

NOTAS PARA EL CURSO DE ESTADÍSTICA - 2013
LICENCIATURA EN RECURSOS NATURALES
RIVERA

Alejandro Cholaquidis

Facultad de Ciencias
Universidad de la República

Estas notas fueron hechas para el curso de Estadística de la Licenciatura en Recursos Naturales, dictado en el año 2012. Han sido re escritas para el curso 2013. No pretenden ser un sustituto de las clases, sino más bien una guía para que el estudiante sepa que temas se dieron en las clases, y una orientación del estilo y metodología usados en las mismas. Las definiciones se introducen de manera intuitiva, omitiendo las construcciones teóricas rigurosas que la teoría de la probabilidad requiere. La bibliografía de cada tema figura al final. Las erratas que hubieren, se agradece comunicarlas a acholaquidis@hotmail.com.

Índice general

1. Conteo, Probabilidad, Variables Aleatorias	5
1.1. Conteo	5
1.1.1. Permutaciones	5
1.1.2. Arreglos	6
1.1.3. Combinaciones	7
1.2. Probabilidad	8
1.2.1. Casos favorables sobre casos posibles: Tirar dos dados	9
1.2.2. Independencia	11
1.3. Variables aleatorias discretas	13
1.3.1. Distribución Binomial	13
1.3.2. Distribución Geométrica	15
1.3.3. Distribución Hipergeométrica. Extracciones sin reposición	16
1.3.4. Distribución de Poisson	17
1.4. Variables aleatorias absolutamente continuas	18
1.4.1. Distribución Uniforme	18
1.4.2. Densidad asociada a una variable aleatoria	20
1.4.3. Distribución Normal	21
1.4.4. Distribución Exponencial	23
1.4.5. Distribución T de Student	24
1.4.6. Distribución <i>Chi-cuadrado</i> : χ_k^2	24
1.5. Esperanza y Varianza de una variable aleatoria	25
1.5.1. Esperanza de $X \sim Bin(n, p)$	27
1.5.2. Varianza	28
1.6. Independencia de Variables	29
1.6.1. Suma de variables aleatorias	30
2. Intervalos de confianza, Pruebas de hipótesis.	32
2.1. Estimación de $E(X)$ y $Var(X)$	33
2.2. Intervalos de confianza	34
2.2.1. Intervalos de confianza para datos normales, σ conocido.	34
2.2.2. Intervalos de confianza para datos normales, σ desconocido.	36
2.2.3. Intervalo de confianza para datos cualquiera, σ conocido	37
2.2.4. Intervalo de confianza para datos cualquiera, σ desconocido	37
2.2.5. Intervalos de confianza para proporciones	38
2.3. Pruebas de hipótesis	38
2.3.1. Pruebas de hipótesis unilaterales, datos normales, σ conocido	39
2.3.2. Pruebas de hipótesis unilaterales, datos normales, σ desconocido	40
2.3.3. Pruebas de hipótesis bilaterales, datos normales, σ conocido	41
2.3.4. Pruebas de hipótesis bilaterales, datos normales, σ desconocido	42

2.3.5. Pruebas de hipótesis para proporciones	42
2.4. Pruebas de bondad de ajuste	42
3. Test de aleatoriedad	45
3.1. Introducción	45
3.1.1. Test de Rachas	45
3.1.2. Test de Spearman	47

Capítulo 1

Conteo, Probabilidad, Variables Aleatorias

1.1. Conteo

En la presente sección mencionaremos de forma breve algunos conceptos que serán muy importantes para el cálculo de probabilidades. El entendimiento de los mismos no requieren de conocimientos previos de ningún tipo. Para un estudio más completo de los mismos, una referencia clásica, con ejemplos y ejercicios resueltos es [2].

1.1.1. Permutaciones

Vamos a empezar con un ejemplo muy simple, consideremos las tres primeras letras, A, B y C del abecedario, queremos contar de cuantas maneras distintas se pueden permutar u ordenar. Es claro que las posibilidades son ABC, ACB, BCA, BAC, CBA y CAB. Es decir tenemos 6 formas distintas. Si escribimos todas las formas de ordenar las cuatro primeras letras ABCD, vemos que en este caso nos quedan 24 posibilidades, no obstante esta forma de *contar* se vuelve inapropiada cuando por ejemplo tomamos las 10 primeras letras. Veamos a continuación una forma de contar las maneras de ordenar una cantidad cualquiera n de objetos.

Definición 1.1. Supongamos que tenemos n objetos distintos. Queremos contar la cantidad de formas distintas de ordenarlos en n lugares. En el primer lugar podemos poner cualquiera de los n elementos, en el segundo lugar, dado que ya tomamos un objeto para el primer lugar, nos quedan $n - 1$ objetos. Por lo tanto para el primer y segundo lugar tenemos $n \times (n - 1)$ posibilidades. Razonando de esta manera, llegamos que tenemos

$$n \times (n - 1) \times (n - 2) \times \cdots \times 2 \times 1,$$

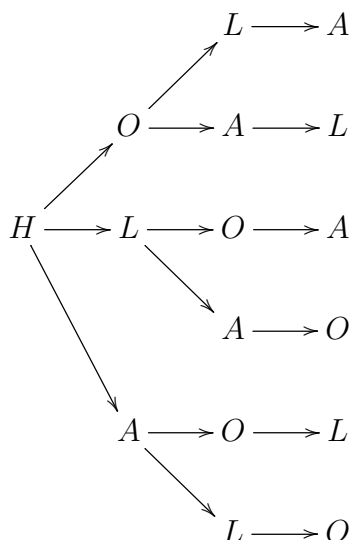
posibles ordenaciones, o *permutaciones* de los n objetos.

En general se emplea la notación $n! := n \times (n - 1) \times (n - 2) \times \cdots \times 2 \times 1$. y se asume, por definición que $0! = 1$.

Observemos que, dado que, por definición $(n - 1)! = (n - 1) \times (n - 2) \times \cdots \times 2 \times 1$ tenemos que $n! = n \times (n - 1)!$.

Ejemplo 1.2. Supongamos que queremos contar de cuantas maneras se pueden ordenar las letras de la palabra *HOLA*. Observemos primero que, dado que no se repiten letras, es un problema de conteo de permutaciones de 4 elementos distintos. En el primer lugar

podemos poner la letra H , la O la L o la A , es decir tenemos 4 posibilidades. En el segundo lugar, dado que elegimos 1 de las letras, tenemos 3 posibilidades, en el tercer lugar tenemos 2 y en el último nos queda 1 sola letra sin colocar. En la Figura (1.2) se muestra el árbol de las posibilidades, si empezamos con la letra H .



Cuadro 1.1: Si comenzamos con H tenemos 6 posibilidades, dado que podemos empezar con cualquiera de las 4 letras (y que comenzar con letras distintas da ordenaciones distintas) tenemos $6 \times 4 = 24 = 4!$

1.1.2. Arreglos

Siguiendo nuestro ejemplo sencillo de las primeras tres letras del alfabeto, supongamos que queremos contar de cuantas maneras se pueden ordenar esas tres letras en 2 lugares. Es claro que ahora las *palabras* van a estar formadas por dos letras, tomadas de las tres que tenemos, es decir las posibilidades son: AB,BA, AC,CA,BC y CB. Nos da 6 igual que antes. Veamos que pasa ahora con las cuatro primeras letras. Aquí las posibilidades son AB,AC,AD,BA,CA,DA,BC,BD,CD,DC,CB,DB. En este caso nos da 12, es decir la mitad que el total de permutaciones de 4 elementos distintos. Comprobar que si tenemos 4 letras, y tomamos 3 lugares, y nos importa el orden, tenemos 4 posibilidades. Veamos como contarlas cuando son más de 4 elementos.

Definición 1.3. Supongamos que tenemos n objetos distintos. Queremos contar la cantidad de formas distintas de ordenarlos en k lugares (con $k \leq n$). Es importante tener en cuenta que importa el orden en que se colocan los objetos, por ejemplo, si queremos ordenar las letras de la palabra *HOLA* en tres lugares, el ordenamiento *OLA* es distinto que *LOA*. En este caso tenemos 4 letras para colocar en el primer lugar, 3 en el segundo lugar y finalmente 2 en el tercero; nos da $4 \times 3 \times 2$ posibilidades. En general, si tenemos n objetos distintos (razonando de la misma manera que antes) en el primer lugar podemos poner cualquiera de los n elementos, en el segundo $n - 1$, en el tercero $n - 2$ y en el lugar k (con $k \leq n$) $n - k + 1$. Es decir en total tenemos

$$n \times (n - 1) \times \cdots \times (n - k + 1).$$

En general se denota $A_k^n := n \times (n - 1) \times \cdots \times (n - k + 1)$.

Ejemplo 1.4. ¿Cuántos números de tres cifras se pueden formar con los dígitos 0, 1, 2, 3 y 4 si no se permite la repetición de dígitos? Observemos que en las unidades podemos poner cualquiera de los 5 dígitos, esto nos deja 4 dígitos para elegir para las decenas, y finalmente, en las centenas podemos poner 3 dígitos. En total son $A_3^5 = 5 \times 4 \times 3$. Observemos además que podemos escribir

$$A_3^5 = 5 \times 4 \times 3 = 5 \times 4 \times 3 \times \frac{2 \times 1}{2 \times 1} = \frac{5!}{2!} = 60$$

Observación 1.5. Observar que

$$A_k^n = \frac{n!}{(n-k)!}$$

en particular $A_n^n = n!$ $A_1^n = n$.

1.1.3. Combinaciones

Continuando con nuestro ejemplo, supongamos que queremos contar la cantidad de maneras distintas en que se pueden ordenar las tres primeras letras del abecedario, en 2 lugares, pero nos es indistinta la ordenación. Es decir, no distinguimos AB de BA, ni AC de CA, ni CB de BC. Observemos que esto es lo mismo que agrupar de a pares de letras, o, lo que es lo mismo, construir conjuntos de dos elementos (ya que en los conjuntos no importa el orden). Tenemos entonces tres conjuntos: el que contiene las letras A y B: $\{A, B\}$, el que contiene las letras A y C: $\{A, C\}$ (observemos que como conjunto es distinto al anterior). Y el que contiene las letras B y C: $\{B, C\}$ que es distinto a los anteriores. Son entonces tres posibilidades. Observemos que nos dio exactamente la mitad que en el caso en que tenemos en cuenta el orden, ¿es esto intuitivo?. Si repetimos el razonamiento con 4 letras, y 2 lugares nos da los conjuntos $\{A, B\}$ $\{A, C\}$, $\{A, D\}$, $\{C, B\}$, $\{C, D\}$ y $\{B, D\}$, es decir, nos da 6, nuevamente nos dio la mitad que teniendo en cuenta el orden. No obstante, si tomamos 4 letras y 3 lugares tenemos 4 conjuntos (escribirlos) y esto es la misma cantidad que si hubiésemos tenido en cuenta el orden. En general no tener en cuenta el orden da lugar a menos posibilidades que al tenerlo en cuenta ya que al no tener en cuenta el orden estamos agrupando ordenaciones (por ejemplo el grupo que contiene A,B y B,A).

Definición 1.6. Supongamos que tenemos n objetos distintos y que queremos contar la cantidad de subconjuntos de k elementos (con $k \leq n$) formado con k de los n elementos. Al hablar de subconjuntos, no importa el orden. Por ejemplo: tenemos $n = 5$ objetos diferentes A,B,C, D y E y queremos contar cuántos subconjuntos de $k = 3$ elementos se pueden formar. En el subconjunto formado por los elementos A, B y C tenemos $3! = 6$ ordenaciones posibles de dichos elementos que producen el mismo subconjunto. Por lo tanto basta contar las ordenaciones de 3 elementos y luego agrupar las que forman el mismo subconjunto. O, lo que es lo mismo, si tenemos el número de subconjuntos, multiplicamos por $3!$ y nos da el número de ordenaciones o arreglos de 3 elementos. Es decir, en base a esto último, el número que queremos (que denotaremos $\binom{n}{k}$) tiene la propiedad de que $\binom{n}{k} \times k! = A_k^n$. Si usamos la observación anterior.

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}. \tag{1.1}$$

Ejemplo 1.7.

- Se juega a un juego del tipo 5 de Oro: hay que acertar 5 números, elegidos dentro de 44 posibilidades. ¿Cuántas jugadas posibles hay?. Lo primero y más importante que hay que observar es que el orden no importa, importa a qué números jugamos, y no en qué orden. Por lo tanto estamos ante un problema de combinaciones. Tenemos $n = 44$ objetos distintos, y queremos contar la cantidad de subconjuntos de $k = 5$ elementos, es decir la respuesta es $\binom{44}{5}$.
- De entre 8 personas debemos formar un comité de cinco miembros. ¿Cuántas posibilidades diferentes existen para formar el comité?. Nuevamente, como hablamos de un comité de personas, no importa el orden en que las elegimos. La respuesta es entonces $\binom{8}{5}$.

Ejercicio 1.8. Veamos un ejemplo que nos será de utilidad más adelante

- Supongamos que tenemos una urna con 6 bolas de color rojo (que denotaremos R_1, R_2, R_3, R_4, R_5 y R_6 y 3 de color negro (N_1, N_2 y N_3). Vamos a extraer 5 bolas. Contaremos la cantidad de formas de extraerlas de modo tal que 3 sean de color rojo, y 2 de color negro. Es decir una posible extracción es R_1, R_2, R_5, N_1, N_3 mientras que otra *distinta* es R_3, R_2, R_5, N_1, N_3 . Por otro lado, no nos importa el orden es decir R_1, R_2, R_5, N_1, N_3 se puede sacar de $5!$ formas distintas, (y lo mismo para cualquier extracción). Como no nos importa el orden, vamos a usar combinaciones. Observemos que tenemos $\binom{6}{3}$ formas *distintas* de extraer 3 bolas rojas, de las 6 bolas rojas que hay en total. Por otro lado, *para cada* extracción de 3 bolas rojas, podemos completar las 5 que tenemos que sacar en total eligiendo 2 de las 3 bolas negras, es decir tenemos $3 = \binom{3}{2}$ formas de extraer dichas bolas. En total tenemos entonces

$$\binom{6}{3} \times \binom{3}{2}.$$

- Supongase que se tiene una urna con 15 bolas de color rojo y 7 de color negro. ¿De cuantas formas se pueden extraer 9 de modo tal que 6 sean de color rojo y 3 de color negro?

Observación 1.9. *Observar que*

- $\binom{n}{n} = 1$ $\binom{n}{1} = n$. Nótese que $A_1^n = \binom{n}{1}$. Sin usar las formulas respectivas, explicar por qué es razonable ese resultado.
- $\binom{n}{k} = \binom{n}{n-k}$. Esta fórmula, que se demuestra fácilmente a partir de (1.1) nos dice que, contar subconjuntos de k elementos, es lo mismo que elegir $n - k$ elementos que no vamos a incluir en el subconjunto.

1.2. Probabilidad

Aquí mencionaremos de forma breve, y sin el rigor que la teoría de la probabilidad requiere, los conceptos de probabilidad, independencia, probabilidad condicional, y algunas de las propiedades de la misma. Un libro recomendable para profundizar en estos temas es [1] o [3].

El objeto de estudio de la probabilidad son los experimentos cuyos resultados no se pueden *determinar* de antemano. Es decir la repetición del experimento no nos conduce necesariamente al mismo resultado (en este sentido es que decimos que el resultado es aleatorio

o depende del azar). Por ejemplo arrojar un dado, una moneda. etc. El ejemplo clásico donde esto no sucede, son los experimentos de la química donde siempre que se combinan, en iguales condiciones, determinadas sustancias, se obtiene el mismo resultado. Por otro lado, en probabilidad supondremos que el conjunto de resultados posibles es conocido de antemano. En el caso del dado, sabemos que solo puede salir 1,2,3,4,5 o 6 en la cara superior. El conjunto de resultados posibles se llama espacio muestral, y suele denotarse con la letra griega Ω . Finalmente, si bien no sabemos con anterioridad el resultado del experimento, supondremos que es uno. Es decir, al arrojar un dado no puede salir 3 y 5 en la cara superior.

De manera intuitiva y a partir de la experiencia cotidiana (que motivó y dio origen a la formalización teórica posterior) la probabilidad de un suceso, es la frecuencia con que este se da. Así por ejemplo, la probabilidad de que salga cara al arrojar una moneda es la mitad, $1/2$, de la cantidad de tiradas. De la misma manera, la probabilidad de que al tirar un dado salga 2, es $1/6$ ya que si arrojamos *muchas veces* el dado, $1/6$ de las mismas saldrá 2 (y es igual a la probabilidad de que salga cualquiera de los 6 números). Observemos que, lo que estamos haciendo es asignar *el mismo* número entre 0 y 1, que llamamos probabilidad, a cada uno de los posibles resultados, y dado que son 6, solamente podemos asignar $1/6$. Esta es la forma más simple de asignar probabilidades a *sucesos*, el modelo se conoce como *casos favorables sobre casos posibles*. Nuestra intuición nos dice que en general no es cierto que los sucesos tengan la misma probabilidad siempre, pensemos por ejemplo que queremos asignar una probabilidad al color en que está la luz del semáforo cuando llegamos a una esquina, intuitivamente es *menos probable* que la luz este en color amarillo, ya que el tiempo de duración del color amarillo es menor. Si pensamos que el tiempo en que esta la luz en amarillo es $1/4$ del tiempo en que esta en rojo, y que el tiempo en que esta en rojo y en verde son iguales, tenemos que, al asignar probabilidades la probabilidad de que sea amarillo es $4/9$ (¿por qué?). Como mencionamos antes, los sucesos a los que vamos a asignar probabilidades (es decir números entre 0 y 1) serán subconjuntos de un cierto conjunto *conocido* Ω . Por ejemplo, en el caso de los resultados de tirar una moneda $\Omega = \{0, 1\}$ donde 0 representa cara, y 1 número. En el caso de los resultados de tirar un dado, $\Omega = \{1, 2, 3, 4, 5, 6\}$ y un suceso es, por ejemplo $A = \{1, 2\}$. En este sentido la probabilidad de A , que denotaremos $P(A)$ es la probabilidad de que al tirar un dado salga 1 o 2, que, intuitivamente es $2/6$. Si tiramos dos dados tenemos que Ω es el conjunto de las 36 parejas (i, j) con $i, j = 1, \dots, 6$ (ver tabla (1.2) más adelante) y A puede ser por ejemplo el suceso {la suma de los resultados es 7}, etc.

