

**Universidad Nacional de Rosario**

FACULTAD DE CIENCIAS EXACTAS, INGENIERÍA Y AGRIMENSURA

# PROBABILIDAD Y ESTADÍSTICA

## *Unidad 4*

Autor del resumen:

Charles Chaplin

Julio 2020

# Contents

<b>1</b>	<b>Variables aleatorias unidimensionales</b>	<b>2</b>
1.1	Noción general de una variable aleatoria . . . . .	2
1.2	Variables aleatorias discretas . . . . .	4
1.3	Características de las variables aleatorias . . . . .	5
1.3.1	El valor esperado de una variable aleatoria . . . . .	5
1.3.2	Propiedades del valor esperado . . . . .	5
1.3.3	La varianza de una variable aleatoria . . . . .	6
1.3.4	Propiedades de la varianza de una variable aleatoria . . . . .	7
1.4	La distribución binomial . . . . .	8
1.5	La distribución de Poisson . . . . .	9
1.5.1	La distribución de Poisson como aproximación de la distribución Binomial. . . . .	9
1.6	La distribución Geométrica . . . . .	10
1.7	La distribución Pascal . . . . .	11
1.8	Relación entre las distribuciones Binomial y Pascal . . . . .	12
1.9	La distribución Hipergeométrica . . . . .	13
1.10	La distribución Multinomial . . . . .	14
1.11	El proceso de Poisson . . . . .	15

# 1 Variables aleatorias unidimensionales

## 1.1 Noción general de una variable aleatoria

En muchas situaciones experimentales deseamos asignar un número real  $x$  a cada uno de los elementos  $s$  del espacio muestral  $S$ . Esto es:

$$X(s) : S \rightarrow \mathbb{R}$$

**Definición.** Sea  $\epsilon$  un experimento y  $S$  el espacio muestral asociado con él. Una función  $X$  que asgna a cada uno de los elementos  $s \in S$ , un numero real  $X(s)$ , se llama **variable aleatoria**.

El espacio  $R_x$ , es decir, el conjunto de todos los valores posibles de  $X$ , algunas veces se le llama recorrido. En cierto sentido podemos considerar a  $R_x$  como otro espacio muestral. El espacio muestral (original)  $S$  corresponde a los resultados no numéricos (posiblemente) del experimento, mientras que  $R_x$ , es el espacio muestral asociado con la variable aleatoria  $X$ , que representa la característica numérica que puede ser de interés. Si  $X(s) = S$ , tenemos  $S = R_x$ .

Una variable aleatoria  $X$  puede ser concebida de dos formas:

- Realizamos el experimento  $\epsilon$  que tiene como resultado  $s \in S$ . Luego evaluamos el número  $X(s)$ .
- Efectuamos  $\epsilon$ , obteniendo el resultado  $s$ , e (inmediatamente) calculamos  $X(s)$ . El número  $X(s)$  se considera entonces como el resultado obtenido en el experimento y  $R_x$  se convierte en el espacio muestral del experimento.

En el primer caso, el experimento termina, de hecho, con la observación de  $s$ . La evaluación de  $X(s)$  se estima como algo que se hace posteriormente y que no se afecta por la aleatoriedad de  $\epsilon$ . En el segundo caso, se considiera que el experimento no está terminado hasta que el número  $X(s)$  se ha calculado y se tiene así el espacio muestral  $R_x$ , como resultado. Al estudiar variables aleatorias estamos más interesados respecto a los valores que toma  $X$  que a su forma funcional. Por lo tanto, en muchos casos ignoraremos por completo el espacio muestral sobre el cual se puede definir  $X$ .

### Ejemplos págs 71-72

En general nos referiremos a las variables aleatorias con letras mayúsculas ( $X, Y, Z$ ) y a sus valores con letras minúsculas ( $x, y, z$ ).

**Definición.** Sea  $\epsilon$  un experimento y  $S$  su espacio muestral. Sea  $X$  una variable aleatoria definida en  $S$  y sea  $R_x$  su recorrido. Sea  $B$  un evento respecto a  $R_x$ ; esto es,  $B \subset R_x$  y sea  $A$  un evento respecto a  $S$  definido como:

$$A = \{s \in S | X(s) \in B\}$$

En palabras,  $A$  consta de todos los resultados en  $S$  para los cuales  $X(s) \in B$ . En este caso decimos que  $A$  y  $B$  son **eventos equivalentes**.

De manera más informal,  $A$  y  $B$  son eventos equivalentes siempre que ocurran juntos. Es importante destacar que en nuestra definición de eventos equivalentes,  $A$  y  $B$  están asociados con espacios muestrales diferentes.

