

**Universidad Nacional de Colombia
Departamento de Estadística**

DIPLOMADO EN PROBABILIDAD E INFERNICIA BÁSICA

Módulo
PROBABILIDAD

Tercera sesión

Modelos discretos

Generalidades

Las variables aleatorias discretas tienen recorridos finitos o numerables

Variable aleatoria discreta

Sea X una variable aleatoria con recorrido es $R_X = \{x_1, x_2, \dots\}$ y función de distribución $F_X(x)$.

Supóngase que para cada valor $x_i \in R_X$, $i = 1, 2, \dots$, se definen las probabilidades

$$p_i := P(X = x_i)$$

Es claro que:

- $0 < p_i < 1$
- $p_1 + p_2 + \dots = 1$
- $p_i = F_X(x_i) - F_X(x_{i-1})$

Las parejas $[x_i; p_i]$ conforman la *distribución de masa de probabilidad* de X .

En este caso, se dice que la variable aleatoria es *discreta*.

Atributos de una VAD

El *valor esperado* de una VAD X es una medida de localización y se define como

$$\begin{aligned}\mu = E(X) &= \sum_{i=1}^{\infty} x_i p_i \\ &= x_1 p_1 + \cdots + x_n p_n + \cdots\end{aligned}$$

La *varianza* de una VAD X es una medida de dispersión y se define como

$$\sigma^2 = V(X) = E(X^2) - \mu^2$$

donde

$$E(X^2) = \sum_{i=1}^{\infty} x_i^2 p_i$$

Tanto $E(X)$ como $V(X)$ son constantes y tienen unidades de medida. Usualmente, se toma como medida de dispersión la desviación estándar de la VAD dada por $\sigma = \sqrt{V(X)}$.

Modelos de variables aleatorias discretas

Se revisan los modelos de variables aleatorias más usados en la práctica cotidiana

Distribución uniforme discreta

Sea X una VAD con recorrido $R_X = \{x_1, \dots, x_n\}$. La variable X tiene distribución *uniforme* discreta, *laplaciana* o *clásica*, si y sólo si su distribución de masa de probabilidad está dada por

$$p_i = P(X = x_i) = p$$

con $i = 1, 2, \dots, n$.

EJEMPLO. Sea X : “la cantidad de puntos en la cara superior de un dado regular”. El rango de la variable es $R_X = \{1, 2, 3, 4, 5, 6\}$ y su distribución está dada por

X	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X) = 3,5 \text{ y } V(X) \cong 2,92$$

Distribución Bernoulli

La VAD X tiene distribución de *Bernoulli* con *probabilidad de éxito* $0 < p < 1$, si y sólo si su distribución de masa de probabilidad está dada por

$$P(X = x) = p^x(1 - p)^{1-x}$$

con $x = 0$ ó 1 .

Simbólicamente, se escribe $X \sim \mathcal{B}(p)$.

X	0	1
$P(X = x)$	$1 - p$	p

$$E(X) = p \text{ y } V(X) = p(1 - p)$$

EJEMPLO. Sea X : “se obtiene CARA cuando se lanza una moneda regular una vez”. Es claro que el evento ($X = 1$) consiste en obtener CARA; luego, $X \sim \mathcal{B}(p = 0,5)$.

Distribución binomial

La VAD X tiene distribución *binomial* con parámetros n y p , simbólicamente $X \sim \mathcal{B}(n; p)$, si y sólo si su distribución de masa de probabilidad está dada por

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

con $x = 0, 1, 2, \dots, n$.

Los atributos numéricos de X son

$$E(X) = np \text{ y } V(X) = np(1 - p)$$

Es claro que

$$X := \sum_{i=1}^n Y_i$$

donde $Y_i \sim \mathcal{B}(p)$, $i = 1, \dots, n$, son variables aleatorias independientes.

Es decir, X representa la cantidad de éxitos en n repeticiones independientes de un mismo ensayo de Bernoulli.