1.2.1. Casos favorables sobre casos posibles: Tirar dos dados

En la tabla (1.2) se muestran los posibles resultados de tirar dos dados, vamos a suponer que podemos distinguirlos entre si, por ejemplo uno tiene los números en rojo y el otro en negro. Para simplificar la explicación, cuando decimos, por ejemplo, que el dado rojo es mayor que el dado negro, nos referimos a que el resultado que sale en la cara superior del dado cuyos números son rojos es mayor que el resultado que sale en la cara superior del dado cuyos números son negros.

Tenemos 36 posibles resultados (casos posibles). Es razonable pensar que la frecuencia con que cualquiera de ellos aparece es la misma, y por lo tanto $1/36$. Así, por ejemplo, la probabilidad de que salga (1, 1) es igual a la de que salga (1, 3) es igual a la de que salga (3, 1) y es $1/36$. Observemos que si queremos calcular la probabilidad de que salga un 1 y un 3, tenemos 2 *casos favorables* a considerar, el (1, 3) y el (3, 1) y por lo tanto la probabilidad es $1/36 + 1/36 = 1/18$, lo que hacemos es *sumar* las probabilidades de los

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Cuadro 1.2: Resultados de tirar 2 dados

sucesos. Si queremos calcular la probabilidad de que los dos números sean iguales, tenemos 6 *casos favorables*, es decir los 6 que forman la diagonal. Por lo tanto la probabilidad es $6/36 = 1/6$. Observemos que lo que hacemos nuevamente es *sumar* las probabilidades de los sucesos (1, 1) hasta (6, 6). Supongamos que queremos calcular la probabilidad de que el número que sale en uno de los dados sea estrictamente mayor que el que sale en el otro. Si observamos la tabla, tenemos que contar todos los elementos, excepto los de la diagonal. Esto nos da $36 - 6 = 30$ *casos favorables*. Es decir la probabilidad es $30/36$. Veamos que este número lo podemos calcular de otra manera: si contamos la cantidad de posibles resultados para los cuales el dado rojo es mayor que el dado negro esto nos da 15 casos (correspondientes a las entradas de la tabla cuya fila es estrictamente menor que la columna), es decir la probabilidad de que el dado rojo sea mayor que el negro es $15/36$. Por otro lado, si contamos la cantidad de casos en los que el dado rojo es menor que el dado negro obtenemos nuevamente 15 casos (observar que corresponden a las entradas de la tabla cuya fila es estrictamente mayor que la columna). Es decir la probabilidad de que el dado rojo sea menor que el negro es $15/36$. Es claro que la probabilidad de que un dado sea menor que el otro, es la *suma* de estos dos sucesos, y es, como habíamos calculado antes $15/36 + 15/36 = 30/36$.

El ejemplo anterior no conduce a una propiedad de la probabilidad, (que será cierta en general, no solo en el modelo de casos favorables sobre casos posibles). Si tenemos A y B dos posibles resultados o conjunto de resultados de un experimento (por ejemplo $A = \{ \text{el dado rojo es menor que el dado negro} \}$ y $B = \{ \text{el dado rojo es mayor que el dado negro} \}$), de modo tal que si siempre que pasa A no sucede B (es decir son disjuntos). Entonces

$$P(A \cup B) = P(A) + P(B). \tag{1.2}$$

Esto quiere decir que la probabilidad de la unión de *dos* sucesos disjuntos (es decir, la probabilidad de que *alguno* de ellos ocurra) es igual a la suma de las probabilidades de los sucesos.

Análogamente si tenemos A, B, C tres sucesos, tal que si sucede uno, no sucede ninguno de los otros 2, entonces

$$P(A \cup B \cup C) = P(A) + P(B) + P(C). \tag{1.3}$$

Por ejemplo, $A = \{ \text{el dado rojo es menor que el dado negro} \}$, $B = \{ \text{el dado rojo es mayor que el dado negro} \}$ y $C = \{ \text{ambos dados son iguales} \}$ (es decir el número que sale es el mismo). Es claro que $A \cup B \cup C$ incluye los 36 casos. Por lo tanto $P(A \cup B \cup C) = 36/36 = 1$, y por lo que calculamos antes esto coincide con $P(A) + P(B) + P(C)$.

Observación 1.10. *Es importante observar que ni (1.2) ni (1.3) son ciertos en general. Pensemos por ejemplo que $A = \{ \text{el dado rojo es menor o igual que el dado negro} \}$, mientras*

que B es el suceso {el dado rojo es mayor o igual que el dado negro }, En este caso $P(A \cup B) = 1$ mientras que $P(A) = 21/36 = P(B)$ es decir $P(A) + P(B) = 42/36 > 1$. Tampoco es cierto en general que si se cumple (1.2) o (1.3) entonces los sucesos son disjuntos.

Observación 1.11. En general se cumple que $P(A \cup B) \leq P(A) + P(B)$ y $P(A \cup B \cup C) \leq P(A) + P(B) + P(C)$. Usando la tabla de tirada de dos dados, encontrar ejemplos de conjuntos A y B donde se cumple el menor estricto.

Ejemplo 1.12. Consideremos el suceso $A = \{ \text{el dado rojo es menor que el dado negro} \}$, y el suceso complementario, que denotaremos $A^c = \{ \text{el dado rojo no es menor que el dado negro} \}$ (es decir es mayor o igual) . Observemos que dichos suceso son disjuntos y que la probabilidad de la unión de los mismos es 1. Por lo tanto si usamos (1.2) obtenemos que

$$1 = P(A \cup A^c) = P(A) + P(A^c).$$

Este razonamiento, que hicimos para un A particular es cierto en general ya que A y su complementario A^c son disjuntos. De dicho resultado se sigue que, para cualquier suceso A

$$P(A) = 1 - P(A^c). \tag{1.4}$$

Ejemplo 1.13. Consideremos el suceso $A = \{ \text{la suma de los resultados es 7} \}$, y $B = \{ \text{el dado rojo es menor que el dado negro} \}$. Observando la tabla vemos que $P(A) = 6/36 = 1/6$ y $P(B) = 15/36$. Consideremos ahora el suceso: {el dado rojo es menor que el dado negro y la suma de los resultados es 7 }. Este suceso, que se denota $A \cap B$, esta formado por los resultados $(6, 1), (5, 2), (4, 3)$, por lo tanto $P(A \cap B) = 3/36 = 1/12$. Por otro lado $P(A) \times P(B) = 1/6 \times 15/36 < 1/12$. Es decir, *no es cierto en general* que para todo A y B $P(A \cap B) = P(A) \times P(B)$, el hecho de que esta última igualdad ocurra es importante y es el objeto de estudio de la siguiente sección.

1.2.2. Independencia

Vamos a introducir ahora uno de los conceptos más importantes del curso, el concepto de independencia de sucesos. Supongamos que tiramos una moneda y sale cara. Luego volvemos a tirar la moneda y sale número. Desde un punto de vista intuitivo, el hecho de que haya salido cara en la primer tirada, *no influye* sobre el resultado de la segunda tirada, dicho de otra manera, para saber cual será el resultado de la segunda tirada, no nos aporta información saber que la primer tirada fue cara. Lo mismo nos dice la intuición respecto al resultado de tirar un dado, y luego otro, etc. Intuitivamente, cuando el resultado de la realización de un experimento no nos aporta *información* que nos permita deducir a priori, el resultado de otro experimento, decimos que estos son independientes. Observemos que no estamos diciendo que el hecho de que se haya dado un determinado resultado es incompatible con que suceda el otro. Veamos como se formaliza en términos matemáticos esta idea intuitiva.

Ejemplo 1.14. Consideremos nuevamente los 36 resultados de tirar 2 dados, el suceso $A = \{ \text{la suma de los dados es 7} \}$, y el suceso $B = \{ \text{el dado rojo es 1} \}$. Sabemos que $P(A) = 1/6$ y que $P(B) = 1/6$, además, la probabilidad de que la suma sea 7 y que el dado rojo sea 1 (es decir $A \cap B$) es $1/36$, observemos que

$$P(A \cap B) = P(A) \times P(B). \tag{1.5}$$

Definición 1.15. Si tenemos A y B dos sucesos tal que la identidad (1.5) se cumple, decimos que los sucesos A y B son independientes.

Ejemplo 1.16. Veamos que se confirma lo que nos dice la intuición respecto a la independencia de la primer y segunda tirada. Si en lugar de tirar los dos dados a la vez, tiramos primero el dado negro y luego el rojo la tabla de posibilidades es (1.2). Es razonable pensar que aquí también cualquier pareja de resultados tiene la misma probabilidad, es decir $1/36$. Si calculamos la probabilidad de que, por ejemplo, el dado rojo sea i (donde i es cualquier número entre 1 y 6), tenemos 6 casos (correspondientes a la columna i de la tabla) por lo tanto el suceso $A = \{ \text{el dado rojo es } i \}$, tiene probabilidad $1/6$. Por otro lado, si consideramos el suceso $B = \{ \text{el dado negro es } j \}$ (donde j es cualquier número entre 1 y 6) tenemos que $P(B) = 1/6$ (basta considerar los 6 casos de la fila j de la tabla). El suceso $A \cap B$ que corresponde a que el dado rojo sea i y que el dado negro sea j nos da únicamente la pareja (j, i) que tiene probabilidad $1/36$. Es decir se da la igualdad (1.5) y por lo tanto son independientes.

Definición 1.17. Consideremos tres sucesos A , B y C , si

$$\begin{aligned} P(A \cap B \cap C) &= P(A) \times P(B) \times P(C) \\ P(A \cap B) &= P(A) \times P(B) \\ P(A \cap C) &= P(A) \times P(C) \\ P(B \cap C) &= P(B) \times P(C) \end{aligned}$$

decimos que son independientes.

Ejercicio 1.18. Demostrar, a partir de los 6 posibles resultados de tirar 1 moneda 3 veces, que las tiradas son independientes.

Observemos que en el Ejemplo (1.13) tenemos dos sucesos que *no* son independientes, (como vimos, no se cumple (1.5)). Esto nos conduce al siguiente concepto:

Definición 1.19. Probabilidad Condicional Cuando dos sucesos no son independientes, es decir saber que pasó un suceso A nos aporta información sobre la probabilidad de que pase otro suceso B tiene interés calcular la probabilidad de que pase B *dado que* pasó A , esto se denota $P(B|A)$. Si, por ejemplo, siempre que pasa A pasa B , es intuitivo que $P(B|A)$ tiene que ser 1. En el caso de que sean independientes, es razonable también que $P(B|A) = P(B)$ (saber que pasó A no me aporta información y por lo tanto no hace a B más o menos probable).

Consideremos el suceso $A = \{ \text{la suma de los dados es } 7 \}$ y $B = \{ \text{el dado rojo es par} \}$. En este caso $P(A) = 1/6$ y $P(B) = 1/2$. Observemos que $P(B|A)$ es la probabilidad de que el dado rojo sea par, *dado que* la suma fue 7. Es decir, sabemos que la suma fue 7 queremos ver que tan probable es que el dado rojo sea par. Si sabemos que la suma fue 7 tenemos que *restringirnos* a las 6 posibles tiradas que hacen que la suma sea 7, es decir, el conjunto de casos posibles tiene 6 elementos. De esos 6 solamente 3 resultados tienen el dado rojo par: $(5, 2)$, $(3, 4)$ y $(1, 6)$, es decir tenemos 3 casos favorables, por lo tanto $P(B|A) = 3/6 = 1/2$. Observemos que $P(A \cap B) = 3/36 = 1/12$ y que se cumple que

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \quad (1.6)$$

La identidad (1.6) *define* la probabilidad condicional $P(B|A)$, observemos que para que dicha expresión tenga sentido es necesario que $P(A) \neq 0$.

1.3. Variables aleatorias discretas

1.3.1. Distribución Binomial

En esta sección vamos a centrarnos en el estudio de un tipo de experimento muy particular en el cual existen exactamente 2 posibles resultados, por ejemplo los resultados que surgen de tirar una moneda. Otro ejemplo simple consiste en considerar si al tirar un dado sale o no un cierto número. Usualmente se llama *éxito* a uno de estos resultados y *fracaso* al otro (éxito podría ser que salga 5 al tirar un dado). Es claro que como tenemos 2 posibilidades podemos decir que la frecuencia con que ocurre éxito si repetimos el experimento muchas veces (esto es $P(\text{éxito})$) es un número p donde $0 \leq p \leq 1$, mientras que la frecuencia con que *no* ocurre el éxito (es decir ocurre el *fracaso*) es $1 - p$ (deducirlo a partir de (1.4)). Lo que nos interesará es, fijada de antemano la cantidad de veces que repetiremos el experimento (y bajo la hipótesis de que *los experimentos se realizan de forma independiente*), contabilizar que tan probable es obtener determinada cantidad de éxitos. Por ejemplo, si tiramos 50 veces el dado, que tan probable es que el 5 salga exactamente 10 veces, es decir, la probabilidad de que la cantidad de éxitos sea 10, o, lo que es lo mismo, que la cantidad de fracasos sea exactamente 40.

Veamos como es la construcción del marco teórico de este modelo. Es natural que, para modelar el experimento que consiste en tirar 3 veces un dado y mirar la cantidad de veces que salió el número 5 pensemos en tomar Ω como el conjunto de ternas de ceros y unos, donde 1 es éxito, es decir salió el 5. Así por ejemplo un posible resultado es $(1, 0, 0)$, que representa el experimento en el cual en la primer tirada salió el 5 mientras que en las dos siguientes no salió el 5. Es claro que dicho resultado es distinto de $(0, 0, 1)$ o de $(1, 1, 1)$. Observemos que Ω tiene 6 elementos. Observemos además que, si las tiradas son independientes, el suceso $(1, 0, 0)$ tiene probabilidad $1/6 \times 5/6 \times 5/6$ ya que $P((1, 0, 0)) = P(\{\text{En la primer tirada sale el 5}\} \cap \{\text{En la segunda tirada no sale el 5}\} \cap \{\text{En la tercer tirada no sale 5}\})$ y como son independientes, el resultado se sigue de (1.17).

Por otro lado el el suceso $(1, 1, 1)$ tiene menor probabilidad ($P((1, 1, 1)) = 1/6 \times 1/6 \times 1/6$, esto se sigue razonando de forma análoga a como hicimos con $(1, 0, 0)$). De esto último deducimos que *no* estamos ante un modelo de casos favorables sobre casos posibles.

Ejercicio 1.20. Asignar a cada una de las 6 ternas de ceros y unos sus probabilidades.

Volviendo al problema que mencionamos antes, queremos calcular la probabilidad de tener una determinada cantidad de éxitos. En el caso de repetir 3 veces el experimento de tirar el dado, si éxito es que salga el número 5 es claro que podemos tener:

- 1) 0 éxito (como en $(0, 0, 0)$),
- 2) exactamente 1 éxito, (como en $(1, 0, 0)$ o $(0, 1, 0)$ o $(0, 0, 1)$),
- 3) exactamente 2 éxitos (como en $(1, 1, 0)$ o $(0, 1, 1)$ o $(1, 0, 1)$),
- 4) 3 éxitos (como en $(1, 1, 1)$)

Si lo que nos interesa es contar la cantidad de éxitos, los 3 resultados posibles del caso 2 se pueden agrupar en el suceso {exactamente 1 éxito} y podemos asignarle a esos 3 elementos de Ω el número 1. Análogamente, los 3 elementos del caso 3 se pueden agrupar en el suceso {exactamente 2 éxitos} y podemos asignarle a esos 3 elementos de Ω el número 2. Finalmente, razonando de manera análoga, al elemento del caso 1 le asignamos el 0 mientras que al del caso 4 le asignamos el 3. Ver figura (1.1) Esta función (la cual denotaremos con la

letra X) que hemos definido, de Ω en los números $\{0, 1, 2, 3\}$ se llama **variable aleatoria con distribución binomial**. En general, una variable aleatoria será una función X de Ω en los números reales \mathbb{R} . Y nos interesará calcular cosas como $P(X = t)$ donde t es un número real, entendiendo por eso la probabilidad de el conjunto de los elementos de Ω que hacen que $X = t$.

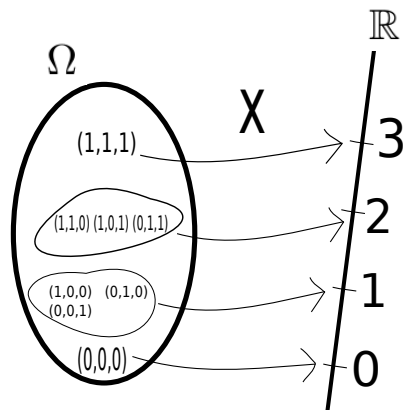


Figura 1.1: Variable aleatoria Binomial en el caso de 3 experimentos

En nuestro ejemplo anterior si X tiene distribución binomial, $P(X = 2) = P((1, 1, 0) \cup (0, 1, 1) \cup (1, 0, 1)) = P((1, 1, 0)) + P((0, 1, 1)) + P((1, 0, 1))$ donde en esta igualdad hemos usado la propiedad (1.3). Dado que $P((1, 1, 0)) = P((0, 1, 1)) = P((1, 0, 1)) = (1/6)^2 \times 5/6$ tenemos que

$$P(X = 2) = 3 \times (1/6)^2 \times 5/6. \quad (1.7)$$

De la misma forma se llega a que

$$P(X = 1) = 3 \times (1/6) \times (5/6)^2 \quad (1.8)$$

Vamos a calcular $P(X = 0)$ tenemos un solo caso (caso 1), que tiene probabilidad $(5/6)^3$ observemos que esto es lo mismo que hacer

$$P(X = 0) = 1 \times (1/6)^0 \times (5/6)^3 \quad (1.9)$$

Nótese que en (1.7), (1.8) y (1.9) obtuvimos, para $k = 0, 1, 2, 3$ éxitos:

$$P(X = k) = (\text{formas de poner } k \text{ unos en } 3 \text{ lugares}) \times (1/6)^k \times (5/6)^{3-k}$$

Veamos como contar en términos de combinaciones la cantidad de formas de poner k unos en 3 lugares (con $k = 0, 1, 2, 3$). Recordemos que $\binom{n}{k}$ cuenta la cantidad de subconjuntos de k elementos que podemos hacer con n objetos diferentes. Podemos pensar que los $n = 3$ objetos distintos son los lugares donde vamos a poner los k unos. Es decir, en la primer tirada, en la segunda, o en la tercera. Por ejemplo, para $k = 2$ elegir *primer tirada* y *tercer tirada* produce la misma cantidad de éxitos que elegirlos en el orden inverso: *tercer tirada* y *primer tirada*, pero es distinto que elegir *primer tirada* y *segunda tirada*. Es decir no importa el orden, y tenemos $\binom{3}{2}$ formas de elegir esos lugares.

