Análisis de Datos Categóricos

Momentos & MLE

Mauricio Bucca
github.com/mebucca
mebucca@uc.cl

26 August, 2024

Valor Esperado de variables discretas

Valor Esperado

El valor esperado de una variable es el análogo teórico de un promedio. Los posibles valores de la variable se ponderan por su probabilidad de ocurrencia. En el caso de variables discretas:

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_i x_i \times \mathbb{P}(X = x_i) \\
&\equiv \sum_i x_i \times f(x_i)
\end{aligned}
$$

Es teórico porque esta información la podemos saber *a priori*, sin necesidad de datos.

Análogamente, para variables continuas:

$$\mathbb{E}(X) = \int x f(x) dx$$

Valor Esperado

Por ejemplo, supongamos que Y es una variable que resulta de tirar un dado "justo". ¿Cuál es el valor esperado de Y?

$$\mathbb{E}(Y) = \sum_i y_i \times \mathbb{P}(Y = y_i)$$

$$= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6}$$

$$= 3.5$$

Valor Esperado, algunas propiedades útiles

1) El valor esperado de una constante es una constante.

$$\mathbb{E}(c) = c$$

2) Si X es una variable aleatoria y c una constante, entonces

$$\mathbb{E}(X + c) = \mathbb{E}(X) + c$$

3) Si X es una variable aleatoria y c una constante, entonces

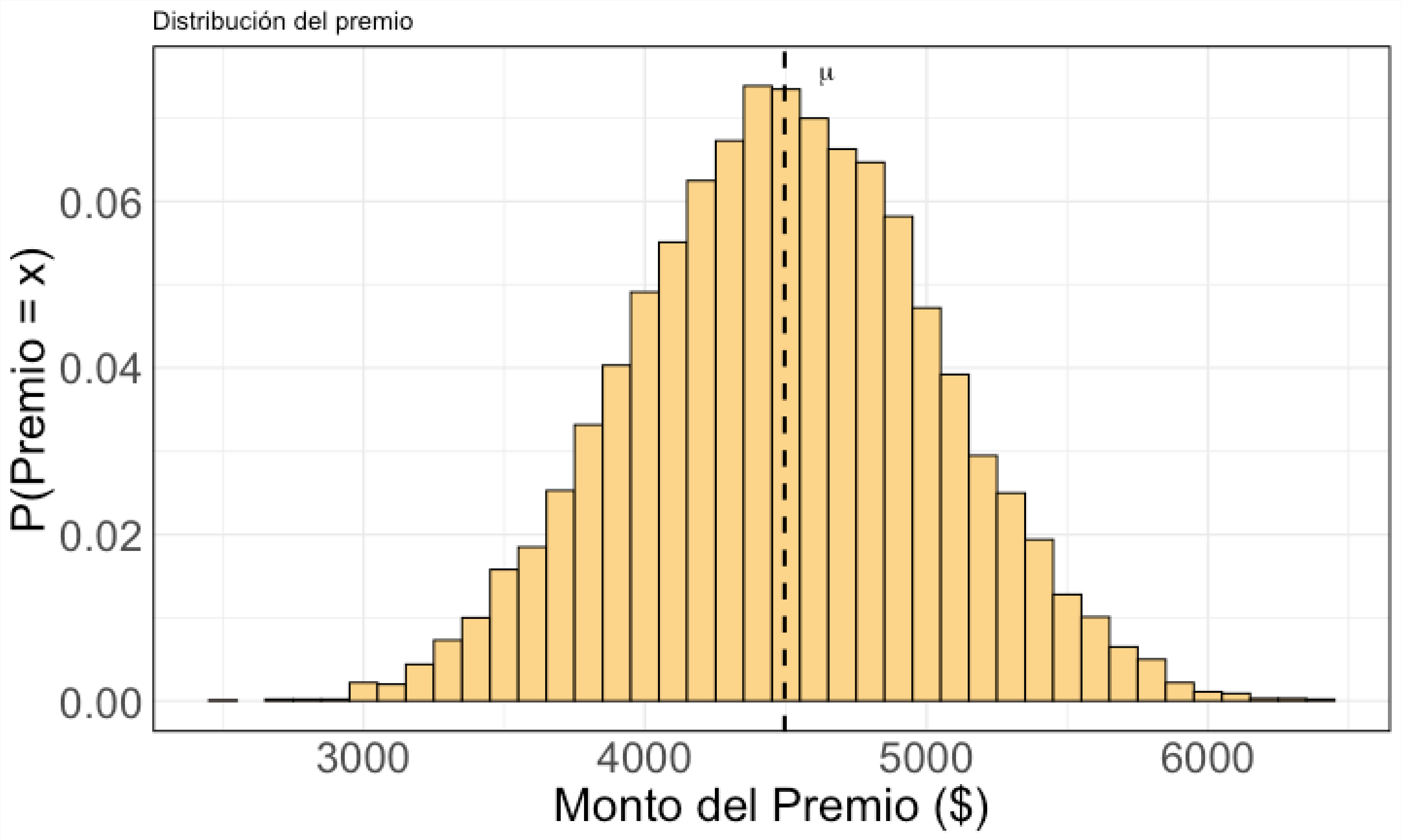
$$\mathbb{E}(cX) = c\mathbb{E}(X)$$

4) Si X e Y son variables aleatorias (sin importar si $X \perp Y$ o no), entonces

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

Valor Esperado, ejemplo

Por ejemplo, supongamos que X_i es la variable que resulta de tirar un dado "justo". Participamos de un concurso que consiste en tirar el mismo dado 10 veces. El premio (G) es \$ 1000 de base, más el resultado de cada dado i multiplicado por 100. ¿Cuánto es el premio esperado?



