Variable aleatoria univariante: principales distribuciones discretas

ESCUELA POLITÉCNICA SUPERIOR DE CÓRDOBA
Universidad de Córdoba
DEPARTAMENTO DE ESTADÍSTICA

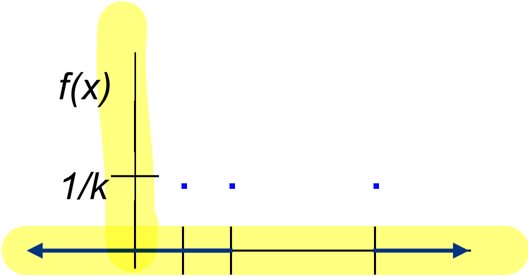
- **Distribución Uniforme.**

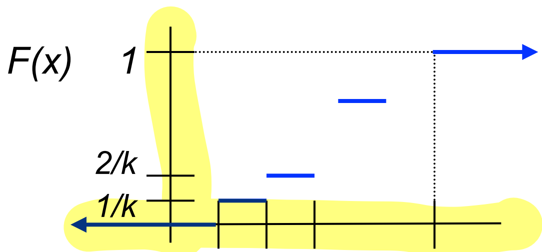
"La v.a. $\mathcal{X}$ toma valores x_1, x_2, ..., x_k con igual probabilidad"

$$f(x) = f(x,k) = \frac{1}{k} \quad \forall\ x = x_1, x_2, \ldots, x_k \in \mathcal{U}(k)$$

$$\mu = \frac{\sum_{i=1}^{k} x_i}{k} \quad ; \quad \sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}$$

En el caso de que $S_{\mathcal{X}} = \{ 1,2,\ldots,k \}$, $\mu = \dfrac{k+1}{2}$; $\sigma^2 = \dfrac{k^2 - 1}{12}$





En el caso particular de ser $S_{\mathcal{X}} = \{ x_0 \}$, la distribución de $\mathcal{X}$ se dice <u>degenerada</u> o <u>singular</u>.

- **Distribución de Bernouilli o Binaria.**

V / F , cara/cruz ...

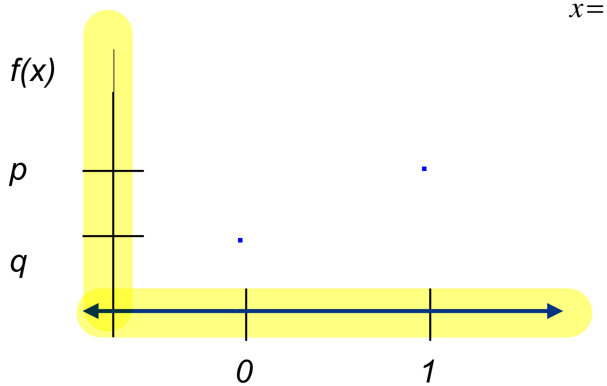
"Experimento aleatorio con dos posibles resultados": $S_x = \{0,1\}$

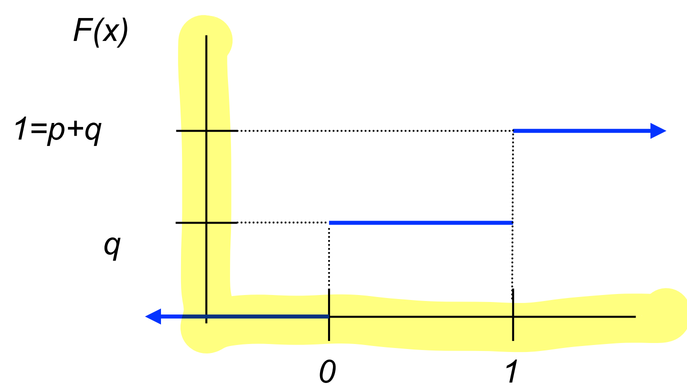
A = {"Éxito"} = {1} ; P(A) = p *B = A^c = {"Fracaso"} = {0} ; P(B) = q = 1-p*

$$f(x) = f(x,p) = p^x(1-p)^{1-x} \quad \in \mathcal{B}(p)$$

$$E[X] = \sum_{x=0}^{1} x p^x (1-p)^{1-x} = 0 + p = p$$

$$V[X] = E\left[X^2\right] - \left(E[X]\right)^2 = \sum_{x=0}^{1} x^2 p^x (1-p)^{1-x} - p^2 = p - p^2 = p(1-p)$$