Definición 1.21. Bin(n,p): Siguiendo el razonamiento que hicimos antes, si tenemos n experimentos, y $p = P(\text{éxito})$ decimos que $X : \Omega \rightarrow \{0, 1, 2, \dots, n\}$ tiene distribución binomial con parámetros n, p que se denota $X \sim \text{Bin}(n, p)$ si

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, 2, \dots, n \quad (1.10)$$

Ejercicio 1.22. Usando la expresión anterior calcular la probabilidad de que al tirar 50 veces un dado salgan exactamente 10 cincos.

Ejercicio 1.23. Supongamos que $X \sim Bin(n, 1/2)$

- Intuitivamente, sin calcular, ¿Cuánto da $P(X = 0) + P(X = 1) + P(X = 2) + \dots + P(X = n)$?
- Usando la parte anterior y (1.10) deducir que $\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = 2^n$ y observar que 2^n es exactamente la cantidad total de secuencias de largo n que podemos formar con ceros y unos.

1.3.2. Distribución Geométrica

Consideremos que estamos interesados en repetir un experimento, que tiene dos posibles resultados, uno de ellos que llamaremos éxito, y el otro fracaso, hasta que se da el éxito por primera vez. Supondremos además que el experimento lo repetimos de forma independiente de lo que sucedió antes. Por ejemplo podemos pensar que tiramos una moneda hasta que salga cara, o un dado hasta que salga 5 etc. En el caso del dado, el 5 puede aparecer por primera vez en la primer tirada, o puede que en la primer tirada no salga 5 pero en la segunda si, etc. La probabilidad de que salga 5 en la primer tirada es $1/6$. La probabilidad de que salga por primera vez en la segunda tirada es la probabilidad de la intersección de los sucesos $A = \{ \text{no sale 5 en la primer tirada} \}$ y $B = \{ \text{sale 5 en la segunda tirada} \}$. Dado que estamos suponiendo que los experimentos se realizan de forma independiente, es decir saber que salió 5 en la primer tirada *no influye* sobre el resultado de la segunda, tenemos que $P(A \cap B) = P(A) \times P(B)$. Como $P(A) = 5/6$ mientras que $P(B) = 1/6$ tenemos que la probabilidad de que salga el 5 por primera vez en la segunda tirada es $5/6 \times 1/6$.

Ejercicio 1.24. Probar que la probabilidad de que salga 5 por primera vez en la tercer tirada es $(5/6)^2 \times 1/6$.

En general, razonando de esta manera, es fácil ver que la probabilidad de que salga 5 por primera vez, en la tirada k es la intersección de los sucesos independientes $A = \{ \text{no sale 5 en las tiradas } 1, 2, \dots, k-1 \}$ y $B = \{ \text{sale 5 en la tirada } k \}$. Dicha intersección tiene probabilidad $(5/6)^{k-1} \times 1/6$. Esto nos conduce a la siguiente definición:

Definición 1.25. Decimos que una variable aleatoria X tiene **distribución geométrica** de parámetro p con $0 \leq p \leq 1$ (que denotaremos $X \sim Geo(p)$) si

$$P(X = k) = (1 - p)^{k-1} \times p \quad k = 1, 2, 3, \dots$$

Observemos que, a diferencia de lo que sucede con las variables con distribución binomial de parámetros n, p , una variable con distribución geométrica es una función que toma infinitos valores $(1, 2, 3, \dots)$ mientras que una variable con distribución binomial de parámetros n, p toma valores $0, 1, 2, \dots, n$, donde n es un número fijado de antemano.

Ejercicio 1.26. Supongamos que tiramos un dado hasta que sale 3

Calcular la probabilidad de que salga por primera vez en la primer tirada o en la segunda.

Calcular la probabilidad de que haya que realizar más de 2 tiradas.

1.3.3. Distribución Hipergeométrica. Extracciones sin reposición

Veremos ahora otro tipo de variable aleatoria. Supongamos que tenemos una urna con 6 bolas de color rojo y 3 de color negro como en (1.8). Vamos a suponer que la diferencia es únicamente el color. Si extraemos una bola al azar, como la diferencia está solamente en el color, podemos agarrar cualquiera de las 9 bolas pero, intuitivamente, la probabilidad de extraer una bola roja es mayor (ya que son más) que la de extraer una bola negra. Dado que cualquiera de las 9 bolas tiene la misma probabilidad (es decir $1/9$), el espacio muestral Ω esta formado por los 9 posibles resultados. Por lo tanto el suceso $A = \{ \text{la bola extraída es roja} \}$ tiene probabilidad $6/9 = 2/3$, mientras que el suceso $B = \{ \text{la bola extraída es negra} \}$ tiene probabilidad $3/9$. Consideremos ahora el experimento que consiste en sacar 5 bolas (al mismo tiempo) de las 9, supongamos que queremos calcular la probabilidad de extraer 3 bolas de color rojo, y 2 de color negro. Como las bolas son todas idénticas, cualquier conjunto de 5 bolas tiene la misma probabilidad, por lo tanto en total tenemos $\binom{9}{5}$ (*casos posibles*) formas de extraer las 5 bolas. Como vimos en (1.8), los *casos favorables* son $\binom{6}{3} \times \binom{3}{2}$ es decir, la probabilidad que queremos calcular es

$$\frac{\binom{6}{3} \times \binom{3}{2}}{\binom{9}{5}}.$$

Ejemplo 1.27. Razonando de la misma forma que antes, si en lugar de sacar 5 bolas sacamos 3 y queremos tener 2 rojas y 1 negra, obtenemos que la probabilidad es

$$\frac{\binom{6}{2} \times \binom{3}{1}}{\binom{9}{3}} = 15/28 \approx 0,5357. \quad (1.11)$$

Una pregunta que surge naturalmente es qué sucede si en lugar de sacar las 5 bolas al mismo tiempo, las sacamos de a una (y las que vamos sacando no las devolvemos a la urna, es decir son *extracciones sin reposición*). Veamos, para el caso en que extraemos 3 bolas (sólo a efectos de hacer menos engorrosas las cuentas), que el resultado es *el mismo* que el que obtuvimos en el ejemplo (1.11). Observemos que tenemos que sumar la probabilidad de los sucesos (disjuntos) $A = \{ \text{Sale roja en la primer extracción, roja en la segunda, y negra en la tercera} \}$, $B = \{ \text{Sale roja en la primer extracción, negra en la segunda, roja en la tercera} \}$ y $C = \{ \text{Sale negra en la primer extracción, roja en la segunda, y roja en la tercera} \}$. Vamos a calcular la probabilidad de C y quedará como ejercicio verificar que D y H tienen la misma probabilidad.

Observemos que la probabilidad de que salga negra en la primer extracción $3/9$. En la segunda extracción, *dado que* sacamos una bola negra, nos quedan 8 bolas en la urna, 6 son de color rojo y 2 de color negro, la probabilidad de sacar una bola roja ahora es $6/8$. Finalmente, en la tercer extracción, *dado que* sacamos una bola negra en la primera y una roja en la segunda, tenemos 7 bolas en la urna, de las cuales 5 son rojas y 2 son negras, por lo tanto la probabilidad de sacar una bola roja es $5/7$. De esta manera, llegamos a que $P(C) = 3/9 \times 6/8 \times 5/7$. Como dijimos antes, A, B y C tienen la misma probabilidad (verificarlo!) por lo tanto la probabilidad que queremos es $3 \times P(C) = 3 \times (3/9 \times 6/8 \times 5/7) = 15/28$ que es lo mismo que obtuvimos en el ejemplo (1.11). La pregunta que nos queda por responder ahora es: ¿por qué multiplicamos las probabilidades que fuimos calculando?. Veamos que esto sale de la aplicación *en cadena* de la fórmula de probabilidad condicional dada en (1.6). Antes de eso observemos que si despejamos en (1.6) obtenemos que.

$$P(B|A) \times P(A) = P(B \cap A) \quad (1.12)$$

Volviendo a lo que queríamos calcular:

$$P(C) = P(\{\text{negra en la primera}\} \cap \{\text{roja en la segunda}\} \cap \{\text{roja en la tercera}\})$$

si tomamos en (1.12) $A = \{\text{negra en la primera}\} \cap \{\text{roja en la segunda}\}$ y $B = \{\text{roja en la tercera}\}$ obtenemos que

$$P(C) = P(\{\text{roja en la tercera}\} \mid \{\text{roja en la segunda}\} \cap \{\text{negra en la primera}\}) \times P(\{\text{roja en la segunda}\} \cap \{\text{negra en la primera}\}) \quad (1.13)$$

Si volvemos a aplicar (1.12) con $B = \{\text{roja en la segunda}\}$ y $A = \{\text{negra en la primera}\}$ obtenemos que

$$P(\{\text{roja en la segunda}\} \cap \{\text{negra en la primera}\}) = P(\{\text{roja en la segunda}\} \mid \{\text{negra en la primera}\}) \times P(\{\text{negra en la primera}\}) = 6/8 \times 3/9 \quad (1.14)$$

Por otra parte

$$P(\{\text{roja en la tercera}\} \mid \{\text{roja en la segunda}\} \cap \{\text{negra en la primera}\}) = 5/7 \quad (1.15)$$

De (1.15) (1.14) y (1.13) obtenemos $6/8 \times 3/9 \times 5/7$ que es a lo que habíamos llegado antes con lo cual queda demostrado que es lo mismo extraer las 3 bolas al mismo tiempo o de a una, siempre y cuando no se vuelvan a colocar en la urna. El problema de hacer el cálculo mediante la segunda forma está en que hay que contar todas las posibles formas de sacar una bola roja, una negra, y otra roja (en algún orden), que dio lugar a los sucesos A , B y C . En este ejemplo contarlos es simple. Luego de que sabemos *cuantos* son, basta calcular la probabilidad de uno de ellos y multiplicarlo por el número de casos ya que como los diferentes casos tienen siempre la misma probabilidad.

Definición 1.28. En general, cuando tenemos un conjunto de N objetos, de los cuales $0 \leq d \leq N$ son de un tipo (llamémosle A) y $N - d$ son de otro, la probabilidad de obtener x elementos de A (con $x = 0, 1, 2, \dots, d$) en una extracción de n elementos es

$$\frac{\binom{d}{x} \times \binom{N-d}{n-x}}{\binom{N}{n}}.$$

Decimos que una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ tiene **distribución hipergeométrica** (que denotaremos $X \sim \text{HipGeo}(d, n, N)$) si

$$P(X = x) = \frac{\binom{d}{x} \times \binom{N-d}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2, \dots, \text{mín}\{n, d\}.$$

1.3.4. Distribución de Poisson

Hasta ahora hemos definido variables aleatorias a partir de cierto tipo de experimentos, vimos que ellas encierran la información que nos permite calcular probabilidades relacionadas a ellos. Por ejemplo, la variable aleatoria con distribución binomial de parámetros n y p la asociamos al modelo que nos permite contar la cantidad de éxitos en la repetición de n experimentos independientes en los cuales la probabilidad de éxito es p . La distribución

geométrica a contar intentos hasta obtener éxito, y la hipergeométrica a contar extracciones de una urna que contiene dos tipos de elementos. En el caso de la distribución de Poisson, que definiremos más adelante, encontrar experimentos naturales que siguen dicha distribución requiere de herramientas teóricas que aún no tenemos. Por lo tanto, lo que haremos es dar la definición y hacer algunos comentarios.

Definición 1.29. Una variable aleatoria con **distribución de Poisson** es un función $X : \Omega \rightarrow \mathbb{R}$ que toma solamente los valores $0, 1, 2, 3, \dots$ con ciertas probabilidades que dependen de un parámetro $\lambda > 0$, esto se denota $X \sim Poisson(\lambda)$. Las probabilidades son:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, 3, \dots \quad (1.16)$$

Observemos que, al igual que lo que sucedía con la distribución geométrica, el recorrido de la variable es infinito (todos los números naturales). A diferencia de la geométrica (que tomaba valores a partir de 1) la Poisson toma el valor 0 con probabilidad $e^{-\lambda}$ como se sigue inmediatamente a partir de tomar $k = 0$ en (1.16). Si $X \sim Poisson(\lambda)$, dado que los *únicos* valores que toma son $0, 1, 2, 3, \dots$ tenemos que $P(X = 0) + P(X = 1) + P(X = 2) + \dots = 1$ (esta suma tiene infinitos sumandos). De esto último se deduce que si k tiende a infinito, $P(X = k)$ tiende a cero muy rápidamente. Una pregunta natural es: ¿Qué papel juega λ en la definición de la variable aleatoria?. Supongamos que tomamos, para fijar ideas $\lambda = 3$. El concepto de probabilidad como frecuencia nos dice que si sorteamos por ejemplo $n = 1$ millón de números con distribución de $Poisson(3)$, aproximadamente $e^{-3} \times n$ de veces va a salir el 0 (es decir la proporción de veces que sale 0 es $P(X = 0)$), aproximadamente $\lambda e^{-\lambda} \times n$ veces va a salir 1 (es decir la proporción de veces que sale 1 es $P(X = 1)$), aproximadamente $(\lambda^2/2) \times e^{-\lambda} \times n$ veces va a salir 2, etc. Si promediamos esos números, es decir sumamos los n números que salieron (de modo tal que si el 0 salió 1560 veces, sumamos 1560 veces 0, si el 1 salió 3450 sumamos 3450 veces 1 etc) y dividimos entre 1 millón, el número que vamos a obtener va a estar próximo a λ que en nuestro caso es 3. Es decir, λ es un parámetro (que se denomina *valor esperado*) que tiene que ver con lo que obtenemos si promediamos *muchos* datos que siguen la distribución $Poisson(\lambda)$. El concepto de valor esperado es importante en la teoría de la probabilidad, y aparecerá a lo largo del texto en muchas oportunidades.

1.4. Variables aleatorias absolutamente continuas

1.4.1. Distribución Uniforme

Hasta ahora hemos visto variables que toman una cantidad finita, o numerable de valores. Como vimos con la distribución geométrica y de Poisson, si toma infinitos valores, no puede tomarlos con la misma probabilidad ya que la suma total de las probabilidades de los valores que toma es 1. Esto quiere decir en particular que nunca vamos a poder sortear al azar, de forma equiprobable, un número natural. En este capítulo vamos a introducir un nuevo tipo de variable aleatoria, que, no solo puede tomar infinitos valores, sino que nos permitirá definir qué quiere decir sortear al azar un número *entre* 0 y 1. Dos preguntas que surgen naturalmente son

- 1) ¿Puede la variable tomar todos los números entre 0 y 1 con *la misma probabilidad* y además que esta sea no nula?

2) ¿En caso de que 1) no sea posible, puede tomar todos los números entre 0 y 1 con probabilidad no nula (aunque no sea la misma)?

Vamos a intentar responder estas preguntas, empecemos por la primera que es la más fácil. Es claro que entre 0 y 1 tenemos infinitos números, por ejemplo $1/2, 1/3, 1/4, 1/5, 1/6, \dots$. Al igual que lo que sucede con los números naturales, no podemos asignarle probabilidad positiva y *la misma* a dichos números. ¿Por qué? Supongamos que le asignamos a los infinitos números de la forma $1/n$ con $n = 2, 3, 4, 5, \dots$ probabilidad $p > 0$, es decir tenemos una variable aleatoria X tal que $P(X = 1/n) = p$ para todo $n > 1$. Entonces

$$P(X = 1/2) + P(X = 1/3) + P(X = 1/4) + \dots + P(X = 1/n) = n \times p,$$

como p es un número no nulo, fijo, $n \times p$ se puede hacer tan grande como se quiera al ir agregando sumandos (y por lo tanto mayor que 1). Con lo cual tiene que ser $p = 0$. Esto responde la pregunta 1), veamos la 2). La pregunta dos es mucho más difícil de responder y simplemente daremos una idea intuitiva de dónde está el problema. Observemos que ahora, el problema no está en que sean infinitos números ya que una variable aleatoria con distribución geométrica toma infinitos valores ($1, 2, 3, \dots$) y sin embargo pudimos asignarle probabilidad positiva a cada uno de ellos: $(1 - p)^k \times p$, sino en que entre 0 y 1 hay *muchos más* valores que puede tomar. Esto tiene que ver con el hecho de que los números entre 0 y 1 no se pueden poner en una lista y contarlos, son *no numerables*. Esto hace que la respuesta a la pregunta formulada en 2) sea negativa.

Volviendo al problema de sortear un número entre 0 y 1 de manera equiprobable, podemos decir que intuitivamente, si nuestro número a sortear puede ser cualquier número entre 0 y 1 y queremos que el sorteo sea *equiprobable* es razonable pensar que cualquier número entre 0 y $1/2$ tiene la misma probabilidad de salir que un número entre $1/2$ y 1. A su vez, la suma de las probabilidades de dichos intervalos es 1. Si bien no son disjuntos ($[0, 1/2] \cap [1/2, 1] = 1/2$) la probabilidad de que X tome el valor $1/2$ es 0, por lo que vimos antes (no puede tomar ningún valor con probabilidad positiva. Por lo tanto $P(X \in [0, 1/2]) = 1/2 = P(X \in [1/2, 1])$. Razonando de la misma manera $P(X \in [0, 1/4]) = 1/4 = P(X \in [1/4, 1/2]) = P(X \in [1/2, 3/4]) = P(X \in [3/4, 1])$, y lo mismo podemos hacer para los n intervalos $[0, 1/n], [1/n, 2/n], \dots, [(n-1)/n, 1]$ y concluir que tienen que tener probabilidad $1/n$. De esta manera vemos que la probabilidad de que la variable pertenezca a un intervalo $[a, b]$ depende *únicamente* de la longitud del intervalo: $b - a$. Consideremos la función f constante e igual a 1, definida del intervalo $[0, 1]$ en los reales. Por lo que acabamos de ver, $P(X \in [a, b])$ no es otra cosa que *el área* de la función f entre los puntos a y b . Ver figura (1.2).

Definición 1.30. Una variable aleatoria tiene **distribución uniforme** en $[0, 1]$ (que denotaremos $X \sim U([0, 1])$) si para todo $0 \leq x \leq y \leq 1$ tenemos que

$$P(X \in [x, y]) = y - x$$

Observemos que en este caso $P(X \in [x, y])$ coincide con área encerrada por el gráfico de la función constante $f = 1$ y el eje x . Observemos además que la probabilidad de que la variable pertenezca a un determinado intervalo depende *únicamente* de la longitud del intervalo.

