Estadística Inferencial

Jackson M'coy Romero Plasencia

Universidad Nacional de San Cristóbal de Huamanga

Departamento Académico de Matemática y Física

Ayacucho 2020

Retroalimentación

Distribución de la Media Muestral

Teorema Central del Límite

Retroalimentación

Definición: Muestra Aleatoria

Una muestra aleatoria de tamaño n de una variable aleatorio X es un conjunto de n - variables aleatorias $X_1, X_2, ..., X_n$, tal que:

- Todas $X_1, X_2, ..., X_n \quad i = \overline{1, n}$ son independientes
- $X_1, X_2, ..., X_n$ tienen la misma distribución, es decir:

$$F_{X_i}(t) = F_X(t) \quad \forall\ i = \overline{1, n}$$

$$\longrightarrow f_{X_1, X_2, ..., X_n}(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

Ejercicio propuesto:

Si X es una variable aleatoria, con:

$$f_X(x,p) = f_X(x) = p^x(1-p)^{1-x}, \quad x = 0,1 \quad 0 < p < 1$$

y sea una muestra aleatoria de tamaño n . Hállese la distribución.

$$Y = \sum_{i=1}^{n} X_i$$

Función Generatriz de Momentos: $M_X(t)$

$$fgm(x) = M_X(t) = \mathbb{E}[e^{tx}]$$

Primero se determinará la $M_X(t)$ en la población

$$f_X(x,p) = f_X(x) = p^x(1-p)^{1-x}, \quad x = 0,1 \quad 0 < p < 1$$

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tx}] = \sum_{x=0}^{1} e^{tx}\, p^x\, (1-p)^{1-x} \\ &= \sum_{x=0}^{1} (pe^t)^x\, (1-p)^{1-x} \\ &= ((1-p) + pe^t) \end{aligned}$$

Primera forma: Usando la Función de Generatriz de Momentos: $Y = \sum_{i=1}^{n} X_i$

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{ty}] = \mathbb{E}[e^{t(X_1+X_2+\ldots+X_n)}] \\ &= \mathbb{E}[e^{t(X_1)}\, e^{t(X_2)} \ldots e^{t(X_n)}] \\ &= \mathbb{E}[e^{t(X_1)}]\, \mathbb{E}[e^{t(X_2)}] \ldots \mathbb{E}[e^{t(X_n)}] \\ &= \left(\mathbb{E}[e^{t(X_i)}]\right)^n \\ &= \left((1-p) + pe^t\right)^n \longrightarrow Bin(n,p) \end{aligned}$$

Segunda forma:

$$f_X(x, n, p) = f_X(x) = \binom{n}{p} p^x (1-p)^{1-x}, \quad x = \overline{0, n} \quad 0 < p < 1$$

Es la función de probabilidad de la Distribución Binomial.

$$\mathbb{E}(X) = np \quad \text{y} \quad \mathbb{V}(X) = npq$$

$$\begin{aligned}
\mathbb{E}(Y) &= \mathbb{E}(X_1 + X_2 + ... + X_n) &= \mathbb{E}(X_1) + ... + \mathbb{E}(X_n) \\
& &= \underbrace{p + p + ... + p}_{n\ veces} = np \\
\mathbb{V}(Y) &= \mathbb{V}(X_1 + X_2 + ... + X_n) &= \mathbb{V}(X_1) + ... + \mathbb{V}(X_n) \\
& &= \underbrace{pq + pq + ... + pq}_{n\ veces} = npq
\end{aligned}$$

Distribución de la Media Muestral

Distribución de la Media Muestral

Si $X_1, X_2, ..., X_n$ es unas muestra aleatoria de una población infinita, con media μ y varianza $\sigma^2 < \infty$, entonces:

$$\mathbb{E}(\overline{X}) = \mu \qquad \mathbb{V}(\overline{X}) = \frac{\sigma^2}{n}$$

Demostración:

Como $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$, entonces:

$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \sum_{i=1}^{n} \frac{\mathbb{E}(X_i)}{n} = \sum_{i=1}^{n} \frac{\mu}{n} = \mu$$

$$\mathbb{V}(\overline{X}) = \mathbb{V}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \sum_{i=1}^{n} \frac{\mathbb{V}(X_i)}{n^2} = \sum_{i=1}^{n} \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Por consiguiente: $\overline{X} \longrightarrow N(\mu, \frac{\sigma^2}{n})$

Ejemplo: Una fábrica textil tiene 5 operarios. Los años de servicio en la fábrica de estos operarios son:

$$3, 4, 7, 9, 12$$

- **a.** Calcule la media y varianza de la población
- **b.** Determine la distribución de la media de las muestras de tamaño 2 de la población(reposición)
- **c.** Determine la distribución de la media de las muestras de tamaño 2 de la población(sin reposición)

Solución:

- **a.** Media y varianza de la población

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i f(x_i) \quad \text{y} \quad \mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i f(x_i) = 3(1/5) + 4(1/5) + ... + 12(1/5) = 7$$

$$\mathbb{E}(X^2) = \sum_{i=1}^{n} x_i^2 f(x_i) = 3^2(1/5) + 4^2(1/5) + ... + 12^2(1/5) = 59.8$$

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 59.8 - 7^2 = 10.8$$

luego:

$$\mathbb{E}(X) = \mu = 7 \qquad \text{y} \qquad \mathbb{V}(X) = \sigma^2 = 10.8$$

En R:

```
pob<-c(3,4,7,9,12)
mean(pob)
```

```
## [1] 7
```

```
var(pob) #varianza con denominador n-1
```

```
## [1] 13.5
```

```
var(pob)*4/5 #barianza con denomiador n
```

b. Muestras de tamaño 2, con reposición

3,3	3,4	3,7	3,9	3,12	3.0	3.5	5.0	6.0	7.5
4,3	4,4	4,7	4,9	4,12	3.5	4.0	5.5	6.5	8.0
7,3	7,4	7,7	7,9	7,12	5.0	5.5	7.0	8.0	9.5
9,3	9,4	9,7	9,9	9,12	6.0	6.5	8.0	9.0	10.5
3,3	3,4	3,7	3,9	12,12	7.5	8.0	9.5	10.5	12.0

$$\mathbb{E}(\overline{X}) = \sum_{i=1}^{n} \overline{x} f(\overline{x}) \quad \text{y} \quad \mathbb{V}(\overline{X}) = \mathbb{E}(\overline{X}^2) - (\mathbb{E}(\overline{X}))^2$$

$$\mathbb{E}(\overline{X}) = \sum_{i=1}^{n} \overline{x} f(\overline{x}) = 3(1/25) + 3.5(2/25) + ... + 12(1/25) = 7$$

$$\mathbb{E}(\overline{X}^2) = \sum_{i=1}^{n} \overline{x}^2 f(\overline{x}) = 3^2(1/25) + ... + 12^2(1/25) = \frac{1360}{25}$$

$$\mathbb{V}(\overline{X}) = \mathbb{E}(\overline{X}^2) - (\mathbb{E}(\overline{X}))^2 = \frac{1360}{25} - 7^2 = 5.4$$

Se debe tener en cuenta, que:

$$\mathbb{V}(\overline{X}) = \sigma^2_{\overline{x}} = \frac{\sigma^2}{n} = \frac{10.8}{2} = 5.4$$

En R:

```
pob<-c(3,4,7,9,12)
a<-sapply(1:1e4,function(x){mean(sample(pob,2,T))})
mean(a)
```

```
## [1] 6.9692
```

```
var(a)
```

```
## [1] 5.335535
```

c. Muestras de tamaño 2, sin reposición

3,4	3,7	3,9	3,12	3.5	5.0	6.0	7.5
4,3	4,7	4,9	4,12	3.5	5.5	6.5	8.0
7,3	7,4	7,9	7,12	5.0	5.5	8.0	9.5
9,3	9,4	9,7	9,12	6.0	6.5	8.0	10.5
3,3	3,4	3,7	3,9	7.5	8.0	9.5	10.5