#### Ejemplo pág. 74

**Definición.** Sea  $B$  un evento en el recorrido  $R_x$ , entonces definimos  $P(B)$  como:

$$P(B) = P(A), \text{ donde } A = \{s \in S | X(s) \in B\}$$

En palabras, definimos  $P(B)$  igual a la probabilidad del evento  $A \subset S$ . Por tanto, la definición anterior hace posible asignar probabilidades a eventos asociados con  $R_x$  en términos de las probabilidades definidas en  $S$ .

#### Ejemplo pág. 75

### Tips para entender mejor

Puesto que en la formulación de la ecuación los eventos  $A$  y  $B$  se refieren a espacios muestrales diferentes, en realidad deberíamos usar una notación diferente cuando nos referimos a las probabilidades definidas en  $S$  y para las definidas en  $R_x$ , por ejemplo, algo como  $P(A)$  y  $P_x(B)$ . Sin embargo, no haremos esto sino que continuaremos escribiendo simplemente  $P(A)$  y  $P(B)$ . El contexto dentro del cual aparezcan estas expresiones hará evidente la interpretación.

Las probabilidades asociadas con eventos en el espacio muestral  $S$  (original) están en un sentido, determinadas por “fuerzas que escapan a nuestro control” o, como se dice algunas veces “por naturaleza”. Cuando introducimos una variable aleatoria  $X$  y su recorrido asociado  $R_x$ , **inducimos** probabilidades sobre los eventos asociados con  $R_x$  que se determinan estrictamente si las posibilidades asociadas con los eventos en  $S$  están especificadas.

## 1.2 Variables aleatorias discretas

Ojo... hay que entender los conceptos, los nombres no son intuitivos.

**Definición.** Sea  $X$  una variable aleatoria. Si el número de valores posibles de  $X$  (esto es,  $R_x$ , el recorrido) es finito o infinito numerable, llamamos a  $X$  una variable aleatoria discreta.

**Ejemplo pág. 76**

**Definición** Sea  $X$  una variable aleatoria discreta. Con cada resultado posible  $x_i$  asociamos un número  $p(x_i) = P(X = x_i)$ , llamado probabilidad de  $x_i$ . Los números  $p(x_i)$ ,  $i = 1, 2, \dots$  deben satisfacer las condiciones siguientes:

- $p(x_i) > 0 \forall_i$
- $\sum_i^\infty p(x_i) = 1$ .

La función  $p$  que antes se definió, se llama **función de probabilidad** (o función de probabilidad puntual) de la variable aleatoria  $X$ . La colección de pares  $(x_i, p(x_i))$ ,  $i = 1, 2, \dots$ , se la llama **distribución de probabilidad de  $X$** .

Sea  $B$  un evento asociado con la variable aleatoria  $X$ . Esto es,  $B \subset R_x$ . Específicamente, supongamos que  $B = \{x_{i_1}, x_{i_2}, \dots\}$ . Por lo tanto,

$$P(B) = P[s \in S | X(s) \in B] = P[s \in S | X(s) = x_{i_j}, j = 1, 2, \dots] = \sum_j^\infty p(x_{i_j})$$

En palabras, la probabilidad de un evento  $B$  es igual a la suma de las probabilidades de los resultados individuales asociados con  $B$ .

### Observaciones.

Si  $X$  toma un número infinito numerable de valores, entonces es imposible tener todos los resultados igualmente probables, porque quizá no podamos satisfacer la condición si hemos de tener  $p(x_i) = c$  para toda  $i$ .

En cada intervalo finito habrá cuando mucho un número finito de valores posibles de  $X$ . Si uno de esos intervalos no contiene ninguno de los valores posibles, le asignamos probabilidad cero. Esto es, si  $R_x = \{x_1, x_2, \dots, x_n\}$  y si ningún  $x_i \in [a, b]$  entonces  $P[a \leq X \leq b] = 0$ .

**Ejemplo pág. 78-79**

## 1.3 Características de las variables aleatorias

### 1.3.1 El valor esperado de una variable aleatoria

En los modelos matemáticos no deterministas o aleatorios que hemos considerado, los parámetros pueden usarse para señalar la distribución de probabilidades. Con cada distribución de probabilidades podemos asociar ciertos parámetros que dan información valiosa acerca de la distribución (tal como la pendiente de una recta proporciona una información útil acerca de la relación lineal que representa).

#### Ejemplo 7.3 pág. 155

**Definición.** Sea  $X$  una variable aleatoria discreta con valores posibles  $x_1, x_2, \dots, x_n, \dots$  y sea  $p(x_i) = P(X = x_i), i = 1, 2, \dots, n, \dots$ . Entonces el **valor esperado** de  $X$  (o esperanza matemática de  $X$ ), que se denota con  $E(X)$ , se define como

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i),$$

si la serie converge absolutamente, es decir, si  $\sum_{i=1}^{\infty} |x_i| p(x_i) < \infty$ . Este número se designa como **valor promedio** de  $X$ .