Vamos a definir formalmente que quiere decir que una variable X tenga distribución uniforme en $[a, b]$, con $a < b$:

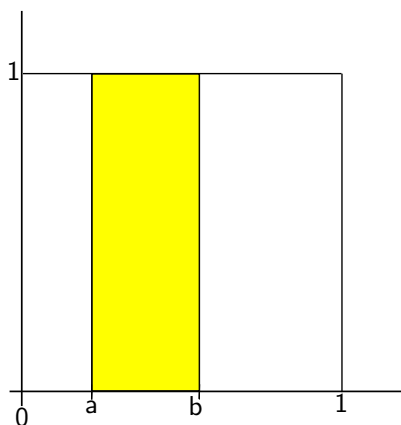


Figura 1.2: Grafico de la densidad de una variable con distribución uniforme en $[0, 1]$

Definición 1.31. Una variable aleatoria tiene **distribución uniforme** en $[a, b]$ (que denotaremos $X \sim U([a, b])$) si para todo $a \leq x \leq y \leq b$ tenemos que

$$P(X \in [x, y]) = \frac{y - x}{b - a}.$$

Observemos que en este caso $P(X \in [x, y])$ coincide con área encerrada por el gráfico de la función constante $f = 1/(b - a)$ y el eje x . Nuevamente la probabilidad de que la variable pertenezca a un determinado intervalo depende *únicamente* de la longitud del intervalo y como caso particular, cuando $a = 0$ y $b = 1$, tenemos la distribución uniforme en $[0, 1]$.

Ejercicio 1.32. Hemos definido variable aleatoria uniforme en un intervalo fijo $[a, b]$ ¿Se puede definir una variable aleatoria con distribución uniforme en \mathbb{R} ?

1.4.2. Densidad asociada a una variable aleatoria

Una función $f \geq 0$ que tiene la propiedad de que $P(X \in [a, b])$ es igual al área que encierra la gráfica de f y el eje de las x , en el intervalo $[a, b]$ se llama *función de densidad*. A veces, en lugar de definir cual es la variable, diremos cual es la densidad. Observemos que, para las variables que toman una cantidad finita (como la binomial o la hipergeométrica) o numerable de valores (como la geométrica y la Poisson) *no* se puede definir una densidad. Veamos (de manera intuitiva) por qué es esto en el caso de la binomial. Si tenemos $X \sim \text{Bin}(4, 1/3)$ sabemos que $P(X = 3) = 4 \times (1/3)^3 \times 2/3$. Si existiese una tal f tendría que tener la propiedad de que $4 \times (1/3)^3 \times 2/3 = P(X \in [3 - 1/n, 3 + 1/n])$ es igual al área de la función f en el intervalo $[3 - 1/n, 3 + 1/n]$ y esto tiene que ser cierto para todo n . Por otro lado observemos que la longitud del intervalo $[3 - 1/n, 3 + 1/n] = 2/n$ que tiende a 0 cuando n tiende a infinito, por lo tanto para que la función tenga área fija e igual a $4 \times (1/3)^3 \times 2/3$ tendría que tomar valores arbitrariamente grandes en dicho intervalo, que conduce a que en 3 tenga que valer infinito.

Una propiedad importante de las densidad es que el área total que queda bajo el gráfico de la misma es 1 (¿por qué?), y una consecuencia inmediata es que si $f(x)$ es una densidad, $f(x)$ tiende a 0 cuando x tiende a $+\infty$ y a $-\infty$.

1.4.3. Distribución Normal

Veremos ahora una de las distribuciones más importantes del curso (y probablemente de la probabilidad toda), la **distribución normal** (o distribución gaussiana). No vamos a ahondar aquí sobre la importancia de dicha distribución, respecto de lo cual se ha escrito mucho, en diversas áreas del conocimiento, desde la sociología hasta la matemática, pasando por la economía. Su origen data de 1733 y hasta la actualidad se siguen descubriendo propiedades y definiendo generalizaciones de la misma. Vamos a empezar con el caso más simple, la distribución normal de parámetros 0 y 1.

Definición 1.33. Decimos que una variable aleatoria X tiene distribución normal de parámetros 0 y 1 (que se denota $X \sim N(0, 1)$) si su función de densidad es

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (1.17)$$

El gráfico de la función f se muestra en la Figura (1.3).

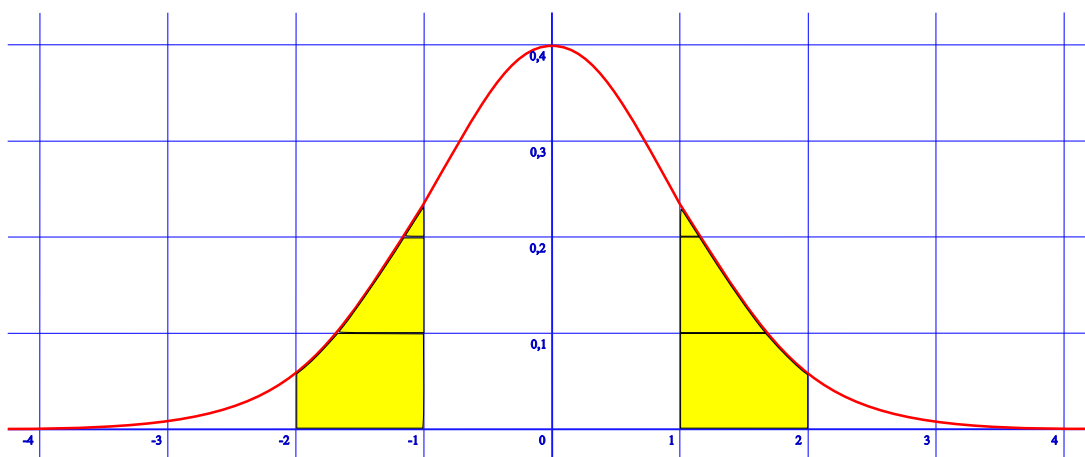


Figura 1.3: Gráfico de la densidad de una variable con distribución Normal(0,1), observese que las áreas en amarillo son iguales.

Veamos algunas propiedades que se deducen de la *forma* de la f . Como se ve en (1.17) la función es simétrica respecto de 0 es decir $f(x) = f(-x)$ (esto se conoce como función par). Esto hace que si $0 < a < b$ entonces $P(X \in [a, b]) = P(X \in [-b, -a])$, además, también por la propiedad de simetría respecto de 0 se obtiene que

$$P(X > a) = P(X < -a), \quad (1.18)$$

en particular $P(X < 0) = P(X > 0) = 1/2$. En general la propiedad (1.18) va a ser cierta para cualquier variable cuya densidad sea par. Por otro lado es claro que el máximo de la función se da en 0 (y es $1/\sqrt{2\pi}$ como se sigue de (1.18)). De donde se deduce que para todo $-a < 0 < a$, $P(X \in [-a, a]) > P(X \in [l, s])$ para *cualquier* intervalo $[l, s]$ que tenga longitud $2a$ (ya que el área va a ser más pequeña). Si bien el área que encierra la gráfica de f con el eje x no se puede calcular de manera simple a partir de (1.17), (como hacíamos con la función de densidad de una variable con distribución uniforme) existen tablas que dan para ciertos valores de $t > 0$ el valor de $P(X < t)$. En general se denota $\Phi(t) = P(X \leq t)$ con $X \sim N(0, 1)$.

Observación 1.34. *Observemos que si $a < b$, $\Phi(b) - \Phi(a) = P(X \in [a, b])$.*

Ejercicio 1.35. Deducir, a partir de (1.18) que

$$\Phi(t) = 1 - \Phi(-t)$$

Esta propiedad es *muy* importante a la hora de usar una tabla de la distribución normal ya que como dijimos antes, dichas tablas solo dan los valores de $\Phi(t)$ para algunos $t > 0$. Dado que la función definida en (1.33) tiende a 0 muy rápido cuando x tiende a $+\infty$ o a $-\infty$ es que en general las tablas no dan el valor $\Phi(t)$ para valores mayores a 4 ($\Phi(4) = 0,999$).

El lector ya sospechará que el 0 en la definición de $N(0, 1)$ tiene que ver con el hecho de que sea simétrica respecto de 0, efectivamente esto es así. Esto está relacionado con un fenómeno relativo al comportamiento que tienen los números que se *sortean* con distribución $N(0, 1)$. En una muestra de muchos datos que siguen dicha distribución, y bajo el supuesto de que los datos son independientes unos de otros, vamos a obtener aproximadamente la misma cantidad de números positivos que negativos (no solo la misma cantidad sino también iguales en magnitud, pero de signo opuesto). Esto hace que si los promediamos de un valor próximo a 0. Por ahora no vamos a justificar mucho más ese parámetro más adelante veremos que es lo que se denomina *valor esperado*. La relación del 1 en $N(0, 1)$ con el comportamiento de los datos es un poco más difícil de explicar a partir de (1.17), tiene que ver con qué tanto se alejan de 0 esos números sorteados con distribución normal. Si en lugar de 1 ponemos 1 millón, vamos a ver que los números se hacen mucho más grandes en magnitud. En virtud de esto es que vamos a definir:

Definición 1.36. Una variable aleatoria X tiene distribución $N(\mu, \sigma^2)$ con $\sigma > 0$ si su densidad es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1.19)$$

La gráfica de esta función, para $\sigma = 1$ es idéntica a la que se muestra en la Figura (1.3) salvo por el hecho de que el punto de simetría es μ . Esto hace que promediar *muchos* datos independientes, con distribución $N(\mu, \sigma^2)$ da un valor próximo a μ . Se puede demostrar, y será lo que se usará para calcular probabilidades de una variable con distribución $N(\mu, \sigma^2)$ el siguiente resultado:

Teorema 1.37. *Si X tiene distribución $N(\mu, \sigma^2)$ entonces*

$$\frac{X - \mu}{\sigma} \sim N(0, 1) \quad (1.20)$$

De éste último se sigue que, si $X \sim N(\mu, \sigma^2)$ entonces

$$\Phi(t) = P\left(\frac{X - \mu}{\sigma} \leq t\right) = P(X \leq \sigma t + \mu),$$

Si tomamos ahora $t = \frac{1}{\sigma}(u - \mu)$ tenemos que

$$\Phi\left(\frac{u - \mu}{\sigma}\right) = P(X \leq u)$$

Ejercicio 1.38. Sea $X \sim N(1, 4)$. Calcular $P(X < 2)$. Usando la igualdad anterior con $u = 1$ tenemos que $P(X < 2) = \Phi\left(\frac{2-1}{2}\right) = \Phi(1/2)$ y, usando la tabla $\Phi(1/2) \approx 0,691$.

1.4.4. Distribución Exponencial

Vamos a definir ahora una nueva variable aleatoria a partir de su densidad: la variable aleatoria con distribución exponencial de parámetro λ . Dicha variable se emplea típicamente para modelar el tiempo entre la llegada consecutiva de dos personas a una fila (y por lo tanto es una variable que toma únicamente valores positivos). Tiene algunas propiedades que la hacen importante en la teoría de probabilidad, una de ellas, quizás la mas conocida es la de *perdida de memoria* que, a grandes rasgos dice que, haber esperado un tiempo h para que suceda un cierto fenómeno no nos aporta información de si este sucederá o no en tiempo $t + h$. Dicho de otra manera, el fenómeno *se olvida* de lo que pasó hasta h . En términos de probabilidad condicional esto se escribe diciendo que

$$P(X > t + h | X > h) = P(X > t). \tag{1.21}$$

Veamos cual es la definición formal

Definición 1.39. Decimos que una variable aleatoria X tiene **distribución exponencial** de parámetro $\lambda > 0$ (que denostarmos $X \sim Exp(\lambda)$) si su densidad es

$$f(x) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{en otro caso} \end{cases} \tag{1.22}$$

El gráfico de dicha densidad, para 3 valores de λ , se muestra en la Figura (1.4).

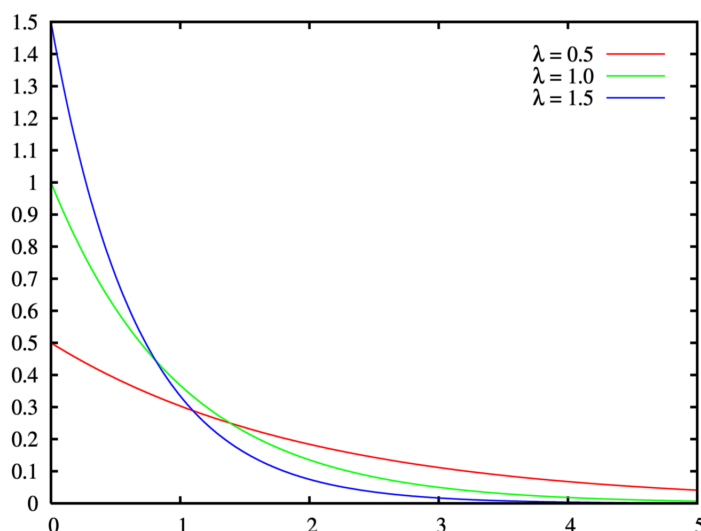


Figura 1.4: Gráficos de las densidades de la variable exponencial, para diferentes valores de λ

Observemos que $f(0) = \lambda$. El parámetro λ , al igual que el μ de la distribución normal, tiene que ver con el promedio de *muchos* datos independientes, que tienen distribución exponencial. Se puede demostrar, aunque no lo haremos ahora que el promedio de esos datos está próximo a $1/\lambda$. Es decir cuanto mas grande es λ más chico es el promedio. Esto coincide con el hecho, observable en la Figura (1.4) de que al aumentar λ la gráfica toma valores más grandes cerca de 0 y valores más chicos hacia infinito.

Observación 1.40. Se puede ver, (con un poco de cálculo integral) que si $X \sim Exp(\lambda)$ entonces

$$P(X < t) = \begin{cases} 1 - e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases} \tag{1.23}$$

Teorema 1.41. *Vamos a demostrar que se cumple (1.21). Observemos que de (1.23) tenemos que, para todo $h \geq 0$*

$$P(X > h) = e^{-\lambda h}.$$

Por definición

$$P(X > t + h | X > h) = \frac{P(X > t + h \cap X > h)}{P(X > h)},$$

Observemos que $P(X > t + h \cap X > h) = P(X > t + h)$ ya que si pasó más de $t + h$, seguro que paso más de h (en términos de sucesos lo que tenemos es que el suceso $\{X > t + h\} \subset \{X > h\}$ por lo tanto $\{X > t + h\} \cap \{X > h\} = \{X > t + h\}$). Si usamos eso tenemos que

$$P(X > t + h | X > h) = \frac{P(X > t + h)}{P(X > h)} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda h}} = \frac{e^{-\lambda t} e^{-\lambda h}}{e^{-\lambda h}} = e^{-\lambda t} = P(X > t)$$

Ejercicio 1.42. Con los mismos argumentos que usamos en la demostración anterior se puede ver que la distribución geométrica *también* tiene la propiedad de perdida de memoria (1.21).

1.4.5. Distribución T de Student

Definiremos ahora una variable aleatoria que aparecerá en el próximo capítulo. Diremos que T tiene **distribución T de Student con $k > 0$ grados de libertad** (que denotamos T_k) si su densidad es

$$f_{T_k}(x) = \frac{C(k)}{\left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}}$$

donde $C(k)$ es una constante (que depende de k) que hace que $f_T(x)$ sea una densidad, es decir que el área bajo la gráfica de f_T sea 1. Se puede dar una fórmula explícita para $C(k)$ que aquí no daremos ya que no haremos uso de ella.

Observemos que por la forma de f_{T_k} , dicha función, al igual que (1.17) es simétrica respecto de 0. Si bien, a diferencia de lo que sucedía con (1.17) existen fórmulas explícitas para $P(T_k < x)$ es decir el área de f_{T_k} hasta x . No obstante lo que se hace en general es usar una tabla que da, para algunos valores de t y de k , los valores de $P(T_k < t)$. Se puede demostrar que para valores de k grandes, $P(T_k < x) \approx P(N(0, 1) < x)$ para todo x más aun

$$P(T_k < x) \rightarrow P(N(0, 1) < x) \text{ si } k \text{ tiende a infinito.}$$

La Figura (1.5) muestra, para diferentes valores de k el gráfico de f_{T_k} .

Una propiedad que nos será de utilidad más adelante, que se deduce de que T_k es simétrica respecto de 0 es que, si llamamos $F_{T_k}(x)$ a la función que (al igual que Φ para la normal) nos da el área hasta x , es decir si $P(T_k \leq x) = F_{T_k}(x)$ entonces, para todo x

$$F_{T_k}(x) = 1 - F_{T_k}(-x) \tag{1.24}$$

1.4.6. Distribución **Chi-cuadrado**: χ_k^2

Finalmente, otra distribución que aparecerá, pero que no estudiaremos en profundidad, es la **distribución Chi-cuadrado con k grados de libertad**, que se denota usualmente χ_k^2 . Diremos que una variable X tiene distribución χ_k^2 si su densidad es

$$f_X(x) = C(k)x^{k/2-1}e^{-x/2} \quad \text{si } x > 0 \tag{1.25}$$

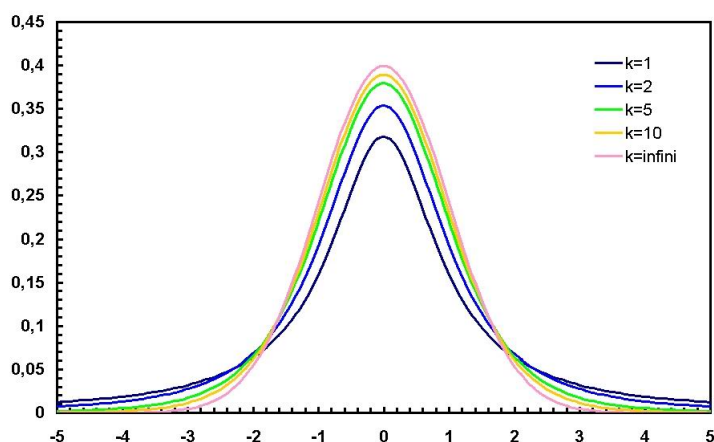


Figura 1.5: Gráficos de las densidades de una variable con distribución T_k para diferentes valores de k

donde $C(k)$ es una constante (que depende de k) que hace que el área bajo la gráfica de f_X sea uno. (No es la misma constante que para la T de student pero también existen fórmulas explícitas para $C(k)$) Observemos que la densidad f_X está definida únicamente para valores de x positivos (al igual que lo que sucedía con la distribución exponencial), es decir si una variable tiene distribución χ_k^2 esta sólo toma valores positivos. La Figura (1.6) muestra el gráfico de f_X para diferentes valores de k .

