# UNIDAD TEMÁTICA 4 Lección 4
# VARIABLE ALEATORIA

## ENUNCIADO 1

*Se hacen n lanzamientos independientes con un dado ordinario de 6 lados. Calcula la probabilidad que:*

*(a) El mayor de los números obtenidos sea k con k ∈ {1, 2, ..., 6}.*
*(b) El menor de los números obtenidos sea k con k ∈ {1, 2, ..., 6}.*

**Resolución:**

La variable aleatoria que se enuncia en este ejercicio es claramente de naturaleza discreta, que tiene asociado un espacio probabilístico (espacio muestral), $\Omega = \{1,2,...,6\}^n$, con $\mathcal{F} = \mathcal{P}(\Omega)$. Es decir, si $\omega \in \Omega$ será $P(\{\varpi\}) = \frac{1}{6^n}$, si se considera un dado con lanzamientos independientes.

(**a**) Sea $X_i \triangleq$ "el valor del i-ésimo lanzamiento", $\forall i \in \{1,2,...,6\}$. Entonces, se entiende fácilmente que

$$P_a = P\left[\max_{i=1,2,...,n}\{X_i\} \le k\right] = P\left[X_i \le k \ \forall i = 1,2,...,n\right]$$

que teniendo en cuenta la independencia entre lanzamientos se verifica que

$$P_a = \prod_{i=1}^{n} P\left[X_i \le k\right]$$

Además, como las variables $X_i$ son equiprobables se deduce que

$$P\left[X_i \le k\right] = P\left[\bigcup_{i=1}^{k}\{X_i = k\}\right] = \sum_{i=1}^{k} P\left[X_i = k\right] = \sum_{i=1}^{k}\frac{1}{6} = \frac{k}{6}$$

Por lo tanto:

$$P_a = P\left[\max_{i=1,2,...,n}\{X_i\} \le k\right] = \left(\frac{k}{6}\right)^n, \ \forall k = 0,1,...,6$$

Además, se observa la siguiente regla de recurrencia:

# UNIDAD TEMÁTICA 4 Lección 5-6
# DISTRIBUCIONES DE PROBABILIDAD

## ENUNCIADO 1

*Según una encuesta el número de personas que contrajeron una determinada enfermedad no contagiosa en un país de 10 millones de habitantes el último año fue de 235000, habiéndose repartido la incidencia de la enfermedad por igual en todo el territorio. Calcula: (**a**) De una familia de 10 personas, la probabilidad de que menos de 2 hayan contraído la enfermedad. (**b**) De un pueblo de 10000 personas, al menos 220 hayan contraido la enfermedad.*

**Resolución:**

Antes de proceder a contestar a las cuestiones planteadas ha de estimarse la proporción de contagios que ha tenido lugar, $\hat{\pi} = \frac{235\times10^3}{10^6} = 0.0235$, junto con el error probable de la estimación, $\sigma_{\hat{\pi}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.0235\times0.9765}{10^6}} = 4.7904\times10^{-5}$ (este resultado pertenece a otro tema, pero se cometa aquí para que el/la alumn@ se comience a familiarizar con las técnicas inferenciales). Dicha estimación se supondrá constante para los cálculos ques siguen. Sea X := "*número de personas que contrajeron una determinada enfermedad no infecciosa*", la variable binomial correspodiente al enunciado.

(a) $$p(X<2)\Big|_{\substack{n=10\\p=0.0235}} = 1-\sum_{k=0}^{1}\binom{n}{k}p^n(1-p)^{n-k} = \begin{cases} binomial: & 0.978078353 \\ Poisson: & 0.976354999 \end{cases}$$

Se puede aplicar Poisson ya que $\lambda = np = 0.235 \le 4$ (o equivalentemente, p pequeño y n grande ($n \gg 50$)).

(**b**) Se observa que no podemos hacer uso ni de las tablas ni de la expresión de la distribución binomial dada la naturaleza de los datos que intervienen, puesto que se pide calcular la $p(X \ge 220)$. Del enunciado, se puede utilizar la distribución normal ($np = 235 > 4$, $nq = 9675 >> 4$) de características

$$\mathcal{N}\left(\mu = np = 235, \sigma = \sqrt{npq} = 15.1485\right)$$

para efectuar la aproximación a la cuestión planteada. Entonces:

$$p(X \ge 220) \underset{\substack{correccion\\por\ continuidad}}{=} p\left(Z \ge z_1 = \frac{219.5-235}{\sigma}\right) = p\left(Z \ge z_1 = -1.02320\right) = 15.3106042\%$$

$$= p\left(Z \le z_1 = 1.02320\right) = 84.6893958\%$$

## ENUNCIADO 2

*De una muestra de 2000 personas de una población, 3 han mostrado reacción adversa a un determinado fármaco.*

