Media muestral: es un estimador de $E[X]$.

Lo vamos a denotar $\bar{X}(n)$

$$\bar{X}(n) = \frac{X_1 + X_2 + \cdots + X_n}{n}$$; toma muestras de tamaño n $X_1, X_2, \ldots, X_n$.

Si las X_i son indep, igualmente distribuidas:

$$E[\bar{X}(n)] = E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} n E[X]$$
$$= E[X]$$

- $\hat{\theta} = \bar{X}(n)$ es un estimador insesgado de $E[X]$.
- $Var(\hat{\theta}) = Var(\bar{X}(n)) = Var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \boxed{\frac{Var(X)}{n}}$

Una primera solución para disminuir $Var(\bar{X}(n))$ es aumentando el tamaño de la muestra.

- En las simulaciones, un argumento para detener las simulaciones puede ser acotar $Var(\bar{X}(n)) = \frac{Var(X)}{n}$.

$$Var(X) = \sigma^2$$

Usualmente también se desconoce la varianza σ^2. Luego también es necesario estimarla.

Estimador de $Var(X) = S^2(n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}(n))^2$; insesgado

Para controlar las iteraciones y detenerse cuando $\frac{S^2(n)}{n} < d$ (para cierto d) es necesario mantener los valores $X_1, X_2, \dots, X_n$; porque son necesarios para calcular $S^2(n+1) = \frac{1}{n} \sum (X_i - \bar{X}(n+1))^2$.

Por esto se utilizan fórmulas recursivas para $\bar{X}(n)$ y $S^2(n)$.

Para $\bar{X}(n+1)$

$$\boxed{\bar{X}(n+1)} = \frac{1}{n+1} \sum_{i=1}^{n+1} X_i = \frac{1}{n+1} \left(\sum_{i=1}^{n} X_i + X_{n+1} \right)$$

$$= \frac{1}{n+1} \left(n \cdot \bar{X}(n) + X_{n+1} \right) = \frac{1}{n+1} \left((n+1-1) \bar{X}(n) + X_{n+1} \right)$$

$$= \boxed{\bar{X}(n) + \frac{X_{n+1} - \bar{X}(n)}{n+1}}$$

Para $S^2(n+1)$

$$n\,S^2(n+1) = \sum_{i=1}^{n+1}\left(X_i - \bar{X}(n+1)\right)^2$$

$$= \sum_{i=1}^{n}\left(X_i - \bar{X}(n) + \bar{X}(n) - \bar{X}(n+1)\right)^2 + \left(X_{n+1} - \bar{X}(n+1)\right)^2$$

$$= \sum_{i=1}^{n}\left(X_i - \bar{X}(n)\right)^2 + \sum_{i=1}^{n}\left(\bar{X}(n) - \bar{X}(n+1)\right)^2 + 2\underbrace{\sum_{i=1}^{n}\left(X_i - \bar{X}(n)\right)}_{= n\bar{X}(n) - n\bar{X}(n) = 0}\left(\bar{X}(n) - \bar{X}(n+1)\right) + \left(X_{n+1} - \bar{X}(n+1)\right)^2$$

$$= (n-1)S^2(n) + n\left(\bar{X}(n) - \bar{X}(n+1)\right)^2 + \left(X_{n+1} - \bar{X}(n+1)\right)^2$$

$$X_{n+1} - \bar{X}(n+1) = \frac{(n+1)}{(n+1)}X_{n+1} - \left(\frac{n+1}{n+1}\bar{X}(n) + \frac{X_{n+1} - \bar{X}(n)}{n+1}\right) = \frac{n}{n+1}\left(X_{n+1} - \bar{X}(n)\right) = n\left(\bar{X}(n+1) - \bar{X}(n)\right)$$

$$n\, S^2(n+1) = (n-1)\, S^2(n) + \underbrace{(n+n^2)}_{n(n+1)} \left(\bar{X}(n) - \bar{X}(n+1)\right)^2$$

$$\boxed{S^2(n+1) = \left(1 - \frac{1}{n}\right) S^2(n) + (n+1)\left(\bar{X}(n) - \bar{X}(n+1)\right)^2}$$

```
def Media_Muestral_X(d):
    'Estimación del valor esperado con ECM<d'
    Media = simular X # X(1)
    Scuad, n = 0, 1 #Scuad = S^2(1)
    while n <= 100 or sqrt(Scuad/n) > d:
        n += 1
        simular X
        MediaAnt = Media
        Media = MediaAnt + (X - MediaAnt) / n
        Scuad = Scuad * (1 - 1 /(n-1)) + n*(Media - MediaAnt)**2
    return Media
```

Estimador de una proporción

Lo usamos cuando queremos estimar una probabilidad.