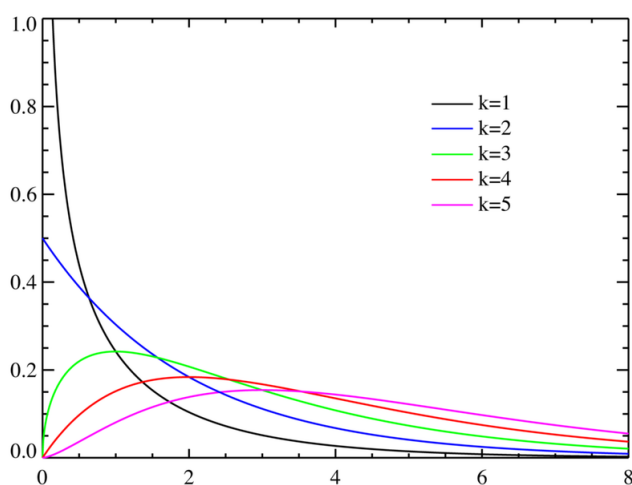


Figura 1.6: Gráficos de las densidades de una variable con distribución χ_k^2 para diferentes valores de k

1.5. Esperanza y Varianza de una variable aleatoria

En esta sección vamos a definir dos valores asociados a una variable aleatoria que serán de importancia en el resto del curso. No daremos las definiciones formales sino simplemente la idea intuitiva de qué son, para qué se usan, y cómo se calculan para las variables aleatorias que hemos definido. Vamos a comenzar con el concepto de esperanza de una variable aleatoria. Al comienzo del apartado sobre probabilidad dijimos que, intuitivamente, la

probabilidad de que se de un determinado resultado en un experimento, está asociado a lo que sucede cuando repetimos ese experimento muchas veces. Si tiramos una moneda mil veces es intuitivo que *aproximadamente* la mitad de las veces va a salir cara, y la otra mitad va a salir número, si representamos cada una de estas ocurrencias con un 1 cuando sale cara y 0 cuando sale número y promediamos la secuencia de ceros y unos que obtenemos vamos a tener que sumar cerca de 500 veces 0 y cerca de 500 veces uno, y dividir sobre el total de tiradas, es decir el promedio es *aproximadamente* $500/1000 = 1/2$. Consideremos la variable X que toma los valores 0 y 1 con probabilidad $1/2$, observemos que el promedio que calculamos es $0 \times P(X = 0) + 1 \times P(X = 1) = 1/2$. Esto se expresa diciendo que el *valor esperado* de X es $1/2$. El nombre valor esperado puede dar lugar a confusión, *no* quiere decir que si arrojamus una moneda (cuyos resultados son únicamente 0 (cara) y 1 (número) vamos a obtener $1/2$ (que en términos del experimento no representa nada) sino que el promedio se está proximo a $1/2$.

Veamos otro ejemplo: la probabilidad de que al tirar un dado salga 5 es $1/6$ porque intuitivamente, si tiramos un dado, digamos mil veces para fijar ideas, $1/6$ de las mismas va a salir 5 ($1000/6 \approx 166$ de veces). Veamos que pasa si cada vez que obtenemos un resultado, lo anotamos, y luego promediamos los mil resultados que obtuvimos, pudimos haber obtenido: 1, 3, 3, 3, 3, 4, 5, 3, 5, 6, 6, 6, 4, 2, 1, 1, 4, 1, 2, 1, 1.... Como dijimos, el 5 sale aproximadamente $1/6$ de las veces, es decir tenemos que sumar 166 veces 5. Es razonable suponer que lo mismo va a pasar con el 1, con el 2, con el 3 con el 4 y con el 6. Es decir, vamos a tener que sumar 166 veces 1, 166 veces 2, 166 veces 3, 166 veces 4 y 166 veces 6. Por lo tanto el promedio que queremos hacer va a dar aproximadamente:

$$\frac{1 \times 166 + 2 \times 166 + 3 \times 166 + 4 \times 166 + 5 \times 166 + 6 \times 166}{1000} \approx 3,5,$$

que es igual a:

$$1 \times \frac{166}{1000} + 2 \times \frac{166}{1000} + 3 \times \frac{166}{1000} + 4 \times \frac{166}{1000} + 5 \times \frac{166}{1000} + 6 \times \frac{166}{1000} \quad (1.26)$$

a su vez, como $\frac{166}{1000} \approx 1/6$, (1.26) es una *aproximación* de

$$1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}.$$

Vamos a definir la variable aleatoria X que sea i si salió i al tirar el dado, es claro que X toma valores 1,2,3,4,5,6 con probabilidad $1/6$ cada uno. Lo que escribimos antes no es otra cosa que

$$1 \times P(X = 1) + 2 \times P(X = 2) + \dots + 6 \times P(X = 6). \quad (1.27)$$

Observemos que en (1.27) desapareció completamente el número mil, que era el numero de veces que tirábamos el dado. Es decir el numero que se calcula en (1.27) es una propiedad de la variable X , que, como vimos, esta relacionada con lo que pasa al promediar los resultados de experimentos *independientes* que siguen la distribución de X . Dicho número se llama valor esperado de X . Es claro que no todas las variables toman los valores 1 a 6 (una variable con distribución geométrica toma infinitos valores) ni los toman con la misma probabilidad. No obstante, el valor esperado se calcula de forma parecida a (1.27), de tal manera que represente el número al que se aproxima el promedio de los resultados de repetir el experimento de forma independiente. Esto se debe a que en el promedio de los resultados, cada número aparece tantas veces como la probabilidad de que aparezca, multiplicado por la cantidad de experimentos que se hacen. La tabla (1.5) tiene los valores esperados de las variables aleatorias que hemos presentado hasta ahora.

Distribución	Valor esperado
$Bin(n, p)$	np
$Geo(p)$	$1/p$
$Hipgeo(d, n, N)$	$\frac{nd}{N}$
$Poisson(\lambda)$	λ
$U[a, b]$	$(a + b)/2$
$N(\mu, \sigma^2)$	μ
$Exp(\lambda)$	$1/\lambda$
T_k	0
χ_k^2	k

Cuadro 1.3: Esperanza de algunas variables

1.5.1. Esperanza de $X \sim Bin(n, p)$

Veamos ahora cómo calcular la esperanza de una variable con distribución binomial. Para fijar ideas tomemos $n = 4$ y $p = 1/6$ y denotemos $X \sim Bin(4, 1/6)$. Recordemos que la distribución binomial cuenta la cantidad de éxitos que se tienen en la repetición n veces de un experimento cuya probabilidad de éxito es p . Si repetimos muchas veces la secuencia de 4 experimentos, el valor esperado será próximo a la cantidad de éxitos promedio que obtenemos. Al igual que como hicimos con los dados, el valor esperado de X que se denota $E(X)$ es

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) + 2 \times P(X = 2) + 3 \times P(X = 3) + 4 \times P(X = 4),$$

Usando (1.21) con $n = 4$ y $p = 1/6$ tenemos que

$$E(X) = \binom{4}{1} \times \frac{1}{6} \times \left(\frac{5}{6}\right)^3 + \binom{4}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2 + \binom{4}{3} \times \left(\frac{1}{6}\right)^3 \times \frac{5}{6} + \binom{4}{4} \times \left(\frac{1}{6}\right)^4 = 4/6$$

Observemos que nos dio exactamente np . En general si $X \sim Bin(n, p)$ se tiene que $E(X) = np$, por ejemplo si $p = 1$ el número *promedio* esperado de éxitos es n , lo cual es muy razonable dado que $p = P(\text{éxito})$. Si $p = 0$ da 0 que también es muy intuitivo.

No vamos a hacer la cuenta para calcular $E(X)$ para la distribución geométrica, que implica sumar los infinitos valores:

$$E(X) = 1 \times P(X = 1) + 2 \times P(X = 2) + 3 \times P(X = 3) + 4 \times P(X = 4) + \dots$$

Se puede demostrar que dicho valor es $1/p$. Aquí también, valores grandes de p hacen que $E(X)$ sea chico, esto quiere decir que el tiempo que, *en promedio* tenemos que esperar hasta que ocurra el primer éxito se hace menor a medida que hacemos más probable que este ocurra. Muy razonable, ¿no?. Por otro lado $1/p$ tiende a infinito si p tiende a 0. Lo cual significa que si es poco probable que se de el éxito, vamos a tener que esperar *en promedio* mucho hasta que se de por primera vez.

Definición 1.43. Vamos ahora a definir cómo se calcula la esperanza de una variable aleatoria que toma valores $x_1, x_2, \dots, x_k, \dots$ (una cantidad finita o infinita). Supongamos que $P(X = x_i) = p_i > 0$ y que $p_1 + p_2 + p_3 + \dots = 1$ es decir esos son todos los valores que toma. Entonces:

$$E(X) = x_1p_1 + x_2p_2 + x_3p_3 + \dots + x_kp_k + \dots,$$

Siempre que la suma (o serie en el caso de infinitos sumandos) anterior sea finita. Como pasaba con el dado, cuyo valor esperado era 3,5, el valor esperado de una variable no necesariamente es uno de los valores que toma. Tampoco es cierto que en general tenga que ser positivo ya que los x_i pueden ser negativos. No obstante, se puede demostrar que esta entre el mínimo y el máximo de los valores que toma X . Esto último es muy intuitivo si lo pensamos en términos de promediar las veces que salieron los x_i . El promedio nunca va a dar más de lo que da si siempre sumamos el mayor de los datos ni menos que lo que daría si sumamos el más chico.

En el caso de variables que tienen densidad, las formulas anteriores no se aplican ya que la variable, como vimos, toma cualquier número con probabilidad 0 y además toma una cantidad *no numerable* de valores. En estos casos se usan otros métodos para calcular la esperanza, que aquí no explicaremos.

1.5.2. Varianza

Supongamos que tenemos mil datos independientes, X_1, \dots, X_{1000} que corresponden a observaciones de una variable aleatoria X que sabemos que tiene distribución normal con esperanza 10 y varianza σ^2 desconocida. Por lo que dijimos antes, si promediamos dichos datos nos da próximo a 10 ya que la esperanza de $X \sim N(10, \sigma^2)$ es 10. Pensemos que esos datos son resultados de mediciones de algo cuyo valor real sabemos que tiene que ser 10, con un aparato que tiene un cierto error al medir. Lo que queremos ahora es tener una idea de qué tanto se equivoca al medir el aparato. Es decir, que tan alejados de 10 son los valores que medimos. La varianza σ^2 nos da una idea de la *dispersión* de los datos entorno al valor esperado. Cuanto más grande es la varianza, más dispersos son los datos, y, en nuestro ejemplo, más grande es el error que comete el aparato. En el caso extremo en que $\sigma^2 = 0$ es decir la varianza es nula, los datos van a ser todos iguales a 10, en cuyo caso el error es 0. La definición de la varianza se hace por medio de la esperanza de una nueva variable. Lo que hacemos es, dada la variable X restarle la esperanza, es decir tomar $X - 10$ (con lo cual los datos que antes tenían esperanza $E(X)$ ahora, *centrados*, tienen esperanza 0) y luego promediar las nuevas observaciones $X_1 - 10, X_2 - 10, \dots, X_{1000} - 10$ pero de modo tal que los datos que son muy grandes *influyan más* en el promedio que los datos pequeños. Es decir, en lugar de promediar $X_1 - 10, \dots, X_{1000} - 10$ que sabemos que nos va a dar próximo a 0, promediamos $(X_1 - 10)^2, (X_2 - 10)^2, \dots, (X_{1000} - 10)^2$. Observemos que la función x^2 hace justamente lo que decíamos antes. Si una observación, supongamos la X_4 para fijar ideas, es tal que $X_4 - 10$ es muy grande (que quiere decir que ese dato se alejó mucho de la esperanza), al elevar al cuadrado da un número aún mayor. Mientras que si $X_4 - 10 < 1$, entonces $(X_4 - 10)^2 < (X_4 - 10)$. La definición formal es la siguiente:

Definición 1.44. Sea X una variable aleatoria tal que $E(X^2) < \infty$ entonces

$$Var(X) = E((X - E(X))^2)$$

Veamos como se calcula la varianza a partir de la fórmula que dimos en (1.43), para variables que toman una cantidad finita o numerable de valores:

Observación 1.45. *Vamos ahora a calcular la varianza de una variable aleatoria que toma valores $x_1, x_2, \dots, x_k, \dots$ (una cantidad finita o infinita). Supongamos que $P(X = x_i) = p_i > 0$ y que $p_1 + p_2 + p_3 + \dots = 1$ es decir esos son todos los valores que toma. Sabemos, por (1.43) que*

$$E(X) = x_1p_1 + x_2p_2 + x_3p_3 + \dots + x_kp_k + \dots,$$

De la formula anterior tenemos que para calcular $E((X - E(X))^2)$ tenemos que ver que valores toma $(X - E(X))^2$ y con qué probabilidades, y luego usar dicha fórmula. Observemos que $(X - E(X))^2$ toma los valores $(x_1 - E(X))^2$ con probabilidad p_1 , $(x_2 - E(X))^2$ con probabilidad p_2 , etc, $(x_k - E(X))^2$ con probabilidad p_k , ... por lo tanto, usando la formula anterior

$$E((X - E(X))^2) = (x_1 - E(X))^2 p_1 + (x_2 - E(X))^2 p_2 + (x_3 - E(X))^2 p_3 + \dots + (x_k - E(X))^2 p_k + \dots$$

La observación anterior nos da la forma de calcular, por ejemplo, al varianza para la binomial, la geométrica, la hipergeométrica y la Poisson. Para las variables que tienen densidad no es posible usar dicha formula (no obstante la definición (1.44) sigue siendo cierta, pero se calculan de otra forma. La tabla (1.5.2) contiene los resultados de las varianzas de las variables que hemos visto hasta ahora.

Distribución	Varianza
$Bin(n, p)$	$np(1 - p)$
$Geo(p)$	$(1 - p)/p^2$
$Hipgeo(d, n, N)$	$\frac{nd(N-d)(N-n)}{N^2(N-1)}$
$Poisson(\lambda)$	λ
$U[a, b]$	$(b - a)^2/12$
$N(\mu, \sigma^2)$	σ^2
$Exp(\lambda)$	$1/\lambda^2$
T_k con $k > 2$	$k/(k - 2)$
χ_k^2	$2k$

Cuadro 1.4: Varianza de algunas variables

1.6. Independencia de Variables

En la sección sobre independencia de sucesos vimos qué quería decir que dos sucesos fuesen independientes. En esta sección definiremos qué quiere decir que dos variables X e Y sean independientes. La idea intuitiva de fondo es la misma que para sucesos, lo que decimos es que saber el resultado de una variable no nos aporta información para deducir el resultado que tendrá la otra. Por ejemplo, si X es una variable que dada una persona nos da la altura (es decir definida en $\Omega = \{\text{conjunto de todas las personas}\}$, e Y es una variable (definida en el mismo Ω que dada la persona nos da el color de ojos (supongamos que los hemos numerado 1= verdes, 2 azul, etc) es razonable pensar que dichas variables son independientes en el sentido de que saber el color de ojos de una persona nada nos dice de la altura, y que saber su altura nada nos dice del color de ojos que tendrá. En términos matemáticos estamos diciendo, por ejemplo

$$P(1,60 < X < 1,75 | Y = 1) = P(1,60 < X < 1,75)$$

O, como vimos, (dado que $P(Y = 1) > 0$), considerando los sucesos $\{1,60 < X < 1,75\}$ y $\{Y = 1\}$ lo anterior es equivalente a decir que

$$P(\{1,60 < X < 1,75\} \cap \{Y = 1\}) = P(\{1,60 < X < 1,75\}) \times P(\{Y = 1\})$$

y esto, por definición, no es otra cosa que decir que los sucesos $\{1,60 < X < 1,75\}$ y $\{Y = 1\}$ son independientes. En virtud del ejemplo anterior, vamos a introducir la siguiente definición

Definición 1.46. Dos variables X e Y son independientes si, para todo par de intervalos $[a, b]$ y $[c, d]$ se cumple que

$$P(\{X \in [a, b]\} \cap \{Y \in [c, d]\}) = P(\{X \in [a, b]\}) \times P(\{Y \in [c, d]\}),$$

es decir los sucesos $\{X \in [a, b]\}$ y $\{Y \in [c, d]\}$ son independientes, *para todo* a, b, c, d .

La definición anterior se generaliza de manera intuitiva a n variables X_1, \dots, X_n . Muchas veces, cuando tenemos una muestra de n datos, vamos a hacer el supuesto de que son independientes. Es un supuesto *muy* fuerte pero sin el cual muchos razonamientos que haremos no son ciertos. En la siguiente sección vamos a ver algunas propiedades de las variables independientes.

1.6.1. Suma de variables aleatorias

Hasta ahora hemos hablado de sumar datos, promediarlos, etc. Vamos a introducir aquí un poco más de formalización respecto de lo que significan esos procedimientos. Supongamos que tiramos 3 veces un dado, y que éxito es que salga 5 en la cara superior. Contar la cantidad de éxitos nos lleva (o debería) a pensar en una variable binomial de parametros $n = 3$, $p = 1/6$. Sabemos que cada una de las 3 tiradas se hacen de forma independiente (porque es parte de los supuestos del experimento). Por otro lado, si representamos éxito con un 1, y fracaso con un 0, contar la cantidad de éxitos no es otra cosa que contar la cantidad de unos que que hay en las ternas de ceros y unos. Es decir, si ahora consideramos las tres variables: X_1 que vale 1 si salió éxito en el primer experimento y 0 en caso contrario, X_2 la variable que vale 1 si salió éxito en el segundo experimento y 0 en caso contrario X_3 la variable que vale 1 si salió éxito en el tercer experimento y 0 en caso contrario, es claro que dichas variables son independientes. Si llamamos X a la variable que cuenta la cantidad de éxitos, de la cual sabemos que $X \sim Bin(3, 1/6)$, tenemos que la cantidad de éxitos no es mas que $X_1 + X_2 + X_3$. Es decir

$$X = X_1 + X_2 + X_3.$$

Veamos algunos cálculos para convencernos de esto: $P(X = 3) = P(X_1 + X_2 + X_3 = 3)$ y esto se da únicamente si $X_1 = X_2 = X_3 = 1$. Usando la formula de $Bin(3, 1/6)$ tenemos que $P(X = 3) = (1/6)^3$.