*(**a**) Estimar la proporción de personas con reacción adversa en dicha población indicando la desviación típica de la estimación.*

*(**b**) En una muestra de 1000 sujetos, ¿cuál es la probabilidad de que exactamente 3 personas muestren reacción adversa?, ¿menos de 2 personas muestren reacción adversa?*

*(**c**) De una muestra de 100000 personas, ¿cuál es la probabilidad de que menos de 140 personas reacciones adversamente a dicho fármaco?*

**Resolución:**

(a) En este caso la variable aleatoria binomial es X ≜ "*número de personas que han mostrado reacción adversa a un determinado fármaco*". Se trata de una estimación puntual de proporciones (grandes muestras, $n \gg 30$):

$$\hat{\pi} = p = \frac{3}{2000} = 0.0015$$

con error probable (o muestra)

$$\sigma_{\hat{\pi}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.0015 \times 0.9985}{2000}} = 0.000865376$$

(b) Se ha de suponer que la tasa de reacciones adversas se mantiene, y como se tiene que no podemos aplicar las tablas de la distribución dicotómica (que acaban en n = 20). Por otra parte, se observa que $\lambda = np = 1000 \times 0.0015 = 1.5 \Rightarrow$ se puede aproximar la distribución original $\mathcal{B}(n = 2000, p = 0.0015)$ por una distribución de Poisson / $\mathcal{P}(\lambda = 1.5)$. Entonces:

$$p(X = 3) \overset{tablas}{=} 0.934 - 0.809 = 0.125$$

$$p(X < 3) \overset{tablas}{=} 0.558$$

(**c**) Ahora se sigue que $\begin{array}{l} \mu = np = 100000 \times 0.0015 = 150 > 4 \\ nq = 100000 \times 0.9985 = 99850 > 4 \end{array} \Rightarrow$ aproximar la distribución binomial original por la distribución normal con

$$\mathcal{N}(\mu = 150, \sigma = \sqrt{npq} = \sqrt{150 \times 0.9985} = 12.2383)$$

Entonces, la probabilidad que se pide es:

$$p(X<140) \underset{\substack{correccion\ por \\ continuidad}}{\simeq} p(X \leq 139.5) = p\left(Z \leq \frac{139.5-150}{12.2383}\right) = p(Z \leq -0.85796) =$$
$$= 1 - p(Z \leq 0.85796) = 1 - 0.80454277 = 0.19545723$$

## ENUNCIADO 3

*Se sabe que la probabilidad de que un estudiante de enseñanza primaria presente escoliosis (curvatura de la espina dorsal) es 0.004. De los siguientes 1875 estudiantes que se revisen en busca de escoliosis:*

*(**a**) ¿Cuál es el número medio de estudiantes que presentan escoliosis?*
*(**b**) ¿Cuál es la desviación estándar?*
*(**c**) Encuentra la probabilidad de que ocho, nueve o 10 estudiantes presenten el problema.*

**Resolución:**

Se trata de una variable aleatoria de Poisson ya que p = 0.004 << 1 y n = 1875 > 50 (es decir, es un "suceso raro"). Así pues:

(a) La cantidad $\lambda = np = 7.5\ estudiantes$ es el número medio que presenta escoliosis,

(b) Siendo su desviación estándar $s = \sqrt{\lambda} = 2.7386\ estudiantes$.

(c) Entonces, se deduce que la probabilidad pedida es:

$$p(8\ estudiantes \leq X \leq 10\ estudiantes) =$$
$$p(X \leq 10\ estudiantes) - p(X \leq 7\ estudiantes) = 0.8622 - 0.5246 = 0.3376$$

valores que se obtienen interpolando en las tablas de trabajo existentes.

