Probabilidades y estadística

5

MINISTERIO de EDUCACIÓN
CIENCIA y TECNOLOGÍA
PRESIDENCIA de la NACIÓN

serie/ciencias para la educación tecnológica

Especialista de contenido:

- Juan Foncuberta

Asesor científico:

- Luis Santaló

Ciencias para la educación tecnológica

serie/ciencias para la educación tecnológica

1. Análisis matemático. Sus aplicaciones
2. Física. Ondas electromagnéticas
3. Física. El campo magnético
4. Física. Interacciones a distancia
5. Probabilidades y estadística

Índice

La serie "Ciencias para la educación tecnológica"

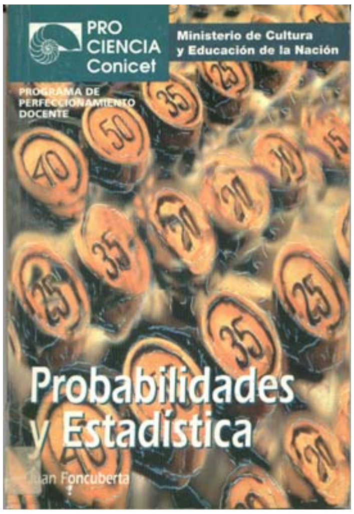
Con el título **Ciencias para la Educación Tecnológica***,* estamos planteando desde el CeNET una serie de publicaciones que convergen en el objetivo de:

Acompañar a nuestros colegas docentes en la adquisición de contenidos científicos que les permitan una mejor definición, encuadre y resolución de los problemas tecnológicos que se enseñan en la escuela.

Porque, en Educación Tecnológica las ciencias básicas ocupan "una posición importante aunque subalterna e instrumental (...) La tecnología es un modo de ver el fenómeno de la artificialidad, y de analizar 'sistémicamente' los objetos tecnológicos desde su finalidad y no desde los fundamentos científicos en que se basa su funcionamiento". (Ministerio de Cultura y Educación de la Nación. Consejo Federal de Cultura y Educación. Contenidos Básicos Comunes para la Formación Docente de Grado. Tercer Ciclo de la Educación General Básica y Educación Polimodal. "Contenidos Básicos Comunes del Campo de la Formación de Orientación de la Formación Docente de Educación Tecnológica". Buenos Aires.)[1]

PRO CIENCIA Conicet
Programa de perfeccionamiento docente
Ministerio de Cultura y Educación de la Nación

Probabilidades y estadística, el material que usted tiene en sus manos, es una versión digital de la publicación del mismo nombre que, en 1996, elaboró el Programa de Perfeccionamiento Docente Prociencia-CONICET[2], del Ministerio de Cultura y Educación de la Nación Argentina, al que desde el CeNET nos proponemos continuar distribuyendo[3].



[1] Puede usted encontrar la versión completa en el sitio web del Ministerio de Educación, Ciencia y Tecnología; específicamente, en la página:
- http://www.me.gov.ar/curriform/servicios/publica/publica/fordoc/index.html

[2] El CONICET es el Consejo Nacional de Investigaciones Científicas y Técnicas, de la Secretaría de Ciencia, Tecnología e Innovación Productiva –Ministerio de Educación, Ciencia y Tecnología–.

[3] La versión de este libro en soporte papel corresponde al ISBN 950-687-019-5, desarrollado con la coordinación de Mirta Grines de Hanfling, y con la participación del equipo docente, técnico y administrativo del Programa Prociencia.

MÓDULO 1

En este módulo se abordan las nociones básicas de probabilidad y sus relaciones con otros temas de matemática.

En su desarrollo, se comparan diferentes definiciones de probabilidad y se ubican históricamente diversos temas.

Se trata, también, el cálculo de la cantidad de sucesos favorables o posibles mediante la combinatoria, y la aplicación de algunas fórmulas elementales de probabilidades; asimismo, se distingue entre sucesos dependientes e independientes.

Buscamos, finalmente, relacionar los temas dados con los temas de matemática de la escuela media.

Recordando a Roberto J. Payró

Por aquellos años, en Pago Chico, solamente había dos profesores de Matemática: Don Joaquín Requejo y Don Hermenegildo Gómez. Mucho nos gustaría presentar una edición –por decirlo así, “bilingüe”– donde se reflejaron los particulares enfoques de tan preclaros profesores al encarar un mismo tema. La providencial tacañería de los editores nos ha llevado a elegir la presentación de las clases de don Joaquín, aunque aquí y allí deslicemos alguna opinión de don Hermenegildo, que bien merece ser recordado.

Pues bien, don Joaquín llegaba a su clase y comentaba con sus alumnos que la noche anterior había jugado a los dados en el Club del Progreso, y les preguntaba:

- Si tiramos un dado, ¿quién creéis que tiene más chance de ganar: El que apuesta a que salga el “as” o el que lo hace por que salga “par”? Aclaro que los dos juegan un peso fuerte cada uno y, si aciertan, ganan un peso.

Enseguida, contestaba un muchacho:

- El que apuesta a “par” tiene más chance.
- Muy bien. Pero, como yo os doy matemática, quiero enseñaros a medir esa chance.

(Aquí, observará el lector, que el lenguaje de Joaquín se pone un tanto pesado con tanto “os”; de manera que trataremos de abreviar y traducir.)

1. La probabilidad es una medida

La definición histórica de probabilidad

"Aunque vuesa merced los ve [estos naipes] astrosos y maltratados, usan de una maravillosa virtud con quien los entiende, que no alzará que no quede un as debajo..."
(Rinconete y Cortadillo. Miguel de Cervantes Saavedra.)

Históricamente, la teoría de la probabilidad se inició con el estudio de los juegos de azar.

No sabemos en qué medida se corresponde con la historia; pero, a los efectos didácticos, nos parece adecuado comenzar el estudio de las probabilidades con juegos de dados.

Sabemos que, al arrojar un dado, el conjunto E de resultados posibles es:

$$E = (1,2,3,4,5,6)$$

Laplace, en su teoría analítica de las probabilidades (1812), dio la siguiente definición de probabilidad de un suceso:

Si al llevarse a cabo un experimento (por ejemplo: tirar un dado, hacer girar a ruleta...), sobre n casos igualmente posibles hay s casos favorables al suceso 1, diremos que:

$$\text{probabilidad de } A = \frac{s}{n} = \frac{\text{n° casos favorables a A}}{\text{n° casos posibles}}$$

De esta manera, la probabilidad de que salga el 4 es $P(4) = \frac{1}{6}$

La probabilidad de que salga un número par es $P(\text{par}) = \frac{3}{6} = \frac{1}{2}$

Actividad 1.1.

En el caso ideal, proponemos indicar, con relación a un mazo de cartas españolas:

1. La probabilidad de que sea un oro.
2. La probabilidad de que sea un siete.
3. La probabilidad de que no sea una copa.
4. La probabilidad de que sea un as o un siete.
5. La probabilidad de que sea un oro o un siete.
6. La probabilidad de que sea oro, copa, espada o basto.
7. La probabilidad de que sea un quince.

Estos ejemplos les permitirán observar algunas propiedades.

La probabilidad de un suceso es un número comprendido entre cero y uno. Se pueden buscar, para el mazo de cartas españolas, sucesos con probabilidad cero.

Actividad 1.2.

La probabilidad de que se dé un suceso A o un suceso B se obtiene, a veces, sumando P(A) + P(B) y otras veces no. ¿En qué casos se obtiene sumando?

Actividad 1.3.

Se suele anotar $P(A \cup B) = P(A) + P(B)$ sólo si A y B...

Otra forma de calcular probabilidad

Historia de un alumno difícil

El alumno Timoteo Ladrillo, ingenioso inventor de trampas para pájaros, había agotado la paciencia de don Hermenegildo, que lo había expulsado de su clase. Recaló en el aula de don Joaquín, quien conoció su voz al oír la pregunta:

– ¿Qué quiere decir que, en un dado, la probabilidad de 4 es 1/6?

El maestro, que se jactaba de su habilidad para tratar a estos tipos difíciles, le respondió:

– Te vas a tu casa y echas el dado 240 veces; cuentas cuántas veces sale el 4 y mañana me das el resultado.

El bueno de Timoteo hizo como se le dijo y, al día siguiente, afirmó que el 4 había salido 44 veces, poniendo como testigo a Ramoncito López.

– Ahora, divide 44 por 240.

Ladrillo que, por obra de don Hermenegildo, era rápido para las cuentas, respondió:

– 0,18333. Y eso, ¿qué tiene que ver con 1/6?

Don Joaquín lo miró desafiante; pero, Timoteo sacó un papel no muy limpio, diciendo:

– Mire, hice más de lo que me pidió y escribí esta tabla: El 1 salió 38 veces; el 2, 42 veces; el 3, 40 veces; el 4 salió 44 veces; el 5, 39; y, el 6, 37 veces.
– Divide, hijo, como hicimos antes y obtendrás las frecuencias relativas.

Nadie pudo adelantarse al ex-alumno de don Hermenegildo, que escribió:

1	2	3	4	5	6
0,1583	0,175	0,1666	0,1833	0,1625	0,1541

– Y eso, ¿qué tiene que ver con 1/6?

El maestro no contestó; pero, un alumno se apresuró a decir:

– Todos los resultados se parecen a 1/6 = 0,166...
– Sería cosa buena echar el dado 2000 veces.

¿Ya intentó su propia definición?

Cuando se repite un experimento n veces y aparece el suceso A s veces, el número $\frac{s}{n}$ se llama **frecuencia relativa** de A.

La razón $\frac{s}{n}$ tiende a estabilizarse, cuando el número de veces n se agranda.

Quiere decir que $\frac{s}{n}$ se aproxima a un límite. La teoría prueba que ese límite es la probabilidad del suceso A.

En los casos de dados, monedas, juegos ideales, podemos hallar la probabilidad. Pero, en muchos problemas prácticos, sólo podemos estimar probabilidades recurriendo a la estadística. Por ejemplo: ¿Cuál es la probabilidad de accidentes en ruta? ¿Cuál es la probabilidad de que un medicamento produzca efectos secundarios?

Laplace

Los párrafos de la biografía de Laplace que aparecen a continuación han sido extraídos del trabajo de Luis Santaló: *Laplace* (Centro Editor de América Latina.)

> (...) Laplace (1749-1827) no fue un gran creador de teorías ni un crítico de los fundamentos. Fue un extraordinario continuador, hasta límites insospechados, de las teorías iniciadas por otros, sea la mecánica de Newton o las probabilidades de Bernoulli y De Moivre; pero, no se detiene a analizar con demasiado cuidado los fundamentos. Las bases de las que parte son, a veces, endebles, pues las toma tal como las encuentra en sus predecesores, pero con su inmensa capacidad e insuperable inventiva para artificios de cálculo, saca de las teorías el máximo provecho. Fue un virtuoso del análisis matemático.
>
> En la teoría de las probabilidades, por ejemplo, pueden criticarse a Laplace tanto la definición misma de probabilidad como el haber continuado con ideas ambiguas (la esperanza moral de Daniel Bernoulli, por ejemplo).
>
> (...) La definición de probabilidad adoptada por Laplace es la clásica: ‘Fracción cuyo numerador es el número de casos favorables y cuyo denominador es el número de todos los casos posibles’, definición que el mismo Laplace se apresura en contemplar con la condición de que ‘sean igualmente posibles los diversos casos’. Pero, ¿qué quiere decir igualmente posibles? Evidentemente, la definición no resiste un examen lógico riguroso, pero es difícil iniciarse en la teoría de las probabilidades sin tener esta definición –aún sabiéndola incorrecta– como punto de partida y faro de referencia. Por otra parte, la definición conjuntista de Kolmogoroff (1933), actualmente la más aceptada –con algunas variantes de un autor a otro–, no es más que la abstracción y axiomatización de tal definición intuitiva.

> Las teorías matemáticas nunca se han construido, al comienzo, sobre bases firmes e inmutables. A las altas cumbres se sube únicamente con equipaje ligero y saltando peligrosamente sobre inciertos puntos de apoyo; de aquí que haga falta mucha habilidad e ingenio. Más tarde, ya plantada la bandera en la elevada cresta, vendrán las legiones de pacientes trabajadores para rellenar huecos y apuntalar columnas; no faltará alguno que critique al iniciador por haberse apoyado en cimientos movedizos e inseguros, y que no perciba que, sin tal audacia, su laboriosa tarea de relleno carecería de sentido, pues todo fundamento supone el conocimiento previo de lo que se desea fundamentar: Laplace, efectivamente, no edificó la teoría de probabilidades sobre bases firmes; pero, los métodos que creó y los resultados a que llegó siguen siendo válidos y útiles, después de todas las ulteriores fundamentaciones.

Elementos de la definición axiomática

Lo que sigue es para usted y sólo para algún alumno que le plantee la inquietud.

En el ejercicio anterior, señalamos el defecto de la definición clásica. Para corregir ese defecto, habrá que proponer una definición axiomática. Al realizar un experimento, existe un conjunto de resultados. El conjunto de todos estos sucesos elementales se llama **espacio muestral**. En un dado común, el espacio muestral es E = (1, 2, 3, 4, 5, 6).

Nos interesa. por ejemplo, la probabilidad de que salga “cuatro”:

$$P(4)$$

Pero, el espacio muestral no basta.

También podemos jugar a que salga “impar”:

$$P(\text{impar}) = P(\{1,3,5\})$$

{1,3,5} es un subconjunto del espacio muestral E.

O bien podemos jugar a que salga “mayor de dos”:

$$P\ (\text{mayor que}\ \ 2) = P(\{3,4,5\})$$
$$\{3,4,5,6\} \subset E$$
$$\{1,3,5\} \subset E$$
$$\{4\} \subset E$$

La inventiva de los apostadores no tiene límite. En realidad, en el caso de una tirada del dado, tiene límite, porque los **sucesos** son todos los subconjuntos de E.

¿Cuántos son? Se podrán traducir algunos de ellos al lenguaje verbal e indicar su probabilidad, de acuerdo con la definición clásica.

La probabilidad es un número que asignamos a cada subconjunto de E. Es deseable que, conocidas las probabilidades de los resultados o sucesos elementales, podamos deducir las probabilidades de cualquier suceso.

Para ello, proponemos la siguiente definición axiomática:

P es una función de probabilidad definida en un espacio de sucesos (E) si:

1. La probabilidad de todo suceso es un número real mayor o igual que cero, y menor o igual que 1.
2. P(E)=1
3. Si A y B son subconjuntos de E tales que $A \cap B = \emptyset$, entonces $P(A \cup B) = P(A) + P(B)$

Estos axiomas nos parecen naturales... porque ya estábamos familiarizados con la noción intuitiva.

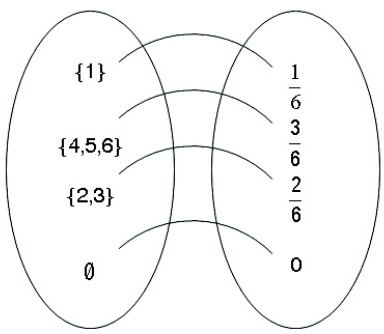
El mismo Santaló nos dice:

> "...la definición conjuntista no es más que la abstracción y axiomatización de la definición intuitiva".

No creemos necesario, en la enseñanza media, considerar este enfoque. Sin embargo, los docentes debemos conocer algo más que los alumnos.

Actividad 1.4.

En el caso de un dado ideal, podemos hacer el siguiente diagrama de la función de probabilidad:



Este diagrama no está completo porque faltan otros subconjuntos como, por ejemplo, {1,3,4,5,6} o {2,4,6} con sus respectivas probabilidades: 5/6 o 3/6. ¿Por qué deben ser éstas sus respectivas probabilidades?

¿Cuál debe ser –según este diagrama- P(5)? ¿Por qué? ¿Y P (par)?

Actividad 1.5.

Ahora, le proponemos que exprese una función de probabilidad para un extraño dado que tiene:

- dos "unos",
- dos "dos",
- un "tres",
- un "cuatro".

Damos una parte de la respuesta en la tabla:

{1}	{2}	{3}	{4}	{1,2}	{1,3}	{2,3,4}	...
1/3	1/3	1/6	1/6	2/3	1/2	2/3	...

¿Por qué decimos “una” función y no “la” función?

Actividad 1.6.

Un alumno distraído, al completar la tabla anterior, calculó así:

$$P(\{1,2,3,4\}):$$

$$\{1,2,3,4\} = \{1,2\} \cup \{2,3,4\}$$

$$\therefore P(\{1,2,3,4\}) = P(\{1,2\}) + P\{2,3,4\})$$

$$\frac{2}{3} + \frac{2}{3}$$

$$= \frac{4}{3}$$

¡No puede ser!

¿Por qué no puede ser?
¿Cuál es el error del razonamiento del alumno?

2. Cálculo combinatorio

Historia de un alumno difícil

Timoteo Ladrillo, ingenioso inventor de trampas para pájaros, había faltado, de modo que cuando volvió a clase, don Joaquín había dado vuelta la página y proponía a los alumnos este problema:

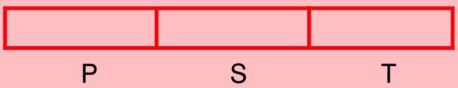
- Hay cinco candidatos para formar listas de presidente, secretario y tesorero del Club del Progreso. ¿Cuántas listas pueden formarse si nadie puede ocupar más de un cargo?
- ¿Cómo se llaman los candidatos?
- No necesitas saberlo; pero, ya te lo puedes imaginar: Siempre son los mismos.

Los muchachos se fueron agrupando, discutiendo entre sí, proponiendo ideas; algunos trabajaban silenciosamente, otros hacían bastante ruido. Había quienes se interesaban poco. Pasaron muchos minutos con el problema. Don Hermenegildo reprochaba a su colega esta pérdida de tiempo.

Uno de los alumnos dio como respuesta: 60, mostrando una larga lista: ABC, ABD ... BCA... y aclarando que habían intervenido varios en el trabajo.

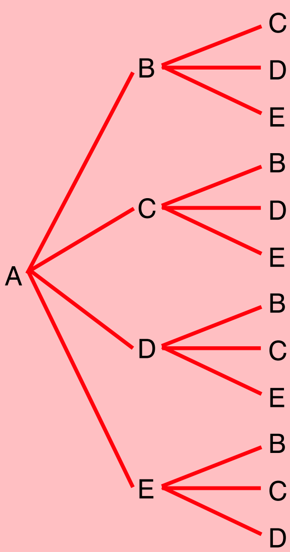
Otro de los alumnos propuso la siguiente explicación:

– Debemos llenar 3 cargos o casillas, con nombres. Para la primera casilla, disponemos de 5 personas; para la segunda casilla nos quedan 4; y, por fin, tenemos 3 para la tercera casilla. Multiplicamos 5.4.3 = 60 y ya está.



– ¿Por qué se multiplica? Es lo único que no entiendo –dijo Ladrillo.

Como había varios alumnos que no entendían, Joaquín les enseñó a construir un diagrama de árbol:



La primera rama corresponde a A presidente, B secretario, C tesorero. La última rama corresponde a A presidente, E secretario, D tesorero.

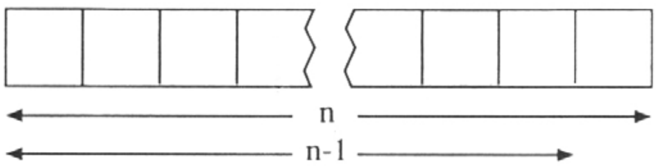
- Este diagrama nos dice que con A de presidente hay 12 listas posibles.
- ¿Qué título ponemos en el cuaderno?
- Pongamos “Variaciones”. Pero, lo importante no es el nombre sino resolver problemas.
- Y, ¿qué son “variaciones”?

Si debemos colocar m objetos en n casillas y debe distinguirse el orden, cada uno de los agrupamientos posibles se llama **variación o arreglo**.

El número de variaciones de m objetos en n casillas será:

$$\boxed{V_{m.n} = m(m-1)(m-2)...[m-(n-1)]}$$

Esta fórmula “habla”: Cuando vamos a llenar la primera casilla, tenemos m objetos disponibles; para la segunda nos quedan (m-1)... Al llegar a la última nos hemos gastado (n-1) objetos. Para la última casilla nos quedan para elegir m -(n-1).



Permutaciones

Si se trata de colocar m objetos en m casillas, sin repetición, es obvio que el número de posibilidades será:

$$V_{m.m} = m(m-1)(m-2)(m-3)...3.2.1$$

Al producto de los m primeros números naturales lo llamamos **factorial** de m y lo simbolizamos así: **m!**

Además, como si hay m objetos y m casillas, lo único que podemos hacer es permutar, decimos que se trata de permutaciones y escribimos:

$$\boxed{P_m = m!}$$

¿De cuántas maneras pueden disponerse en fila los señores A, B y C?

ABC ACB BAC BCA CAB CBA $P_3 = 3! = 3.2.1 = 6$

Las calculadoras de bolsillo suelen tener una tecla dedicada a n!

Combinaciones

Don Joaquín propuso el siguiente problema: Tenemos 5 jugadores de pato. ¿Cuántos equipos distintos de tres jugadores podemos formar?

- 5.4.3 = 60 –saltó el negro Liborio.
- No por mucho madrugar...

Como de costumbre, en la clase de don Joaquín, la mayoría se puso a trabajar hasta que alguno dijo:

- Don Joaquín, ¿el equipo JKL es el mismo que el LJK? Sí... Entonces ya lo tengo. Quedan 10: JKL, JKM, JKN, JLN, JMN, KLM, KLN, KMN, LMN.
- ¿Por qué dices “quedan” 10?
- Digo que, de todas las variaciones, quedan 10 porque no importa el orden (un cambio de orden no me da otro equipo).
- Cuando se tienen m objetos diferentes para llenar n casillas sin que interese el orden, cada agrupamiento se llama **combinación**. ¿Cómo haríais para calcular las combinaciones de 5 objetos en 3 casillas sin escribir JKL,...etc?

El domingo siguiente, don Hermenegildo no podía creer lo que afirmaba don Joaquín: Un alumno había descubierto la fórmula. Después de un rato, había encontrado que:

$$C_{5,3} = \frac{V_{5,3}}{P_3}$$

La fórmula general es:

$$\boxed{C_{m,n} = \frac{V_{m,n}}{P_n}}$$

Don Joaquín exclamó entusiasmado: Esta fórmula “habla”. Si tienes las variaciones y debes calcular las combinaciones, como no te interesa el orden, debes dividir por las permutaciones del número de casillas.

En la mayoría de las aplicaciones, el orden no interesa; por ejemplo, al elegir enfermos para un tratamiento, al elegir alumnos para un control, al elegir personas para una encuesta.

“Las fórmulas sólo deben ser una guía. Cualquiera puede inventar sus propias fórmulas para resolver los problemas. Hay que leer bien los enunciados. El camino del infierno del conocimiento está empedrado de memoristas.” (Don Joaquín Requejo, pedagogo pagochiquense)

Actividad 2.1.

1. ¿De cuántas maneras pueden repartirse 3 premios diferentes entre 10 personas que, a lo sumo, pueden recibir un premio cada una?
2. ¿Cuántos números de cuatro dígitos distintos pueden escribirse con las cifras de 1 a 9?
3. ¿Cuántos números de cuatro dígitos pueden escribirse con las cifras de 1 a 9?

4. Disponemos de los dígitos 1, 2 ,3, 4, 5, 6, 7, 8. ¿De cuántas maneras podemos alinearlos de a cuatro, si deben alternarse números par impares, empezando por pares?
5. Disponemos de los mismos dígitos del problema anterior. ¿De cuántas maneras podemos alinearlos a todos, si deben alternarse pares e impares, o viceversa?
6. Tenemos tres libros distintos de química y cuatro libros distintos de biología. Se pide determinar: a) ¿De cuántas maneras podemos alinearlos en un estante? b) ¿De cuántas maneras podemos alinearlos, si deben estar agrupados por materias?
7. Alí Babá debía elegir 5 hombres (todos igualmente valientes) para una misión peligrosa. ¿Cuántas eran las elecciones posibles?
8. a) ¿De cuántas maneras se pueden permutar las letras de la palabra "combinar"?
   b) ¿De cuántas maneras se pueden permutar las letras de la palabra "caminar"?
   c) ¿De cuántas maneras se pueden permutar las letras de la palabra "ananá"?

Números combinatorios

Cuando escribimos $\binom{m}{n}$ entenderemos que se trata del número de combinaciones de m objetos en n casillas. siendo $n \leq m$.

Sabemos que:

$$\binom{m}{n} = C_{m.n} = \frac{V_{m.n}}{P_n} = \frac{m(m-1)(m-2)...(m-n+1)}{n!}$$

Si multiplicamos numerador y denominador por (m-n)!, obtenemos:

$$C_{m.n} = \frac{m!}{n!(m-n)!}$$

Pero, entonces, los números combinatorios $\binom{m}{n}$ y $\binom{m}{m-n}$ son iguales.

Por ejemplo: $\binom{47}{42} = \binom{47}{5}$

Esta propiedad tiene un significado muy claro: Cada vez que elegimos 42 personas entre 47, elegimos "no elegir" 5 personas entre 47 (Ideal como destrabalengua).

Admitimos, por definición, que 0! = 1 y que 1! = 1

Actividad 2.2.

Calcule $\binom{m}{0}$ y $\binom{m}{1}$

No nos parece conveniente enseñar estas fórmulas del tipo $C_{m.n} = \dfrac{m!}{n!(m-n)!}$ en un curso elemental. Pueden perturbar la solución de problemas. En general tendremos:

$(a+b)^n = a^n + a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \ldots b^n$

(desarrollo de Newton)

En cuanto don Joaquín consideró que los alumnos sabían el ABC de la combinatoria, les propuso diversos problemas, pidiéndoles que, para calcular la probabilidad del suceso considerado, emplearan la fórmula:

$$\text{Probabilidad} = \frac{\text{N° casos favorables}}{\text{N° casos posibles}}$$

Entre 10 personas (entre las cuales están Timoteo y Liborio) hay que elegir 5 personas al azar. ¿Cuál es la probabilidad de que, en la elección hecha, figuren simultáneamente Timoteo y Liborio?

El número total de casos es $C_{10,5}$. El número de casos en que figuran T y L se obtiene así: Colocamos obligatoriamente a T y L, quedándonos 8 personas para 3 lugares.

Por lo tanto: $P = \dfrac{\binom{8}{3}}{\binom{10}{5}} = \dfrac{2}{9}$

Este resultado puede interpretarse así: Si pusiéramos, en una caja, 10 nombres entre los cuales estuvieran T y L, y extrajésemos 5 papelitos repitiendo esta operación muchas veces, en alrededor de 2/9 de las veces aparecerían en el grupo T y L juntos.

Actividad 2.3.

En una oficina trabajan 12 hombres y 8 mujeres. Hay que elegir 5 personas al azar, entre esas 20. ¿Cuál es la probabilidad de que, en el grupo elegido, figuren exactamente 3 mujeres?

En la caja N° 1 hay 3 bolitas verdes y 2 blancas; en la N° 2 hay 4 verdes y 1 blanca; en la N° 3 hay 2 verdes y 2 blancas. Se saca una bolita de cada caja. ¿Cuál es la probabilidad de que, al finalizar la operación tengamos 2 blancas y 1 verde?

El resultado “dos blancas y una verde”, puede obtenerse extrayendo, por ejemplo, blanca de la caja N°1, blanca de la N°2 y verde de la N°3.

En consecuencia, hay tres sucesos favorables: BBV, BVB, VBB.

La probabilidad de BBV es $\dfrac{2.1.2}{5.5.4} = 1/25$

La probabilidad de BVB es $\frac{2.4.2}{5.5.4} = 4/25$

La probabilidad de VBB es $\frac{3.1.2}{5.5.4} = 3/50$

La probabilidad de obtener dos blancas y una verde al finalizar la operación es

$$\frac{1}{25} + \frac{4}{25} + \frac{3}{50} = \frac{13}{50}$$

3. Sucesos excluyentes y no excluyentes

En la primera sección se ha resuelto, con relación a un mazo de cartas españolas:

P (as o siete) y P (oro o siete)

No pudimos resolverlos de la misma manera: En el primer caso, se trataba de sucesos excluyentes y, en el segundo, de sucesos no excluyentes.

Hasta ahora hemos empleado una fórmula para calcular P(A o B), cuando A y B se excluyen.