Valor Esperado, ejemplo

Por ejemplo, supongamos que X_i es la variable que resulta de tirar un dado "justo". Participamos de un concurso que consiste en tirar el mismo dado 10 veces. El premio (G) es $\$$ 1000 de base, más el resultado de cada dado i multiplicado por 100. ¿Cuánto es el premio esperado?

$$G = 1000 + \sum_{i=1}^{n=10} X_i \times 100 \quad \text{por tanto,}$$

$$\mathbb{E}(G) = \mathbb{E}(1000 + \sum_{i=1}^{n=10} X_i \times 100)$$

$$\mathbb{E}(G) = 1000 + 100 \times \sum_{i=1}^{n=10} \mathbb{E}(X_i)$$

$$\mathbb{E}(G) = 1000 + 10(3.5 + 3.5 + \cdots + 3.5)) = 1000 + 100 \times 10 \times 3.5 = \$4500$$

Valor Esperado de variables discretas

Bernoulli

Si X es una variable Bernoulli, su valor esperado viene dado por:

$$
\begin{aligned}
\mathbb{E}(X) = \sum_i x_i \times \mathbb{P}(X = x_i) &= \sum_i x_i \times p^{x_i}(1-p)^{1-x_i} \\
&= 1 \times p + 0 \times (1-p) \\
&= p
\end{aligned}
$$

Binomial

Si X es una variable Binomial, su valor esperado viene dado por:

$$\mathbb{E}(X) = np$$

Pregunta: ¿Cuántas "Caras" debo esperar si tiro 200 monedas "justas"?

Respuesta: $np = 200 \times 0.5 = 100$

Varianza de variables discretas

Varianza

La varianza de una variable aleatoria es el análogo teórico de la varianza de los datos. Mide cuánta dispersión existe en torno al centro (la media). Formalmente, en el caso de variables aleatorias discretas:

$$\mathbb{V}\text{ar}(X) = \sum_i \left(x_i - \mathbb{E}(X)\right)^2 \times f(x_i)$$

Análogamente, para variables continuas:

$$\mathbb{V}\text{ar}(X) = \int \left(x - \mathbb{E}(X)\right)^2 f(x)dx$$

Equivalentemente,

$$\mathbb{V}\text{ar}(X) = \mathbb{E}\left([X - \mathbb{E}(X)]^2\right)$$

Varianza, algunas propiedades útiles

1) La varianza de una constante es cero.