Observemos que las variables X_1, X_2 y X_3 son independientes, en el sentido de la definición anterior ya que hicimos el supuesto de que los experimentos se realizaban de forma independiente. Por lo tanto

$$P(\{X_1 = 1\} \cap \{X_2 = 1\} \cap \{X_3 = 1\}) = P(\{X_1 = 1\}) \times P(\{X_2 = 1\}) \times P(\{X_3 = 1\}) = 1/6 \times 1/6 \times 1/6,$$

es decir llegamos al mismo resultado. Para calcular $P(X = 2)$ tenemos que considerar *todas* las formas de sumar 2 a partir de las variables X_1, X_2 y X_3 que nos conduce a 3 casos, $((1, 1, 0), (1, 0, 1)$ y $(0, 1, 1))$. Queda como ejercicio hacer la cuenta de que en ese caso también se cumple que $P(X = 2) = P(X_1 + X_2 + X_3 = 2)$. Lo que hicimos fue *sumar* variables, para obtener una nueva variable aleatoria (definida en el mismo Ω). Este procedimiento tiene interés en general. Asimismo a veces puede ser de interés considerar la variable aX que dados los resultados de X los multiplica por una constante a . Estas nuevas variables tienen algunas propiedades interesantes:

Teorema 1.47. ■ *Si X e Y son variables independientes entonces*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y),$$

Es importante aclarar que el resultado anterior no es cierto sin la hipótesis de independencia.

■ *Para cualquier variable X y cualquier número a ,*

$$E(aX) = aE(X) \quad \text{y} \quad \text{Var}(aX) = a^2\text{Var}(X).$$

■ *Para cualquier par de variables X, Y se tiene que*

$$E(X + Y) = E(X) + E(Y).$$

Capítulo 2

Intervalos de confianza, Pruebas de hipótesis.

En este capítulo veremos algunas aplicaciones de los conceptos introducidos en el capítulo anterior. En general se cuenta con datos X_1, \dots, X_n que resultan de realizar n veces un cierto experimento, y se quiere deducir propiedades de la distribución de los mismos: su valor esperado, su varianza, etc. Por ejemplo, si suponemos que los n datos son alturas de cierto grupo de personas, y si suponemos además que los mismos se distribuyen uniformemente en algún intervalo $[a, b]$ que desconocemos, nos puede interesar aproximar a o b o el valor esperado $(a + b)/2$. En otro caso, los datos pueden provenir de mediciones de un fenómeno físico que suponemos sigue una distribución normal, cuya esperanza y varianza desconocemos (llamemos μ y σ^2 respectivamente), y nos interesa estimar dichos parámetros. Por otro lado, tener una aproximación del valor esperado no es de mucha utilidad si la varianza es muy grande ya que nuevos datos nos van a conducir a una aproximación del valor esperado que puede diferir mucho de la que habíamos obtenido con los primeros datos. Por lo tanto es necesario conocer, *con cierta certeza o confianza* un intervalo donde es muy probable que esté el valor esperado. Es decir, no solo nos interesa aproximar μ , por medio de un *estimador* $\hat{\mu}_n$ sino también afirmar que, con cierta probabilidad, μ estará por ejemplo en el intervalo $[3, 4]$, dicho en términos matemáticos, esto no es otra cosa que pedir que:

$$P(3 < \mu < 4) = 1 - \alpha$$

donde α es un número fijado de antemano, que se denomina *nivel*. Lo que sucederá es que si la varianza de la variable es muy grande, el intervalo va a ser grande.

En otro tipo de aplicaciones queremos saber si es razonable afirmar, *con cierta confianza o certeza* que la esperanza *no* supera un valor crítico μ_0 . Pensemos por ejemplo que las variables X_1, \dots, X_n provienen de mediciones del nivel en sangre de colesterol (medido en mg/dl) en una población de n personas y que el valor crítico, es decir lo que se considera alto, es $130mg/dl$. Es claro que es mucho más grave equivocarse afirmando que el nivel de colesterol de la población es normal cuando en realidad supera 130 (y es más grave aún si la patología que estamos estudiando requiere de un tratamiento inmediato), que decir que el nivel es alto cuando en realidad es normal. Lo que queremos es que la probabilidad, *si la muestra es tal que $\mu > 130$* de afirmar que $\mu \leq 130$ sea un valor chico, fijado de antemano (que en general denotaremos con la lera griega α). Vamos a testear dos hipótesis la primera, $H_0 : \mu > 130mg/dl$, (que se denomina hipótesis nula), contra la alternativa $H_1 : \mu \leq 130$. ¿Qué sucede cuando tenemos más datos?. Es razonable pensar que al tener más datos podemos aproximar mejor μ a partir de $\hat{\mu}_n$, construir intervalos de confianza más chicos que nos permitan determinar mejor donde estará μ , afirmar con mayor certeza que μ

supera el valor crítico, o que no lo supera, etc. Este comportamiento, que intuitivamente queremos que se verifique es cierto en general para todo lo que veremos en este capítulo no obstante no siempre daremos demostraciones de que es así dado que algunas requieren de herramientas matemáticas sofisticadas.

Finalmente, y no menos importante, ¿qué hipótesis haremos sobre los datos que tenemos?, por ahora simplemente diremos que, a lo largo de todo el capítulo supondremos que los datos X_1, \dots, X_n son independientes, y que todos siguen una determinada distribución, de la cual en general supondremos que desconocemos ciertos parámetros. Por ejemplo, sabemos que es normal con cierta varianza pero desconocemos μ , o sabemos que es normal y desconocemos tanto μ como σ^2 . En otros ejemplos es razonable suponer que los datos tienen distribución exponencial pero desconocemos λ , etc.

2.1. Estimación de $E(X)$ y $Var(X)$

Como dijimos antes, dados n datos X_1, \dots, X_n independientes e idénticamente distribuidos según una cierta variable X , si queremos estimar $E(X)$ es razonable considerar como *estimador* el promedio \overline{X}_n , dado por

$$\overline{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n},$$

en general se demuestra, bajo ciertas hipótesis, que \overline{X}_n aproxima bien a $E(X)$, es decir

$$\overline{X}_n \rightarrow E(X) \quad \text{cuando } n \text{ tiende a infinito,}$$

El resultado anterior se conoce como ley de los grandes números. Dado que $Var(X)$ lo definimos a partir de la esperanza de una nueva variable, es decir $Var(X) = E((X - E(X))^2)$ tenemos que, si conociéramos $E(X)$ un estimador para $Var(X)$ es

$$\widehat{Var}(X) = \frac{(X_1 - E(X))^2 + (X_2 - E(X))^2 + \dots + (X_n - E(X))^2}{n},$$

en caso de no conocerlo podemos sustituir $E(X)$ por el estimador \overline{X}_n obteniendo así el estimador de $Var(X)$:

$$\widetilde{Var}(X) = \frac{(X_1 - \overline{X}_n)^2 + (X_2 - \overline{X}_n)^2 + \dots + (X_n - \overline{X}_n)^2}{n},$$

en este caso también es posible demostrar, bajo ciertas hipótesis que

$$\widetilde{Var}(X) \rightarrow Var(X) \quad \text{cuando } n \text{ tiende a infinito.}$$

En general, en lugar de estimar $Var(X)$ a partir de $\widetilde{Var}(X)$ usaremos el estimador de $\sqrt{Var(X)}$

$$S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2}$$

es decir $S_n^2 \rightarrow Var(X)$. Observemos que la relación entre $\widetilde{Var}(X)$ y S_n^2 es:

$$\widetilde{Var}(X) = \frac{n-1}{n} S_n^2.$$

Observación 2.1. *Se puede demostrar, y es útil a veces para hacer cuentas, que*

$$S_n^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - n(\overline{X}_n)^2 \right]$$

Vamos a enunciar sin demostrar algunas propiedades de S_n que serán de utilidad para calcular intervalos de confianza.

Teorema 2.2. *Sean X_1, \dots, X_n independientes e idénticamente distribuidas con distribución $N(\mu, \sigma^2)$, entonces*

$$1) \overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

2) \overline{X}_n y S_n^2 son independientes.

$$3) \frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2.$$

$$4) \sqrt{n} \frac{(\overline{X}_n - \mu)}{S_n} \sim T_{n-1}.$$

2.2. Intervalos de confianza

En esta sección vamos a ver como se construyen intervalos de confianza para el valor esperado (que denotaremos μ) de una variable. Primero supondremos que los datos tienen distribución $N(\mu, \sigma^2)$ y que conocemos σ^2 y luego que es desconocido. Al final de la sección veremos como construir intervalos de confianza para el valor esperado cuando los datos no son normales. Y finalmente, el caso particular en que queremos calcular un intervalo de confianza para proporciones, es decir tenemos datos que son ceros o unos, correspondientes a los posibles resultados de un experimento cuyos resultados posibles son éxito o fracaso. Y queremos un intervalo I tal que, fijado un nivel α , $P(p \in I) \approx 1 - \alpha$ siendo $p = P(\text{éxito})$. Cuando los datos son normales (tanto cuando conocemos σ^2 como cuando no) es posible construir intervalos *exactos*, es decir tal que $P(\mu \in I) = 1 - \alpha$. En general, cuando no sabemos que distribución tienen los datos hay que hacer uso de un teorema, denominado teorema central del límite, que aquí no demostraremos, mediante el cual se obtiene un intervalo *aproximado* I_n (que depende de n), tal que $P(\mu \in I_n) \approx 1 - \alpha$. No obstante se puede demostrar que

$$P(\mu \in I_n) \rightarrow 1 - \alpha \quad \text{cuando } n \text{ tiende a infinito.}$$

2.2.1. Intervalos de confianza para datos normales, σ conocido.

Sea $X \sim N(\mu, \sigma^2)$ con σ^2 conocido. Buscamos un intervalo I de la forma $[\overline{X}_n - k, \overline{X}_n + k]$. Debemos hallar k tal que $P(\mu \in I) = 1 - \alpha$, entonces

$$1 - \alpha = P(\overline{X}_n - k \leq \mu \leq \overline{X}_n + k) = P(\mu - k \leq \overline{X}_n \leq \mu + k)$$

Observemos que, por la parte 1) del Teorema (2.2) sabemos que $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, si ahora restamos μ y dividimos entre σ/\sqrt{n} tenemos que

$$P(\mu - k \leq \overline{X}_n \leq \mu + k) = P\left(\frac{\mu - k - \mu}{\sigma/\sqrt{n}} \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu + k - \mu}{\sigma/\sqrt{n}}\right)$$

y de (1.20) tenemos que

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1),$$

con lo cual, de (1.34)

$$P\left(\frac{-k}{\sigma/\sqrt{n}} \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{k}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{k}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-k}{\sigma/\sqrt{n}}\right)$$

Y si ahora usamos que que $\Phi(-t) = 1 - \Phi(t)$ tenemos que lo anterior es

$$2\Phi\left(\frac{\sqrt{nk}}{\sigma}\right) - 1.$$

Por lo tanto obtuvimos que

$$1 - \alpha/2 = \Phi\left(\frac{\sqrt{nk}}{\sigma}\right) \text{ entonces, aplicando } \Phi^{-1} \frac{\sqrt{nk}}{\sigma} = \phi^{-1}(1 - \alpha/2),$$

y por lo tanto tomamos

$$k = \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \alpha/2).$$

Notación: Denotemos $Z_p = \Phi^{-1}(p)$ (en el ejemplo que sigue veremos como se calcula a partir de la tabla de Φ), con esta notación el intervalo de confianza del ejemplo anterior es

$$\left[\overline{X}_n - \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}; \overline{X}_n + \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}\right]. \quad (2.1)$$

Es importante destacar que el intervalo (2.1) es exacto, no es para valores de n grandes. Por otro lado, observemos que la longitud del mismo es $2\frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}$ que tiende a 0 cuando n tiende a infinito. Esto nos dice que a medida que tenemos más datos podemos determinar con más exactitud donde es más probable que esté μ .

Ejemplo 2.3. Veamos un ejemplo de como se calcula. . La siguiente muestra corresponde a los niveles de sodio en sangre de 10 pacientes de una clínica (medidos en milimoles por litro).

139, 13	136, 67	140, 25	140, 57	137, 71	142, 38	142, 56	139, 92	140, 65	140, 35
---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

Asumiendo que los datos provienen de una distribución normal con $\sigma = 2$, dar un intervalo de confianza 90 % para la media.

Observemos primero que como estamos hablando de intervalo de confianza 90 % entonces $\alpha = 0,1$ es decir $1 - \alpha/2 = 0,95$. Para calcular el intervalo definido en (2.1) tenemos que calcular \overline{X}_n y $Z_{0,95}$.

$$\overline{X}_n = \frac{1}{10}(139, 13+136, 67+140, 25+140, 57+137, 71+142, 38+142, 56+139, 92+140, 65+140, 35) \approx 140$$

$Z_{0,95}$ es por definición $\Phi^{-1}(0,95)$, es decir es el valor que hace que el área hasta $Z_{0,95}$ sea 0,95. Si vemos una tabla de la distribución normal, tenemos que $P(N(0, 1) < 1,64) = 0,9494$ y $P(N(0, 1) < 1,65) = 0,9505$ por lo tanto el $Z_{0,95}$ que queremos está entre 1,64 y 1,65, tomemos por 1,65, el intervalo es

$$\left[140 - \frac{2}{\sqrt{10}}1,65; 140 + \frac{2}{\sqrt{10}}1,65\right] = [138,956; 141,043]$$

2.2.2. Intervalos de confianza para datos normales, σ desconocido.

Veamos otro intervalo de confianza, para el caso en que desconocemos σ .

Se $X \sim N(\mu, \sigma^2)$ con σ^2 desconocido. Dado que desconocemos σ pero sabemos por lo dicho en la sección anterior que S_n es un estimador del mismo, lo que se hace es sustituir S_n por σ . Buscamos un intervalo I de la forma

$$[\bar{X}_n - kS_n; \bar{X}_n + kS_n].$$

Tenemos que hallar k tal que $P(\mu \in I) = 1 - \alpha$

$$P(\mu \in I) = P(\mu - kS_n \leq \bar{X}_n \leq \mu + kS_n) = P\left(-k \leq \frac{\bar{X}_n - \mu}{S_n} \leq k\right)$$

si multiplicamos todo por \sqrt{n} tenemos que

$$P(\mu \in I) = P\left(-k\sqrt{n} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq k\sqrt{n}\right)$$

Por el teorema (2.2) 4) sabemos que $\sqrt{n} \frac{\bar{X}_n - \mu}{S_n}$ tiene distribución T_{n-1} de Student con $n - 1$ grados de libertad. Denotemos $F_{T_{n-1}}(x)$ la función que nos da el área bajo el gráfico de la densidad de T_{n-1} hasta x (es la función que juega el papel de Φ) es decir

$$P(T_{n-1} \leq x) = F_{T_{n-1}}(x),$$

Observemos que

$$P(\mu \in I) = P\left(-k\sqrt{n} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq k\sqrt{n}\right) = F_{T_{n-1}}(k\sqrt{n}) - F_{T_{n-1}}(-k\sqrt{n}).$$

Si usamos la propiedad (1.24) de las variables con distribución T de Student tenemos que

$$\begin{aligned} P(\mu \in I) &= F_{T_{n-1}}(\sqrt{n}k) - F_{T_{n-1}}(-\sqrt{n}k) \\ &= 2F_{T_{n-1}}(\sqrt{n}k) - 1 = 1 - \alpha, \end{aligned}$$

de donde

$$k = \frac{F_{T_{n-1}}^{-1}(1 - \alpha/2)}{\sqrt{n}} = \frac{t_{1-\alpha/2}(n-1)}{\sqrt{n}},$$

donde usamos la notación $F_{T_{n-1}}^{-1}(p) = t_p(n-1)$ siendo $n-1$ son los grados de libertad. Por lo tanto el intervalo de confianza para μ al nivel $1 - \alpha$ es

$$I = \left[\bar{X}_n - \frac{S_n}{\sqrt{n}} t_{1-\alpha/2}(n-1); \bar{X}_n + \frac{S_n}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right].$$

La longitud de dicho intervalo es

$$2 \frac{S_n}{\sqrt{n}} t_{1-\alpha/2}(n-1)$$

que nuevamente tiende a 0 cuando n tiende a infinito (ya que $t_{1-\alpha/2}(n-1) \rightarrow Z_{1-\alpha/2}$).

Ejemplo 2.4. Veamos, para los datos del ejemplo anterior, y $\alpha = 0,1$ como queda el intervalo, si usamos la fórmula para S_n nos da que $S_n^2 = 3,36074$ y por lo tanto $S_n = 1,8332$. Nos resta calcular $t_{0,95}(9)$ para eso necesitamos una tabla análoga a la que usabamos para Φ , en la tabla de la distribución t de Student tenemos en general, en la primer columna los posibles grados de libertad, por lo tanto nos fijamos en la fila 9. Las diferentes columnas corresponden a los diferentes α . Observemos que si bien trabajamos con $\alpha = 0,1$, lo que queremos es el valor $t_{0,95}(9)$ es decir

$$P(T_9 < t_{0,95}(9)) = 0,95,$$

que, uso de tabla mediante, nos da $t_{0,95}(9) = 1,83$. El intervalo nos queda entonces

$$\left[140 - \frac{1,8332}{\sqrt{10}} 1,83; 140 + \frac{1,8332}{\sqrt{10}} 1,83 \right] = [138,94; 141,06]$$

Observemos que este intervalo es levemente más largo que el que obtuvimos antes. Lo cual es muy razonable si pensamos que en este caso teníamos menos información (desconocíamos σ).

2.2.3. Intervalo de confianza para datos cualquiera, σ conocido

Ahora los datos son X_1, \dots, X_n independientes, pero no supondremos que tienen distribución normal, no obstante, supondremos que conocemos σ . Queremos encontrar nuevamente, dado $\alpha > 0$ un intervalo I tal que $P(\mu \in I) = 1 - \alpha$, siendo $\mu = E(X)$. En este caso no sera posible encontrar un intervalo exacto sino que encontraremos un interval I_n tal que $P(I_n) \rightarrow 1 - \alpha$. Aquí simplemente diremos como queda el intervalo y haremos un ejemplo simple. La forma del intervalo es *exactamente la misma* que para datos normales con σ conocido. Es decir:

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}; \bar{X}_n + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \right]. \quad (2.2)$$

2.2.4. Intervalo de confianza para datos cualquiera, σ desconocido

Aquí la forma del intervalo es análoga a la anterior (nuevamente es un intervalo aproximado) pero sustituyendo σ por su estimador S_n , es decir

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}; \bar{X}_n + \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} \right]. \quad (2.3)$$

Veamos un ejemplo para el caso en que los datos tienen distribución exponencia.

