$$P(X = 1) = p \qquad P(X = 0) = 1 - p$$

Utilizando el método de la transformada inversa, se pueden generar valores de X mediante el siguiente esquema:

1. Generar un número aleatorio u
2. Si $u \leq p$, hacer $X = 1$. En caso contrario, tomar $X = 0$.

3.2. Distribución binomial

Sea $X \sim \mathcal{B}i(n,p)$. La función de probabilidad puntual de X es:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Puesto que una variable $\mathcal{B}i(n,p)$ es la repetición de n experimentos de Bernoulli de probabilidad de éxito p, podemos generar valores de esta distribución del siguiente modo:

1. Hacer $x = 0$, $i = 1$.
2. Si $i \leq n$ ir al paso 3. En caso contrario ir al paso 5
3. Generar un número aleatorio u_i. Si $u_i \leq p$, hacer $x = x + 1$.
4. Hacer $i = i + 1$. Ir al paso 2.
5. x es el valor generado de la variable X

Igualmente, puesto que una variable $\mathcal{B}i(n,p)$ es la suma de n variables aleatorias independientes $\mathcal{B}(p)$, se podrían generar n valores independientes de una distribución $\mathcal{B}(p)$, y posteriormente sumar tales valores.

En cualquier caso, el método anterior requiere la generación de n números aleatorios y n comparaciones. Por ello, en general, es más eficiente la aplicación del método de inversión de la función de distribución teniendo en cuenta la siguiente relación recursiva que verifica la función puntual de probabilidad de un variable $\mathcal{B}i(n,p)$:

$$P(X = i+1) = \frac{(n-i)p}{(i+1)(1-p)} P(X = i).$$

El esquema de generación sería el siguiente:

1. Generar un número aleatorio u.
2. Hacer $x = 0$, $P = F = (1-p)^n$.
3. Si $u < F$, ir al paso 5.
4. Hacer $P = \frac{(n-i)p}{(x+1)(1-p)}P$, $F = F + p$, $x = x + 1$. Ir al paso 3.
5. x es el valor generado de X.

Con el esquema anterior es necesario generar un único número aleatorio y el número de comparaciones es uno más que el valor generado de X. Así pues, en promedio se realizan $1 + np$ comparaciones.

3.3. Distribución de Poisson

Sea $X \sim \mathcal{P}o(\lambda)$. La función de probabilidad puntual de X es:

$$P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}, \quad \forall x \geq 0$$

La distribución de Poisson está relacionada con la distribución exponencial, de modo que el número de veces que ocurre un determinado suceso en un intervalo de tiempo de longitud unidad sigue una distribución de Poisson de parámetro λ si y sólo si los tiempos entre sucesos son independientes y se distribuyen según una distribución $Exp(\lambda)$. Haciendo uso de esta propiedad, podemos generar valores de una distribución $Po(\lambda)$ mediante el siguiente esquema:

1. Hacer $x = 0$, $h = 0$.
2. Generar un valor y de una distribución exponencial de parámetro λ. Hacer $h = h + y$. Si $h > 1$ ir al paso 4.

3. Hacer $x = x + 1$ e ir al paso 2.
4. x es el valor generado de la distribución $Po(\lambda)$.

Como aplicación del método de inversión de la función de distribución, se vio que se pueden generar valores de una distribución $Y \sim Exp(\lambda)$ mediante la expresión $Y = -\frac{1}{\lambda} \ln U$.

En el esquema anterior se cuentan los sucesos que ocurren hasta que $\sum_i Y_i > 1$. Se puede simplificar este esquema teniendo en cuenta que:

$$\sum_i Y_i > 1 \Leftrightarrow -\frac{1}{\lambda} \sum_i \ln U_i > 1 \Leftrightarrow \ln \prod_i U_i < -\lambda \Leftrightarrow \prod_i U_i < e^{-\lambda} \Leftrightarrow e^{\lambda} \prod_i U_i < 1$$

De este modo, el algoritmo quedaría

1. Hacer $x = 0$, $h = e^{\lambda}$.
2. Generar un número aleatorio $u \in \mathcal{U}(0, 1)$. Hacer $h = h \cdot u$. Si $h < 1$ ir al paso 4.
3. Hacer $x = x + 1$ e ir al paso 2.
4. x es el valor generado de la distribución $Po(\lambda)$.

Al igual que con la distribución exponencial, podemos aplicar el método de inversión de la función de distribución teniendo en cuenta la siguiente relación recursiva que verifica la función puntual de probabilidad de un variable $\mathcal{P}o(\lambda)$:

$$P(X = i + 1) = \frac{\lambda}{i + 1} P(X = i).$$

El esquema de generación sería el siguiente:

1. Generar un número aleatorio u.

2. Hacer $x = 0$, $P = F = e^{-\lambda}$.
3. Si $u < F$, ir al paso 5.
4. Hacer $P = \frac{\lambda}{x+1}P$, $F = F + p$, $x = x + 1$. Ir al paso 3.
5. x es el valor generado de X.