Cabe destacar que  $E(X)$  es un número (parámetro) asociado con una distribución de probabilidades teórica (recomendado leer observaciones pág 156).

#### Ejemplo 7.4 pág. 157

### 1.3.2 Propiedades del valor esperado

#### Propiedades

- Si  $X = C$ , donde  $C$  es una constante, entonces  $E(X) = C$ .
- Supongamos que  $C$  es una constante y  $X$  es una variable aleatoria. Entonces,  $E(CX) = CE(X)$ .
- Sea  $(X, Y)$  una variable aleatoria bidimensional con una distribución de probabilidades conjunta. Sean  $Z = H_1(X, Y)$  y  $T.V = H_2(X, Y)$ . Entonces,  $E(Z + W) = E(Z) + E(W)$ . PARA UNIDAD 6.
- Sean  $X$  y  $Y$  dos variables aleatorias cualesquiera. Entonces,  $E(X + Y) = E(X) + E(Y)$ .
- Sean  $X_1, X_2, \dots, X_n$   $n$  variables aleatorias.

Entonces,  $E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$

- Sea  $(X, Y)$  una variable aleatoria bidimensional y supongamos que  $X$  e  $Y$  son independientes. Entonces,  $E(XY) = E(X)E(Y)$ . PARA UNIDAD 6.

Todas las demostraciones se encuentran entre las págs 169 y 171.

Supongamos que para una variable aleatoria  $X$  encontramos que  $E(X)$  es igual a 2. Es preciso que no se atribuya más importancia a esta información que la justificada. Significa sencillamente, que si consideramos un gran número de valores de  $X$ , digamos  $x_1, x_2, \dots, x_n$ , y los promediamos, este resultado estará cercano a 2 si  $n$  es grande

### 1.3.3 La varianza de una variable aleatoria

Supóngase que  $X$  representa la duración de una bombilla que se recibe de un fabricante, y que  $E(X) = 1000$  horas. Esto podría significar una de varias posibilidades. Podría significar que se espera que la mayor parte de las bombillas dure entre 900 y 1100 horas. También podría significar que las bombillas que se entregan son de dos tipos diferentes: alrededor de la mitad son de muy alta calidad y con duración de casi 1300 horas, mientras que la otra mitad son de muy mala calidad y tienen una duración de cerca de 700 horas.

Hay una necesidad obvia de presentar una medida cuantitativa que distinga entre estas situaciones.

**Definición** Sea  $X$  una variable aleatoria. Definamos la **varianza** de  $X$ , que se denota con  $V(X)$  o  $\sigma_X^2$  como sigue:

$$V(X) = E[X - E(X)]^2$$

La raíz cuadrada positiva de  $V(X)$  se llama **desviación estándar** de  $X$  y se designa con  $\sigma_X$ .

El cálculo de  $V(X)$  se simplifica con la ayuda del resultado siguiente:

**Teorema.**  $V(X) = E(X^2) - [E(X)]^2$ .

**Demostración.**

$$V(X) = E[X - E(X)]^2$$

$\Longleftrightarrow$

$$E\{X^2 - 2XE(X) + E(X)^2\}$$

$\Longleftrightarrow$

$$E(X^2) - E(2XE(X)) + E(E(X)^2)$$

$\Longleftrightarrow$

$$E(X^2) - 2E(X)E(X) + E(X)^2$$

$\Longleftrightarrow$

$$E(X^2) - E(X)^2$$

**Ejemplos págs 177-179**

### 1.3.4 Propiedades de la varianza de una variable aleatoria

#### Propiedades

- Si  $C$  es una constante,  $V(X + C) = V(X)$
- Si  $C$  es una constante,  $V(XC) = C^2V(X)$
- Sea  $(X, Y)$  es una variable aleatoria bidimensional, y si  $X$  y  $Y$  son independientes.

Entonces  $V(X + Y) = V(X) + V(Y)$ . PARA UNIDAD 6

- Sean  $X_1, X_2, \dots, X_n$ ,  $n$  variables aleatorias independientes.

Entonces,  $V(X_1 + X_2 + \dots + X_n) = V(X_1) + \dots + V(X_n)$

#### Demostraciones y ejemplos págs 179-182

Tal como en los modelos deterministas, en los cuales ciertas relaciones funcionales desempeñan un papel importante (tales como lineales, cuadráticas, exponenciales, trigonométricas, etc..), al elaborar modelos no deterministas para fenómenos observables, también encontramos que ciertas distribuciones de probabilidades aparecen más a menudo que otras. Una razón de esto es que, como en el caso determinista, algunos modelos matemáticos relativamente simples parecen ser capaces de describir un gran número de fenómenos.

Es por esto que estudiaremos alguna de ellas en los siguientes apartados ...



