UNIDAD DIDÁCTICA 4

DISTRIBUCIONES DE PROBABILIDAD – PARTE 1

Objetivo: El objetivo de esta Unidad Didáctica es introducir de forma resumida los conceptos elementales sobre las distribuciones de probabilidad y la esperanza matemática indispensables para el resto de la asignatura, así como presentar sucintamente los modelos más importantes en la práctica para variables aleatorias discretas (Binomial y Poisson) y continuas (Uniforme, Exponencial y Normal).

De acuerdo con el enfoque general de la asignatura, el tratamiento de los conceptos expuestos se realiza a un nivel elemental e intuitivo, pero suficiente para que un ingeniero pueda entenderlos y aplicarlos a los problemas reales que puedan aparecer en su ejercicio profesional.

Contenido

Fuente: Romero, R.; Zúnica, L. R. *Métodos estadísticos en ingeniería*. I.S.B.N 84-9705-727-9

En esta primera parte de la Unidad Didáctica se expone el concepto de **distribución de probabilidad**, distinguiendo entre variables aleatorias discretas y continuas. A continuación, se introduce el cálculo de la **media** y la **varianza** de dichas distribuciones. Por último se presentan dos de los modelos de distribuciones para variables discretas más importantes por su aplicación práctica: la distribución **Binomial** y la distribución de **Poisson**.

1. Introducción y conceptos básicos

1.1. Variables aleatorias. Distribuciones de probabilidad

En el la UD 1 se introdujo el concepto de variable aleatoria asociada a una población, como una característica expresable numéricamente cuyo valor fluctúa según el individuo considerado de la población.

Nota: En este tema nos referiremos a una variable aleatoria utilizando una letra mayúscula en negrita, por ejemplo, variable aleatoria **X**.

La probabilidad de que una variable aleatoria tome valores comprendidos en un determinado intervalo puede interpretarse intuitivamente como la proporción de individuos de la población en los que el valor que toma dicha característica pertenece al intervalo considerado.

A toda variable aleatoria le corresponde, por tanto, una determinada forma de distribuirse dichas probabilidades en el conjunto de valores posibles. Se utiliza, en consecuencia, el término de **distribución de probabilidad** (o simplemente distribución) como sinónimo del término variable aleatoria

En el presente tema consideraremos sólo el caso de variables aleatorias unidimensionales, en el que el conjunto de valores posibles se halla contenido en la recta real $\Re_1$.

1.2. Distribuciones de probabilidad discretas

Cuando el conjunto de los valores que puede tomar una determinada variable aleatoria es discreto (es decir, finito o infinito numerable) se dice que dicha variable, o distribución de probabilidad, es de tipo discreto.

Ejemplos de variables discretas con un número finito de valores posibles serían el número de puntos al lanzar un dado (6 valores posibles) o el número de piezas defectuosas en una muestra de 20 piezas seleccionadas al azar de una partida que tiene un 10% de piezas defectuosas (número de valores posibles 0, 1, ..., 20, o sea 21 en total).

Ejemplos de variables discretas con un número infinito (o no determinado) de valores posibles serían el número de veces que hay que lanzar un dado hasta obtener un 5, o el número de accidentes mortales en los fines de semana en las carreteras españolas.

<u>Función de probabilidad</u>

La forma de caracterizar la distribución de probabilidad de una determinada variable discreta es a partir de la <u>función de probabilidad</u>, también denominada a veces <u>función de cuantía</u> o <u>función de masa</u>, función que da la probabilidad de cada uno de los valores posibles x_i de **X**. Se simboliza como:

$$\boxed{\mathbf{P(X)}}$$

La función de probabilidad da la probabilidad de que ($\mathbf{X} = x_i$) para todo x_i cuya probabilidad es > 0

Ejercicio 1: Sea la población constituida por todos los posibles lanzamientos de dos monedas simétricas. A cada individuo de la población, es decir a cada lanzamiento de las dos monedas, se le asocia la variable aleatoria "número de caras obtenidas".

¿Cuál es el conjunto de valores posibles de la variable aleatoria? ¿Es dicha variable discreta? Calcular la función de probabilidad de dicha variable aleatoria.