$$\mathbb{V}\mathrm{ar}(c) = 0$$

2) Si X es una variable aleatoria y c una constante, entonces

$$\mathbb{V}\mathrm{ar}(X + c) = \mathbb{V}\mathrm{ar}(X)$$

3) Si X es una variable aleatoria y c una constante, entonces

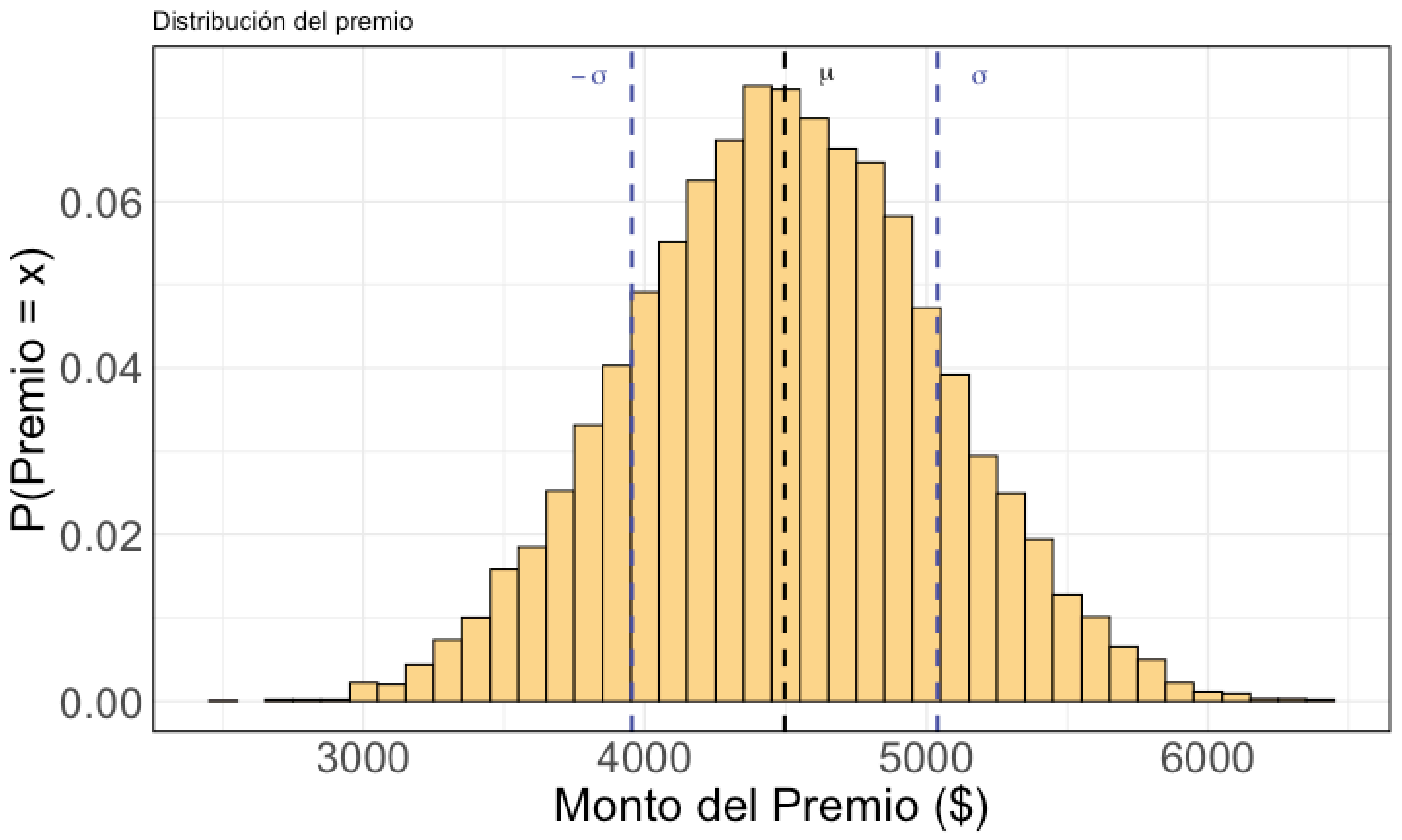
$$\mathbb{V}\mathrm{ar}(cX) = c^2\mathbb{V}\mathrm{ar}(X)$$

4) Si X e Y son dos variables aleatorias **independientes**, entonces

$$\mathbb{V}\mathrm{ar}(X \pm Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$$

Varianza, ejemplo

Por ejemplo, supongamos que X_i es la variable que resulta de tirar un dado "justo". Participamos de un concurso que consiste en tirar el mismo dado 10 veces. El premio (G) es $ 1000 de base, más el resultado de cada dado i multiplicado por 100. ¿Cuánto es la desviación estándar del premio?



Varianza, ejemplo

Por ejemplo, supongamos que X_i es la variable que resulta de tirar un dado "justo". Participamos de un concurso que consiste en tirar el mismo dado 10 veces. El premio (G) es \$ 1000 de base, más el resultado de cada dado i multiplicado por 100. ¿Cuánto es la desviación estándar del premio?