## 1.4 La distribución binomial

### Ejemplo ilustrador pág 80

**Definición.** Consideremos un experimento  $\epsilon$  y sea  $A$  un evento asociado con  $\epsilon$ . Supongamos que  $P(A) = p$  y, por lo tanto,  $P(A^C) = 1 - p$ . Consideremos  $n$  repeticiones independientes de  $\epsilon$ . Por lo tanto, el espacio muestral consiste en todas las sucesiones posibles  $\{a_1, a_2, \dots, a_n\}$ , donde cada  $a_i$  es  $A$  o  $A^C$ , según  $A$  o  $A^C$  ocurra en la  $i$ -ésima repetición de  $\epsilon$ . (Hay  $2^n$  de tales sucesiones). Aún más, supongamos que  $P(A) = p$  es el mismo para todas las repeticiones. Definamos la variable aleatoria  $X$  como sigue:  $X$  = número de veces que ocurrió el evento  $A$ . Llamamos a  $X$  una variable aleatoria binomial con los parámetros  $n$  y  $p$ . Sus valores posibles obviamente son  $0, 1, 2, \dots, n$ . (Decimos en forma equivalente que  $X$  tiene una distribución binomial- $X \sim Bi(n, p)$ .) Las repeticiones individuales de  $\epsilon$  se llaman ensayos de Bernoulli.

**Teorema.** Sea  $X$  una variable binomial con base en  $n$  repeticiones. Entonces

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{(n-k)}, \quad k = 0, 1, \dots, n.$$

### Demostración pág 82 -recomendable de leer-

**Teorema.** Sea  $X$  una variable aleatoria distribuida binomialmente con parámetro  $p$ , con base en  $n$  repeticiones de un experimento. Entonces

$$E(x) = np$$

### Demostración pág. 157

### Análisis de varianza pág. 181

**Varianza:**  $V(X) = np(1-p)$

## 1.5 La distribución de Poisson

**Definición.** Sea  $X$  una variable aleatoria que toma los valores posibles:  $0, 1, \dots, n, \dots$ . Si

$$P(X = k) = \frac{e^{-\alpha} \alpha^k}{k!}, \quad k = 0, 1, \dots, n, \dots,$$

decimos que  $X$  tiene una **distribución** de Poisson con parámetro  $\alpha > 0$

**Teorema** Si  $X$  tiene una distribución de Poisson con parámetro  $\alpha$ , entonces  $E(X) = \alpha$  y  $V(X) = \alpha$   
**Demostración** pág. 210

### 1.5.1 La distribución de Poisson como aproximación de la distribución Binomial.

#### Análisis detallado págs 211-213

**Teorema.** Sea  $X$  una variable aleatoria distribuida binomialmente con parámetro  $p$  (con base en  $n$  repeticiones del experimento). Esto es,

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

Supóngase que cuando  $n \rightarrow \infty, np = \alpha$ , o equivalentemente, cuando  $n \rightarrow \infty, p \rightarrow 0, np \rightarrow \alpha$ . En estas condiciones tenemos

$$\lim_{n \rightarrow \infty} P(X = k) = \frac{e^{-\alpha} \alpha^k}{k!}$$

la distribución de Poisson con parámetro  $\alpha$

#### Observaciones.

- El teorema anterior esencialmente dice que podemos aproximar las probabilidades binomiales con las probabilidades de la distribución de Poisson siempre que  $n$  sea grande y  $p$  pequeña.
- La distribución binomial se caracteriza por dos parámetros,  $n$  y  $p$ , mientras que la distribución de Poisson se caracteriza por un solo parámetro,  $\alpha = np$ , que **representa al número esperado de éxitos por unidad de tiempo (o por unidad de espacio en algún otro caso)**. Este parámetro también se designa como **intensidad de la distribución**. Es importante distinguir entre el número esperado de ocurrencias por unidad de tiempo y el número esperado de ocurrencias en el tiempo especificado.

#### Ejemplos págs 214-217

## 1.6 La distribución Geométrica

Supóngase que efectuamos un experimento  $\epsilon$  y que estamos interesados sólo en la ocurrencia o no ocurrencia de algún evento  $A$ . Supóngase, como en la presentación de la distribución binomial, que repetidamente efectuamos  $\epsilon$ , que las repeticiones son independientes y que en cada una de ellas  $P(A) = p$  y  $P(A^c) = 1 - p = q$  permanecen constantes. Supóngase que repetimos el experimento hasta que  $A$  ocurre por primera vez. (Aquí nos apartamos de las hipótesis que conducen a la distribución binomial. Allí el número de repeticiones era predeterminado, mientras que aquí es una variable aleatoria.)

Definamos la variable aleatoria  $X$  como el número de repeticiones necesarias hasta incluir la primera ocurrencia de  $A$ . Así  $X$  toma los valores  $1, 2, \dots$ ; puesto que  $X=k$  si y sólo si las primeras  $(k - 1)$  repeticiones de  $\epsilon$  producen  $A^c$ , mientras que la  $k$ -ésima da por resultado  $A$ , tenemos

$$P(X = k) = q^{k-1}p, \quad k = 1, 2, \dots$$

Se dice que una variable aleatoria con una distribución de probabilidades de esta forma tiene una **distribución geométrica**.

