Solutions Problem Set 2

Tatiana Rosá Alejo Eyzaguirre

22nd September 2021

Outline

Pregunta 1: Encuestas

Hay 2 candidatos, A y B, presentándose para ser elegidos presidente en las próximas elecciones en Chile. Una encuestadora quiere tomar una muestra de los votantes lo suficientemente grande para que, si 52% de la verdadera población vota a A, la probabilidad que en la muestra se encuentre un porcentaje devotos en favor de A de menos de 50% sea menos de 1%.

1. Asumimos que en la población la probabilidad de que un individuo vote A es p = 0.52, y que lo que vote un individuo es independiente de lo que vote otro. Pruebe que la probabilidad de que la encuesta de un resultado menor a $\alpha\%$ es:

$$\sum_{k=0}^{\alpha N} \frac{N!}{k!(N-k)!} 0.52^k 0.48^{N-k} \tag{1}$$

Donde N es el tamaño de la muestra.

Pregunta 1.1

En una muestra aleatoria de N votantes, el número de los votos a favor de A sigue una binomial(N,p). Por lo tanto, la PDF corresponde a:

$$f(k) = \binom{N}{k} p^k (1-p)^{n-k}$$

Es decir la probabilidad de que k personas voten por el candidato A, corresponde a f(k). Dicho en otras palabras, la probabilidad de que A tenga éxito k veces de un total de N, equivale a f(k). Por lo tanto, lo que se nos pide en

el enunciado equivale a agregar todos los casos en que αN o menos personas, voten por el candidato A. Esto equivale a hacer una sumatoria de la probabilidad de 0 hasta αN de la PDF, con lo cual obtendremos la CDF de la binomial.

$$F(\alpha N) = \sum_{k=0}^{\alpha N} \binom{N}{k} 0.52^k 0.48^{N-k}$$

Q.E.D.

Pregunta 1.2

Asumiendo que aumentar el tamaño de la muestra tiene un costo, encuentra la ecuación que define el N óptimo, N^* que satisface la condición de la encuestadora (que en la muestra se encuentre un porcentaje de votos en favor de A de menos de 50% sea menos de 1%.).

Dado que aumentar el N de la encuesta es costoso, hay que plantear un problema de optimización que nos permita encontrar el mínimo tamaño de la muestra, que cumpla con que la probabilidad de que A NO obtenga mayoría simple sea de a lo menos 1%. Usando la CDF, planteamos:

$$N^* = \underset{F(0.5N) \leq 0.1}{\arg\min} \; N$$

Pregunta 1.3

Podría encontrar el N^*. Piense en su reloción de forma analítica y numérica. No hace falta resolverlo.

Forma Analítica: La solución al problema de optimización se puede obtener usando un software matemático. La lógica para resolverlo es hacer un *loop* e ir iterando con distintos valores de N hasta que (*while not* sigo iterando) se cumpla la condición.

N	2	10	50	100	200
F(αN)	0.73	0.57	0.44	0.38	-

Forma Numérica: Sería vía la utilización de ciertos algoritmos de optimización numéricos, como por ejemplo Gradient Descent, BHHH o Newton Raphson (los últimos dos no se podrían usar porque requieren las segundas derivadas).

Pregunta 1.4

Usando un argumento asintótico, encuentre una aproximación para el tamaño óptimo de la muestra.

Como bien sabemos, por Ley de los Grandes Números podríamos aplicar el TLC, para la proporción de votos que obtiene A, en la votación. Como esta proporción es una muestra extraída de una población que distribuye Bernoulli(p), entonces tiene la siguiente distribución asintótica.

$$\mathcal{N}\left(p, \frac{p(1-p)}{N}\right)$$

Por lo tanto, tomando la CDF (estandarizada) de la distribución asintótica, tenemos:

$$F(\alpha N) \approx \Phi\left(\frac{0.5 - 0.52}{\sqrt{\frac{0.52 \times 0.48}{N}}}\right) = \tilde{F}(\alpha N) \tag{2}$$

Pregunta 1.4

Siguiendo la misma línea de la pregunta anterior podemos obtener el N de manera analítica y de forma "numérica". Para la forma analítica, seguimos el mismo *approach* que en el acápite anterior:

N	2	10	50	100	200
$\tilde{F}(\alpha N)$	0.48	0.45	0.39	0.34	0.29

Note que la aproximación no es muy buena para valores pequeños de N, y esta mejora cuando N aumenta. La forma numérica de resolver este problema sería despejando N de la ecuación 2, restringiendo que $\tilde{F}(\alpha N)$ sea igual a 0,01.

$$N^* = \left(\frac{\sqrt{0.52 \times 0.48}\Phi^{-1}(0.01)}{0.5 - 0.52} \right)^2 \approx 3.377$$

Pregunta 2: Distribuciones de la Media Muestral

Consideremos un muestra aleatoria de tamaño N de la distribución de $X \sim X_1^2$; $(X_1, X_2, ..., X_N)$. En este ejercicio el uso de las tablas de las distribuciones estándar pueden ser útiles.

1. Encuentre la distribución exacta de $\bar{X}$.

   Note que $\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$, donde cada X_i fue extraído de una población que distribuye $\mathcal{X}_1^2$. Por lo tanto, tenemos que $N \cdot \bar{X} \sim \mathcal{X}_N^2$, ya que eso distribuye la $\sum_{i=1}^{N} X_i$ (suma de variables aleatorias que distribuyen $\mathcal{X}_1^2$).

2. Calcule $P(\bar{X} > 1)$ para N = 1, 5, 10, 100.

   Note que: $P(\bar{X} > 1) = P(N \cdot X > N) = 1 - G_N(N)$. Donde G_N es la CDF de una distribución $\mathcal{X}_N^2$. Luego usando las tablas de la distribución $\mathcal{X}^2$ vemos que:

N	1	5	10	100
$1 - G_N(N)$	0.317	0.416	0.440	0.481

Pregunta 2: Distribuciones de la Media Muestral

3. Encuentre la distribución asintótica de $\bar{X}$.

   Tomando en cuenta que $E(X) = 1$ y que $Var(X) = 2$ y aplicando el TLC, tenemos que:

$$\bar{X} \overset{a}{\sim} \mathcal{N}\left(1, \frac{2}{N}\right)$$

4. Calcule $P(\bar{X} > 1)$ usando la distribución asintótica para N = 1, 5, 10, 100.

   Note que: $P(\bar{X} > 1) = 1 - \Phi(0)$, ya que la estandarización deja en el numerador un 0. Por lo tanto usando la distribución asintótica, independiente del valor de N que se utilice: $P(\bar{X} > 1) = 0.5$

5. Compare las probabilidades obtenidas con la aproximación asintótica y con la distribución exacta. ¿Apartir de que tamaño muestral considera usted que la aproximación es buena?

   Note que a partir de N=10, la aproximación es buena (0.44 vs 0.5). Por lo tanto, en este ejemplo, pese a que la distribución es solo asintótica (válida cuando $N \rightarrow \infty$), es una buena guía para la distribución exacta de la media muestral en este ejemplo.

Gracias!

jeeyzaguirre@uc.cl