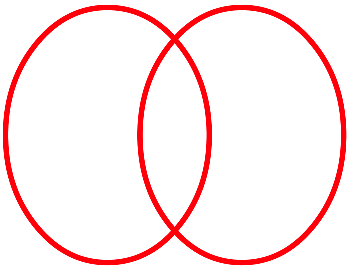
Para el caso de que A y B no sean excluyentes, podemos escribir:

$$A \cup B = A \cup (B - A)$$

Recordemos que

Como A y B-A son excluyentes:

$$P(A \cup B) = P(A) + P(B \cap \overline{A}) \qquad (1)$$



$$A \cup B = A \cup (B - A)$$
$$B = (B \cap \overline{A}) \cup (A \cap B)$$

Debemos tratar que aparezca P(B) y eliminar $P(B \cap \overline{A})$

$$B = (B \cap \overline{A}) \cup (A \cap B)$$
$$P(B) = P(B \cap \overline{A}) + P(A \cap B) \qquad (2)$$

A la igualdad (1) le restamos la (2) y obtenemos, después de operar:

$$\boxed{P = (A \cup B) = P(A) + P(B) - P(B \cap A)}$$

Por ejemplo, sean A={1,2,3], B={3,4,5,6} sucesos que corresponden a la tirada de un dado.

$$P = (A \cup B) = P(A) + P(B) - P(B \cap A) = 3/6 + 4/6 - 1/6$$

Justifique el resultado obtenido en 1.1. ítem 5:

$$P(\text{oro o siete}) = \frac{13}{40}$$

Esta fórmula es también válida para el caso particular de A y B excluyentes. ¿Por qué?

De 10 personas (entre las que están A y B), debemos elegir una comisión de 4. Si la elección se realiza al azar, ¿cuál es la probabilidad de que, en la comisión elegida, esté al menos uno de los nombrados?

Llamemos A al suceso “está A” y B al suceso “está B”.

$$P(A) = \binom{9}{3} : \binom{10}{4} = 0{,}4 \qquad P(B) = \binom{9}{3} : \binom{10}{4} = 0{,}4$$

Los sucesos A y B no son excluyentes o incompatibles porque hay comisiones donde están A y B. ¿En cuántas comisiones están A y B? En $\binom{8}{2}$ comisiones. ¿Por qué?

Luego:

$$P(A \cap B) = \binom{8}{2} : \binom{10}{4} = 0{,}133$$ (probabilidad de que estén A y B)

Por lo tanto:

$$P = (A \cup B) = P(A) + P(B) - P(A \cap B) = 0{,}4 + 0{,}4 - 0{,}133 = 0{,}667$$

El problema puede resolverse directamente, de la siguiente manera: ¿En cuántas comisiones no estarán ni A ni B? En $\binom{8}{4}$ comisiones. ¿Por qué?

Entonces, la posibilidad de que no estén ni A ni B es:

$$\binom{8}{4} : \binom{10}{4} = 0{,}333$$

Por lo tanto, la probabilidad de que estén al menos A o B es 1 - 0,333 = 0,667.

Es conveniente alentar la presentación de distintas soluciones para un mismo problema. Hacemos notar que es más fácil hallar la probabilidad de que no estén A ni B que calcular los casos en que están A o B.

Si **p** es la probabilidad de un suceso A, la probabilidad del suceso contrario es **1 - p = q**

Por ejemplo:

p = 1/6 es la probabilidad de que salga el “seis” en el dado.
1-p = q = 1-1/6 = 5/6 es la probabilidad de que “no salga el seis”.

Un día de lluvia

El agua lavaba las ventanas de la vieja escuela. No llegaban a cinco los alumnos presentes. Don Joaquín les propuso este problema:

– Si ponemos en una caja las letras de la palabra "víbora" y sacamos sucesivamente cuatro letras sin reponer, ¿cuál es la probabilidad de que salga la palabra "rabo"?

No tardaron mucho en contestar: $1/V_{6,4} = 1/360$

– Y, ¿cuál es la probabilidad de que salgan las letras que palabra "rabo"?

$$V_{4,4}/V_{6,4} = P_4/V_{6,4} = 24/360$$

Liborio no estuvo de acuerdo. Su respuesta fue: 0. Don Joaquín se acercó al banco y vio escrita la palabra "vívora".

Pero, ni corto ni perezoso, propuso el siguiente problema: ¿Cuántas permutaciones diferentes se pueden hacer con las letras de la palabra "caminata"?

– Si las letras fuesen distintas, sería muy fácil 8! = 43320 -dijo un muchacho-. Pero, deben ser muchas menos porque...

Después de muchas deliberaciones y pruebas entre los presentes, hallaron que las permutaciones de "caminata" son: $\frac{8!}{3!} = \frac{40320}{6} = 6720$

Actividad 3.1.

1. ¿De cuántas maneras se pueden permutar una bandera roja, una azul, una blanca, una amarilla, una verde y una celeste?
2. ¿De cuántas maneras se pueden permutar tres banderas rojas exactamente iguales, una azul, una blanca y una amarilla?
3. ¿De cuántas maneras se pueden permutar dos banderas rojas exactamente iguales, dos azules exactamente iguales, una blanca y una amarilla?

Probabilidades geométricas

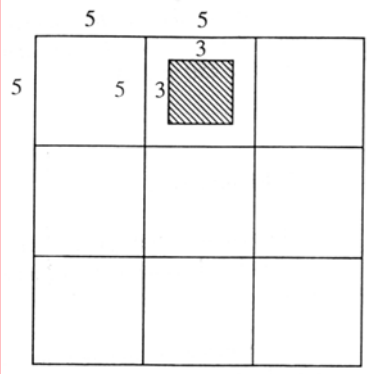
En algunas situaciones, los casos favorables y los posibles son infinitos. No se puede hablar, entonces, de

$$\frac{\text{N° de casos favorables}}{\text{N° de casos posibles}}$$

La siguiente historia nos muestra cómo "se las arreglan" los matemáticos.

Día de cuadreras en Pago Chico

El pueblo estaba alborotado porque era domingo y había cuadreras. Don Joaquín encontró a su alumno Timoteo rodeado de gente. Frunció la nariz y se acercó; observó que había colocado, en el suelo, un cartón en el cual había dibujado –con más prolijidad que de costumbre– un cuadriculado. Proponía que tirasen una moneda de 1 centavo sobre el cuadriculado. Si no cortaba a ninguna línea, les daba 2 centavos; de lo contrario, se quedaba con el centavo. El maestro calculó que Timoteo hacía una buena ganancia. ¡Para eso servían sus enseñanzas! Pero, ya tendría que rendir cuentas; por lo pronto, al día siguiente le comunicó que debía calcular la probabilidad de ganancia de un jugador en ese tablero, so pena de recibir algunos azotes (aclaramos que, para ese entonces, no se habían abolido en Pago Chico los castigos corporales).



Con gran temor, trató el muchacho de responder a la pregunta; pero, fue comprendiendo que de nada le servirían permutaciones, combinaciones y demás yerbas.

El negro Líborio acudió en su ayuda y, al día siguiente, respondió:

- Un apostador tiene probabilidad 0,35 de ganar.
- ¿Cómo lo calcularon?
- Tiramos 100 veces la moneda...
- Yo... que me había hecho la ilusión de que recordabais un poco de geometría...

Esta observación molestó a los muchachos que, luego de la clase, se reunieron a deliberar. Un tercero, de los que nunca faltan, sugirió:

- Para que una recta toque a una circunferencia, el centro debe encontrarse a una distancia menor o igual que la longitud del radio.
- Vemos que cada cuadrado tiene 5 cm. Como la moneda tiene 1 cm de radio, si el centro cae dentro del cuadrado sombreado, gana el jugador. En consecuencia, la probabilidad de ganancia del jugador es:

$$P = \frac{\text{área del cuadrado chico}}{\text{área del cuadrado grande}} = 9/25 = 0{,}36$$

- Al final resultó que la geometría sirve para algo

Don Joaquín, que oía satisfecho estas reflexiones, casi se desmaya al oír el final y agregó:

- No es un juego equitativo, si el premio es igual a la apuesta.

Actividad 3.2.

1. Si en el interior de un segmento se selecciona un punto al azar, ¿cuál es la probabilidad de que dicho punto esté más cerca del punto medio del segmento que de sus extremos?
2. Si una línea telefónica entre dos puntos A y B distantes 2 km, se corta en un punto desconocido, ¿cuál es la probabilidad de que el lugar afectado no se halle a más de 450 m de A?

El número π

Los alumnos trabajan con π desde la escuela primaria; se puede aprovechar la probabilidad geométrica para hacer una aproximación experimental del número π.

Proponga a los alumnos que construyan un cuadrado de 10 cm de lado y un círculo inscripto en el mismo. Si se dejan caer porotos o lentejas sobre el dibujo, puede esperarse aproximadamente la siguiente proporción:

$$\frac{\text{N° de lentejas dentro del círculo}}{\text{N° de lentejas en el cuadrado}} = \frac{\text{Área del círculo}}{\text{Área del cuadrado}}$$

La razón aproximada será: $\dfrac{\pi\ .\ 25}{100} = \dfrac{\pi}{4}$

Inecuaciones

Para este tema se pueden dar ejercicios de probabilidad geométrica como el siguiente:

> Para ir a la escuela tengo que tomar dos ómnibus. La espera máxima para el primero es de 10 minutos y, para el segundo, de 8 minutos. Sabiendo que el tiempo de viaje efectivo es de 20 minutos, ¿qué probabilidad tengo de llegar a la hora sí salgo de casa 30 minutos antes?

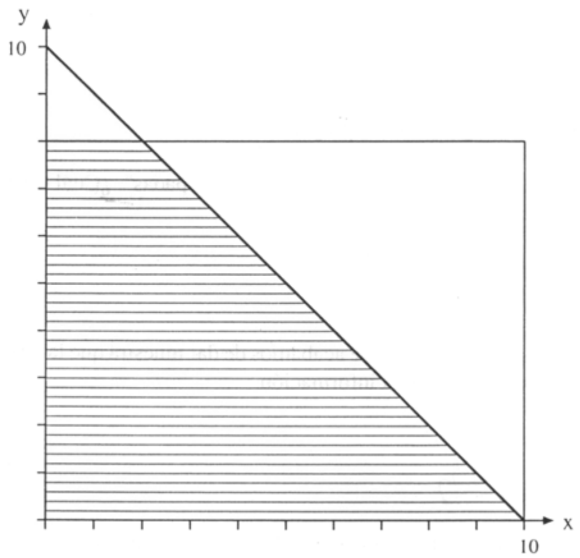
Llamando x al tiempo de espera para el primer ómnibus e y al tiempo de espera del segundo, tenemos las condiciones $x \leq 10,\ \ y \leq 8.$ Los casos posibles son, por tanto, todos los puntos del área del rectangulo de lados 10 y 8, o sea, 80.

Los casos favorables corresponden a los puntos del rectángulo para los cuales es $x + y \leq 10$, los cuales son los que constituyen el área rayada en la figura:

$$\frac{(2+10)\ .\ 8}{2} = 48$$

Por lo tanto, la probabilidad buscada es:

p = 48/80 = 0,6



4. Sucesos dependientes e independientes

Probabilidad condicional

Suponemos que una persona que no vemos, nos dice:

- Tengo un mazo de 40 cartas españolas. Voy a extraer una al azar. ¿Cuál es la probabilidad de que sea el as de espadas?
- 1/40 –contestaríamos-.

Si, luego, la persona nos pregunta:

- Ya he sacado la carta y resultó ser “espadas”. ¿Cuál es la probabilidad de que sea el as?
- 1/10.

Un ejemplo como el que acabamos de dar, muestra que la asignación de probabilidades varía según la información.

Llamaremos probabilidad condicional del suceso B con respecto a A (o probabilidad de que se dé B, habiéndose dado A) al número:

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \qquad (1)$$

Si pensamos en frecuencias relativas, resulta:

$$P(B/A) = \frac{N^\circ \text{ de elementos de } A \cap B}{N^\circ \text{ de elementos de } A}$$

Dividiendo numerador y denominador por el número de elementos de E, resulta:

$$P(B/A) = \frac{P(B \cap A)}{P(A)}$$

¿Por qué?

¿Esta fórmula cumple las condiciones de probabilidad? Veamos:

$P(A/A) = 1$ ¿Lo verificó?

Además, debe verificarse que si $A_i \cap A_j = \varnothing$

Entonces:

$$P(A_i \cup A_j / B) = P(A_i / B) + P(A_j / B)$$

Trate de probarlos.

Veamos, por fin, si la definición "funciona" en el ejemplo anterior.

Sea A: "Sale espadas". B: "Sale as".

$P(A) = 1/4$ $\quad\quad\quad$ $P(A \cap B) = 1/40$

¿Por qué?

Luego:

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{1/40}{1/4} = 1/10$$

que es la respuesta que dimos al comenzar el capítulo.

Quiere decir que, si tenemos la información de que salió "espadas", la probabilidad de "as de espadas" es 1/10.

> En la caja N° 1 hay dos bolitas blancas y 8 verdes. En la caja N° 2 hay 4 blancas y 6 verdes. Una persona elige al azar (por ejemplo, tirando una moneda) una de las cajas y extrae de ella una bolita que resulta ser blanca. Nos comunica el resultado preguntándonos de qué caja salió la bolita. ¿Qué probabilidad de acertar tenemos, si decimos "pertenece a la primera caja"?

$A_1 = \text{Caja N°1}$

$P(A_1) = 0{,}5$

$A_2 = \text{Caja N°2}$

$P(A_2) = 0{,}5$

$B = \text{Blanca}$

$V = \text{Verde}$

Sabemos que salió blanca, de manera que hay que determinar:

$$P(A_1/B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{2/20}{6/20} = 1/3$$ [1]

Como verificación, calculemos:

$$P(A_2/B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{4/20}{6/20} = 2/3$$

[1] *Para calcular la probabilidad de "blanca" hemos tenido en cuenta que hay 6 blancas sobre un total de 20; para calcular* $P(A_1 \cap B)$ *consideramos que sobre 20 bolitas había sólo 2 que fuesen, al mismo tiempo, de la primera caja y "blanca".*

Obsérvese que, una vez que sabemos que salió blanca, o proviene de la N° 1 o de la N° 2, no hay más casos; por eso:

$$P(A_1/B) + P(A_2/B) = 1$$

En una bolsa hay 3 papeles rojos y 7 azules. ¿Cuál es la probabilidad de que al extraer dos papeles uno tras otro al azar sin reposición[2], el primero sea rojo y el segundo azul?

¡Recuerde! Intente resolverlo usted, antes de leer nuestra propuesta. Si no se anima, le damos un empujoncito.

Llamemos R_1 al suceso "rojo en primer lugar" y A_2 al suceso "azul en segundo lugar". Tenemos que calcular $P(R_1 \cap A_2)$.

¿Ya lo hizo? Quizás, haya despejado en la fórmula:

$$P(A_2 / R_1) = \frac{P(R_1 \cap A_2)}{P(R_1)}$$

y haya obtenido:

$$P(R_1 \cap A_2) = P(R_1).P(A_2 / R_1)$$

La probabilidad de rojo en primer lugar es 3/10. La probabilidad $P(A_2 / R_1)$ es 7/9 pues, habiendo salido el rojo, quedan 9 papeles de los cuales 7 son azules. En consecuencia, la respuesta será 3/10 . 3/10 . 7/9 = 7/30.

Si nos pidiesen la probabilidad de sacar un papel rojo y uno azul, sin especificar el lugar, tendríamos:

$$\underbrace{3/10 \,.\, 7/9}_{\text{rojo-azul}} + \underbrace{7/10 \,.\, 3/9}_{\text{azul-rojo}} = 14/30$$

La misma bolsa del problema anterior (como en las obras de teatro). Si extraemos un papel, lo reponemos y luego extraemos otro, ¿cuál es la probabilidad del par rojo-azul? ¿Cuál es la probabilidad de obtener un papel azul y otro rojo, si extraemos dos papeles juntos?

0,3 . (0,7) = 0,21
3/10 . 7/9 + 7/10 . 3/9 = 7/15

[2] *Sin reposición: Después de cada extracción, el papel no se repone en la bolsa.*

El teorema de Bayes

Tres máquinas producen, respectivamente 40%, 35% y 25% del total de artículos en una fábrica, con porcentajes de defectos de producción de 3%, 4% y 5%, cada una. ¿Cuál es la probabilidad de que, si se selecciona un artículo al azar, éste sea defectuoso? Si el artículo seleccionado resultó defectuoso, ¿cuál es la probabilidad de que haya sido producido por la primera máquina?

Examinemos las características del problema.

- Hay un **suceso** A: Un artículo seleccionado al azar resulta defectuoso.
- Este suceso puede darse solamente **acompañado del suceso** B_1, B_2 ... B_n: Producido por la primera máquina o la segunda o la tercera.
- Estas alternativas son **excluyente**s: El artículo fue producido por una de las máquinas. A se da con B_1 o (excluyente) con B_2 o...
- Además, B_1 ... B_n completan el espacio: 40% + 35% + 25% = 100%.

Para resolver problemas con estas características, nos será útil la definición de probabilidad condicional.

Hagamos el planteo general y, por cuestiones de simplicidad, trabajemos con dos sucesos B_1 y B_2.

Podemos escribir, entonces:

$$A = (A \cap B_1) \cup (A \cap B_2)$$
$$P(A) = (A \cap B_1) + (A \cap B_2) \qquad (1)$$

¿Por qué?

Como $P(A / B_i) = \dfrac{P(A / B_i)}{P(B_i)}$ resulta $P(A / B_i).P(B_i) = P(A \cap B_i)$

Sustituimos en (1) y obtenemos:

$$P(A) = P(A / B_1).P(B_1) + P(A / B_2).P(B_2)$$

Por otra parte:

$$P(B_1 / A) = \frac{P(A \cap B_1)}{P(A)}$$

Si hacemos las sustituciones adecuadas:

$$\boxed{P(B_1 / A) = \frac{P(A / B_1).P(B_1)}{P(A / B_1).P(B_1) + P(A / B_2).P(B_2)}}$$

que es la fórmula de Bayes.

Análogamente, se calcula $P(B_2 / A)$.

La fórmula puede generalizarse si A está acompañado de B_1, B_2 ... B_n. Tendremos, entonces:

$$P(B_i / A) = [P(A / B_i).P(B_i)] : \left(\sum_{1}^{n} P(A / B_h).P(B_h) \right)$$

En la caja N° 1 hay 2 bolitas blancas y 8 azules. En la caja N° 2 hay 6 blancas y 4 azules. En otra parte, se tienen dos papeles con la inscripción Caja N° 1 y un papel con la inscripción Caja N° 2. Una persona extrae un papel; luego, de la caja cuyo número indica el papel, extrae una bolita. Nos informa que la bolita es blanca. ¿Cuál es la probabilidad de acertar si apostamos a que la bolita salió de la caja N° 1?

Examine las características del problema. ¿Cuál es ahora el suceso que, en general, llamamos A? ¿Cuáles son los que hemos llamado B_1 y B_2?

Llamamos C_1: Extraer un papel que indica Caja 1.
Llamamos C_2: Extraer un papel que indica Caja 2.

$$P(C_1) = \frac{2}{3} \qquad P(C_2) = \frac{1}{3}$$

B: Extraer una bolita blanca.
A: Extraer una bolita azul.

$$P(B / C_1) = \frac{2}{10} \qquad P(B / C_2) = \frac{6}{10}$$

Tenemos que calcular $P(C_1 / B)$:

$$P(C_1 / B) = \frac{2/10 \ . \ 2/3}{2/10 \ . \ 2/3 + 6/10 \ . \ 1/3} = 2/5$$

Como prueba, hagamos:

$$P(C_2 / B) = \frac{6/10 \ . \ 1/3}{2/10 \ . \ 2/3 + 6/10 \ . \ 1/3} = 3/5$$

Obsérvese que, una vez que sabemos que salió blanca, la bolita debe ser de la Caja N°1 o de la Caja N°2 (no hay otra alternativa; por lo tanto, la suma de probabilidades debe dar 1).

Pedimos calcular P (C_1/A) y P (C_2/A).

Actividad 4.1.

Resuelva el problema de las máquinas.

Sucesos independientes

- Voy a comprar un número para esa rifa. Elijo el 025, en una rifa de 1000 números.
- ¡No! ¡No elijas ése! Es justo el que salió en la jugada anterior. Tenés poca probabilidad de que salga de nuevo.

Hay dos puntos de vista para enfrentar la decisión de jugar o no al número que ya salió.

- El que haya salido o no el 025, no influye para su probabilidad de salir en el nuevo sorteo: se trata de sucesos independientes.

$$P\ (025) = 0{,}001$$

- La persona que piensa que es más difícil, se coloca al principio de la cadena, cuando todavía no salió 025 por primera vez y, entonces, tiene:

$$P(025 \text{ y } 025) = (0{,}001)^2$$

Diremos que A es independiente de B, si y sólo si:

$$P(B/A) = P(B)$$

Presentemos otro ejemplo, más evidente, de sucesos independientes:

Si preguntamos a nuestros alumnos:
- Al tirar un buen dado, ¿cuál es la posibilidad de que salga “as”?
- 1/6 -contestaron.
- Ahora, tiro una moneda y sale “cara”. Después tiro el dado. ¿Cuál es la probabilidad de que salga “as”?
- ¿Qué tiene que ver? –protestaron algunos, indignados– sigue siendo 1/6.

Quiere decir que P(B/A) = P(B), lo que es razonable porque “A no influye sobre B”.

$$P(\text{as}) = \frac{1}{6} = P(\text{as} / \text{cara})$$

La independencia de sucesos es simétrica; quiere decir que si A es independiente de B, entonces B es independiente de A. ¿Podrá, el lector, probarlo? Si los sucesos son independientes, se verifica que:

$$P(A \cap B) = P(A).P(B)$$

¿Por qué?

Esta fórmula sólo puede emplearse si los sucesos son independientes.

La caja N°1 contiene 5 bolitas verdes y 4 rojas. La caja N°2 contiene 4 verdes, 2 rojas y 6 negras. Sacamos una bolita de la caja N°1 y, luego, una de la N°2. ¿Cuál es la probabilidad de que la primera sea verde y la segunda roja?

A: “Sacar verde de la primera caja”.
B: “Sacar roja de la segunda caja”. A y B son independientes. Por lo tanto:

$$P(A \cap B) = P(A).P(B) = (5/9)(2/12) = 5/54$$

Actividad 4.2.

1. Con los mismos datos del problema anterior, si usted extrae una bolita de la caja N° 2 y, luego, otra de la caja N° 1, ¿cuál es la posibilidad de obtener el par ordenado verde-roja?
2. De la caja N° 1 (5 verdes, 4 rojas) extraemos una bolita x. La reponemos. Luego, extraemos otra bolita. ¿Cuál es la probabilidad del par ordenado verde-roja?
   Sea: A: Verde en la primera extracción.
   B: Roja en la segunda extracción.
   La reposición de la primera bolita hace que A y B sean independientes. Encomendamos el cálculo al lector.
3. De la caja N° 1 extraemos una bolita que no reponemos; de la misma caja extraemos una segunda bolita. ¿Cuál es la posibilidad del par ordenado verde-roja?
   Aquí, los sucesos no son independientes. La solución es: (5/9)(4/8) = 5/18
4. En la misma situación del último problema, ¿cuál es la probabilidad del par roja-verde?

Conocidas las fórmulas de probabilidad compuesta, muchos problemas de la primera parte pueden encararse sin emplear el cálculo combinatorio.

Si disponemos de 4 hombres y 5 mujeres para elegir una comisión de 3, ¿cuál es la probabilidad de que la comisión esté formada sólo por mujeres?

Deben llenarse 3 lugares; la probabilidad de “mujer” en primer lugar es 5/9; la probabilidad de “mujer“ en segundo lugar es 4/8; en tercer lugar, es 3/7.

La respuesta será:

$$\frac{5}{9} \cdot \frac{4}{8} \cdot \frac{3}{7} = \frac{\binom{5}{3}}{\binom{9}{3}}$$

Con referencia al problema anterior, ¿cuál es la probabilidad de que sean elegidos 2 hombres y 1 mujer?

$$p = \frac{4}{9}\cdot\frac{3}{8}\cdot\frac{5}{7} + \frac{4}{9}\cdot\frac{5}{8}\cdot\frac{3}{7} + \frac{5}{9}\cdot\frac{4}{8}\cdot\frac{3}{7}$$

$$p = \frac{\binom{4}{2}\binom{3}{1}}{\binom{9}{3}}$$

Hemos observado, en la práctica, que muchos alumnos prefieren esta forma y la usan con mayor éxito que las fórmulas combinatorias.

La independencia suele presentar problemas formidables en las investigaciones. El investigador puede creer que está trabajando con sucesos independientes, cuando entre ellos existe una relación oculta. (Por ejemplo: Sabemos que la probabilidad de un producto defectuoso, para cierta máquina, es de 0,05. Nos preguntamos cuál es la probabilidad de 2 defectuosos en una muestra de 10. Si la aparición del defectuoso no influye sobre la del otro, los sucesos son independientes; pero, puede suceder que un defectuoso perjudicara a la máquina de modo que aumentara la probabilidad de un nuevo defectuoso, en cuyo caso los sucesos no serían independientes.)

Una parte importante de la estadística trata de descubrir dependencias más o menos escondidas: ¿Será independiente la condición "fumador" y la de enfermo de pulmón? Las dificultades para el aprendizaje, ¿son independientes del grupo social?

Apuntes didácticos de Don Joaquín

"Es deber del docente no dedicar sus esfuerzos a que los alumnos resuelvan con absoluta seguridad algunos problemas como si preparásemos animalitos de circo. Nuestro deber, diría yo, es más bien plantearles interrogantes, llevándolos poco a poco a transitar laberintos más complicados e interesantes. El placer está en el riesgo y la aventura, no en la seguridad. Un problema resuelto ya no es un problema. Pero, entiéndase bien, el problema debe también estar al alcance de los alumnos: Si esta condición no se da, si es inalcanzable, tampoco es un problema.

A veces, bastará cambiar la redacción, el modo de preguntar, para acostumbrarlos a un esfuerzo que bien vale la pena.

También, a los futuros docentes pagochiquenses que han de sucederme a mí (y también al bueno de don Hermenegildo a quien no puedo convencer), les pido que alienten la solución de un mismo problema por distintos caminos. Por favor, no nos atravesemos en medio; no seamos un obstáculo para que los ideales de civilización y progreso cultural aniden en Pago Chico.

Suponga usted que A, B y C llegan a resultados coincidentes por el mismo camino. Es posible que sean correctos; pero, debemos revisarlos. Si A, B y C llegan a resultados iguales por caminos distintos, la posibilidad de acierto es mayor que en el primer caso. De todos modos, encomendaremos a algunos alumnos la revisión de métodos y cálculos. ¿Y el maestro? El maestro no debe intervenir demasiado, sino poco y bien. Puede que A, B y C lleguen a resultados distintos. Seguramente, no todos los resultados son correctos; pero, puede que todos estén mal. Aprovechad esto para dar abundante trabajo a los tribunales inferiores: Aprenderán mucho. Aspiro a que, al fin, la Justicia resplandezca."[1]

---

[1] *A pesar de su estilo barroco, no dejamos de coincidir con las observaciones de don Joaquín (Aunque, ya conoce el lector nuestra debilidad por él).*

Hagamos los problemas dados, de otra forma...

La caja N° 1: 5 verdes y 4 rojas; la caja N° 2: 4 verdes, 2 rojas y 6 negras. Sacamos una bolita de la primera caja y, luego, una de la segunda. ¿Cuál es la probabilidad de que se dé el par ordenado (verde, roja)?

Calculemos el número total de casos: 108 (cada bolita de la primera caja con cada bolita de la segunda; se trata del cardinal de un producto cartesiano).

Calculemos el número de casos favorables: 10 (cada verde de la primera caja con cada roja de la segunda).

En consecuencia, P (verde, roja) = 10/108.
Coteje el lector con el resultado que ya hemos hallado.

De la caja N° 1 extraemos una bolita que no reponemos; de la misma caja, extraemos una segunda bolita. ¿Cuál es la probabilidad del par ordenado (verde, roja)?

El número total de casos es 72 (cada una de las nueve bolitas de la caja puede formar par con cada una de las ocho que restan, después de sacar la primera bolita).