Proceso de Bernouilli: / Experimento tipo Bernouilli

1. La probabilidad de éxito permanece constante para cada uno de los intentos.
2. Los intentos repetidos son independientes.

- **Distribución binomial.**

Dadas *n* pruebas de Bernouilli repetidas e independientes:

$$X_1, X_2, \cdots, X_n \in \mathcal{B}(p)$$

La v.a. X "número de éxitos total en las *n* pruebas" se dice binomial de parámetros *n* y *p*.

→ probabilidad constante en las repeticiones

→ nº de repeticiones

$$X = \sum_{i=1}^{n} X_i \in b(n, p)$$

$$f(x) = f(x, n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E[X] = E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n] = n\,p$$

$$V[X] = V[X_1 + X_2 + \ldots + X_n] = 1^2 V[X_1] + 1^2 V[X_2] + \ldots + 1^2 V[X_n] = n\,p\,q$$

"Asociada a la idea de muestreos **con reemplazamiento**".

En un proceso de control de calidad, en cada lote de 24 envases, se eligen dos al azar. Si ambos están correctos, se acepta el lote completo. Se supone que en el lote hay 5 envases defectuosos. Se pide calcular la probabilidad de aceptar el lote completo.

Para calcular la probabilidad pedida, habrá que determinar en primer lugar la distribución de la variable aleatoria:

$\mathcal{X}$: “Número de envases defectuosos en la muestra”

Distribución que dependerá de cómo se realice la extracción de los dos envases que conforman la muestra.

- Supongamos muestreo <u>con reemplazamiento</u>

Se elige un envase al azar, y después de examinado es devuelto antes de la extracción del segundo envase.

Si se entiende por “Exito” : “Envase defectuoso”, es posible definir dos variables aleatorias con distribución

$$\mathcal{X}_i \in \mathcal{B}\left(p = \frac{5}{24}\right) \qquad i = 1,2$$

Si sumamos estas dos variables, dado que las extracciones son independientes,

$$\mathcal{X} = \sum_{i=1}^{2} \mathcal{X}_i \in b\left(n = 2, p = \frac{5}{24}\right) \qquad \text{con} \qquad S_X = (0,1,2)$$

Luego la probabilidad pedida será: $P(\mathcal{X} = 0) = \binom{2}{0}\left(\frac{5}{24}\right)^{0}\left(\frac{19}{24}\right)^{2-0} = 0.6267$

- **Distribución hipergeométrica**.

"La v.a. $\mathcal{X}$ nº de éxitos en una muestra aleatoria de tamaño *n* seleccionada de entre *N* resultados posibles, de los cuales *a* son éxitos y *b* fracasos (*a*+*b*=*N*), se dice hipergeométrica".

$$\mathcal{X} \in h(n,a,b)$$

$$f(x) = f(x,n,a,b) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}} \qquad \begin{cases} x \le a \\ n-x \le b \\ n \le a+b \end{cases}$$

$$\mathrm{E}\left[\mathcal{X}\right] = \mathrm{np}\,;\ \mathrm{V}\left[\mathcal{X}\right] = \frac{\mathrm{N-n}}{\mathrm{N-1}}\,\mathrm{n}\frac{\mathrm{a}}{\mathrm{N}}\frac{\mathrm{b}}{\mathrm{N}}$$

"Asociada a la idea de muestreos $\boxed{\textbf{sin}\ \text{reemplazamiento}}$".