$$G = 1000 + \sum_{i=1}^{n=10} X_i \times 100 \quad \text{por tanto,}$$

$$\mathbb{V}\text{ar}(G) = \mathbb{V}\text{ar}(1000 + \sum_{i=1}^{n=10} X_i \times 100)$$

$$\mathbb{V}\text{ar}(G) = \mathbb{V}\text{ar}(1000) + 100^2 \times \sum_{i=1}^{n=10} \mathbb{V}\text{ar}(X_i)$$

$$\mathbb{V}\text{ar}(G) = 0 + 100^2 \times 10 \times 2.9167 = \$291,670$$

$$\sigma_G = \sqrt{0 + 100 \times 100 \times 2.91670} = \$539.88$$

Varianza de variables discretas

Bernoulli

Si X es una variable Bernoulli, su varianza viene dada por:

$$\mathbb{V}\text{ar}(X) = \sum_i \left(x_i - \mathbb{E}(X)\right)^2 \times f(x_i)$$

$$= (1 - \mathbb{E}(X))^2 \times \mathbb{P}(X = 1) + (0 - \mathbb{E}(X))^2 \times \mathbb{P}(X = 0)$$

$$= (1 - p)^2 \times p + (0 - p)^2 \times (1 - p)$$

$$= p^2 - p^3 + p - 2p^2 + p^3$$

$$= p(1 - p)$$

Varianza de variables discretas

Binomial

Si X es una variable Binomial, su varianza viene dada por:

$$\mathbb{V}\mathrm{ar}(X) = n \times p(1-p)$$

Pregunta: ¿Cuánta variabilidad debo esperar en el número de "Caras" si tiro 200 monedas "justas"?

Respuesta: varianza es $n \times p(1-p) = 200 \times 0.5 \times 0.5 = 50$. La desviación estándar es $\sqrt{50} = 7.01$.

Varianza de variables Binomial

Ilustración via Monte Carlo simulation

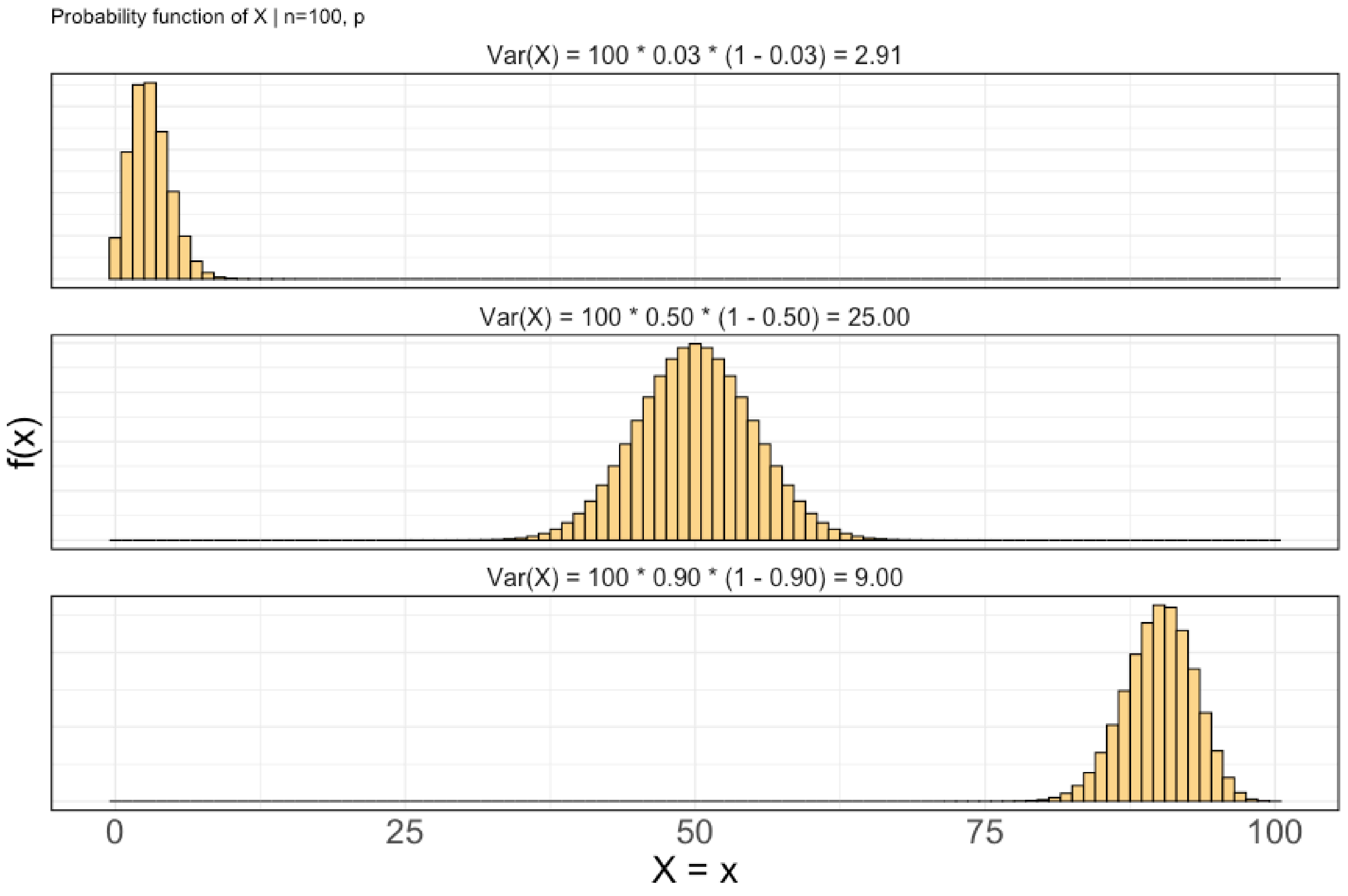
```r
# Repeat experiment of tossing 200 coins 10000 times
coins200 <- rbinom(10000, size=200, p=0.5)
glimpse(coins200)
```

```
##  int [1:10000] 103 86 92 112 99 105 94 112 95 95 ...
```