El número de casos favorables es 20 (cada una de las 5 verdes de la caja N° 1 puede formar par con cada una de las 4 rojas).

P (verde. roja) = 20/72 = 5/18 (véase el resultado que obtuvimos anteriormente).

Utilidad de los diagramas de árbol

Tenemos una caja N°1 con 7 bolitas verdes y 3 blancas; una caja N°2 con 4 verdes y 6 blancas. Extraemos al azar dos bolitas de la caja N°1 y las colocamos en la N°2. Luego, de esta última caja, sacamos una bolita. ¿Cuál es la probabilidad de que esta bolita sea verde?

Parece difícil porque no hicimos ninguno similar. ¿Por qué no cierra el módulo e intenta hacerlo?

¿Qué puede suceder?

$$\text{verde} - \text{verde} \quad \left(\frac{7}{10}\cdot\frac{6}{9}\right) \to (2) \to \text{verde} \left(\frac{6}{12}\right)$$

$$\text{verde} - \text{blanca} \quad \left(\frac{7}{10}\cdot\frac{3}{9}\right) \to (2) \to \text{verde} \left(\frac{5}{12}\right)$$

$$\text{blanca} - \text{verde} \quad \left(\frac{3}{10}\cdot\frac{7}{9}\right) \to (2) \to \text{verde} \left(\frac{5}{12}\right)$$

$$\text{blanca} - \text{blanca} \quad \left(\frac{3}{10}\cdot\frac{2}{9}\right) \to (2) \to \text{verde} \left(\frac{4}{12}\right)$$

Las ramas corresponden a sucesos excluyentes; por lo tanto, la probabilidad de verde en la caja N° 2 se obtiene sumando las probabilidades de cada rama del árbol:

$$\frac{7}{10}\cdot\frac{6}{9}\cdot\frac{6}{12}+\frac{7}{10}\cdot\frac{3}{9}\cdot\frac{5}{12}+\frac{3}{10}\cdot\frac{7}{9}\cdot\frac{5}{12}+\frac{3}{10}\cdot\frac{2}{9}\cdot\frac{5}{12}$$

Todos entendemos que si de la primera caja sacamos una bolita y, luego, otra sin reposición, la probabilidad del suceso verde-verde es (7/10).(6.9). En cambio, cuando extraemos dos bolitas juntas dudamos (¿por qué no hacer (7/10). (6/9)?). Lo que caracteriza a la reposición es que una misma bolita puede extraerse dos veces. Al extraer las dos juntas, equivale a extraer sin reposición.

Un poco de historia

Timoteo protesta

- Don Joaquín, permítame... Según usted, la probabilidad de extraer dos ases de un mazo de 40 cartas al sacar dos cartas, es:

  sin reponer o juntas (4/40) . (3/39) = 0,00769
  con reposición (4/40) . (4/40) = 0,01

- Has entendido bien.
- Lo que no veo es la necesidad de tanta exactitud; los resultados son parecidos.

Los muchachos creyeron que don Joaquín fulminaría al atrevido; pero, el maestro aprovechó para contarles una historia...

En el siglo XVII, muy lejos de este pago, en Francia, un caballero, muy jugador el mozo, observó que las personas que apostaban a obtener un “doble seis” **por lo menos** al tirar 24 veces dos dados, en una larga serie de juegos, tenían pérdidas. En cambio, los que apostaban a obtener por lo menos un “seis” en 4 tiradas, a la larga obtenían ganancias.

El caballero razonaba: Al tirar un dado, el número de resultados posibles es 6; al tirar dos dados, ese número es 36. Por lo tanto, 4 tiros es a 24 como 6 es a 36 y la probabilidad debe ser la misma.Pero, sus registros estadísticos decían lo contrario. No pudiendo resolver el problema, se lo propuso a Pascal (1623-1662), quien lo resolvió así:

- P(al menos un “seis” en 4 tiros) = 1 -$(5/6)^4$ = 1 - 0,48225 = 0,51775
- (5/6) . (5/6) . (5/6) . (5/6) es la probabilidad de que en 4 tiros no salga el “seis”
- P (al menos “doble seis” en 24 tiros) = 1 - $(35/36)^{24}$ = 1 - 0,50859 = 0,49141

He aquí la razón de lo observado por el caballero: Cuando la probabilidad en un juego es mayor que 0,5, el juego nos es favorable. En caso contrario, desfavorable.

Aunque los resultados son aproximadamente iguales, uno conduce a la fortuna y otro a la ruina. De ahí que el cálculo de probabilidades debe hacerse con la mayor fineza posible. De paso, observamos que debemos ser prudentes en el uso de la regla de tres.

"Mi hermano nació en Clermont, el 19 de junio de 1623. Así, desde su infancia, cuando no se le daban buenas razones, él las buscaba por sí mismo y, cuando se ocupaba de algo, no lo abandonaba hasta quedar satisfecho (...) Su genio para la geometría fue evidente a los doce años (...) Mi padre le dio, para sus horas de recreo, los *Elementos de Euclides* (...) Participaba en las reuniones semanales que se hacían en París, donde las personas presentaban sus descubrimientos (...) Mi padre estaba muy contento con los progresos de mi hermano pero no advirtió que podía perjudicar su salud, que comenzó a alterarse a los 18 años. A los 23 años, habiendo observado las experiencias de Torricelli, realizó sus últimos trabajos en las ciencias exactas y naturales (...) Desde entonces, su espíritu se dedicó a los pensamientos e instituciones religiosas que tanto influyeron en su época (...) Falleció en 1662." (*Vida de Pascal por su hermana*).

Actividad 4.3.

1. Se mezclan 4 billetes falsos y 6 auténticos. Se elige al azar un billete. ¿Cuál es la probabilidad de que sea falso? Se mezclan 4 billetes falsos y 6 auténticos. Si se eligen dos billetes, ¿cuál es la probabilidad de que uno sea falso y el otro bueno?
2. De un mazo de 52 cartas (entre ellas 4 ases), se sacan cartas una tras otra. ¿Cuál es la probabilidad de que el primer as aparezca en la undécima extracción?
3. En una reunión de 30 personas, ¿cuál es la probabilidad de que al menos dos de ellas cumplan años el mismo día?
4. Consideremos que la probabilidad de nacimiento de un varón es 0,51 y la de una niña es de 0,49. En un conjunto de familias de dos hijos: a) ¿Cuál es la probabilidad de que una familia elegida al azar tenga dos varones?b) ¿Al menos un varón?e) ¿Dos hijos varones –si sabemos que ya tiene un varón–?
5. Un alumno se ha equivocado en 2 de 5 problemas. Las soluciones están en hojas separadas y el profesor las corrige eligiéndolas al azar. ¿Cuál es la probabilidad de que la última hoja revisada corresponda a una solución incorrecta?

El caso de la laguna

Don Joaquín y don Hermenegildo, a pesar de sus profundas diferencias en materia de metodología, eran buenos amigos y en sus ratos libres mantenían largas charlas sobre problemas matemáticos, políticos y sociales. En eso estaban un día, en la confitería de Cármine, cuando un tipo en la mesa de al lado dijo:

– ¿Cuántos peces habrá en la laguna?

A los pocos días decidieron hacer un paseo conjunto con sus dos cursos, aunque don Hermenegildo refunfuñaba diciendo que no había que meter a los alumnos en estas cosas.

Con gran entusiasmo y cuidado marcaron 400 peces, y los devolvieron vivitos y coleando a la laguna. Al día siguiente, regresaron, sacaron 200 peces entre los cuales contaron 23 marcados. Devolvieron la mayoría al agua y se quedaron con algunos para freír y chuparse los dedos.

Mientras el fuego hacía su trabajo, don Joaquín preguntó:

– ¿Cuántos peces había en la laguna?

Uno de los muchachos comenzó a hacer cálculos, empleando una ramita sobre el cuaderno de la tierra. Por fin, dijo:

– Yo sé cuántos peces hay.
– ¿Qué vas a saber vos?
– Hay 3.478.
– Efectivamente –dijo Joaquín–. Conviene que agregues: “aproximadamente”. Y ¿cómo lo hiciste?
– Pensé: Llamemos x al número tal de peces de la laguna. Si saco todos, hay 400 marcados. Entonces, me sirve la regla del tres:
  200 ———————— 23
  X ———————— 400
– Regla de tres, –corrigió don Hermenegildo.
– Yo ya sé por qué hay que decir “aproximadamente” –dijo otro.
– Porque si, en vez de 23 hubiésemos sacado 25, el cálculo cambiaba un poco; pero, siempre anda alrededor de esa cifra.

5. Estadística

Valor medio

Suponemos que en una evaluación se han obtenido las notas siguientes: 6; 8; 5; 5; 5; 4; 2; 8; 4; 1; 2; 2; 6; 6; 7 ¿Cómo calculamos el promedio?

Para calcular el promedio, sumamos y dividimos por 15.

También podríamos proceder así:

$$\frac{1(1)+2(3)+3(0)+4(2)+5(3)+6(3)+7(1)+8(2)+9(0)+10(0)}{15}$$

O bien: 1 (1/15) + 2 (3/15) + ...+ 10 (0/15)

Obsérvese que 1/15, 3/15, ... son las frecuencias relativas de cada nota.

La varianza y la desviación

El Sr. J., gobernante de un remoto país, se jactaba de que el salario medio era de 3000 dólares. El Señor Modulador, de visita por esos lugares observó las extrañas costumbres de sus habitantes: vestían muy mal, comían peor, se guarecían donde podían y padecían graves enfermedades. Como el señor Modulador es sumamente ingenuo, creyó que en ese país al ahorro, más que una virtud era una obsesión; pero, por más que preguntó, nadie supo sacarlo de su perplejidad. Hasta que cierto día encontró a un colega Modulador de aquellas regiones que le aclaró el misterio: Sucedía que entre un millón de habitantes, la renta se distribuía así:

X	Renta	n
Señor J.	1.800.000.000	1
Allegados a J.	1.000.120.000	999
Resto	199.880.000	999.000

Desde entonces, el Sr. Modulador desconfía mucho del valor medio.

No tiene la misma situación una nación con salario medio 1000 con valores máximo y mínimo 1300 y 800, que otra nación con el mismo valor medio pero con valores extremos 50 y 10000. En el primer caso, diremos que los valores están menos dispersos que en el segundo.

Debemos estudiar una medida para la desviación o dispersión de los datos.

Variación y desvío

Consideramos las notas de tres cursos que, para simplificar supondremos de 3 alumnos cada uno:

A	B	C
6-6-6	6-8-4	10-0-8

Para los tres cursos el valor medio es 6.

La varianza se calcula haciendo el promedio de los cuadrados de los desvíos respecto del valor medio.Así, para el curso A, tendremos:

$$V(A) = \frac{(6-6)^2 + (6-6)^2 + (6-6)^2}{3} = 0$$

Para B:

$$V(B) = \frac{(6-6)^2 + (8-6)^2 + (4-6)^2}{3} = \frac{8}{3}$$

Para C:

$$V(C) = \frac{(10-6)^2 + (0-6)^2 + (8-6)^2}{3} = \frac{16+36+4}{3} = \frac{56}{3}$$

Se llama **desvío estándar típico** a la raíz cuadrada positiva de la varianza.

$$\sigma(A) = 0$$
$$\sigma(B) = 1{,}63$$
$$\sigma(C) = 4{,}32$$

Las calculadoras de bolsillo suelen tener una parte dedicada a cálculos estadísticos.

En general, debemos comenzar por colocar a la calculadora en posición, para que pueda ocuparse de cuestiones de estadística. Esto se logra, en algunas calculadoras, mediante una palanquita; en otras apretando INV y luego MODE. En el visor suele aparecer SD. A continuación se ingresan los datos. Ejemplo: Para ingresar 3,5,8 se pulsan las teclas: 3, M+, 5, M+, 8, M+. Si, ahora, pulsamos n aparece un 3 en el visor (n es el número de datos); si apretamos $\overline{X}$ aparece 5,333 (valor medio); pulsando σ aparece 2,054 (desviación).

Trabajo estadístico

Un profesor de Educación Física registra los tiempos de carrera (400 metros) de diferentes alumnos, con estos registros en segundos. ¿Cómo hallar el valor medio?

48	67.3	57.1	61,5	56	55.1	61.2	57.8	55.4	58	58.5	51.2
55.7	60.1	53	50.5	60.4	54.2	61.8	49.2	54	59	59	53.9
67.8	64.2	54.5	57.7	61.5	54	61.9	66	55	64.8	58.1	60.7
60.9	56.2	61.6	61.8	58.2	54.1	52.4	55.8	60.3	61.4	67.3	57.4
51.4	57.9	57	57.7	64.8	67.1	65	64.3	52.4	53.4	58	58.3
56.1	66.1	65	59.4	58.2	50.1	59.6	56	51.6	53	65.1	59.8
55	63.6	51.2	61.5	**51.7**	**60.8**	58.4	54.1	62.8	57.5	52.1	61.8
58.1	55.8	67	56.9	50.4	66.5	59.7	65.8	52.1	59.5	47.1	56.2
56.8	**54.5**	53.8	67.2	52.4	52.8	**63.5**	58.5	53.9	**61.8**	63.5	50.3
61.8	62.4	52	61.9	61.7	57.6	56	57.5	61.4	56	**49.2**	53.9
64.7	61.3	56.9	63.5	52.8	63.5	50.9	60.8	55.8	58.1	64.2	57.6
62	55.9	62.8	60.4	53.8	**64.6**	54.6	61.4	59.2	60.1	55	58.2
59	64.8	66	**61.6**	53.8	63.2	54.7	55	56	57.8	60.6	59.2
55.1	62.8	53.7	59.8	57.1	58.6	**52.7**	58.9	55.8	56.5	56.2	56.1

Tantos números nos marean; por esta razón, los agrupamos en:

Tiempo (segundos)	N° alumnos (frecuencia)
47-50	4
50-53	20
53-56	35
56-59	40
59-62	37
62-65	20
65-68	12
	168

Para hallar el valor medio $\overline{X}$, asignaremos a cada alumno el valor central del intervalo. Así, a los 4 alumnos del primer intervalo les asignaremos 48,5 (en lugar de los valores registrados 48; 49,2; 47,1 y 49,2).

Tendremos, así:

$$\overline{X} = \frac{48{,}5.4 + 51{,}5.20 + 54{,}5.35 + 57{,}5.40 + 60{,}5.37 + 63{,}5.20 + 66{,}5.12}{168}$$

$$\overline{X} = 57{,}96$$

En relación con este agrupamiento en clases, resulta evidente que se pierde parte de la información recogida. A mayor amplitud de los intervalos, más información perdemos. También se presentan problemas en los extremos.

¿A qué intervalo asignan el valor 53? Una solución puede ser adjudicar cada unidad que pertenece a un extremo de clase, por mitades, a las clases contiguas. (En nuestro agrupamiento no hemos empleado este criterio, para simplificar los cálculos). A pesar de estos inconvenientes, los agrupamientos son necesarios porque tienen la ventaja de presentar, de modo más claro para el análisis, el material reunido. La pericia de los entendidos permite hallar un compromiso aceptable entre claridad y precisión.

Si el lugar $\overline{X}$ sobre la tabla agrupada, lo hubiésemos calculado según la lista completa de datos, el resultado hubiera sido $m = 58{,}20$ con $\sigma = 4{,}60$.

El símbolo E (x) = m se reserva para las distribuciones teóricas o para el valor medio de la población. Lo mismo sucede con el desvío σ.

En cambio $\overline{X}$ y s se refieren a estimadores de m y σ, obtenidos a partir de muestras.

¿Cómo lograr una estimación de parámetros a partir de muestras?

La población (formada por los tiempos de carrera) de nuestro ejemplo tiene un tamaño n = 168 que no es demasiado grande. Sin embargo, en general, en la práctica, el número de datos suele ser mucho mayor (por ejemplo: los consumos de electricidad de todas las familias de Tucumán) en cuyo caso, por razones de costo y trabajo, es mejor estimar los parámetros m y σ a través del estudio de muestras.

Con relación a nuestro cuadro, tomaremos una muestra de N = 10 valores, para lo cual los elegiremos al azar. Las calculadoras nos proveen números al azar apretando "INV RAN" o "SHIFT RAN". En este momento, voy a obtener 10 números al azar con mi calculadora:

0,820 ; 0,586 ; 0,968 ; 0,467 ; 0,633 ; 0,706 ; 0,615 ; 0,881 ; 0,434 ; 0,457

Los números al azar pertenecen al conjunto de números racionales de tres cifras decimales situados entre 0 y 1. Para adaptarlos a nuestra población, debemos multiplicarlos por 168 y redondear:

138; 98; 163; 78; 106; 119; 103; 148; 73; 77

Buscamos en la tabla los tiempos que corresponden a estos lugares y obtenemos los números destacados en la tabla:

55; 51,7; 60,8 ; 54,5 ; 63,5 ; 61,8 ; 49,2; 64,6; 61,6; 52,7

La media muestral es: $\overline{X} = 57{,}54$

El desvío muestral es: $S = 5{,}23 \ (\sigma \text{ de la calculadora})$

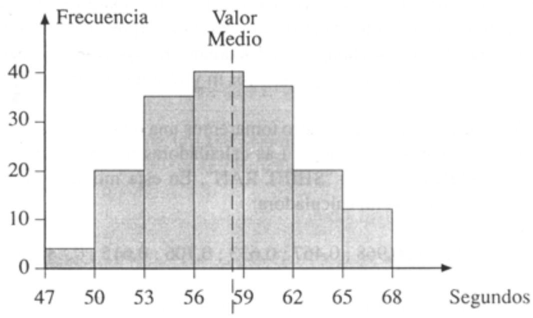
Si el tamaño de la muestra es inferior a 30, es mejor emplear σ que nos da 55,1.

Para pasar de σ_n a σ_{n-1} se emplea la fórmula:

$$\boxed{\sigma_{n-1} = \sigma_n \sqrt{\frac{n}{n-1}}}$$

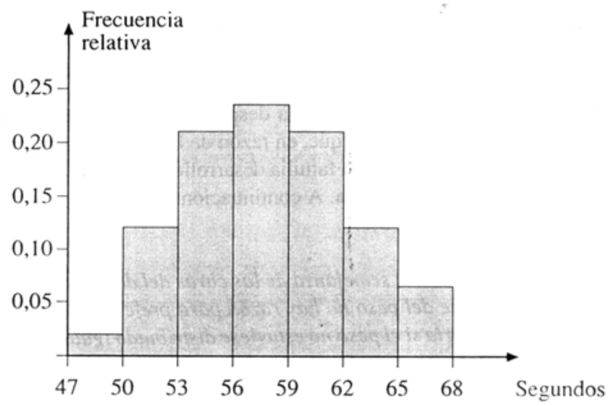
Histogramas

Podemos hacer una representación gráfica de nuestros datos agrupados:



También podemos representar las frecuencias relativas:

Clase 1: 4/168 = 0,024
Clase 2: 20/168 = 0,12
Clase 3: 35/168 = 0,21
Clase 4: 40/168 = 0,24
Clase 5: 37/168 = 0,22
Clase 6: 20/168 = 0,12
Clase 7: 12/168 = 0,07



Las frecuencias relativas pueden traducirse a porcentaje, multiplicando por 100.

Actividad 5.1.

1. Se ha tomado una muestra de cinco tabletas de chocolate, obteniéndose los siguientes pesos: 18,5 gramos; 19,8; 20,4; 20,1 y 19,7. Hallar el valor medio $\overline{X}$ y la desviación s_{n-1}.
2. Se ha tomado una muestra de 6 lámparas para registrar su duración. Se obtuvieron estos resultados: 1000 horas; 1150; 1220; 1400; 1250; 1320. Averigüe $\overline{X}$ y s_{n-1}.

Dibuje un ángulo en el pizarrón y proponga que los alumnos estimen su medida en grados, a ojo. El curso hará una estadística de las distintas respuestas. Se pueden hacer histogramas. Calcular $\overline{X}$ y s_{n-1}. Por fin, medir el ángulo y comentar los resultados.

Los Bernoulli

Al comenzar el siglo XVI, Flandes era el centro del comercio y la banca; y, en particular, el puerto de Amberes que había despojado a Brujas del predominio financiero. Allí vivían los Bemoulli que, en razón de las persecuciones políticas y religiosas, pasaron a Suiza donde la familia desarrolló una importante actividad científica y, en particular, matemática.

A continuación transcribimos una página de Jacobo Bemoulli (1654-1705):

"En virtud de la semejanza de las caras del dado y de la distribución uniforme del peso, no hay razón para preferir una de las caras, como sucedería si el peso no estuviese distribuido igualmente o si las caras tuviesen distinta forma. De la misma manera, si conocemos el número de bolillas blancas y el de bolillas negras de una caja, podemos calcular probabilidades. Pero, yo pregunto a Usted, ¿quién de los mortales puede conocer, por ejemplo, el número de enfermedades que invaden las innumerables partes del cuerpo humano a cualquier edad y pueden causar su muerte? ¿Quién puede afirmar qué enfermedad puede matar más fácilmente a un hombre (la plaga o la hidropesía, la hidropesía o la fiebre) para predecir la vida o la muerte?

¿Quién puede registrar los innumerables cambios de la atmósfera para predecir lo que sucederá dentro de un mes? ¿Quién conoce suficientemente la admirable estructura de nuestra mente o nuestro cuerpo, para predecir un ganador en juegos que dependen de la agudeza del ingenio o de la agilidad del cuerpo? Puesto que tales cosas dependen del ingenio o de causas ocultas en virtud de la gran variedad de combinaciones que escapan a nuestros esfuerzos de cálculo, sería decididamente insano intentar un conocimiento de este modo.

Sin embargo, hay otra manera de obtener lo que queremos. Lo que es imposible a priori puede encontrarse, al menos, a posteriori; es decir, registrando los resultados de observaciones realizadas muchas veces. Porque, puede suponerse que un hecho puede ocurrir, en el futuro, tantas veces como haya aparecido en el pasado, bajo condiciones similares. Por ejemplo, si en el pasado, 300 hombres de la misma edad y estado general que tiene Tito en este momento fueron investigados y se encontró que (en el término de 10 años) 200 murieron, podemos pensar que Tito tiene una probabilidad de 2/3 de pagar, en la próxima década, su deuda con la vida. De la misma manera, se pueden llevar registros sobre el tiempo (bueno, lluvioso, etc.) o sobre una larga serie de juegos entre dos personas.

Esta manera de determinar razones entre el número de casos favorables y el número total de casos no es nueva. Aunque prácticamente todos la conocemos, la demostración sobre bases científicas no es de ninguna manera fácil. Todavía queda por ver algo que quizá nadie ha pensado y es determinar si al número de observaciones aumenta la probabilidad de obtener una razón (f:n) genuina o si, por el contrario, llega un punto en que no es posible mejorar el resultado.

Si al aumentar el número de observaciones no puedo mejorar la apreciación, confieso que nuestra tarea es inútil. Si sucede lo contrario, serán muchas las aplicaciones y podremos aplicar la teoría con la misma confianza que en los juegos de azar. Por ejemplo, si en lugar de una caja con bolillas extraemos aire del cuerpo humano que contiene causas de varias enfermedades como una caja contiene bolillas podremos, por observaciones repetidas, determinar cuándo una enfermedad puede darse con mayor probabilidad que otra. Con este problema he luchado durante 20 años."[1]

Jacobo Bernoulli jamás hubiera tenido éxito con uno de esos profesores que dicen:

– Al primero que termina le pongo una buena nota.

Miren que tardar 20 años... Debía ser medio tonto.

[1] *Uspensky, J. V.* Introduction to Mathematical Probability. *Mc. Graw-Hill.*

Clave de respuestas

1.1.

1. p(oro) $= \dfrac{10}{40} = \dfrac{1}{4}$

2. p (siete) $= \dfrac{4}{10} = \dfrac{1}{10}$

3. p (no copa) $= \dfrac{30}{10} = \dfrac{3}{4}$

4. p (as o siete) $= \dfrac{2}{10} = \dfrac{1}{5}$

5. p (oro o siete) $= \dfrac{13}{40}$

6. 1

7. 0; pues esta carta no existe en el mazo.

1.2. p (A o B) = p (A) + p (B), sólo cuando A y B no pueden darse simultáneamente; es decir, cuando A y B son sucesos excluyentes.

p (as o siete) = p (as) + p (siete) $= \dfrac{1}{10} + \dfrac{1}{10} = \dfrac{1}{5}$

pues, una carta no puede ser, a la vez, as y siete.

En cambio, en 1.1. no es p (as o siete) = p (oro) + p (siete), pues oro y siete no se excluyen: una carta puede ser el siete de oro.

1.3. A y B son sucesos excluyentes; $A \cap B = \varnothing$

1.4. $P(5) = \dfrac{1}{6}$. Una manera de obtenerla, por el axioma 3, es:

$$P(\{2,3,5\}) = P(\{2,3,5\}) + P(\{5\})$$

$$\frac{3}{6} = \frac{2}{6} + P(\{5\})$$

Datos en el diagrama

$$\therefore P(\{5\}) = \frac{1}{6}$$

$$P(\text{par}) = \frac{3}{6}$$

1.5. Podría no ser única, una función que cumpliera con la tabla dada y con los tres axiomas de probabilidad.

1.6. Está mal aplicado el axioma 3, pues $\{1,2\}$ y $\{2,3,4\}$ no son conjuntos disjuntos.

2.1.

1. 10.9.8 = 720

2. 9.8.7.6 = 3024

3. 9.9.9.9 = 6521

4. 4 3 3 2. Comenzando con impar, hay 4 elecciones posibles en primera casilla y 3 en tercera casilla; para la segunda casilla tenemos tres pares y, para la cuarta, dos pares. Empezando con par, tenemos:
3 (4) (2) (3) En total: 4.3.3.2 + 3.4.2.3 = 144

5. Hay que empezar, forzosamente, por impar:
4 (3) (3) (2) (3) (1) (1) = 144

6. a) 5040
   b) 4! 3! 2! = 288

7. $\binom{40}{5} = 658008$

8. a) 8! = 40320

   b) $\dfrac{8!}{2!} = 20160$

   c) $\dfrac{8!}{2!\ .\ 3!} = 10$

2.2. a) 1
   b) m

2.3. 2)

$$P = \frac{\binom{8}{3} . \binom{12}{2}}{\binom{20}{5}} = 0{,}238$$

3.1.

1. 6!

2. 6! / 3! = 120

3. 6! / 2! 2! = 180

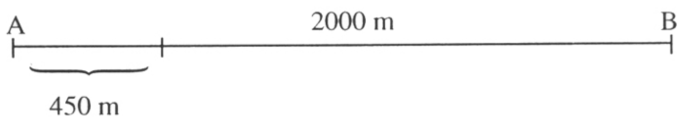
3.2.

1. punto medio de $\overline{AB}$
   punto medio de $\overline{AO}$
   punto medio de $\overline{OB}$

Todos los puntos de $\overline{RS}$ están más cerca de 0, que de A o B.

$$\therefore P = \frac{\text{long.}\overline{RS}}{\text{long.}\overline{AB}} = \frac{1}{2}$$

2.



$\frac{450}{2000}$, suponiendo que los tramos de la línea son homogéneos.

4.1. Llamando D al suceso "Seleccionado un artículo al azar, resulta defectuoso", será:

$$P(D) = P(M_1).P(D/M_1) + P(M_2).P(D/M_2) + P(M_3).P(D/M_3)$$

(Siendo M_i el artículo que sale de la máquina i)

$$P(D) = \frac{40}{100}.\frac{3}{100} + \frac{35}{100}.\frac{4}{100} + \frac{25}{100}.\frac{5}{100} = \frac{385}{10000}$$

$$P(M_1/D) = \frac{P(D/M_1).P(M_1)}{\frac{385}{10000}} = \frac{\frac{3}{100}.\frac{40}{100}}{\frac{385}{10000}} = \frac{120}{385}$$

4.2.

1. C: "Sacar verde de la caja N°2"
   D: "Sacar roja de la caja N°1"

   $P(C \cap D) = P(C).P(D) = \frac{4}{12}.\frac{4}{9} = \frac{4}{27}$, pues C y D son independientes.

2. $P(A \cap B) = \frac{5}{9}.\frac{4}{9} = \frac{20}{81}$

3. Resuelto en el texto.

4. P (roja o verde) $= \frac{4}{9}.\frac{5}{8} = \frac{5}{18}$

4.3.

