

# UN EJEMPLO PRÁCTICO DE CÁLCULO DE PROBABILIDADES

Francisco José Mora Lizán, Mar Pujol López, Ramón Rizo Aldeguer

*Grupo de Investigación (Visión, Gráficos, e Inteligencia Artificial)  
Departamento de Ciencia de la Computación e Inteligencia Artificial  
Universidad de Alicante  
e-mail: [mora, mar, rizo]@dccia.ua.es*

**Resumen:** En este artículo no haremos un desarrollo exhaustivo de los objetivos de la asignatura, ni en la metodología utilizada, ya que estos temas han sido abordados suficientemente en otros artículos. Aquí vamos a describir un proyecto, creado para los alumnos de las ingenierías informáticas, para compaginar con las prácticas de la asignatura. ¿Qué mejor forma de aprender estadística que practicando con casos reales y que son atractivos para el alumno?. Nuestra innovación consiste en que vamos a utilizar el ejemplo desde el principio hasta el final, practicando con toda la parte teórica de la asignatura.

## 1.- INTRODUCCIÓN.

Estadística I es una asignatura troncal perteneciente a primer curso de las titulaciones de Ingeniero Técnico en Informática de Gestión (ITIG), Ingeniero Técnico en Informática de Sistemas (ITIS) e Ingeniería Informática (II). Esta asignatura pertenece al segundo cuatrimestre y tiene una carga lectiva de 4.5 créditos teóricos y 1.5 créditos prácticos.

Intentamos impartir la asignatura utilizando una fuerte dosis de experimentabilidad, aplicando una metodología de enseñanza activa, práctica y crítica, rehusando una formación demasiado teórica y enciclopédica. El programa de curso que seguimos aparece en la **tabla 1**.

Nosotros planteamos un problema bastante conocido y tratado superficialmente en muchos cursos de Estadística: las **Quinielas de Fútbol**. No olvidemos tampoco lo difícil que es buscar ejemplos que el alumno

conozca y le resulten atractivos. Nosotros trataremos de despertar su curiosidad a través de ampliaciones sucesivas del problema. Nuestra innovación consiste en que vamos a utilizar el ejemplo desde el principio hasta el final, **practicando con toda la parte teórica de la asignatura**. También nos apoyaremos en un programa informático de Gestión de Quinielas desarrollado por los autores del presente artículo, donde el alumno podrá ir realizando simulaciones de juego y viendo los premios que se consiguen. Esto crea una rivalidad entre los alumnos, los cuales llegan a interesarse por la asignatura al verle una aplicabilidad inmediata. Los fuentes del programa de Gestión de Quinielas también están disponibles para el alumno, a fin de que pueda ver la utilización en algoritmos de los contenidos impartidos.

0. Análisis combinatorio. Variaciones ordinarias. Variaciones con repetición. Permutaciones ordinarias. Permutaciones con repetición. Combinaciones ordinarias. Combinaciones con repetición. Algunas series. Problemas.	3. Esperanza. 3.1. Esperanza de una variable. 3.2. Esperanza de variables discretas no negativas. 3.3. Varianza 3.4. Momentos. 3.5. Covarianza y correlación. 3.6. Esperanza condicional. 3.7. Media Muestral. 3.8. Problemas.
1. Cálculo de probabilidades. 1.1. Experimentos y sucesos. 1.2. Definición de probabilidad. 1.3. Permutaciones ordinarias. 1.4. Permutaciones con repetición. 1.5. Combinaciones ordinarias. 1.6. Combinaciones con repetición.	4. Distribuciones especiales. 4.1. Distribuciones de Bernoulli y Binomial. 4.2. Distribución de Poisson. 4.3. Distribución Normal. 4.4 Teorema central del límite. 4.5. Problemas
2. Variables aleatorias. 2.1. Introducción. 2.2. Función de distribución. 2.3. Variables discretas. 2.4. Variables continuas. 2.5. Variables continuas bidimensionales. 2.6. Funciones de una variable aleatoria 2.7. Problemas.	

