Bajo la situación actual en la que se pide a los habitantes de la ciudad que no salgan de sus casas a menos que sea necesario, y que no se realicen reuniones sociales, se desea estimar la proporción de personas que no cumplen con dicha directiva y realizan reuniones sociales. Sea θ la verdadera proporción de personas que participó en reuniones sociales en el último mes. Se desea por lo tanto, estimar θ.

El experimento consiste en seleccionar n personas al azar y preguntarles si realizaron o no reuniones sociales en el último mes. A partir de la respuesta del encuestado a la pregunta "¿Participó de reuniones sociales en el último mes?", se definen las variables

$$X_i = \begin{cases} 1 & \text{si} \quad \text{la i-ésima persona responde que sí} \\ 0 & \text{si} \quad \text{la i-ésima persona responde que } \textbf{NO} \end{cases}$$

Si todos los encuestados dicen la verdad, resulta

$$X_i \sim Bi(1, \theta)$$

Sin embargo, se sospecha que no todas las personas que asistieron a reuniones sociales dicen la verdad por vergüenza o miedo de responder. Llamemos π_{M} a la probabilidad de que una persona que participó de una reunión social haya respondido que no asistió al ser encuestada. Asumimos que si una persona no estuvo en una reunión social no miente, es decir, que la probabilidad de decir que sí asistió es 0.

Por lo tanto, definimos las siguientes variables:

$$V_i = \begin{cases} 1 & \text{si} \quad \text{la i-ésima persona estuvo en una reunión social,} \\ 0 & \text{si} \quad \text{el i-ésimo encuestado persona } \textbf{NO} \text{ estuvo en una reunión social} \end{cases}$$

$$L_i = \begin{cases} 1 & \text{si} \quad \text{la i-ésima persona miente} \\ 0 & \text{si} \quad \text{la i-ésima persona } \textbf{NO} \text{ miente} \end{cases}$$

a) Expresar $\mathbb{P}(X_i = 1)$ en función de θ y π_{M}.

b) Sea $\widehat{\theta} = \overline{X}$, decidir si $\widehat{\theta}$ es un estimador insesgado y consistente para θ.

<u>Método de respuesta aleatorizada</u>

Teniendo en cuenta que las personas entrevistadas que participaron en reuniones sociales pueden mentir, se propone el siguiente método para estimar θ.

A cada entrevistado se le entrega un dado, que tiene la palabra "sí" en dos de sus caras, y "no" en las 4 caras restantes. Se le pide que tire el dado y diga si la respuesta que observa en el dado coincide con su respuesta a la pregunta "¿Participó de reuniones sociales el último mes?", es decir,

- Si asistió a reuniones sociales y al tirar el dado sale una cara con "sí", va a responder "sí", mientras que si sale una cara con "no" va a responder "no"
- Por otra parte, si no asistió a reuniones sociales y al tirar el dado sale una cara con "sí", va a responder "no", mientras que si sale una cara con "no" va a responder "sí"

Un punto importante en este procedimiento es que el entrevistador no ve el resultado del dado, de esa manera la persona entrevistada no tiene ninguna razón para mentir.

Esté método ofrece una alternativa para estimar a θ. Para ver esto, defina las variables aleatorias

$$Y_i = \begin{cases} 1 & \text{si} \quad \text{la i-ésima persona responde que "sí" al ver el resultado del dado,} \\ 0 & \text{si} \quad \text{si responde que la respuesta que observa en el dado } \textbf{NO} \text{ coincide} \\ & \qquad \text{con su respuesta a la pregunta } \textit{¿Participó de reuniones sociales el último mes?} \end{cases}$$

$$E_i = \begin{cases} 1 & \text{si} \quad \text{la i-ésima persona obtuvo "sí" en el dado} \\ 0 & \text{si} \quad \text{al i-ésimo encuestado le salió "no" en el dado} \end{cases}$$

La muestra $Y_1, \dots, Y_n$ son los únicos datos que tiene el encuestador para estimar θ. En base a esta nueva información

c) Encontrar la expresión que relaciona $p = \mathbb{P}(Y = 1)$ con θ.

d) Encontrar el estimador de momentos, $\widehat{\theta}_{\mathrm{MO}}$, para θ

   (I) Es un estimador razonable si $\overline{Y} > 2/3$ o $\overline{Y} < 1/3$?

   (II) Decidir si es insesgado y consistente.

e) Encontrar el estimador de máxima verosimilitud, $\widehat{\theta}_{\mathrm{MV}}$, para θ.

   Hint: Trate con cuidado los casos $1/3 < \overline{Y} < 2/3$, $\overline{Y} > 2/3$ *y el caso* $\overline{Y} < 1/3$.

   (I) Coincide el estimador de momentos con el estimador de máxima verosimilitud?

   (II) Muestre que $|\widehat{\theta}_{\mathrm{MV}} - \theta| \leq |\widehat{\theta}_{\mathrm{MO}} - \theta|$. Decidir si el estimador de máxima verosimilitud es consistente.

Para resolver en R

Vamos a estudiar ahora como se comporta el error cuadrático medio, a medida que aumentamos el tamaño de la muestra, para los tres estimadores definidos a lo largo del trabajo.

Supongamos que un 75 % de los encuestados que asistieron a reuniones sociales dicen la verdad, o sea, que $\pi_{\mathrm{M}} = 0.25$

f) Implementar en R funciones que calculen el error cuadrático medio de $\widehat{\theta}$, $\widehat{\theta}_{\mathrm{MO}}$ y de $\widehat{\theta}_{\mathrm{MV}}$ en función del número de encuestados n y de la verdadera proporción θ de personas que participó de reuniones sociales en el último mes.

g) Graficar el ECM de $\widehat{\theta}$, $\widehat{\theta}_{\mathrm{MO}}$ y de $\widehat{\theta}_{\mathrm{MV}}$ como función de n, suponiendo que $\theta = 0.2$

h) Graficar el ECM de $\widehat{\theta}$, $\widehat{\theta}_{\mathrm{MO}}$ y de $\widehat{\theta}_{\mathrm{MV}}$ como función de θ, suponiendo que $n = 100$