1. $P(\text{falso}) = \frac{4}{10}$

   $P(\text{uno falso y uno bueno}) = \frac{48}{100}$

2. P (as recién en la 11° extracción) $= \frac{48}{52}.\frac{47}{51}.\frac{46}{50}\cdots\frac{39}{43}.\frac{4}{42}$

3. 1-P (ninguna cumple en el mismo día) $= 1 - \frac{365}{365}.\frac{364}{365}.\frac{363}{365}\cdots\frac{336}{365}$

4. a) $0,51^2$
   b) 0,51
   c) $0,51^2$

5. $\frac{4}{10}$

5.1.

1. 19,85; 0,90

2. 1223; 138,9

MÓDULO 2

En este módulo se tratan los siguientes temas:

- Pruebas repetidas.
- Distribución binomial.
- Aproximación de Poisson.
- Desigualdades de Tchbyccheff y Bernoulli.
- Empleo de números aleatorios

El tratamiento es circular, procurando repasar las nociones del primer módulo y ampliándolas.

Así como prácticamente todo el contenido del primer módulo puede volcarse al aula en la enseñanza media, este segundo módulo apunta a profundizar temas seguramente vistos pero, acaso, olvidados.

El profesor que tiene una visión amplia de la asignatura sabe a dónde debe apuntar sus objetivos. Es, además, probable que los temas tratados aparezcan en los contenidos definitivos de las polimodales.

Si algunos temas nos resultan difíciles, leamos con detenimiento y sentido crítico: Es una aptitud que debemos desarrollar en los alumnos. La tarea formativa no debe consistir en una fábrica de robots calculadores; en esto no podremos nunca sustituir a Casin, Apple, IBM, Macintosh...

6. Fórmula de Bernoulli

Pruebas repetidas

Pensemos en un experimento aleatorio que puede repetirse, en las mismas condiciones, n veces. Suponemos, además, que el resultado de cada experimento no depende de los demás. Por ejemplo, podemos tirar un dado cinco veces y registrar, en cada tirada, si sale o no sale el "cuatro". Para facilitar nuestros registros, escribiremos 1 si sale el cuatro y cero en caso contrario.

Nosotros obtuvimos la secuencia: 00010.

Podríamos preguntarnos qué probabilidad tiene nuestra secuencia, antes iniciar las tiradas. ¿Por qué no cierra el módulo y lo piensa?

La probabilidad de que:

- en la primera tirada, no salga el 4 es 5/6;
- en la segunda no salga es 5/6;
- no salga en la tercera es 5/6;
- aparezca el 4 en la cuarta tirada es 1/6;
- no salga en la quinta es 5/6.

Como un resultado no influye sobre el otro, multiplicamos:

$$\left(5/6\right)\left(5/6\right)\left(5/6\right)\left(1/6\right)\left(5/6\right) = 0{,}080$$

¿Puede obtener el mismo resultado como cociente entre el número de casos favorables y el número total de casos?

Actividad 6.1.

Con relación al párrafo anterior, calcular la probabilidad de las secuencias:

a. 00000
b. 11111
c. 00101
d. 01001
(1= "Sale cuatro" 0 = "No sale cuatro").

Actividad 6.2.

Calcular la probabilidad de que, al tirar 5 veces el dado, aparezca:

a. Exactamente una vez el cuatro.
b. Exactamente dos veces el cuatro.

En el problema anterior hemos calculado la probabilidad de una secuencia determinada. Planteamos ahora el siguiente problema:

Si tiramos cinco veces el dado, ¿cuál es la probabilidad de obtener exactamente un "cuatro"? El problema es distinto al pedir una secuencia determinada, ¿por qué?

Hemos visto que la probabilidad de obtener la secuencia 00010 –es decir, de obtener un "cuatro" en la cuarta tirada y no sacar "cuatro" en las restantes– es $(1/6)\ (5/6)^4$.

La probabilidad de obtener exactamente un "cuatro" (no importa en qué tirada) será $5\ (1/6)\ (5/6)^4$, porque el cuatro puede salir en la primera tirada, en la segunda...

Y, ¿cuál será la probabilidad de obtener exactamente dos "cuatros" en cinco tiros?

Sabemos que la probabilidad de una secuencia cualquiera con dos "cuatros" (por ejemplo, 01001) es $(1/6)^2\ (5/6)^3$.

Hay muchas secuencias que tienen esa misma probabilidad (11000, 01100, 10001...). Debemos averiguar cuántas secuencias con exactamente dos "cuatros" hay. Razonemos así: Tenemos 5 lugares y debemos elegir dos lugares para ubicar los "cuatros".

Es lo mismo que elegir dos lugares entre cinco; el número será $\binom{5}{2}$.

Por lo tanto, la probabilidad de obtener exactamente dos "cuatros" al tirar 5 veces el dado será $10\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$.

Vale la pena escribir las secuencias que contienen exactamente dos "cuatros" en 5 tiradas: 11000; 01100; 00110; 10100; 01010; 00101; 10010; 01001; 00011; 10001.Cada una de ellas tiene probabilidad $(1/6)^2\ (5/6)^3$.

Nos resultará fácil ahora calcular las probabilidades de obtener: Ningún "cuatro"; exactamente un "cuatro":

$$P_5(0) = (5/6)^5 = 0{,}40187$$

$$P_5(1) = 5\ (1/6)\ (5/6)^4 = 0{,}40187$$

$$P_5(2) = \binom{5}{2}(1/6)^2\ (5/6)^3 = 0{,}16075$$

$$P_5(3) = \binom{5}{3}(1/6)^3\ (5/6)^2 = 0{,}03215$$

$$P_5(4) = \binom{5}{4}(1/6)^4\ (5/6) = 0{,}00321$$

$$P_5(5) = (1/6)^5 = 0{,}00013 \qquad \text{Suma} = 0{,}99998\ (\text{aprox.}1)$$

Si tiramos n veces un dado ideal, podremos escribir la siguiente fórmula:

$$P_n(h) = (1/6)^h (5/6)^{n-h}$$

$P_n(h)$ significa "Probabilidad de que en n tiradas salga exactamente h veces una determinada cara del dado".

En general, si repetimos n veces un experimento aleatorio, de modo tal cada repetición el suceso A tenga probabilidad constante p, tendremos:

$$\boxed{P_n(h) = \binom{n}{h} p^h\, q^{n-h}} \text{ donde } q = 1 - p$$

Actividad 6.3.

Tiramos un dado 10 veces. ¿Cuál es la probabilidad de obtener exactamente 4 ases?

Actividad 6.4.

En una caja hay tres bolitas blancas, 2 verdes y 5 azules. Extraemos una bolita al azar y la volvemos a la caja. Hacemos esto 6 veces. ¿Cuál es la probabilidad de sacar exactamente 4 veces "blanca"?

Actividad 6.5.

Tiramos una moneda 8 veces. ¿Cuál es la probabilidad de obtener 3, 4 o 5 caras?

Actividad 6.6.

Tiramos 5 veces una moneda. ¿Cuál es la probabilidad de obtener 0, 1, 2, 3, 4, 5 caras?

Si tiramos n veces el dado y nos interesa el suceso "Sale as", este suceso puede aparecer 0 vez, 1 vez.... (n-1) veces. No hay otra alternativa; en consecuencia, la suma de las probabilidades debe dar 1. Es fácil probarlo. La probabilidad de sacar el as es 1/6 y la contraria es 5/6. Podemos escribir:

$$(1/6 + 5/6)^n = (1/6)^n + \binom{n}{1}(1/6)^{n-1}(5/6) + \binom{n}{2}(1/6)^{n-2}(5/6)^2 + \ldots + (5/6)^n$$

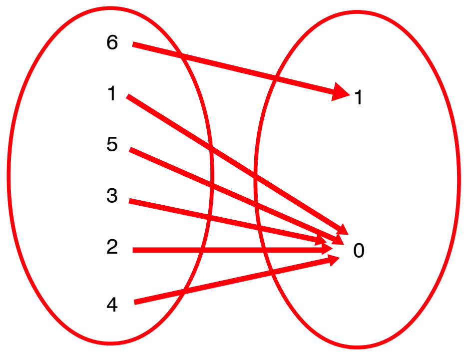
¿Por qué la suma de todos estos términos debe dar 1? En general, si p es la probabilidad de un suceso A, en n pruebas repetidas independientes, tendremos:

$$(p+q)^n = \sum_{h=0}^{n} \binom{n}{h} p^{n-h}\, q^h = 1$$

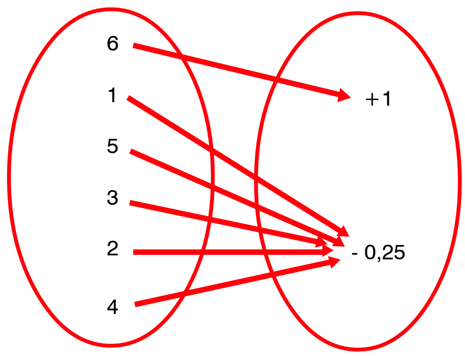
La fórmula de probabilidad en n pruebas repetidas independientes con probabilidad constante se denomina **fórmula de Bernoulli**.

Variables aleatorias

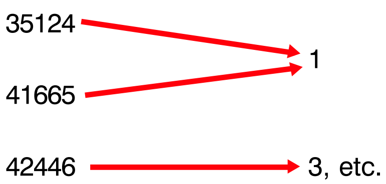
Tiramos un dado y observamos si sale o no sale "seis". Si sale "seis", anotamos 1; si no sale, registramos 0. Podemos definir la función:



Una función que, a cada resultado de un experimento aleatorio le asigna número real, es una **variable aleatoria**. Simbolizaremos las variables aleatorias con X, Y etc. Claro está que –si salir "seis" nos proporciona una ganancia de $1 y no salir "seis" una pérdida de $ 0,25– podríamos considerar la función:



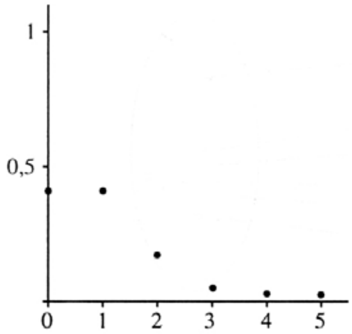
Consideremos, nuevamente, el caso del dado que tiramos 5 veces para observar cuántas veces sale el "cuatro". El espacio muestral, es decir el conjunto de todos los resultados posibles, está formado por las 5-uplas (35124; 41665...). Pero, como a nosotros nos interesa solamente el número de "cuatros", nuestra variable aleatoria X sólo puede tomar los valores 0,1,2,3,4,5, según el esquema:

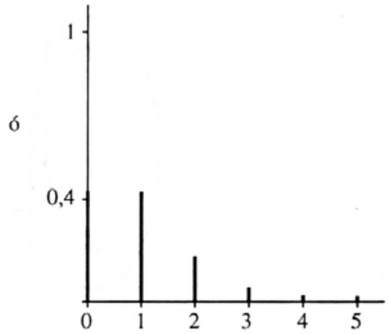


Podemos definir una función que a cada valor de la variable aleatoria le asigne su probabilidad. En el caso del dado que se tira 5 veces para ver cuántas veces sale el "cuatro" tendremos, de acuerdo con los cálculos anteriores:

P_5 (0) = 0,40187
P_5 (1) = 0,40187
P_5 (2) = 0,16075
P_5 (3) = 0,3215
P_5 (4) = 0,00321
P_5 (5) = 0,00013

Hagamos la gráfica:





A esta función la llamamos **función de densidad de probabilidad** o **función probabilidad** f (x). En símbolos, f (x) = P (X = x).

Por ejemplo, f (5) es igual a probabilidad de que la variable aleatoria tome el valor 5. El gráfico nos dice que, al tirar 5 veces el dado, es poco probable obtener 3, 4 ó 5 “cuatros”.

Consideremos el siguiente problema:

En una caja hay 3 bolitas blancas, 2 verdes y 3 azules. Extraemos una bolita al azar y no la reponemos. Hacemos esto 6 veces. ¿Cuál es la probabilidad de sacar exactamente 3 blancas?

En este caso, no podemos emplear la fórmula binomial porque la probabilidad de blanca no es constante (no hay reposición).

7. Funciones de distribución

Continuamos con el ejemplo que estamos trabajando, para preguntamos:

¿Cuál es la probabilidad de que, al tirar 5 veces un dado, el número de "cuatros" sea 0,1, 2 o 3 ? Esto equivale a preguntarse: ¿Cuál es la probabilidad de que la variable cumpla la condición $0 \leq X \leq 3$?

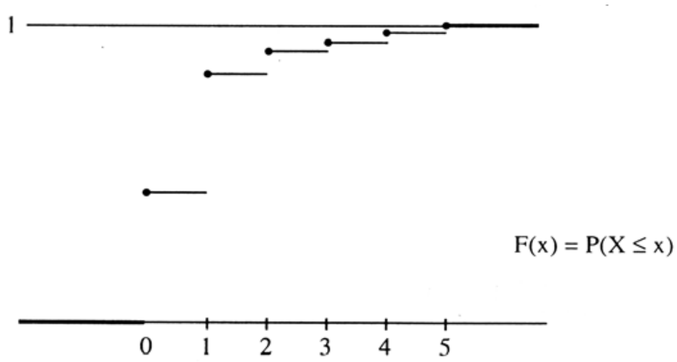
La respuesta es 0,40187 + 0,40187 + 0,16075 + 0,03215.

Podemos, entonces, definir una función acumulando las probabilidades. En nuestro ejemplo, tendremos la siguiente tabla:

X	**F(x)**
0	0,40187
1	0,80374
2	0,96449
3	0,99664
4	0,99985
5	1,00000

Es posible extraer esta función para cualquier x real. Así, por ejemplo, la probabilidad de que, al realizar el experimento, la variable aleatoria X (número de veces que sale 4) tome un valor menor que cero es, evidentemente, cero. La probabilidad de que X tome un valor menor o igual que 8 es 1.

Esta función $F(x) = P(X \leq x)$ de probabilidad acumulada se llama **función de distribución de la variable aleatoria X**. Hagamos la representación gráfica:



No hay que pasar este gráfico sin examinarlo con cuidado. Observe, por ejemplo, que usted no puede sacar menos 2 (-2) veces el cuatro; en consecuencia, la probabilidad es cero y esto sucede desde (-infinito) hasta casi, casi el cero. Pero, usted puede sacar cero veces el cuatro. De ahí el salto que se observa en X = 0. Las probabilidades de $X \leq 1{,}26$ $X \leq 1{,}37$ etc., son iguales a la probabilidad de X = 0.

En el punto X = 1, la gráfica pega un salto porque se ha acumulado la probabilidad de 1 a la probabilidad de 0.

Para X = 5 se produce el último salto. Piense que $P(X \leq 6) = 1$. En efecto, al tirar 5 veces un dado, el número de "cuatros" que obtenemos es siempre menor o igual que 6.

Antes de seguir adelante, tómese un recreo:

Tire 5 veces el dado y repita 100 veces (o tire 100 veces 5 dados). Anote el número de “cuatros”. Espere un rato; ya le daré mis resultados.

La prueba consiste en tirar una moneda 6 veces y contar el número de caras. Nuestra variable aleatoria X = número de caras, puede tomar los valores 0, 1, 2, 3, 4, 5, 6. Pedimos:

a) Completar la tabla con error menor que 0,001.
b) Representar f (x).
c) Representar F (x).

X	0	1	2	3	4	5	6
F (x)	0,016	0,094	0,234	0,312	0,234	0,094	0,016
F (x)	0,016	0,110	0,344	0,656	0,890	0,984	1,000

– *Hay algo que me cuesta comprender. En estas funciones de las que usted habla hay dos “equis”, una grande y una chica. Me parece que no se trata de funciones de dos variables, porque puede representarlas sin dificultad en el papel.*
– *La variable independiente es la x minúscula. La X es la variable aleatoria que sirve como intermediaria para dar valores a la función. Así, por ejemplo, si tiro 6 veces la moneda para observar el número de caras, tendré: f(1) = P (X=1) = 0,094, es decir f(1) será la probabilidad de que el número de caras sea igual a 1. Si entendió, podrá decirme cuál es el valor de f(7).*
– *Evidentemente, 0; porque, si tiro 6 veces la moneda no pueden salir 7 caras. Ahora, explíqueme el significado de F(x).*
– *Lea bien* $F(x) = P(X \leq 3)$. *Quiere decir que F(3) es la probabilidad de que el número de caras sea menor o igual a 3.*
– *¡Está claro! Para hallar F(3), debo sumar las probabilidades de 0, 1, 2, 3 caras.*
– *¿Cuál será, entonces, F(10)?*
– *F(10) = 1; pues,* $P(X \leq 10) = 1$, *ya que si tiro 6 veces la moneda, el número de caras será ciertamente menor o igual que 10.*

Limitaciones

– *Se ve que estoy algo viejo porque el trabajo del recreo me dejó algo preocupado. Miré la tabla que obtuve al tirar 5 dados para observar el suceso “4”. Repetí el experimento 100 veces.*

X = Número de 4	0	1	2	3	4	5
frecuencia	40	38	15	3	0	0

Hay resultados que parecen encajar y otros no. Un compañero me pasó otra tabla:

X = Número de 4	0	1	2	3	4	5
frecuencia	40	38	15	3	0	0

Si no entendí mal, de acuerdo con los cálculos teóricos, deberíamos haber obtenido: 40, 40, 16, 3...

- *No olvide usted que hemos hecho los cálculos teóricos pensando en monedas o dados ideales.*
- *¡Ah! Ya veo; yo usé un dado de carne y hueso. Mejor dicho, de hueso...*
- *O de cuerno; algo lejos del dado de Platón...*
- *Pero, entonces, no veo para qué pueden servirnos los razonamientos y fórmulas que buen trabajo nos han dado. Aunque, por otra parte, no hay gran discrepancia entre el comportamiento de mi dado y el de Platón.*
- *Por ahí anda la cosa...*
- *Mejor, déjeme decir lo que pienso...; pero, no se ría. El dado ideal es un modelo, como el péndulo ideal. Aunque sólo existe en nuestra mente, permite anticipar aproximadamente el comportamiento de los pobres dados mortales...*
- *Siempre que esos dados no se alejen demasiado del dado ideal.*
- *Pero, si la discrepancia fuese grande, ¿qué hacer? Además, ¿cuál será la discrepancia tolerable?*
- *Lo felicito por sus explicaciones. En cuanto a su última pregunta: 1) Si la discrepancia fuese grande, repetiríamos, de ser posible, la experiencia; en caso de observar nuevamente discrepancias grandes, habría que pensar que el modelo no sirve para describir el comportamiento de nuestro dado. Si usted se diese una vuelta por Neptuno y observase que la suma de los ángulos de un triángulo material es 120° y para otro triángulo de 118°, tendría que abandonar el modelo de Euclides. 2) En cuanto a la discrepancia tolerable, ya tendremos tiempo de tratarla, más adelante.*

Valor medio para la distribución binomial

Recordamos cómo se calcula el valor medio para la tabla:

x	f
5	10
6	25
7	15

(1)

X puede interpretarse como el puntaje obtenido en una prueba; f como frecuencia o número de alumnos con ese puntaje. El valor medio será:

$$E(x) = m = \frac{5.10 + 6.25 + 7.15}{50} = 6{,}1 \qquad (A)$$

Si en lugar de frecuencias se dan probabilidades (tabla 2), obtendremos el valor medio mediante el cálculo:

x	f
0	0,3
1	0,5
2	0,2

(2)

$$E(x) = 0(0{,}03) + 1(0{,}5) + 2(0{,}02) = 0{,}9 \qquad (B)$$

La tabla (2) podría interpretarse así: Con referencia a un equipo de fútbol, la probabilidad de ganar es 0,2; la probabilidad de perder es 0,3.

Cada vez que se da una tabla, es conveniente hallar posibles interpretaciones para la misma.

Obsérvese que (A) y (B) son equivalentes porque (A) podría escribirse:

$$m = 5(0{,}2) + 6(0{,}5) + 7(0{,}3) = 6{,}1$$

Calculemos, ahora, el valor medio para un solo experimento, en el cual observamos un suceso de probabilidad p.

$$E(x) = 1p + Oq = p$$

1	P
0	q

(3)

Para n experimentos independientes, el valor medio será:

$$\boxed{E(x) = np}$$

Si tiramos 180 veces un dado para observar la aparición del “as”. ¿Cuál será el valor medio o esperado?

El valor medio o esperado será np = 180 (1/6) = 30

Esto significa que, en 180 tiradas del dado, esperamos que el número de “ases” no se aparte demasiado de 30.

Varianza en la distribución binomial

Recordemos que para calcular la varianza correspondiente a la tabla 1, procedemos así:

$$V(x) = \frac{(5-6{,}1)^2.10 + (6-6{,}1)^2.25 + (7-6{,}1)^2.15}{50} = 0{,}49$$

O bien:

$$V(x) = (5-6{,}1)^2.0{,}2 + (6-6{,}1)^2.0{,}25 + (7-6{,}1)^2.0{,}3 = 0{,}49$$

Para la tabla (3) obtendremos:

$$V(x) = (1-p)^2 p + (0-p)^2 q$$

(un solo experimento)

$$V(x) = q^2 p + p^2 q = pq(q+p) = pq$$

Para n pruebas independientes, la varianza será:

$$V(x) = npq$$

Y, en consecuencia:

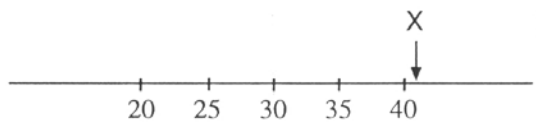
$$\boxed{\text{desvío} = \sqrt{npq}}$$

Si n = 180 tiradas de un dado para observar “as”. ¿Cuál será la varianza?

$$V(x) = 180\left(\tfrac{1}{6}\right)\left(\tfrac{5}{6}\right) = \tfrac{900}{36}$$

$$\text{desvío} = \left(\tfrac{30}{6}\right) = 5$$

Aunque sin pretender una gran precisión, podemos intentar una respuesta a cuáles serán las discrepancias admisibles. En 180 tiradas del dado, si el número de ases se encuentra a distancia mayor de 2 desvíos del valor medio, pondríamos al dado bajo sospecha. Si se encuentra el resultado a distancia mayor de 3 desvíos, la sospecha sobre la honestidad del dado sería vehemente.



Obtener 41 ases sería sospechoso.

Actividad 7.1.

Se tiene un dado con forma de tetraedro con 3 caras azules y una blanca. Lo tiramos cinco veces. La variable aleatoria considerada es X = n° de azules.

a) Calcular las probabilidades para x = 0, 1, 2, 3, 4, 5.
b) Representar la función de probabilidad.
c) Representar la función de distribución.

Actividad 7.2.
Calcular el valor medio, la varianza y el desvío correspondientes al problema anterior.

Actividad 7.3.

La probabilidad de encontrar un producto defectuoso de cierta marca es 0,05. ¿Cuál es la probabilidad de que, al recibir 10 productos de esa marca, aparezcan:

a) ningún producto defectuoso;
b) más de un defectuoso.

Actividad 7.4.

Si dos equipos A y B son teóricamente equivalentes, ¿cuál es la probabilidad de que, en una serie de 8 partidos, A gane todos o A gane 7?

Actividad 7.5.

En una caja hay 3 bolitas rojas y 7 azules. Se extrae una bolita y, si resulta azul, se repone y se agrega otra azul; si sale roja se repone y se agrega otra roja. ¿Cuál es la probabilidad de obtener azul al realizar la tercera extracción?

8. Distribución de Poisson

El número e

La experiencia nos dice que pocos recuerdan el caso $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$ y, seducidos por esa muchacha casquivana (a veces tan certera, a veces tan equivocada) que llamamos intuición, contestan 1.1.1... no puede dar otra cosa que 1.

No es el momento de hacer un tratado de análisis; pero, la calculadora nos dirá que:

$$\left(1+\frac{1}{10}\right)^{10} = 2{,}59 \qquad \left(1+\frac{1}{100}\right)^{100} = 2{,}70$$

Sólo queríamos recordar que:

$$\boxed{\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e}$$

Y también esperamos que usted recuerde las variantes:

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n = \lim_{n\to\infty}\left[\left(1+\frac{1}{-n}\right)^{-n}\right]^{-1} = e^{-1}$$

O bien:

$$\lim_{n\to\infty}\left(1-\frac{k}{n}\right)^n = \lim_{n\to\infty}\left[1+\frac{1}{\frac{-n}{k}}^{\frac{-n}{k}}\right]^{-k} = e^{-k}$$

Es bueno saber que, si la base tiende a 1 y el exponente tiende a infinito, en el asunto está metido el número e.

No olvide, sin embargo que, si la base tiende a 1 pero el exponente es una constante, el límite es 1.

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^k = 1$$

Serie de Maclaurin

Recordemos también el desarrollo en serie de Maclaurin de la función

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

En consecuencia:

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \ldots$$

Distribución de Poisson

Hasta aquí, hemos resuelto problemas sobre pruebas independientes, empleando la fórmula binomial. Sin embargo, la aplicación de esta fórmula resulta complicada cuando el número de pruebas es grande.

Se realizan 1000 pruebas independientes y registramos si aparece o no el suceso A (p = 0,3) para conocer la probabilidad de que el número de veces que se da A está entre 270 y 420.

Deberíamos calcular:

$$\binom{1000}{270} 0{,}3^{270}\ 0{,}7^{730} + \binom{1000}{271} 0{,}3^{271}\ 0{,}7^{729} + \ldots + \binom{1000}{420} 0{,}3^{420}\ 0{,}7^{580}$$

Trabajo que hubiera amedrentado al mismo Hércules. Para enfrentar estos laberintos, se han abierto distintas vías. Veamos una de ellas...

Sabemos que:

$$P(n,h) = \binom{n}{h} p^h\ q^{n-h}$$

$$P(n,h) = \frac{n!}{(n-h)!\,h!} p^h\ q^{n-h} = \frac{n(n-1)\ldots 3.2.1}{[(n-h)(n-h-1)\ldots 3.2.1]\ h!} p^h\ q^{n-h}$$

$$P(n,h) = \frac{n(n-1)\ldots(n-h+1)}{h!} p^h\ q^{n-h}$$

Dividiendo cada factor por n y compensando:

$$P(n,h) = \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{h-1}{n}\right)}{h!} (np)^h (1-p)^{n-h}$$

$$P(n,h) = \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{h-1}{n}\right)}{h!} (np)^h \left(1-\frac{np}{n}\right)^n (1-p)^{-h}$$

Si tomamos límite para n tendiendo a infinito y llamamos λ al límite de np cuando n tiende a infinito y p tiende a cero, tendremos:

$$\boxed{P(n,h) = \frac{\lambda^h}{h!} e^{-\lambda}}$$ Fórmula de Poisson

En el desarrollo anterior hemos supuesto que cuando n tiende a infinito y p tiende a cero, el producto np tiende a una constante λ. Esta suposición implica restricciones en el empleo de la fórmula. En un problema particular, n puede ser grande pero no infinito y p no tiende a cero porque p es constante.

De todos modos si n es grande y p pequeño, la aproximación de Poisson es muy buena.

En una caja hay 500 bolitas (495 rojas y 5 azules). ¿Cuál es la probabilidad de que, si hacemos 2000 extracciones al azar con reposición, obtengamos exactamente 3 azules?

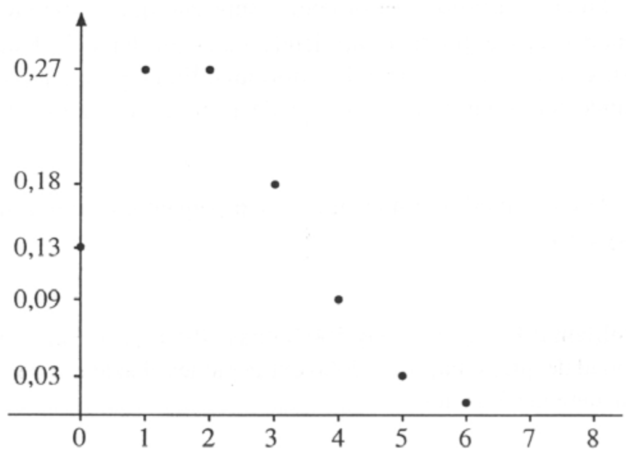
La fórmula de Bernoulli indica que:

$$p = \binom{200}{3} 0{,}01^3.0{,}99^{197} = 0{,}1813$$

Aplicando la fórmula de Poisson:

$$\lambda = np = 200\ .\ 0{,}01 = 2$$

$$P_3 = \frac{\lambda^3}{3!} e^{-\lambda} = \frac{2^3}{3!} e^{-2} = 0{,}1804$$



Representación de la función de probabilidad de Poisson para λ=2

Un dodecaedro tiene 11 caras blancas y una cara verde. ¿Cuál es la probabilidad de que, al tirarlo 10 veces, salgan exactamente dos verdes?