En el ejemplo anterior, supongamos ahora muestreo sin reemplazamiento

Se elige un envase al azar, y después de examinado **no** se devuelve para hacer la extracción del segundo envase.

Si se entiende por **"Exito": "Envase defectuoso"**, la v.a. que nos mide si un envase es defectuoso o No, tiene por distribución:

Primer envase: $\mathcal{X}_1 \in \mathcal{B}\left(p_1 = \frac{5}{24}\right)$ Segundo envase: $\mathcal{X}_2 \in \mathcal{B}\left(p_2 = \frac{5}{23}\right)$ ó $\mathcal{X}_2 \in \mathcal{B}\left(p_2 = \frac{4}{23}\right)$

dependiendo de que el primero sea correcto o defectuoso.

En el conjunto de los dos envases que forman la muestra, la v.a.

$\mathcal{X}$: "Número de envases defectuosos en la muestra",

tiene por distribución: $$\mathcal{X} = \sum_{i=1}^{2} \mathcal{X}_i \in h\left(n = 2, a = 5, b = 19\right)$$

puesto que los dos extracciones son dependientes entre si, siendo su espacio muestral: $S_X = \left(0,1,2\right)$

Luego la probabilidad pedida es: $$P\left(\mathcal{X} = 0\right) = \frac{\binom{5}{0}\binom{19}{2}}{\binom{24}{2}} = 0.6196$$

- **Distribución de Poisson.**

La v.a. $\mathcal{X}$, nº de ocurrencias del suceso *A* (éxito) durante un gran número de pruebas, se dice de Poisson".

En términos coloquiales: "Una binomial con *n* grande y *p* pequeño".

$$\mathcal{X} \in \mathcal{P}(\lambda) \quad ; \quad \lambda = n\,p$$

Matemáticamente, si: $n \to \infty \quad \text{y} \quad p \to 0 \quad \text{y} \quad n\,p = \lambda = cte$

ocurre muy poco

Asociada con "sucesos raros". Algunos casos típicos pueden ser: nº de bacterias en un cultivo, nº de errores mecanográficos por página, nº de llamadas telefónicas recibidas, etc.

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{1}{x!}\frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x}\lambda^x \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x}$$

$$= \frac{\lambda^x}{x!} 1 \underbrace{\left(1-\frac{1}{n}\right)}_{1} \cdots \underbrace{\left(1-\frac{x-1}{n}\right)}_{1} \underbrace{\left[\left(1-\frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda}}_{e^{-\lambda}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{1} \rightarrow \frac{\lambda^x}{x!}e^{-\lambda}$$

Por lo tanto: $f(x) = f(x,\lambda) = \dfrac{\lambda^x}{x!}e^{-\lambda}$

- $E[X] = \lambda \quad ; \quad V[X] = \lambda$

Ej: Se dispone de una disolución de un cierto tipo de bacterias, con una concentración media de tres bacterias por ml. Calcular la probabilidad de que se tome una muestra no contaminada de un ml. de dicha disolución.

Sea X la variable aleatoria que mide **“El número de bacterias por ml. de disolución”.**

Como la media de bacterias por ml. es 3, la variable X sigue una distribución:

$$X \in \mathrm{P}(\lambda = 3)$$

ya que la media de una distribución de Poissón coincide con el valor del parámetro.

En un ml. de disolución, la probabilidad de que no esté contaminado es:

$$\mathrm{P}(X = 0) = \frac{e^{-3}\, 3^0}{0!} = 0.0497$$

- **Distribución geométrica.**

Dado un fenómeno dicotómico con alternativas *A* (con *P(A)=p*) y su complementario A^c ($P(A^c)=1-p=q$) "la v.a. que cuenta el número de la prueba en la que aparece el primer éxito, se dice geométrica" $X \in G(p)$

$$f(x) = f(\underbrace{x}_{\text{éxito}}, p) = \underbrace{q^{x-1}}_{\text{fracaso}} p \qquad E[X] = \frac{1}{p} \quad ; \quad V[X] = \frac{q}{p^2}$$

- **Distribución binomial negativa.**

Dado un fenómeno dicotómico con alternativas *A* (con *P(A)=p*) y su complementario A^c ($P(A^c)=1-p=q$), "la v.a. que cuenta el número de la prueba en la que aparece el éxito número k, se dice binomial negativa".