## ENUNCIADO 4

*(**a**) Supongamos que los diámetros d de los frascos fabricados por una empresa se distribuyen normalmente con $\mu$ = 0.635 cm y $\sigma$ = 0.050 cm. Se considera que un frasco es defectuoso si d $\leq$ 0.508 cm o d $\geq$ 0.711 cm. Hallar el porcentaje de frascos defectuosos fabricados por dicha empresa.*

*(**b**) Supongamos que las 220 erratas de un libro de 200 páginas están distribuidas aleatoriamente por el mismo. Hallar la probabilidad de que una página cualquiera contenga: (1) ninguna errata, (2) no más de 2 erratas, (3) exactamente 2 erratas, (4) 2 o más erratas.*

*(**c**) Supongamos que el 1 % de los objetos producidos por una máquina son defectuosos. Hallar la probabilidad de que en una muestra de (1) 15 objetos (2) 100 objetos, 3 o más sean defectuosos.*

**Resolución:**

(**a**) Al ser una distribución normal de parámetros $\mathcal{N}\left(\mu = 0.635\, cm, \sigma = 0.050\, cm\right)$ se pide calcular la probabilidad

$$\mathbb{P}(D \le d_1 = 0.508\, cm \vee D \ge d_2 = 0.711\, cm) = \mathbb{P}\left( Z \le z_1 = \frac{0.508 - 0.635}{0.050} \vee Z \ge z_2 = \frac{0.711 - 0.635}{0.050} \right) =$$

$$\mathbb{P}\left(Z \le z_1 = 2.54 \vee Z \ge z_2 = 1.52\right) \underset{\substack{por\ ser\ sucesos \\ incompatibles}}{=} \mathbb{P}\left(Z \le z_1 = 2.54\right) + \mathbb{P}\left(Z \ge z_2 = 1.52\right) =$$

$$= 0.00554265 + 0.06425551 = 0.06979816$$

(**a**) Se trata ahora de una variable aleatoria de Poisson de parámetros $\mathcal{P}\left( \lambda = \frac{220}{200} = 1.1\, {}^{erratas}\!/\!_{pagina} \right)$. Así, las probabilidades pedidas son:

(1) $\mathbb{P}\left(X = 0\, erratas\right) \underset{tablas}{=} 0.332871084$

(2) $\mathbb{P}\left(X \le 2\, erratas\right) \underset{tablas}{=} 0.900416281$

(3) $\mathbb{P}\left(X = 2\, erratas\right) \underset{tablas}{=} 0.201387006$

(4) $\mathbb{P}\left(X \ge 2\, erratas\right) \underset{tablas}{=} 1 - \mathbb{P}\left(X < 2\, erratas\right) = 1 - 0.699029276 = 0.300970724$

(c) La variable aleatoria en este caso es binomial de forma que

(1) $\mathcal{B}\left(n = 15, p = 0.01\right)$. Sin embargo, no se puede usar la tabla y como $\lambda = np = 15 \times 0.01 = 0.15$, se puede aproximar por una variable aleatoria de Poisson, de modo que con $\mathcal{P}\left(\lambda = 0.15\, defectos\right)$ la probabilidad pedida es

$$\mathbb{P}\left(Y \ge 3\, defectos\right) = 1 - \mathbb{P}\left(Y \le 2\, defectos\right) = 1 - \sum_{k=0}^{2} \frac{e^{-\lambda}\lambda^k}{k!} = 1 - 0.999497138 = 0.000502862$$

Utilizando la distribución binomial (también con la fórmula) se obtiene que $\mathbb{P}\left(Y \ge 3\, defectos\right) = 0.000415803$.

(2) Ahora es $\mathcal{B}\left(n = 100, p = 0.01\right)$ de la que no disponemos de tablas. Ahora pues, la aproximación se hace con la distribución de Poisson $\mathcal{P}\left(\lambda = np = 100 \times 0.01 = 1\, defectos\right)$. Usando la misma estrategia que en el apartado (c1) se deduce que $\mathbb{P}\left(Y \ge 3\, defectos\right) = 1 - 0.919698603 = 0.080301397$, y si se aplica la distribución binomial se obtiene el resultado $\mathbb{P}\left(Y \ge 3\, defectos\right) = 0.079373202$.

$$P\left[\max_{i=1,2,\ldots,n}\{X_i\}\le k\right]=P\left[\left\{\max_{i=1,2,\ldots,n}\{X_i\}\le k-1\right\}\cup\left\{\max_{i=1,2,\ldots,n}\{X_i\}=k-1\right\}\right]=$$

$$\left(\frac{k-1}{6}\right)^n+P\left[\max_{i=1,2,\ldots,n}\{X_i\}=k\right],\ \forall k=1,\ldots,6$$

Despejando se obtiene el resultado buscado:

$$P\left[\max_{i=1,2,\ldots,n}\{X_i\}=k\right]=\left(\frac{k}{6}\right)^n-\left(\frac{k-1}{6}\right)^n$$