Por la binomial:

$$P_2 = \binom{10}{2}\left(\frac{1}{12}\right)^2\left(\frac{11}{12}\right)^8 = 0{,}15579$$

Empleando Poisson:

$$\lambda = 10.\frac{1}{2} = \frac{5}{6}$$

$$P_2 = \frac{\left(5/6\right)^2}{2!}e^{-5/6} = 0{,}1509$$

Observamos que Poisson da una buena aproximación, aunque ni n es grande ni p muy pequeña.

¿Será la función de Poisson una auténtica función de probabilidad?

Esta función tiene como dominio el conjunto 0, 1, 2... Quiere decir que la variable aleatoria puede tomar infinitos valores. La suma de probabilidades debe dar 1.

Tenemos la serie:

$$e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2}{2!}e^{-\lambda} + \frac{\lambda^3}{3!}e^{-\lambda} + \ldots =$$

$$= e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots\right) = e^{-\lambda}e^{\lambda} = 1$$

Ahora, hallaremos el valor medio de esta variable aleatoria X, según Poisson. Nos ayudará la tabla:

X	0	1	2	3	...
p	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2!}e^{-\lambda}$	$\frac{\lambda^3}{3!}e^{-\lambda}$	...

Podemos calcular:

$$E(X) = 0e^{-\lambda} + 1\lambda e^{-\lambda} + 2\frac{\lambda^2}{2!}e^{-\lambda} + \frac{\lambda^3}{3!}e^{-\lambda} + \ldots =$$

$$= \lambda e^{-\lambda}\left(1 + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots\right) = \lambda e^{-\lambda}e^{\lambda} = \lambda = np$$

Vemos que el valor medio coincide con el valor esperado de la binomial que es la madre de Poisson.

Varianza en la distribución de Poisson

Dada la tabla:

X	1	2	3	4
p	0,1	0,3	0,2	0,4

Calculemos la varianza.

1) $E(X) = 1(0{,}1) + 2(0{,}3) + 3(0{,}2) + 4(0{,}4) = 2{,}9$

2) $V(X) = (1-2{,}9)^2 0{,}1 + (2-2{,}9)^2 0{,}3 + (3-2{,}9)^2 0{,}2 + (4-2{,}9)^2 0{,}4$

También podríamos haber hecho:

$$V(X) = E(X-m)^2 = E(X^2 - 2mX + m^2) = E(X^2) - 2mE(X) + m^2$$
$$V(X) = E(X^2) - 2m^2 + m^2 = E(X^2) - m^2$$

X^2	1	4	9	16
p	0,1	0,3	0,2	0,4

$$E(X^2) = 1(0{,}1) + 4(0{,}3) + 9(0{,}2) + 16(0{,}40) = 9{,}5$$
$$V(X) = E(X^2) - m^2 = 9{,}5 - 2{,}9^2 = 1{,}09$$

Para la distribución de Poisson, tendremos:

$$E(X^2) = e^{-\lambda}\left(0^2 + 1\lambda + 4\frac{\lambda^2}{2!} + 9\frac{\lambda^3}{3!} + \ldots\right)$$
$$E(X^2) = \lambda e^{-\lambda}\left(1 + 2\lambda + \frac{3}{2}\lambda^2 + \frac{2}{3}\lambda^3 + \ldots\right)$$

El último paréntesis es igual a $(1+\lambda)\ e^{\lambda}$.

Entonces:

$$E(X^2) = \lambda(1+\lambda)$$
$$V(X) = E(X^2) - m^2 = \lambda(1+\lambda) - \lambda^2 = \lambda$$

En consecuencia, la varianza en la distribución de Poisson es λ.

En la distribución de Poisson, el valor medio y la varianza son iguales a λ = np.

Durante un año de 365 días, se han registrado las llegadas de buques al puerto de Buenos aires, obteniéndose la tabla:

X= N° de buques por día	N° de días
0	1
1	2
2	7
3	18
4	32
5	44
6	61
7	56

X= N° de buques por día	N° de días
8	45
9	34
10	28
11	15
12	11
13	7
14	3
15	1

La tabla nos dice que hubo 7 días en los cuales llegaron dos buques; hubo 56 de 7 buques por día, 26 días de 10 buques, etc.

Vamos a construir un modelo de Poisson con esos datos.

Como no tenemos el valor de λ lo estimamos, calculando el valor medio de la tabla.

$$\overline{X} = (0{,}1 + 1{,}2 + \ldots + 15{,}1) \ / \ (1 + 2 + 7 + \ldots + 7 + 3 + 1)$$

$$\overline{X} = 7{,}03$$

Construiremos la tabla de frecuencias esperadas, según el modelo de Poisson, tomando $\lambda = 7{,}03$.

X	$P = \frac{X}{X!} e^{-\lambda}$ Modelo Poisson	Frecuencias esperadas P x 365
0	0,0008	0,323
1	0,0062	2,271
2	0,0218	7,981
3	0,0512	18,703
4	0,0900	32,871
5	0,1266	46,217
6	0,1484	54,150
7	0,1490	54,383
8	0,1309	47,789

X	$P = \frac{X}{X!} e^{-\lambda}$ Modelo Poisson	Frecuencias esperadas P x 365
9	0,1023	37,328
10	0,0719	26,242
11	0,0459	16,771
12	0,0269	9,825
13	0,0146	5,313
14	0,0073	2,668
15	0,0034	1,250
16	0,0015	0,549
17	0,0006	0,227
	0,9993	364,858

Observemos que las tablas de frecuencias esperadas y observadas concuerdan bastante (la única diferencia algo pronunciada se da para X = 6). El modelo de Poisson parece ser adecuado. Se trata de una tabla con n bastante grande y p pequeña; en efecto:

$$\lambda = np \qquad 7{,}03 = 365\ p \qquad p = 0{,}02$$

Un estudio como el anterior tiene consecuencias prácticas; puede emplearse para diseñar servicios y analizar costos de operación.

Supongamos que las autoridades estiman que, debido a fluctuaciones del comercio, se espera un incremento del 15% en la llegada de buques. Parece razonable afirmar que el proceso seguirá aproximadamente el modelo de Poisson. Si el año anterior llegaron 2567 buques, esperamos 2952 para el próximo año. Si usted calcula, obtendrá $\lambda = 8{,}09$. ¿Cuál es el porcentaje de aumento de λ?

Como nos han informado que el puerto tiene dificultades para operar encima de 12 llegadas diarias, nos interesan los valores superiores a 12.

$$P_{13} = \frac{8{,}09^{13}}{13!} = 0{,}031 \qquad 0{,}031.365 = 11{,}42$$

$$P_{14} = \frac{8{,}09^{14}}{14!} e^{-8{,}09} = 0{,}018 \qquad 0{,}018.365 = 6{,}60$$

$$P_{15} = \frac{8{,}09^{15}}{15!} e^{-8{,}09} = 0{,}0098 \qquad 0{,}098.365 = 3{,}56$$

$$P_{16} = 0{,}0049 \qquad 0{,}0049.365 = 1{,}80$$

$$P_{17} = 0{,}0023 \qquad 0{,}0023.365 = 0{,}86$$

$$\overline{24{,}24}$$

Quiere decir que, en alrededor de 24 días, tendremos dificultades operativas. El año anterior se presentaron solamente 9 días “difíciles”. Es interesante observar que un incremento de las llegadas del 15% produjo un crecimiento de los inconvenientes del 167%.

- *Señor Modulador, tanto le decimos a los alumnos que no respondan “6,5 botellas” o “3,4 obreros” y ahora me sale con 6,60 buques a puerto.*
- *En la Estadística esto es necesario. No olvide que estamos trabajando con un modelo. Lo que usted enseña a sus alumnos es otra cosa: Ellos no pueden responder: “para transportar 300 toneladas de trigo necesito 21,4 camiones”.*
- *Este tema me resultó muy interesante pero... es evidente que, sin el conocimiento de la noción de límite y el número e, no podríamos encararlo.*
- *Paso a paso, podríamos arreglárnoslas con una tabla de números aleatorios.*

Números aleatorios

Si colocamos en un bolillero diez bolillas numeradas de 0 al 9 y las extraemos con reposición, anotando los números obtenidos, tendremos una sucesión de números aleatorios:

8, 4, 2, 0, 8, 0, 6, 0, 5, 2, 5, 7...

Pero este procedimiento es poco práctico. Podríamos idear una fórmula que permitiese generar números aleatorios; por ejemplo, partir de las cifras de $\sqrt{5}$ que da una calculadora y extraer las sucesivas raíces cuadradas:

2236068   14953488   12228445   1105823 ....

Vemos que aparecen demasiados “unos”. Además, en pocos pasos llegaremos a 1,00000... ¿por qué?

¿Por qué nos importa que aparezcan muchos unos? En una tabla de números aleatorios, "a la larga", debe aparecer aproximadamente cada dígito el mismo número de veces.

Uno de los primeros algoritmos sugeridos para obtener números aleatorios fue el siguiente: Tomar un número de cuatro cifras, elevarlo al cuadrado y quedarse con las cuatro cifras interiores. Luego, repetir la operación. Partiendo del número "semilla" 8237 obtenemos:

8237	8481	9273	9885	7132	8654	8917	5128	2693	7793
7308	4068	5486	0961	9235	2852	1339	7929	8690	5161
6359	4368	0794	6304	7404					

Esta tabla tiene apenas 100 números; de todos modos la sometemos a prueba:

Dígito	0	1	2	3	4	5	6	7	8	9
Frecuencia	7	8	9	12	9	8	10	10	14	13

El resultado es relativamente bueno; pero, la prueba no es suficientemente larga: Debe tener dígitos en cantidades aproximadamente iguales.

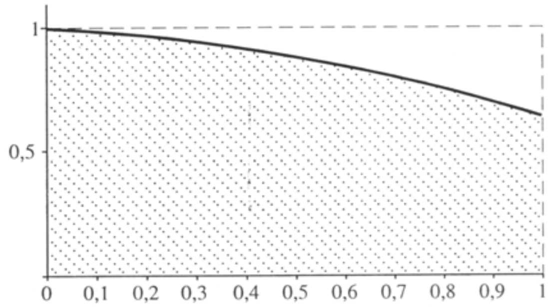
Advertimos que el método depende de la elección del número semilla; hay números sencillos que rápidamente conducen a una repetición de los números. Las calculadoras de bolsillo permiten obtener números aleatorios para lo cual emplean fórmulas complicadas. Se entiende que los números obtenidos mediante fórmulas son números seudoaleatorios, en el sentido de que quien dispusiese de la fórmula podría anticipar los resultados.

Cálculo de integrales mediante números aleatorios

Vamos a calcular el área de la región encerrada por la gráfica $f(x) = e^{-\frac{x^2}{2}}$ con el eje x en el intervalo [0,1].

x	$f(x) = e^{-\frac{x^2}{2}}$
0,05	0,9987
0,10	0,9950
0,15	0,9888
0,20	0,9801
0,25	0,9692
0,30	0,9560
0,35	0,9406
0,40	0,9231
0,45	0,9037
0,50	0,8825

x	$f(x) = e^{-\frac{x^2}{2}}$
0,55	0,8596
0,60	0,8352
0,65	0,8096
0,70	0,7827
0,75	0,7548
0,80	0,7261
0,85	0,6968
0,90	0,6670
0,95	0,6368
1	0,6065



¿Cómo vamos a calcular el integral? Tirando al blanco, a lo loco, sin hacer puntería pero con la suficiente para no salirnos de la zona 1 x 1. Para tirar, empleamos una tabla de números aleatorios o una tecla RAN.

Tenemos a:

5295	x	0969	x
5443	x	6088	x
4073	x	1373	x
1750	x	5124	x
1526	x	4071	x
3806	x	3728	x
1971	x	3354	x
8951	x	6991	No
1024	x	8340	No
4729	x	9362	No

¿Qué hicimos? Analizamos si nuestros “tiros” pegaron en la zona punteada. Las cruces indican impactos.

El área estimada es $\frac{17}{20} = 0{,}85$

El resultado mejoraría si, en lugar de 20 tiros (subintervalos de longitud 0,05), hubiésemos hecho 100 “tiros” (subintervalos de longitud 0,01). Esto puede hacerse fácilmente con el auxilio de una computadora.

El número aleatorio de Champernowne[1]
En la década del ’30, Champernowne propuso el siguiente número aleatorio 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18... Como vemos, basta escribir la sucesión de números naturales; esta sucesión, si bien al principio se desvía de un comportamiento normal (hay números que aparecen con mayor frecuencia), luego, si se consideran muchas cifras, registra un comportamiento aleatorio muy bueno.

Los números aleatorios como los que se obtienen mediante un programa –o el número aleatorio que acabamos de ver– son aleatorios en sentido amplio; no podrían ser empleados en un juego, porque alguien podría descubrir la forma de generarlos, y para él serían previsibles y no aleatorios.

Integremos contenidos a través de estos problemas:

En una caja hay 19 bolitas blancas y una azul. Se hacen 10 extracciones con reposición. ¿Cuál es la probabilidad de obtener, exactamente, 2 azules?

a. Emplear la fórmula binomial,
b. Aplicar la distribución de Poisson.

[1] Investigación y Ciencia*; enero 1980.*

a. $\binom{10}{2}\left(\frac{1}{20}\right)^2\left(\frac{19}{20}\right)^8 = 0{,}0746$

b. $np = 10.\frac{1}{20} = 0.5 = \lambda \qquad P_2 = \frac{\lambda^2}{2!}e^{-\lambda} = \frac{0{,}5^2}{2!}e^{-0{,}5} = 0{,}0758$

Se observa que, aunque n=10 no es muy grande, la aproximación es buena (La probabilidad de azul es pequeña).

Si tiramos un dado 12 veces, ¿cuál es la probabilidad de obtener, exactamente, tres "ases"?

a. Emplear la fórmula binomial,
b. Aplicar la distribución de Poisson.

a. $\binom{12}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^9 = 0{,}1974$

b. $np = 12.\frac{1}{6} = 2 = \lambda \qquad P_3 = \frac{2^3}{3}e^{-2} = 0{,}1804$

En este caso, los resultados difieren bastante; no es aconsejable emplear la distribución de Poisson cuando p es grande.

Un periódico se imprime en forma automática. La probabilidad de que una palabra aparezca con defectos de impresión es 0,0001. Cada edición tiene 30000 palabras. ¿Cuál es la probabilidad de que en una edición aparezcan más de cuatro palabras con defectos de impresión?

$\lambda = np = 30000.10^{-4} = 3$

$P_0 + P_1 + P_2 + P_3 + P_4 = e^{-3}\left[1 + 3 + \frac{9}{2} + \frac{27}{6} + \frac{81}{24}\right] = 0{,}815$

$P(X > 4) = 1 - 0{,}815 = 0{,}185$

El 2% de las piezas producidas por una fábrica es defectuoso. Si examinamos una muestra de 150 piezas, ¿cuál es la probabilidad de que:

a. ¿Ninguna pieza sea defectuosa?
b. ¿2 sean defectuosas?
c. ¿Más de 3 sean defectuosas?

$$\lambda = np \qquad \lambda = 0{,}02(150) = 3$$

$$P_0 = e^{-3} = 0{,}0498$$

$$P_1 = 3e^{-3} = 0{,}1494$$

$$P_2 = \frac{3^2}{2} e^{-3} = 0{,}2241$$

$$P_3 = \frac{3^3}{3!} e^{-3} = 0{,}2241$$

$$P(X > 3) = 1 - (0{,}0498 + 0{,}1494 + 0{,}4482) = 0{,}3526$$

No nos cansaremos de repetir que una buena estrategia de enseñanza implica no cerrar capítulos y abandonarlos: *Ya pasó el cálculo combinatorio, ya pasó la noción de independencia de sucesos...*

Por estas razones, proponemos resolver:

Actividad 8.1.

Entre 10 abogados y 7 ingenieros debemos elegir una comisión de 4 abogados y 3 ingenieros. ¿Cuántas son las elecciones posibles?

Actividad 8.2.

Entre los 10 abogados del problema anterior hay 6 cuyos apellidos comienzan en vocal; y, de los 8 ingenieros, hay 5 cuyos apellidos comienzan en vocal. ¿Cuál es la probabilidad de que la comisión de 4 abogados y 3 ingenieros esté formada sólo por personas cuyos apellidos comienzan en vocal, si la elección es al azar?

Actividad 8.3.

Entre 8 libros, hay 3 repetidos. ¿Cuál es la probabilidad de que, si elegimos de entre los 8, cuatro al azar; no haya repetidos?

Actividad 8.4.

En una compañía están vacantes los cargos de gerente general, gerente de exportación y gerente de comercialización. De entre 20 personas ¿cuál es el número de elecciones posibles, si una persona no puede detentar más de un cargo?

Actividad 8.5.

En una caja ponemos 2 bolitas verdes y 4 amarillas. Extraemos una bolita y, luego, la reponemos. Hacemos lo mismo tres veces. Sea X = número de amarillas, la variable aleatoria. Representar la función de probabilidad y la función de distribución.

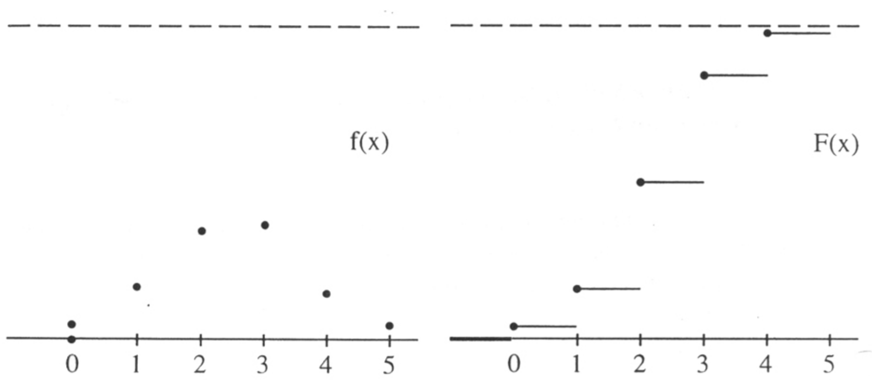
Actividad 8.6.

Un tenista A ha vencido 9 veces a B en 10 partidos. ¿Cuál es la probabilidad de que esto haya sucedido por pura casualidad?

9. Funciones de distribución

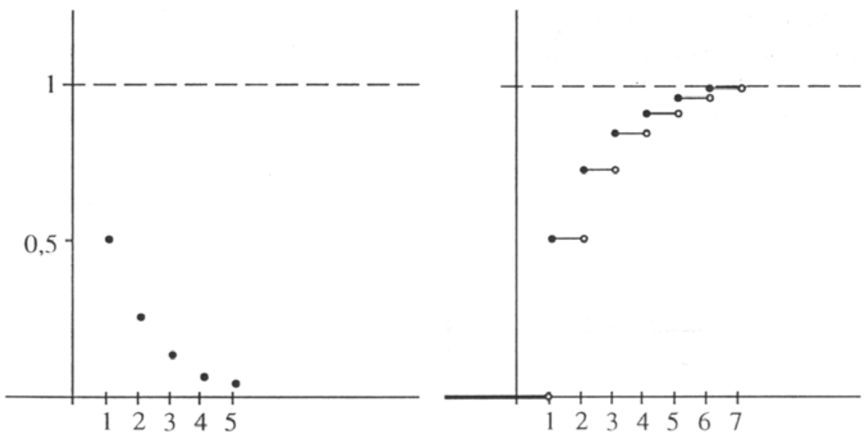
Funciones de distribución discretas y continuas

Hasta ahora hemos considerado variables aleatorias que toman un número finito de valores. Por ejemplo, si tiramos 5 veces una moneda el número x de caras puede ser 0, 1, 2, 3, 4, 5. La función de probabilidad es discreta y la función de distribución presenta “saltos”, según hemos visto:



Consideremos un experimento algo distinto. Tiramos una moneda hasta que salga “cara”. Anotamos, entonces, el número de tirada que corresponde a la salida de la primera cara. En este caso, la variable x puede tomar los valores 1,2,3,4... Aunque sea muy improbable, la primera cara podría salir en la tirada n° 20 o en la n° 1000... Por ejemplo, la probabilidad de que salga la primera cara en el quinto tiro es:

$$\left(\frac{1}{2}\right)^5 = \frac{1}{32}$$



Aunque, en este caso, el dominio de la función de probabilidad es infinito, la función es discreta porque no puede tomar valores intermedios como $\frac{1}{2}, \frac{2}{3}$, etc.

Nos gustaría comprobar que la suma de probabilidades es 1. Tenemos $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\ldots$

¿Se acuerda de las series geométricas? La suma parcial enésima corresponde a la suma de n términos de una progresión geométrica.

$$S_n = a_1 \frac{1-q^n}{1-q}$$

En este caso:

$$S_n = \frac{1}{2}\frac{1-\left(\frac{1}{2}\right)^n}{1-\frac{1}{2}}$$

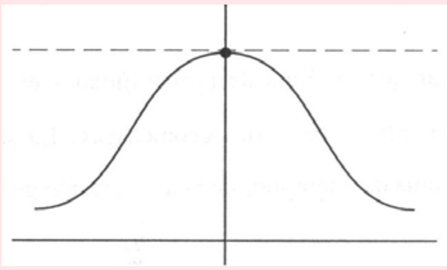
Para tener la suma de la serie, calculamos: $S_n = \frac{1}{2}\ \frac{1-0{,}5^n}{0{,}5} = 1$

Por lo tanto, la suma de probabilidades es 1.

En las funciones de distribuciones discretas se produce un "salto" cada vez que se acumula una probabilidad. Estas funciones son siempre continuas a la derecha y discontinuas a la izquierda, en todo punto en el cual la probabilidad es distinta de cero.

Existen funciones de probabilidad continuas; por ejemplo, el porcentaje de CO_2 presente en la atmósfera puede considerarse como una variable aleatoria que puede tomar todos los valores reales dentro de un cierto intervalo.

Consideramos la función de probabilidad $f(x) = \frac{1}{\pi} \cdot \frac{1}{(1+x^2)}$ definida en todo el eje real.



En relación con estas funciones de probabilidad continuas, se presentan varios interrogantes interesantes:

¿Cómo saber si la suma de las probabilidades es 1?

Recordemos la definición de función de distribución: $F(x) = P(X \le x)$

En nuestro ejemplo, F (2) es la probabilidad de que la variable aleatoria X tome cualquier valor comprendido entre $-\infty$ y 2; en este caso continuo, la suma es ahora una integral:

$$F(2) = \int_{-\infty}^{2} f(x)dx = \int_{-\infty}^{2} \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{1}{\pi} \operatorname{arctg} x \Big|_{-\infty}^{2} =$$

$$= \frac{1}{\pi} \operatorname{arctg} 2 - \frac{1}{\pi} \operatorname{arctg} (-\infty) = \frac{1}{\pi} . 1{,}107 - \frac{1}{\pi}\left(-\frac{\pi}{2}\right) = 0{,}85$$

La calculadora debe colocarse en “RADIANES”. El lector puede comprobar que f (x) cumple la condición $\int_{-\infty}^{\infty} f(x) = 1$

En las funciones de probabilidad continuas, la probabilidad de x = a es igual a cero. Por esta razón, las preguntas más frecuentes son del tipo: ¿Cuál es la probabilidad de que x esté comprendida entre 1 y 2,8?

- *Señor Modulador, pongo a prueba mis conocimientos. He entendido que, en el caso continuo, f (x) debe ser la derivada de F (x).*
- *Parece cierto.*
- *No entiendo cómo la probabilidad de una función de un punto dado sea igual a cero.*
- *Compare los gráficos de una función discreta y una continua.*

  ....
- *En el caso discreto, observo que cada vez que aparece un punto con cierta probabilidad, la función de distribución pega un salto. Si, en el caso continuo, un punto tuviese probabilidad distinta de cero, la función de distribución saltaría y... no habría continuidad.*
- *Me asombra que, con mis pobres explicaciones, haya comprendido tanto.*
- *No se admire, lo que no comprendo es cómo, si sumo probabilidades iguales a cero puedo llegar al ansiado 1.*
- *Estimado, obispo Berkeley, el problema que presenta es el de la integral, o el del Aquiles y la tortuga, el problema del límite.*

  ....

  Debo repasar mi análisis; pero, quiero preguntar: ¿Cómo definimos la función de distribución en el caso continuo?

Si f (x) es la función de probabilidad de una variable aleatoria X, su función de distribución es:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\ dt$$

Esperanza matemática y varianza para distribuciones continuas

Para funciones continuas, definimos:

$$E(x) = \int_{-\infty}^{\infty} xf\ (x)\ dx \qquad E(x)^2 = \int_{-\infty}^{\infty} x^2 f(x)\ dx$$

$$D^2(x) = \int_{-\infty}^{\infty} (x-m)^2 f\ (x)\ dx$$

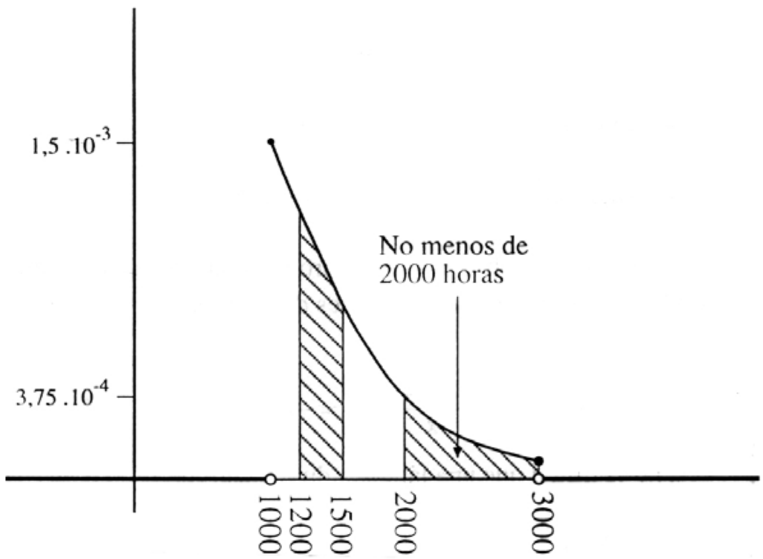
La duración, en horas, de las lámparas fabricadas por una empresa es una variable aleatoria cuya función de densidad de probabilidad es $f(x) = \frac{1500}{x^2}$. Si $1000 \leq 3000$, siendo f(x) = 0 para x no perteneciente a ese intervalo, calcular la probabilidad de que:

a. una lámpara tomada al azar tenga una duración comprendida entre 1200 y 1400 horas;
b. una lámpara tomada al azar dure no menos de 2000 horas.
c. Calcular la duración media.

$$a.\ P(1200 \leq X \leq 1400) = \int_{1200}^{1400} \frac{1500}{x^2} dx = -\frac{1500}{x}\Bigg|_{1200}^{1400} = -1{,}07 + 1{,}25 = 0{,}18$$

$$b.\ P(\geq 2000) = \int_{2000}^{3000} \frac{1500}{x^2} dx = -\frac{1500}{x}\Bigg|_{2000}^{3000} = -0{,}5 + 0{,}75 = 0{,}25$$

$$c.\ E(X) = \int_{-\infty}^{+\infty} xf\ (x)\ dx = \int_{1000}^{3000} x.\frac{1500}{x^2} dx = \int_{1000}^{3000} \frac{1500}{x} dx \cong 1648$$



- *Con relación al problema anterior... Seguramente, lo tomó de la propaganda de las lámparas "Penunbrillo". Mire el folleto. ¿No es hermoso? Aquí dice: "No se engañe, señor; nosotros lo engañamos mejor".*
- *Cosas de la propaganda, usted sabe... El caso es que, he comprado... la señorita que hacía la propaganda por TV de las "Penumbrillo" era tan... ¿Dónde estábamos?*
- *Me decía que había comprado...*
- *5 lámparas de esa marca y duraron menos de 1 300 horas cada una.*
- *Vaya a saber cómo registró la duración.*
- *De eso se encargó el abuelo que es inventor y muy minucioso; por ese lado, estoy seguro. Bueno, habré tenido mala suerte. Hablemos de otra cosa.*
- *No; ahora que la cosa se está poniendo interesante...*
- *Y, ¿qué podemos hacer?*
- *Vamos a medir su mala suerte.*
- *Mire que es gracioso, señor Modulador.*
- *Manos a la obra... Calcule la probabilidad de que la lámpara dure menos de 1300 horas, según las afirmaciones de la empresa.*
- $\int_{1000}^{1300} \frac{1500}{x^2} dx = \left.\frac{-1500}{x}\right|_{1000}^{1300} = -1{,}15 + 1{,}5 = 0{,}35$. *Creí que era menos; entonces, no es difícil que una lámpara dure menos de 1300 horas. No pensemos más.*
- *Pero, usted me dijo que esto había sucedido con 5 lámparas. ¿Puede calcular la probabilidad de lo que le pasó a usted con su adquisición? Si no puede hacerlo, la verdad que no sé para qué hace el curso.*
- *No me agravie porque, ¿a quién deberíamos echarle la culpa? Pero, bien que sé responder. Si, para una lámpara la probabilidad de menos de 1300 horas de vida es 0,35; para 5 lámparas (es evidente que la duración de una no influye sobre la otra) es $0{,}35^5 = 0{,}00525$*
- *Su mala suerte parece increíble; porque, la probabilidad de que cinco lámparas hayan estado por debajo de 1300 horas es muy baja. Más bien, habría que pensar que las afirmaciones de la publicidad de "Penumbrillo" son falsas.*
- *Para empezar, les escribiré una carta no muy cordial. Pero, ahora, me gustaría calcular la varianza en el problema de las lámparas.*

$$V(X) = \int_{1000}^{3000} (x-m)^2 f(x)dx = \int_{1000}^{3000} (x-1648)^2 \frac{1500}{x^2} dx =$$

$$= 1500 \int_{1000}^{3000} \left(1 - \frac{3296}{x} + \frac{1648^2}{x^2}\right) dx = 1500 \left[x - 3296 \ \text{In} \ x - \frac{1648^2}{x} \right]_{1000}^{3000}$$

$$= 283500$$

El desvío es $\sigma = \sqrt{283500} = 532$.