$X \in bn(k,p)$

$$f(x) = f(x,k,p) = \binom{x-1}{k-1} q^{x-k} p^k$$

$$E[X] = \frac{k}{p} \quad ; \quad V[X] = \frac{kq}{p^2}$$

Ej^o

Una máquina automática produce tornillos de uno en uno. Cada tornillo fabricado tiene una probabilidad 0.01 de ser defectuoso. Los tornillos fabricados, son defectuosos o no, independientemente de los demás tornillos. Calcular la Probabilidad de que el tornillo número 20 de los fabricados sea el quinto que sale defectuoso.

$\underbrace{\text{4D} \qquad \text{nº20 = D}}_{\text{20 torn = 5D}}$

Sea X_i la v.a. **"Tornillo fabricado por la máquina, es defectuoso"**

$$X_i \in \mathcal{B}(p = 0.01)$$

Si el tornillo número 20 es el quinto defectuoso, significa que entre los diecinueve primeros, 4 han sido defectuosos y 15 no, o dicho de otro modo, el quinto "éxito" (tornillo defectuoso) se ha obtenido en la repetición número veinte del experimento

la v.a. X que nos medirá en que repetición del experimento se obtiene el éxito número k, es:

$$X \in bn\,(k = 5, p = 0.01)$$

Luego:

$$P(X = 20) = \begin{pmatrix} 20-1 \\ 5-1 \end{pmatrix} \times 0.01^4 \times 0.99^{15} \times 0.01 = \begin{pmatrix} 19 \\ 4 \end{pmatrix} \times 0.01^5 \times 0.99^{15} = 3.33 \times 10^{-7}$$

Para el ejemplo anterior, calcular la probabilidad de que el primer tornillo defectuoso que se fabrique, sea el que hace el número 50.

Si el tornillo número 50 debe ser el primer defectuoso, significa que los 49 anteriores son buenos

la v.a. $\mathcal{X}$, que nos medirá **en que repetición del experimento aparecerá el primer "éxito"** (Tornillo defectuoso), tendrá por distribución:

$$\mathcal{X} \in \mathcal{G}\left(\mathrm{p}=0.01\right)$$

Por tanto

$$\mathrm{P}\left(\mathcal{X}=50\right)= 0.99^{49} \times 0.01=6.11\times 10^{-3}$$

ESCUELA POLITÉCNICA SUPERIOR DE CÓRDOBA
Universidad de Córdoba
DEPARTAMENTO DE ESTADÍSTICA

APROXIMACIÓNES ENTRE LAS DISTRIBUCIONES

- Aproximación de la D. Hipergeométrica a la D. Binomial

$P(X=a) \approx P(X'=a)$

$$\mathcal{X} \in h(n,a,b) \to \mathcal{X}' \in b\left(n; p = \frac{a}{N} \right)$$

- Si $a+b=N \to \infty$ o bien $a+b >> n$

En la práctica: $n \leq 0.05(a+b)$

- Aproximación de la D. Binomial a la D. Poissón

$$\mathcal{X} \in b(n,p) \to \mathcal{X}' \in \mathcal{P}(\lambda = np)$$

Si $\begin{cases} n \geq 50 \\ p \leq 0.1 \end{cases}$ o bien $\lambda < 18$ (*) / otras $\begin{cases} n \geq 20 \\ p \leq 0.5 \end{cases}$; $\lambda \leq 10$