Andrés Gutiérrez Rojas
Centro de Investigaciones y Estudios Estadísticos
Facultad de Estadística
Universidad Santo Tomás

Email: hugogutierrez@usantotomas.edu.co
Email: contacto@gutierrezandres.com

Considere el algoritmo de selección de Fan-Muller-Rezucha. Para demostrar que este método efectivamente induce muestras con probabilidad de selección igual a $Pr(S=s)=\binom{N}{n}^{-1}$.

1. Sustente por qué $Pr(k\in S|n_k)=\frac{n-n_k}{N-k+1}=c_k$. Donde n_k es el número de individuos seleccionados antes del k-ésimo elemento...

   Empezamos por notar que $Pr(k\in S|n_k)$ no es igual a π_k. Ahora, la anterior probabilidad está supeditada a la inclusión del elemento k dado que n_k ya han sido seleccionados. Es algo así como la probabilidad de selección en el algoritmo. Y ese algoritmo está dado por la generación de un $\xi_k\sim U(0,1)$. El algoritmo afirma que si $\xi_k<c_k$, entonces el elemento queda seleccionado. Por lo tanto, dada la distribución de ξ_k, se tiene que $Pr(k\in S|n_k)=Pr(\xi_k<c_k)=\frac{n-n_k}{N-k+1}$.

2. Ahora, explique por qué $Pr(S=s)=\prod_{k\in s}Pr(k\in s|n_k)\prod_{k\notin s}Pr(k\notin s|n_k)$...

   En primer lugar, es claro que la ocurrencia del evento $S=s$, está dado por la ocurrencia de N eventos dados por $I_1=I_1(s),I_2=I_2(s),\ldots,I_N=I_N(s)$. Es decir, que una muestra sea realizada implica que los I_k sean unos o ceros. Ahora, utilizando el condicionamiento sucesivo[1] y notando que el evento n_k está directamente relacionado con la realización de los I_k anteriores, se tiene que

$$
\begin{aligned}
Pr(S=s)&=Pr(I_1=I_1(s),I_2=I_2(s),\ldots,I_N=I_N(s))\\
&=Pr(I_N=I_N(s)\mid I_{N-1}(s),\ldots,I_2(s),I_1(s))Pr(I_{N-1}=I_{N-1}(s)\mid I_{N-2},\ldots,I_2(s),I_1(s))\\
&\times\cdots\times Pr(I_2=I_2(s)\mid I_1(s))Pr(I_1=I_1(s))\\
&=\prod_{k\in U}Pr(I_k=I_k(s)\mid n_k)\\
&=\prod_{k\in s}Pr(I_k=I_k(s)\mid n_k)\prod_{k\notin s}Pr(I_k=I_k(s)\mid n_k)=\prod_{k\in s}Pr(k\in s|n_k)\prod_{k\notin s}Pr(k\notin s|n_k)
\end{aligned}
$$

3. Con base en los anteriores puntos, demuestre que

$$
Pr(S=s)=\left(\prod_{k\in U}(N-k+1)\right)^{-1}\left(\prod_{k\in s}(n-n_k)\right)\left(\prod_{k\in U-s}(N-k+1-n+n_k)\right)
$$

[1] Esta propiedad afirma que $Pr(A,B,C)=Pr(C\mid B,A)Pr(B\mid A)Pr(A)$.

Utilizando el desarrollo anterior, se tiene que

$$\begin{aligned}
Pr(S=s) &= \prod_{k\in s} Pr(k \in s|n_k) \prod_{k\notin s} Pr(k \notin s|n_k) \\
&= \prod_{k\in s}\left(\frac{n-n_k}{N-k+1}\right) \prod_{k\notin s}\left(1-\frac{n-n_k}{N-k+1}\right) \\
&= \frac{\prod_{k\in s}(n-n_k)}{\prod_{k\in s}(N-k+1)} \frac{\prod_{k\notin s}(N-k+1-n+n_k)}{\prod_{k\notin s}(N-k+1)} \\
&= \frac{\prod_{k\in s}(n-n_k)\prod_{k\notin s}(N-k+1-n+n_k)}{\prod_{k\in U}(N-k+1)} \\
&= Pr(S=s) = \left(\prod_{k\in U}(N-k+1)\right)^{-1}\left(\prod_{k\in s}(n-n_k)\right)\left(\prod_{k\in U-s}(N-k+1-n+n_k)\right)
\end{aligned}$$

4. Explique por qué se tienen las siguientes igualdades

   - $\prod_{k\in U}(N-k+1) = N!$

     Se tiene que

     $$\prod_{k\in U}(N-k+1) = \prod_{k=1}^{N}(N-k+1) = N(N-1)(N-2)\cdots(2)(1) = N!$$

   - $\prod_{k\in s}(n-n_k) = n!$

     Nótese que para todos los individuos en la muestra realizada s, se tiene que $n_k \neq n_l$ $(\forall k \neq l)$ y además $n_k \in \{0, 1, \ldots, n-1\}$. Luego,

     $$\prod_{k\in s}(n-n_k) = (n-0)(n-1)\cdots(n-n+2)(n-n+1) = n!$$

   - $\prod_{k\notin s}(N-k+1-n+n_k) = (N-n)!$

     Para desarrollar esta productoria, nótese que $N-k+1-n+n_k = (N-k+1)-(n-n_k)$. Es claro que $N-k+1$ describe el número de elementos en la población a los cuales no se les ha aplicado el algoritmo, incluyendo el k-ésimo. Ahora $n-n_k$ denota el número de elementos que faltan por incluir en la muestra. De tal forma que la resta de estos dos términos resulta ser el número de elementos que no serán incluidos en la muestra, desde k en adelante. Por lo tanto, para el primer elemento en el complemento de la muestra, se tiene que la resta es $N-n$; para el segundo elemento en el complemento de la muestra, se tiene que la resta es $N-n-1$ y así sucesivamente hasta el último elemento, en donde la resta será 1.

5. Concluya que $Pr(S = s) = \binom{N}{n}^{-1}$.

   Del punto anterior, se tiene que

$$\begin{aligned}
Pr(S = s) &= \left(\prod_{k\in U}(N - k + 1)\right)^{-1}\left(\prod_{k\in s}(n - n_k)\right)\left(\prod_{k\in U-s}(N - k + 1 - n + n_k)\right) \\
&= \frac{n!(N - n)!}{N!} = \frac{1}{\binom{N}{n}}
\end{aligned}$$