Ejemplo 2.5. Se tiene una muestra de 1000 datos de una distribución exponencial de parámetro λ y se sabe que la suma de los datos es 492 y la suma de los cuadrados 467, calcular un intervalo de confianza al 95 % para λ .

Observemos primero que nos piden un intervalo para λ y no para el valor esperado de la exponencial, que como sabemos es $1/\lambda$. No obstante si tenemos $I = [a, b]$ con $0 < a < b$ un intervalo de confianza para el valor esperado, es claro que $J = [1/b, 1/a]$ es un intervalo de confianza para λ . Veamos como calcular entonces un intervalo para el valor esperado. Observemos que estamos en el caso de datos con distribución cualquiera, y σ desconocido. Buscamos un intervalo de la forma $I = \left[\bar{X}_n - \frac{S_n}{\sqrt{n}} Z_{(1-0,05/2)}, \bar{X}_n + \frac{S_n}{\sqrt{n}} Z_{(1-0,05/2)} \right]$, si usamos

la Observación (2.1) tenemos que $S_n^2 = 1/(999)(467 - 1000(0,492)^2) = 0,22516$, (y $S_n = 0,4745$) por otro lado, $\alpha = 0,05$ es decir $1 - \alpha/2 = 0,097$ y de la tabla de la distribución normal $Z_{0,975} = 1,96$. El intervalo para el valor esperado es entonces

$$I = \left[0,467 - \frac{0,4745}{31,62} 1,96; 0,467 + \frac{0,4745}{31,62} 1,96 \right] = [0,4376; 0,4964].$$

Por lo tanto el intervalo de confianza para λ es

$$J = [1/0,4964; 1/4376] = [2,0144; 2,2852].$$

2.2.5. Intervalos de confianza para proporciones

Si bien este es un caso particular del caso anterior, vale la pena estudiarlo por separado. Lo que hacemos es considerar un experimento cuyos resultados son 0 o 1 (éxito o fracaso), y tenemos una muestra X_1, \dots, X_n independientes, que corresponden a un 0 o un 1 según si fue éxito o no. Queremos un intervalo para la probabilidad de éxito. Veamos un ejemplo

Ejemplo 2.6. Para estimar la proporción de roedores de una cierta especie que padecen determinada infección se realiza un examen histológico a 182 individuos y se encuentra que 72 están infectados. Dar un intervalo de confianza 95% para la proporción total de roedores infectados.

Observemos primero que aquí los 182 datos son 0 o 1 según el roedor está o no infectado. Si llamamos X a una variable que toma el valor 1 con probabilidad $p = P(\text{infectado})$ y 0 con probabilidad $1 - p$ tenemos que, usando la fórmula para el cálculo de la esperanza de una variable

$$E(X) = 0 \times (1 - p) + 1 \times p = p$$

Es decir el valor esperado de dicha variable es p . Por otro lado es fácil ver que $Var(X) = p(1 - p)$ Lo que nos pide el ejercicio es un intervalo de confianza para p . Observemos que un estimador de p es \bar{X}_n , es fácil ver que el estimador $\widetilde{Var}(X)$ que definimos en la sección anterior es

$$\widetilde{Var}(X) = \bar{X}_n(1 - \bar{X}_n).$$

En este caso, dado que sabemos \bar{X}_n en lugar de usar S_n usamos $\sqrt{\widetilde{Var}(X)}$ por lo tanto el intervalo nos queda

$$\left[\bar{X}_n - \frac{\sqrt{\bar{X}_n(1 - \bar{X}_n)}}{\sqrt{n}} Z_{1-\alpha/2}; \bar{X}_n + \frac{\sqrt{\bar{X}_n(1 - \bar{X}_n)}}{\sqrt{n}} Z_{1-\alpha/2} \right]. \quad (2.4)$$

En nuestro ejemplo $\bar{X}_n = 72/182 \approx 0,3956$ y $\sqrt{182} \approx 13,49$

$$I = \left[0,3956 - \frac{\sqrt{0,3956(1 - 0,3956)}}{13,49} 1,96; 0,3956 + \frac{\sqrt{0,3956(1 - 0,3956)}}{13,49} 1,96 \right]$$

2.3. Pruebas de hipótesis

Como dijimos al comienzo del capítulo, a veces queremos no solo un intervalo para la esperanza de una cierta variable X , sino poder afirmar con cierta certeza (de modo que la probabilidad de equivocarnos sea α , con α chico) si la misma supera o no un valor crítico μ_0 . La idea intuitiva es que, si tenemos n datos X_1, \dots, X_n , estimamos $E(X)$ mediante

el promedio \overline{X}_n y rechazamos que supera μ_0 si el promedio es menor que $\mu_0 - \delta$ donde $\delta > 0$ es un valor que depende de la cantidad de datos que tenemos, de α etc. Es intuitivo que si queremos tener más seguridad de que no nos estamos equivocando (es decir que δ sea más chico) vamos a necesitar más datos. Lo que se plantea en estos casos son dos hipótesis, la *hipótesis nula* (que usualmente se denota como H_0) y que en el caso del valor crítico es $H_0 : \mu > \mu_0$ y la *hipótesis alternativa* $H_1 : \mu \leq \mu_0$. Tenemos dos tipos de errores posibles, o bien la media μ de la población era mayor que μ_0 pero con la información que tenemos rechazamos que lo sea (es decir rechazamos H_0 cuando en realidad era cierta). O bien $\mu < \mu_0$ pero *no rechazamos* H_0 (es decir no rechazamos H_0 cuando en realidad era cierta H_1).

Es importante tener en cuenta en las aplicaciones que si μ_0 es un valor crítico que superarlo implica algún tipo de riesgo o costo grande, es *más grave* el primer error que dijimos, es decir que la media μ superaba el valor crítico, pero aún así rechazamos que lo superara. La probabilidad de dicho error es el nivel (que usualmente se denota con la letra griega α) de la prueba, es decir

$$\alpha = P_{H_0}(\text{rechazar } H_0). \quad (2.5)$$

Por otro lado, el segundo tipo de error, que también queremos que sea chico, está relacionado con la *potencia* de la prueba, usualmente se denota

$$\beta = P_{H_1}(\text{no rechazar } H_0).$$

Dado que no sabemos el valor exacto de μ (sino simplemente rechazaríamos H_0 si $\mu \leq \mu_0$) lo estimamos por medio de \overline{X}_n . Tenemos que ver cómo determinar *en que región* de valores de \overline{X}_n rechazamos H_0 . Es intuitivo que si queremos testear $H_0 : \mu > \mu_0$ y calculamos \overline{X}_n y nos da mayor que μ_0 no rechazamos H_0 . Pero, ¿qué pasa si nos da $\mu_0 - 1/10^9$?, en este caso si bien no es mayor que μ_0 la diferencia es muy chica como para que estemos seguros de que no estamos cometiendo un error producto de que tenemos pocos datos. En virtud de esto, es razonable pensar que la *región crítica* donde rechazaremos H_0 es de la forma $RC = \{\overline{X}_n < \mu_0 - \delta\}$. Y tenemos que determinar cuál es la tolerancia δ permitida. Si tenemos en cuenta (2.5) queremos encontrar δ tal que

$$P_{H_0}(\overline{X}_n < \mu_0 - \delta) = \alpha.$$

Observemos que una vez que determinamos δ rechazamos H_0 si \overline{X}_n cayó en la región crítica, es decir $\overline{X}_n < \mu_0 - \delta$. Y *no rechazamos* H_0 en caso contrario.

Aquí lo que haremos es considerar diferentes pruebas de hipótesis y veremos cuáles son los δ que hay que tomar. No vamos a demostrar que dichos valores cumplen los requisitos necesarios.

2.3.1. Pruebas de hipótesis unilaterales, datos normales, σ conocido

En el caso de que X_1, \dots, X_n tengan distribución $N(\mu, \sigma^2)$ con σ conocido, las pruebas de hipótesis

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases}$$

con $\mu_1 > \mu_0$ tienen región crítica

$$RC = \left\{ \bar{X}_n > \mu_0 + \frac{Z_{1-\alpha}\sigma}{\sqrt{n}} \right\}.$$

Por otro lado, las pruebas de hipótesis

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases}$$

Con $\mu_1 < \mu_0$. Tienen región crítica

$$RC = \left\{ \bar{X}_n < \mu_0 - \frac{Z_{1-\alpha}\sigma}{\sqrt{n}} \right\}.$$

Ejemplo 2.7. Consideremos los datos del ejemplo (2.3) correspondientes al nivel de sodio en sangre de 10 personas (observemos que aquí estamos asumiendo que los datos tienen distribución normal con σ conocido). Vamos a considerar que el nivel es crítico si es mayor o igual que $\mu_0 = 143$. Trabajaremos con $\alpha = 0,05$ Planteamos la prueba de hipótesis

$$\begin{cases} H_0 : \mu \geq 143 \\ H_1 : \mu < 143 \end{cases}$$

Planteamos la región crítica

$$RC = \left\{ \bar{X}_n < 143 - 2 \frac{Z_{0,95}}{\sqrt{10}} \right\}.$$

$Z_{0,95}$ es por definición el valor de t (que se obtiene mediante la tabla de la distribución normal) tal que $\Phi(t) = 0,95$, usando la tabla $Z_{0,95} = 1,65$ entonces

$$RC = \left\{ \bar{X}_n < 141,93 \right\}$$

Como $\bar{X}_n = 140 < 141,93$ estamos en la región crítica, por lo tanto rechazamos H_0 .

2.3.2. Pruebas de hipótesis unilaterales, datos normales, σ desconocido

En el caso de que X_1, \dots, X_n tengan distribución $N(\mu, \sigma^2)$ con σ desconocido, las pruebas de hipótesis

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases}$$

con $\mu_1 > \mu_0$ tienen región crítica

$$RC = \left\{ \bar{X}_n > \mu_0 + \frac{t_{1-\alpha}(n-1)S_n}{\sqrt{n}} \right\}.$$

Por otro lado, las pruebas de hipótesis

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases}$$

Con $\mu_1 < \mu_0$. Tienen región crítica

$$RC = \left\{ \bar{X}_n < \mu_0 - \frac{t_{1-\alpha}(n-1)S_n}{\sqrt{n}} \right\}.$$

Ejemplo 2.8. La siguiente muestra registra los niveles en sangre de colesterol *LDL*, medidos en *mg/dl*, correspondientes a diez pacientes de una clínica.

134,3	133,7	129,5	131,2	128,4	125,4	135,0	130,4	133,0	132,7
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Clínicamente, se consideran altos los niveles de colesterol que están por encima de 130 mg/dl. Asumiendo que los datos tienen distribución normal con media μ y desviación estándar σ (que por el momento suponemos desconocida), implemente (con nivel $\alpha = 0,05$) la siguiente prueba de hipótesis para decidir si los niveles de colesterol son razonables:

$$\begin{cases} H_0 : \mu \geq 130 \\ H_1 : \mu < 130 \end{cases}$$

Observemos que como desconocemos σ tenemos que estimarlo mediante S_n , haciendo cuentas $\bar{X}_n = 131,36$, $S_n^2 = 8,9493$. Por otro lado, de la tabla de la *t*, tenemos que $t_{0,95}(9) = 1,83$

$$RC = \left\{ \bar{X}_n < 130 - 1,83 \frac{2,991}{\sqrt{10}} \right\} = \left\{ \bar{X}_n < 128,27 \right\}$$

Como $131,36 > 128,27$ no estamos en la región crítica y por lo tanto no se rechaza H_0 .

2.3.3. Pruebas de hipótesis bilaterales, datos normales, σ conocido

En el caso de que X_1, \dots, X_n tengan distribución $N(\mu, \sigma^2)$ con σ conocido, la prueba de hipótesis:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

tiene región crítica

$$RC = \left\{ \left| \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma} \right| \geq Z_{1-\alpha/2} \right\}$$

Ejemplo 2.9. Cada luciérnaga tiene un modo peculiar de centellear; un destello corto de luz es seguido por un período de reposo. Los siguientes datos corresponden a los períodos de reposo entre centelleos (medidos en segundos) para una muestra de 11 luciérnagas:

4,05	3,95	3,74	3,33	3,94	4,04	3,73	3,75	3,88	3,50	3,59
------	------	------	------	------	------	------	------	------	------	------

Se asume que los datos corresponden a una distribución normal de valor esperado μ y desvío $\sigma = 0,06$. Realice una prueba de hipótesis al nivel 0,1 para decidir entre las siguientes hipótesis

$$\begin{cases} H_0 : \mu = 4 \\ H_1 : \mu \neq 4 \end{cases}$$

$\bar{X}_n = 3,773$, $\sigma = 0,06$. Haciendo cuentas tenemos que

$$\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma} = \frac{\sqrt{11}(3,773 - 4)}{0,06} = -12,714$$

y al tomar valor absoluto tenemos que $|-12,714| = 12,714$ que es mayor que $Z_{0,95} = 1,65$ y por lo tanto estamos en la región crítica y se rechaza H_0 .

2.3.4. Pruebas de hipótesis bilaterales, datos normales, σ desconocido

En el caso de que X_1, \dots, X_n tengan distribución $N(\mu, \sigma^2)$ con σ desconocido, la pruebas de hipótesis

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu \neq \mu_0 \end{cases}$$

tiene región crítica

$$RC = \left\{ \left| \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n} \right| \geq t_{1-\alpha/2}(n-1) \right\}$$

2.3.5. Pruebas de hipótesis para proporciones

Supongamos que queremos saber si una moneda está balanceada o no. Se tira 100 veces y obtenemos 54 caras, debemos tomar una decisión entre

$$\begin{cases} H_0 : & p = 1/2 \quad \text{donde } p = P(\text{cara}) \\ H_1 : & p \neq 1/2. \end{cases}$$

Es razonable pensar que rechazaremos H_0 si el promedio de veces que salió cara es mucho menor que 1/2 o mucho mayor. Es decir planteamos una región crítica de nivel α de la forma

$$RC = \left\{ \left| \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma} \right| \geq Z_{1-\alpha/2} \right\},$$

Observemos que, bajo H_0 estamos en el caso en que $\mu = p = 1/2$ por lo tanto $\sigma^2 = p(1-p)$. En el caso de la moneda, supongamos que trabajamos con $\alpha = 0,05$, $\bar{X}_n = 54/100$, $Z_{1-\alpha/2} = Z_{0,975} = 1,96$. Haciendo cuentas

$$\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma} = \frac{10(54/100 - 1/2)}{1/2} = 0,8$$

Como $0,8 < 1,96$ no estamos en la región crítica y por lo tanto no se rechaza H_0 .

2.4. Pruebas de bondad de ajuste

Hasta ahora hemos hecho pruebas de hipótesis a partir de una muestra X_1, \dots, X_n de n observaciones de una variable X asumiendo que la misma tiene distribución normal cuya media μ desconocemos. En el primer caso supusimos que conocíamos la varianza σ^2 y en el segundo que no la conocíamos. Finalmente vimos pruebas de hipótesis para proporciones, es decir los datos X_i son ceros o unos. En todos los casos considerados estamos suponiendo que, a menos de algunos parámetros, conocemos la *forma* de la distribución. Es claro que esta situación ideal no es ciera en general en la práctica, donde solo contamos con los datos. Por lo tanto nos puede interesar decidir, con cierto nivel de confianza, si efectivamente dicha muestra tiene esa distribución, es decir, si tiene distribución $N(1, 4)$ o no, si tiene distribución uniforme en el intervalo $[0, 1]$ o no, etc. Veamos con un ejemplo como se construye el test y la región crítica.

Ejemplo 2.10. Se tiene una muestra que registra la altura en centímetros de 100 niños de siete años de determinada población. Se intenta ver si es razonable afirmar que la misma corresponde a una variable aleatoria normal con media 118 centímetros y desvío estándar $\sigma = 5$ centímetros. Para ello, se resumen en una tabla las observaciones:

altura en cm.	n. observado (o_i)
menos de 111	7
entre 111 y 115	17
entre 115 y 119	29
entre 119 y 122	30
más de 122	17

Suponer que los datos tienen la distribución de una variable $X \sim N(118, 25)$ significa, por ejemplo, que la probabilidad de que la altura de un niño sea menor que 111 es $P(X < 111)$, análogamente la probabilidad de que la altura de un niño este entre 111 y 115 es $P(111 \leq X \leq 115)$, etc. Observemos que estos valores los podemos calcular a partir de la tabla de la distribución normal. Esto nos dice que, si dicha hipótesis fuese cierta, deberíamos tener, entre las 100 alturas, *aproximadamente*:

- $100 \times P(X < 111) = 100 \times 0,0808$ alturas menores que 111.
- $100 \times P(111 \leq X \leq 115) = 100 \times 0,1935$ alturas entre 111 y 115.
- $100 \times P(115 < X \leq 119) = 100 \times 0,305$ alturas entre 115 y 119.
- $100 \times P(119 < X \leq 122) = 100 \times 0,2088$ alturas entre 119 y 122.
- $100 \times P(X > 122) = 100 \times 0,2119$ alturas mayores que 122

Por lo tanto una idea intuitiva para testear si los datos de la tabla provienen efectivamente de la distribución $N(118, 25)$ es compararlos con las frecuencias anteriores. Por ejemplo considerar

$$(100 \times 0,0808 - 7)^2 + (100 \times 0,1935 - 17)^2 + (100 \times 0,305 - 29)^2 + (100 \times 0,2088 - 30)^2 + (100 \times 0,2119 - 17)^2$$

Donde las diferencias, al igual que en el caso de la definición de la varianza, se elevan al cuadrado para que los valores más grandes *pesen más* que los mas chicos. Para darle mayor peso a las diferencias correspondientes a intervalos de probabilidad pequeña vamos a dividir cada sumando por la *frecuencia teórica* del intervalo, es decir consideraremos

$$\frac{(100 \times 0,0808 - 7)^2}{100 \times 0,0808} + \frac{(100 \times 0,1935 - 17)^2}{100 \times 0,1935} + \frac{(100 \times 0,305 - 29)^2}{100 \times 0,305} + \frac{(100 \times 0,2088 - 30)^2}{100 \times 0,2088} + \frac{(100 \times 0,2119 - 17)^2}{100 \times 0,2119} = 5,315 \quad (2.6)$$

Observemos que si llamamos I_1 al intervalo $(-\infty, 111]$, I_2 al intervalo $(111, 115]$, I_3 al intervalo $(115, 119]$, I_4 al intervalo $(119, 122]$ e I_5 al intervalo $(122, +\infty]$ y n al número de observaciones (en nuestro caso $n = 100$) lo que calculamos en (2.6) no es otra cosa que

$$T_n = \sum_{i=1}^5 \frac{(n \times P(X \in I_i) - [\text{nro de observaciones en } I_i])^2}{n \times P(X \in I_i)} \quad (2.7)$$

La región crítica, es decir los valores de X_1, \dots, X_n que hacen que rechazemos la hipótesis (nula) de que la muestra tiene distribución $N(118, 25)$ serán aquellos que hacen que T_n sea *grande*, por lo tanto es razonable plantear una región crítica de la forma $RC = \{T_n > h\}$ donde h es un valor que se determina en función del nivel de confianza en el que estemos trabajando. Es decir, queremos que encontrar h tal que $\alpha = P_{H_0}(T_n > h)$. Se puede demostrar que si los datos efectivamente provienen de la distribución $N(118, 25)$ entonces

$$P_{H_0}(T_n > h) \rightarrow P(\chi_4^2 > h)$$

donde χ_4^2 fue definida en (1.25), con $k = 4$. Por lo tanto para determinar h tenemos que plantear

$$\alpha = P(\chi_4^2 > h).$$

Por último determinamos el valor de h a partir de la tabla de la distribución χ_4^2 . Por ejemplo si estamos trabajando a nivel $\alpha = 0,05$ (95% de confianza) y $k = 4$ la tabla nos dice que $h = 9,488$, como $5,315 < 9,48$ no rechazamos la hipótesis de que los datos provengan de una variable con distribución $N(118, 25)$.