Teorema de Tchbyccheff

Sea X una variable aleatoria con valor medio m y desviación σ.

Entonces:

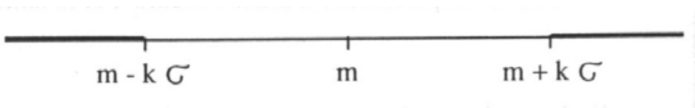
$$P(|X - m| > k\sigma)\frac{1}{k^2}$$

(k es un número positivo arbitrario.)

Leamos con cuidado. Sabemos que $|X - m| \leq k\sigma$ implica:

$$-k\sigma \leq X - m \leq k\sigma \ \ (\text{es decir}: m - k\sigma \leq x \wedge X \leq +k\sigma)$$

Entonces $|X - m| > k\sigma$ implica $X < m - k\sigma \wedge X > m + k\sigma$



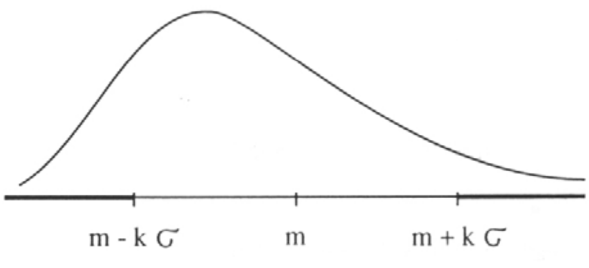
El teorema se refiere a la probabilidad de que la variable aleatoria X tome valores en la zona "oscura", alejados del valor medio.

Si X tiene una distribución continua, podemos escribir:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - m)^2 f(x) dx$$

Como sólo nos interesa la zona oscura:

$$\sigma^2 > \int_{-\infty}^{m-k\sigma} (x - m)^2 f(x) dx + \int_{m+k\sigma}^{\infty} (X - m)^2 f(x) dx$$



Además, en esa zona se verifica que $|X - m| > k\sigma$ equivalente a $(X - m)^2 k^2 \sigma^2$

En consecuencia:

$$\sigma^2 > k^2\sigma^2 \left(\int_{-\infty}^{m-k\sigma} f(x) dx + \int_{m+k\sigma}^{\infty} f(x) dx \right)$$

Pero, el paréntesis es, justamente, la probabilidad de que el valor que tome la variable aleatoria pertenezca a la zona oscura. Por lo tanto: $\frac{1}{k^2} > P(|X - m| > k\sigma)$

En lenguaje simple, este teorema nos dice que es improbable que la variable aleatoria tome valores alejados del valor medio. Cuanto mayor sea k (más lejos de m), menor es la probabilidad de que se dé este resultado.

En una caja hay 3 bolitas blancas y 2 azules. Si se hacen 1000 extracciones con reposición, ¿qué afirma la desigualdad de Tchebyccheff acerca de la probabilidad de que el número de azules esté comprendido entre 369 y 431?

Si X = n° azules

$$E(X) = np = 1000.(0,4) = 400$$

$$\sigma = \sqrt{npq} = \sqrt{1000\ (0,4)\ (0,6)}$$
$$\sigma = 15,5$$

$$369 = 400 - 2\sigma$$
$$431 = 400 + 2\sigma$$

La desigualdad de Tchebyccheff afirma que:

$$P(|X - m| > 2\sigma) < \frac{1}{4} \qquad \text{Entonces } P(|X - m| \le 2\sigma) \ge \frac{3}{4}$$

Quiere decir que la probabilidad de que el número de azules esté entre 369 y 431 es mayor que $\frac{3}{4}$.

En este ejemplo, la variable aleatoria es discreta. La desigualdad puede probarse, también, en este caso, empleando sumas en lugar de integrales.

Volvemos a la demostración anterior.

Escribamos $\sigma^2 = \int_{-\infty}^{\infty} (x - m)^2 f(x)dx$

Empleando la hipótesis $|x - m| > k\sigma$, resulta :

$$\sigma^2 > \int_{-\infty}^{\infty} k^2\sigma^2 f(x)dx \qquad \sigma^2 > k^2\sigma^2 \int_{-\infty}^{\infty} f(x) \qquad \text{pero} \int_{-\infty}^{\infty} f(x)dx = 1$$

por ser f(x) una función de probabilidad.

$$\sigma^2 > k^2\sigma^2 \ (\text{absurdo si } k > 1)$$

¿Dónde está el error?

Teorema de Bernoulli

Si efectuamos n pruebas repetidas independientes, en las cuales la probabilidad p de un suceso A es constante, podemos tomar como variable aleatoria X la frecuencia relativa f/n (Por ejemplo, si tiramos una moneda y salen 40 caras, la frecuencia relativa es 40/87).

Recordemos que E (f) = np y que, por lo tanto, $E\left(\frac{f}{n}\right) = p$

Por otra parte, la varianza V (f) = npq

Se puede probar que $V(f\ /\ n) = \frac{pq}{n}$

La desigualdad de Tchebyccheff afirma que $P(|\, x - m\, |\geq k\sigma) < 1/k^2$

En particular, para X = f/n, tenemos:

$$P\left(\left|\frac{f}{n} - p\right| > k\sqrt{\frac{pq}{n}}\right) < \frac{1}{k^2}$$

Si hacemos $k\sqrt{\frac{pq}{n}} < \varepsilon$,tendremos :

$$P\left(\left|\frac{f}{n} - p\right| > \varepsilon\right) < \frac{pq}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

fácil probar que $pq \leq 1/4.$

En el idioma habitual, el teorema dice que:

La probabilidad de que la frecuencia relativa difiera considerablemente de la probabilidad es muy pequeña, cuando el número n de pruebas es grande.

Este resultado responde al problema que tanto preocupó a nuestro colaborador Bernoulli y del que hablamos en el primer módulo.

A Jacobo Bernoulli la prueba le resultó más difícil, porque no disponía de la desigualdad de Tchebyccheff. Y, ¿cómo no se le ocurrió a Bernoulli la desigualdad? Es muy fácil comentar el partido después de jugado. ¿Hacemos ver esto a nuestros alumnos? ¿Les hacemos comprender que los problemas son interesantes y que el premio está en la búsqueda? Que todos podemos intentar resolverlos...

La prueba de Bernoulli fue publicada en 1713, después de su muerte, en su obra *Ars Conjectandi.*

Actividad 9.1.

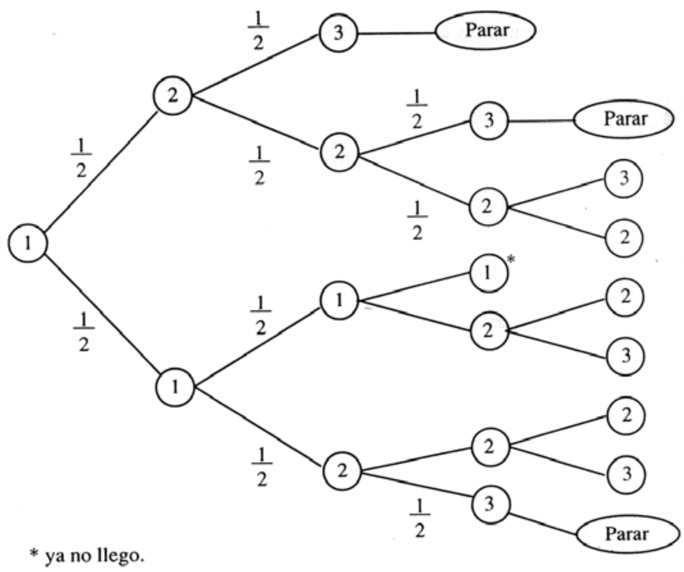
1. Encuentre ejemplos de sucesos excluyentes o incompatibles. Busque contraejemplos.
2. ¿Es siempre verdadera la igualdad ¿Puede dar un contraejemplo?
3. ¿Cómo explicaría la probabilidad condicional e independencia de sucesos?
4. Tiramos una moneda cinco veces:
   a. ¿Cuál es la probabilidad de obtener 3 caras y 2 cruces?
   b. ¿Cuál es la probabilidad de obtener al menos una cara?
   c. ¿Cuál es la probabilidad de obtener 3 caras y 2 cruces, si sabemos que salió cruz en el primer tiro?

Actividad 9.2.

De entre cuatro chinos y seis nigerianos, debemos elegir una comisión de 3 personas. ¿Cuál es la probabilidad de que la comisión esté formada por dos chinos y un nigeriano? Resolver el problema de dos maneras diferentes.

Vamos a tirar una moneda con certificado de buena conducta. Si, al llegar al quinto tiro o antes, obtenemos tres caras, ganamos. Ya hemos hecho un tiro y ha salido cara. ¿Cuál es la probabilidad de que ganemos?

En el diagrama, los números indican cuántas caras tenemos:



La probabilidad de llegar a 3 (¡Gano!) es:

$$\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{8}=\frac{11}{16}$$

Haga el mismo problema empleando la fórmula binomial

La distribución de Poisson y el flujo de sucesos

Consideremos un conmutador telefónico al que pueden llegar o no llamadas en un instante dado. A partir de un cierto instante inicial, comenzamos a contar las llamadas. Nos interesaría obtener una función que nos diera la probabilidad de un número k de llamadas en función del tiempo.

Como hacemos siempre, haremos ciertas simplificaciones para poder abordar el problema.

Supondremos que la probabilidad de una llamada exactamente en el lapso Δt es igual a $\lambda(\Delta t)$ donde λ es constante. (Ésta es una simplificación porque λ podría ser variable). Por otra parte, como Δt se supone pequeño, la probabilidad de más de una llamada puede despreciarse. Designaremos con $P_0(t)$ a la función que da la probabilidad de que no haya llamadas al llegarse a un instante t (a partir del inicial).

Por lo tanto, $P_0(t + \Delta t)$ es la probabilidad de que no haya llamadas hasta el instante $(t + \Delta t)$. Podemos escribir:

$$P_0(t+\Delta t) = P_0(t)(1-\lambda\Delta t) \qquad (1)$$

Para que no haya llamadas hasta el instante $t + \Delta t$, no deben haberse producido llamadas hasta el instante t y tampoco debe haber llamadas entre t y $t + \Delta t$. Debemos multiplicar las probabilidades de estos sucesos como hicimos en (1), porque suponemos que esos sucesos son independientes.

A partir de (1), podemos escribir:

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

Si hacemos que Δt tienda a cero, en el primer miembro tenemos la derivada:

$$\frac{dP_0(t)}{P_0(t)} = -P_0 dt$$

Nos ha nacido una ecuación diferencial como sucede en los cursos de Física o de Química. Obsérvese la diferencia con lo que suele suceder en muchos cursos de matemática: *Resuelva, efectúe, realice, calcule...*

$$\frac{dP_0(t)}{P_0(t)} = -\lambda dt$$

Integramos:

$$\ln P_0(t) = -\lambda t + C$$

$$P_0(t) = e^{-\lambda t + C}$$

Tenemos que determinar la constante C. Sabemos que $P_0(0) = 1$ porque la probabilidad de 0 llamadas para t = 0 debe ser 1. Esto se cumple si C = 0. Por lo tanto: $P_0(t) = e^{-\lambda t}$.

Obsérvese que esta fórmula es muy aceptable porque, como todas las fórmulas "habla", y nos dice que, a medida que pasa el tiempo, la probabilidad de que no haya llamadas tiende a cero.

Tratemos de determinar $P_1(t)$

$$P_1(t+\Delta t) = P_1(t)(1-\lambda\Delta t) + P_0(t)\lambda(\Delta t)$$

¿Qué nos dice esta desigualdad? Puede suceder que hasta t haya una llamada y no haya llamadas entre t y $t+\Delta t$; pero, también puede darse el caso de que hasta t no haya llamadas y aparezca una entre t y $t+\Delta t$.

$$\frac{P_1(t+\Delta t) - P_1(t)}{\Delta t} = -\lambda P_1(t) + P_0(t)\lambda$$

Tomando límite para Δt tendiendo a cero, nos queda:

$$\frac{dP_1(t)}{dt} + \lambda P_1(t) = \lambda e^{-\lambda t} \quad (*)$$

Donde sustituimos $P_0(t) = e^{-\lambda t}$

La ecuación es del tipo: $\dfrac{dy}{dx} + yP(x) = Q(x)$

La solución de (*) es $P_1(t) = (\lambda t)e^{-\lambda t}$ (Pruebe si es solución, sustituyendo). Un análisis análogo lleva a los siguientes resultados:

$$P_2(t) = \frac{(\lambda t)^2}{2!}e^{-\lambda t} \qquad P_3(t) = \frac{(\lambda t)^3}{3!}e^{-\lambda t}$$

$$\boxed{P_n(t) = \frac{(\lambda t)^n}{n!}e^{-\lambda t}}$$

Una experiencia de Rutherford y Geiger

En muchas cuestiones de la Física se dan excelentes condiciones para la aplicación del cálculo de probabilidades, porque el número de partículas es muy grande.

Rutherford y Geiger observaron el número de partículas alfa desprendido de una sustancia radioactiva en intervalos de 7,5 segundos. Hicieron 2608 observaciones independientes y obtuvieron la tabla:

X = N° de partículas emitidas	N° de períodos observados
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	2

La tabla nos dice que, por ejemplo, hubo 525 períodos donde se observó la emisión de 3 partículas alfa.

Pedimos al lector, como ejercicio, que:

1. Obtenga el valor medio (de la tabla) para estimar λ.
2. Emplee ese λ para calcular las frecuencias esperadas, según el modelo de Poisson.

Observará un buen acuerdo entre las frecuencias esperadas y las observadas.

Haremos una parte:

$\overline{X} = 3{,}87$ partículas cada 7,5 segundos.

Para calcular las frecuencias esperadas, empleamos:

$$P_h = \frac{(\lambda t)^h}{h!} e^{-\lambda t}$$

Por ejemplo:

$$P_5 = \frac{(3{,}87)^5}{5!} e^{-3{,}87} = 0{,}1509$$

La frecuencia esperada es 0,1509 x 2608 = 393,54

$$P_9 = \frac{(3{,}87)^9}{9!} e^{-3{,}87} = 0{,}0112 \qquad \text{Frecuencia esperada} = 29{,}20$$

Actividad 9.3.

Dada la función:

$$f(x) = \begin{cases} 0 \ \text{si} \ x < 0 \\ \dfrac{3x^2}{8} \text{si} \ 0 \leq x < 2 \\ 0 \ \text{si} \ 2 \leq 0 \end{cases}$$

Probar que cumple la condición de las funciones de probabilidad:

$$\int_{-\infty}^{\infty} f \ (x) \ dx = 1$$

Calcular E(x) y Var(x)

Actividad 9.4.

Aplicar el teorema de Tchebyccheff para estimar la probabilidad de que, al tirar 120 veces un dado, el número de ases esté entre 11 y 29.

Actividad 9.5.

La probabilidad de efectos nocivos al aplicar una vacuna es de 5 en 10000. ¿Cuál es la probabilidad de que en una muestra de 2000 individuos haya más de dos infectados?

Actividad 9.6.

En un cierto lugar caen estrellas fugaces a razón de 6 por hora. ¿Cuál es la probabilidad de observar exactamente 3 estrellas fugaces en un cuarto de hora?

Clave de respuestas

6.1.

a. $\left(\frac{5}{6}\right)^5$

b. $\left(\frac{1}{6}\right)^5$

c. $\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3$

6.2.

a. $\binom{5}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^4$

b. $\binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$

6.3.

$$\binom{10}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^6$$

6.4.

$$\binom{6}{4}\,0{,}3^4\;0{,}7^6$$

6.5.

$$\binom{8}{3}0{,}5^8+\binom{8}{4}0{,}5^8+\binom{8}{5}0{,}5^8$$

6.6.

$$\binom{5}{0}0{,}5^5;\binom{5}{1}0{,}5^5;\binom{5}{2}0{,}5^5;\binom{5}{3}0{,}5^5;\binom{5}{4}0{,}5^5;\binom{5}{5}0{,}5^5$$

7.1.

a.

$P(X=0)=0{,}00098 \quad P(X=1)=0{,}01465 \quad P(X=2)=0{,}08789$

$P(X=3)=0{,}26367 \quad P(X=4)=0{,}39551 \quad P(X=5)=0{,}23730$

b.

$$\text{Si } -\infty < x < 0 \quad F(X) = 0$$

$$\text{Si } 0 \leq x < 1 \quad F(X) = 0{,}00098$$

$$\text{Si } 1 \leq x < 2 \quad F(X) = 0{,}01563$$

$$\text{Si } 2 \leq x < 3 \quad F(X) = 0{,}10352$$

$$\text{Si } 3 \leq x < 4 \quad F(X) = 0{,}36719$$

$$\text{Si } 4 \leq x < 5 \quad F(X) = 0{,}7627$$

$$\text{Si } X \geq 5 \quad F(X) = 1$$

7.2.

$$E(x) = np = 5 \,.\, 0{,}75 = 3{,}75$$

$$V(x) = npq = 5(0{,}75)(0{,}25) = 0{,}9375$$

$$\sigma = \sqrt{npq} = 0{,}9682$$

7.3.

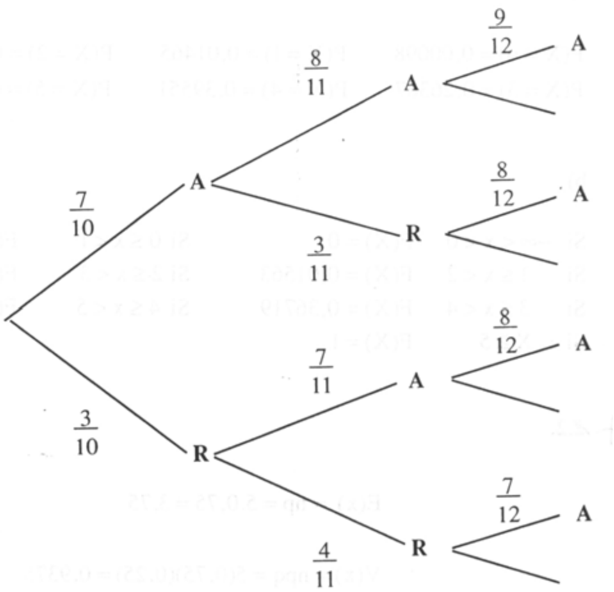
$$\text{a. } 0{,}95^{10} = 0{,}5987$$

$$\text{b. } 1 - \left(0{,}95^{10} + \binom{10}{1} 0{,}95^{9}\ 0{,}05 \right) = 1 - (0{,}5987 + 0{,}3151) = 0{,}0862$$

7.4.

$$\text{a. } \binom{8}{8} 0{,}5^{8} + \binom{8}{7} 0{,}5^{7}\ 0{,}5 = 0{,}0039 + 0{,}03125 = 0{,}03515$$

7.5.



$$\frac{7}{10}\cdot\frac{8}{11}\cdot\frac{9}{12}+\frac{7}{10}\cdot\frac{3}{11}\cdot\frac{8}{12}+\frac{3}{10}\cdot\frac{7}{11}\cdot\frac{8}{12}+\frac{3}{10}\cdot\frac{4}{11}\cdot\frac{7}{12}=0{,}7$$

(Probabilidad de azul en tercera extracción)

8.1.

$$\binom{10}{4}\binom{7}{3}$$

8.2.

$$\binom{6}{4}\binom{5}{3}/\binom{10}{4}\binom{7}{3}$$

8.3.

$$3\binom{6}{4}/\binom{8}{4}$$

8.4. $V_{20,3} = 20 \cdot 19 \cdot 18$

8.5.

$$f(0) = \left(2/6\right)^3$$
$$f(0) = 0{,}0370$$

$$f(1) = 3\left(4/6\right)\left(2/6\right)^2$$
$$f(1) = 0{,}2222$$

$$f(2) = 3\left(4/6\right)^2\left(2/6\right)$$
$$f(2) = 0{,}4444$$

$$f(3) = \left(4/6\right)^3$$
$$f(3) = 0{,}2963$$

$F(x) = 0 \quad$ Si $x < 0$

$F(x) = 0{,}0370 \quad$ Si $0 \le x < 1$

$F(x) = 0{,}2592 \quad$ Si $1 \le x < 2$

$F(x) = 0{,}7036 \quad$ Si $2 \le x < 3$

$F(x) = 1 \quad$ Si $x \ge 3$

8.6. $\binom{10}{9} \; 0{,}5^9 \; 0{,}5 = 0{,}0098$

9.1.
4.

a. $\binom{5}{3} \; 0{,}5^3 \; 0{,}5^2$

b. $1 - P(0) = 1 - 0{,}5^5$

c. $\binom{4}{1} \; 0{,}5^1 \; 0{,}5^3$

9.2.

$$\binom{4}{2}\binom{6}{1}/\binom{10}{3} = 0{,}3$$

$$\left(\underset{CH}{\frac{4}{10}}.\underset{CH}{\frac{3}{9}}.\underset{N}{\frac{6}{8}}\right).3 = 0{,}3$$

Se multiplica por 3 porque N puede salir primera, segunda, tercera.

9.3.

$$E(x) = \int_0^2 x.\frac{3x^2}{8}dx = \frac{3x^4}{32}\Bigg|_0^2 = \frac{48}{32} = 1{,}5$$

$$\text{Var X} = \int_0^2 (x-1{,}5)^2\,\frac{3x^2}{8}\,dx =$$

$$= \frac{3}{8}\int_0^2 (x^2-3x+2{,}25)x^2dx = \frac{3}{8}\int_0^2 (x^4-3x^3+2{,}25x^2)dx =$$

$$= \frac{3}{8}\left[\frac{x^5}{5}-\frac{3x^4}{4}+\frac{2{,}25x^3}{3}\right]_0^2 = \frac{3}{8}[6{,}4-12+6] = 0{,}15$$

9.4.

$$m = 20 \qquad P(|\,x-m\,|\,k\sigma) < \frac{1}{k^2}$$

$$\text{DESV} = \sqrt{120.(1/6).(5/6)} = 4{,}08$$

$$20+2(4{,}08) = 28{,}16 \qquad P(|\,x-20\,|> 8{,}16) < \frac{1}{4}$$

$$20-2(4{,}08) = 11{,}84$$

El teorema de Tchebyccheff “estima” que la probabilidad $11 \le X \le 24$ es, aproximadamente ¾.

9.5.

$$\lambda = np = 2000\ .\ 0{,}0005 = 1$$

$$P(2000,\ 0) = e^{-1} \quad = 0{,}368$$

$$P(2000,\ 1) = e^{-1} \quad = 0{,}368$$

$$P(2000,\ 2) = e^{-1}/2 = 0{,}184$$

$$\text{Suma} = 0{,}92$$

La probabilidad de más de 2 es 0,08.

9.6.
En ¼ hora, el valor medio es 1,5.

$$P(3 \text{ en } 1/4 \text{ hora}) = \frac{1^3}{3!} e^{-1} = 0{,}06$$

MÓDULO 3

En este módulo se introduce la aproximación normal de la distribución binomial. Se trata, también, la correlación entre dos variables, y una introducción a la teoría y práctica con muestras.

Algunas de estas cuestiones quizás no pueden llevarse al aula (salvo, el caso de estudios especializados); pero, es bueno que el profesor tenga un panorama de las posibilidades que ofrecen estos temas.

Como en los módulos anteriores, se insiste en la revisión de temas hasta que queden incorporados de un modo natural y, también, en las diferentes maneras de encarar un mismo problema.

10. La distribución normal

Recordamos la distribución binomial referida a pruebas repetidas con probabilidad constante.

En una caja hay 4 bolitas verdes y 1 blanca. Si se hacen tres extracciones con reposición:

a. ¿Cuál es la probabilidad de obtener exactamente tres verdes?

b. ¿Por lo menos una verde?

a. $\binom{3}{3}\ 0{,}8^3\ 0{,}2^0 = 0{,}512$

b. $1-\binom{3}{0}\ 0{,}8^0\ 0{,}2^3 =$

Al repasar, no se trata dc hacer muchos ejercicios del mismo tipo. Mejor es hacer muchas preguntas sobre un problema. Por ejemplo:

¿Cuál es la probabilidad de 2 verdes y 1 blanca?

$$\binom{3}{2}\ 0{,}8^2\ 0{,}2^1 = 3\ .\ 0{,}64\ .\ 0{,}2 =$$

¿De dónde proviene este cálculo? La probabilidad de la secuencia VVB es 0,8 (0,8) (0,2). ¿Por qué multiplicamos por 3? Porque hay 3 lugares para B.

No podríamos emplear la distribución binomial si no hubiese reposición en cada extracción; porque, en tal caso, la probabilidad no sería constante. Si no hubiese reposición, las respuestas serían:

a. $\frac{4}{5}.\frac{3}{4}.\frac{2}{3} = 0{,}4$

b. $P(0\ \text{verde}) = 0$

Sin embargo, si la población fuese grande y el número de extracciones relativamente pequeño, podríamos emplear la distribución binomial.

Si en una caja hay 4000 bolitas verdes y 1000 blancas.

a. ¿Cuál es la probabilidad de, exactamente, 3 verdes en 3 extracciones con reposición?

b. Y, ¿sin reposición?

a. $\binom{3}{3} 0{,}8^3\ 0{,}2^0 = 0{,}512$

b. $\dfrac{4000}{5000} \cdot \dfrac{3999}{4999} \cdot \dfrac{3998}{4998} = 0{,}5119$

Resultados prácticamente iguales.

Aproximación de Poisson

Hemos visto que, si la probabilidad es pequeña y n grande, la distribución binomial puede aproximarse mediante la fórmula de Poisson.

$$P(n,h) = \frac{\lambda^h}{h!} e^{-\lambda} \qquad (\lambda = np)$$

(n es el número de intentos; h el número de "éxitos"; entendemos por éxito la aparición del suceso investigado).

Si la probabilidad de padecer la enfermedad A en una comunidad es 0;01 (1%), ¿cuál es la probabilidad de que al examinar una muestra de 200 individuos no aparezcan enfermos?