**Teorema** Si  $X$  tiene una distribución geométrica, entonces

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

El hecho de que  $E(X)$  sea el recíproco de  $p$  es interesante intuitivamente, puesto que dice que con valores pequeños de  $p = P(A)$  se necesitan muchas repeticiones para que ocurra  $A$  (valor  $E(X)$  muy grande).

**Análisis** pág 225

**Ejemplos** págs 225-226

**Teorema.** Supóngase que  $X$  tiene una distribución geométrica. Entonces para dos enteros positivos cualesquiera  $s$  y  $t$ ,

$$P(X \geq s + t | X > s) = P(X \geq t)$$

El teorema anterior indica que la distribución geométrica “no tiene memoria” en el sentido siguiente. Supongamos que el evento  $A$  no ha ocurrido durante las primeras repeticiones de  $\epsilon$ . Entonces, la probabilidad de que no ocurra durante las próximas  $t$  repeticiones es la misma que la probabilidad de que no ocurra durante las primeras  $t$  repeticiones. En otras palabras, la información de ningún éxito es “olvidada” en lo que se refiere a cálculos subsecuentes.

**Ejemplo** pág 227

## 1.7 La distribución Pascal

Si planteamos la siguiente cuestión, surge una generalización obvia de la distribución geométrica. Supongamos que un experimento se continúa hasta que un evento particular A ocurre por r-ésima vez. Si

$$P(A) = p, P(A^c) = q = 1-p$$

en cada una de las repeticiones, definimos la variable aleatoria Y como sigue.

Y es el número de repeticiones necesarias para que A ocurra exactamente r veces. Necesitamos la distribución de probabilidades de Y. Debe ser evidente que si  $r = 1$ , Y tiene una distribución geométrica.

Ahora  $Y = k$  si y sólo si A ocurre en la k-ésima repetición y precisamente A ocurrió (r - 1) veces en las (k - 1) repeticiones previas. La probabilidad de este evento es simplemente  $p \binom{k-1}{r-1} p^{r-1} q^{k-r}$ , puesto que lo que sucede en las primeras (k - 1) repeticiones es independiente de lo que sucede en la k-ésima repetición. Luego

$$P(Y = k) = \binom{k-1}{r-1} p^r q^{k-r}$$

Una variable aleatoria que tenga la distribución de probabilidades descrita anteriormente se la conoce como que tiene una **distribución Pascal**.

**Teorema.** Si Y tiene una distribución de Pascal, entonces

$$E(Y) = \frac{r}{p} \text{ y } V(Y) = \frac{rq}{p^2}$$

De esta demostración podemos observar que si X es una distribución geométrica, y sea  $Y = X$  ocurre r veces, entonces Y es una distribución pascal

**Ejemplo pág 229**

## 1.8 Relación entre las distribuciones Binomial y Pascal

Sea  $X$  una distribución binomial con parámetros  $n$  y  $p$  (es decir,  $X$  = número de éxitos en  $n$  ensayos de Bernoulli con  $P(\text{éxito}) = p$ ). Sea  $Y$  una distribución de Pascal con parámetros  $r$  y  $p$  (es decir,  $Y$  = número de ensayos de Bernoulli necesarios para obtener  $r$  éxitos con  $P(\text{éxito}) = p$ ). Por tanto, se establece la siguiente relación:

$$\text{a) } P(Y \leq n) = P(X \geq r)$$

$$\text{b) } P(Y > n) = P(X < r)$$

1. Si hay  $r$  o más éxitos en los primeros  $n$  ensayos, entonces es necesario  $n$  o menos ensayos para obtener los primeros  $r$  éxitos.
2. Si hay menos que  $r$  éxitos en los primeros  $n$  ensayos, entonces se necesitan más de  $n$  ensayos para obtener  $r$  éxitos.

**Observación.** las propiedades anteriores hacen posible el empleo de la distribución binomial tabulada para evaluar probabilidades asociadas con la distribución de Pascal

**Ejemplo observación a) pág 230**

## 1.9 La distribución Hipergeométrica

Supóngase que tenemos un lote de  $N$  artículos, de los cuales  $r$  son defectuosos y  $(N - r)$  son no defectuosos. Supóngase que escogemos al azar  $n$  artículos del lote ( $n \leq N$ ), sin reposición. Sea  $X$  el número de artículos defectuosos encontrados. Puesto que  $X = k$  si y sólo si obtenemos exactamente  $k$  artículos defectuosos (de los  $r$  defectuosos del lote) y exactamente  $(n - k)$  no defectuosos [de los  $(N - r)$  no defectuosos del lote], tenemos

$$P(X = k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \dots$$

Se dice que una variable aleatoria discreta que tiene esta distribución de probabilidades tiene una **distribución hipergeométrica**.