La prueba que explicamos a partir de un ejemplo se conoce como Prueba chi-cuadrado de Pearson, veamos ahora la explicación general. Supondremos primero que tenemos k intervalos I_1, \dots, I_k que forman una partición de todos los números reales, es decir $I_1 = (-\infty, a_1]$, $I_2 = (a_1, a_2]$, $I_3 = (a_2, a_3]$ hasta $I_k = (a_{k-1}, +\infty)$, con $a_1 < a_2 < a_3 < \dots < a_{k-1}$. Vamos a suponer que tenemos una variable X de la cual queremos saber si tiene una cierta distribución conocida, que denotaremos F_0 o no, es decir, la prueba que planteamos es

$$\begin{cases} H_0 : & X \sim F_0 \\ H_1 : & X \text{ no tiene distribución } F_0. \end{cases}$$

Vamos a llamar p_i a la probabilidad de que X pertenezca al intervalo I_i si es cierto H_0 es decir $p_i = P_{H_0}(X \in I_i)$. Asumiremos que los intervalos son tomados de forma tal que $p_i > 0$ para todo i . Es importante observar además que como estamos suponiendo que F_0 la conocemos (por ejemplo es $N(118, 25)$) los valores p_i se pueden calcular. Construimos ahora el *estimador* que usaremos, que será la generalización del T_n que definimos antes:

$$T_n = \sum_{i=1}^k \frac{(n \times p_i - [\text{nro de observaciones en } I_i])^2}{n \times p_i} \quad (2.8)$$

Nuevamente la región crítica es $RC = \{T_n > h\}$. Por último, para determinar h de modo tal que RC tenga nivel α usamos el resultado

$$P_{H_0}(T_n > h) \rightarrow P(\chi_{k-1}^2 > h)$$

del cual se sigue que h es el valor (obtenido mediante el uso de la tabla de la distribución χ_{k-1}^2) que deja probabilidad α a la derecha de h , es decir:

$$\alpha = P(\chi_{k-1}^2 > h).$$

Capítulo 3

Test de aleatoriedad

3.1. Introducción

Para calcular intervalos de confianza, hacer pruebas de hipótesis, y hacer pruebas de bondad de ajuste hemos asumido que los datos son independientes e idénticamente distribuidos. Sin dicha hipótesis muchos de los cálculos que hemos hecho hasta ahora no son válidos. En éste capítulo veremos algunos test que permiten chequear cuándo una muestra X_1, \dots, X_n cumple dichas hipótesis. Como no haremos suposiciones respecto de la distribución de las X_i veremos métodos que se basan en la forma en que las variables están ordenadas.

3.1.1. Test de Rachas

Supongamos que tenemos 10 mediciones de temperatura hechas a la misma hora, durante 20 días, es decir tenemos X_1, \dots, X_{10} es claro que si los datos son crecientes, no van a ser independientes, ya que saber el valor de la temperatura en el día i -ésimo nos aporta información respecto de lo que pasará con la temperatura el día siguiente, nos dice que será mayor. Lo mismo diríamos si los valores son decrecientes o si alternan (crece: $X_1 < X_2$, decrece: $X_3 < X_2$, crece: $X_3 < X_4$, etc). Esta idea intuitiva que relaciona el crecimiento-decrecimiento de los datos es la que usaremos para testear la aleatoriedad de la muestra. Lo que hacemos es, dados los datos X_1, \dots, X_n contruir $n - 1$ variables Y_1, \dots, Y_{n-1} donde la variable Y_i es 1 si $X_i < X_{i+1}$ y 0 en otro caso. Por ejemplo, si las X_i son mediciones (en grados celcius) de temperatura dadas por la siguiente tabla:

1	22,67	6	17,88
2	21.66	7	23.17
3	16.31	8	24,85
4	15.95	9	15,17
5	15,15	10	23,19

Las variables Y_1, \dots, Y_9 son

1	0	6	1
2	0	7	1
3	0	8	0
4	0	9	1
5	1		

En general, supongamos que tenemos variables aleatorias X_1, X_2, \dots, X_n y queremos chequear si son i.i.d, para eso definimos las variables Y_1, \dots, Y_{n-1} de la siguiente forma

$$Y_i = \mathbb{I}_{\{X_i < X_{i+1}\}}.$$

Lo que haremos es estudiar el número de rachas totales de Y_1, \dots, Y_{n-1} o lo que es lo mismo

$$\hat{R}_n = 1 + \sum_{i=1}^{n-2} \mathbb{I}_{\{Y_i \neq Y_{i+1}\}}.$$

Observemos que en general $\hat{R}_n \leq n - 1$.

En el caso del ejemplo tenemos que $n = 10$ y $\hat{R}_n = 4$ Si n es chico ($n \leq 25$) la distribución de \hat{R}_n esta tabulada, y rechazamos la hipótesis de ser i.i.d., a nivel α , si el valor observado \hat{R}_n cumple que $|\hat{R}_n| > R_{\alpha/2}$ donde $R_{\alpha/2}$ se obtiene de la tabla. La tabla nos da, para diferentes valores de n la probabilidad de que se obtengan k rachas, con k de 0 a $n - 1$ veamos en el ejemplo. En nuestro caso vamos en la tabla a $n = 10$ y obtenemos

n	\hat{R}_n	Left tail P	\hat{R}_n	Right tail P
10	1	0.0000		
	2	0.0003	9	0.0278
	3	0.0079	8	0.1671
	4	0.0633	7	0.4524
	5	0.2427	6	0.7573

Esta tabla hay que interpretarla de la siguiente manera, si \hat{R}_{10} denota la variable aleatoria que indica el número de rachas posibles cuando tenemos $n = 10$ datos (que es un valor como dijimos, que puede ser 1, 2, 3, ..., 9), tenemos que la columna *Left tail* indica las probabilidades hasta ese valor de \hat{R}_{10} , y por lo tanto va creciendo a medida que \hat{R}_{10} aumenta, es decir

$$P(\hat{R}_{10} \leq 1) = 0,0000, P(\hat{R}_{10} \leq 2) = 0,0003, P(\hat{R}_{10} \leq 3) = 0,0079, P(\hat{R}_{10} \leq 4) = 0,0633, P(\hat{R}_{10} \leq 5) = 0,2427.$$

La columna *Right tail* nos da la probabilidad de obtener valores mayores o iguales que el valor, por lo tanto va creciendo a medida que \hat{R}_{10} decrece, es decir

$$P(\hat{R}_{10} \geq 6) = 0,7573, P(\hat{R}_{10} \geq 7) = 0,4524, P(\hat{R}_{10} \geq 8) = 0,1671 \text{ y } P(\hat{R}_{10} \geq 9) = 0,0278.$$

Supongamos que nuestro nivel de confianza es $\alpha = 0,1$. Vamos a rechazar H_0 (que los datos son i.i.d) si obtenemos valores de \hat{R}_{10} muy grandes, o muy chicos. Dado que $P(\hat{R}_{10} \leq 3) = 0,079 < \alpha/2$ y $P(\hat{R}_{10} \leq 4) = 0,0633 > \alpha/2$, la región crítica incluye los valores de $\hat{R}_n = 1, 2, 3$. Por otro lado $P(\hat{R}_{10} \geq 9) = 0,0278 < \alpha/2$ mientras que $P(\hat{R}_{10} \geq 8) = 0,1671$. Finalmente, en virtud de esto, la región crítica es

$$RC = [1, 3] \cup [9]$$

como obtuvimos $\hat{R}_{10} = 4$ no estamos en la región crítica y por lo tanto no se rechaza la hipótesis (nula) de que los datos son i.i.d.

Ejercicio 3.1.

- Usando la tabla de \hat{R}_n demostrar que la región crítica para $\alpha = 0,10$ y $n = 15$ datos es

$$RC = [1, 5] \cup [14, 15]$$

- Usando la tabla de \hat{R}_n demostrar que la región crítica para $\alpha = 0,10$ y $n = 23$ datos es

$$RC = [1, 11] \cup [19, 22]$$

mientras que si tomamos $\alpha = 0,05$ la región crítica (para $n = 23$) es

$$RC = [1, 10] \cup [20, 22]$$

Para valores grandes de n ($n > 25$) usamos que

$$P \left(\frac{\hat{R}_n - (2n - 1)/3}{\sqrt{\frac{16n-29}{90}}} \leq t \right) \rightarrow \Phi(t)$$

por lo tanto vamos a rechazar H_0 si

$$\left| \frac{\hat{R}_n - (2n - 1)/3}{\sqrt{\frac{16n-29}{90}}} \right| > Z_{\alpha/2}$$

3.1.2. Test de Spearman

El test de Spearman sirve, al igual que el test de rachas, para decidir si una muestra de datos X_1, \dots, X_n es i.i.d. Veamos como se construye el estadístico que usaremos. A partir de una muestra X_1, \dots, X_n ordenamos los datos y obtenemos $X_{(1)} \leq \dots \leq X_{(n)}$. Por ejemplo en el caso de la tabla con las 10 mediciones que usamos para el test de rachas, los datos ordenados son

$$X_{(1)} = 15, 15, X_{(2)} = 15, 17, X_{(3)} = 15, 95, X_{(4)} = 16, 31, X_{(5)} = 17, 88, X_{(6)} = 21, 66, X_{(7)} = 22, 67, X_{(8)} = 23, 17, X_{(9)} = 23, 19, X_{(10)} = 24, 85$$

Construimos los estadísticos de rangos R_1, \dots, R_n donde

$$R_i = \sum_{j=1}^n \mathbb{I}_{\{X_j \leq X_i\}}.$$

Es decir, para cada dato X_i , R_i cuenta la cantidad de datos menores o iguales que X_i . Por ejemplo, como $X_1 = 22,67 = X_{(7)}$ tenemos que $R_1 = 7$, como $X_2 = 21,66 = X_{(6)}$ $R_2 = 6$ y así sucesivamente, obtenemos que

$$R_1 = 7, R_2 = 6, R_3 = 4, R_4 = 3, R_5 = 1, R_6 = 5, R_7 = 8, R_8 = 10, R_9 = 2, R_{10} = 9$$

Calculamos el estadístico mediante la fórmula

$$\rho_s = 1 - \frac{6D}{n(n^2 - 1)},$$

con $D = \sum_{i=1}^n (R_i - i)^2$. En el caso de nuestro ejemplos es

$$\rho_s = 1 - \frac{6D}{10 \times 99} \quad (3.1)$$

y

$$D = (7-1)^2 + (6-2)^2 + (4-3)^2 + (3-4)^2 + (1-5)^2 + (5-6)^2 + (8-7)^2 + (10-8)^2 + (2-9)^2 + (9-10)^2 = 126$$

Es decir, $\rho_1 = 0,2364$.

Se sabe, aunque no lo veremos aquí, que ρ_s es una variable aleatoria discreta que tiene una distribución simétrica y toma valores entre -1 y 1 . Se puede demostrar además que $E(\rho_s) = 0$ y $Var(\rho_s) = 1/(n-1)$. Definimos la región crítica $RC = \{(X_1, \dots, X_n) / |\rho_s| > c\}$, tenemos que determinar (al igual que como hacíamos en el test de rachas con \hat{R}_n) el valor de c , que dependerá del nivel de significación de la prueba. Para valores de n menores o iguales que 10 existen tablas con la distribución de ρ_s que nos dan, para diferentes valores k_i de ρ_s entre 0 y 1 (en el caso de que el estadístico de un valor negativo usamos la simetría de ρ_s) la probabilidad de obtener un valor mayor o igual que k_i , es decir $P(\rho_s \geq k_i)$. En el caso por ejemplo de $n = 5$ la tabla es la siguiente

n	R	P
5	1.000	0.008
	0.900	0.042
	0.800	0.067
	0.700	0.117
	0.600	0.175
	0.500	0.225
	0.400	0.258
	0.300	0.342
	0.200	0.392
	0.100	0.475
0.000	0.525	

Esta tabla nos dice en particular que si $n = 5$ $P(\rho_s \geq 1) = 0,008$, $P(\rho_s \geq 0,900) = 0,042$, $P(\rho_s \geq 0,800) = 0,067$ etc. Si tomáramos $\alpha = 0,1$ y tuviéramos 5 datos, la región crítica es entonces $RC = [-1, -0,9] \cup [0,9, 1]$. rechazamos a nivel α si nuestro estadístico ρ_s calculado mediante (3.1) nos da mayor o igual que 0,900 o menos que $-0,900$. En el caso de $n = 10$ y $\alpha = 0,1$ la región crítica es

$$RC = [-1; -0,564] \cup [0,564; 1]$$

Como obtuvimos $\rho_s = 0,2364$ no estamos en la región crítica y por lo tanto no se rechaza la hipótesis de que los datos son *i.i.d.*

Ejercicio 3.2.

- Usando la tabla de ρ_s demostrar que la región crítica para $\alpha = 0,10$ y $n = 7$ datos es

$$RC = [-1; -0,714] \cup [0,714; 1]$$

- Usando la tabla de ρ_s demostrar que la región crítica para $\alpha = 0,05$ y $n = 9$ datos es

$$RC = [-1; 0,683] \cup [0,683; 1]$$

De $n = 11$ a $n = 30$ la forma de la tabla cambia, veamos como es para algunos valores de n

n	0.100	0.050	0.025	0.010	0.005	0.001
11	0.427	0.536	0.618	0.709	0.764	0.855
12	0.406	0.503	0.587	0.678	0.734	0.825
30	0.241	0.307	0.363	0.426	0.467	0.548

Supongamos que $n = 12$, esta tabla nos dice que $P(\rho_s \geq 0,406) = 0,1$, $P(\rho_s \geq 0,503) = 0,05, \dots$, $P(\rho_s \geq 0,825) = 0,001$. Por lo tanto nuestra región crítica si $\alpha = 0,1$ es $RC = [-1; -0,503] \cup [0,503, 1]$. Si fuese $n = 30$ y $\alpha = 0,1$ es $RC = [-1; -0,307] \cup [0,307; 1]$ mientras que si $n = 30$ y $\alpha = 0,05$ la región crítica es $RC = [-1; -0,363] \cup [0,363; 1]$. Observe que para el mismo α a medida que n crece la longitud de la región crítica crece, ¿por qué?.

Para valores grandes de n ($n > 30$) usamos que

$$P(\sqrt{n-1}\rho_s \leq t) \rightarrow \Phi(t)$$

por lo tanto vamos a rechazar H_0 si

$$|\sqrt{n-1}\rho_s| > Z_{\alpha/2}.$$

Índice alfabético

Arreglos, 5

Combinaciones, 6

extracciones sin reposición, 12

Independencia, 9

Intervalos de confianza

datos cualquiera

σ conocido, 19

σ desconocido, 19

datos normales

σ conocido, 17

σ desconocido, 18

para proporciones, 20

intervalos de confianza, 15

Permutaciones, 4

Probabilidad, 7

de la unión, 8

del complemento, 8

pruebas de bondad de ajuste - chi cuadrado,
23

pruebas de hipótesis, 20

datos normales

σ conocido, 21

σ desconocido, 21

para proporciones, 23

Test de Rachas, 26

Test de Spearman, 28

variable aleatoria, 10

con distribución

binomial, 10

de Poisson, 13

geométrica, 11

hipergeométrica, 12

esperanza y varianza

estimación, 16

Bibliografía

- [1] D. Freedman, R. Pisani, and R. Purves. *Estadística. 2ed.* Antoni Bosch, Editor., 1993.
- [2] R. Grimaldi. *Matemáticas discreta y combinatoria. 3ed.* Addison Wesley, 1997.
- [3] B. Shahbaba. *Bioestadistics with R.* Springer, 2012.

Índice alfabético

Arreglos, 5

Combinaciones, 6

extracciones sin reposición, 12

Independencia, 9

Intervalos de confianza

datos cualquiera

σ conocido, 19

σ desconocido, 19

datos normales

σ conocido, 17

σ desconocido, 18

para proporciones, 20

intervalos de confianza, 15

Permutaciones, 4

Probabilidad, 7

de la unión, 8

del complemento, 8

pruebas de bondad de ajuste - chi cuadrado,
23

pruebas de hipótesis, 20

datos normales

σ conocido, 21

σ desconocido, 21

para proporciones, 23

Test de Rachas, 26

Test de Spearman, 28

variable aleatoria, 10

con distribución

binomial, 10

de Poisson, 13

geométrica, 11

hipergeométrica, 12

esperanza y varianza

estimación, 16