1.3. Distribuciones de probabilidad continuas

A nivel intuitivo podemos decir que una variable aleatoria es continua si su conjunto de valores posibles es un infinito continuo (en la práctica, si sus valores podrían apreciarse con un gran número de decimales si se dispusiera de un aparato de medida suficientemente preciso).

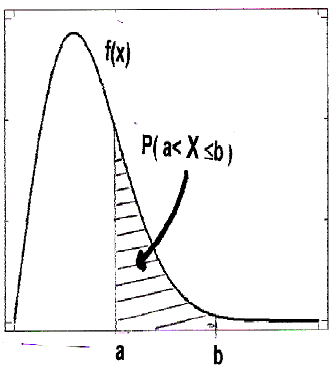
Nota: Sin embargo, esta definición, que puede ser útil en la práctica en la inmensa mayoría de los casos, no es completamente correcta. A diferencia de lo que sucede con las variables discretas, cuyo tratamiento puede llevarse a cabo con herramientas matemáticas muy sencillas, el estudio de las variables aleatorias continuas exige el recurso a herramientas de análisis más sofisticadas, que en este texto se expondrán sólo de una forma elemental y poca rigurosa.

<u>Función de densidad</u>

La forma de caracterizar la distribución de probabilidad de una determinada variable continua **X** es a partir de su función de densidad**:**

$$\boxed{\mathbf{f(x)}}$$

Su propiedad fundamental es que **el área comprendida bajo la función de densidad de una variable aleatoria entre dos valores "a" y "b", coincide con la probabilidad de que la variable aleatoria tome valores en dicho intervalo**.



$$\int_a^b f(x)dx = P(a < X \leq b)$$

> **Nota**: en la expresión anterior el signo "≤" podría sustituirse por "<", dado que para variables continuas la probabilidad de un valor exacto es nulo, puesto que el área sobre un punto es nula.

Un resultado de gran importancia práctica es que, tanto para variables aleatorias discretas como continuas, la probabilidad de que una variable tome valores dentro de un intervalo [a, b] se puede obtener como:

$$\mathbf{P(a < X \leq b) = P(X \leq b) - P(X \leq a)}$$

La relación entre el concepto de función de densidad y el de histograma de frecuencias se ilustra en el siguiente ejemplo.

Considérese la variable aleatoria "dureza de los respaldos de poliuretano de los vehículos fabricados por cierta factoría" (expresada en Newtons).

Si representamos los valores de esta variable mediante un histograma de frecuencias relativas, la barra trazada sobre cada tramo tiene un área igual a la proporción de observaciones en dicho tramo (**Figura 1**), pero se irá aproximando a la función de densidad a medida que aumenta el número de tramos (**Figura 2** y **Figura 3**).

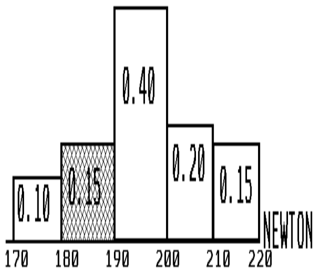


Figura 1

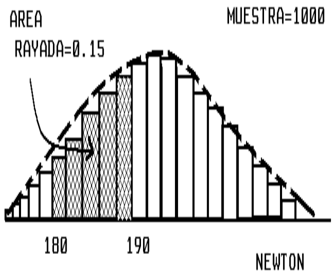


Figura 2

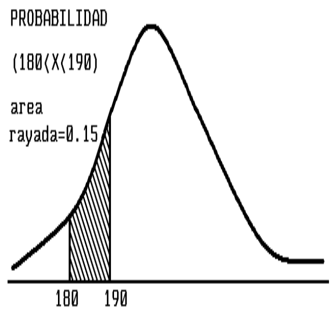


Figura 3

1.4. Esperanza matemática

El concepto de media aritmética, o promedio, de un conjunto de valores observados, definido como la suma de todos ellos dividida por el número de valores, tiene una clara interpretación intuitiva. Una idealización de dicho concepto lleva a la definición de la **Esperanza Matemática**, o **Valor Medio**, de una función h(**X**) de una determinada variable aleatoria **X**.

1.4.1. Variables discretas

En el caso de que **X** sea una variable aleatoria discreta, la esperanza matemática, o valor medio, de dicha variable aleatoria se expresará como **E(h(X))** y se define como:

$$E(h(X)) = \sum_i h(x_i) P(X = x_i) \qquad (1)$$

sumatorio que se extiende para todos los valores x_i.

