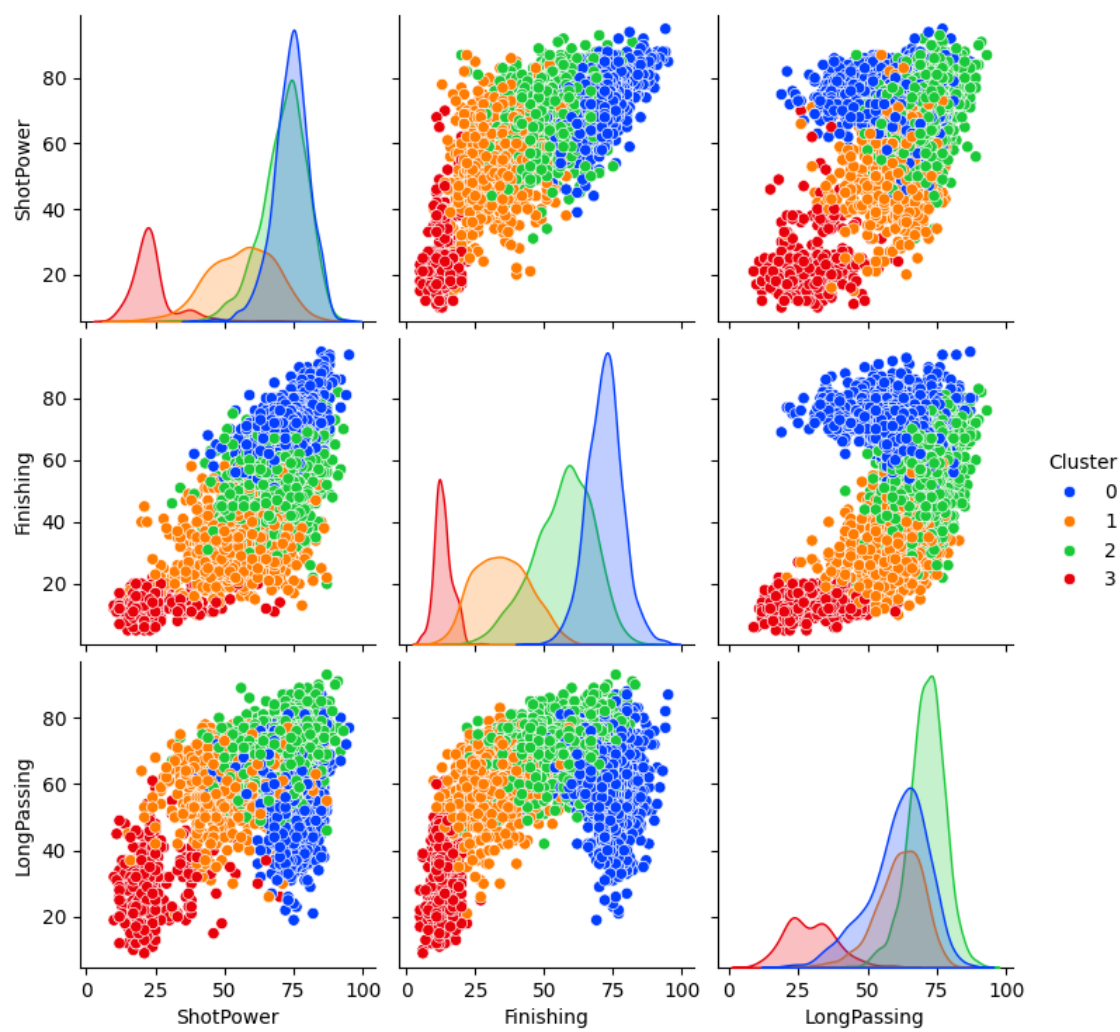


# **Elementos de probabilidad y estadística**

Dra. Ana Georgina Flesia

12 de agosto de 2024



# Índice general

<b>1. Introducción</b>	<b>7</b>
1.1. Introducción . . . . .	7
1.2. Un poco de historia . . . . .	9
<b>2. Teoría de Conteo de conjuntos discretos</b>	<b>11</b>
2.1. Teoría de Conjuntos . . . . .	11
2.1.1. Noción intuitiva de conjunto . . . . .	11
2.1.2. Operaciones entre conjuntos . . . . .	13
2.1.3. Diagramas de Venn . . . . .	13
2.2. Conjuntos finitos: Combinatoria . . . . .	16
2.2.1. Variaciones . . . . .	19
2.2.2. Permutaciones . . . . .	20
2.2.3. Combinaciones . . . . .	22
2.2.4. Bolas en celdas . . . . .	25
<b>3. Espacios de Probabilidad</b>	<b>31</b>
3.1. Modelo matemático para una población . . . . .	31
3.2. Medidas de Probabilidad . . . . .	39
3.2.1. Espacios discretos: Cálculo de la probabilidad de un evento . . . . .	44
3.3. Probabilidad condicional . . . . .	54
3.3.1. Independencia . . . . .	62
3.4. Variables aleatorias . . . . .	65
3.4.1. Media y Varianza . . . . .	68
3.4.2. Correlación . . . . .	70
3.4.3. Coeficiente de Correlación lineal . . . . .	72
<b>4. Variables aleatorias discretas</b>	<b>75</b>
4.1. Introducción . . . . .	75
4.2. Cómputos con densidades . . . . .	77
4.3. Vectores aleatorios . . . . .	79
4.4. Variables aleatorias independientes . . . . .	81
4.4.1. Esperanza de una variable aleatoria discreta . . . . .	82
4.4.2. Propiedades de la esperanza . . . . .	84
4.4.3. Varianza de una variable discreta . . . . .	85
4.5. Distribuciones discretas clásicas . . . . .	87
4.5.1. Variable aleatoria Bernoulli . . . . .	87
4.5.2. Distribución Binomial . . . . .	87
4.5.3. Distribución Hipergeométrica . . . . .	89
4.5.4. Distribución Geométrica . . . . .	91

4.5.5.	Distribución Binomial negativa . . . . .	93
4.5.6.	Distribución Poisson . . . . .	95
4.5.7.	Distribución Multinomial . . . . .	97
4.5.8.	Suma de variables aleatorias independientes . . . . .	99
<b>5.</b>	<b>Variables aleatorias continuas</b>	<b>101</b>
5.1.	Introducción . . . . .	101
5.2.	Distribución de probabilidad continua . . . . .	102
5.2.1.	Densidad de probabilidad . . . . .	103
5.2.2.	Esperanza de una variable aleatoria continua . . . . .	105
5.2.3.	Cómputo de densidades . . . . .	106
5.3.	Distribuciones continuas clásicas . . . . .	109
5.3.1.	Distribución Uniforme . . . . .	109
5.3.2.	Distribución Exponencial . . . . .	111
5.3.3.	Distribución Normal . . . . .	113
5.3.4.	Uso de la tabla . . . . .	116
5.4.	Distribución conjunta . . . . .	117
5.4.1.	Independencia . . . . .	119
5.4.2.	Propiedades de la esperanza . . . . .	121
<b>6.</b>	<b>Distribuciones de Muestreo</b>	<b>125</b>
6.1.	Introducción a la metodología estadística . . . . .	125
6.2.	Distribuciones del muestreo . . . . .	132
6.2.1.	Tipos de Muestreo . . . . .	132
6.2.2.	Muestreo de una distribución teórica . . . . .	133
6.3.	Aproximaciones . . . . .	138
6.4.	Ley de los grandes números . . . . .	140
6.4.1.	Desigualdad de Chebyshev . . . . .	141
6.4.2.	Ley de los grandes números . . . . .	144
6.4.3.	Proporciones . . . . .	146
6.5.	Teorema Central del Límite . . . . .	147
6.5.1.	Aproximaciones normales: Binomial . . . . .	147
6.5.2.	Teorema central del límite para una muestra aleatoria simple . . . . .	150
<b>7.</b>	<b>Inferencia estadística</b>	<b>155</b>
7.1.	¿Cómo se hacen las inferencias? . . . . .	155
7.2.	Estimar con confianza . . . . .	156
7.2.1.	Estimadores puntuales . . . . .	156
7.2.2.	Propiedades de un estimador . . . . .	157
7.2.3.	Algunos estimadores puntuales muy usados . . . . .	158
7.3.	Estimación por intervalo . . . . .	159
7.4.	Confianza estadística . . . . .	160
7.4.1.	Varianza desconocida . . . . .	164
7.5.	Ajuste de datos: un caso de distribución Poisson . . . . .	167
<b>8.</b>	<b>Nociones de test de hipótesis</b>	<b>171</b>
8.1.	Introducción . . . . .	171
8.2.	La lógica del test de hipótesis . . . . .	172
8.2.1.	¿Como planteamos las hipótesis? . . . . .	173

8.2.2.	¿Como tomamos la decisión?	173
8.3.	P-valores	174
8.3.1.	Significado estadístico	175
8.4.	$z$ test de nivel $\alpha$	178
8.5.	Abuso de los test de hipótesis	181
8.5.1.	Inferencia estadística no corrige faltas de diseño	181
8.5.2.	Cuidado con la búsqueda de valores significativos	182
<b>9.</b>	<b>Inferencia sobre distribuciones normales</b>	<b>185</b>
9.1.	Procedimientos basados en la distribución $t$	185
9.1.1.	Muestras Apareadas	187
9.2.	Comparando dos medias	188
9.2.1.	El $z$ -test para dos muestras: $\sigma_1$ y $\sigma_2$ conocidos	189
9.2.2.	El test $t$ para dos muestras: $\sigma_1 = \sigma_2$ desconocido	191
9.2.3.	El test $t$ para dos muestras: $\sigma_1$ y $\sigma_2$ desconocidos	192



# Capítulo 1

## Introducción

### 1.1. Introducción

La estadística es un sustantivo tanto singular como plural. Como plural significa hechos numéricos sistemáticamente recogidos y como singular es la ciencia de la recopilación, clasificación y utilización de estadísticas.

¿Qué hace la estadística?

1. Permite presentar hechos en una forma precisa y definida que ayudan a obtener una comprensión adecuada de lo que se dice. Los hechos exactos son más convincentes que las declaraciones vagas.
2. Ayuda a condensar la masa de datos en unas pocas medidas numéricas, es decir, resume los datos y presenta una información general significativa sobre una masa de los datos.
3. Ayuda a encontrar relaciones entre diferentes factores involucrados en la prueba de la validez de una relación asumida.
4. Ayuda a predecir los cambios de un factor debido a los cambios en otro factor.
5. Ayuda en la formulación de planes y políticas que requieren el conocimiento de nuevas tendencias;
6. Las estadísticas desempeñan un papel vital en la adopción de decisiones.

Las técnicas estadísticas se utilizan para solucionar problemas prácticos en casi todos los aspectos de la vida. ¿Como lo hacen? ¿Como se utilizan estas técnicas? Antes de contestar estas preguntas examinemos algunos ejemplos de interés para el usuario de la estadística.

#### Ejemplo 1.1.1

- Se diseñan encuestas para recaudar información previa al día de las elecciones y así predecir el resultado de las mismas.
- Se seleccionan al azar consumidores para obtener información con el fin de predecir la preferencia con respecto a cierto producto.
- El médico que investiga realiza experimentos para determinar el efecto de ciertos medicamentos y de condiciones ambientales controladas en los humanos para así determinar el método apropiado para curar cierta enfermedad.
- El ingeniero muestrea las características de calidad de un producto, junto con otras variables controlables del proceso, para distinguir que variables están más relacionadas con dicha calidad y usarlas para predecir si un lote está en condición de ser vendido o no.

- El economista considera varios índices de la situación económica durante cierto período y los usa para predecir la situación económica futura.

Los ejemplos antes citados son de naturaleza y complejidad variada, pero cada uno de ellos implican predicción y toma de decisiones. También, cada uno de estos ejemplos implican muestreo.

**Definición 1.1.1** *El gran conjunto de datos que es centro de nuestro interés se denomina **población** y el subconjunto de allí seleccionado se denomina **muestra**.*

### Ejemplo 1.1.2

- Las decisiones de los votantes por un candidato a la gobernación (Márquez) expresadas de manera cuantitativa (1 por sí, 0 por no) determinan una población real y finita de gran interés para Márquez. El desearía considerar el conjunto total de votantes para determinar la fracción a favor de su elección.
- Las mediciones efectuadas en pacientes sometidos a un experimento representan una muestra de una población conceptual (teórica) de todos los pacientes que sufren una misma enfermedad y de los que la padecerían en un futuro próximo.

El objetivo de la estadística es hacer inferencias (predicciones, decisiones) sobre la población basada en la información de una muestra. La teoría estadística propone dos pasos para realizar inferencias, el diseño de técnicas específicas para un tipo de experimento, y la cuantificación del ajuste del modelo a la situación real. Por ejemplo, es de interés conocer el valor de la probabilidad con que la estimación obtenida con la muestra se aproxima al valor real. Intuitivamente podemos afirmar que cuanto mayor sea el número de votantes incluidos en la muestra mayor será la probabilidad de una estimación exacta.

El mecanismo para realizar inferencias es la teoría de la probabilidad. El probabilista usa la información acerca de una población modelo para caracterizar el resultado del experimento, la muestra. En cambio el estadístico utiliza la teoría de probabilidad para calcular la probabilidad de una muestra observada y inferir de ella las características de una población desconocida. Siguiendo nuestro ejemplo de los votantes, un probabilista puede hacer una familia de modelos que permita conocer la probabilidad de que un número  $n$  de votantes (entre  $N$  elegidos al azar) esté a favor de Márquez, si la verdadera proporción de votantes a favor de Márquez es un número  $p$ . Un estadístico puede inferir que si el porcentaje de votantes a favor de Márquez en la muestra observada es del 100 %, Márquez tiene grandes chances de ganar las elecciones. ¿Porque? Pues la probabilidad de obtener esa muestra si Márquez no fuera a ganar es muy pequeña.

Este es un ejemplo simple que ilustra porqué la teoría de la probabilidad es el fundamento de la teoría de la estadística. Para el estadístico es de primordial importancia conocer la probabilidad de un evento, pues si realiza una inferencia usando la premisa de "supongo que ocurre lo mas probable", el error que comete es proporcional a cuan improbable es el evento complementario.

En esta introducción estamos usando palabras como *evento* y *evento complementario* que refieren a sucesos cuantificables mediante conjuntos de una población, finita o infinita. En el próximo capítulo enunciaremos algunas reglas de la teoría de conjuntos que son necesarias para luego definir la noción de probabilidad.



## 1.2. Un poco de historia

Las antiguas civilizaciones, como la egipcia, la china y la azteca, ya hacían estadísticas sobre el número de personas que vivían en las ciudades, normalmente para organizar el pago de impuestos y el ejército. En general a lo largo de toda la historia los gobiernos y dirigentes de las distintas naciones han procurado disponer de datos sobre la población con fines organizativos.

Una recopilación sistemática de datos sobre la población y la economía se inició en las ciudades-estados italianas de Venecia y Florencia durante el Renacimiento. El término estadística, derivado de la palabra estado, se utilizó para referirse a una colección de hechos de interés para el estado. La idea de recopilar datos se extendió desde Italia a otros países de Europa occidental. De hecho, en la primera mitad del siglo XVI, era común que los gobiernos europeos exigieran a las parroquias que registraran nacimientos, matrimonios y muertes. Debido a las malas condiciones de salud pública, esta última estadística fue de particular interés.

La alta tasa de mortalidad en Europa antes del siglo XIX se debía principalmente a enfermedades epidémicas, guerras y hambrunas. Entre las epidemias, las peores fueron las plagas. A partir de la Peste Negra en 1348, las plagas se repitieron frecuentemente durante casi 400 años. En 1562, como una forma de alertar a la corte del rey para que considere trasladarse al campo, la ciudad de Londres comenzó a publicar noticias semanales de mortalidad. Inicialmente, estos decretos de mortalidad enumeraban los lugares de muerte y si una muerte había resultado de la peste. A partir de 1625, los decretos fueron ampliados para incluir todas las causas de muerte. En 1662 el comerciante inglés John Graunt publicó un libro titulado *Natural and Political Observations Made on the Bills of Mortality*.

Graunt utilizó las cuentas de mortalidad de Londres para estimar la población de la ciudad y utilizó esta estimación para proyectar una cifra para toda Inglaterra. En su libro señaló que estas cifras serían de interés para los gobernantes del país, como indicadores tanto del número de hombres que podrían ser considerados para servir en el ejército y el número de hombres que podían pagar impuestos. Graunt también utilizó los cálculos de la mortalidad de Londres, y algunas suposiciones inteligentes sobre qué enfermedades mataron a quién y a qué edad, para inferir las edades de la muerte. (Recuerde que en las cuentas de mortalidad solo se enumeran las causas y los lugares de muerte, no las edades de los que mueren.) Graunt luego utilizó esta información para calcular las tablas que dan la proporción de la población que muere a diferentes edades. Las estimaciones de Graunt de las edades a las que la gente estaba muriendo eran de gran interés para aquellos en el negocio de la venta de anualidades. Las anualidades son lo opuesto a los seguros de vida, en que se paga una suma fija como inversión y luego se reciben pagos regulares durante toda la vida. El trabajo de Graunt sobre las tablas de mortalidad inspiró el trabajo de Edmund Halley en 1693. Halley, el descubridor del cometa que lleva su nombre (y también el hombre que fue más responsable, tanto por su aliento como por su apoyo financiero, de la publicación de la famosa *Principia Mathematica* de Isaac Newton), utilizó tablas de la mortalidad para calcular las probabilidades de que una persona de cualquier edad viviera a cualquier otra edad particular. Halley fue influyente en convencer a los aseguradores de la época de que una prima de seguro de vida anual debería depender de la edad de la persona asegurada. Después de Graunt y Halley, la recopilación de datos aumentó constantemente durante el resto del siglo XVII y hasta el siglo XVIII. Por ejemplo, la ciudad de París comenzó a dictar certificados de mortalidad en 1667; y en 1730 se había convertido en práctica común en toda Europa registrar las edades en el momento de la muerte.