λ = 200 (0,01) = 2 (Si en cada 100 personas hay, en promedio, un enfermo, es razonable que, en 200, el valor esperado sea 2.)

$$P(200;0) = \frac{2^0}{0!} e^{-2} = 0{,}14$$

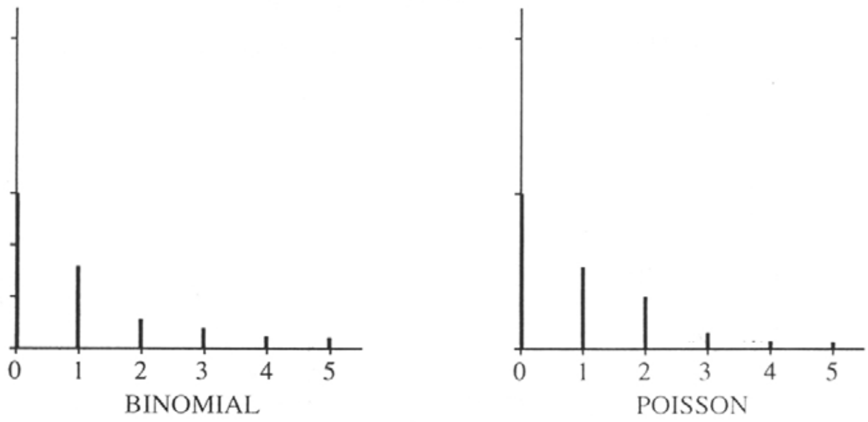
Si la probabilidad es pequeña, aunque n sea pequeño, Poisson da resultados razonablemente aproximados.

Actividad 10.1

Si p = 0,1 n = 5, calcule el lector las probabilidades para X = 0, 1, 2, 3, 4, 5 "éxitos", empleando:

a. la binomial,
b. Poisson.

Si hacemos los diagramas correspondientes, obtendremos:



Podemos observar que, para probabilidades pequeñas, las distribuciones binomial y la de Poisson son asimétricas.

Poisson y el flujo estacionario de sucesos

Si interviene el tiempo, empleamos la fórmula:

$$P(t) = \frac{(\lambda t)^{\lambda t}}{h!} e^{-\lambda t}$$

(λ es el promedio aritmético o valor medio).

Se sabe que, a la guardia del hospital de un pequeño pueblo, llega un promedio 5 personas por hora.

a. ¿Cuál es la probabilidad de que, en media hora, lleguen cuatro personas?
b. ¿Cuál es la probabilidad de que, en una hora, lleguen exactamente cinco personas?

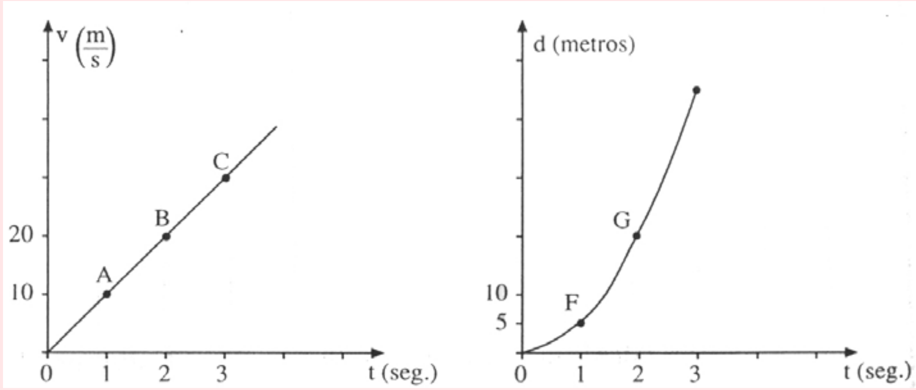
$$P(0{,}5 \text{ horas}) = \frac{(5 \ . \ 0{,}5)^{0{,}25}}{4!} e^{-2{,}5} = 0{,}034$$

$$P(1 \text{ hora}) = \frac{5^5}{5!} e^{-5} = 0{,}175$$

Realicemos una revisión de análisis:

En la figura 1, representamos la función v (t) = 10 (t) que podemos interpretar como la velocidad de un cuerpo que cae libremente.

En la figura 2, representamos las áreas acumuladas bajo la recta de la figura 1.



En la figura 1: Área 01A = 5; Área 02B = 20; Área 03C = 45
En la figura 2: Área 01A = 1F; Área 02B = 2G

Recordamos que las áreas bajo la recta v(t) = 10 t pueden calcularse por integración:

$$\int_0^1 10\ tdt = 5t^2\Big|_0^1 = 5 \qquad \int_0^2 10\ t = 5t^2\Big|_0^2 = 20\ldots$$

Podemos, entonces, definir la función de la figura 2, escribiendo $F(t) = \int_0^1 10\ xdx$ (Si damos a t los valores l; 2; 3... obtenemos 5, 20, 45...)

Aproximación de la binomial mediante la normal

Si consideramos 1000 lanzamientos, no tiene mucho sentido calcular la probabilidad de exactamente 500 caras; esa probabilidad es muy pequeña y no arriesgaríamos dinero para ganar sólo en el caso de obtener exactamente 500 caras en 1000 tiros. Cuando el número de pruebas es grande, las preguntas interesantes son del tipo: ¿Cuál es la probabilidad de que, en 1000 tiros, el número de casos esté comprendido entre 480 y 510? El cálculo mediante la binomial sería formidable:

$$0{,}5^{1000}\left[\binom{1000}{480}+\binom{1000}{481}+\ldots+\binom{1000}{510}\right]$$

Puede demostrarse que la distribución normal es una buena aproximación cuando n es grande.

A continuación, introducimos la función normal.

Sabemos que $\binom{1000}{480} = \dfrac{1000!}{480!\ 520!}$

El estudio de la función normal tiene su base en la fórmula de Stirling que permite aproximar factoriales:

$$\boxed{n! = \sqrt{2\ \pi\ n}\left(\frac{n}{e}\right)^n}$$

Por ejemplo:

n	n!	Stirling
10	3628800	3598695,6
20	2,4329 / 18	2,42278 / 18
30	2,65252 / 32	2,64517 / 32

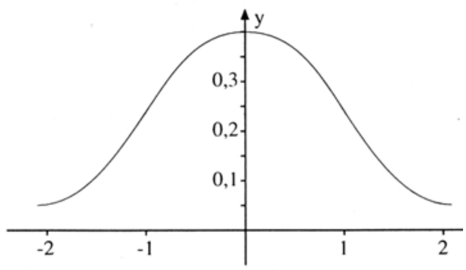
Al aumentar n, disminuye el error relativo.

La función normal

Representemos la función normal de densidad:

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

x	y
0	0,40
1	0,24
-1	0,24
2	0,05
-2	0,05

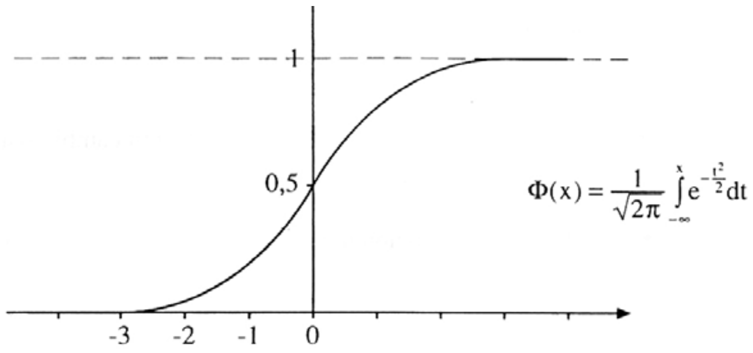


Esta función, según vemos, tiene gran importancia en la teoría de probabilidades y en la estadística.

Podríamos representar la función de áreas acumuladas bajo esta curva, para lo cual deberíamos calcular la integral de esta función para distintos límites. Pero, según hemos visto en el módulo 2, estas integrales sólo pueden calcularse mediante métodos aproximados. Puede el lector volver a repasar esas páginas y recordar que:

$$0{,}399\int_0^1 e^{-\frac{x^2}{2}}dx = 0{,}34$$

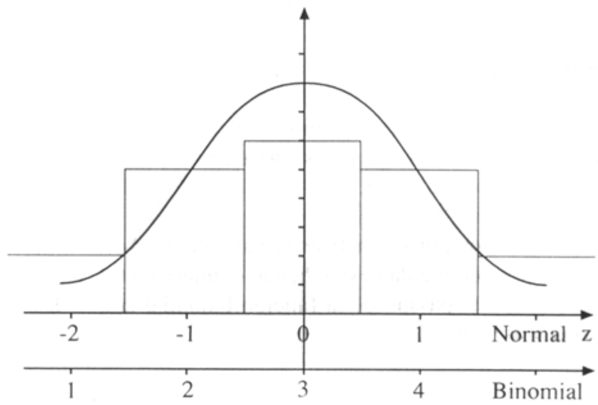
En el apéndice 4, damos una tabla de la función normal. Sólo consideramos valores no negativos de x porque, para valores negativos, tendremos en cuenta la simetría de la curva. Empleando los valores de esta tabla, dibujamos la función de distribución (o de acumulación):



Función de distribución normal

Aproximación de la binomial mediante la normal

En la figura hemos superpuesto a la distribución binomial para N = 6 y P = 0,5 (6 tiradas de una moneda), la curva normal:



El acuerdo no es bueno porque n es pequeño. Imaginemos los sucesivos gráficos de la binomial para n = 10, 20... tiradas de una moneda.

La demostración matemática obliga a que las longitudes de los intervalos disminuyan, de modo tal que los rectángulos son cada vez más "flaquitos". En el límite, si n tiende a infinito, obtenemos la curva normal.

.

Pero... observamos en el gráfico que las escalas no coinciden (si se tratase de 18 tiros de una moneda, el 0 de la normal coincidirá con el 9 de la binomial).

Es necesario, entonces, para resolver problemas de distribución binomial mediante la normal, "correr el origen". Por otra parte, habrá que tener en cuenta la desviación. Recordamos que, para la distribución binomial, el desvío es igual a $\sqrt{npq}$ (en el caso de n = 6 p = 0,5 desvío 1,22; en cambio, para n = 18 p = 0,5 desvío 2,12).

Para aplicar la distribución normal, es necesario hacer el cambio de variable:

$$Z = \frac{X - m}{\sigma}$$

Obsérvese que, cuando x coincide con el valor medio, resulta Z = 0.

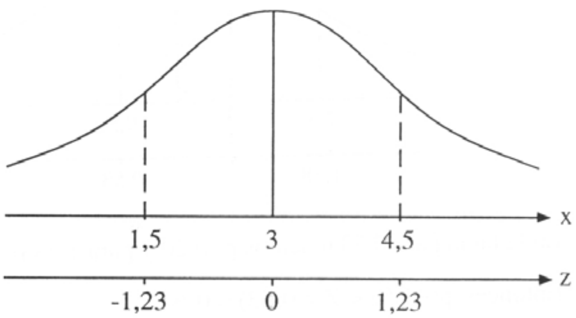
A este cambio se lo denomina **normalización de la variable**.

Aunque n es muy pequeño, vamos a calcular la probabilidad de que, en 6 tiros de la moneda, el número de caras esté entre 2 y 4 $(2 \leq X \leq 4)$, empleando la aproximación normal.

Comenzaremos por normalizar la variable X. Además, haremos la llamada corrección por continuidad, extendiendo el intervalo 0;5 a ambos lados (no olvidemos que hemos pasado de la función binomial, que es discreta, a la función normal, que es continua).

Quiere decir que en lugar de 2 escribiremos 1,5; y 4,5 en lugar de 4. Resultará:

$$Z_1 = \frac{1{,}5 - 3}{1{,}22} = -1{,}23 \qquad Z_2 = \frac{4{,}5 - 3}{1{,}22} = 1{,}23$$



En la tabla, encontramos para Z = 1,23 el área 0,3907 (Área a partir de Z = 0 hasta Z = 1,23). Entonces, la probabilidad es:

$$P(2 \leq X \leq 4) = P(-1{,}23 \leq Z \leq 1{,}23) = 0{,}39\ .\ 2 = 0{,}7814$$

Si empleamos la fórmula binomial, tendremos:

$$\left[\binom{6}{2} + \binom{6}{3} + \binom{6}{4}\right] 0{,}5^6 = 0{,}78125$$

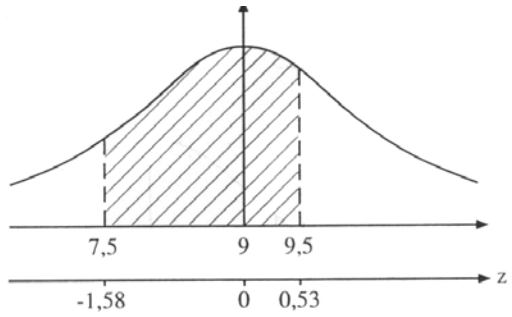
La aproximación es muy buena a pesar de que n es pequeña, lo que es explicable por la simetría que implica p = 0,5.

Veamos otro caso:

Un suceso tiene probabilidad constante p = 0,9. ¿Cuál es la probabilidad de que, en 10 tiros, obtengamos entre 8 y 9 éxitos?

En este caso $np = 10\ (0,9) = 9$ $\qquad \lambda = \sqrt{npq} = 0,95$

$$Z_1 = \frac{7,5-9}{0,95} = -1,58 \qquad Z_2 = \frac{9,5-9}{0,95} = 0,53$$



En la tabla, para 0,53 tenemos p = 0,20; y, para 1,58, p = 0,44.

Entonces $p(-1,58 < Z < 0,53) = 0,64$

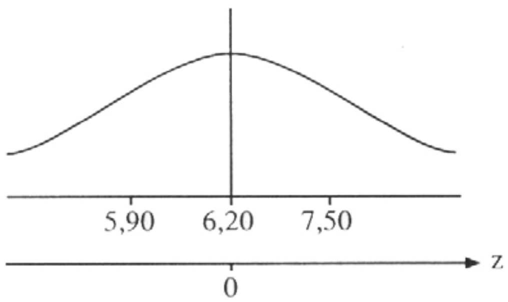
Si aplicamos el cálculo binomial: $\binom{10}{8}\ 0,9^8\ 0,1^2 + \binom{10}{9}\ 0,9^9\ 0,1\ = 0,58$

La aproximación no es buena.

En un colegio con una población de 2000 alumnos, el promedio de calificaciones en historia es 6,20, con un desvío = 1,32. Si suponemos normal la distribución de notas, ¿cuál es la probabilidad de que, elegido un alumno al azar, su nota esté entre 5,90 y 7,50? (Las calificaciones de los alumnos se aproximan al centésimo).

Calculamos los Z correspondientes:

$$Z_1 = \frac{5,90-6,20}{1,32} = -0,23 \qquad Z_2 = \frac{7,50-6,20}{1,32} = 0,98$$



En la tabla obtenemos los valores de probabilidad correspondientes:

0,0910 y 0,3365

Por lo tanto, la probabilidad pedida es: 0,4275

Con relación al problema anterior: ¿Cuántos alumnos, aproximadamente, estarán comprendidos entre 5,90 y 7,50?

$$0{,}4275 \times 2000 = 855$$

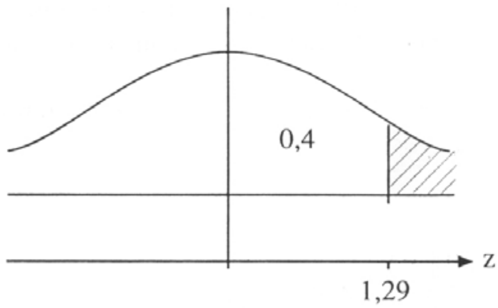
Con relación al problema anterior: ¿Cuál es la nota mínima para que un alumno pertenezca al 10 % de rendimiento superior?

En la tabla hallamos, para la probabilidad 0,4, el valor Z = 1,29

Podemos escribir:

$$1{,}29 = \frac{X - 6{,}20}{1{,}32}$$

$$7{,}90 = X$$



Actividad 10.2.

1. a) ¿De cuántas maneras diferentes pueden alinearse las letras a, b, c, d, e, f?
   b) ¿De cuántas maneras diferentes pueden alinearse las letras a, b, a, d, e, f?
   c) ¿De cuántas maneras diferentes pueden alinearse las letras a, a, c, d, e, a?
2. En tarjetas figuran los nombres de 9 personas; el nombre de una de ellas figura en dos tarjetas, de modo que hay 10 tarjetas. Hay que elegir al azar tres personas distintas. Si aparecen nombres repetidos, se anula la elección. ¿Cuál es la probabilidad de anulación?
3. Hay un grupo formado por 6 chicos y 6 chicas. Este grupo de 12 personas se vuelve a dividir al azar. ¿Cuál es la probabilidad de que la partición esté formada por varones, por un lado, y mujeres, por el otro?
4. Se tira un dado 2880 veces para observar la aparición de ases. ¿Cuál es la probabilidad de que el número de ases esté comprendido entre 470 y 490?
5. Se hace girar una ruleta de cinco letras AAABB. ¿Cuál es la probabilidad de que, en 1000 giros, el número de A obtenidos esté entre 580 y 630?
6. La probabilidad de que una persona padezca cierta enfermedad es 0,001. ¿Cuál es la probabilidad de que, en un grupo de 2000 personas, haya exactamente cuatro que padezcan la enfermedad? ¿Cuál es la probabilidad de que, en un grupo de 1000 personas, exactamente una tenga la enfermedad?

6 varones y 6 mujeres forman un solo grupo. Este grupo se subdivide en dos grupos de seis. Si la subdivisión se hace al azar, ¿cuál es la probabilidad de que cada grupo esté formado por tres personas de cada sexo?

Hay doce personas; en cuanto se eligen 6 para formar un grupo, el otro grupo queda determinado automáticamente. Entonces, el número de casos posibles es $\binom{12}{6}$. Los casos en los cuales cada grupo está formado por 3 mujeres y 3 hombres son $\binom{6}{3}\binom{6}{3}$

$$p = \frac{\binom{6}{3}\binom{6}{3}}{\binom{12}{6}} = \frac{400}{924} = 0{,}433$$

También podemos pensarlo de la siguiente manera: Formemos el primer grupo con 3 hombres y 3 mujeres; el otro grupo queda formado automáticamente.

Razonemos así: La probabilidad de que un hombre sea elegido al azar para formar el primer grupo es $\frac{6}{12}$

Consecuentemente:

$$\left(\underset{H}{\frac{6}{12}} \cdot \underset{H}{\frac{5}{11}} \cdot \underset{H}{\frac{4}{10}} \cdot \underset{M}{\frac{6}{9}} \cdot \underset{M}{\frac{5}{8}} \cdot \underset{M}{\frac{4}{7}} \right)$$

A este resultado hay que multiplicarlo por $\binom{6}{3}$ que es el número de lugares que los hombres pueden ocupar en la secuencia.

En consecuencia p = 0,433

Pueden repetirse los cálculos para 8 varones y 8 mujeres, p = 0,38. ¿Cuál es la tendencia, a medida que aumenta el número?

Puede intentarse, formando el primer y segundo grupo, eligiendo alternativamente los miembros de los mismos, hasta cumplir las exigencias.

11. Distribuciones en dos dimensiones

Recordamos que hemos simbolizado a la esperanza matemática o valor medio de la variable aleatoria X con E(X) o m.

Si calculamos la esperanza matemática de $(X\text{-}m)^2$, tenemos la varianza, que simbolizamos con:

$$E(X-m)^2 = Var(x) = D^2(x)$$

Un ejemplo. Para los valores X = 5; 4; 6 tenemos: E (X) = 5; siendo los valores de X - m:0; -1; 1 y $(X - m)^2$ = 0; 1; 1.

Por tanto, $E(X-m)^2 = (0+1+1):3 = 0{,}667$

Distribuciones para dos variables: Covarianza

Un interesante problema puede presentarse en estadística cuando, por ejemplo, tratamos de relacionar el consumo de gas con la temperatura ambiente. En este caso, tendremos dos variables: X = temperatura Y = consumo de gas. Damos un cuadro con pocos valores, como ilustración.

Para cada variable podemos hallar el valor medio y la varianza:

X	**2**	**10**	**18**	**26**
Y	8	6	3	1

$$E(X) = 14 = m_1 \;\; Var(X) = 80$$

$$E(Y) = 4{,}5 = m_2 \;\; Var(X) = 7{,}25$$

Pero, estos valores no relacionan X con Y.

Calcularemos, ahora:

$$E(x-m_1).(y-m_2) =$$
$$= \left[(2-4)(8-4{,}5)+(10-14)(6-4{,}5)+(18-14)(3-4{,}5)+(26-14)(1-4{,}5)\right]:4 =$$
$$= -24$$

Este valor se denomina covarianza de X e Y.

Correlación de variables

Consideremos la función y = 2x + 5

(Esta función puede interpretarse como el costo de producción de x unidades, siendo 5 el costo fijo)

X	Y
2	9
3	11
5	15

Con relación a la tabla:

E(x) = 3,33   E(y) = 11,67
V(x) = 1,56   V(y) = 6,22

$$3E(x-m_1)(y-m_2) = (2-3{,}33)(9-11{,}67) + (3-3{,}33)(11-11{,}67) + (5-3{,}33)(15-11{,}67)$$
$$\text{Co var ianza}(x,y) = 3{,}5 + 0{,}22 + 5{,}56 = 3{,}09$$

Vamos a dividir la covarianza por el producto de los desvíos:

$$\frac{3{,}09}{\sqrt{1{,}56}\sqrt{6{,}22}} = \frac{3{,}09}{3{,}11} \text{ aproximadamente } 1$$

Al número covarianza (x,y)/(desvío x, desvío y), se lo llama **coeficiente de correlación lineal** de las variables x e y. En este ejemplo, las variables están completamente correlacionadas por la fórmula y = 2x + 5. En este caso, si se eliminaran los errores de redondeo, el coeficiente sería uno.

Se puede probar que el coeficiente de correlación r toma valores en el intervalo |-1,1|.

Actividad 11.1

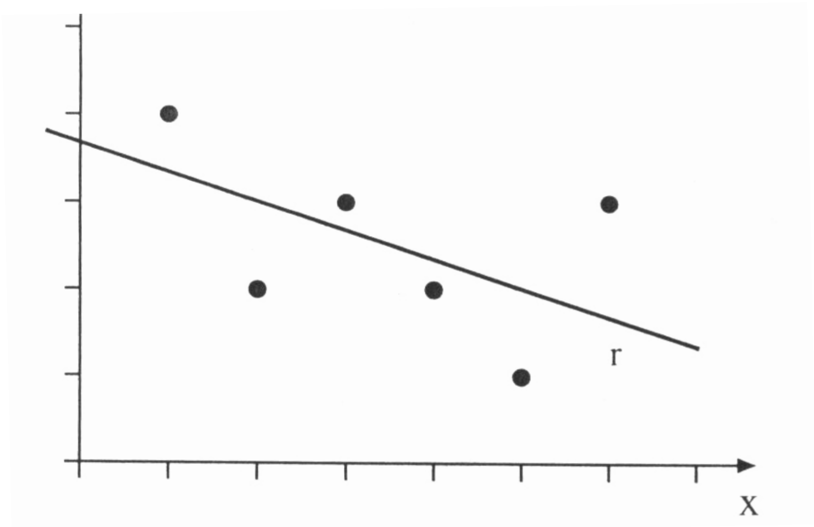
Calcular el coeficiente de correlación lineal para la tabla:

X	**2**	**10**	**18**	**26**
Y	8	6	3	1

Recta de regresión

Representamos los puntos correspondientes a la tabla:

X	1	2	3	4	5	6
Y	4	2	3	2	1	3



Como los puntos no están alineados, la correlación no es completa. Se presenta el problema de encontrar una recta que aproxima lo mejor posible a esa "nube" de puntos.

Supongamos que la recta r es la solución. La ecuación de r tendrá la forma y = ax + b. Nuestro trabajo consiste en determinar a y b. Se podrían enunciar distintos criterios para determinar esa recta. Por múltiples razones, generalmente, se adopta el siguiente: Hay que hacer mínima la suma de los cuadrados de las diferencias entre los valores de las ordenadas de la tabla y de la recta. Es decir, hay que hacer mínimo el valor de la función:

$$f = [(a.1+b)-4]^2 + [(a.2+b)-2]^2 + \ldots + [(a.6+b)-3]^2$$

Se trata de una función de dos variables. Hechos los cálculos de rutina, resulta:

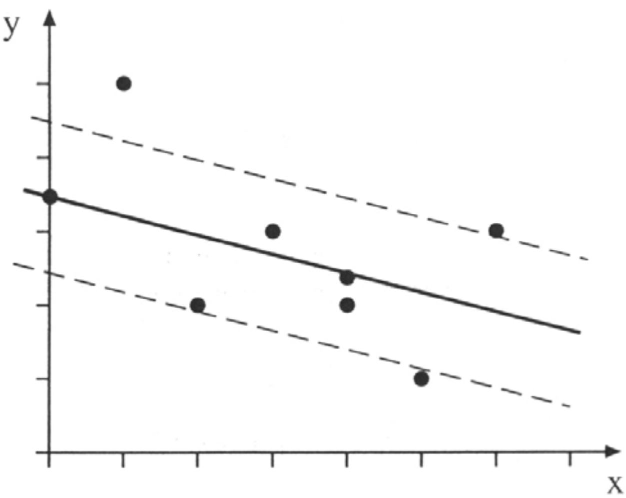
$$f = 91a^2 + 42ab + 6b^2 - 96a - 30b + 43$$

Para hallar el mínimo se calculan las derivadas parciales y se resuelve el sistema:

$$\begin{cases} 182a + 42b = 96 \\ 42a + 12b = 30 \end{cases} \qquad \text{Solución} \quad a = -0{,}257 b = 3{,}4$$

La recta que mejor aproxima los datos según este criterio de mínimos cuadrados es:

Y = -0,257 x + 3,4



X	1	2	3	4	5	6
Y	4	2	3	2	1	3
$\hat{y}$	3,14	2,89	2,63	2,37	2,12	1,86

A partir de la recta de regresión, podemos hacer algún trabajo; por ejemplo, comparar las estimadas por la recta con las Y de los datos.

Podemos calcular el error de la estimación, hallando:

$$[(4-3{,}14)^2 + (2-2{,}89)^2 + (3-2{,}63)^2 + (2-2{,}37)^2 + (1-2{,}12)^2 + (3-1{,}86)^2] : 6 = 0{,}73$$

Entonces, $s = \sqrt{0{,}73} = 0{,}85$

Las líneas de puntos están a la distancia 0,85 de la recta. Cuando el número de datos es suficientemente grande, se espera que alrededor del 68% de ellos caiga dentro de la banda de paralelas.

En este ejemplo, como los datos son apenas seis, conviene emplear s corregida, dividiendo por (n-2); es decir:

$$s = \sqrt{1{,}09} = 1{,}05$$

La recta de regresión puede emplearse para predecir pares que no figuran en la tabla. Sin embargo, no sería prudente calcular y para x = 20.

Por ejemplo, para x = 2,6, predecimos:

$$y = -0{,}257(2{,}6) + 3{,}4 = 2{,}73$$

Muchas calculadoras de bolsillo permiten calcular rápidamente los coeficientes de las rectas de regresión (Hay que fijarse si ofrecen LR *–línear regresión–*).

Más allá de los cálculos, es necesario conocer, aproximadamente, cómo funciona este tema.

Supongamos que se hacen mediciones en un laboratorio para estudiar la dilatación lineal del acero. Se obtendrán valores finales de la medida de una varilla de longitud inicial l_0 para distintos incrementos de temperatura. Se tendrá una nube de puntos que sugieren una relación lineal. ¿Por qué los puntos no están totalmente alineados?

Porque, por más cuidadoso que sea el trabajo experimental, se deslizan errores aleatorios.

La recta de regresión será, aproximadamente, $l = l_0 + l_0\ \alpha\ (\Delta t)$

Éste es el modelo teórico de dilatación lineal, siendo α el coeficiente de dilatación del acero.

Podemos asimilar $l = Y \qquad (\Delta t) = X$

Si se experimenta, para estudiar los metros recorridos en la caída libre al variar el tiempo t, la nube de puntos no sugerirá alineación sino una parábola. Siguiendo las mismas pautas, puede obtenerse una parábola de regresión.