**Ejemplo pág. 231**

**Teorema.** Sean  $X$  una distribución hipergeométrica y  $p = \frac{r}{N}$ ,  $q = 1 - p$ . Entonces tenemos

- a)  $E(X) = np$ ;
- b)  $V(X) = npq \frac{N-n}{N-1}$ ;
- c)  $P(X = k) \simeq \binom{n}{k} p^k (1 - p)^{n-k}$

**Observación** La propiedad c) del teorema establece que si el tamaño  $N$  del lote es suficientemente grande, la distribución de  $X$  puede ser aproximada por la distribución binomial. Esto es razonablemente intuitivo. La distribución binomial se aplica cuando muestreamos con sustitución (puesto que en ese caso la probabilidad de obtener un artículo defectuoso permanece constante), mientras que la distribución hipergeométrica se aplica cuando muestreamos sin sustitución. Si el tamaño del lote es grande, no hay gran diferencia si un artículo en particular se devuelve o no al lote antes de escoger el siguiente. La propiedades simplemente una expresión matemática de ese hecho. Nótese también que el valor esperado de una variable aleatoria hipergeométrica,  $Y$ , es el mismo que el de la variable aleatoria correspondiente distribuida binomialmente, mientras que la varianza de  $X$  es un poco más pequeña que la correspondiente en el caso binomial. El “término de corrección”  $(N - n)/(N - 1)$  es aproximadamente igual a 1, para un  $N$  grande.

**Ejemplo observación c) pág. 232**

## 1.10 La distribución Multinomial

Consideremos una variable aleatoria discreta importante de mayor dimensión que se puede concebir como una generalización de la distribución binomial. Consideremos un experimento  $\epsilon$ , su espacio muestral  $S$  y una partición de  $S$  en  $k$  eventos mutuamente excluyentes  $A_1, \dots, A_k$  (es decir, cuando se efectúa  $\epsilon$  uno y sólo uno de los eventos  $A_i$  ocurre.) Considérese  $n$  repeticiones independientes de  $\epsilon$ . Sea  $p_i = P(A_i)$  y supóngase que  $p_i$  permanece constante durante todas las repeticiones. Desde luego tenemos que  $\sum_{i=1}^k p_i = 1$ . Definamos las variables aleatorias  $X_1, \dots, X_k$  como sigue.

$X_i$  es el número de veces que  $A_i$  ocurre entre las  $n$  repeticiones de  $\epsilon$ ,  $i = 1, \dots, k$ .

Las  $X_i$  no son variables aleatorias independientes, puesto que  $\sum_{i=1}^k X_i = n$ . Entonces, tan pronto como el valor de cualquiera de las  $(k-1)$  variables aleatorias es conocido, se determina el valor de la otra.

Ahora si... presentamos el teorema

**Teorema.** Si  $X_i, i = 2, \dots, k$  están definidas como antes, tenemos que

$$P(X_1 = n_1, \dots, X_k = n_k) = \frac{n!}{n_1!n_2!\dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}, \text{ donde } \sum_{i=1}^k n_i = n.$$

**Demostración** pág 233

La distribución anterior se conoce como **distribución multinomial de probabilidades**. Recordemos que los términos de la distribución binomial se obtuvieron del desarrollo de la expresión binomial  $[p + (1 - p)]^n = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k}$ . De manera análoga, las probabilidades anteriores pueden obtenerse de un desarrollo de la expresión multinomial  $(p_1 + p_2 + \dots + p_k)^n$

**Teorema.** Supóngase que  $(X_1, \dots, X_k)$  tiene una distribución multinomial. Entonces

$$E(X_i) = np_i \text{ y } V(X_i) = np_i(1 - p_i), i = 1, 2, \dots, k.$$

**Demostración** pág. 234

**Ejemplo** pág 234.

## 1.11 El proceso de Poisson

Consideremos una fuente de material radiactivo que emite partículas  $\alpha$ . Sea definida  $X_t$  como el número de partículas emitidas durante un periodo de tiempo específico  $[0, t]$ . Vamos a hacer algunas hipótesis acerca de la variable aleatoria (discreta)  $X_t$  que nos permitirán determinar la distribución de probabilidades de  $X_t$ . La posibilidad de estas hipótesis (recordando lo que  $X_t$  representa) se justifica por el hecho de que la evidencia empírica sostiene una cantidad considerable de resultados teóricos que vamos a derivar.