El término estadística, que fue utilizado hasta el siglo XVIII como abreviatura para la ciencia descriptiva de los estados, en el siglo XIX se identificó cada vez más con los números. En la década de 1830 el término fue considerado casi universalmente en Gran Bretaña y Francia como sinónimo de la ciencia numérica de la sociedad. Este cambio de significado fue causado por la gran disponibilidad de registros del censo y otras tabulaciones que comenzaron a ser sistemáticamente recogidas y publicadas por los gobiernos de Europa occidental y los Estados Unidos a partir de alrededor de 1800. A lo largo del siglo XIX, aunque la teoría de la probabilidad había sido desarrollada por matemáticos como Jacob Bernoulli, Karl Friedrich Gauss y Pierre Simon Laplace, su uso en el estudio de los hallazgos estadísticos era casi inexistente, ya que la mayoría de los estadísticos sociales de la época estaban contentos de dejar que los datos hablaran por sí

mismos.

En particular, en esa época los estadísticos no estaban interesados en extraer inferencias sobre los individuos, sino que se preocupaban por la sociedad en su conjunto. Por lo tanto, no estaban interesados en la toma de muestras, sino que trataban de obtener censos de toda la población. Como resultado, la inferencia probabilística de muestras a una población era casi desconocida en las estadísticas sociales del siglo XIX.

No fue hasta finales de los años 1800 que las estadísticas se interesaron en inferir conclusiones a partir de datos numéricos. El movimiento comenzó con el trabajo de Francis Galton sobre el análisis de características hereditarias a través del uso de lo que ahora llamaríamos análisis de regresión y correlación y obtuvo gran parte de su inspiración del trabajo de Karl Pearson. Pearson, quien desarrolló el test de bondad-de-ajuste chi-cuadrado, fue el primer director del laboratorio de Galton, fundado por Francis Galtón en 1904. Allí Pearson originó un programa de investigación destinado a desarrollar nuevos métodos de uso de estadísticas en inferencia. Su laboratorio invitó a estudiantes avanzados de la ciencia y la industria a aprender métodos estadísticos que luego podrían ser aplicados en sus campos. Uno de sus primeros investigadores visitantes fue W. S. Gosset, un químico por formación, que mostró su devoción a Pearson publicando sus propias obras bajo el nombre de Estudiante. (Una famosa historia tiene que Gosset tenía miedo de publicar bajo su propio nombre por temor a que sus empleadores, la cervecería Guinness, no estuviesen contentos de descubrir que uno de sus químicos estaba haciendo investigación en estadísticas.) Gosset es famoso por su desarrollo de la prueba  $t$ .

Dos de las áreas más importantes de las estadísticas aplicadas a principios del siglo XX fueron la biología de la población y la agricultura. Esto se debió al interés de Pearson y otros en su laboratorio y a los notables logros del científico inglés Ronald A. Fisher. La teoría de la inferencia desarrollada por estos pioneros, incluyendo, entre otros, el hijo de Karl Pearson, Egon y el estadístico matemático polaco Jerzy Neyman, fue lo suficientemente general como para abordar una amplia gama de problemas cuantitativos y prácticos. Como resultado, después de los primeros años de este siglo, un número cada vez mayor de personas en la ciencia, los negocios y el gobierno comenzaron a considerar las estadísticas como una herramienta capaz de proporcionar soluciones cuantitativas a los problemas científicos y prácticos.

Hoy en día las ideas de las estadísticas están en todas partes. Las estadísticas descriptivas aparecen en todos los periódicos y revistas. La inferencia estadística se ha vuelto indispensable para la salud pública y la investigación médica, para el marketing y el control de calidad, para la educación, la contabilidad, la economía, las previsiones meteorológicas, las votaciones y las encuestas, los deportes, los seguros, los juegos de azar, y para toda la investigación que afirma ser científica. De hecho, las estadísticas se han arraigado en nuestro patrimonio intelectual.

La Estadística como ciencia experimentó un gran avance gracias al desarrollo de la Matemáticas, en especial de la Teoría de la Probabilidad, cuyas bases no fueron establecidas hasta el siglo XVII por los matemáticos franceses Pierre de Fermat y Blaise Pascal. ¿Y qué tiene que ver la Probabilidad con la Estadística? ¡Mucho! Hemos dicho que la Estadística es un instrumento para el estudio de un fenómeno cuando no se conocen que leyes lo rigen, y si hay fenómenos que no están regidos por leyes (aunque esto no es del todo cierto como veremos más adelante) eso son los fenómenos aleatorios. Y la base matemática para el estudio estadístico de los fenómenos aleatorios la proporciona la Teoría de la Probabilidad.

# Capítulo 2

## Teoría de Conteo de conjuntos discretos

### 2.1. Teoría de Conjuntos

Para realizar un desarrollo ordenado de la teoría de probabilidad, se necesitan conceptos básicos de teoría de conjuntos.

#### 2.1.1. Noción intuitiva de conjunto

**Definición 2.1.1** *Un conjunto es la reunión en un todo de objetos bien definidos y diferenciables entre sí, que se llaman elementos del mismo. Si  $a$  es un elemento del conjunto  $A$  se denota con la relación de pertenencia  $a \in A$ . En caso contrario, si  $a$  no es un elemento de  $A$  se denota  $a \notin A$ .*

Esta definición tan simple produce, sin embargo, una paradoja, que fue discutida por primera vez por Bertrand Russell. Este afamado filósofo logicista, cuando estaba estudiando las clases de conjuntos posibles comentó,

*"Me parece que una clase a veces es, y a veces no es, un miembro de sí misma. La clase de las cucharitas de té, por ejemplo, no es otra cucharita de té, pero la clase de cosas que no son cucharitas de té es una de las cosas que no son cucharitas... [esto] me condujo a considerar las clases que no son miembros de sí mismas; y éstas, parecía, debían formar una clase. Me pregunté si esta clase es o no un miembro de sí misma. Si es un miembro de sí misma, debería poseer las propiedades que definen a dicha clase, que consisten en no ser miembros de sí mismas. Si no es un miembro de sí misma, no debe poseer la propiedad definitoria de la clase, y por tanto debe ser un miembro de sí misma. Así cada alternativa lleva a su opuesta y existe una contradicción*

Para poder enunciar en forma más simple esta contradicción se realizaron ejemplos mas sencillos como el que sigue a continuación, lo cual se conoce como la paradoja del barbero:

*En un lejano poblado de un antiguo emirato había un barbero llamado As-Samet diestro en afeitar cabezas y barbas, maestro en escamondar pies y en poner sanguijuelas. Un día el emir se dio cuenta de la falta de barberos en el emirato, y ordenó que los barberos sólo afeitaran a aquellas personas que no pudieran hacerlo por sí mismas. Cierta día el emir llamó a As-Samet para que lo afeitara y él le contó sus angustias:*

*—En mi pueblo soy el único barbero. Si me afeito, entonces puedo afeitarme por mí mismo, por lo tanto no debería de afeitarme el barbero de mi pueblo ¡que soy yo! Pero si por el contrario, no me afeito, entonces algún barbero me debe afeitar ¡pero yo soy el único barbero de allí!*

*El emir pensó que sus pensamientos eran tan profundos, que lo premió con la mano de la más virtuosa de sus hijas. Así, el barbero As-Samet vivió por siempre feliz.*

Esta paradoja involucra a los conjuntos complementarios, es difícil definirlos si no se tiene un marco de referencia, por lo cual en esta obra consideraremos conjuntos que son partes de uno dado  $\mathbb{U}$ , llamado conjunto universal o de referencia. En estadística es una restricción muy razonable porque la noción de población está siempre presente. Para más información sobre la teoría logicista ver Mosterín(2000).

### Ejemplo 2.1.1 Ejemplos de conjuntos.

- $\emptyset$  : el conjunto vacío, que carece de elementos. Se define como  $[x \in \mathbb{U}] \wedge [x \notin \mathbb{U}]$ , una contradicción. Todas las contradicciones resultan en el conjunto vacío.
- $\mathbb{N}$ : el conjunto de los números naturales.
- $\mathbb{Z}$ : el conjunto de los números enteros.
- $\mathbb{Q}$  : el conjunto de los números racionales.
- $\mathbb{R}$ : el conjunto de los números reales.
- $\mathbb{C}$ : el conjunto de los números complejos.

**Definición 2.1.2** Se puede definir un conjunto: **por extensión**, enumerando todos y cada uno de sus elementos, o **por comprensión**, diciendo cuál es la propiedad que los caracteriza.

Un conjunto se suele denotar encerrando entre llaves a sus elementos, si se define por extensión, o su propiedad característica, si se define por comprensión. Por ejemplo:

$$A := \{1, 2, 3, 4, 5\} = \{\text{los números naturales del 1 al 5}\} \quad B := \{p \in \mathbb{Z} \mid p \text{ es par}\}$$

El conjunto  $A$  es un conjunto finito, puede escribirse por extensión o comprensión, pero el conjunto  $B$  es un conjunto infinito, y no puede escribirse en forma completa por extensión.

### Definición 2.1.3 .

1. Se dice que  $A$  está contenido en  $B$  (también que  $A$  es un subconjunto de  $B$  o que  $A$  es una parte de  $B$ ), y se denota  $A \subset B$ , si todo elemento de  $A$  lo es también de  $B$ , es decir,  $a \in A \Rightarrow a \in B$ .
2. Dos conjuntos  $A$  y  $B$  se dicen iguales, y se denota  $A = B$ , si simultáneamente  $A \subset B$  y  $B \subset A$ ; esto equivale a decir que tienen los mismos elementos (o también la misma propiedad característica).
3. El conjunto formado por todos los subconjuntos de uno dado  $A$  se llama partes de  $A$ , y se denota  $\mathcal{P}(A)$ . Entonces, la relación  $B \subset A$  es equivalente a decir  $B \in \mathcal{P}(A)$ .

**Observación 2.1.1** Para cualquier conjunto  $A$  se verifica que

1.  $\emptyset \subset A$ ,  $A \subset A$  y  $A \subset \mathbb{U}$ ;
2.  $B \subset A$  es un subconjunto propio de  $A$  si  $B \neq \emptyset$  y  $B \neq A$ .

### Ejemplo 2.1.2 Partes de un conjunto.

1. Si  $A = \{a, b\}$  entonces  $\mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, A\}$ .
2. Si  $a \in A$  entonces  $\{a\} \subset \mathcal{P}(A)$ .

### 2.1.2. Operaciones entre conjuntos

**Definición 2.1.4** *Dados dos conjuntos  $A$  y  $B$ , se llama diferencia al conjunto*

$$A - B := \{a \in A \mid a \notin B\}.$$

*Asimismo, se llama diferencia simétrica entre  $A$  y  $B$  al conjunto*

$$A \triangle B := (A - B) \cup (B - A).$$

*Si  $A \in \mathcal{P}(\mathbb{U})$ , a la diferencia  $\mathbb{U} - A$  se le llama complemento de  $A$  respecto de  $\mathbb{U}$ , y se denota abreviadamente por  $A^c$  ( $\mathbb{U}$  se supone fijado de antemano).*

**Lema 2.1.1** *Si  $A$  y  $B$  son subconjuntos cualesquiera de  $\mathbb{U}$  se verifica:*

- i)  $\emptyset^c = \mathbb{U}$ ,
- ii)  $\mathbb{U}^c = \emptyset$ ,
- iii)  $(A^c)^c = A$ ,
- iv)  $A \subset B$  si y solamente si  $B^c \subset A^c$

**Definición 2.1.5** *Se llama unión de dos conjuntos  $A$  y  $B$  al conjunto formado por objetos que son elementos de  $A$  ó de  $B$ , es decir:*

$$A \cup B := \{x \mid (x \in A) \text{ ó } (x \in B)\}.$$

*Se llama intersección de dos conjuntos  $A$  y  $B$  al conjunto formado por objetos que son elementos de  $A$  y de  $B$ , es decir:*

$$A \cap B := \{x \mid (x \in A) \text{ y } (x \in B)\}.$$

Si  $A$  y  $B$  son subconjuntos de un cierto conjunto universal  $\mathbb{U}$ , entonces es fácil ver que  $A - B = A \cap B^c$ .

**Ejemplo 2.1.3** Sea  $A$  el conjunto de aquellas personas que tienen sobrinos y  $B$  la colección de personas que tienen hermanos. Entonces el conjunto  $A \cap B$  consta de aquellas personas que tienen hermanos y tienen sobrinos, mientras que  $A \cap B^c$  está constituido por aquellas personas que tienen sobrinos pero no tienen hermanos. Observe que cada persona es un elemento de alguno de los siguientes conjuntos  $A \cap B$ ,  $A \cap B^c$ ,  $A^c \cap B$  ó  $A^c \cap B^c$ .

### 2.1.3. Diagramas de Venn

Los conjuntos se suelen representar gráficamente mediante "diagramas de Venn", con una línea que encierra a sus elementos. Así, todas las operaciones entre conjuntos se pueden representar gráficamente con el fin de obtener una idea más intuitiva.

En la figura 2.1 vemos el conjunto resultante de unir dos conjuntos, y en la figura 2.2 se ve el conjunto diferencia  $A - B = A \cap B^c$ .

Es importante recordar que la llamadas operaciones booleanas (unión e intersección) verifican las siguientes propiedades, si  $A$  y  $B$  son subconjuntos de un cierto conjunto universal  $\mathbb{U}$ , :

**Lema 2.1.2** *Propiedades*

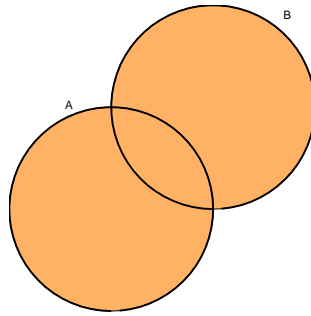


Figura 2.1: Diagrama de Venn que representa la unión de dos conjuntos

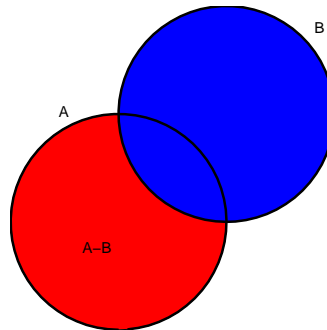


Figura 2.2: Diagrama de Venn que representa la diferencia A-B

1. **Idempotencia**  $A \cup A = A, A \cap A = A$ .
2. **Conmutatividad**  $A \cup B = B \cup A, A \cap B = B \cap A$ .
3. **Asociatividad**  $A \cup (B \cup C) = (A \cup B) \cup C, A \cap (B \cap C) = (A \cap B) \cap C$
4. **Absorción**  $A \cup (A \cap B) = A, A \cap (A \cup B) = A$
5. **Distributividad**  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
6. **Complementaridad**  $A \cup A^c = \mathbb{U}, A \cap A^c = \emptyset$

### Demostración:

Las dos primeras propiedades, la idempotencia y la conmutatividad resultan consecuencia directa de la definición de conjunto. La tercera propiedad, la asociatividad de la unión es mas delicada pero aun así es simple. Si  $x \in A \cup (B \cup C)$  quiero ver que  $x \in (A \cup B) \cup C$ . Sea entonces  $x \in A \cup (B \cup C)$ , por definición,  $x \in A$  o  $x \in B \cup C$ . Si ocurre esto último,  $x \in B$  o  $x \in C$ , por lo cual  $x \in A \cup B$  o  $x \in C$  y se da la inclusión. La otra inclusión sale en forma parecida.

La cuarta propiedad, la absorción es consecuencia directa de la distributividad, que probaremos primero. Si  $x \in A \cup (B \cap C)$ , entonces  $x \in A$  o  $x \in B \cap C$ . Si  $x \in A$ , entonces  $x \in A \cup B$  y  $x \in A \cup C$ , por lo cual está en la intersección de los dos. Si  $x \in B \cap C$ , entonces  $x \in B$ , por lo cual  $x \in A \cup B$ . También, si  $x \in C$ , resulta  $x \in A \cup C$ . Entonces  $x \in (A \cup B) \cap (A \cup C)$ . La otra inclusión se razona igual.

La complementaridad, la última propiedad se verifica observando que  $A \cup A^c$  es el conjunto de todos los puntos de  $\mathbb{U}$  que están en  $A$  o los que no están en  $A$ , por lo cual  $A \cup A^c = \mathbb{U}$ . También,  $A \cap A^c$  es el conjunto

de todos los puntos de  $\mathbb{U}$  que están en  $A$  y los que no están en  $A$ , lo cual es una contradicción que define al conjunto vacío, y  $A \cap A^c = \emptyset$ .

**Lema 2.1.3** *Además de éstas, se verifican también las siguientes propiedades:*

1.  $A \cup \emptyset = A$
2.  $A \cap \emptyset = \emptyset$  ( $\emptyset$  es el elemento nulo).
3.  $A \cup \mathbb{U} = \mathbb{U}$ ,  $A \cap \mathbb{U} = A$  ( $\mathbb{U}$  es el elemento universal).
4.  $(A \cup B)^c = A^c \cap B^c$ ,  $(A \cap B)^c = A^c \cup B^c$  (leyes de De Morgan).

### Demostración:

Veamos una de las leyes de De Morgan, del ítem 4, el resto de los ítems son simples derivaciones directas de las definiciones. Las leyes de De Morgan van a ser muy usadas en la parte de probabilidad, y es importante recordarlas. Pero un simple dibujo les servirá para recordar su enunciado.

Para ver que el complemento de la unión es la intersección de los complementos,  $(A \cup B)^c = A^c \cap B^c$  debemos ver las dos inclusiones

- i)  $(A \cup B)^c \subset A^c \cap B^c$
- ii)  $A^c \cap B^c \subset (A \cup B)^c$

Veamos entonces la primera inclusión  $(A \cup B)^c \subset A^c \cap B^c$ .