Tabla 1. Contenidos de Estadística 1.

## 2.- PRIMERAS APROXIMACIONES. VARIACIONES Y PERMUTACIONES.

**a) Problemas con variaciones:** El primer ejemplo a ver es el más sencillo y directo. Vamos a ver cuantas combinaciones son necesarias para acertar 14 resultados en una quiniela, es decir el número de combinaciones que existen. Como tenemos 14 partidos y 3 posibles resultados (1, X y 2), los cuales se pueden repetir hemos de emplear la formula para Variaciones con

Repetición:  $VR_{n,h} = n^h$ ,  $VR_{3,14} = 3^{14} = 4.782.969$ . Es decir que tenemos más de 4 millones de combinaciones y suponiendo un precio de 50 Pts. por combinación nos tocaría gastar 239.148.450 Pts.

**b) Problemas con permutaciones:** Imaginemos lo interesante que puede resultar si hacemos un estudio y observamos el número medio de unos, equis y doses que salen en cada quiniela. Si queremos saber el número de quinielas que son necesarias rellenar para acertar 14 resultados sabiendo que salen X unos, Y equis y Z doses, utilizaremos la fórmula de permutaciones con repetición:  $m!/(a!*b!*...*r!)$ . Imaginemos que sabemos que van a salir 6 unos, 5 doses y 3 equis. Entonces el resultado será el siguiente:  $14!/(6!*5!*3!)$ . Veremos más adelante, al hablar de los intervalos de confianza, que interesará trabajar sobre un rango de unos, equis y doses. También vemos que el número de combinaciones es muy amplio y que es necesario **estudiar los resultados de los partidos y de las quinielas anteriores**, para reducir el número de combinaciones, quedándonos con las de mayor probabilidad.

### 3.- APROXIMACIÓN A RESULTADOS EN PARTIDOS (SEGUIMIENTO DE LA LIGA).

**a) Definición de espacio muestral:** Ateniéndonos a la definición de espacio muestral como conjunto de todos los posibles resultados de la experiencia aleatoria, el espacio muestral en un partido vendrá dado por tres posibles valores (resultados) 1, X o 2, es decir que un equipo gane, empate o pierda. Dicho espacio será discreto y finito como podemos ver trivialmente.

Esta no es la única forma de definir el espacio muestral, ya que podíamos haberlo definido como un par ordenado (x, y) donde "x" representaría los goles del un equipo e "y" los goles del equipo contrario. Así el espacio muestral sería discreto e infinito numerable. Nosotros utilizaremos la primera definición dada.

**b) Definición de función cuantía/densidad:** Hablaremos de función de cuantía ya que tratamos con variables aleatorias discretas. Definiremos la función de cuantía asignando la misma probabilidad a 1, X y a 2, es decir:

$$f_x(1)=P(x=1)=0.33; f_x(X)=P(x=X)=0.33; f_x(2)=P(x=2)=0.33;$$

Como podemos ver se cumplen las propiedades de la función de cuantía:

1)  $f_x(x) \geq 0$ .

2)  $\sum f_x(x) = 1$  con  $x=1, X, 2$ .

Más adelante podremos ver otras formas más interesantes de definir la función de cuantía. Por ejemplo:

- Asignar un valor según el número de unos, equis y doses que salen en las jornadas ligueras. Si salen más unos le daremos mayor valor al 1 etc...
- Asignar un valor según el número de unos, equis y doses que salen en las quinielas, así si calculamos los 14 posibles resultados de una quiniela, se tenderá a que se rellene de forma que el número de unos, equis y doses coincida con los que suelen salir semanalmente.
- Asignar diferentes valores según los equipos que juegan, es decir nos basaremos en el histórico de partidos.

