Capítulo 4

Variables aleatorias discretas

4.1. Introducción

Recordemos que una **variable aleatoria** es una variable cuyo valor depende del resultado de un experimento aleatorio. En el tema 1 (apartado 1.1.) se vieron diversos ejemplos y se distinguió entre variables cualitativas (o categóricas) y cuantitativas. Dentro de éstas últimas, diferenciamos entre:

- **variables discretas**: toman un conjunto finito o infinito numerable (que se pueden contar) de valores
- **variables continuas**: su espacio muestral está formado por un conjunto infinito de valores que no podemos contar

En el tema anterior consideramos básicamente variables categóricas, mientras que en este tema nos centraremos en las variables aleatorias discretas, y el tema siguiente lo dedicaremos a las variables aleatorias continuas.

En este tema veremos unos modelos matemáticos concretos que nos darán la pauta de variabilidad asociada a una variable aleatoria. Estos modelos matemáticos se llaman distribuciones de probabilidad. Una **distribución de probabilidad** es un conjunto de probabilidades para los posibles distintos sucesos que pueden darse en un experimento aleatorio, en otras palabras, lo que nos proporciona es cómo se *distribuye* la probabilidad entre los sucesos que pueden producirse.

[Nota: sólo veremos el caso univariante (una única variable), sin embargo también existen modelos que consideran varias variables conjuntamente].

En el presente tema, estudiaremos 3 modelos (hay muchos más, en los complementos del tema puedes encontrar algunos), que corresponderán a la consideración de experimentos con determinadas características. El fin de estos modelos teóricos es la descripción razonable de algunos fenómenos aleatorios. Son modelos aleatorios o estocásticos, que se diferencian de los modelos matemáticos determinísticos. Para los modelos determinísticos, los resultados se encuentran predeterminados por las condiciones bajo las cuales se verifica el experimento, es decir, dada una entrada su salida (resultado) queda determinada. Por ejemplo, una fuente de alimentación (E) suministra corriente a circuito de resistencia eléctrica (R), el modelo matemático que nos describiría el flujo de corriente viene dado por la Ley de Ohm I=E/R. El modelo suministaría el valor de I tan pronto como se dieran los valores de E y R. Sin embargo, para los experimentos aleatorios, los resultados nos pueden predecirse con certeza.

Los tres modelos que estudiaremos son: la uniforme discreta, la Binomial y la Poisson. Tanto la distribución Binomial como la de Poisson tiene aplicación en fiabilidad y en control de calidad (cómo se verá en la práctica 3 con el Statgraphics). La fiabilidad estudia la probabilidad de funcionamiento de una unidad, entendida no sólo como parte indescomponible de un sistema, sino también como un sistema o subsistema en sí. En el tema 3 (ejemplo 3.8.), ya vimos en clase algunos casos sencillos. El control de calidad, justamente se encarga de *controlar la calidad*, en la práctica 3 puedes encontrar más detalles.

Distribución uniforme discreta: es la distribución que sigue una variable aleatoria X que toma n posibles valores x_1, x_2, ..., x_n con la misma probabilidad. Por tanto,

$$P(X = x_i) = \frac{1}{n} \qquad \text{i = 1, ..., n}$$

Ejemplo 3.1.: X=”resultado al lanzar un dado no trucado”

P(X = 5) =

P(1 < X ≤ 4) = P(X=2) + P(X=3) + P(X=4) =

P(X > 4) = P(X=5) + P(X=6) =

P(X ≥ 4) = P(X=4) + P(X=5) + P(X=6) =

P(X < 6) = 1 - P(X ≥ 6) = 1 - P(X=6) =

[Fíjate que para calcular las probabilidades de los sucesos que nos interesen (por ejemplo

A = (1 < X ≤ 4) = {2, 3, 4}) sumaremos las probabilidades de los puntos muestrales que forman el suceso (como vimos en el tema anterior). Recuerda que la suma de todas las probabilidades de todos los puntos muestrales es 1.]

4.2. Binomial

Esta distribución tiene una amplia gama de aplicaciones, sobre todo cuando se trata de realizar pruebas cuyo resultado sólo puede adoptar dos valores: "éxito" o "fracaso".

Supongamos que llevamos a cabo un **proceso de Bernoulli**, es decir, una serie de n pruebas. Cada prueba puede resultar en un "éxito" o en un "fracaso". La probabilidad de éxito es la misma cantidad, p, para cada prueba, sin importar los resultados de las otras pruebas, o sea, las pruebas son independientes.