Si  $x \in (A \cup B)^c$  entonces  $x \notin A \cup B$ , por lo cual  $x \notin A$  y  $x \notin B$  y  $x \in A^c$  y  $x \in B^c$  por la definición de complemento, por lo cual  $x \in A^c \cap B^c$ .

Veamos entonces la segunda inclusión  $A^c \cap B^c \subset (A \cup B)^c$

Si  $x \in A^c \cap B^c$  entonces  $x \in A^c$  y  $x \in B^c$  por lo cual  $x \notin A$  y  $x \notin B$  lo cual implica que  $x \notin A \cup B$  y resulta  $x \in (A \cup B)^c$ .

Por último, los conceptos anteriores pueden generalizarse a familias de conjuntos, un grupo de conjuntos numerados que podría ser infinito. La definición formal es la siguiente:

**Definición 2.1.6** *Si para cada elemento  $i$  de un conjunto (de índices)  $I$ , finito o a lo sumo numerable, se tiene un conjunto  $A_i$ , entonces se define el conjunto  $\{A_i : i \in I\}$  y se lo denomina familia de conjuntos indexada por  $I$ . También se suele denotar por  $\{A_i\}_{i \in I}$ . De forma análoga se define una familia de elementos  $\{a_i\}_{i \in I}$ .*

**Definición 2.1.7** *Dada una familia de conjuntos  $\{A_i\}_{i \in I}$  se definen:*

- $\bigcup_{i \in I} A_i := \{a : \exists i \in I \text{ tal que } a \in A_i\}$
- $\bigcap_{i \in I} A_i := \{a : a \in A_i, \forall i \in I\}$

Las propiedades de la unión e intersección siguen siendo válidas para familias de conjuntos, y en particular las leyes de De Morgan:

$$\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c \quad \left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} A_i^c$$

**Definición 2.1.8** *Sean  $A_1, \dots, A_n$  subconjuntos de un conjunto universal  $\mathbb{U}$ . Decimos que  $A_1, \dots, A_n$  forman una partición de  $\mathbb{U}$  si*

$$a) \mathbb{U} = \bigcup_{i=1}^n A_i.$$

$$b) A_i \cap A_j = \emptyset, \forall i \neq j.$$

**Definición 2.1.9** Se llama cardinal de un conjunto  $A$  al número de elementos de dicho conjunto, y se lo denota como  $\#A$ .

**Observación 2.1.2** Un conjunto con infinitos elementos tiene cardinal infinito. Sin embargo, hay conjuntos que a pesar de tener infinitos elementos tienen menos elementos que otros conjuntos infinitos. Por ejemplo,  $\mathbb{N} \subset \mathbb{Q} \subset \mathbb{R}$ , y la inclusión es estricta, y todos son conjuntos infinitos. Por lo cual hay diversas clases de cardinales infinitos. En este curso haremos énfasis en los conjuntos finitos definidos por comprensión y sus cardinales, y aprenderemos técnicas para calcular dichos números sin listar los elementos.

**Ejemplo 2.1.4** Sea  $\mathbb{U}$  un conjunto finito. Probar que

1. Si  $A \subset B \subset \mathbb{U}$ ,  $\#(B - A) = \#B - \#A$ .
2. Si  $A \subset \mathbb{U}$ ,  $\#A^c = \#\mathbb{U} - \#A$ .
3. Si  $A_1, \dots, A_n$  forman una partición del conjunto  $\mathbb{U}$ , entonces  $\#\mathbb{U} = \#A_1 + \dots + \#A_n$ .
4. Si  $A$  y  $B$  son conjuntos de  $\mathbb{U}$  entonces  $\#(A \cup B) = \#A + \#B - \#(A \cap B)$ .

### Resolución

1) Si  $A$  es parte del conjunto  $B$ , entonces los elementos de  $B - A$  son todos los elementos de  $B$  que no están en  $A$  y  $\#B = \#(B - A) + \#A$ , por lo cual, despejando resulta  $\#(B - A) = \#B - \#A$ .

2) Si  $\mathbb{U}$  es un conjunto finito, entonces  $\mathbb{U} = A \cup A^c$  en forma disjunta, lo puedo partir en dos montones diferentes, las cosas que están en  $A$  y las que no están en  $A$  y por supuesto, la cantidad de elementos total será la suma de los elementos de cada montón que armé,  $\#\mathbb{U} = \#A + \#A^c$ . Una vez más despejando resulta  $\#A^c = \#\mathbb{U} - \#A$ .

3) Lo que describimos anteriormente pasa para una colección de conjuntos disjuntos cuya unión genera el espacio. Esta colección se llama partición. Entonces queda claro que la cantidad de elementos de  $\mathbb{U}$  es la suma de los elementos en cada montón, esto es,  $\#\mathbb{U} = \#A_1 + \dots + \#A_n$ .

4) Observemos que  $(A \cup B) = A \cup (B - A)$  en forma disjunta por lo cual el número de elementos de  $(A \cup B)$  es el número de elementos de  $A$  más el número de elementos de  $B - A$ . Como  $B - A = B - (A \cap B)$ , el número de elementos de  $B - A$  es el número de elementos de  $B$  menos el número de elementos de  $A \cap B$ . Por lo cual

$$\#(A \cup B) = \#A + \#(B - A) = \#A + \#B - \#(A \cap B)$$

**Definición 2.1.10** Se dice que un conjunto  $U$  es discreto si  $\#U$  es finito o  $U$  es un conjunto infinito pero numerable.

## 2.2. Conjuntos finitos: Combinatoria

La Combinatoria es la parte de las Matemáticas que se dedica al estudio de los conjuntos finitos. Puesto que la propiedad principal de estos conjuntos es que se puede representar su número de elementos mediante un número natural (llamado cardinal de dicho conjunto), la tarea básica de la Combinatoria es precisamente el cálculo del cardinal de dichos conjuntos.



**Ejemplo 2.2.1** Supongamos que el conjunto  $A$  está formado por todos los números naturales de tres cifras que pueden formarse con los dígitos 1 y 5. Listemos el conjunto y veamos cuántos hay.

**Resolución**

Un diagrama de árbol ayuda a visualizar el conteo de la cantidad de casos posibles, el diagrama correspondiente a este ejemplo puede verse en la figura 2.3.

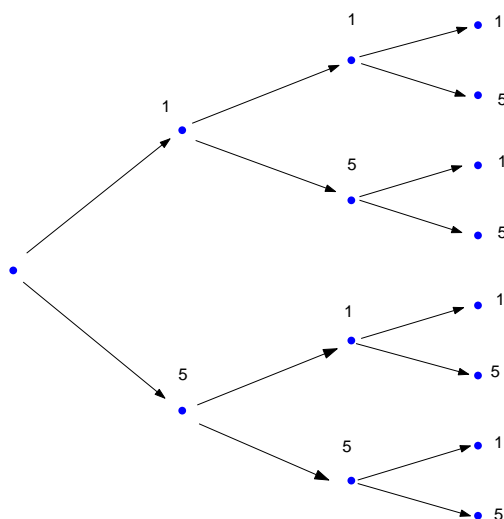


Figura 2.3: Diagrama de árbol correspondiente al ejemplo A

La lista de todos los números posibles se obtiene siguiendo las ramas del árbol desde la raíz hasta llegar a una hoja. La cantidad total de hojas es la cantidad total de números que podemos formar, o cardinal del conjunto. En este caso, el cardinal es 8.

**Observación 2.2.1** ¿Podemos contar cuantos hay sin tener que listar todos los casos? Si un número de tres cifras genérico se escribe como ABC, entonces tenemos dos posibilidades para la A, por cada una de ellas tenemos dos para la B, y por cada una de las AB tenemos dos para la C. Por lo cual el número total de resultados es  $2 \times 2 \times 2 = 2^3$ .

Esta mecánica basada en el sentido común nos provee de una regla útil para contar. En nuestro caso, en cada paso la división en el árbol era la misma, teníamos dos posibilidades por cada dígito, pero en general, en cada paso puede haber más o menos posibilidades nuevas. El resultado final va a ser el producto de los casos posibles en cada paso.

**Principio general de enumeración**

**Lema 2.2.1** Si una experiencia  $E_1$  arroja  $n_1$  resultados posibles y por cada resultado de  $E_1$  se realiza una experiencia  $E_2$ , que puede arrojar  $n_2$  resultados posibles, entonces la realización **en sucesión** de  $E_1$  y  $E_2$  arroja un número total de  $n_1 \cdot n_2$  resultados posibles. Esto puede generalizarse a la realización de  $E_1, \dots, E_k$  experiencias en sucesión, que van a generar un número total de

$$n_1 \times n_2 \times \cdots \times n_k$$

resultados posibles.

**Ejemplo 2.2.2** ¿Cuántas parejas de baile pueden formarse en un salón donde se encuentran 20 personas de un género y 17 del otro?

**Resolución**

Las parejas de baile se forman eligiendo un varón y una mujer. Tengo 20 elecciones para el varón y por cada elección hecha tengo 17 para la mujer, en total  $20 \times 17 = 340$ .

**Ejemplo 2.2.3** Un investigador quiere determinar el efecto de tres variables, presión, temperatura y el tipo de catalizador, en la producción de un proceso de refinación. Si el investigador tiene la intención de utilizar tres temperaturas, tres presiones y dos tipos de catalizadores, ¿cuántos experimentos habría que hacer si quisiera incluir todas las posibles combinaciones de presión, temperatura y tipos de catalizador?

**Resolución**

Tenemos tres temperaturas, tres niveles de presión y dos catalizadores, serían  $3 \times 3 \times 2 = 18$  experimentos.

**Ejemplo 2.2.4** Tincho es un jugador de poker profesional. En este momento, él necesita sacar dos diamantes en forma sucesiva. Se encuentra sentado en la mesa mirando su mano de cartas y las cartas que están boca arriba en la mesa. Tincho ve 11 cartas. De esas cartas 4 son diamantes. El mazo completo posee 14 diamantes entre 52 cartas, por lo cual 9 de las 41 cartas que todavía están en el mazo son diamantes. La cantidad de formas de sacar dos diamantes en sucesión del mazo sería  $9 \times 8 = 72$  formas distintas pues para el segundo diamante tengo menos diamantes en el mazo. ¿Cuál sería la cantidad de formas que puede Tincho sacar dos cartas cualesquiera de ese mazo? Tendría  $41 \times 40 = 1640$  formas distintas pues quedan  $52 - 11 = 41$  cartas no vistas en la mesa para sacar la primera, y luego quedan 40 de donde sacar la segunda. ¿Que le parece, tiene alguna chance de sacar dos diamantes en sucesión Tincho o no?

**Ejemplo 2.2.5** ¿Cuántos números de cinco cifras **pares distintos** pueden formarse con los números 1,2,3,4,5,7 y 9, si no quiero que las cifras de los números se repitan?.

**Resolución**

Si pienso los números 1,2,3,4,5,7 y 9 como objetos de una bolsa que debo poner en una fila, es evidente que no los puedo repetir, si puse el siete en la fila no puede haber más setes después porque en la bolsa no hay más setes para poner. También, me piden que cuenten todos los números pares que puedo formar. Los números pares terminan en un dígito par, por lo cual estamos buscando números de 5 cifras con dígitos distintos (de la lista dada) que terminen en cuatro o en 2.

Escribiendo un número genérico como ABCD2, tenemos 6 dígitos posibles (de la bolsa) para la A, 5 para la B, 4 para la C y 3 para la D, lo cual da un total de  $6 \cdot 5 \cdot 4 \cdot 3 = 360$  posibilidades. El otro caso ABCD4 se razona de la misma forma, dando 360 posibles resultados también, por lo cual la cantidad total de números posibles es  $360 + 360 = 720$ .

Este ejemplo nos cuenta que si podemos particionar nuestro conjunto de resultados en dos subconjuntos de resultados distintos, la cantidad de resultados es la suma de la cantidad de resultados de cada subconjunto. Esto se puede generalizar a una partición de varios casos distintos más simples.

**Principio de adición**

**Lema 2.2.2** *Supongamos que nuestro experimento consiste en realizar una experiencia  $E_1$  que arroja  $n_1$  resultados posibles, o una experiencia  $E_2$  que puede arrojar  $n_2$  resultados posibles, entonces la realización excluyente de  $E_1$  o  $E_2$  arroja un número total de  $n_1 + n_2$  resultados posibles. Esto puede generalizarse a la realización de  $E_1, \dots, E_k$  experiencias excluyentes, que van a generar un número total de*

$$n_1 + n_2 \cdots + n_k$$

*resultados posibles.*

**Ejemplo 2.2.6** ¿Cuántas formas hay de coser cinco botones rojos y cinco azules a una camisa si los colores deben estar alternados?

**Resolución**

Primero hay que preguntar si los botones son distintos. Usualmente son de una misma clase todos los de color rojo y de la misma clase todos los de color azul. Pero en estos casos se suele poner la palabra indistinguibles para estar seguros. La camisa empieza con el primer botón del cuello y luego continúa la fila de botones que deben ser colocados en forma alternada. El primer botón entonces puede ser rojo o azul, dos eventos excluyentes, por la cual la cantidad total de resultados será la cantidad de filas que puedo hacer empezando con un botón rojo más la cantidad de filas que puedo hacer empezando con un botón azul. Lo interesante es que solo puedo hacer una fila con botones alternados en cada caso, por lo cual tengo solo dos formas posibles de coser los botones.

Existe notación especial para escribir las cantidades que resultan de estos ejemplos, los números combinatorios y factoriales. En las siguientes secciones estudiaremos las variaciones, permutaciones y combinaciones de elementos de un conjunto, cuántas de ellas se pueden formar, y como usar esta información para resolver preguntas sobre cardinales de conjuntos.

### 2.2.1. Variaciones

**Ejemplo 2.2.7** Consideremos el experimento que consiste en registrar el cumpleaños de 20 personas elegidas al azar. Si ignoramos años bisiestos y suponemos que hay 365 posibles cumpleaños, ¿Cuántos resultados posibles tendría este experimento?

**Resolución**

La respuesta es  $365^{20} = 365 \times 365 \times 365 \cdots \times 365$ , pues por cada persona de las 20 tengo 365 posibilidades, y por el principio general de enumeración estas cantidades se multiplican.

**Definición 2.2.1** *Supongamos tener  $n$  elementos diferentes. Llamamos variación con repetición (o muestreo con repetición) a una lista ordenada de  $r$  elementos tomados del conjunto total considerando repeticiones.*

**Lema 2.2.3** *El número total de variaciones de  $n$  en  $r$  con repetición es*

$$VR_r^n = n^r$$

*y se demuestra en forma parecida a la resolución del ejemplo de los cumpleaños.*

**Definición 2.2.2** . *Números factoriales.*

Se define  $n!$  mediante la ley de recurrencia

$$n! = n \times (n-1)!$$

y la condición inicial  $0! := 1$ . De forma iterativa, se tiene

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1$$

**Ejemplo 2.2.8** ¿Cuántas posibles listas habría si las 20 personas cumplen años en días distintos?

**Resolución**

Para la primera persona tenemos 365 posibilidades, para la segunda, tenemos 364 posibilidades, para la tercera 363, y asísiguiendo, la número 20 va a tener  $365 - 19 = 346$ . Por lo cual el número de listas posibles es

$$365 \times 364 \times 363 \times \cdots \times 346 = \frac{365!}{345!}$$

**Definición 2.2.3** Supongamos tener  $n$  elementos diferentes. Llamamos *variación* (o *muestreo sin repetición*) a una lista ordenada de  $r$  elementos tomados del conjunto total sin considerar repeticiones.

**Lema 2.2.4** El número total de variaciones de  $n$  en  $r$  sin repetición es

$$V_r^n = \frac{n!}{(n-r)!}$$

y la demostración es similar a la resolución del ejemplo anterior.

## 2.2.2. Permutaciones

**Corolario 2.2.1**  $P_n = n!$  es el número de permutaciones de  $n$  elementos, es decir, es el número total de formas de realizar listas ordenadas con  $n$  elementos.

**Demostración:** La cantidad total de permutaciones de  $n$  elementos es una variación sin repetición de  $n$  en  $n$ , por lo cual  $P_n = V_n^n = n!$ .

**Ejemplo 2.2.9** ¿De cuántas formas pueden ordenarse en hilera 5 cubos de colores diferentes?.

**Resolución**

Si una hilera de cubos se escribe ABCDE, entonces tenemos cinco posibilidades para la A, cuatro para la B, tres para la C, dos para la D y una para E. Por lo cual el número total de resultados es  $5 \times 4 \times 3 \times 2 \times 1$ . La clave de este resultado es el hecho de que no puedo repetir elementos, solo estoy **permutando** de lugar los 5 elementos que tengo.

**Ejemplo 2.2.10** ¿De cuántas maneras pueden acomodarse 7 personas?

1. en una fila de siete sillas
2. alrededor de una mesa redonda

**Resolución**

Una fila de siete sillas puede representarse por una sucesión de siete casilleros

— — — — —

por lo cual la forma de llenarlos es una permutación  $7! = 5040$ . Si e vez de una hilera de sillas (que tiene principio y fin) tuviera una mesa redonda, debería fijar una persona y permutar las restantes alrededor de ella para obtener todas las formas distintas de sentarse, por lo cual al respuesta es  $6! = 720$ .

**Ejemplo 2.2.11** Supongamos que la reunión anual de tomadores de mate amargo tiene confirmada la asistencia de 10 delegaciones de 3 miembros cada una. Los mismos van a sentarse en una gran mesa circular para debatir el memorándum de la reunión. Si los miembros de cada delegación no se separan, de cuantas formas pueden ubicarse alrededor de la mesa?.