$$\mathbb{E}(\overline{X}) = \sum_{i=1}^{n} \overline{x} f(\overline{x}) \quad \text{y} \quad \mathbb{V}(\overline{X}) = \mathbb{E}(\overline{X}^2) - (\mathbb{E}(\overline{X}))^2$$

$$\mathbb{E}(\overline{X}) = \sum_{i=1}^{n} \overline{x} f(\overline{x}) = 3.5(2/20) + 5(2/20) + ... + 10.5(2/20) = 7$$

$$\mathbb{E}(\overline{X}^2) = \sum_{i=1}^{n} \overline{x}^2 f(\overline{x}) = 3.5^2(2/20) + ... + 10.5^2(2/20) = \frac{1061}{20}$$

$$\mathbb{V}(\overline{X}) = \mathbb{E}(\overline{X}^2) - (\mathbb{E}(\overline{X}))^2 = \frac{1061}{20} - 7^2 = 4.05$$

Se debe tener en cuenta, que:

$$\mathbb{V}(\overline{X}) = \sigma_{\overline{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{10.8}{2}\left(\frac{5-2}{5-1}\right) = 4.05$$

En R:

```
pob<-c(3,4,7,9,12)
a<-replicate(1e5, mean(sample(pob,2,F)) )
mean(a)
```

```
## [1] 6.99492
```

```
var(a)
```

```
## [1] 4.044285
```

Notas

1. La aproximación de $\overline{X}$ a la distribución normal$(N(\mu, \sigma^2/n))$ es buena, si $n \geq 30$, sin importa si la distribución en discreta o continua.
2. Si la muestra aleatoria es escogida de una población normal $N(\mu, \sigma^2)$, entonces, la distribución de $\overline{X}$ es exactamente normal $N(\mu, \sigma^2/n)$, para cualquier tamaño de muestra $n \geq 2$
3. Si el muestro es sin reeemplazo en una población finita de tamaño N, entonces, la varianza de la media muestral es:

$$\mathbb{V}(\overline{X}) = \sigma^2_{\overline{x}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

   El coeficiente $\left(\frac{N-n}{N-1}\right)$ se denomina factor de correción para población finita
4. La desviación estándar de una estadística es conocidad como error típico o error estándar

Ejemplo: El número de automóviles por familia es una variable aleatoria X cuya distribución de probabilidad es como sigue:

x	0	1	2	3	4
$f_X(x)$	$4/12$	$4/12$	$2/12$	$1/11$	$1/12$

- Halle la media y la varianza de la población del número de automóviles por familia.Si se escoge una muestra de 49 familias. ¿Cuál es la probabilidad de que la media muestral de autos por familia esté entre 1 y 2?

En R:

```
#a.
poba<-c(0,0,0,0,1,1,1,1,2,2,3,4)
mean(poba)
```

```
## [1] 1.25
```

```
var(poba)*(48/49)
```

```
## [1] 1.625232
```

```r
#b.
poba<-c(0,0,0,0,1,1,1,1,2,2,3,4)
mues<-sapply(1:1e4, function(x) {
mean(sample(poba,49,T))} )
mean(mues)
```

```
## [1] 1.248696
```

```r
var(mues)
```

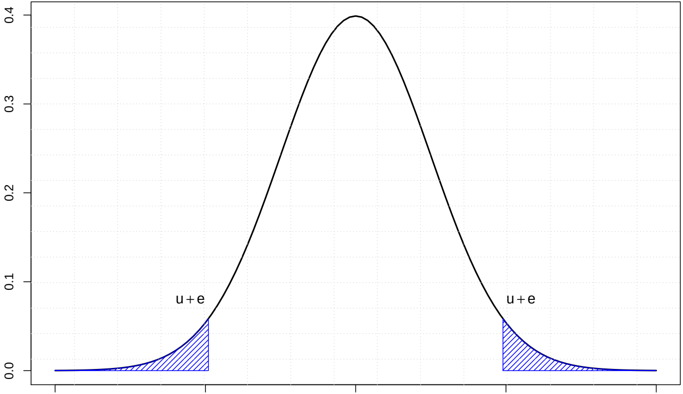
```
## [1] 0.03100554
```

Teorema Central del Límite

Desigualdad de Chebyshev

Sea X una variablea aleatoria, con media $\mathbb{E}(X) = \mu$ y varianza $\mathbb{V}(X) = \sigma^2$, entonces:

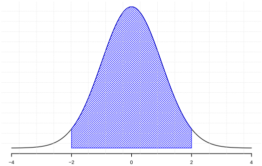
$$\forall\, e > 0, P(|X - \mu| \geq e) \leq \frac{\sigma^2}{e^2}$$



Otras formas de la desigualdad de Chebyshev

- $\forall\, e > 0, P(|X - \mathbb{E}(X)| \geq e) \leq \dfrac{\mathbb{V}(X)}{e^2}$
- $\forall\, k \geq 1, P(|X - \mu| \geq k\,\sigma) \leq \dfrac{1}{k^2}$

Ejemplo: Sea X una variable aleatorio con distribución normal estándar $N(0,1)$. Hallar el valor de: $P(-2 \leq X \leq 2)$




```
pnorm(2)-pnorm(-2)
```

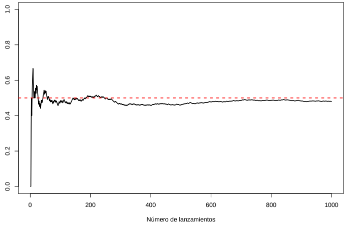
```
## [1] 0.9544997
```

Ley de los grandes números

Sea $\{X_n\}_{n \in N}$ una sucesión de variables aleatorias independientes e indénticamente distribuídas con media $\mu < \infty$. Entonces:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{} \mu$$

en donde la convergencia se cumple en el sentido casi seguro(ley fuerte) y también en probabilidad (ley débil) Seam $X1_, X_2, \ldots$ v.a con distribución de *Bernoulli*$(p = 0.5)$



Convergencia en Probabilidad y Casi segura

Sean $X_1, X_2, \ldots$ variablea aleatorias y sea X una variable aleatoria.

- Ley Débil de Khintchin: Convergencia en probabilidad

X_n converge a X en probabilidad, si para todo $\epsilon > 0$

$$P(|X_n - X| \geq \epsilon) \longrightarrow 0 \quad \text{cuando} \quad n \rightarrow \infty$$

$$X_n \xrightarrow{p} X$$

- Ley fuerte de Kolmogorov: convergenvia casi segura

X_n converge a X casi seguramente , si $P(X_n \longrightarrow X)$

cuando $n \rightarrow \infty$, es decir, el evento: $A = \{\omega : X_n(\omega)$

$\longrightarrow X(\omega)\}$ tiene probabilidad 1

Ejemplo: Ley fuerte de los GN:Lanzamientos de una moneda: Las v.a. $X_n = Sn/n$ forman un sucesión de valores números del resultado de los lanzamiento, entonces $X_n(\omega) = (1/n)*$(el número de resultados posibles). luego:

$$X_n \longrightarrow 1/2$$

Ejemplo: Ley débil de los GN: luego:

$$S_n = \sum_{i=1}^{n} \frac{X_i}{n}, \text{ entonces}$$

$$\mathbb{E}(S_n) = \mu \quad \mathbb{V}(S_n) = \frac{\sigma^2}{n}$$

Usando la desigualdad de chebyshev:

$$P(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon) \leq \frac{\sigma^2}{n\,\epsilon^2} \longrightarrow 0$$

Teorema Central del Límite

Convergencia en Distribución: Sea $X_1, X_2, \dots$ variables aleatorias y sean F y F_n las funciones de Distribución de X y X_n, respectivamente. Decimos que X_n converge a X en Distribución, y escribimos:

$$X_n \xrightarrow{\mathcal{D}} X$$

si,

$$F_n(x) \longrightarrow F(x)$$

Si $X_1, X_2, \dots, X_n$ contituye una m.a. de tamaño n de una población infinita que tiene media $\mathbb{E}(X) = \mu$ y varianza $\mathbb{V}(X) = \sigma^2$, entonces:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1)$$