Ejemplo

En una pequeña ciudad, el 53% de los nacimientos son niñas. Una pareja de recién casados planifica tres nacimientos. Sea X la variable aleatoria que representa la “*cantidad de niñas en los tres nacimientos que planifica la pareja*”. Es claro que $X \sim \text{Bin}(n = 3; p = 0,53)$; luego,

$$P(X = x) = \binom{3}{x} 0,53^x 0,47^{3-x}$$

con $x = 0, 1, 2$ ó 3 .

La distribución de probabilidades tiene la forma

X	$P(X = x)$
0	0,1038
1	0,3512
2	0,3961
3	0,1489

La probabilidad de que la pareja tenga por lo menos una niña es

$$P(X \geq 1) = 1 - P(X = 0) = 0,8962$$

Continuación del ejemplo

La cantidad esperada de niñas en tres nacimientos es

$$E(X) = 3 \times 0,53 = 1,59 \text{ niñas}$$

La pareja “espera” tener entre 1 y 2 niñas. (Entre muchas parejas que planifican tres nacimientos la cantidad más frecuente de niñas es de 1 ó 2, pero más 2 que 1. Equivalentemente, a largo plazo la cantidad media de niñas está entre 1 y 2).

La varianza de la cantidad de niñas en tres nacimientos es

$$V(X) = 3 \times 0,53 \times 0,47 = 0,7473$$

La desviación estándar es igual a $\sigma \cong 0,86$ niñas. Luego, la pareja tendrá aproximadamente entre 1 y 2 niñas (0,73; 2,45). Este cálculo se obtiene de construir el intervalo de valores que se encuentran a una desviación estándar de la media.

Distribución binomial negativa

La VAD X tiene distribución **binomial negativa** con parámetros r y p , simbólicamente $X \sim \text{bn}(r; p)$, si y sólo si su distribución de masa de probabilidad está dada por

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

con $x = r, r+1, \dots$

La distribución binomial negativa se obtiene de fijar la cantidad de éxitos r a obtener en una cantidad variable X de repeticiones independientes de un mismo ensayo de Bernoulli con probabilidad de éxito igual a p . Se tiene que

$$E(X) = \frac{r}{p} \quad \text{y} \quad V(X) = \frac{r(1-p)}{p^2}$$

Ejemplo

Supóngase que la pareja desea tener su segunda hija en el tercer nacimiento. Sea X la variable aleatoria que representa la “*cantidad de nacimientos necesarios para tener dos niñas*”. En este caso se tiene que $X \sim \text{Bin}(r = 2; p = 0,53)$. Luego,

$$P(X = x) = (x - 1)0,53^2 0,47^{x-2}$$

con $x = 2,3,\dots$

(a) La probabilidad de que la pareja tenga su segunda hija en el tercer nacimiento es igual a

$$P(X = 3) = 2 \times 0,53^2 \times 0,47 = 0,264$$

(b) La probabilidad de que la pareja tenga las dos hijas que desea en el menor tiempo posible es

$$P(X = 2) = 0,53^2 = 0,2809$$

Continuación del ejemplo

La cantidad esperada de nacimientos para que nazca la segunda niña es

$$E(X) = \frac{2}{0,53} \cong 3,77$$

Es decir, la segunda niña nacerá, en términos medios, entre el tercer y cuarto nacimientos.

La varianza de la cantidad de nacimientos necesarios para que nazca la segunda niña es

$$V(X) = \frac{2 \times 0,47}{0,53^2} \cong 3,346 \text{ nac}^2$$

La desviación estándar es $\sigma \cong 1,83$ nac. Luego, la segunda niña nacerá aproximadamente entre el segundo y sexto nacimientos (cantidad de nacimientos que caen en el intervalo $3,77 \pm 1,83$).

Distribución geométrica

Sea $X \sim \text{bn}(r; p)$. Si se hace $r = 1$, se obtiene la variable con distribución *geométrica* de probabilidades

$$P(X = x) = (1 - p)^{x-1} p$$

con $x = 1, 2, \dots$

Simbólicamente, se escribe $X \sim G(p)$.

Hacer $r = 1$ en la definición de la distribución binomial negativa implica saber en cuál intento se puede obtener el primer éxito.