**Resolución**

Si ubico primero las delegaciones, tengo un total de  $9!$  formas diferentes, al fijar una delegación para evitar las permutaciones circulares que no cambian la ubicación relativa. Luego, por cada configuración de las delegaciones tengo  $3!$  permutaciones de los miembros de cada delegación. Esto da un total de

$$(3!)^{10} 9!$$

**Ejemplo 2.2.12** ¿Cuántos números de 5 dígitos capicúas pueden formarse con los números 1,2,3,4,5,6,7,8?

**Resolución**

Un número capicúa de cinco dígitos es de la forma

$$xyzyx$$

Por lo cual el problema se reduce a ver cuántos números de tres dígitos pueden formarse con aquellos dígitos. Exac-  
tamente  $8^3$ .

**Ejemplo 2.2.13** ¿Cuántas permutaciones pueden formarse con la palabra Micaela?

**Resolución**

Si escribo

$$M i c a_1 e l a_2$$

todas las letras son distintas, luego hay  $7!$  permutaciones, pero cada par de permutaciones del tipo

$$\begin{array}{c} - - - a_1 - - a_2 \\ - - - a_2 - - a_1 \end{array}$$

coinciden, por lo tanto tengo que dividir por 2 el número total de permutaciones. El resultado es

$$\frac{7!}{2}$$

**Ejemplo 2.2.14** ¿Cuántas palabras pueden formarse con las letras de la palabra

*ARQUITECTURA*

si pedimos que la letra Q esté seguida de la letra U, y esta, a la vez, seguida solamente de la E o de la I?

### Resolución

Escribiendo la palabra anterior así

$a_1 r_1 q u_1 i t_1 e c t_2 u r_2 a_2$

tengo 12 letras, pero si tengo en cuenta las restricciones tengo dos casos

$a_1 r_1 (qui) t_1 e c t_2 u r_2 a_2$

$a_1 r_1 (que) t_1 i c t_2 u r_2 a_2$

cada uno con un número total de  $10!$  permutaciones, pues estoy fijando la Q con una de las U's, y la siguiente letra puede ser una I o una E. Por lo cual hasta el momento el número de permutaciones es

$$2 \times 10!$$

Pero permutando las  $a_i$ , o las  $r_i$ , o las  $t_i$ , sin mover las otras letras, obtengo palabras iguales. No es el caso de las  $u$ 's, pues al estar una fija al lado de la  $q$ , formando un bloque, tengo solo una  $u$  libre para mover.

Como hay  $2!$  permutaciones de  $a_i$ ,  $2!$  de  $r_i$ , y  $2!$  de  $t_i$ , queda

$$\frac{2 \times 10!}{2!2!2!}$$

### 2.2.3. Combinaciones

**Ejemplo 2.2.15** ¿De cuántas formas puedo armar un grupo de 3 personas de un conjunto de 20 personas?

### Resolución

El número de subconjuntos posibles es el número de listas ordenadas de 3 personas elegidas de las 20 (sin reposición) dividido el número de permutaciones de tres elementos.

$$\frac{20!}{3!17!}$$

**Definición 2.2.4** Supongamos tener  $n$  elementos diferentes. Llamamos combinación (o muestreo sin orden) a un subconjunto de  $r$  elementos tomados del conjunto total. La gran diferencia entre las variaciones y las combinaciones radica en que estas últimas no se consideran orden ni repeticiones posibles.

**Lema 2.2.5** El número total de combinaciones de  $n$  en  $r$  es

$$C_r^n = \frac{n!}{r!(n-r)!} = \binom{n}{r}$$

denotado por el número combinatorio.

**Lema 2.2.6** El número de formas en que  $n$  objetos pueden ser agrupados en  $r$  clases con  $n_j$  en el  $j$ -ésima clase,  $j : 1, \dots, r$  y  $\sum_{j=1}^r n_j = n$  es

$$\binom{n}{n_1!n_2!\cdots n_r!} = \frac{n!}{n_1!n_2!\cdots n_r!}$$

**Proposición 2.2.1** Los coeficientes binomiales cumplen las siguientes propiedades:

a)  $\binom{n}{k} = \binom{n}{n-k}$

b)  $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$

c) Binomio de Newton

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

### Cantidad de subconjuntos de un conjunto finito

Como aplicación de los números combinatorios y del Binomio de Newton, podemos contar el número total de subconjuntos que tiene un conjunto  $A$  con  $n$  elementos, es decir, el cardinal de partes de  $A$ .

**Lema 2.2.7** Si  $A$  tiene  $n$  elementos, entonces  $\#\mathcal{P}(A) = 2^n$

#### Demostración:

Para ello, notemos primero que si particiono un conjunto en union de conjuntos disjuntos, entonces el cardinal de la unión es la suma de los cardinales de los conjuntos que forman la partición.

Luego observemos que el conjunto  $\mathcal{P}(A)$  se puede escribir como la unión disjunta de los subconjuntos de cero elementos, los subconjuntos de exactamente 1 elemento, los de exactamente dos elementos, los de exactamente tres, hasta los de exactamente  $n$  elementos

$$\mathcal{P}(A) = \cup_{k=0}^n \{\text{subconjuntos de exactamente } k \text{ elementos}\}$$

Por lo cual el cardinal de  $\mathcal{P}(A)$  es

$$\begin{aligned} \#\mathcal{P}(A) &= \sum_{k=0}^n \#\{\text{subconjuntos de exactamente } k \text{ elementos}\} \\ &= \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} \end{aligned}$$

Pero esta cantidad corresponde a desarrollar mediante el binomio de Newton la expresión

$$(1+1)^n = 2^n$$

Así pues se obtiene que  $\#\mathcal{P}(A) = 2^n$  si  $\#A = n$ .

### Principio de Exclusión-Inclusión

Sea  $\mathbb{U}$  un conjunto finito y sea  $A$  un subconjunto de  $\mathbb{U}$ . Sean  $A_1, \dots, A_n$  subconjuntos de  $\mathbb{U}$  que cumplen ciertas propiedades  $p_1, \dots, p_n$ . El principio en cuestión da una fórmula para el número de elementos de  $\mathbb{U}$  que **no** satisfacen ninguna de las propiedades  $p_1, \dots, p_n$ .

**Ejemplo 2.2.16** Supongamos que queremos contar la cantidad de formas en que podemos repartir 7 facturas en 3 platos sin dejar ninguno vacío. Vamos a definir como conjuntos  $A_1, A_2$  y  $A_3$  a los eventos “el primer plato está vacío”, “el segundo plato está vacío”, “el tercer plato está vacío”, respectivamente. El evento  $A$  “ningún plato está vacío” se escribe como el complemento de la unión de  $A_1, A_2$  y  $A_3$

$$A = (A_1 \cup A_2 \cup A_3)^c$$

Por lo cual

$$\#A = \#(\text{ninguno vacío}) = \#(\text{todas las formas de repartir}) - \#(\text{alguno vacío}) = \#\mathbb{U} - \#(A_1 \cup A_2 \cup A_3)$$

por lo cual debemos encontrar alguna forma de calcular

$$\#(A_1 \cup A_2 \cup A_3)$$

**Lema 2.2.8** *El cardinal de  $A_1 \cup A_2 \cup A_3$  es*

$$\#(A_1 \cup A_2 \cup A_3) = \#A_1 + \#A_2 + \#A_3 - \#(A_1 \cap A_2) - \#(A_1 \cap A_3) - \#(A_2 \cap A_3) + \#(A_1 \cap A_2 \cap A_3)$$

**Demostración:**

Supongamos primero que  $n = 2$ . Quiero probar que

$$\#(A \cup B) = \#A + \#B - \#(A \cap B)$$

Observemos que  $A \cup B = A \cup (B \cap A^c)$  y estos últimos conjuntos son disjuntos. Entonces

$$\#(A \cup B) = \#A + \#(B \cap A^c)$$

Por otra parte  $B = (B \cap A) \cup (B \cap A^c)$ , unión disjunta también, por lo cual

$$\#B = \#(A \cap B) + \#(B \cap A^c)$$

y resulta

$$\#(B \cap A^c) = \#B - \#(A \cap B)$$

Reemplazando obtenemos

$$\#(A \cup B) = \#A + \#(B \cap A^c) = \#A + \#B - \#(A \cap B)$$

Si tenemos 3 conjuntos  $A, B, C$ ,

$$\begin{aligned} \#(A \cup B \cup C) &= \#[A \cup (B \cup C)] \\ &= \#A + \#(B \cup C) - \#(A \cap (B \cup C)) \\ &= \#A + \#B + \#C - \#(B \cap C) - \#[(A \cap B) \cup (A \cap C)] \\ &= \#A + \#B + \#C - \#(B \cap C) - [\#(A \cap B) + \#(A \cap C) - \#(A \cap B \cap C)] \\ &= \#A + \#B + \#C - \#(B \cap C) - \#(A \cap B) - \#(A \cap C) + \#(A \cap B \cap C) \end{aligned}$$

Usando esta fórmula en el ejemplo anterior, resulta

$$\begin{aligned} \#(A_1 \cup A_2 \cup A_3) &= \#A_1 + \#A_2 + \#A_3 - \#(A_1 \cap A_2) - \#(A_1 \cap A_3) - \#(A_2 \cap A_3) + \#(A_1 \cap A_2 \cap A_3) \\ &= \binom{7+2-1}{7} + \binom{7+2-1}{7} + \binom{7+2-1}{7} - 1 - 1 - 1 + 0 \end{aligned}$$



**Ejemplo 2.2.17** Determinar el número de permutaciones de  $\{1, 2, 3\}$  que no fijan ningún punto. O sea, si  $f$  es una permutación de  $\{1, 2, 3\}$  entonces  $f(i) \neq i$  cualquiera sea  $i \in \{1, 2, 3\}$ . Por ejemplo,  $f(1) = 2, f(2) = 3, f(3) = 1$  y  $g(1) = 3, f(2) = 1, f(3) = 2$  son algunas de tales permutaciones. Defino  $A_i$  el conjunto de las permutaciones que fijan el valor  $i$ , con  $i = 1 : 3$ . Es claro que  $A_1 \cup A_2 \cup A_3$  es el conjunto de todas las permutaciones que fijan **algún** elemento. Nos interesa calcular el número de elementos del complemento de ese conjunto. Se tiene entonces

$$\begin{aligned}\#(A_1 \cup A_2 \cup A_3) &= \#A_1 + \#A_2 + \#A_3 - \#(A_1 \cap A_2) - \#(A_1 \cap A_3) - \#(A_2 \cap A_3) + \#(A_1 \cap A_2 \cap A_3) \\ &= 2 + 2 + 2 - 1 - 1 - 1 + 1 = 4\end{aligned}$$

Como  $\#U$  es  $3!$ , hay solo  $6-4=2$  permutaciones que no fijan ningún elemento, y son las dos que hemos dado como ejemplo.

## 2.2.4. Bolas en celdas

### Bolas indistinguibles

Supongamos tener un grupo de bolas rojas indistinguibles y celdas donde distribuirlas. Para esquematizar tal distribución denotamos la cajas con barras  $|$  y con puntos a la izquierda los objetos dentro de esa caja, salvo la última que no es necesario describirla.

$$\cdot | | | \dots$$

Por ejemplo, la configuración anterior describe la distribución de 8 bolas en 6 cajas de la siguiente forma, la primera caja tiene un elemento, las siguientes dos cajas están vacías, la cuarta tiene dos elementos, la quinta esta vacía y la sexta tiene cinco elementos.

En este tipo de representación aparecen 5 barras y 8 puntos, un punto por cada bola, pero una barra menos que el numero de celdas. Por lo cual el número de configuraciones posibles es el número de permutaciones de todos los elementos, puntos y barras, dividido por el número de permutaciones de las barras (porque son indistinguibles) y por el número de permutaciones de los puntos que también son indistinguibles. Este número es

$$\frac{(6-1+8)!}{8!(6-1)!} = \binom{13}{8} = \binom{13}{5}$$

En general, si llamamos  $n$  al número de celdas y  $k$  al de bolas, el número de formas que podemos poner  $k$  bolas indistinguibles en  $n$  celdas es

$$\frac{(n-1+k)!}{k!(n-1)!} = \binom{n-1+k}{k} = \binom{n-1+k}{n-1}$$

¿Por qué es interesante saber este esquema de bolas en celdas? Pues permite contar configuraciones de forma fácil, sin tener que listarlas a todas. Veamos los siguientes ejemplos antes de pasar al siguiente caso, donde las bolas son distinguibles.

**Ejemplo 2.2.18** 1. ¿De cuantas maneras pueden disponerse 7 figuras de gel indistinguibles en 10 ventanas?

2. ¿Cuantas maneras hay si no se permite colocar más de una figura por ventana?

**Resolución**

Si las figuras de gel son las bolas y las ventanas las celdas, tenemos que contar las configuraciones posibles de 7 bolas en 10 celdas, que es

$$\binom{10-1+7}{7} = \binom{16}{7} = 11440$$

En el caso de que solo se permita una figura por ventana, son todas las formas de elegir 7 ventanas entre 10, no distinguiendo que pongo en cada ventana.

$$\binom{10}{7} = 120$$

**Ejemplo 2.2.19** ¿De cuantas formas se pueden ubicar 3 loros y 20 palomas en 5 jaulas numeradas con la condición de que cada jaula debe contener a lo sumo un loro y por lo menos tres palomas?

### Resolución

En este tipo de ejercicios debemos primero ordenar lo que sabemos. Hay 5 jaulas, 3 loros y 20 palomas, y nos dan restricciones de como poner los animales en las jaulas. Si hay por lo menos 3 palomas por cada una de las 5 jaulas, en realidad nos quedan 5 palomas para repartir, las otras 15 están colocadas tres por jaula. Si no podemos poner más de un loro por jaula, entonces debemos elegir en que jaulas van a ir los loros primero y después colocar las 5 palomas restantes.

Por lo cual tenemos

$$\binom{5}{3} = 10$$

formas de elegir las jaulas donde va a ir un loro en cada una (ya pusimos tres palomas en cada una de ellas), y por cada una de esas elecciones, las formas de colocar las 5 palomas restantes son

$$\binom{5-1+5}{5} = \binom{9}{5} = 126$$

Por lo cual el resultado final es  $10 \times 126 = 1260$ .

**Ejemplo 2.2.20** ¿De cuantas formas se pueden ordenar en una fila diez botones rojos idénticos y ocho azules idénticos si no se quiere que dos de los azules estén juntos?

### Resolución

Este es un ejercicio de bolas en celdas, que necesita ingenio para decidir cuantas celdas y cuantas bolas son las que tenemos. Razonemos de la siguiente forma, para que no se junten dos botones azules, tenemos que colocar botones azules y rojos alternados. Hay 8 botones azules, necesitamos 7 botones rojos para separar los azules, por lo cual nos quedan solo 3 botones rojos para colocar libremente. Esas son nuestras bolas. ¿Cuantas celdas hay?, son los lugares a la derecha e izquierda de los hermanitos Azul-Rojo que ya están fijos. Miremos el diagrama que nos queda

$$-AR-AR-AR-AR-AR-AR-AR-A-$$

Tenemos 9 lugares donde poner los 3 botones rojos restantes, esas son nuestras celdas. Entonces, las formas de colocar los botones con la restricción pedida es

$$\binom{9-1+3}{3} = \binom{11}{4} = 165$$

**Bolas distinguibles**

¿Que pasa cuando las bolas son distinguibles? El problema corresponde al experimento que consta de tirar bolas de distinto color en cajas numeradas y anotar **que bolas** caen en cada caja (en vez de **cuantas bolas** caen en cada caja, como en el caso anterior). Para contar cuantas configuraciones tenemos, debemos distinguir dos casos, existe orden dentro de cada caja o no.

Estudiemos el siguiente ejemplo. Supongamos tener 3 bolas de diferentes colores  $R, A, V$  (roja, amarilla y verde) y cuatro celdas. Tengo dos posibilidades

- a) Me interesa distinguir si la bola amarilla y la verde (AV) cayeron en la segunda celda o si la roja y la verde (RV) lo hicieron. Es decir, las configuraciones

$$R|AV| \neq A|RV|$$

son distintas pero

$$R|AV| = R|VA|$$

no distingo orden dentro de cada celda.

- b) Me interesa distinguir si la bola amarilla y la verde (AV) cayeron en la segunda celda o si la roja y la verde (RV) lo hicieron, **y en qué orden**. Es decir, las configuraciones

$$R|AV| \neq R|VA| \neq A|RV| \neq A|VR|$$

Es la forma de contar anagramas especiales, donde el orden de las letras es importante.

Entonces, en el primer caso, cuando no distingo orden dentro de la celda, sino solo qué bola cayó en ella, tengo  $4 * 4 * 4 = 4^3$  posibles configuraciones, pues tengo 4 celdas posibles para la primera bola, 4 para la segunda, y 4 para la tercera bola y por el principio de enumeración, el total es  $4^3 = 64$ .

En el segundo caso, cuando distingo el orden en el que cayeron en la celda, uso el esquema de puntos (bolas) y barras (celdas menos una), es decir tengo  $4 - 1 + 3$  objetos **ordenados** para permutar, pero divido solo por las permutaciones de las barras, pues mantengo el orden de las bolas dentro de cada celda. Este número es

$$\frac{(4 - 1 + 3)!}{(4 - 1)!} = \frac{6!}{3!} = 120$$

**Ejemplo 2.2.21** ¿De cuántas maneras pueden ubicarse 5 juguetes distintos en tres cajas diferentes?

**Resolución**

Si consideramos a los juguetes como bolas y a las cajas como celdas, todos distinguibles pero sin orden dentro de las cajas, tendremos, 3 cajas posibles para el primer juguete, 3 para el segundo, 3 para el tercero, 3 para el cuarto y 3 para el quinto. Por lo cual el resultado final es  $3^5 = 243$ .

**Ejemplo 2.2.22** ¿De cuantas maneras pueden ubicarse 5 libros distintos en tres estantes diferentes? Ahora estamos interesados en el orden de los libros en cada estante.

**Resolución**

Como nos interesa el orden dentro de cada estante, tenemos un esquema de 5 bolas y 3 celdas, con orden dentro de cada celda. El número de formas distinta será

$$\frac{(3 - 1 + 5)!}{(3 - 1)!} = \frac{7!}{2!} = 2520$$

**Ejemplo 2.2.23** ¿Cuántas palabras distintas (anagramas) pueden formarse con las letras de la palabra MARTIN con la condición de que las consonantes mantengan el orden original?