**c) Definición de esperanza matemática:** Como  $x$  (variable aleatoria) es discreta definiremos la esperanza matemática como:

$E(h(x)) = \sum h(x_i) \cdot P(x=x_i)$ , donde  $h(x)$  es una función uniforme de  $x$ . La forma más sencilla de definir la esperanza matemática es definiéndola como media o valor esperado  $E(x)$ . Imaginemos que damos los valores 1, 2, 3 a 1, X y 2 respectivamente. Si calculamos la media nos dará el resultado más probable o esperado. También obtendremos un dato de sumo interés y es la distancia hacia nuestros valores (1,2,3). Si por ejemplo obtenemos un valor 1,5 sabemos que el resultado puede ser 1 o X. Por otro lado si obtenemos un valor 1.08 sabemos que el resultado tiende claramente a 1.

Pero no acaban aquí nuestros problemas, ya que en un partido juegan 2 equipos y estamos obteniendo la esperanza para cada equipo independientemente. ¿Cómo uniremos estos resultados? ¿Tiene el mismo valor ganar por goleada que por un gol de diferencia? ¿Juegan los equipos igual en casa que fuera?. Responder a todas estas preguntas no es sencillo. El cálculo automático de pronósticos se basa en los datos estadísticos de los equipos. Se trata de obtener la probabilidad de que se produzca un 1, una X o un 2 Para obtener estos porcentajes se utilizan tres variables de cálculo:

- Goles:** Goles del equipo de casa y del equipo visitante en esta temporada.
- Partidos ganados, empatados y perdidos (GEP):** en función de los partidos ganados, empatados y perdidos por cada equipo en esta temporada.
- Historia:** aplicando exactamente los criterios anteriores pero para ligas anteriores y puntuando más los datos más recientes.

El porcentaje final de cada pronóstico se obtiene como una composición de estos porcentajes parciales. Para calcular estos porcentajes utilizaremos una serie de parámetros, que a continuación comentamos.

- **Porcentaje Partidos:** el porcentaje parcial se forma como un compendio el porcentaje de Goles y el porcentaje de ganados, empatados y perdidos. Entre los dos suman 100%. Lo normal es repartir equitativamente entre los dos (50% para cada uno).
- **Porcentaje del equipo directo:** los porcentajes se calculan teniendo en cuenta que el equipo de casa juega en casa y el de fuera juega fuera. Pero

también se puede hacer que influya en el resultado final la posibilidad inversa; es decir, que el equipo de casa jugase fuera y el visitante jugase en casa. Lo normal es dar mucho más porcentaje a los cálculos directos (80% por ejemplo) y dejar el resto a los cálculos inversos.

- Porcentaje para 1, X y 2: Podemos calcular el valor como el número medio de unos, equis y doses que salen en las quinielas o darle un valor fijo a cada uno teniendo en cuenta que la suma de los porcentajes debe ser 100.

**d) Definición de desviación típica:** Es la medida de dispersión más utilizada; viene a ser la media cuadrática de las desviaciones de los valores de la variable respecto a su media aritmética:  $\varphi = \sqrt{\sum f_i \cdot |x_i - \text{media}|^2 / \sum f_i}$ , donde  $f_i$  es la frecuencia con la que ha salido cada resultado (1, X, 2).

También se puede definir en relación a la varianza como:

$$\varphi = \text{RAIZ\_QUADRADA}(\text{varianza})$$

**e) Definición de varianza:** La varianza mide la dispersión media de los valores de una variable respecto de un valor medio. La fórmula de la varianza es la siguiente:  $\text{var}(x) = E(x^2) - E(x)^2$

Calculemos la varianza para el caso de darle la misma probabilidad a los 3 valores y asignando los valores 1, 2, 3 a 1, X Y 2 respectivamente.

$$\text{var}(x) = E(x^2) - E(x)^2 = 14/3 - 2^2 = 14/3 - 12/3 = 2/3$$

**f) Intervalos de confianza:** Al calcular la esperanza de un resultado nos da un posible valor que se aproxima en diferente medida a los elementos del espacio muestral. Pero ¿cuanto se debe de aproximar a un elemento para tener claro su valor?