Los atributos numéricos de la distribución geométrica son

$$E(X) = \frac{1}{p} \quad \text{y} \quad V(X) = \frac{1-p}{p^2}$$

Ejemplo

Supóngase ahora que la pareja desea saber cuando más posiblemente tiene su primera hija. Sea X la variable aleatoria que representa la “*cantidad de nacimientos necesarios para tener la primera niña*”. En este caso, se tiene que $X \sim G(p = 0,53)$. Luego,

$$P(X = x) = 0,53 \times 0,47^{x-1}$$

con $x = 1, 2, \dots$

La distribución de masa de probabilidad de X es entonces

X	$P(X = x)$
1	0,5300
2	0,2491
3	0,1171
4	0,0550
5	0,0259
6	0,0122
7	0,0057
8	0,0027

Continuación del ejemplo

La cantidad esperada de nacimientos hasta la primera niña es igual a

$$E(X) = \frac{1}{0,53} \cong 1,89 \text{ nac}$$

A largo plazo, parejas con el mismo deseo tendrán su primera niña más probablemente entre el primer y segundo nacimiento.

La varianza de la cantidad de nacimientos hasta la primera niña es

$$V(X) = \frac{0,47}{0,53^2} \cong 1,673 \text{ nac}^2$$

La desviación estándar es $\sigma \cong 1,29 \text{ nac}$. Luego, la primera niña nacerá aproximadamente entre el primer y tercer nacimientos (cantidad de nacimientos que caen en el intervalo $1,89 \pm 1,29$).

Distribución hipergeométrica

Se tiene N objetos de los cuales K ($K \leq N$) poseen una característica de interés. Sea X : “*la cantidad de objetos en una muestra de tamaño n , tomada sin reemplazamiento, con la característica de interés*”. Se dice que X tiene distribución hipergeométrica con distribución de masa de probabilidad

$$P(X = x) = \frac{\binom{K}{x} \binom{N - K}{n - x}}{\binom{N}{n}}$$

con $x = 0, 1, 2, \dots, \min(n, K)$.

Simbólicamente, se escribe $X \sim \mathcal{H}(N, K, n)$.

El valor esperado de la variable X es

$$E(X) = n \frac{K}{N}$$

y la varianza de X es

$$V(X) = n \frac{K}{N} \left(1 - \frac{K}{N}\right) \left(\frac{N - n}{N - 1}\right)$$

TEOREMA. *Aproximación de hipergeométrica por binomial.* Sea $X \sim \mathcal{H}(N, K, n)$. Si $\frac{K}{N} \rightarrow p$, entonces X sigue aproximadamente una distribución binomial con parámetros n y p .

Ejemplo

Se capturaron y etiquetaron 5 ejemplares de un tipo de animal en vía de extinción, para luego liberarlos y que se mezclaran con la población existente.

Después de un tiempo, se toma una muestra de 10. Si se sabe que la población total consta de 25 animales, se pide hallar la probabilidad de que exactamente 2 de los ejemplares marcados estén en la muestra.

Sea X : “la cantidad de animales etiquetados en la muestra de 10”. Es claro que

$X \sim \mathcal{H}(N = 25, K = 5, n = 10)$; luego,

$$P(X = x) = \frac{\binom{5}{x} \binom{20}{10-x}}{\binom{25}{10}}$$

con $x = 0, 1, 2, 3, 4, 5$.

Se pide: $P(X = 2) = 0,385$

Continuación del ejemplo

La cantidad esperada de animales etiquetados en la muestra seleccionada de 10 ejemplares es

$$E(X) = 10 \times \frac{5}{25} = 2$$

A largo plazo, en cada muestra de 10 animales que se seleccionen de un total de 25 habrá, en promedio, dos ejemplares etiquetados.

La varianza de la cantidad de animales etiquetados en la muestra de 10 es

$$V(X) = \frac{10 \times 5}{25} \left(1 - \frac{5}{25}\right) \left(\frac{25 - 10}{25 - 1}\right) = 1$$

La desviación estándar es $\sigma = 1$ y la cantidad aproximada de animales etiquetados en la muestra de tamaño 10 está entre 1 y 3 (valores que se encuentran en el intervalo $2,00 \pm 1,00$).