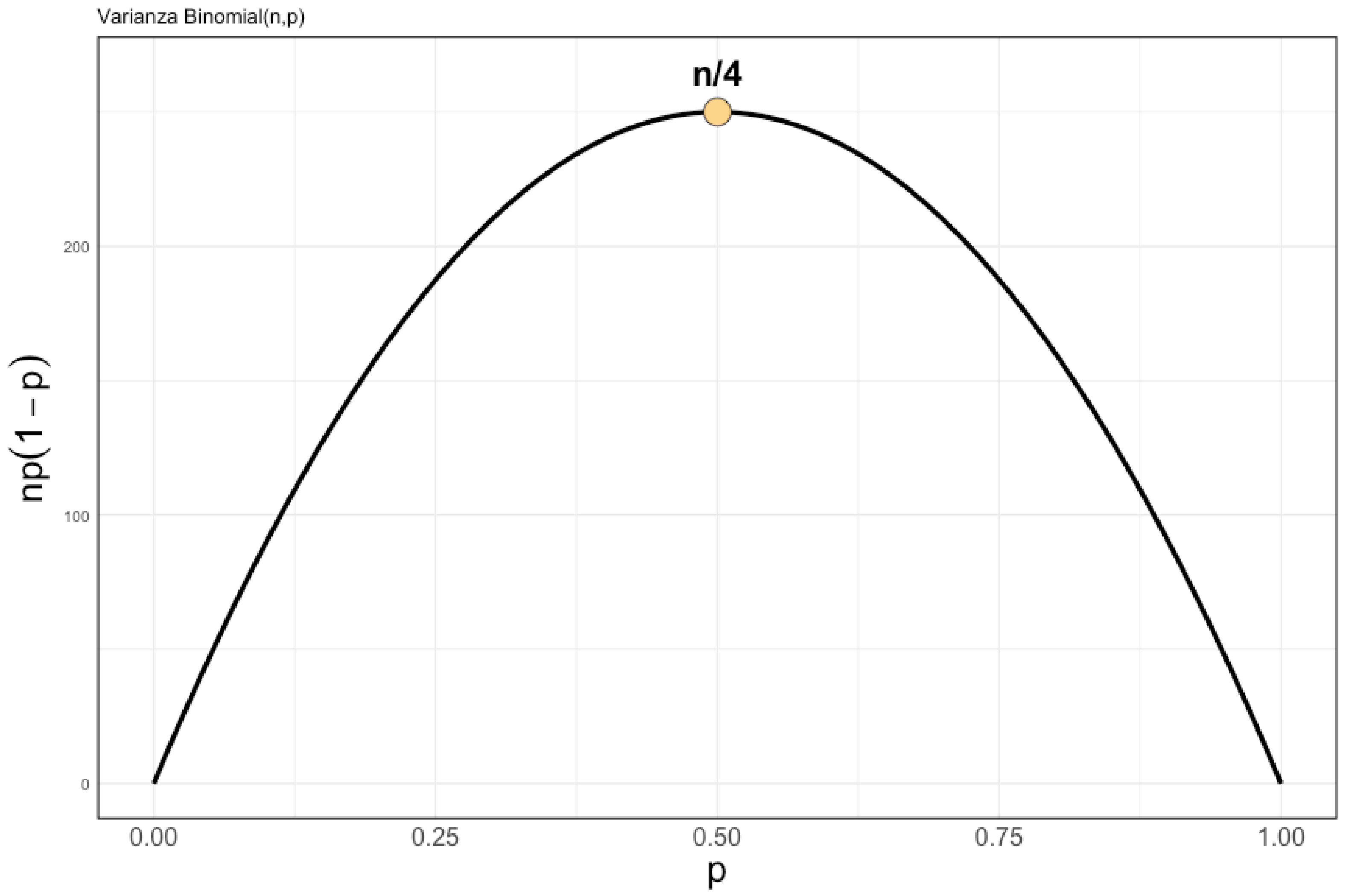
```r
moments = list(mean=mean(coins200), var=var(coins200))
print(moments)
```

```
## $mean
## [1] 99.997
##
## $var
## [1] 49.68256
```