(**b**) Tomaremos como punto de partida que $\min_{i=1,2,\ldots,n}\{X_i\}=\max_{i=1,2,\ldots,n}\{7-X_i\}$. En consecuencia:

$$P\left[\min_{i=1,2,\ldots,n}\{X_i\}=k\right]=P\left[7-\max_{i=1,2,\ldots,n}\{X_i\}=k\right]=P\left[\max_{i=1,2,\ldots,n}\{X_i\}=7-k\right]$$

o lo que es lo mismo

$$P\left[\min_{i=1,2,\ldots,n}\{X_i\}=k\right]=\left(\frac{7-k}{6}\right)^n-\left(\frac{7-k-1}{6}\right)^n=\left(1-\frac{k-1}{6}\right)^n-\left(1-\frac{k}{6}\right)^n$$

Si uno se da cuenta, de esta relación se puede calcular

$$P_b=P\left[\min_{i=1,2,\ldots,n}\{X_i\}\ge k\right]=P\left[X_i\ge k\ \forall i=1,2,\ldots,n\right]$$

y razonar como en el apartado (**a**). Más aún, el método utilizado para calcular la distribución del máximo se puede utilizar para cualquier familia de variable aleatorias iid (independientes idénticamente distribuidas) y lo mismo para el mínimo.

## ENUNCIADO 2

*Si Y es una variable aleatoria que sigue una distribución uniforme en [0; 5], ¿cuál es la probabilidad de que ambas raíces de la ecuación*

$$4x^2 + 4xY + Y = 0$$

*sean reales?*

**Resolución:**

La definición de la variable aleatoria continua de este ejercicio es Y $\triangleq$ "variable aleatoria uniforme en [0, 5] con densidad de probabilidad" $f_Y[x]:=\frac{1}{5}1_{[0,5]}(x),\ \forall x\in\mathbb{R}$.

Al considerar que Y es una función medible que está defnida en cierto espacio muestral $\Omega$ y a valores en $\mathbb{R}$, se tendrá que dado un $\omega \in \Omega$ la ecuación en la indeterminada x $4x^2 + 4xY(\varpi) + Y(\varpi) = 0$ tiene sus dos raíces reales si y solo si se cumple la siguiente condición del discriminante $[4Y(\varpi)]^2 - 16Y(\varpi) \geq 0 \Leftrightarrow [Y(\varpi)-1]Y(\varpi) \geq 0$. De forma equivalente, al definir en $\Omega$ el suceso

$$R := \{\omega \in \Omega : \text{la ecuacion } 4x^2 + 4xY(\varpi) + Y(\varpi) = 0 \text{ tiene ambas raices reales}\}$$

Entonces:

$$R := \{\omega \in \Omega : [Y(\varpi)-1]Y(\varpi) \geq 0\} \qquad \textbf{[1]}$$
$$\{\omega \in \Omega : Y(\varpi) \geq 0 \wedge Y(\varpi) \geq 1\} \cup \{\omega \in \Omega : Y(\varpi) \leq 0 \wedge Y(\varpi) \leq 1\}$$

Pero como $P(\{\omega \in \Omega : Y(\varpi) \leq 0 \wedge Y(\varpi) \leq 1\}) = P(\{\omega \in \Omega : Y(\varpi) \leq 0\}) = \int_{-\infty}^{0} f_Y(x)dx = 0$.

De [1] se concluye que

$$P(R) = P(\{\omega \in \Omega : Y(\varpi) \geq 0 \wedge Y(\varpi) \geq 1\}) = P(\{\omega \in \Omega : Y(\varpi) \geq 1\})$$
$$= P(Y \in [1, +\infty]) = \int_{1}^{+\infty} f_Y(x)dx = \int_{1}^{5} \frac{1}{5}dx = \frac{4}{5}$$

## ENUNCIADO 3

*Consideremos el experimento (E') de tirar dos veces e independientemente un dado y a continuación sumar los resultados obtenidos en cada tirada. Se pide que el alumno:*

*(a) Modele el experimento (E') como la evaluación de cierta variable aleatoria X definida en cierto espacio probabilístico (Ω, A, P) y da valores en cierto espacio de probabilidad (Ω', A').*

*(b) Determine en (E') la probabilidad de que ocurra el suceso "la suma de los resultados obtenidos en cada tirada es par".*

*(c) Determine si en (E') el suceso "los resultados obtenidos en cada tirada del dado son números consecutivos" puede o no ser descrito a partir de los valores que toma la variable aleatoria X.*

Este ejercicio se deja sin resolver por el momento. Aquel grupo de alumnos que deseen resolverlo y entregarlo pueden hacerlo y entregarlo al profesor a través del correo electrónico (con el formato que se ha solicitado en la Tarea TIC 4), siendo la fecha tope el **lunes 10 de abril de 2006 a las 14:30 horas**.