1.4.2. Variables continuas

Si la variable **X** es continua, en la expresión del valor medio hay que sustituir la probabilidad P(**X**=x_i) por el valor f(x) de la función de densidad en x, y el sumatorio por una integral definida en la que los límites de integración se limitarán finalmente a la región en la que f(x) sea diferente de cero:

$$E(h(\mathbf{X})) = \int_{-\infty}^{+\infty} h(x)f(x)dx \qquad (2)$$

1.5. Valor medio: concepto y propiedades

En el apartado anterior se ha definido el concepto de esperanza matemática de una función cualquiera h(**X**) de una variable aleatoria **X**. En el caso particular de que sea h(**X**) = **X**, a la esperanza matemática **E(X)** se le denomina **valor medio**, o simplemente **media**, de dicha variable aleatoria **X**. Por tanto:

Valor medio de una variable discreta: $E(\mathbf{X}) = \sum_i x_i P(X = x_i)$

Valor medio de una variable continua: $E(\mathbf{X}) = \int_{-\infty}^{+\infty} x f(x)dx$

El **valor medio** o **media** se simboliza como **μ** o **m**.

Ejercicio 2: Hallar el valor medio de la variable aleatoria “Número de caras obtenidas” definida en el Ejercicio 1

Propiedades del valor medio

Una propiedad fundamental del valor medio es que es un operador lineal, o sea que la media de una combinación lineal de variables aleatorias es la combinación lineal de las medias de las mismas:

$$\mathbf{E(a_0 \pm a_1X_1 \pm \ldots \pm a_nX_n) = a_0 \pm a_1E(X_1) \pm \ldots \pm a_nE(X_n)}$$

En particular se cumplirá, por tanto, que

si $\mathbf{Y} = a\mathbf{X} \pm b \Rightarrow E(\mathbf{Y}) = aE\mathbf{X}) \pm b$

si $\mathbf{Y} = \mathbf{X_1} \pm \mathbf{X_2} \Rightarrow E(\mathbf{Y}) = \mathbf{E}(\mathbf{X_1}) \pm E(\mathbf{X_2})$

Ejercicio 3: Calcular el valor medio de la variable aleatoria “Nº de puntos obtenidos al lanzar dos dados simétricos”

1.6. Varianza: concepto y propiedades

Siendo m la media de una variable aleatoria X, se denomina **varianza** de dicha variable (y se simboliza como σ^2) a la esperanza matemática de la función $h(\mathbf{X}) = (\mathbf{X} - m)^2$

$$\text{Varianza} = \boldsymbol{\sigma^2} = E(\mathbf{X}\text{-}m)^2$$

Por tanto:

Varianza de una variable discreta: $\sigma^2(\mathbf{X}) = \sum_i (x_i - m)^2 P(X = x_i)$

Varianza de una variable continua: $\sigma^2(\mathbf{X}) = \int_{-\infty}^{+\infty} (x - m)^2 f(x)dx$

A la raíz cuadrada positiva σ de la varianza se le denomina **desviación típica**.

Tanto la varianza como la desviación típica, como ya se expuso en la UD 2, tienen una gran importancia en Estadística pues cuantifican el grado de **dispersión** de la variable alrededor de su valor medio.

Propiedades de la varianza

$$\boxed{\boldsymbol{\sigma^2(a + bX) = b^2\sigma^2(X)}}$$

Si **X** e **Y** son independientes, entonces:

$$\boxed{\boldsymbol{\sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y)}}$$

Nota. En general, la varianza de una suma de variables aleatorias se puede obtener como:

$$\sigma^2(a_0 + a_1X_1 + a_2X_2) = a_1^2\sigma^2(X_1) + a_2^2\sigma^2(X_2) + 2a_1a_2\,\mathrm{Cov}(X_1,X_2)$$

Ejercicio 4: Calcular la varianza de la variable aleatoria "Nº de puntos obtenidos al lanzar dos dados simétricos"

2. Principales distribuciones de probabilidad

Se introducen en este apartado los conceptos de las cinco distribuciones de probabilidad más importantes en la práctica: la distribución Binomial, la distribución de Poisson, la distribución Uniforme, la distribución Exponencial y la distribución Normal.