Por ejemplo, $P(X>3)$.

$$Y = \begin{cases} 1 & \text{si } X>3 \\ 0 & \text{si } X<3 \end{cases} \qquad E[Y] = P(X>3)$$

$Y \sim B(p)$ (Bernoulli)

$$Y = \begin{cases} 1 & p \\ 0 & 1-p \end{cases}$$

$$E[Y] = p\cdot 1 + (1-p)\cdot 0 = p$$

$$\begin{aligned} Var(Y) &= E[Y^2] - E[Y]^2 \\ &= p - p^2 = p(1-p) \end{aligned}$$

Para estimar la proporción usamos $\hat{p} = \bar{X}(n) \left(\frac{Y_1 + Y_2 + \cdots + Y_n}{n} \right)$

Un estimador para la varianza de Y es $\hat{p}(1-\hat{p})$;

y entonces la varianza del estimador de p: $Var(\hat{p}) = \frac{p(1-p)}{n}$

Para controlar la simulación $\left(\frac{Var(Y_i)}{n} < d\right)$ usamos

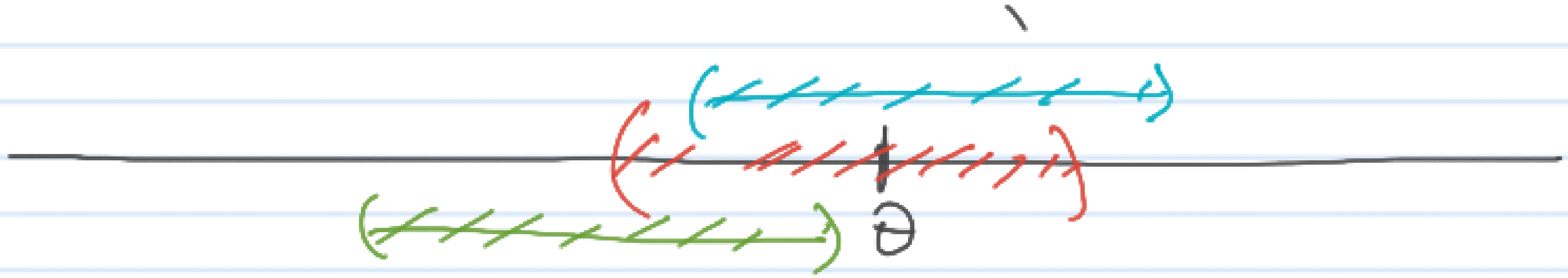
$$\frac{\hat{p}(1-\hat{p})}{n} = \frac{\bar{X}(n)(1-\bar{X}(n))}{n}$$

```
def estimador_p(d):
    'Estimación de proporción con ECM<d'
    p = 0
    n = 0
    while n <= 100 or sqrt( p * (1-p) / n) > d:
        n += 1
        Simular X
        p = p + (X - p) / n
    return p
```

$p = \bar{X}(n)$

Estimador por intervalos

Se quiere estimar $E[X] = \Theta$; más precisamente determinar un intervalo que contenga al parámetro Θ con nivel de confianza $(1-\alpha)$



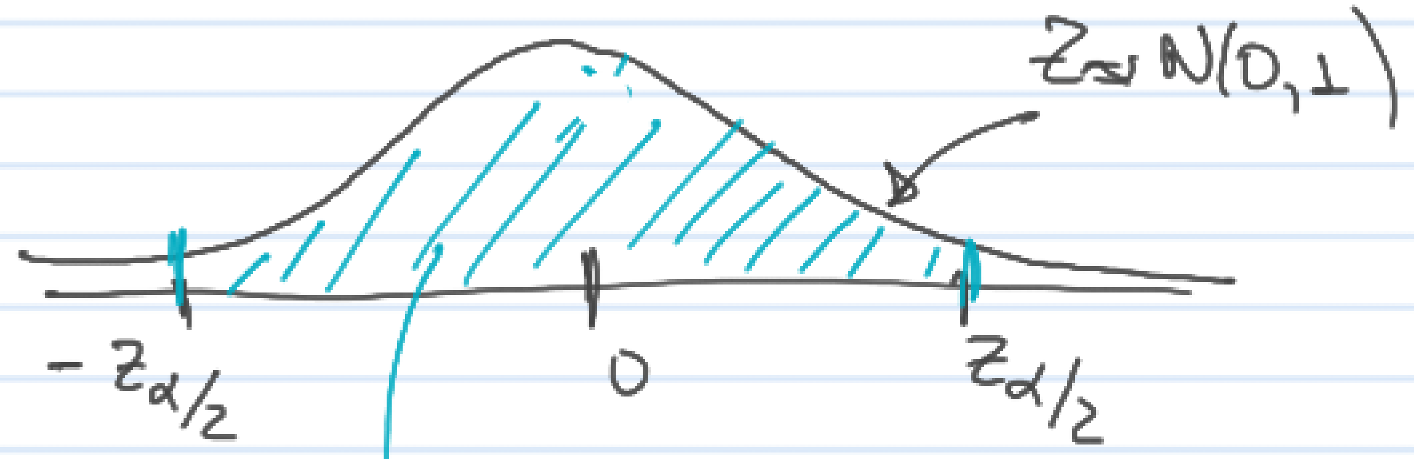
Para construir estos intervalos usamos que