En el esquema anterior es necesario la generación de un único número aleatorio y, en promedio, serán necesarias $1 + \lambda$ comparaciones.

Ejemplo 3.1. Generación de valores de $Po(\lambda)$ por medio del método de composición.

3.4. Distribución geométrica

Sea $X \sim \mathcal{G}(p)$. X representa el número de intentos necesarios para alcanzar el primer éxito en las realizaciones independientes sucesivas de un experimento de Bernoulli de probabilidad de éxito p. Su función puntual de probabilidad viene dada por:

$$P(X = x) = (1-p)^{x-1}p, \quad \forall x \geq 1$$

La función de distribución de X es:

$$F(x) = P(X \leq x) = \sum_{j=1}^{x} P(X = j) = \sum_{j=1}^{x}(1-p)^{j-1}p = \frac{(1-p)^x - 1}{(1-p) - 1}p = 1-(1-p)^x, \quad \forall x \geq 1$$

Podemos utilizar el método de la transformada inversa para generar valores de la distribución geométrica del siguiente modo. Dado un número aleatorio u, asignamos a X el valor x si y sólo si $F(x-1) \leq x < F(x)$.

$$\begin{aligned}
F(x-1) \leq u < F(x) &\Leftrightarrow 1-(1-p)^{x-1} \leq x < 1-(1-p)^x \Leftrightarrow (1-p)^{x-1} \geq 1-u > (1-p)^x \\
&\Leftrightarrow (x-1)\ln(1-p) \geq \ln(1-u) > x\ln(1-p) \Leftrightarrow x-1 \leq \frac{\ln(1-u)}{\ln(1-p)} < x \\
&\Leftrightarrow x = 1 + \left[\frac{\ln(1-u)}{ln(1-p)}\right]
\end{aligned}$$

Obsérvese que si u es una realización de una variable aleatoria $\mathcal{U}(0,1)$, entonces $1-u$ también lo es. Por lo tanto, el algoritmo queda:

1. Generar un número aleatorio u
2. El valor generado de X es $1+\left[\frac{\ln u}{\ln(1-p)}\right]$.

Otra posibilidad para generar valores de una distribución geométrica es utilizar la propia definición de la misma.

1. Hacer $x=1$.
2. Generar un número aleatorio u. Si $u \leq p$ ir al paso 4. En caso contrario ir al paso 3.
3. Hacer $x=x+1$. Ir al paso 2.
4. x es el valor generado de la variable aleatoria X.

3.5. Distribución binomial negativa

Sea $X \sim \mathcal{BN}(n,p)$. X representa el número de intentos necesarios para alcanzar el n-ésimo éxito en las realizaciones independientes sucesivas de un experimento de Bernoulli de probabilidad de éxito p. Su función puntual de probabilidad viene dada por:

$$P(X=x)=\binom{x-1}{n-1}p^n(1-p)^{x-n}, \quad \forall x=n,n+1,\ldots$$

Es fácil demostrar que una variable $\mathcal{BN}(n,p)$ se puede expresar como la suma de n variables independientes $\mathcal{G}(p)$. Por lo tanto, el algoritmo se podría esquematizar del siguiente modo:

1. Generar $x_1, x_2, \ldots, x_n$ valores independientes de una distribución $\mathcal{G}(p)$

2. Tomar $x = \sum_{i=1}^{n} x_i$ como valor generado de X

3.6. Distribución hipergeométrica

En una urna hay N unidades de las cuales D son defectuosas y $C = N - D$ son correctas. Se extraen al azar n unidades sin reemplazamiento. Sea X: número de unidades defectuosas extraídas. Entonces, $X \sim \mathcal{HG}(N, D, n)$. Su función puntual de probabilidad es:

$$P(X = x) = \frac{\binom{D}{x}\binom{C}{n-x}}{\binom{N}{n}}, \quad \forall\, \text{máx}\{0, n - C\} \leq x\, \text{mín}\{D, n\}.$$

Mediante el siguiente esquema se puede generar valores de la distribución hipergeométrica:

1. Hacer $x = 0$, $D = d$,$c = C = N - D$, $i = 1$

2. Si $i \geq n$, ir al paso 3. En caso contrario, ir al paso 6.

3. Generar un número aleatorio u_i. Si $u_i \leq \frac{d}{N}$, ir al paso 4. En caso contrario, ir al paso 5.

4. Hacer $x = x + 1$, $d = d - 1$, $N = N - 1$, $i = i + 1$. Ir al paso 2.

5. Hacer $c = c - 1$, $N = N - 1$, $i = i + 1$. Ir la paso 2.

6. x es el valor generado de X.