La variable aleatoria X que representa el número de éxitos observados en un proceso de Bernoulli tiene una **distribución binomial**.

Ejemplo 4.1.: El ejemplo por excelencia de variable aleatoria distribuida como una binomial, sería X = "número de caras obtenidas al lanzar una moneda no trucada 5 (por ejemplo) veces", en este caso $n = 5$ y $p = 0.5$. O bien, X = "número de caras obtenidas al lanzar una moneda trucada (de forma que la probabilidad de salir cara sea 0.7) 10 (por ejemplo) veces", en este caso $n = 10$ y $p = 0.7$.

Si la variable X sigue (se distribuye como) una distribución binomial de parámetros n y p (siendo n el número de pruebas y p la probabilidad de éxito), que representaremos como $X \sim Bi(n, p)$, las probabilidades se distribuyen de la siguiente manera (considerando combinatoria podría deducirse):

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, ..., n, \qquad q = 1 - p \qquad \text{, donde}$$

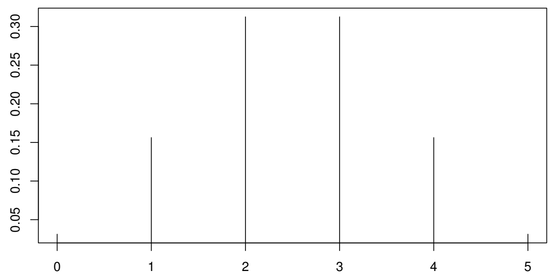
$$\binom{n}{x} = \frac{n!}{x! \cdot (n-x)!} \qquad \text{siendo} \qquad n! = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot 2 \cdot 1$$

[Fíjate que una variable $X \sim Bi(n, p)$, sólo puede tomar un número de valores FINITO, de 0 a n.]

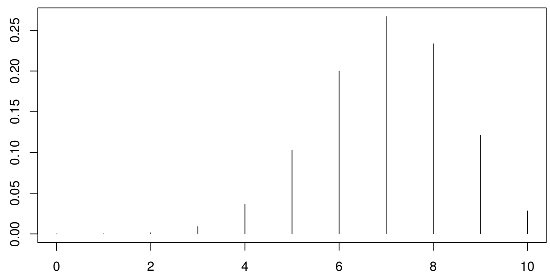
Calculadora: para calcular $\binom{n}{x}$ puede emplearse la teclas $\boxed{nCr}$ o bien $\boxed{x!}$. Recuerda también que $\binom{n}{0} = \binom{n}{n} = 1$, $\binom{n}{1} = \binom{n}{n-1} = $ n, $0! = 1$ y $1! = 1$.

Ejemplo 4.1.: Las siguientes gráficas muestran cómo se distribuye (reparte) la probabilidad entre los puntos muestrales, para las dos variables de este ejemplo. Fíjate que si sumamos todas las probabilidades de los puntos muestrales obtendremos 1.

Binomial(5, 0.5)



Binomial(10, 0.7)



En cada problema, debe especificarse qué quiere decir "éxito". "Éxito" puede ser "salir cara" como en el ejemplo 4.1, o bien, por ejemplo "ser defectuoso", "ser satisfactorio", o "cumplir las especificaciones", etc.

Ejemplo 4.2.: La calidad del acabado de cierto producto se califica como: alta, media y baja, de manera que el 70 % son de alta calidad, el 20 % de calidad media y el 10 % baja. Si seleccionamos 10 productos aleatoriamente, ¿cuál sería la probabilidad de que

8 sean de alta calidad?

al menos 3 sean de alta calidad?

4 sean de calidad media?

como máximo 3 sean de calidad baja?

[Presta atención al significado de las expresiones que nos interesen: *al menos* (quiere decir $\geq$), *como máximo, o como mucho* (quiere decir $\leq$), *menos de* (quiere decir $<$), exactamente (quiere decir $=$), etc. Además, recuerda que los contrarios de $<$ es $\geq$, de $>$ es $\leq$. Cuando convenga, usa lo que ya sabes sobre los sucesos contrarios, y así te ahorrarás muchos cálculos.]

Puesto que en este tema estamos estableciendo modelos teóricos que describan el *comportamiento* de ciertas variables aleatorias, también podremos establecer cuál sería la media poblacional, μ, y la varianza poblacional, σ^2 (recuerda el apartado 1.5. del tema 1), usando estos modelos.