**Resolución**

Martín tiene 6 letras, dos de ellas vocales. Si las consonantes tienen que mantener el orden original, hay 5 lugares donde ubicar las vocales

$$-M-R-T-N-$$

Si consideramos los lugares como celdas y las vocales como bolas **distinguibiles** el número de configuraciones posibles es

$$\frac{(2+5-1)!}{(5-1)!} = \frac{6!}{4!}$$

Otra forma posible es elegir primero donde poner las consonantes (como subconjunto, pues hay una única ordenación posible), que resulta

$$\binom{6}{4}$$

y por cada configuración, tengo  $2!$  formas de poner las vocales.

$$2! \binom{6}{4} = \frac{2!6!}{2!4!} = \frac{6!}{4!}$$

**Ejemplo 2.2.24** ¿De cuántas formas pueden descender 10 pasajeros de un ascensor que para en 12 pisos? Y si no se hace distinción de personas?

**Resolución**

Las personas son siempre distinguibles, por lo cual este problema es equivalente a repartir 10 bolas distinguibles en 12 celdas. El número de configuraciones es  $12^{10}$ .

Si consideramos las personas indistinguibles, entonces la cantidad de configuraciones posibles es

$$\binom{10+12-1}{10}.$$

## Particiones ordenadas

**Ejemplo 2.2.25** ¿En cuántas formas es posible descomponer al número natural  $n$  en suma de  $k$  sumandos enteros mayores o iguales a cero? Por ejemplo, si  $k = 2$  y  $n = 4$  se tienen las siguientes particiones:

$$4 = 0 + 4 = 4 + 0 = 1 + 3 = 3 + 1 = 2 + 2$$

**Resolución**

Si distinguimos el orden de los sumandos, se trata de colocar  $n$  objetos en  $k$  celdas. Cada sumando representa el número de objetos en esa celda.

$$\binom{n+k-1}{n}$$

**Ejemplo 2.2.26** ¿En cuántas formas es posible descomponer al número natural  $n$  como suma de  $k$  números naturales?.

**Resolución**

En este caso el cero no está permitido, por lo cual si coloco un objeto en cada celda, me quedan  $n - k$  objetos indistinguibles para repartir. Este número es

$$\binom{n-k+(k-1)}{n-k}$$

**Ejemplo 2.2.27** ¿En cuantas formas pueden desfilan  $n$  personas en  $k$  grupos alineados no vacíos?

XXXX  
XXX  
XXXX

**Resolución**

Considero primero las personas como indistinguibles. Como los grupos son no vacíos, asigno  $k$  personas, una por grupo y luego asigno las  $(n - k)$  restantes sin restricciones. Tengo entonces

$$\binom{(n-k) + (k-1)}{n-k}$$

Ahora, como las personas son distinguibles, tengo  $n!$  posibles reordenaciones de cada configuración. La cantidad de posibles resultados en

$$n! \binom{n-1}{n-k}$$

**Ejemplo 2.2.28** ¿De cuantas formas podemos ordenar 7 niños y 5 niñas en hilera si no queremos que dos o más niñas estén juntas?

**Resolución**

Consideremos los niños y niñas indistinguibles. Si ubicamos 4 niños alternados con las 5 niñas, tenemos la siguiente configuración

-MV-MV-MV-MV-M-

en la cual hay 6 lugares libres donde ubicar los 3 varones restantes. Por lo cual el número de configuraciones es

$$\binom{6+3-1}{3} = \binom{8}{3}$$

Como los niños y niñas son distinguibles, debemos contar las reordenaciones posibles entre las niñas y los niños, por lo cual el número de ordenaciones posibles es

$$\binom{8}{3} 7! 5!$$

**Ejemplo 2.2.29** ¿De cuántas formas pueden ubicarse 6 hombres y 6 mujeres en un teatro que tiene 11 filas de 12 asientos cada una, si todos deben sentarse en la misma fila y dos personas del mismo sexo no deben tener asientos contiguos?

**Resolución**

Elijo primero la fila donde deben sentarse las personas, tengo 11 posibilidades. Luego elijo si comienzo por un hombre o una mujer, tengo 2 posibilidades. Y fijado esto, la configuración alternada de hombre y mujer es una sola. Como hombres y mujeres son siempre distinguibles, por cada posibilidad de fila y configuración tengo  $6!$  movimientos entre las mujeres y  $6!$  movimientos entre los hombres. El resultado final es  $11 \times 2 \times 6! \times 6! = 11,404,800$ .

**Ejemplo 2.2.30** ¿Cuántos números impares de 6 cifras se pueden formar con los dígitos 1,2,3 y 6 de manera que el 2 aparezca una cantidad impar de veces?

**Resolución**

Empecemos por estudiar la restricción. El número dos solo puede aparecer un número impar de veces, es decir puede haber números con un solo dos, con tres dos o con cinco dos. Además el número final debe ser impar, por lo cual el

dos y el seis no pueden estar en la última posición. Entonces debemos partir en casos nuestro problema y la suma de los resultados será el resultado final.

Caso 1: Solo un dos en el número final. Elijo una posición para el dos, que no puede ser la última, por lo cual tengo 5 posibilidades de elección. Quedan 5 posiciones para llenar. La última posición tiene dos posibilidades solamente, el 1 y el tres pues el número final tiene que ser impar, y las restantes tienen 3 posibilidades cada una, el seis el uno y el tres. El resultado final es  $5 \times 2 \times 3 \times 3 \times 3 \times 3 = 810$ .

Caso 2: Si el número tiene tres dos, y el último dígito de los seis no puede ser un dos, tengo que elegir tres posiciones de cinco para ubicar esos dos,  $5!/2!3! = 10$  posibilidades. Fijados los dos, quedan tres dígitos por llenar, y el último no puede ser un seis, por lo cual tengo solo dos posibilidades para ese dígito y tres para los dos restantes. El resultado final es  $10 \times 2 \times 3 \times 3 = 180$ .

Caso 3: Si hay 5 dos en un número que no es par, deben estar todos en los primeros cinco dígitos, el último no puede ser seis por lo cual hay solo dos posibilidades para este número.

El resultado final es  $810 + 180 + 2 = 992$ .

**Ejemplo 2.2.31** ¿Cuántas ensaladas de frutas puedo formar si tengo una banana, una manzana, una pera, un durazno, una ciruela y una rodaja de melón con tres jugos: naranja, frutilla y limón, si sólo permito mezclar a lo sumo dos de estos jugos? (La ensalada se forma con al menos una fruta).

#### Resolución

Una ensalada de frutas posible es un subconjunto no vacío del conjunto de frutas. Si tengo seis frutas tengo  $2^6 - 1$  subconjuntos no vacíos posibles. Ahora cada subconjunto puede ser mezclado con a lo sumo dos jugos, es decir, puedo no mezclarlo con jugos, puedo agregarle un jugo o puedo agregarle dos jugos. Lo que no puedo hacer es agregarle los tres jugos. Entonces el número de jugos que puedo agregar es  $2^3 - 1$ . El resultado final es  $(2^6 - 1) \times (2^3 - 1) = 441$ .

**Ejemplo 2.2.32** Suponga que tiene 10 pares de zapatos en su botinero. Diga cuántos subconjuntos distintos de cuatro zapatos puede elegir. Diga cuántos conjuntos de cuatro zapatos contienen por lo menos un par de zapatos.

#### Resolución

Tengo 20 zapatos en total, puedo elegir  $20!/16!4! = 4845$  subconjuntos de cuatro elementos. Ahora si quiero pares de zapatos, preciso elegir un par de 10 y luego de los 9 pares restantes, tengo dos opciones, elijo otro par, 9 posibilidades, (y divido por dos porque quiero subconjuntos) o elijo un subconjunto que no sea un par. Si tengo 18 zapatos, tengo 18 posibilidades para el primer zapato y 16 para el segundo (automáticamente estoy rechazando el compañero del primero). Divido por las permutaciones de dos (son subconjuntos) y resulta  $10 \times 9/2 + 10 \times (18 \times 16/2) = 45 + 1440 = 1485$ .

# Capítulo 3

## Espacios de Probabilidad

El avance de la ciencia y la rapidez de las computadoras han provocado que muchas situaciones puedan ser predichas de antemano. Si se deja caer una pelota desde la ventana de un tercer piso de un edificio hacia la vereda, usando las leyes elementales de la física se puede predecir cuanto va a tardar en llegar al piso. Si se observa la posición y el movimiento de un asteroide, usando leyes más avanzadas de la física se puede predecir donde va a estar posicionado dentro de un año. En cambio otras cosas son impredecibles, como el resultado de tirar una moneda o el tiempo que puede tardar un consumidor en ser atendido en la cola de un supermercado. Sin embargo, hay una regularidad en esos fenómenos que los ubica entre la predicción perfecta y la total aleatoriedad. A pesar de que un resultado particular puede ser desconocido, en varias repeticiones puede surgir un patrón de resultados. No se puede decir si va a salir cara o número al tirar una moneda de 50 centavos, pero se intuye que si se tira muchísimas veces la moneda a la larga se obtendrán tantas caras como números. Ese es el patrón que emerge del fenómeno observado.

También, hay fenómenos que pueden observarse en forma repetida y hay otros que no es posible hacerlo. Sacar una carta de un mazo de naipes, tirar una moneda, sacar una bola de una urna son experiencias que pueden ser controladas y repetidas sin problemas. Prender una computadora hasta que falle para estudiar cuanto dura es un poco más complicado, dejando de lado el costo. Lo mismo mantener una lámpara encendida hasta que se queme para estudiar su durabilidad. Observar el efecto del estrés en el deambular de una codorniz en un laboratorio también es un experimento y es mucho más difícil de controlar y repetir que sacar un naipe de un mazo. En este capítulo vamos a estudiar mediante ejemplos de diversa naturaleza los fenómenos aleatorios y su relación con preguntas como ¿Cuál es la probabilidad de sacar una lámpara de un lote, probarla y que se queme en menos de dos horas, si el fabricante dice que menos del 3% de sus lámparas se quema en menos de dos horas? Pero primero vamos a estudiar ejemplos totalmente controlados, más simples, por tradición relacionados con juegos de azar.

### 3.1. Modelo matemático para una población

**Definición 3.1.1** Un **experimento** es el proceso por el cual se obtienen las observaciones que se desean estudiar.

**Ejemplo 3.1.1** Ejemplos típicos de experimentos son:

- Tirar una moneda, sacar una carta de un mazo, una bola de una urna.
- Mediciones de cantidad de lluvia mensual.
- Observaciones sobre el estado del tiempo: nublado, parcialmente nublado, soleado, lluvioso, con lluvias intermitentes, etc.
- Entrevistas de preferencia sobre productos diversos.

- Inspecciones a tubos de rayos catódicos de televisores para identificar defectuosos.

Notemos que las observaciones no siempre son numéricas, pueden ser categóricas. Observemos también que una población de mediciones resulta de repetir el experimento infinitas veces. Por ejemplo, estamos interesados en la vida útil de un tubo de rayos catódicos para televisor producido en tal fábrica durante cierto período. La inspección de 1 tubo para detectar cuanto tiempo falla es un experimento simple, mientras que la repetición de dicha experiencia con todos los tubos producidos en ese período genera la población completa. Una muestra es el conjunto de datos obtenidos al repetir el experimento un número pequeño de veces.

Trataremos de analizar los experimentos para poder construir un modelo matemático que lo describa en el cual se asigne probabilidad a cada resultado posible.

**Definición 3.1.2** El **Espacio muestral** es el conjunto de los resultados posibles de un experimento y se denota con la letra griega Omega  $\Omega$ . Un **evento** es un subconjunto del espacio muestral. Un punto muestral es un elemento del espacio muestral. Si  $\Omega$  es un conjunto **finito o numerable**, el espacio muestral se llama **discreto**.

Es muy importante observar que los eventos son conjuntos de puntos muestrales, por lo cual si tomamos un par de eventos, las operaciones comunes de conjuntos, como la intersección y la unión, vuelven a ser un evento del espacio muestral. Lo mismo ocurre con el complemento; el complemento de un evento es un evento y se llama evento complementario.

**Ejemplo 3.1.2** El experimento consiste en arrojar un dado honesto y observar que número salió. ¿Cuál es el espacio muestral? ¿Y los eventos  $A =$  se observa un número par,  $B =$  se observa un número menor o igual que 4?. Liste el evento  $C = B \cap A^c$ , se observa un número menor o igual que cuatro que no sea par.

### Resolución

El espacio muestral consta de los valores que puede tomar el dado,  $\Omega = \{1, 2, 3, 4, 5, 6\}$  y es un espacio discreto. El evento  $A$  consta de los números  $A = \{2, 4, 6\}$  y el  $B = \{1, 2, 3, 4\}$ . El evento  $C$  es  $A^c \cap B = \{1, 3\}$ .

Veamos ahora un ejemplo un poco más complicado.

**Ejemplo 3.1.3** Una urna tiene 5 bolas, 3 blancas numeradas  $b_1, b_2, b_3$  y dos negras numeradas  $n_1, n_2$ . Describa el espacio muestral de los experimentos

1. Se sacan dos bolas en sucesión, devolviendo la primera bola a la urna para después extraer la otra
2. Se sacan dos bolas en sucesión, sin devolver la primera bola a la urna para después extraer la otra
3. Se sacan dos bolas al mismo tiempo

### Resolución

Sean  $b_1, b_2, b_3$  las bolas blancas y  $n_1, n_2$  las bolas negras. Entonces

1. El primer experimento corresponde al llamado muestreo con reposición, es decir, repito el mismo experimento, sacar una bola, dos veces consecutivas en forma exacta. Tiene como espacio muestral al conjunto de pares ordenados tales que el primer elemento del par representa la primera bola extraída y el segundo elemento del par la segunda bola extraída. Como todo conjunto puede ser listado o descrito por comprensión. Como es un espacio moderadamente grande lo describiremos por comprensión:

$$\Omega_1 = \{(x, y), x, y = b_1, b_2, b_3, n_1, n_2\} \quad \# \Omega_1 = 5 \times 5 = 25$$

No es necesario listar todos los espacios, pero si es necesario describir como son sus puntos en forma clara y contar cuantos puntos hay en total.



2. El segundo experimento corresponde al llamado muestreo sin reposición, es decir, repito “casi” el mismo experimento, sacar una bola, dos veces consecutivas. Al no volver a poner la primer bola que saque, la urna no es la misma al sacar la segunda bola, y los puntos son pares ordenados con la restricción de que la primera bola no puede ser igual a la segunda.

$$\Omega_2 = \{(x, y), x, y = b_1, b_2, b_3, n_1, n_2, \text{ pero } x \text{ no puede ser igual a } y\} \quad \#\Omega_2 = 5 \times 4 = 20$$

El espacio muestral tiene menos puntos que el anterior, los puntos  $(b_1, b_1), (b_2, b_2), (b_3, b_3), (n_1, n_1), (n_2, n_2)$  no están en este espacio pues no son observables.

3. El experimento sacar dos bolas al mismo tiempo tiene como espacio muestral al conjunto de subconjuntos de dos bolas de las 5 que tengo, es mucho más chico que los anteriores, pues no se repitieron experimentos en sucesión, se realizó la extracción al mismo tiempo. El espacio descrito por comprensión es

$$\Omega_3 = \{\{x, y\}, x, y = b_1, b_2, b_3, n_1, n_2, x \neq y\} \quad \#\Omega_3 = \binom{5}{2} = 10$$

Este es un espacio más chico y se puede listar

- |                   |                    |
|-------------------|--------------------|
| 1. $\{b_1, b_2\}$ | 6. $\{b_2, n_1\}$  |
| 2. $\{b_1, b_3\}$ | 7. $\{b_2, n_2\}$  |
| 3. $\{b_1, n_1\}$ | 8. $\{b_3, n_1\}$  |
| 4. $\{b_1, n_2\}$ | 9. $\{b_3, n_2\}$  |
| 5. $\{b_2, b_3\}$ | 10. $\{n_1, n_2\}$ |

Veamos ahora un ejemplo de subconjuntos elegidos al azar, pero en vez de bolas en una urna vamos a tener zapatos en el ropero.

**Ejemplo 3.1.4** El botinero tiene 2 pares de zapatos de hombre y 2 par de zapatos de mujer. Si se eligen 4 zapatos al azar, describa el espacio muestral que corresponde a la experiencia y encuentre el evento “se sacaron dos pares de zapatos”.

### Resolución

Los zapatos son usualmente distinguibles entre sí, pues los modelos son distintos, y por cada par está el zapato izquierdo y el derecho. Si los llamamos  $H_{11}, H_{12}, H_{21}, H_{22}$  a los zapatos de hombre y  $M_1, M_2$  a los zapatos de mujer, el espacio muestral resulta el de los posibles subconjuntos de cuatro elementos formados de un conjunto de seis,  $6!/4!2! = 15$  suconjuntos en total:

- |   |                                      |                                    |
|---|--------------------------------------|------------------------------------|
| 1. $\{H_{11}, H_{12}, H_{21}, H_{22}\}$ | 6. $\{H_{11}, H_{12}, H_{21}, M_2\}$ | 11. $\{H_{11}, H_{21}, M_1, M_2\}$ |
| 2. $\{H_{11}, H_{12}, H_{21}, M_1\}$    | 7. $\{H_{11}, H_{12}, H_{22}, M_2\}$ | 12. $\{H_{11}, H_{22}, M_1, M_2\}$ |
| 3. $\{H_{11}, H_{12}, H_{22}, M_1\}$    | 8. $\{H_{11}, H_{22}, H_{21}, M_2\}$ | 13. $\{H_{12}, H_{21}, M_1, M_2\}$ |
| 4. $\{H_{11}, H_{22}, H_{21}, M_1\}$    | 9. $\{H_{22}, H_{12}, H_{21}, M_2\}$ | 14. $\{H_{12}, H_{22}, M_1, M_2\}$ |
| 5. $\{H_{22}, H_{12}, H_{21}, M_1\}$    | 10. $\{H_{11}, H_{12}, M_1, M_2\}$   | 15. $\{H_{21}, H_{22}, M_1, M_2\}$ |

El evento  $A =$  “se sacaron dos pares de zapatos” es

$$A = \{\{H_{11}, H_{12}, H_{21}, H_{22}\}; \{H_{11}, H_{12}, M_1, M_2\}; \{H_{21}, H_{22}, M_1, M_2\}\}$$

**Ejemplo 3.1.5** Una urna tiene 5 bolas, 3 blancas indistinguibles y dos negras indistinguibles. Escriba el espacio muestral de los experimentos

1. Se sacan dos bolas con repetición (se vuelve a poner la primera en la urna, se repite el experimento).

2. Se sacan dos bolas sin repetición (la primera bola no se vuelve a poner en la urna, no se repite el experimento, se cambia).
3. Se sacan dos bolas al mismo tiempo.