Para cada una de ellas se da su definición y sus principales propiedades, así como, en su caso, el manejo de ábacos o tablas para obtener sus probabilidades, y se ilustran sus posibles aplicaciones con algunos ejemplos escogidos.

2.1. La distribución Binomial

Sea **A** un suceso de probabilidad **p** asociado a un determinado experimento aleatorio. Supongamos que llevamos a cabo **n** repeticiones independientes del experimento en cuestión, y sea **X** el número de veces que se presenta el suceso **A**.

La variable **discreta X** así definida tiene como valores posibles los enteros 0, 1, 2,..., **n** siendo la probabilidad de cada uno de ellos función de los valores concretos de **n** y **p**. Dicha variable sigue una distribución denominada **distribución binomial** que depende de los dos parámetros mencionados **n** y **p**, y nos referiremos a ella como una variable **B(n,p)**

$$\boxed{\mathbf{X \sim B(n,p)}}$$

Ejemplos de variables aleatorias binomiales son:

- El número de cincos que se obtienen al lanzar 10 dados simétricos será una variable binomial B(n = 10, p= 1/6)

- El número de piezas defectuosas (X) que se obtienen al extraer al azar una muestra de 10 piezas de un almacén. Se sabe que el 30% de las piezas son defectuosas. De esta forma X será una variable binomial B(n = 10, p = 0.3)

- Un caso especial, denominado **distribución de 2 puntos**, o **distribución de Bernouilli** lo constituyen las variables binomiales con n=1. Estas variables sólo tienen 2 valores posibles, 0 y 1, siendo su función de probabilidad P(**X**=1) = p y P(**X**=0) = 1-p

Se demuestra que la función de probabilidad de una variable **X** que sigue una distribución Binomial de parámetros **n** y **p** es:

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

en la que $\binom{n}{x}$ son las combinaciones de n elementos tomados de x en x.

La función de probabilidad de una X~B(n=11,p=0,2) tiene la siguiente forma (**Figura 4**):

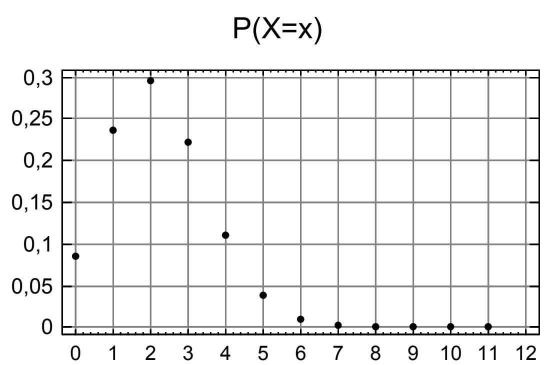


Figura 4. Función de probabilidad de una variable aleatoria X~B(n=11,p=0,2)

Ejercicio 5: Para controlar la calidad de las partidas que recibe, el servicio de control de calidad de una empresa extrae una muestra de 50 piezas, aceptando la partida si en la muestra hay menos de 2 unidades defectuosas. ¿Qué probabilidad tiene este plan de inspección de aceptar partidas que tengan un 1% de unidades defectuosas? (Asumir que el número de unidades de la partida es grande, de modo que los resultados de las diferentes extracciones pueden considerarse independientes)

Ejercicio 6: Calcular la probabilidad de obtener al menos un seis en seis lanzamientos de un dado simétrico.

Se demuestra que la **media** y la **varianza** de una distribución binomial vienen dadas por las expresiones:

$$\mathbf{E(X) = np} \qquad \boldsymbol{\sigma^2(X) = np(1\text{-}p)}$$

Ejercicio 7: Calcular número medio de "piezas defectuosas" del Ejercicio 5 y el número medio de "seises obtenidos" del Ejercicio 6.

El cálculo de la probabilidad $P(B(n,p) \leq x)$ puede ser laborioso. Una propiedad muy útil al respecto es que cuando el producto **np(1-p)** es grande (del orden de 9 o más) las probabilidades correspondientes a una variable binomial pueden también aproximarse usando las tablas de la distribución normal, según se expondrá en el apartado 2.5.