$$\lim_{n \to \infty} \frac{X_1 + \dots + X_n}{n}$$ tiene una distribución aproximadamente normal.

$$E\left[\frac{X_1+\dots+X_n}{n}\right] = E[X] = \theta \qquad \left(E[\bar{X}(n)] = \theta \right)$$

$$Var\left(\frac{X_1+\dots+X_n}{n}\right) = \frac{Var(X)}{n} = \sigma^2/n$$

$$\boxed{\frac{\bar{X}(n)-\theta}{\sigma/\sqrt{n}} \simeq N(0,1)}$$



$$P(|Z| < z_{\alpha/2}) = (1-\alpha)$$

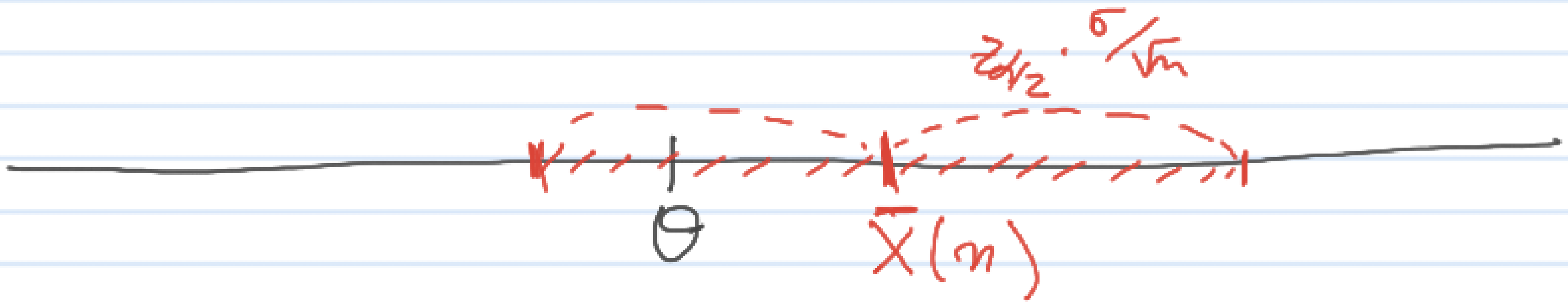
Luego: $$P\left(\left|\frac{\bar{X}(n)-\theta}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-z_{\alpha/2} < \frac{\theta-\bar{X}(n)}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$

Equivalentemente:

$$P\left(\bar{X}(n) - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \theta < \bar{X}(n) + z_{\alpha/2} \cdot \sigma/\sqrt{n}\right) = 1-\alpha$$

Intervalo de confianza: $\left(\bar{X}(n) - z_{\alpha/2} \cdot \sigma/\sqrt{n}, \quad \bar{X}(n) + z_{\alpha/2}\, \sigma/\sqrt{n}\right)$



Si no conocemos, usamos $\sqrt{S^2(n)}$

Con α: definimos el nivel de confianza.

Ejemplo: $z_{\alpha/2} = 1.96$ entonces $1-\alpha = 95\%$

Incrementando n: mejoramos la varianza y disminuimos el tamaño del intervalo.

<u>Ancho del intervalo</u> $= 2 \cdot \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha/2}$

ó $2\sqrt{S^2(n)/n} \cdot z_{\alpha/2}$

ó $2\sqrt{\frac{\bar{X}(n)(1-\bar{X}(n))}{n}} \cdot z_{\alpha/2}$