Puede ser útil señalar que en la deducción de cualquier resultado matemático debemos aceptar algunos postulados o axiomas fundamentales. En la búsqueda de axiomas para describir fenómenos observables, algunos axiomas pueden ser más apropiados (y menos arbitrarios) que otros. Por ejemplo, al describir el movimiento de un objeto impulsado hacia arriba con cierta velocidad inicial, podríamos suponer que la distancia sobre el suelo, llamémosla  $S$ , es una función cuadrática del tiempo  $t$ ; es decir,  $s = at^2 + bt + c$ . Esta sería difícilmente una hipótesis intuitiva, de acuerdo con nuestra experiencia. En su lugar, podríamos suponer que la aceleración es una constante y luego de esto deducir que  $S$  debe ser una función cuadrática de  $t$ . Lo importante es por supuesto que si debemos suponer algo con el propósito de elaborar nuestro modelo matemático, preferiríamos suponer lo que es apropiado, en vez de lo que no lo es.

La variable aleatoria  $X_t$  antes definida puede tomar los valores  $0, 1, 2, \dots$ ; sea  $P_n(t) = P[X_t = n]$ ,  $n = 0, 1, 2, \dots$

### Hipótesis

- $A_1$  El número de partículas emitidas durante intervalos de tiempo **no sobrepuestos** son variables aleatorias **independientes**.
- $A_2$  Si  $X_t$  se define como antes y si  $Y_t$  es igual al número de partículas emitidas durante  $[t_1, t_1 + t]$ , para cualquier  $t_1 > 0$ , las variables aleatorias  $X_t$  y  $Y_t$  tienen la misma distribución de probabilidades. (En otras palabras, la **distribución** del número de partículas emitidas durante cualquier intervalo **depende sólo de la longitud del intervalo** y no de los puntos extremos.)
- $A_3$   $p_1(\Delta t)$  es igual aproximadamente a  $\lambda \Delta t$ , si  $\Delta t$  es suficientemente pequeña, donde  $\lambda$  es una constante positiva. Esto lo escribimos como  $p_1(\Delta t) \sim \lambda \Delta t$  (Esta hipótesis expresa que si el intervalo es suficientemente pequeño, la probabilidad de obtener exactamente una emisión durante ese intervalo es directamente proporcional a la longitud del intervalo.).
- $A_4$   $\sum_{k=2}^{\infty} p_k(\Delta t) \sim 0$  (Esto implica que  $p_k(\Delta t) \rightarrow 0, k \geq 2$ ). Esto significa que la probabilidad de obtener dos o más emisiones en un intervalo suficientemente pequeño es despreciable.
- $A_5$   $A_0 = 0$ , o de manera equivalente  $P_0(0) = 1$ . Esto equivale a una condición inicial para el modelo que estamos describiendo.



### Conclusiones de estas hipótesis

a) Las hipótesis  $A_1$  y  $A_2$  juntas implican que la variable aleatoria  $X_t$  y  $[X_{t+\Delta t} - X_t]$  son variables aleatorias independientes con la misma distribución de probabilidades.

b) De las hipótesis  $A_3$  y  $A_4$  podemos concluir que  $p_0(\Delta t) = 1 - p_1(\Delta t) + \sum_{k=2}^{\infty} p_k(\Delta t) \sim 1 - \lambda \Delta t$

c) Podemos escribir

$$\begin{aligned} P_0(t + \Delta t) &= P[X_{t+\Delta t} = 0] \\ &= P[X_t = 0 \text{ y } (X_{t+\Delta t} - X_t) = 0] \\ &= p_0(t)p_0(\Delta t) && [\text{conclusión a)}] \\ &\sim p_0(t)[1 - \lambda \Delta t] && [\text{conclusión b)}] \end{aligned}$$

d) De la ecuación anterior obtenemos que

$$\frac{P_0(t+\Delta t) - P_0(t)}{\Delta t} \sim -\lambda p_0(t)$$

Haciendo  $\Delta t \rightarrow 0$ , y observando que el lado izquierdo representa el cociente de la diferencia de la función  $p_0$  y, por tanto, tiende a  $p'_0(t)$  (más precisamente, a la derivada por la derecha, puesto que  $\Delta t \geq 0$ ), tenemos.

$$p'_0(t) = -\lambda p_0(t) \text{ o equivalentemente, } \frac{p'_0(t)}{p_0(t)} = -\lambda$$

Integrando ambos miembros respecto a  $t$ , obtenemos  $\ln p_0(t) = -\lambda t + C$ . De la hipótesis  $A_5$  encontramos, al hacer  $t = 0$ , que  $C = 0$ . Luego,  $p_0(t) = e^{-\lambda t}$ .