Para una variable Binomial, $X \sim Bi(n,p)$, se tiene $\mu = \mathrm{n} \cdot \mathrm{p}$, y $\sigma^2 = \mathrm{n} \cdot \mathrm{p} \cdot \mathrm{q}$. La media, μ también se llama esperanza matemática (de hecho en el problema 14 del tema 3, realizado en clase, ya se introdujo este concepto).

Ejemplo 4.2.:¿Cuál sería el número esperado de productos de alta calidad?

4.3. Poisson

Consideremos ahora una serie de experimentos que consisten en observar el número de ocurrencias de un hecho en un intervalo de tiempo o espacio determinado. Por ejemplo:

Ejemplo 4.3.: número de fallos de un equipo industrial durante 5 años.

Ejemplo 4.4.: número de defectos de fabricación por cada 1000 metros de cable.

Una variable aleatoria X sigue una **distribución de Poisson**, si cuenta el número de ocurrencias por unidad de magnitud, cuando:

- el número de ocurrencias en un intervalo de tiempo o del espacio es independiente del número de ocurrencias en otro intervalo disjunto (proceso sin memoria).
- Además, la probabilidad de que haya una sola ocurrencia en un intervalo muy corto es proporcional a la amplitud del intervalo y
- la probabilidad de que haya más de una ocurrencia en un intervalo muy corto es despreciable.

Si la variable X sigue (se distribuye como) una distribución Poisson de parámetro λ ($X \sim \mathrm{Po}(\lambda)$), donde λ indica el número medio de ocurrencias por unidad de magnitud y suele denominarse parámetro de intensidad, las probabilidades se distribuyen de la siguiente manera (también podemos probarla, aunque no lo haremos en clase):

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0,\ 1,\ 2,\ 3,\ ... \qquad (x \in \mathbb{N})$$

[Fíjate que una variable $X \sim \mathrm{Po}(\lambda)$, puede tomar un número INFINITO NUMERABLE (contable) de valores.]

En el caso de una variable Poisson, $X \sim \text{Po}(\lambda)$, se tiene que $\mu = \lambda$ y $\sigma^2 = \lambda$.

✍ **Ejemplo 4.5:** Supongamos que el número de defectos que aparecen en una lámina de $1m^2$ puede modelizarse como una variable Poisson de parámetro 3. Determina:

¿cuál es el número medio de defectos en una lámina de $1m^2$?

la probabilidad de encontrar un defecto: P(X = 1) =

la probabilidad de encontrar como mucho un defecto: P(X $\leq$ 1) = P(X =0) + P(X = 1) =

la probabilidad de encontrar algún defecto: P(X $\geq$ 1) = P(X = 1) + P(X = 2) + ... (suma infinita) = suceso contrario = 1 - P(X < 1) = 1 - P(X = 0) =

la probabilidad de encontrar más de uno pero menos de cuatro defectos: P(1 < X < 4) = P(X = 2) + P(X = 3) =

Si ahora consideramos el mismo tipo de láminas, pero de $5m^2$. Determina:

¿cuál es la variable de interés?¿cuál es el número medio de defectos en una lámina de $5m^2$?

la probabilidad de no encontrar ningún defecto: P(X = 0) =

la probabilidad de encontrar algún defecto: P(X $\geq$ 1) = 1 - P(X = 0) =

la probabilidad de que no haya más de dos defectos: P(X $\leq$ 2) = P(X = 0) + P(X = 1) + P(X = 2) =

Si ahora, consideramos 6 láminas de $1m^2$. Determina:

¿cuál es la variable de interés? ¿cuál es el número medio esperado de láminas con defectos?

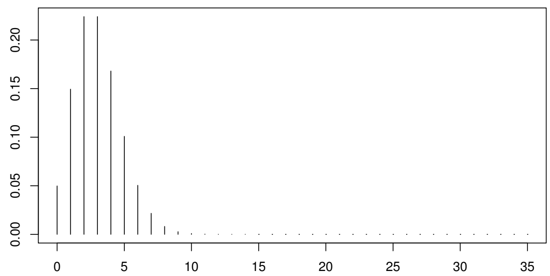
la probabilidad de tener exactamente 1 lámina con defectos: P(X = 1) =

la probabilidad de tener a lo sumo una lámina con defectos: P(X ≤ 1) = P(X = 0) + P(X=1)=

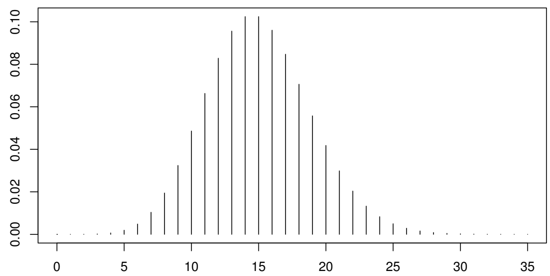
Si la retirada de una lámina defectuosa, supone un gasto de 15 €, ¿cuál es la probabilidad de que para esas 6 láminas tengamos que gastar por lo menos 30 €?