Ejemplo de la aproximación

En una ciudad de dos millones de habitantes, el 60% pertenece a cierto partido político. Se eligen al azar 10 personas. Se quiere calcular la probabilidad de que por lo menos la mitad de los seleccionados en la muestra pertenezcan al partido.

Sea X la variable que denota la cantidad de militantes en una muestra aleatoria de 10 ciudadanos. Es claro que X sigue una distribución hipergeométrica con parámetros $N, K \rightarrow \infty$; luego, la probabilidad requerida se puede “aproximar” utilizando la distribución binomial con parámetros $n = 10$ y $p = 0.6$ de la siguiente manera

$$P(X \geq 5) \cong \sum_{x=5}^{10} \binom{10}{x} 0,6^x 0,4^{10-x} = 0,8338$$

Distribución Poisson

La variable aleatoria X tiene distribución **Poisson** con parámetro o rata de ocurrencias $\lambda > 0$, si y sólo si su distribución de masa de probabilidad es

$$P(X_t = x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}$$

con $x = 0, 1, 2, \dots$ y t es un soporte continuo donde tienen lugar las ocurrencias de interés.

Simbólicamente, se escribe $X_t \sim \mathcal{P}(\lambda t)$.

Es claro que para cualquier variable con distribución Poisson se tiene que

$$P(X_t = 0) = e^{-\lambda t}$$

Además,

$$E(X_t) = V(X_t) = \lambda t$$

TEOREMA. *Aproximación de binomial por Poisson.* Sea $X \sim \mathcal{B}(n; p)$. Si $n \rightarrow \infty$ y $p \rightarrow 0$, entonces $X \sim \mathcal{P}(\lambda t = np)$.

Ejemplo

Una fábrica asegura que ocho de los componentes que produce fallan antes de cumplir 100 horas de funcionamiento ($\lambda = 0,08$). Si la cantidad de fallas X_t es una variable aleatoria tal que $X_t \sim \mathcal{P}(0,08t)$, entonces:

- (a) La probabilidad de que un componente falle antes de 25 horas ($t = 25$) es

$$P(X_{25} = 1) = \frac{2^1}{1!} e^{-2} = 0,2707$$

Continuación del ejemplo

(b) Así mismo, la probabilidad de que no más de dos componentes fallen antes de 50 horas ($t = 50$) es

$$P(X_{50} \leq 2) = \sum_{x=0}^2 \frac{4^x}{x!} e^{-4} = 0,2381$$

(c) La probabilidad de que fallen por lo menos diez componentes en 125 horas es

$$P(X_{125} \geq 10) = 1 - \sum_{x=0}^9 \frac{10^x}{x!} e^{-10} \cong 0,5421$$

Ejemplo de la aproximación

Supóngase que sólo el 0,1% de todas las computadoras de cierto tipo experimentan fallas del CPU durante el período de garantía. Considérese una muestra de 10000 computadores.

Sea X : “la cantidad de computadoras que experimentan fallas en la muestra disponible”. Bajo ciertas condiciones de regularidad, se podría asumir que $X \sim \mathcal{B}(n = 10000; p = 0,001)$. Se quiere hallar la probabilidad de que por lo menos 10 de computadoras tengan el defecto.

Se busca

$$P(X \geq 10) = \sum_{x=10}^{10000} \binom{10000}{x} 0,001^x 0,999^{10000-x}$$

En este caso, es claro que $n \rightarrow \infty$ y $p \rightarrow 0$ y se puede evaluar $P(X \geq 10)$ usando la aproximación $X \sim \mathcal{P}(\lambda t = np = 10)$ así

$$\begin{aligned} P(X \geq 10) &= 1 - P(X < 10) = \\ &= 1 - \sum_{x=0}^9 \frac{10^x}{x!} e^{-10} = 0,542 \end{aligned}$$

G R A C I A S

Diagramado en