Así, nuestras hipótesis nos han conducido a una expresión para  $P[X_t = 0]$ . Ahora obtengamos  $p_n(t)$ ,  $n \geq 1$

e) Considerando  $p_n(t + \Delta t) = P[X_{t+\Delta t} = n]$

Ahora,  $X_{t+\Delta t} = n$  sí y sólo sí  $X_t = x$  y  $[X_{t+\Delta t} - X_t] = n - x$ ,  $x = 0, 1, \dots, n$ . Utilizando las proposiciones  $A_1$  y  $A_2$ , tenemos

$$\begin{aligned} p_n(t + \Delta t) &= \sum_{x=0}^n p_x(t)p_{n-x}(\Delta t) \\ &= \sum_{x=0}^{n-2} p_x(t)p_{n-x}(\Delta t) + p_{n-1}(t)p_1(\Delta t) + p_n(t) + p_0(\Delta t) \end{aligned}$$

Utilizando las hipótesis  $A_3$  y  $A_4$  y c), obtenemos

$$p_n(t + \Delta t) \sim p_{n-1}(t)\lambda \Delta t + p_n(t)[1 - \lambda \Delta t]$$

Luego,

$$\frac{p_n(t+\Delta t) - p_n(t)}{\Delta t} \sim \lambda p_{n-1}(t) - \lambda p_n(t)$$

Nuevamente haciendo  $\Delta t \rightarrow 0$ , y observando otra vez que el lado izquierdo representa el cociente diferencial de la función  $p_n$ , obtenemos

$$p'_n(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad n = 1, 2, \dots$$

Esta representa un sistema infinito de ecuaciones lineales diferenciales de diferencias. El lector interesado puede verificar que si definimos la función  $q_n$  por la relación  $q_n(t) = e^{\lambda t} p_n(t)$ , el sistema anterior se transforma en  $q'_n(t) = \lambda q_{n-1}(t)$ ,  $n = 1, 2, 3, \dots$ ; puesto que  $p_0(t) = e^{-\lambda t}$ , encontramos que  $q_0(t) = 1$ . [Nótese también que  $q_n(0) = 0$  para  $n \neq 0$ .] Así obtenemos, recursivamente,

$$\begin{aligned} q'_1(t) &= \lambda & \text{y, por tanto, } q_1(t) &= \lambda t \\ q'_2(t) &= \lambda q_1(t) = \lambda^2 t & \text{y, por tanto, } q_2(t) &= \frac{(\lambda t)^2}{2} \end{aligned}$$

En general,  $q'_n(t) = \lambda q_{n-1}(t)$  y por lo tanto,  $q_n(t) = \frac{(\lambda t)^n}{n!}$ . Al recordar la definición de  $q_n$ , finalmente obtenemos

$$p_n(t) = e^{-\lambda t} (\lambda t)^n / n!, \quad n = 0, 1, 2, \dots$$

Hemos demostrado así que el número de partículas emitidas durante el intervalo de tiempo  $[0, t)$  de una fuente radiactiva, con las suposiciones hechas anteriormente, es una variable aleatoria con una distribución de Poisson con parámetros  $(\lambda t)$ .

### Observaciones.

- a) Es importante darse cuenta de que la distribución de Poisson apareció como una consecuencia de ciertas suposiciones que hicimos. Esto significa que cada vez que dichas suposiciones sean válidas (o al menos lo sean aproximadamente) la distribución de Poisson debe usarse como un modelo apropiado. Resulta que hay una gran cantidad de fenómenos para los cuales es adecuado el modelo de Poisson. **Ejemplos pág 222 i,...,iv**
- b) La constante  $\lambda$  apareció originalmente como una constante de proporcionalidad en la hipótesis A3. Vale la pena mencionar las siguientes interpretaciones de  $\lambda$ : si  $X_t$  representa el número de ocurrencias de un evento durante un intervalo de tiempo de longitud  $t$ , entonces,  $E(X_t) = \lambda t$  y, por tanto,  $\lambda = [E(X_t)/t]$  representa la razón esperada con la cual se emiten las partículas. Si  $X_V$  representa el número de ocurrencias de algún evento dentro de un volumen especificado  $V$ , entonces  $E(X_V) = \lambda V$  y, por lo tanto,  $\lambda = [E(X_V)/V]$  representa la densidad esperada con la cual aparecen las estrellas.
- c) Es importante señalar que nuestra exposición en esta sección no se refirió sólo a una variable aleatoria  $X$  que posee una distribución de Poisson, sino que para cada  $t > 0$ , encontramos que  $Xt$  tenía una distribución de Poisson con un parámetro dependiente de  $t$ . Tal colección (infinita) de variables aleatorias también se conoce como proceso de Poisson. (De igual forma, se genera un proceso de Poisson cada vez que ocurre un evento en algún intervalo de tiempo de modo que se satisfagan las hipótesis  $A_1$  hasta  $A_5$ .) De manera análoga podemos definir un proceso de Bernoulli: si  $X_1, X_2, \dots, X_n, \dots$ , son los números de ocurrencias de eventos en  $1, 2, \dots, n$ , ensayos de Bernoulli, entonces la colección de variables aleatorias  $X_1, X_2, \dots, X_n, \dots$  se llaman proceso de Bernoulli.

### Ejemplos 223-224