### Resolución

Si denotamos por  $b$  y  $n$  a las bolas blancas y negras:

1. Los dos primeros experimentos tienen como espacio muestral al conjunto de pares ordenados en los cuales solo se registra el color de la bola extraída.

$$\Omega = \{(n, n), (n, b), (b, n), (b, b)\}$$

2. El tercer experimento tiene como espacio muestral al conjunto de subconjuntos, pues no se tiene registro de que bola se extrajo primero, solo el color de las bolas extraídas.

$$\Omega_2 = \{\{b, b\}, \{b, n\}, \{n, n\}\}$$

Los tres espacios son espacios muestrales discretos. Los primeros dos espacios muestrales son idénticos, el tercero es más chico, pero los dos primeros experimentos no son iguales, ¿porqué no se vé eso en la descripción de las observaciones?. Porque la diferencia se vá a ver cuando terminemos de definir el espacio de probabilidad, cuando veamos medidas de probabilidad. Sigamos explorando espacios muestrales.

**Ejemplo 3.1.6** Las alturas de las personas de una población son cantidades físicas exactas, pero al medirlas (aún haciéndolo cuidadosamente) se suele introducir un error de medición que se considera aleatorio. Si elijo una persona adulta al azar y le mido su altura, el conjunto de posibles resultados del experimento es

$$\Omega = \{h \in \mathbb{R} \mid 2,5 > h > 0,5\}$$

aunque suele representarse por un conjunto más amplio, el de los reales positivos

$$\Omega = \{h \in \mathbb{R} \mid h > 0\}$$

Este no es un espacio muestral discreto, dado que no es finito ni numerable. No es algo muy serio considerar un espacio muestral como más grande de lo que realmente debería ser, aunque ninguna persona mide más de tres metros, pero si debemos estar seguros de que no olvidamos ningún resultado que sea posible de ser observado.

En el siguiente ejemplo revisaremos el ejemplo de muestreo con repetición con tres bolas iguales  $c$  y tres bolas iguales  $d$ , y tres extracciones. Sin embargo el contexto es totalmente distinto. Es porque nuestros ejemplos de estudio en realidad se repiten en una gran variedad de situaciones. Podemos estudiarlos mejor si la descripción es más precisa, como con la urna, en la cual sé exactamente cuantas bolas de cada color tengo.

**Ejemplo 3.1.7** Un conductor en su camino al trabajo pasa por una sucesión de tres intersecciones con semáforos. En cada intersección el conductor se detiene  $d$  o continua  $c$  su marcha de acuerdo con la luz del semáforo. El espacio muestral de todos los posibles resultados es

$$\Omega = \{ccc, ccd, cdd, cdc, ddd, ddc, dcc, dcd\}$$

donde  $cdc$  representa el evento que el conductor continúe en la primera intersección, se detenga en la segunda y continúe en la tercera. Sea  $A$  el evento el conductor se detiene en la primera intersección, y  $B$  el evento se detiene en

la tercera intersección, entonces el evento  $C$  que se detenga en la primera y la tercera es la intersección de los eventos  $A$  y  $B$ .

$$A = \{ddd, ddc, dcc, dcd\} \quad B = \{ddd, dcd, ccd, cdd\} \quad C = \{ddd, dcd\} = A \cap B$$

¿Cuál es el evento que no se detenga en la primera intersección?. Es  $A^c$ , y consiste de los siguientes puntos.

$$A^c = \{ccc, ccd, cdd, cdc\}$$

Ahora vemos un ejemplo del tipo de bolas extraídas al mismo tiempo, solo que en este caso son medias.

**Ejemplo 3.1.8** Un cajón tiene 2 pares de medias negras, 1 par de medias azules y 1 par de medias verdes. Si todas las medias están sueltas y mezcladas en el cajón, y se sacan dos medias en el oscuro, encuentre el espacio muestral resultante al experimento que registra los colores de las medias extraídas. ¿Sería igual si las medias fueran distinguibles? ¿Cuál sería el cardinal del espacio de medias distinguibles?

### Resolución

Si denotamos a las medias azules con  $A$ , las verdes con  $V$  y a las negras con  $N$ , el espacio muestral es el de los subconjuntos que puedo formar con dos medias elegidas al azar **de tres posibles**, considerando que no distingo las medias de un mismo color. Y si tengo más medias no se modifica, solo si agrego colores pues lo único que registra este experimento son los colores de las medias extraídas.

$$\Omega = \{AA, AV, AN, VN, VV, NN\}$$

Si las medias fueran distinguibles, el espacio sería mucho mayor, pues cada uno de los puntos listados anteriormente sería un evento. Por ejemplo, el par de medias azules involucraría 2 medias azules distintas, y el de verdes dos medias verdes distintas, por lo cual  $AV$  se transformaría en

$$AV = \{A_1V_1, A_1V_2, A_2V_1, A_2V_2\}$$

Observemos que siguen siendo subconjuntos, no saco una media primero y otra despues, que involucraría distinguir el orden en que se sacaron las medias. Por lo cual, el cardinal del espacio de medias distinguibles es la cantidad de subconjuntos de dos elementos que puedo formar con seis medias distintas,  $6!/3!2! = 6 \times 5 \times 2 = 60$  puntos.

Ahora vamos a estudiar otro espacio correspondiente a un experimento controlado, donde las cajas o urnas son más de una y las bolas pueden ser distinguibles o indistinguibles.

**Ejemplo 3.1.9** El experimento consiste en distribuir tres bolas **distinguibles** en tres cajas. El espacio muestral del experimento  $\Omega$  consta de los puntos

- |                 |                |                |
|-----------------|----------------|----------------|
| 1. $abc  -   -$ | 10. $c ab  -$  | 19. $- b ac$   |
| 2. $- abc  -$   | 11. $- ab c$   | 20. $a  -  bc$ |
| 3. $-  -  abc $ | 12. $b ac  -$  | 21. $- a bc$   |
| 4. $ab c  -$    | 13. $b ac  -$  | 22. $a b c$    |
| 5. $ab  -  c$   | 14. $a bc  -$  | 23. $a c b$    |
| 6. $ac b  -$    | 15. $- bc a$   | 24. $b a c$    |
| 7. $ac  -  b$   | 16. $c  -  ab$ | 25. $b c a$    |
| 8. $bc a  -$    | 17. $- c ab$   | 26. $c a b$    |
| 9. $bc  -  a$   | 18. $b  -  ac$ | 27. $c b a$    |

Enumerar los conjuntos que corresponden a los siguientes eventos:

1. primera caja vacía

2. dos bolas en la primera caja
3. ninguna caja ocupada por más de una bola
4. primera caja vacía y ninguna caja ocupada por más de una bola.

### Resolución

El evento  $A$ , “la primera caja vacía” consta de los puntos  $\{ 2. (-|abc|-), 3. (-|-|abc|), 11. (-|ab|c), 15. (-|bc|a), 17. (-|c|ab), 19. (-|b|ac), 21. (-|a|bc) \}$ .

El evento  $B$ , “dos bolillas en la primera caja” consta de los puntos  $\{ 4. (ab|c|-), 5. (ab|-|c), 6. (ac|b|-), 7. (ac|-|b), 8. (bc|a|-), 9. (bc|-|a) \}$ .

El evento  $C$ , “ninguna caja ocupada por más de una bolilla” consta de los puntos  $\{ 22. (a|b|c), 23. (a|c|b), 24. (b|a|c), 25. (b|c|a), 26. (c|a|b), 27. (c|b|a) \}$ .

El evento  $D = A \cap C$ , “primera caja vacía y ninguna caja ocupada por más de una bolilla” es el evento vacío.

**Ejemplo 3.1.10** El experimento consiste en distribuir tres bolas **indistinguibles** en tres cajas. El espacio muestral del experimento  $\Omega$  consta de los siguientes puntos

- |            |           |
|------------|-----------|
| 1. 000 — — | 6. 0 00 — |
| 2. — 000 — | 7. — 00 0 |
| 3. — — 000 | 8. 0 — 00 |
| 4. 00 0 —  | 9. — 0 00 |
| 5. 00 — 0  | 10. 0 0 0 |

Enumerar los conjuntos que corresponden a los siguientes eventos:

1. dos bolas en la primera caja
2. ninguna caja vacía
3. alguna caja ocupada por más de una bola.

### Resolución

El evento  $A$ , “dos bolas en la primera caja” consta de los puntos  $\{ 4. (00|0|—) 5. (00|—|0) \}$ .

El evento  $B$ , “ninguna caja vacía” consta del punto  $\{ 10. (0|0|0) \}$ .

El evento  $C$ , “alguna caja ocupada por más de una bola.” consta de todos los puntos menos el punto  $\{ 10. (0|0|0) \}$ .

En el siguiente ejemplo vamos a armar espacios grandes, que usualmente no se listan, sino se les calcula el tamaño y quizás se liste algún evento importante. Vamos a hacer uso de las nociones de combinatoria obtenidas en el capítulo anterior.

**Ejemplo 3.1.11** Un kiosquero compra un kilo de caramelos de goma de cuatro sabores, banana, frutilla, pera y naranja. El kiosquero arma bolsitas de 10 caramelos cada una venderle a los niños de una escuela cercana.

- a) Si los caramelos se consideran distinguibles entre sí, arme el espacio muestral de la experiencia armar bolsitas eligiendo caramelos al azar. ¿Cuántos puntos muestrales hay?
- b) Si los caramelos de cada sabor se consideran indistinguibles, entonces dos bolsitas son iguales si la cantidad de caramelos de distinto sabor son iguales. Arme el espacio muestral de la experiencia armar bolsitas distintas eligiendo caramelos al azar. ¿Cuántos puntos muestrales hay?

**Resolución**

a) Si los caramelos son distinguibles, el espacio muestral es el conjunto de todas las bolsitas posibles de 10 caramelos. Cada bolsita es un subconjunto de 10 elementos elegidos al azar de un conjunto de 400 elementos. Por lo cual el cardinal del espacio muestral es

$$\binom{400}{10}$$

b) Como los caramelos son indistinguibles si tienen el mismo sabor, hay tantas bolsitas distintas como 4-uplas ordenadas  $(k_b, k_f, k_p, k_n)$  con  $k_b + k_f + k_p + k_n = 10$ , pues  $k_b, k_f, k_p, k_n$  indican el número de caramelos de cada sabor que tiene la bolsita. El cardinal del conjunto es

$$\binom{10+4-1}{10}$$

todas las formas de escribir el número 10 como suma de cuatro sumandos mayores o iguales a cero, considerando el orden de los sumandos importante (pues indica el sabor del caramelo).

**Ejemplo 3.1.12** Supongamos que el experimento consiste en tirar un dardo al azar en un disco de radio  $r$  y denotar el lugar donde éste cayó. ¿Cuál es el espacio muestral del experimento?

**Resolución**

El espacio muestral  $\Omega$  es el conjunto de puntos del plano real contenidos en el disco de radio  $r$ . Es un conjunto infinito no numerable, por lo cual no es un espacio discreto. Como son los eventos de este espacio muestral? Todos los círculos, cuadrados y rombos que están dentro del disco considerado son eventos. Es un espacio continuo, no discreto como los considerados anteriormente.

Ahora consideraremos experimentos donde se repiten en sucesión experiencias simples distintas.

**Ejemplo 3.1.13** Hacemos el experimento siguiente: lanzar una moneda y un dado.

1. Describir el espacio muestral.
2. Describir los eventos:  $A$  = “aparecen caras y un número par”,  $B$  = “aparece un número primo” y  $C$  = “aparecen número y un número impar”
3. Describir los eventos:  $A$  o  $B$  suceden;  $B$  y  $C$  suceden; no sucede  $B$ .

**Resolución**

Si denotamos a la salida de una cara por  $C$  y a la de número por  $X$ , el espacio muestral es

- |          |           |
|----------|-----------|
| 1. $C 1$ | 7. $X 1$  |
| 2. $C 2$ | 8. $X 2$  |
| 3. $C 3$ | 9. $X 3$  |
| 4. $C 4$ | 10. $X 4$ |
| 5. $C 5$ | 11. $X 5$ |
| 6. $C 6$ | 12. $X 6$ |

El evento  $A$  = “aparecen caras y un número par” consta de los puntos  $C|2, C|4, C|6$ , y el evento  $B$  = “aparece un número primo” consta de los puntos  $C|1, X|1, C|3, X|3, C|5, X|5$ . El evento  $C$  = “aparecen número y un número impar” tiene los puntos  $X|1, X|3, X|5$ .

El evento  $A$  o  $B$  suceden es la union de ambos eventos  $A \cup B$  y consta de los puntos  $C|2, C|4, C|6, C|1, X|1, C|3, X|3, C|5, X|5$ . El evento  $B$  y  $C$  suceden es la interseccion de ambos eventos  $B \cap C$ , es exactamente el evento  $C$ . El evento “no sucede  $B$ ” consta de los puntos  $C|2, X|2, C|4, X|4, C|6, X|6$ .

El siguiente ejemplo es un ejemplo distinto a los anteriores y más complejo de analizar. Los datos no salen de una experiencia de laboratorio sino de un encuesta. Dentro de las preguntas hechas a las señoras se encuentran su estado civil anterior y actual y que tipos de estudios tienen. En la tabla siguiente solo se presentan la cantidad de personas que coinciden en sus respuestas, en cuatro categorías, y la cantidad de personas que respondieron en cada categoría. Este es un ejemplo de un estudio real y de como se identifica la información recolectada. En principio, como no se tiene una lista de los sujetos que son el espacio muestral, se trabaja con porcentajes, porción del total que cumple alguna condición.

**Ejemplo 3.1.14** En un estudio demográfico de 1436 mujeres que se casaron al menos una vez se obtuvieron los siguientes estratos.

Educación	Casada una vez	Casada más de una vez	Total
Universitaria	550	61	611
No universitaria	681	144	825
Total	1231	205	1436

1. Identifique el espacio muestral.
2. ¿Cuál es la proporción de mujeres que se caso más de una vez?
3. ¿Cuál es la proporción de mujeres casadas una vez que tienen educación universitaria?
4. ¿Cuál es la proporción de mujeres casadas más de una vez que no tienen educación universitaria?
5. De las mujeres que tienen estudios universitarios, ¿cuál es la proporción que se casó más de una vez?
6. De las mujeres que se casaron solo una vez, ¿cuál es la proporción que tiene estudios universitarios ?

### Resolución

El espacio muestral es el conjunto de las 1436 mujeres que se casaron al menos una vez. Claramente no puedo listar los puntos por mas que enumere a las señoras, y en la tabla dada estan presentados los cardinales de algunos eventos importantes y sus intersecciones. En muchos casos es útil listar las proporciones del total de señoras que cumplen alguna condición.

La proporción de señoras que se casó más de una vez es la cantidad de señoras que se casó más de una vez dividido la cantidad de señoras que se encuestó,

$$\frac{205}{1436} = 0,143.$$

La proporción de mujeres casadas una vez que tienen educación universitaria es el cardinal del conjunto intersección, 550, dividido el total de encuestadas, 1436

$$\frac{550}{1436} = 0,383.$$

De las mujeres que tienen estudios universitarios, la proporción de mujeres casadas más de una vez que no tienen educación universitaria es el cardinal del conjunto intersección, 61, dividido el total de **universitarias** encuestadas , 611

$$\frac{61}{611} = 0,099.$$

De las mujeres casadas más de una vez, la proporción de señoras que no tienen educación universitaria es el cardinal del conjunto intersección, 681, dividido el total de encuestadas de esa clase 1231

$$\frac{681}{1231} = 0,55.$$

## 3.2. Medidas de Probabilidad

La noción de probabilidad más arraigada en el subconsciente de las personas corresponde con la idea intuitiva de frecuencia relativa de un evento. Un evento es probable, más probable, poco probable, muy probable que ocurra si es de nuestro conocimiento que en varias repeticiones del experimento ese evento se observe algunas veces, varias veces, pocas veces o muchas veces. Un periodista deportivo puede decir que es muy probable que Boca gane el domingo pues en los últimos 20 encuentros con ese equipo gana 15 y perdió 5 partidos, o porque viene con una racha ganadora de 10 partidos sin perder, etc.

¿Cuál es entonces la definición de probabilidad?. ¿Hay una única probabilidad o depende de la subjetividad de cada uno?

La respuesta es compleja, pues en realidad hay infinitas medidas de probabilidad para un mismo espacio muestral, pero en muchos casos la medida más razonable, la que ajusta mejor a nuestras nociones intuitivas, que modela mejor el problema, es una sola.