Varianza variable Bernoulli/Binomial



Varianza variable Binomial



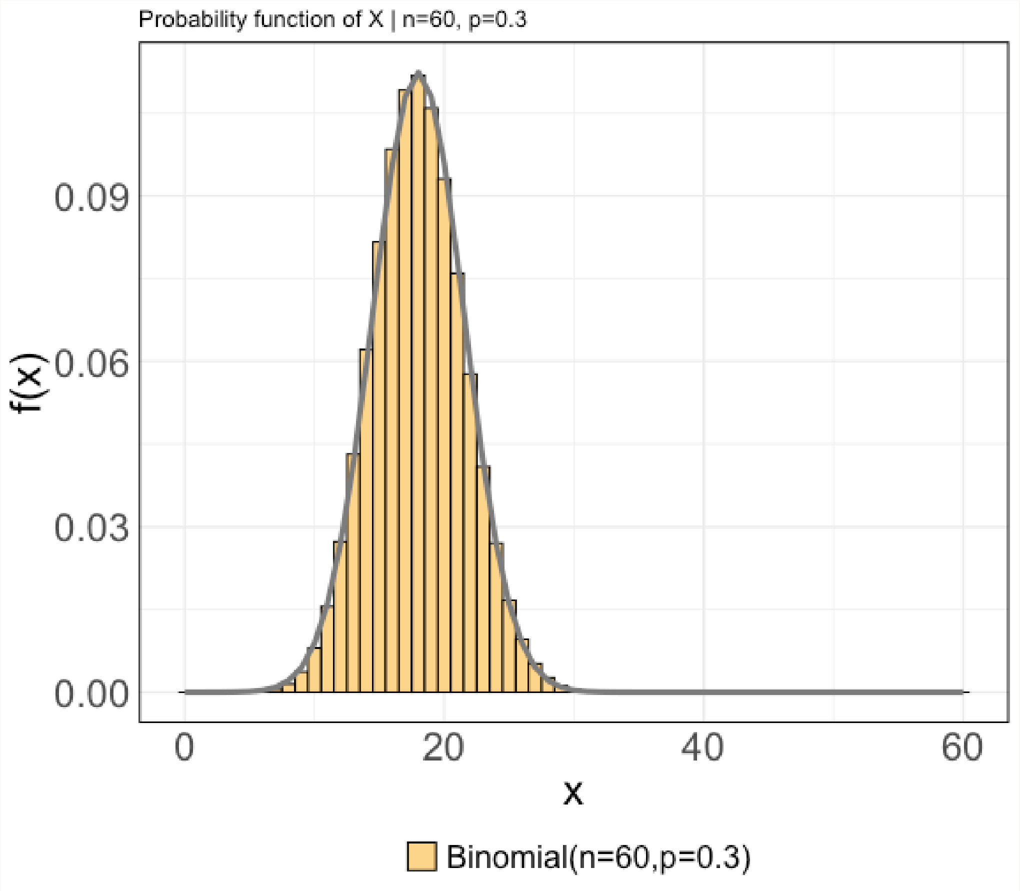
Distribución Binomial es asintóticamente Normal

- Si $X \sim \text{Binomial}(n, p)$, entonces:

$$X \xrightarrow{d} \text{Normal}(\mu = np, \quad \sigma = \sqrt{np(1-p)})$$

(cuando $n \to \infty$)