En las quinielas estos intervalos podrían definirse de diferentes formas dando lugar a las diferentes jugadas para un partido. Se utilizarían los MÚLTIPLES, es decir no damos una solución única a un partido sino que lo hacemos mediante dobles y triples (1X, 12, X2, 1X2). Hacemos notar que el valor 12 sería bastante difícil de encontrar en un cálculo de intervalos.

#### 4.- APROXIMACIÓN A LOS RESULTADOS DE LA QUINIELA.

**a) Definición de espacio muestral:** El espacio muestral vendrá dado por todas las quinielas posibles. Utilizando la fórmula de variación con repetición, podemos ver que hay 4.782.969 posibles resultados. No tenemos en cuenta el Pleno al quince de las quinielas, por simplificación del problema. Como podemos ver, **el espacio muestral es discreto y finito**. Todos los elementos del espacio muestral tienen la misma probabilidad matemática de salir (1/4.782.969), pero si hacemos un estudio de las columnas ganadoras, comprobaremos que existen unos haces prácticamente inútiles en el juego. Para eliminar estas columnas utilizaremos la opción de

“Condicionar combinación”. En esta podremos elegir el rango de equis, doses, variantes, equis seguidas, doses seguidos y variantes seguidas que permitimos en nuestras columnas. Existe una ampliación que permite aplicar lo anterior a un grupo de equipos en lugar de a todos. Estos métodos son conocidos en el mundo quinielístico como métodos DERIVATIVOS.

**b) Definición de función cuantía/densidad:** En este caso también hablaremos de función de cuantía, al ser el espacio muestral discreto. En principio podríamos definir el valor de la función de cuantía de forma similar para cada una de las 4.783.969 columnas. Esta sería una forma muy sencilla, que cumpliría todas las propiedades, pero a su vez sería muy ineficiente. En el apartado anterior ya nos adelantábamos un poco a la solución: los métodos derivativos. Otra forma de definir la función de cuantía es como la distancia al número de unos, equis y doses que suelen salir etc... Decir que es bastante complicado definir una función de cuantía y que se sigan cumpliendo las propiedades.

**c) Definición de esperanza matemática.**

-Número de aciertos esperados: Cabe destacar la forma de jugar en reducción donde con menos dinero mantenemos la misma esperanza de premios al 13 (100%), pero rebajamos la probabilidad de acertar 14 (normalmente al 11,11%).

- Cantidad de premios (pesetas) esperados: Haremos uso de la media para saber el valor esperado de los premios. Una vez sabemos los premios esperados, podremos calcular lo que podremos cobrar. Es bastante importante hacer un estudio de intervalos, es decir los premios máximos, mínimos y medios por categorías.

**d) Definición de desviación típica:** Para solucionar la desviación existen métodos como son la **CORRECCIÓN de fallos**. Supongamos que realizamos una columna y suponemos que podemos tener 2 fallos o errores. Utilizando este método aún teniendo 2 fallos en la columna base tendríamos 14 aciertos.

**e) Intervalos de confianza:** Los intervalos de confianza se deben calcular haciendo uso de la desviación típica de la quiniela y método de corrección de fallos. Mediante este último método los intervalos de confianza se amplían (es decir podemos tener más confianza de acertar).

## **5.- REFERENCIAS.**

- [1] Introducción a la estadística. Wonnacott y Wonnacott. Limusa.
- [2] Tratamiento estad. de datos. Lebart, Morineam y Fenelon. Marcombo.
- [3] Estadística modelos y métodos. Alianza universal textos.
- [4] Problemas de probabilidad y estadística. Volumen 1 y 2. Cuadras. UPV.
- [5] Problemas de estadística. Ruiz Maya
- [6] Como acertar las quinielas mediante sistemas basados en estadísticas. Equipo de expertos 2100. Vecchi.