A continuación daremos la definición matemática establecida por el matemático ruso Kolmogorov en 1933. Estas reglas fueron derivadas del estudio cuidadoso y reflexivo de experimentos simples, como la tirada de una moneda. ¿Cuál debería ser la verdadera probabilidad de observar una cara?. La proporción de caras observadas (número de caras dividido número de tiradas, llamado también frecuencia relativa) es un valor intuitivo, pero depende del número de tiradas realizado. Estos son algunos ejemplos de tiradas de la moneda famosas:

- El naturalista francés Buffon (1707-1788) tiró la moneda 4040 veces, obteniendo 2048 caras. La frecuencia relativa del número de caras es  $2048/4040=0.5069$ .
- El estadístico inglés Karl Pearson, en los principios del siglo 20, tiró la moneda 24000 veces, resultando 12012 caras y una frecuencia relativa de 0.5005.
- El matemático inglés John Kerrich, cuando estuvo prisionero de los alemanes en la segunda guerra mundial, tiró la moneda 10000 veces, con un resultado de 5067 caras, y una frecuencia relativa de 0.5067.

Observemos que en todos estos casos la frecuencia relativa es mayor a 0.5, pero no es siempre la misma, ¿debemos concluir que no sabemos exactamente cuanto es pero debe ser mayor a 0.5?. En realidad no, porque un nuevo ejemplo puede dar menor a 0.5. ¿Cuál es entonces la probabilidad de sacar una cara? Bueno, si no hay razones para pensar que sacar cara es más fácil que sacar número, la probabilidad de sacar cara debería ser igual a la de sacar número. Y como las frecuencias relativas son menores que 1, sería razonable pensar que ambas probabilidades deberían ser menores que uno. Y como son complementarias (o se observa cara o se observa número), su suma debería ser dos veces una de ellas.

$$1 = 2k \text{ implica } k = \frac{1}{2}$$

Por lo cual solo puede ser 0.5 cada una. Esto es lo que querían comprobar los matemáticos con sus experimentos, que este razonamiento estaba de acuerdo con los resultados que obtenían. Y Kolmogorov tradujo este razonamiento en una descripción matemática, en reglas que todas las probabilidades deberían cumplir, con respecto a conjuntos de resultados de experimentos.

**Definición 3.2.1** *Supongamos observar un experimento con un espacio muestral  $\Omega$ . Una medida de probabilidad es una función  $P$  definida en eventos del espacio muestral  $\Omega$  que da como resultado un número real y que cumple los siguientes axiomas:*

- Si  $E$  es un evento, entonces  $P(E) \geq 0$ .

- $P(\Omega) = 1$ , la probabilidad del conjunto total es 1.
- Si  $E_1$  y  $E_2$  son eventos de  $\Omega$  disjuntos entonces la probabilidad de la unión es la suma

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

- Más generalmente, si  $\{E_i\}_{i \in I}$  es una colección de eventos mutuamente excluyentes (no se intersecan entre sí), y  $E = \cup_{i \in I} E_i$  la unión de todos los conjuntos, entonces  $P(E) = \sum_{i \in I} P(E_i)$ , la suma de las probabilidades de los conjuntos.

Un modelo probabilístico asociado a un experimento consta de un espacio muestral y una probabilidad asociada al espacio muestral.

Veamos un ejemplo de espacio muestral y probabilidad asociada.

**Ejemplo 3.2.1** Supongamos que seleccionamos un Rocklet de un paquete al azar y registramos su color. El espacio muestral es

$$S = \{\text{rojo}, \text{azul}, \text{amarillo}, \text{verde}, \text{naranja}\}$$

La probabilidad de cada resultado sería la proporción de Rocklets de cada color producida en la fábrica. El fabricante puede decirnos que la medida de probabilidad correcta para cada uno de los colores es

Color	rojo	azul	amarillo	verde	naranja
Probabilidad	0.3	0.2	0.2	0.2	0.1

¿Cual sería la probabilidad de sacar un Rocklet rojo o naranja? La suma de las probabilidades

$$P(\text{rojo o naranja}) = P(\text{rojo}) + P(\text{naranja}) = 0,3 + 0,1 = 0,4$$

En este ejemplo modelamos el experimento diciendo que la proporción de Rocklet de cada color debería ser la probabilidad de sacar un rocklet de ese color, y el fabricante nos dió ese valor. ¿Qué podemos hacer para saber la probabilidad de sacar un Rocklet de un determinado color si no tenemos acceso al fabricante? Bueno, no podemos decir que nuestro paquete es un conjunto representativo de los colores de los Rocklets, pero si abrimos el paquete y contamos cuantos colores tenemos y cuantos Rocklets hay de cada color, podemos decir cuál es la probabilidad de sacar un rojo o un amarillo en una extracción de ese paquete.

Digamos que el paquete trae 20 Rocklets, de los cuales 3 son amarillos, 4 son rojos, 6 son verdes, 5 son naranja y dos son azules. La probabilidad que voy a asignar a cada color es la proporción de Rocklets que hay de cada color

Color	rojo	azul	amarillo	verde	naranja
Probabilidad	$\frac{3}{20}$	$\frac{4}{20}$	$\frac{6}{20}$	$\frac{5}{20}$	$\frac{2}{20}$

¿Porqué esta probabilidad y no otra? Pues si la extracción del Rocklet es hecha al azar, todos los Rocklets deberían tener igual probabilidad de ser elegidos, y por lo cual la probabilidad de ver un color determinado esta relacionada con cuantos Rocklets hay de ese color.

Y si en vez de sacar un Rocklet saco dos en forma sucesiva, o dos al mismo tiempo? Como resulta el espacio muestral? ¿Cuál es la probabilidad de sacar un Rocklet rojo y luego un amarillo? Estas preguntas vamos a responderlas luego de estudiar el caso general de extracción de bolas de colores de una urna.



Ahora vamos a estudiar un ejemplo de espacio que no es discreto, pues el experimento no toma un número finito de valores.

### Ejemplo 3.2.2

Supongamos que el experimento consiste en tirar un dardo al azar en un disco de radio  $r$  y denotar el lugar donde éste cayó. ¿Cuál es una asignación de probabilidades razonable para el experimento?

### Resolución

El espacio muestral  $\Omega$  es el conjunto de puntos del plano real contenidos en el disco de radio  $r$ . Es un conjunto infinito no numerable, por lo cual si a cada punto le asignamos una probabilidad finita no nula no podemos sumar las probabilidades y no se cumplen los axiomas de probabilidad. Aún así, nuestra intuición nos dice que si tiramos el dardo sin mirar, todos los "lugares" del disco deberían tener la misma probabilidad de ser elegidos. Diremos entonces que todo subconjunto del disco (que representa a un evento) tiene asignada la probabilidad

$$P(A) = \frac{\text{área de } A}{\text{área del círculo}}$$

y como el área de un punto es cero por definición, no tengo el problema de poder sumar. Solo las regiones con área tienen probabilidad distinta de cero, y regiones con igual área tienen igual probabilidad. Esta es la distribución de probabilidad uniforme en el espacio muestral  $\Omega$ .

**Definición 3.2.2** Sea  $\Omega$  un espacio infinito no numerable con una noción de área, y sea  $|A| = \text{área de } A$ . Entonces la medida de probabilidad uniforme sobre  $\Omega$  es aquella que le asigna a todo evento  $A$  del espacio muestral  $\Omega$  la probabilidad

$$P(A) = \frac{|A|}{|\Omega|}$$

En el caso de  $\Omega$  finito, esta probabilidad le asigna a cada punto el mismo valor  $1/\#\Omega$ , y el espacio se llama equiprobable. Esta asignación de probabilidades fue definida por el matemático Francés Pierre Simon Laplace (1749-1827).

### Definición 3.2.3 Probabilidad de Laplace

Si  $\Omega = \{\omega_i, i \in I\}$  es un espacio muestral finito cuyos puntos son equiprobables (ningún punto debería ser más probable que otro por la definición del experimento), entonces la probabilidad de cada punto muestral se define como

$$P(\omega_i) = \frac{1}{n^0 \text{ de puntos del espacio muestral}} = \frac{1}{\#\Omega}$$

y la probabilidad del evento  $A$  resulta:

$$P(A) = \frac{n^0 \text{ de puntos del evento}}{n^0 \text{ de puntos del espacio muestral}} = \frac{\#A}{\#\Omega}$$

Supongamos que existe una partición del espacio  $\Omega$ , conjuntos  $A_i$  disjuntos entre si tal que la unión me da todo  $\Omega$ , si definimos un nuevo espacio muestral como

$$\tilde{\Omega} = \{A_1, \dots, A_n\}$$

Entonces la probabilidad de cada punto de  $\tilde{\Omega}$  es la probabilidad del evento  $A_i$

$$P(A_i) = \frac{\#A_i}{\#\Omega}$$

Esto es lo que ocurre cuando elegimos un elemento al azar de una caja y miramos una característica que involucra a más de un elemento de la caja, por ejemplo, los Rocklets observados por color, los cangrejos diferenciados por sexo, los niños diferenciados por color de ojos, etc. Esta es la probabilidad frecuentista, si repetimos la experiencia veremos que los valores de las frecuencias relativas se acercan cada vez mas a los valores predichos.

Los axiomas que definen las probabilidades son reglas definidas que no deben ser probadas, en cambio las propiedades son resultados derivados de los axiomas. El siguiente lema enuncia las propiedades más comunes derivadas de los axiomas de probabilidad, que se usan repetidamente en los ejemplos para calcular probabilidades de diferentes eventos. La prueba de estas propiedades usa los conceptos que estudiamos en la sección introductoria sobre teoría de conjuntos.

### Lema 3.2.1 Propiedades

1. Sea  $A$  un evento, entonces  $P(A^c) = 1 - P(A)$ .
2. Si  $A = \emptyset$ , entonces  $P(A) = 0$ .
3. Si  $A \subset B$  son eventos entonces  $P(A) \leq P(B)$ .
4. Si  $A$  y  $B$  son eventos entonces  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

### Demostración:

1. Como  $A$  y  $A^c$  son eventos disjuntos con  $A \cup A^c = \Omega$ , resulta  $P(A) + P(A^c) = P(\Omega) = 1$ . Despejando resulta

$$P(A^c) = 1 - P(A)$$

2. También  $P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 1 - 1 = 0$ .

3. Si  $A \subset B$  entonces podemos escribir  $B = A \cup (B \cap A^c)$  la cual es una unión disjunta. Por lo cual  $P(B) = P(A) + P(B \cap A^c)$  y

$$P(A) = P(B) - P(B \cap A^c) \leq P(B)$$

4. Por último, observemos que  $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$  en forma disjunta. También,  $A = (A \cap B) \cup (A \cap B^c)$  en forma disjunta, y  $B = (A \cap B) \cup (B \cap A^c)$  por lo cual

$$\begin{aligned} P(A) + P(B) &= P(A \cap B) + P(B \cap A^c) + P(A \cap B) + P(A \cap B^c) \\ &= P(A \cup B) + P(A \cap B) \end{aligned}$$

y resulta

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

### Ejemplo 3.2.3

Una moneda está cargada de tal manera que la posibilidad de salir cara es dos veces la de salir seca. Calcular ambas probabilidades.

### Resolución

Sea  $p = P(\{\text{salió seca}\})$ . Entonces  $P(\{\text{salió cara}\}) = q = 2.p$ . Como ambos eventos son mutuamente excluyentes,  $1 = p + q = p + 2p = 3p$ , por lo cual  $p = \frac{1}{3}$  y  $q = \frac{2}{3}$ .

**Ejemplo 3.2.4** Si un experimento consiste en arrojar un dado honesto, ¿cuál sería una asignación de probabilidades razonable?.

**Resolución**

Primero, es razonable pensar que todos los valores 1,2,3,4,5,6 tienen igual probabilidad de salir, por lo cual sus probabilidades deberían ser iguales. Si denotamos con  $P$  a dicha asignación, entonces

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6} \quad (A)$$

¿Como debería de ser la asignación de probabilidad del evento  $E$  “salió un número par” o “salió un número menor que 5”? Si listamos el evento, entonces podemos encontrar su probabilidad sumando las probabilidades de cada uno de sus puntos

$$P(E) = P(\{2, 4, 6, 1, 3\}) = \frac{5}{6}$$

También podemos usar una de las propiedades que recién probamos,  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . Si  $A$  es “salió un número par” y  $B$  es “salió un número menor que 5”,  $P(A) = P(\{2, 4, 6\}) = \frac{3}{6}$ ,  $P(B) = P(\{1, 2, 3, 4\}) = \frac{4}{6}$  y  $P(A \cap B) = P(\{2, 4\}) = \frac{2}{6}$ , por lo cual  $P(E) = \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$ .

**Ejemplo 3.2.5**

Cuatro empresas hacen ofertas separadas sobre 3 contratos. Cada una de las empresas puede acceder a lo sumo a un contrato y cada contrato es diferente, por lo cual asignar el contrato  $C_1$  a la empresa  $A$  es diferente a asignar el contrato  $C_2$  a la empresa  $A$ .

1. ¿Cuántas formas de asignar contratos a empresas existen?
2. Si los contratos fueron asignados al azar, encuentre la probabilidad de que  $A$  obtenga algún contrato.

**Resolución**

Tenemos cuatro empresas distintas llamadas  $A, B, C$  y  $D$  y tres contratos distintos,  $C_1, C_2$  y  $C_3$ . La cantidad de asignaciones  $N$  es la cantidad de subconjuntos de tres empresas entre cuatro, (a esas les asigno algún contrato) multiplicado por las permutaciones de tres (pues a esas les puedo dar cualquiera de los tres contratos). Esto es

$$N = \binom{4}{3} 3! = 4! = 24$$

Como las asignaciones fueron hechas al azar, es razonable considerar que el espacio es equiprobable, ninguna de las asignaciones es más probable que otra, ninguna empresa tiene prioridad. Por lo cual la probabilidad de Laplace resulta la mejor regla de probabilidad y la probabilidad de cada asignación es

$$P(\omega) = 1/4! = 1/24.$$

Ahora, la probabilidad de asignarle a la empresa  $A$  algún contrato es mejor calcularla via el complemento. La probabilidad de que  $A$  no obtenga ningún contrato se calcula como el número de puntos del evento dividido 24, y si  $A$  no obtiene ningún contrato, todos los contratos los obtienen las otras tres empresas, por lo cual la cantidad de asignaciones posibles es una permutación de tres elementos,  $3!$ . Entonces

$$P(\text{ningun contrato}) = \frac{3!}{4!} = \frac{1}{4}.$$

Por lo cual, la probabilidad del complemento es

$$P(\text{algun contrato}) = 1 - \frac{1}{4} = \frac{3}{4}.$$

**Ejemplo 3.2.6**

Una clase consta de diez hombres y veinte mujeres. La mitad de cada grupo tiene ojos oscuros y la otra mitad claros. Hallar la probabilidad de que una persona elegida al azar sea hombre o tenga ojos oscuros.

**Resolución**

Comencemos realizando una tabla de doble entrada con los eventos considerados, hombre, mujer, ojos oscuros, ojos claros, y sus intersecciones. La información que tenemos es el cardinal de cada conjunto intersección, pues la mitad de cada grupo tiene ojos oscuros.

	hombre	mujer
ojos oscuros	5	10
ojos claros	5	10

Como elegimos una persona al azar todas las personas son igualmente posibles de ser elegidas y la probabilidad de Laplace, la uniforme, es la mas acertada. Por lo cual la probabilidad de cada evento considerado es el número de elementos del evento dividido el número de personas de la clase. Esto es

	hombre	mujer
ojos oscuros	5/30	10/30
ojos claros	5/30	10/30

Queremos calcular la probabilidad de que una persona elegida al azar sea hombre o tenga ojos oscuros, y como ambos eventos no son disjuntos, debemos usar la regla de la suma para conjuntos en general

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

De la tabla vemos que  $P(\text{hombre}) = P(\text{hombre con ojos oscuros}) + P(\text{hombre con ojos claros}) = 5/30 + 5/30 = 1/3$ ,  $P(\text{ojos oscuros}) = P(\text{hombre con ojos oscuros}) + P(\text{mujer con ojos oscuros}) = 5/30 + 10/30 = 1/2$  y  $P(\text{hombre con ojos oscuros}) = 5/30$ . Por lo cual

$$P(\text{hombre o ojos oscuros}) = P(\text{hombre}) + P(\text{ojos oscuros}) - P(\text{hombre con ojos oscuros}) = 1/3 + 1/2 - 5/30 = 20/30 = 2/3$$

Hasta ahora hemos trabajado con ejemplos donde hemos asignado solo la medida de probabilidad uniforme, es decir, si el espacio era por diseño equiprobable, como en el caso de las elecciones de elementos al azar, hemos asignado la probabilidad que cuenta el número de casos posibles sobre número de casos probables. Pero no todos los experimentos tienen como resultado puntos igualmente probables, ¿como asignamos probabilidades en este caso?. Si el espacio es finito, entonces veremos que con solo asignar probabilidades a los puntos muestrales basta para caracterizar toda la distribución de probabilidad.

**3.2.1. Espacios discretos: Cálculo de la probabilidad de un evento**

Usando las propiedades derivadas de los axiomas, y el hecho de que el conjunto de resultados posibles del experimento considerado es finito o numerable se puede probar el siguiente lema, que es muy útil a la hora de calcular probabilidades de eventos.

**Lema 3.2.2** Sea  $\Omega$  un espacio muestral discreto, es decir, un conjunto a lo sumo numerable. Un modelo probabilístico para un espacio muestral discreto  $\Omega$  se construye asignando probabilidades a cada punto de  $\Omega$ .

Esto es, si  $\Omega = \{\omega_1, \dots, \omega_n\}$ , definimos  $P(\{\omega_i\}) = p_i$ , con  $p_i \in [0, 1]$ , tal que  $\sum_{i=1}^n p_i = 1$ . También, para  $E$  un evento cualquiera de  $\Omega$ , definimos

$$P(E) = \sum_{\omega_i \in E} p_i$$

Entonces  $P$  es una probabilidad sobre  $\Omega$ . Es más, dada una sucesión de valores  $p_i \in [0, 1]$ ,  $1 \leq i \leq n$ , que cumplen  $\sum_{i=1}^n p_i = 1$ , existe una única  $P$  tal que  $P(\{\omega_i\}) = p_i$ , para todo  $i: 1, \dots, n$ .