Ejemplo 4.5.: Las siguientes gráficas muestran cómo se distribuye (reparte) la probabilidad entre los puntos muestrales, para las dos variables Poisson de este ejemplo. Fíjate que si sumamos todas las probabilidades de los puntos muestrales obtendremos 1.

Poisson(3)



Poisson(15)



Propiedad Sea $X \sim Bi(n, p)$. Si n es grande, p pequeña y $n \cdot p$ moderada, podremos aproximar X por una Poisson $(n \cdot p)$.

Ejemplo 4.6: Un comprador de grandes cantidades de circuitos acepta un envío si en el lote no hay más de 2 circuitos defectuosos. Se sabe que el 1 % de los circuitos son defectuosos, y los lotes constan de 120 circuitos.

¿Cuál es la probabilidad de aceptar el lote?

¿Cuál es la probabilidad que haya exactamente 60 circuitos defectuosos en el lote?

4.4. Función de probabilidad. Función de distribución acumulada

En fotocopias.

Problemas del tema 4

Ejemplo Una industria recibe productos semielaborados en lotes de un gran número de unidades. Se desea preparar un plan de control de calidad de tal forma que tomando n unidades del lote, si se observa alguna unidad defectuosa, se rechace el lote. Determina el valor de n para que si el lote tiene un 5 por mil o más de unidades defectuosas, la probabilidad de aceptarlo sea menor del 1 %.

La variable que nos interesa es X = "número de unidades defectuosas en una muestra de tamaño n". Esta es una variable Binomial(n,p), siendo p la proporción de defectuosas en el lote, y n el tamaño que hemos de determinar.

La probabilidad de aceptar el lote será: P(X = 0), que debe ser < 0.01, si p ≥ 0.005, con lo cual:

$$P(X=0) = \binom{n}{0} p^0 \cdot q^n = 1 \cdot 1 \cdot (1-p)^n$$

Como p ≥ 0.005, $P(X=0) = (1-p)^n \leq (1-0{,}005)^n < 0.01 \rightarrow 0{,}995^n < 0.01 \rightarrow$

tomando logaritmos, $n \cdot ln(0.995) < ln(0.01) \rightarrow$ -0.00501· n < -4.605 →

despejamos n y tenemos en cuenta el signo negativo de -0.00501, $n >$ -4.605/-0.00501 = 919.16

1. La probabibilidad de tener una unidad defectuosa en una línea de encaje es de 0.05. Si el número de unidades acabadas constituye un conjunto de pruebas independientes y tenemos 10 unidades:

   a) ¿Cuál es la probabilidad que dos sean defectuosas?
   b) ¿Cuál es la probabilidad que como máximo dos sean defectuosas?
   c) ¿Cuál es la probabilidad que al menos dos sean defectuosas?

   (Sol: 0.075, 0.9888, 0.0862)

2. El número de componentes que fallan antes de cumplir las 100 horas de operación se comporta como una variable Poisson. Si el número medio de éstas es 8:

a) ¿Cuál es la probabilidad que falle un componente en 25 horas de funcionamiento?
b) ¿Cuál es la probabilidad que fallen como máximo dos componentes en 50 horas?
c) ¿Cuál es la probabilidad que fallen al menos 10 en 125 horas?

(Sol: 0.27067, 0.2381, 0.54207)

3. La probabilidad que un chip resulte defectoso es de 0.001. Calcula la probabilidad que al tomar una muestra al azar de 2000 chips, resulten defectosos:

   a) al menos 3 chips
   b) exactamente 2 chips
   c) como máximo 1 chip
   d) exactamente 5 chips

   (Sol: $\sim$ 0.3233, $\sim$ 0.2708, $\sim$ 0.4059, $\sim$ 0.0361)

4. Si la probabilidad de que cierta columna de ala ancha caiga bajo una carga axial dada es de 0.05. ¿Qué probabilidad hay de que entre 16 columnas de este tipo

   a) caigan como mucho 2?
   b) caigan al menos 4?