La recta de regresión puede obtenerse empleando la fórmula:

$$\frac{y - m_2}{\sigma_2} = \rho \cdot \frac{n - m_1}{\sigma_1}, \text{siendo } \rho = \text{coeficiente de correlación}$$

Se puede obtener otra recta de regresión, intercambiando el papel de las variables (X dependiente, Y independiente)

En un grupo de 10 alumnos, se comparan los notas obtenidas en el examen de ingreso en matemática con la nota final del curso de primer año:

X	5	8	4	7	4	6	8	7	6	7
Y	6	6	6	8	5	5	9	7	9	6

X = Nota de ingreso
Y = Nota final

Hallar la recta de regresión Y = a X + b

$$\text{coeficiente de correlación} = 0{,}43$$

$$\overline{X} = 6{,}1 \qquad \overline{Y} = 6{,}7$$

$$\text{desv.}(x) = 1{,}37 \qquad \text{desv.}(y) = 1{,}42$$

$$Y = 0{,}44x + 4{,}02$$

Actividad 11.2.

La siguiente tabla relaciona el número de años que el paciente ha fumado con el daño sufrido en los pulmones según evaluación médica (de 0 a 100)[1].

N°	Años X	Daño Y
1	25	55
2	36	60
3	22	50
4	15	30
5	48	75
6	39	70
7	42	70
8	31	55
9	28	30
10	33	35

Se pide:

a. Hallar el coeficiente de correlación.
b. Determinar la recta de regresión de Y sobre X.
c. Calcular el error de estimación, mediante:

$$s = \sqrt{\frac{\text{suma}(y - \hat{y})^2}{n - 2}}$$

Siendo el valor estimado por la ecuación de la recta.

d. Representar.

[1] *Figura en el texto* Introducción a la Probabilidad y la Estadística, *de W. Mendenhall. Grupo Editorial Iberamérica.*

12. Muestras

Muchas veces, el investigador se encuentra en esta posición: Está trabajando con una gran población (seres humanos o peces o automóviles...) y desea conocer algunos valores (promedios, desviaciones, tendencias...) sin que sea posible o económico obtener los registros de toda la población.

Entonces, apela al estudio de muestras. Puede suceder que tenga libertad para elegir el tamaño de la muestra; pero, a veces, tiene que conformarse con datos escasos (una enfermedad rara, piedra preciosa...).

Supongamos que una población está formada por: 100 valores iguales a 1; 200 iguales 2; etc. según la tabla siguiente (1):

X	1	2	3	4	5	6	7	8	9	10
Y	100	200	50	15,0	150	150	50	50	50	50

Nosotros no tenemos acceso a esta tabla; pero, tomaremos una muestra para estimar el valor medio. Vamos a suponer que la población está numerada correlativamente según esos valores. Por ejemplo: el primer número tres es N° 301.

Vamos a elegir N = 10 como tamaño muestral; por lo tanto, buscamos 10 números aleatorios:

202	009	138	969	936	905	830	076	407	838
3	1	2	10	9	9	8	1	4	7

$$\overline{X} = 5{,}4$$

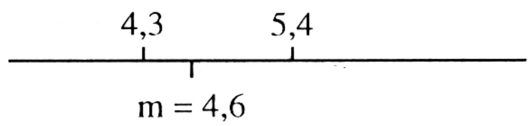
Tomemos otra muestra:

170	419	014	439	290	912	673	955	264	326
2	4	1	4	2	9	6	10	2	3

$$\overline{X} = 4{,}3$$

El que pueda tener acceso a toda la información –es decir, a la tabla (1)–, obtendrá el promedio m = 4,6

El valor $\overline{X}$ que se obtiene para cada muestra es aleatorio. El investigador está "apostando" a acertar "cerca" del valor verdadero.



En nuestro ejemplo, la primera muestra tuvo "más suerte".

Si publicamos un trabajo y afirmamos: "**Estimamos** que el valor medio de la población es 4,3", sería conveniente dar una idea del error posible. Resulta razonable pensar que el error depende del tamaño de la muestra.

Estimadores

Se puede probar que el valor medio muestral cumple la propiedad siguiente:

$$E(\overline{X}) = m$$

Esto quiere decir que: Si en nuestro ejemplo, armáramos todas las muestras posibles e hiciéramos el promedio de los valores medios muestrales, obtendríamos m.

También se puede probar que:

$$\text{Desvío } (\overline{X}) = (\text{Desvío } X) / \sqrt{n}$$

Esto significa que, si calculásemos todos los valores medios muestrales, el desvío de $\overline{X}$ sería igual al desvío de la población (dividido por $\sqrt{n}$; siendo n el tamaño de la muestra).

Esta propiedad es importante porque confirma la idea intuitiva de que cuanto mayor sea el tamaño de la muestra, mejor será la estimación de $\overline{X}$.

Por ejemplo, si formásemos una sola muestra con los veinte valores obtenidos, tendríamos:

$$\overline{X} = (5{,}4 + 4{,}3) : 2 = 4{,}85$$

que está más cerca del verdadero valor.

Estimador de la varianza de la población

Acabamos de ver que el desvío de la media muestral es:

$$(\text{desvío } X) / \sqrt{n} = \sigma / \sqrt{n}$$

Pero, ¿cómo podemos calcular ese desvío si no conocemos σ (desvío de la población) y sólo tenemos una muestra?

Podemos estimar σ calculando el desvío muestral s.

Repasamos, para la primera muestra:

$$\text{Varianza } X = \left[(3 - 5{,}4)^2 + \ldots + (7 - 5{,}4)^2\right] / 10 = 11{,}42$$

$$\text{desvío } X = 3{,}38$$

Análogamente, para la segunda muestra: 2,93

Se puede probar que:

$$\boxed{E(s^2) = \frac{n-1}{n}\sigma^2} \quad (1)$$

lo que quiere decir que, si considerásemos todas las muestras de tamaño 10 de la población y calculásemos el promedio de las $s^2 = E(s^2)$, sería aplicable la fórmula (1).

Si se dispone de s^2 será, entonces, conveniente estimar σ^2 mediante:

$$\frac{n}{n-1}s^2 = \sigma^2$$

Y el desvío mediante:

$$\sigma = \sqrt{\frac{n}{n-1}}\,s$$

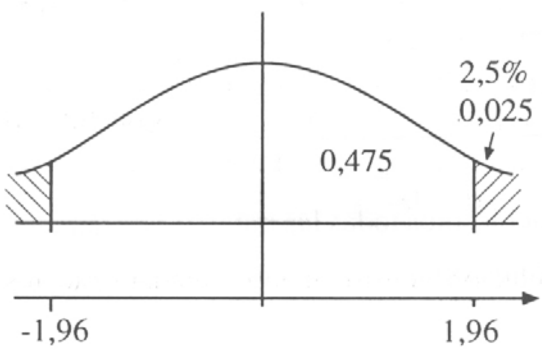
Es evidente que, cuando el tamaño de la muestra es grande, s tiende a σ

Para n pequeño, las calculadoras tienen un registro σ_{n-1} que aplica (1). En el caso de nuestras muestras, tendremos: s = 3,57 y s = 3,09, respectivamente, como estimadores de σ.

Cuando un estimador no requiere corrección –como, en el caso de $\overline{X}$–, decimos que es **insesgado**. En cambio, s es un estimador **sesgado** de σ.

Estimación por intervalos empleando la distribución normal

Para conocer el valor medio de colesterol en una población de 10000 habitantes, se tomó una muestra de 100 habitantes, obteniéndose un valor medio $\overline{X} = 200$ siendo s = 50. Se pide estimar el valor medio de toda la población, usando un intervalo de confianza del 95%.



El área bajo la curva es 1; o, expresada en porcentaje, es 100%.

Para lograr 95%, suprimimos las dos "colitas", porque la estimación $\overline{X} = 200$ podría fallar por exceso o por defecto.

En la tabla encontramos que, para el área 0,475, corresponde Z = 1,96

Sabemos que $Z = \dfrac{x - m}{\text{desvío } (X)}$ (1)

(X corresponde a la variable muestral "valor medio")

Pero, no disponemos de m sino de $\overline{X}$ y no disponemos del desvío de la población sino del desvío s de la muestra. Usaremos s sin corrección, porque el tamaño de la muestra no es pequeño. Como estamos tratando el valor medio, el desvío de este valor está dado por $s / \sqrt{n}$.

Sustituyendo en (1), tendremos:

$$1{,}96 = \frac{\overline{X} - 200}{50 / \sqrt{100}} \qquad \begin{aligned} 200 \pm 1{,}96(5) &= \overline{X} \\ 209{,}8 &= \overline{X} \\ 190{,}2 &= \overline{X} \end{aligned}$$

Quiere decir que, con una probabilidad 0,95, el verdadero valor medio poblacional está en el intervalo [190,2; 209,8].

El significado del resultado es el siguiente: Si, por ejemplo, tomamos 1000 muestras de tamaño 100, en el 95% –aproximadamente– de los casos, el valor medio estará en ese intervalo.

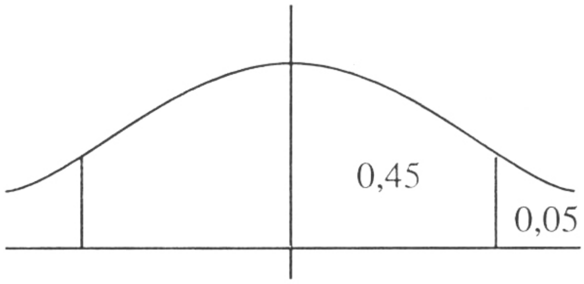
El problema anterior da idea de las dificultades que se presentan al estudiar muestras: El investigador debe arreglárselas con lo que tiene y procurar la mayor precisión posible o, lo que es más claro, señalar el posible margen de error. Pese a las dificultades, los logros teóricos y prácticos de la probabilidad y estadística son formidables (remitimos, por ejemplo, nuevamente, al ejemplo que hemos dado sobre emisión de partículas α).

Las computadoras permiten, por simulación, confirmar, aproximadamente, los resultados.

En una muestra de 300 peces extraídos de una laguna, se obtuvieron los siguientes valores referidos a la longitud: $\overline{X} = 26{,}3$ cm, $s = 8$ cm.

Estimar, con una confianza del 90%, el valor medio de la población.

Estimamos la varianza σ de la población empleando s. Por lo tanto, el desvío de la variable "valor medio" será $8 / \sqrt{300} = 0{,}46$.



El valor Z correspondiente a P = 0,45 es Z = 1,65 (tabla de la normal)

Por lo tanto:

$$1{,}65 = \frac{X - 26{,}3}{0{,}46} \qquad 27{,}06 = X$$

En consecuencia, con el intervalo de confianza 90% es $\left[25{,}54\ ;\ 27{,}06\right]$

Actividad 12.1.

1. Se alinean aleatoriamente los símbolos ABAACDE. ¿Cuál es la probabilidad de que las A figuren en los extremos y en la casilla media?
2. Determinar el coeficiente de correlación y la recta de regresión correspondiente a la tabla:

X	0,5	1,2	1,9	2,5	3,3
Y	0,9	1,18	1,46	1,74	2,02

3. Se ha observado el aumento de peso de 36 pollos sometidos a una dieta durante cierto período de tiempo, obteniéndose $\overline{X} = 0{,}5\ \text{kg}\ ;\ s = 0{,}4\ \text{kg}$. Obtener un intervalo de confianza del 90% para el valor medio.
4. El promedio de llegadas de automóviles a una estación de servicio es 8 por hora. ¿Cuál es la probabilidad de que, en una hora, haya que atender más de 8 automóviles?
5. Calcular el área bajo la curva $y = \sqrt{x}$ en el intervalo $[0{,}1]$, empleando números aleatorios.
6. Al tirar un dado repetidas veces, ¿cuál es la probabilidad de que el primer as recién aparezca en la décima tirada?

Clave de respuestas

10.1.

X	0	1	2	3	4	5	
P	0,60	0,30	0.076	0,013	0,0016	0,00016	Poisson
	0,59	0,33	0,086	0,008	0,0004	10^{-5}	Binomial

10.2.

1. a) 6!
b) 6! / 2!
c) 6! / 3!

2.

$$\binom{8}{1} \Big/ \binom{10}{3}$$

3. Analizado en el texto.

4.

$$m = 480 \qquad \sigma = 20 \qquad Z_1 = \frac{469{,}5 - 480}{20} = -0{,}525$$

$$Z_2 = \frac{490{,}5 - 480}{20} = 0{,}525 \qquad \phi\ (0{,}525) = 0{,}2$$

$$P(470 < X < 490) = 0{,}40$$

5.
(580 – 600) : 15,49 = -1,29 (P = 0,4015)
(630 – 600) : 15,49 = 1,94 (P = 0,4738)
0,4015 + 0,4738 = 0,8753

Si se hace la corrección por continuidad:

(579,5 – 600) : 15,49 = -1,32 (P = 0,4066)
(630,5 – 600) : 15,49 = 1,97 (P = 0,4756)
0,4066 + 0,4756 = 0,8822

6.

$$P(2000{,}4) = \frac{2^4}{4!} e^{-2} = 0{,}09$$

$$P(1000{,}1) = \frac{1}{1} e^{-1} = 0{,}37$$

11.1. r = -0,997

11.2.

$$r = 0{,}7738 \qquad \bar{y} = 11{,}238 + 1{,}309x$$

$$s^2 = \frac{(55-44)^2 + (58-60)^2 + (40-50)^2 + \ldots + (35-54)^2}{10-2}$$

Se han redondeado los cálculos al entero.

S = 11,1

12.1.

1.

$$\frac{P_4}{P_7 : P_3} = \frac{24}{840} = 0{,}029 \qquad \text{o} \qquad \frac{3}{7}\cdot\frac{4}{6}\cdot\frac{3}{5}\cdot\frac{2}{4}\cdot\frac{2}{3}\cdot\frac{1}{2} = 0{,}029$$

2.

r = 1

y = 0,4 x + 0,7 (los puntos están alineados)

3.

$$1{,}65 = \frac{X - 0{,}5}{0{,}4 : \sqrt{36}} \qquad 1{,}1 = X - 0{,}5 \qquad 1{,}6 = X$$

$$\text{El intervalo es } [-0{,}6\ ;\ 1{,}6]$$

4.

$$P(x \le 8) = e^{-8}[1 + 8 + 32 + 85{,}3 + 170{,}67 + 273 + 364 + 416 + 416] = 0{,}58$$

$$P(x > 8) = 0{,}42$$

5.

$\left(\frac{5}{6}\right)^9 \frac{1}{6} = 0{,}03$ (Nueve veces no sale el as; en la décima debe salir)

Modelo de trabajo en clase

Presentar a la clase el siguiente problema[1]:

Se tiran dos dados con forma de dodecaedro. Cada dado tiene 7 caras rojas y 5 blancas. Se pregunta: ¿Cuál es la probabilidad mayor:

- 2 rojas;
- una roja y una blanca;
- 2 blancas?

1. Hacer una estadística de las respuestas. Analizar con los alumnos por qué eligieron la respuesta.

   A continuación dar la respuesta correcta:

   P (2 rojas) = 49/144
   P (1 roja y 1 blanca) = 70/144

2. El problema se puede simular mediante números al azar. Que los alumnos propongan la simulación (Puede hacerse con una calculadora que genere números al azar; por ejemplo: hasta 0,417 se toma como blanca y el resto como roja).

3. Considerar el mismo problema con tres tetraedros, cada uno con 3 caras rojas y 1 blanca, y calcular las probabilidades de los resultados: 3 rojas y 2 rojas y 1 blanca.

Vicisitudes de la enseñanza de la Estadística

La sociedad Estadística de Londres fue fundada en 1834. No manifestó interés por la enseñanza. Recién en 1870 se creó el Comité de Estadística en las Escuelas, siguiendo las indicaciones del Congreso Internacional de Estadística de La Haya que resolvió:

“La enseñanza de la Estadística debe introducirse en las escuelas, en todos los grados que siguen a la enseñanza elemental hasta la universidad”.

Es curiosa la historia del Comité: Tuvo su primera reunión el 14 de Julio de 1870 y su última reunión ocho días después. Literalmente, desapareció (¿No hemos visto esto muchas veces?).

Entre tanto[2]:

- En Hungría, la probabilidad se enseñaba ya en 1849; pero, también desapareció, en poco tiempo, de los programas.
- En Francia se introdujo en 1868; pero, como parte de los programas de geografía.
- En Japón se estableció una cátedra de Estadística en 1882.

---

[1] *Estas actividades han sido sugeridas por la lectura de un artículo de Mike Fletcher, de Oxford, parecido en la revista* Teaching in Statistics *de Verano. 1994. Universidad de Sheffield, Inglaterra*

[2] *Fuente:* Teaching in Statistics*. Universidad de Sheffield, Inglaterra.*

Apéndice 1. Comparación de medidas muestrales

Muchas veces se presenta el problema de analizar si la diferencia entre dos grupos es significativa.

Supongamos que el promedio en matemática de cien alumnos de la escuela A es $\overline{X} = 5{,}20$ con $s^2 = 3{,}24$, siendo los valores para cien alumnos de la escuela B: $\overline{Y} = 5{,}50$ con $s^2 = 4{,}41$. Elegimos como nivel de significación 0,1.

Consideramos una nueva variable $(\overline{X} - \overline{Y})$ que es la diferencia entre las medias. ¿Cuál será la varianza para esa diferencia?

$$\text{Para } \overline{X} \text{ la varianza es } 3{,}24 : 100$$

$$\text{Para } \overline{Y} \text{ la varianza es } 4{,}41 : 100$$

Porque:

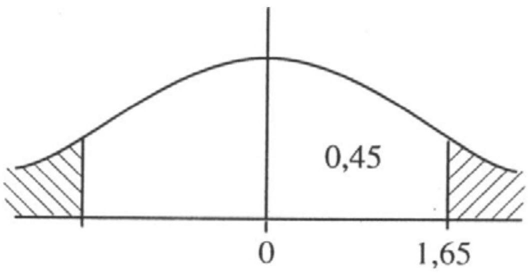
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \text{y} \qquad VAR(X) = \sigma^2 / n$$

Para calcular la varianza estimada de la diferencia, se suman las varianzas (Piense el lector que si se restaran las varianzas, podrían presentarse el caso en que se obtuviera cero). También recomendamos ver el ejemplo que figura abajo.

Dicho esto, es fácil repetir los cálculos habituales. Suponemos que no hay diferencia significativa y, por lo tanto, el valor medio de la diferencia es cero (hipótesis nula).

Tendremos, entonces:

$$\frac{5{,}50 - 5{,}20 - 0}{\sqrt{\frac{3{,}24}{100} + \frac{4{,}41}{100}}} = \frac{0{,}30}{0{,}28} = 1{,}08 \ \text{(En la tabla } P = 0{,}4)$$



Si este valor hubiese superado 1,65, podríamos afirmar que la diferencia es significativa al nivel 0,1. Por lo tanto, en este caso, la diferencia entre A y B no es significativa.

A continuación damos un ejemplo de que:

Varianza (X - Y) = Varianza X + Varianza de Y

	5	8	2	$\overline{X} = 5$	V = 6
$\overline{Y}$	4	6	8	$\overline{Y} = 6$	V = 2,67

A partir esta tabla, podemos formar la población de todas las diferencias (X - Y): 1: -1; -3; 4; 2; 0; -2, -4, -6.

Si calculamos la varianza de estas diferencias, obtendremos 8,67 que es la suma de las varianzas 6 y 2,67.

Con relación al ejemplo anterior, un psicopedagogo afirma: “Al nivel 0,1 no se advierten diferencias en los promedios de matemática de A y B”.

¿Qué quiere decir “al nivel 0,1”? Significa que corre un riesgo de equivocarse porque, para la misma población, si tomase todas las muestras posibles, hubiese encontrado un 10% que presentaría diferencias significativas.

Apéndice 2. Muestras pequeñas y la distribución de Student

Si las muestras son pequeñas (suelen considerarse como tales las de tamaño menor que 30), en lugar de la función normal debe emplearse la distribución t de Student, que tiene en cuenta el tamaño.

Para $n \geq 30$, las distribuciones de Student y la normal convergen hacia los mismos valores.

El índice de contaminación se mide en partes por millón, siendo el tolerable, según la Organización Mundial de la Salud, de 9 p.p.m.

En la ciudad de Buenos Aires, se han tomado los siguientes registros, durante una semana:

11,8; 10,6; 9,3; 8,4; 10,7; 7,4; 6,6

Averiguamos si la diferencia entre el valor medio de la muestra y el tolerable es significativo al nivel 0,1.

Calculamos $\overline{X} = 9{,}26$ $\sigma_{n-1} = 1{,}9$ (corregida)

Procedemos de la misma manera que para calcular Z; pero, empleamos la tabla de la distancia de Student.

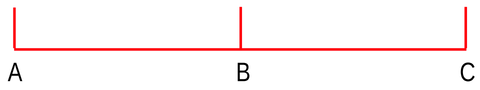
$$t = \frac{9{,}26 - 9}{\frac{1{,}9}{\sqrt{7}}} = 1{,}27$$

En la tabla de Student, debe tenerse en cuenta el número de grados de libertad, que es igual a (n-1).

Para $t_{0,1}$, encontramos 1,44 y n - 1 = 7 - 1 = 6

Como nuestro valor 1,27 < 1,44, diremos que el valor medio de la muestra no es significativamente diferente del valor de tolerancia.

Apéndice 3. Matrices de probabilidad



Consideremos un punto que sale de A y se mueve según estas reglas:

- Si está en A, tiene probabilidad 0,2 de permanecer en A y 0,8 de pasar a B.
- Si está en B, pasa a A con probabilidad 0,4 o a C (p = 0,3) o queda en B (p = 0,3).
- Si está en C, tiene p = 0,6 de permanecer y p = 0,4 de pasar a B.

Estas reglas se pueden resumir en la matriz:

$$\begin{array}{ccc} & & \text{Está} \\ & & \begin{array}{ccc} A & B & C \end{array} \\ \text{Pasa} & \begin{array}{c} A \\ B \\ C \end{array} & \begin{pmatrix} 0{,}2 & 0{,}4 & 0 \\ 0{,}8 & 0{,}3 & 0{,}4 \\ 0 & 0{,}3 & 0{,}6 \end{pmatrix} \end{array}$$

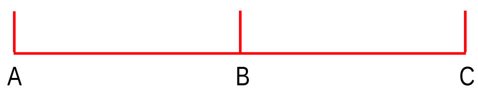
Los sucesivos pasos se hacen de manera aleatoria (empleando papeles marcados, ruletas, etc.).

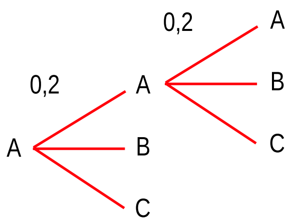
> Si se parte de A, ¿cuál es la probabilidad de estar en A o en B, después de los pasos?

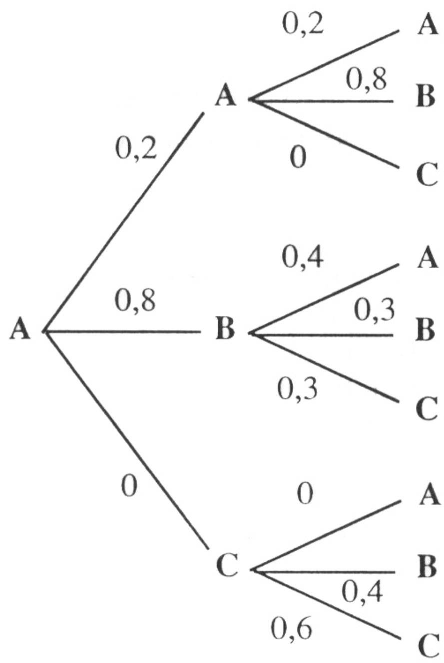
El producto de matrices permite dar la respuesta:

$$\begin{pmatrix} 0{,}2 & 0{,}4 & 0 \\ 0{,}8 & 0{,}3 & 0{,}4 \\ 0 & 0{,}3 & 0{,}6 \end{pmatrix} \begin{pmatrix} 0{,}2 & 0{,}4 & 0 \\ 0{,}8 & 0{,}3 & 0{,}4 \\ 0 & 0{,}3 & 0{,}6 \end{pmatrix} = \begin{pmatrix} 0{,}36 & 0{,}20 & 0{,}16 \\ 0{,}40 & 0{,}53 & 0{,}36 \\ 0{,}24 & 0{,}27 & 0{,}48 \end{pmatrix}$$

Obsérvese que la suma de cada columna da siempre 1. Veámoslo de otra manera.



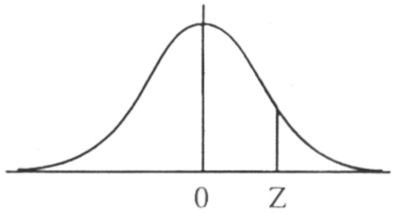




En dos pasos se está en A, con probabilidad = 0,2 . 0,2 + 0,8 . 0,4 = 0,36.

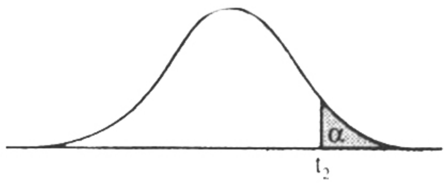
En dos pasos se está en B, con probabilidad = 0,2 . (0,8) + 0,8 . (0,3) = 0,40

Apéndice 4. Tabla de área bajo la función normal



Z	0	1	2	3	4	5	6	7	8	9
0	0	0,004	0,008	0,12	0,016	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,832	0,871	0,910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,131	0,1368	0,1406	0,1443	0,148	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,17	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,195	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,219	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,258	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,291	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,334	0,3365	0,3389
1	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3463	0,3665	0,3686	0,3708	0,3729	0,3749	0,377	0,379	0,381	0,383
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,398	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,437	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,475	0,4756	0,4761	0,4767
2	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,483	0,4834	0,4838	0,4842	0,4846	0,485	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,489
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,492	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,494	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,495	0,4955	0,4956	0,4957	0,4959	0,496	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,497	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,498	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,499	0,499
3,1	0,499	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Apéndice 5. Función de Student



g.1.	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	g.1.
1	3,078	6,314	12,706	31,821	63,657	1
2	1,886	2,92	4,303	6,965	9,925	2
3	1,638	2,353	3,182	4,541	5,841	3
4	1,533	2,132	2,776	3,747	4,604	4
5	1,467	2,015	2,571	3,365	4,032	5
6	1,44	1,943	2,447	3,143	3,707	6
7	1,415	1,895	2,365	2,998	3,499	7
8	1,397	1,86	2,306	2,896	3,355	8
9	1,383	1,833	2,262	2,821	3,25	9
10	1,372	1,812	2,228	2,764	3,169	10
11	1,363	1,796	2,201	2,718	3,106	11
12	1,356	1,782	2,179	2,681	3,055	12
13	1,350	1,771	2,16	2,65	3,012	13
14	1,345	1,761	2,145	2,624	2,977	14
15	1,341	1,753	2,131	2,602	2,947	15
16	1,337	1,746	2,12	2,583	2,921	16
17	1,333	1,74	2,11	2,567	2,898	17
18	1,33	1,734	2,101	2,552	2,878	18
19	1,328	1,729	2,093	2,539	2,861	19
20	1,325	1,725	2,086	2,528	2,845	20
21	1,323	1,721	2,08	2,518	2,831	21
22	1,321	1,717	2,074	2,508	2,819	22
23	1,319	1,714	2,069	2,5	2,807	23
24	1,318	1,711	2,064	2,492	2,797	24
25	1,316	1,708	2,06	2,485	2,787	25
26	1,315	1,706	2,056	2,479	2,779	26
27	1,314	1,703	2,052	2,473	2,771	27
28	1,313	1,701	2,048	2,467	2,763	28
29	1,311	1,699	2,045	2,462	2,756	29
inf.	1,282	1,645	1,96	2,326	2,576	inf.

Bibliografía

- Cramer H. *Elementos de la teoría de probabilidades*. Aguilar.
- Cochran W. *Experimental Designs*. J. Wiley and Sons.
- Feller W. *Introducción a la teoría de probabilidades y aplicaciones.*
- Mendenhall W. *Introducción a la probabilidad y la estadística*. Grupo Editorial Iberoamérica.
- Moroney M. *Hechos y estadísticas*. Eudeba.
- Santaló L. *Probabilidades e inferencia estadística*. OEA.

Revista

- *Teaching Statistics*. Universidad de Sheffield, Inglaterra