2.2. La distribución de Poisson

En algunas aplicaciones se manejan variables aleatorias **discretas** que pueden asimilarse a variables binomiales con un valor muy elevado de **n** y un valor muy bajo de **p**. De dichas variables sólo se conoce su valor medio **np**.

Ejercicio 8: Considérese en la población de todos los coches del aparcamiento de la ETSINF, la variable aleatoria "nº de abolladuras por capó de cada vehículo". Razonar por qué dicha variable podría considerarse como una binomial con una **n** muy elevada (¿qué sería realmente **n**?) y una **p** muy baja (¿qué sería realmente **p**?).

La distribución de Poisson se describe como el límite al que tiende la distribución de una variable binomial cuando $\mathbf{n} \rightarrow \infty$ y $\mathbf{p} \rightarrow 0$ de forma que el producto **np** tiende a una constante λ.

Se demuestra que la función de probabilidad de una distribución de Poisson de parámetro λ viene dada por la expresión:

$$P(X = x) = P\left(Ps(\lambda) = x\right) = e^{-\lambda}\frac{\lambda^x}{x!}$$

La función de probabilidad de una Ps~B(λ=3) tiene la siguiente forma (**Figura 5**):

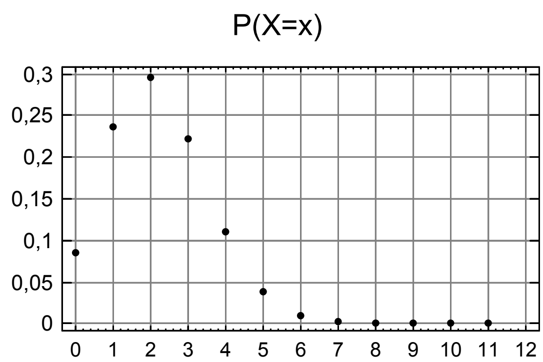


Figura 5. Función de probabilidad de una variable aleatoria P(Ps(λ=3)=x)

Ejercicio 9: Si el número medio de abolladuras por capó de los vehículos definida en el Ejercicio 7 es 0,8. ¿Qué porcentaje de coches no tendrá ninguna abolladura?

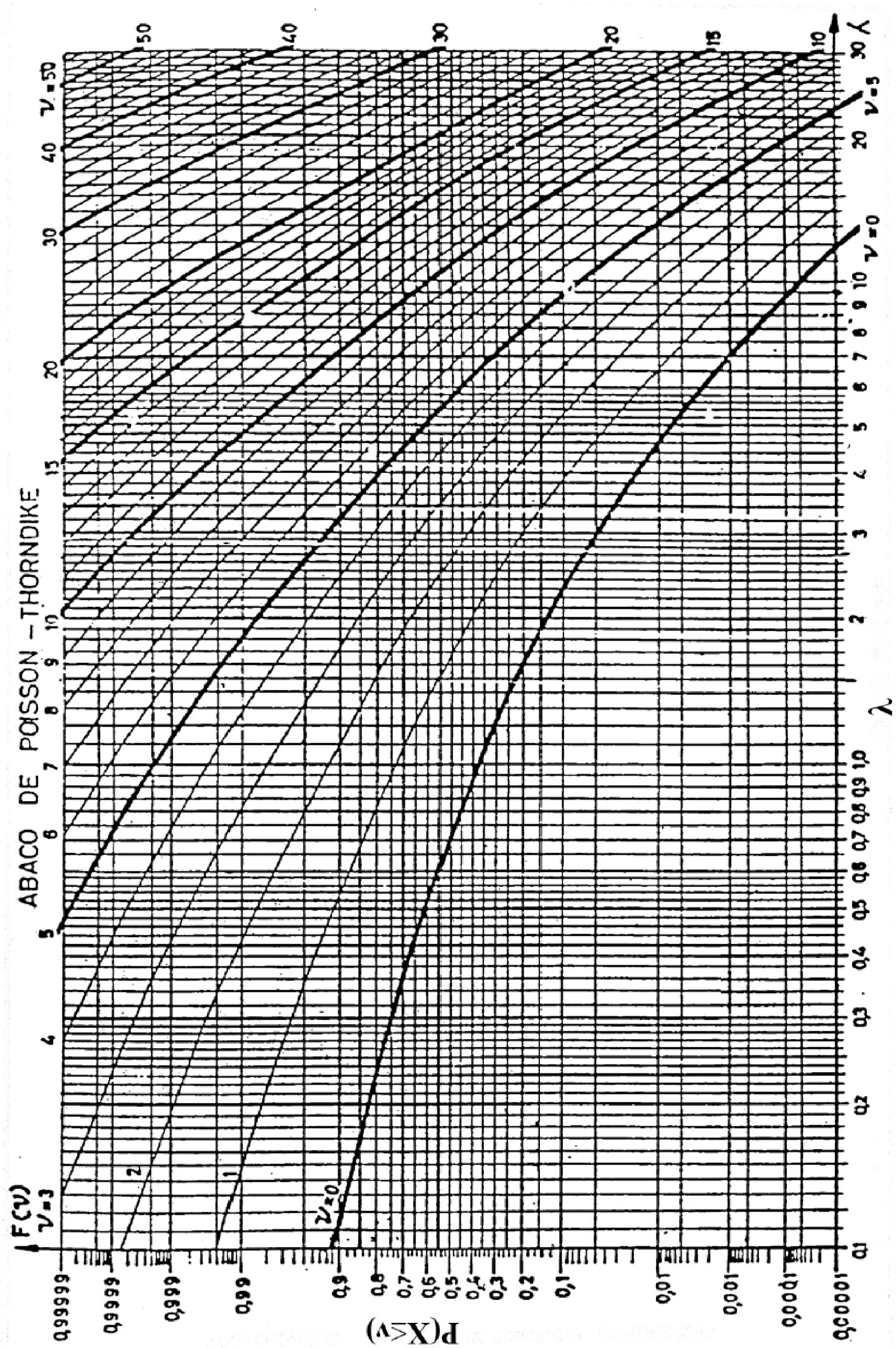
De la definición dada para una distribución de Poisson se deducen inmediatamente los valores de su media y de su varianza. En efecto siendo **X** una variable de Poisson de parámetro λ se tiene:

$$E(\mathbf{X}) = \lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\lambda}} np = \lambda \quad \text{y} \quad \sigma^2(\mathbf{X}) = \lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\lambda}} np(1-p) = \lambda$$

$$\mathbf{E(X) = \sigma^2(X) = \lambda}$$

Para el cálculo de las probabilidades asociadas a una distribución de Poisson se utiliza el ábaco que hay en la página 18 al final del tema que da, en función del parámetro λ y del valor considerado ν, la probabilidad de que una variable de Poisson de parámetro λ sea menor o igual que ν

Ejercicio 10: Siguiendo con la variable del ejercicio anterior, ¿Qué porcentaje de coches tendrá más de una abolladura?. ¿Qué porcentaje de coches tendrá más de cinco abolladuras?


ABACO DE POISSON – THORNDIKE
P(X≤v)
λ
F(v)
ν = 0
ν = 5
ν = 50

Ejercicios resueltos

Apartado 5.A.1 del Capítulo 5 del libro de R. Romero y L.R. Zúnica "Métodos Estadísticos en Ingeniería" SPUPV 637

Apartado 6.A.1 del Capítulo 6 del libro de R. Romero y L.R. Zúnica "Métodos Estadísticos en Ingeniería" SPUPV 637

Apartados 7.A.1 y 7.A.2 del Capítulo 7 del libro de R. Romero y L.R. Zúnica "Métodos Estadísticos en Ingeniería" SPUPV 637

Ver boletín correspondiente en PoliformaT (EST GII: Recursos / 04 | Ejercicios)

Para saber más

- Dibujando interactivamente la función de probabilidad de una variable aleatoria binomial: http://goo.gl/iiuw
- Applet interactivo para dibujar la distribución de Poisson: http://goo.gl/d30R
- El tablero de Galton (Galton's board) es un experimento para ilustrar el comportamiento *casi gaussiano* de una distribución binomial con $p = 1/2$: http://goo.gl/g78V
- Distribución Binomial y aproximación normal. (David Lane, Rice University, Texas (USA)): http://www.ruf.rice.edu/~lane/stat_sim/normal_approx/index.html
- Otras distribuciones discretas que puede ser interesante conocer: http://goo.gl/Ft6U

Fuentes

Métodos Estadísticos en Ingeniería (Romero Villafranca, Rafael)