- Resultado muy conveniente



Estimación

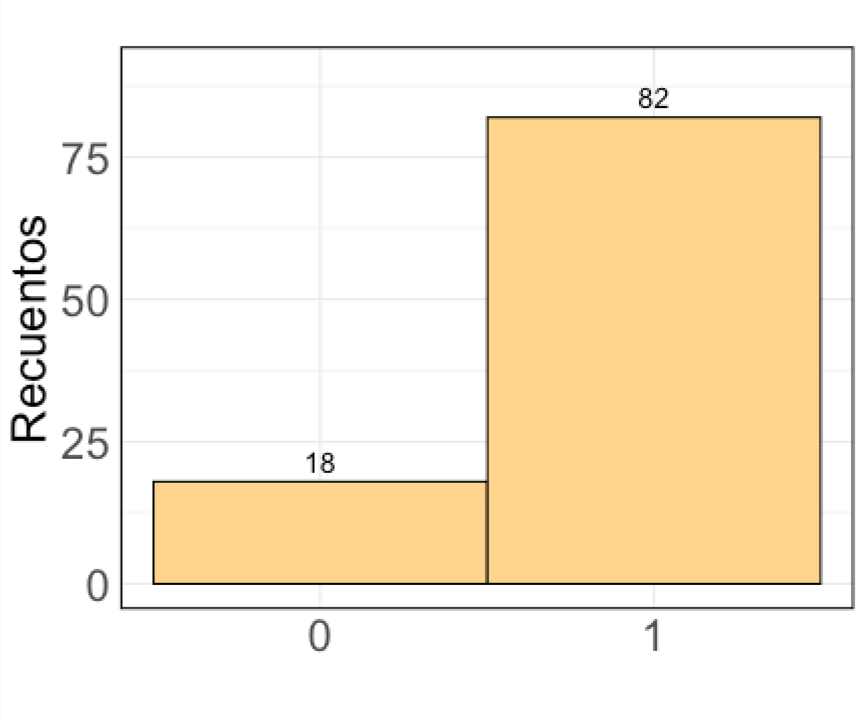
Maximum Likelihood Estimation (MLE)

Estimación

Estimación

Modelos estadísticos: ¿Cuáles son los valores más **plausibles**[1] de los *parámetros* dado los *datos* que observamos?

Ej. Supongamos que alguien lanza 100 veces la misma moneda y registra los resultados en una base de datos. Los datos se ven así:



- Lo que vemos en la izquierda son **datos**
- Datos: realización de n variables aleatorias
- Normalmente *no conocemos* la distribución de las variables
- Datos nos dan una pista sobre cuál podría ser esa distribución
- **Estadística**: aprender de los datos para ***estimar*** los parámetros que los generan

[1] Notar que no dice "más probables"!

Estimación via Maximum Likelihood (MLE)

Previamente lanzamos la misma moneda 100 veces y obtuvimos "Cara" (1) 82 veces. ¿Qué valor de p es más plausible ("likely") que genere estos datos?

MLE es justamente la formalización de esta pregunta. Pasos:

1) Decidir sobre la distribución subyacente que genera los datos. En este caso, podemos asumir que:

- Cada lanzamiento $X_1, X_2, \ldots X_{100} \sim \text{Bernoulli}(p)$, donde X's son iid

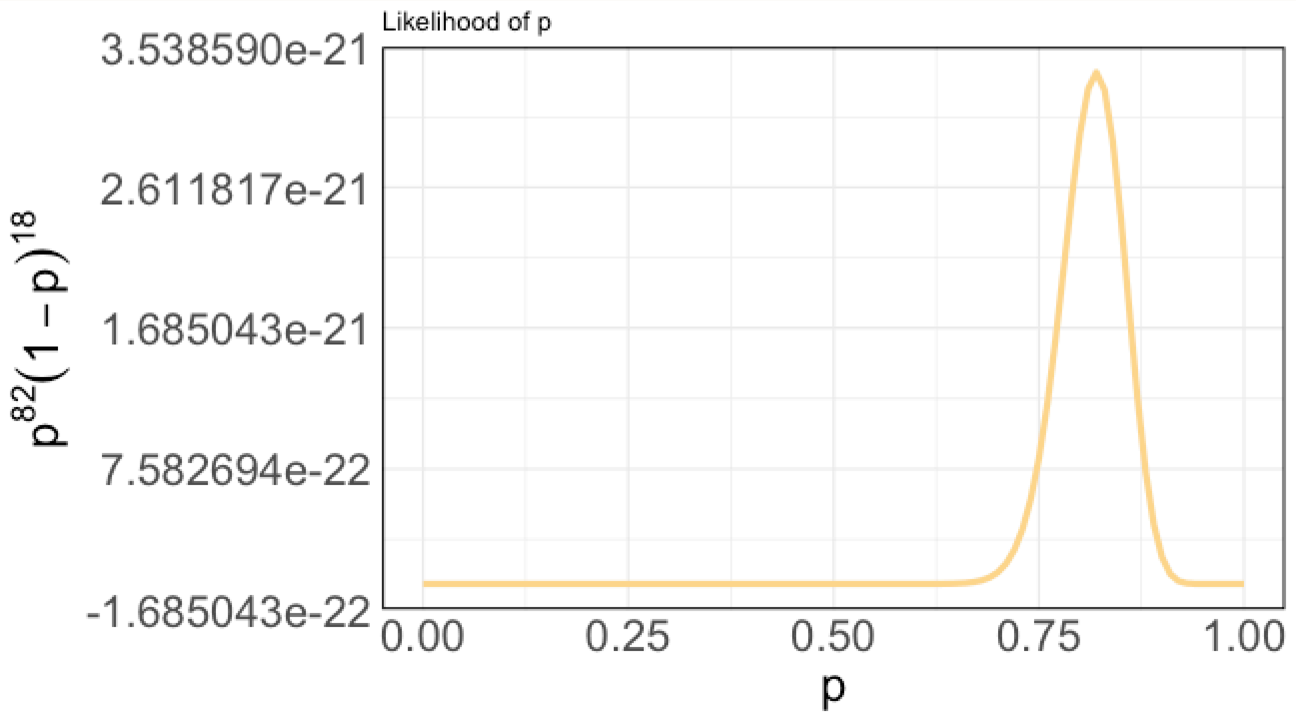
2) Escribir una función que cuantifique la plausibilidad de diferentes valores del parámetro. Dicha función se denomina **likelihood function**:

- $\mathcal{L}(p \mid \text{ Datos}) = \mathbb{P}(\text{ Datos} : \{1{,}0{,}1{,}1{,}....0{,}1\} \mid p)$

- $\mathcal{L}(p \mid \text{ Datos}) = \mathbb{P}(x_1)\mathbb{P}(x_2)\ldots\mathbb{P}(x_{100}) = p^{82}(1-p)^{18}$

Estimación via Maximum Likelihood (MLE)

Podemos inspeccionar visualmente la "likelihood" de diferentes valores p.



Intuitivamente: habiendo obtenido 82 caras, $p = 0.82$ es el valor más plausible de p

Estimación via Maximum Likelihood (MLE)

3) Encontrar matemáticamente el valor de p que maximiza $\mathcal{L}(p \mid \text{Datos})$.

- $\mathcal{L}(p \mid \text{Datos}) = \mathbb{P}(x_1)\mathbb{P}(x_2)\ldots\mathbb{P}(x_n) = \prod_{i=1}^{n} f(x_i) = p^k(1-p)^{n-k} \quad \text{donde } k = \sum x_i$
- Para facilitar el cálculo tomamos logaritmo natural en ambos lados (**log-likelihood**)
  - $\ell\ell(p) = \ln \mathcal{L}(p \mid \text{Datos}) = k\ln(p) + (n-k)\ln(1-p)$
- Calcular la primera* derivada de $\ell\ell(p)$ con respecto a p: pendiente de la curva en cada valor de p.
  - $\ell\ell'(p) = \frac{k}{p} - \frac{n-k}{1-p}$
- Encontrar el máximo de la función $\ell\ell(p)$: valor de p en el cual la curva no cambia, es decir cuando $\ell\ell'(p) = 0$
  - $\frac{k}{p} - \frac{n-k}{1-p} = 0$
- Después de resolver por p obtenemos:

$$p = \frac{k}{n} = \frac{\sum x_i}{n}$$

Estimación via Maximum Likelihood (MLE)

- El estimador ML de p es ....
- $\hat{p} = \frac{\sum x_i}{n}$
- Es decir, el porcentaje de 1's en la muestra!
- Intuitivo y elegante

Estimación via Maximum Likelihood (MLE)

Generalización

$$\hat{\boldsymbol{\beta}}_{MLE} = \arg\max_{\beta} \; \mathcal{L}(\boldsymbol{\beta} \mid \boldsymbol{X})$$

$\hat{\boldsymbol{\beta}}$ es el MLE de $\boldsymbol{\beta}$ si es el(los) valor(es) que maximiza(n) la "likelihood function", condicional en los datos observados.

- Recordar que $\mathcal{L}(\boldsymbol{\beta} \mid \boldsymbol{X}) = \mathbb{P}(\boldsymbol{X} \mid \boldsymbol{\beta})$.
- Requiere especificar de antemano la distribución conjunta de las observaciones (dif. de OLS, por ejemplo).
- ML es probablemente el approach de estimación más popular.
- Intuitivo, pero, por lo general, no tan simple como el ejemplo que vimos hoy.
- Normalmente la maximización se realiza numéricamente (ej. método Newton-Raphson)

Estimación via Maximum Likelihood (MLE)

Hasta la próxima clase. Gracias!

Mauricio Bucca
https://mebucca.github.io/
github.com/mebucca