   (Sol: 0.9571, 0.007)

5. La contaminación es un problema en la fabricación de discos de almacenamiento óptico. El número de partículas contaminantes que aparecen en un disco óptico tiene una distribución de Poisson, y el número promedio de partículas por centímetro cuadrado de superficie del medio de almacenamiento es 0.1. El área de un disco bajo estudio es 100 cm^2. ¿Cuál es la probabilidad de encontrar 12 partículas en el área del disco? ¿Y cero?

   (Sol: 0.095, 4.54 $\cdot 10^{-5}$)

6. Una empresa sigue un proceso de fabricación tal que le permite servir sus pedidos con un retraso inferior a una semana en el 95 % de los casos, lo cual es considerado como satisfactorio. Reclamaciones de ciertos clientes hacen sospechar a la dirección que el porcentaje de retrasos ha aumentado, por lo cual se detiene a estudiar si ello es cierto. La forma de efectuar dicho estudio es la siguiente: se selecciona al azar 10 pedidos y se procede a observar cómo son servidos. Si más de 1 pedido sufre retraso superior a una semana, se revisa el proceso de fabricación y en caso contrario no se revisa. ¿Cuál es la probabilidad de que el proceso sea revisado sin necesidad?

   (Sol: 0.086)

7. Si la probabilidad de que a cualquier persona le desagrade el sabor de una nueva pasta dentífrica es de 0.2, ¿cuál es la probabilidad de que les desagrade a 5 de 18 personas aleatoriamente seleccionadas?

   (Sol: 0.1507)

8. Durante una etapa en la manufactura de chips debe aplicarse a éstos una capa. Si el 70 % de los chips recibe una capa suficientemente gruesa, determina la probabilidad de que entre 15 chips,

   a) al menos 12 tengan una capa suficientemente gruesa.
   b) como máximo 6 tengan una capa suficientemente gruesa.
   c) exactamente 10 tengan una capa suficientemente gruesa.

   (Sol: 0.2969, 0.0152, 0.2061)

9. En una clase el 60 % son mujeres y el resto hombres. Si seleccionamos una muestra al azar de 12 estudiantes [1]:

   a) ¿Cuál es la probabilidad de que la mitad sean mujeres?
   b) ¿Cuál es la probabilidad de que haya más de cuatro hombres?

   (Sol: 0.1766, 0.562)

10. En la inspección de hojalata producida por un proceso electrolítico, se identifican 0.2 imperfecciones en promedio por minuto. Suponiendo que el número de imperfecciones se comporta como una variable Poisson, determina la probabilidad de identificar

    a) una imperfección en 3 minutos.
    b) al menos dos en 5 minutos.
    c) como mucho 1 en 15 minutos.

    (Sol: 0.329, 0.264, 0.199)

11. Este ejercicio ilustra el impacto que la baja calidad puede tener sobre planes y costos. Un proceso de fabricación tiene 100 pedidos en espera de ser surtidos. Cada pedido necesita un componente que se compra a otro proveedor. Sin embargo, lo común es identifcar 2 % de estos componentes como defectuosos; por otra parte, puede suponerse que el estado de cada componente es independiente del de los demás.

    a) Si el inventario del fabricante es de 100 componentes, ¿cuál es la probabilidad de que se puedan surtir los 100 pedidos sin tener que pedir más componentes?
    b) Si el inventario del fabricante es de 102 componentes, ¿cuál es la probabilidad de que se puedan surtir los 100 pedidos sin tener que pedir más componentes?
    c) Si el inventario del fabricante es de 105 componentes, ¿cuál es la probabilidad de que se puedan surtir los 100 pedidos sin tener que pedir más componentes?

    (Sol: 0.13, 0.666, 0.977)

12. Una partida de bujías con alta proporción de inservibles (20 %) sale al mercado en paquetes de 4 unidades y en cajas de 10 paquetes. Cálcula la probabilidad de que:

[1] aunque en realidad la variable sea hipergeométrica, podemos aproximarla por una binomial (mira los complementos del tema 4)

a) Elegido un paquete al azar contenga al menos 2 bujías inservibles.

b) Elegida una caja al azar contenga 39 bujías inservibles.

c) Elegida una caja al azar contenga 3 paquetes sin bujías inservibles.

(Sol: 0.1808, 1.7592 10^{-26}*, 0.2062)*