Tratemos de entender que nos dice este lema. Tiene dos partes, en una nos sugiere como definir una medida de probabilidad y en la otra nos dice que esa medida definida es la única posible. Por ello nos da un método para definir medidas de probabilidad en los ejemplos de espacio muestral que hemos estado viendo.

Estudiemos esto un poco más. Supongamos que queremos conocer la probabilidad de un suceso  $A$ , que es un conjunto finito de posibles resultados de un experimento. El lema permite enunciar el **método de los puntos muestrales**, una sucesión de pasos a seguir para encontrar la probabilidad del evento:

1. Primero, se define el experimento concerniente al evento  $A$ .
2. Segundo se describe el espacio muestral  $\Omega = \{\omega_1, \dots, \omega_n\}$ , que contiene al evento  $A$  que es de interés.
3. Tercero se asigna a cada punto de  $\Omega$  una probabilidad adecuada, verificando que  $P(\omega_i) \geq 0$  y que  $P(\omega_1) + \dots + P(\omega_n) = 1$ .
4. Cuarto se define el evento en estudio como una colección específica de puntos muestrales.
5. Y por último se define  $P(A)$  como la suma de las probabilidades de los puntos del evento  $A$ .

### Ejemplo 3.2.7

Se tiene un dado cargado tal que la probabilidad de cada número es proporcional a ese número (es decir  $P(x) = kx$  con  $k$  fijo). Sean los eventos  $A = \text{"número par"}$ ,  $B = \text{"número primo"}$  (recordar que 1 no es primo), y  $C = \text{"número impar"}$ .

1. Decir como debe ser  $k$  para que  $P(x) = kx$  sea realmente una probabilidad y liste la probabilidad de cada punto del espacio muestral.
2. Hallar  $P(A)$ ,  $P(B \cup C)$ ,  $P(A \cap B^c)$
3. Hallar la probabilidad de que salga
  - I) un número par o primo
  - II) un número impar pero no primo

### Resolución

El espacio muestral del experimento "tiramos un dado cargado" es  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Las probabilidades de cada punto siguen la regla  $P(x) = kx$ , y como la suma de las probabilidades de los puntos de todo el espacio muestral debe ser 1, podemos obtener el valor de  $k$

$$1 = k1 + k2 + k3 + k4 + k5 + k6 = k(1 + 2 + 3 + 4 + 5 + 6) = k21$$

Despejando  $k$  resulta  $k = 1/21$  y

$$P(1) = \frac{1}{21}, P(2) = \frac{2}{21}, P(3) = \frac{3}{21}, P(4) = \frac{4}{21}, P(5) = \frac{5}{21}, P(6) = \frac{6}{21}$$

Si  $A$  = “número par”,  $B$  = “número primo”, y  $C$  = “número impar”, entonces

$$P(A) = P(\{2, 4, 6\}) = \frac{2}{21} + \frac{4}{21} + \frac{6}{21} = \frac{12}{21} = \frac{4}{7}$$

$$P(B \cup C) = P(\{1, 2, 3, 5\}) = \frac{1}{21} + \frac{2}{21} + \frac{3}{21} + \frac{5}{21} = \frac{11}{21}$$

$$P(A \cap B^c) = P(\{\text{un número par que no es primo}\}) = P(\{4, 6\}) = \frac{4}{21} + \frac{6}{21} = \frac{10}{21}$$

Ahora queremos saber la probabilidad de obtener el evento “un número par o primo”, esto es

$$P(A \cup B) = P(\{2, 3, 4, 5, 6\}) = 1 - \frac{1}{21} = \frac{20}{21}$$

La probabilidad del evento “un número impar pero no primo” es

$$P(C \cap B^c) = P(\{1\}) = \frac{1}{21}$$

### Ejemplo 3.2.8

Una urna tiene 6 bolas, 4 blancas numeradas del 1 al 4 y dos negras numeradas del 1 al 2. Asigne probabilidad a los experimentos

1. Saco dos bolas con repetición
2. Saco dos bolas sin reposición
3. Saco dos bolas al mismo tiempo

Asigne probabilidad al evento “ se sacó una bola blanca y una negra” dentro de cada experimento.

### Resolución

Sean  $b_1, b_2, b_3, b_4$  las bolas blancas y  $n_1, n_2$  las bolas negras. Entonces

1. Saco dos bolas con reposición tiene como espacio muestral a conjunto de pares ordenados

$$\Omega_1 = \{(x, y), \text{ donde } x, y \text{ pueden ser } b_1, b_2, b_3, b_4, n_1, n_2\} \quad \#\Omega_1 = 6 \times 6 = 36$$

La probabilidad de cada punto es

$$P((x, y)) = \frac{1}{\#\Omega_1} = \frac{1}{36}$$

En este experimento, la probabilidad del evento “ se sacó una bola blanca y una negra” se puede calcular multiplicando la cantidad de puntos del evento por  $1/36$ . Hay 8 puntos de la forma  $(b, n)$  y hay 8 puntos de la forma  $(n, b)$ , por lo cual hay 16 puntos con una bola blanca y una negra y resulta

$$P(A) = \frac{16}{36}$$

2. Saco dos bolas sin reposición tiene como espacio muestral al conjunto de pares ordenados

$$\Omega_2 = \{(x, y), \text{ donde } x, y \text{ pueden ser } b_1, b_2, b_3, b_4, n_1, n_2, x \neq y\} \quad \#\Omega_2 = 6 \times 5 = 30$$

La probabilidad de cada punto es

$$P((x, y)) = \frac{1}{\#\Omega_2} = \frac{1}{30}$$

En este experimento, la probabilidad del evento “ se sacó una bola blanca y una negra” se puede calcular multiplicando la cantidad de puntos del evento por  $1/30$ . Hay 8 puntos de la forma  $(b, n)$  y hay 8 puntos de la forma  $(n, b)$ , por lo cual hay 16 puntos con una bola blanca y una negra y resulta

$$P(A) = \frac{16}{30}$$

3. Saco dos bolas al mismo tiempo tiene como espacio muestral al conjunto de subconjuntos

$$\Omega_3 = \{\{x, y\}, \text{ donde } x, y \text{ pueden ser } b_1, b_2, b_3, b_4, n_1, n_2, x \neq y\} \quad \#\Omega_3 = \binom{6}{2} = 15$$

La probabilidad de cada punto es

$$P(\{x, y\}) = \frac{1}{\#\Omega_3} = \frac{1}{\binom{6}{2}} = \frac{1}{15}$$

En este experimento, la probabilidad del evento “ se sacó una bola blanca y una negra” se puede calcular multiplicando la cantidad de puntos del evento por 1/15. Hay solo 8 puntos de la forma (b,n) pues no importa el orden y resulta

$$P(A) = \frac{8}{15}$$

Los tres espacios son espacios muestrales discretos **equiprobables**. Esto quiere decir que todos sus puntos tienen igual probabilidad. Debemos notar también que no es necesario listar todos los puntos del espacio muestral, es importante describir este espacio, calcular cuantos puntos tiene y calcular cuantos puntos tiene el evento al que queremos asignar probabilidad. Si el espacio es equiprobable, la asignación de probabilidades que le corresponde tiene nombre, el del renombrado matemático francés Laplace.

**Ejemplo 3.2.9** Una urna tiene 6 bolas, 4 blancas indistinguibles y 2 negras indistinguibles. Asigne probabilidad a los experimentos

1. Saco dos bolas con reposición
2. Saco dos bolas sin reposición
3. Saco dos bolas al mismo tiempo

Asigne probabilidad al evento “ se sacó una bola blanca y una negra” dentro de cada experimento.

### Resolución

Ya habíamos señalado que el espacio muestral de los primeros dos experimentos es el mismo.

$$\Omega = \{(b, n), (n, b), (n, n), (b, b)\}$$

donde  $b$  y  $n$  denotan el color de la bola extraída y el par ordenado el orden de la extracción. Sin embargo, los experimentos son distintos, y la asignación de probabilidad en cada caso reflejar esas diferencias.

Tratemos de razonar como debería ser la probabilidad de cada punto en cada caso, siguiendo la intuición de que la proporción de bolas de cada clase en la urna debería estar relacionado. Eso fue lo que hicimos con los Rocklets, digimos que la probabilidad de sacar un Rocklet de la paquete del color elegido era la proporción de Rocklets de ese color en la paquete. Ahora tenemos dos extracciones de distinto tipo para estudiar.

1. En el caso de muestreo con reposición, donde volvemos a poner la bola en la urna, si queremos calcular la probabilidad del punto  $(b, n)$ , debemos observar que había cuatro bolas de color blanco de donde elegir en la primera extracción, y dos negras para la segunda extracción, por lo cual hay 8 posibilidades de elección de bolas que generan el mismo resultado  $(b, n)$ . Debemos observar también que al tener 6 bolas en la urna en cada extracción (pues devolvemos la bola extraída), hay 36 posibilidades de elegir un par de bolas cualesquiera. Como las elecciones se realizan al azar, cada elección es igualmente probable y es razonable asignar

$$P(b, n) = \frac{8}{36} = \frac{2}{9}$$

El razonamiento para los demás casos es análogo y resulta

$$P(n,b) = \frac{8}{36} \quad P(b,b) = \frac{16}{36} \quad P(n,n) = \frac{4}{36}$$

En este experimento, la probabilidad del evento “ se sacó una bola blanca y una negra” es la suma de las probabilidades de los puntos  $(b,n)$  y  $(n,b)$ , por lo cual

$$P(A) = \frac{8}{36} + \frac{8}{36} = \frac{16}{36}$$

2. En el caso de las extracciones sin reposición, el resultado de la primera extracción afecta las posibilidades de la segunda extracción. En el caso del punto  $(b,n)$ , tengo las mismas posibilidades , 3 blancas y dos negras, para extraer, pero la cantidad de pares distintos que puedo formar con 5 bolas, si extraigo sin reposición no es 36 sino  $6 \cdot 5 = 30$ . por lo cual

$$P(b,n) = \frac{8}{30}$$

El razonamiento para los demás casos es análogo y resulta

$$P(n,b) = \frac{8}{30} \quad P(b,b) = \frac{12}{30} \quad P(n,n) = \frac{2}{30}$$

En este experimento, la probabilidad del evento “ se sacó una bola blanca y una negra” es la suma de las probabilidades de los puntos  $(b,n)$  y  $(n,b)$ , por lo cual

$$P(A) = \frac{8}{30} + \frac{8}{30} = \frac{16}{30}$$

Para el tercer caso, la extracción conjunta, el espacio muestral es un conjunto de subconjuntos

$$\Omega = \{\{b,n\}, \{b,b\}, \{n,n\}\}$$

dado que no se distingue si se sacó primero una negra y después una blanca o fue al revés. Los subconjuntos indican que no hay distinción entre extracciones. La probabilidad del punto  $\{b,n\}$  es el cociente entre la cantidad de subconjuntos que puedo formar con cuatro bolas blancas y dos negras, donde hay solo una blanca y una negra, que son 8, y la cantidad de conjuntos de dos elementos que puedo formar con 6 elementos, que son 15.

$$P(\{b,n\}) = \frac{8}{15}$$

El razonamiento para los demás casos es análogo y resulta

$$P(\{b,b\}) = \frac{6}{15} \quad P(\{n,n\}) = \frac{1}{15}$$

En este experimento, el evento “ se sacó una bola blanca y una negra” es el punto  $\{b,n\}$  por lo cual su probabilidad es

$$P(A) = \frac{8}{15}$$

Cabe observar que estos espacios muestrales son conjuntos de eventos de los respectivos espacios muestrales generales que resultan de considerar las bolas distinguibles entre sí. A pesar de no estar numeradas, las bolas siempre son distintas, pero uno puede considerar un atributo como el color que colapse el espacio muestral general equiprobable en uno particular. Por eso que el evento “ se sacó una bola blanca y una negra” tiene igual probabilidad en ambos espacios (respetando el tipo de muestreo).

También es importante notar que arrojar un dado después del otro, y arrojar dos dados indistinguibles simultáneamente no son experimentos equivalentes a sacar bolas de urnas, y deben ser estudiados en forma particular.



**Ejemplo 3.2.10**

Un fabricante tiene cinco computadoras aparentemente idénticas listas para ser enviadas a destino. El no sabe que dos de ellas son defectuosas. Recibe un pedido especial de dos computadoras y lo surte seleccionando al azar dos de las cinco disponibles. Calcule las probabilidades de los siguientes eventos:

1. de que no entregue ninguna terminal defectuosa.
2. de que entregue una sola defectuosa (y la otra bien).
3. de que entregue ambas defectuosas.

**Resolución**

Denotemos con  $d$  si una computadora es defectuosa, y  $b$  si esta en buen estado. Podemos considerar cualquier punto muestral como un **subconjunto** de dos terminales seleccionadas para el envío, por lo cual el espacio muestral resulta

$$\Omega = \{\{b, d\}, \{b, b\}, \{d, d\}\}$$

Este espacio no es equiprobable, como vimos en el ejemplo anterior, por ello debemos ser cautelosos al calcular la probabilidad de los eventos. Observemos que los eventos pedidos son exactamente los puntos del espacio muestral, y revisando el ejemplo anterior vemos que

1.  $P(\text{ninguna defectuosa}) = P(\{\{b, b\}\}) = 3/10$ , donde el numerador representa la cantidad de subconjuntos de dos elementos que puedo formar con tres computadoras buenas, y el denominador la cantidad de subconjuntos de dos elementos que puedo formar con cinco computadoras.
2.  $P(\text{exactamente 1 defectuosa}) = P(\{\{b, d\}\}) = 6/10$ ,
3.  $P(\text{todas defectuosas}) = P(\{\{d, d\}\}) = 1/10$ .

**Ejemplo 3.2.11** Continuando con el ejemplo del fabricante, supongamos ahora que recibe dos pedidos consecutivos de una computadora cada uno, los de María y José y los surte seleccionando al azar dos de las cinco disponibles. Calcule la probabilidad de que no se entregue ninguna computadora defectuosa y la probabilidad de que el primer pedido, el de María, resulte una computadora defectuosa pero el José no.

**Resolución**

Denotemos con  $d$  si una computadora es defectuosa, y  $b$  si esta en buen estado. En este caso vamos a considerar **pares ordenados** (relacionados con el muestreo sin reposición, para conservar la información acerca del resultado de cada selección (computadora para el pedido de María, computadora para el pedido de José). En este caso el espacio muestral es más grande,

$$\Omega = \{(b, n), (n, b), (n, n), (b, b)\}$$

Este espacio no es equiprobable, pues de las 5 computadoras hay tres en buen estado y dos defectuosas, la probabilidad de sacar una buena debería ser mayor a la de sacar una defectuosa. Calculemos la probabilidad de no enviar ninguna defectuosa, es la probabilidad del punto  $(b, b)$ ,

1.  $P(\text{ninguna defectuosa}) = P((b, b)) = (3 \times 2)/20 = 6/20$
2.  $P(\text{la computadora de María es defectuosa y la de José no}) = P((d, b)) = (2 \times 3)/20 = 6/20$

En este ejemplo queremos tomar una decisión y para ello necesitamos conocer la probabilidad de un evento. Es nuestra primera aproximación a la inferencia usando probabilidades.

**Ejemplo 3.2.12**

Un conflicto laboral ha surgido con respecto a una supuesta distribución desigual de 20 trabajadores en cuatro diferentes trabajos de construcción. El primer trabajo, considerado detestable, requería de seis trabajadores, el segundo, el tercero y el cuarto requerían cuatro, cinco y cinco respectivamente. La controversia surgió con respecto a una supuesta distribución aleatoria de los trabajadores que asignó a los únicos cuatro miembros de un grupo étnico particular al primer trabajo. Un comité conciliatorio pidió conocer la probabilidad del evento observado para establecer si la asignación se hizo o no de manera injusta.

**Resolución**

El  $n^0$  de formas  $n$  en que podemos asignar los 20 trabajadores a los cuatro trabajos es igual al  $n^0$  de formas que se pueden separar 20 personas en cuatro grupos de cardinal 6, 4, 5 y 5 respectivamente. Entonces

$$n = \frac{20!}{6!4!5!5!}$$

La asignación aleatoria de trabajadores a diferentes tareas significa que los  $n$  puntos muestrales son equiprobables con probabilidad  $\frac{1}{n}$ .

Sea  $A$  el evento considerado y  $n_a$  el número de puntos muestrales de  $A$ , entonces la probabilidad de  $A$  es la suma de las probabilidades de los puntos muestrales de  $A$ ,

$$P(A) = \frac{n_a}{n}.$$

Calculemos  $n_a$ ; los cuatro del grupo étnico particular están asignados al primer trabajo, por lo tanto quedan 16 trabajadores para repartir en cuatro trabajos donde el primero tiene ahora 2 plazas y los restantes son iguales, por lo cual:

$$n_a = \frac{16!}{2!4!5!5!}$$

Entonces  $P(A) = \frac{n_a}{n} = 0,0031$ . Esta probabilidad es muy pequeña, por lo cual es razonable dudar que la asignación haya sido hecha al azar.

**Ejemplo 3.2.13**

Tiramos dos dados ¿Cuál es la probabilidad de que la suma sea 5?

**Resolución**

Consideremos el espacio muestral más completo del experimento, suponiendo que tenemos un dado negro y uno blanco, y que los resultados son pares ordenados donde la primera coordenada corresponde al dado negro y la segunda al blanco.

$$\Omega = \{(x_1, x_2) \text{ tal que } x_i = 1, 2, 3, 4, 5, 6, \text{ para } i = 1, 2\}$$

El espacio es equiprobable, y  $\#\Omega = 36$ , por lo cual si  $A = \{(x_1, x_2) \text{ tal que } x_1 + x_2 = 5\}$  resulta

$$P(A) = P(\{(1, 4), (2, 3), (3, 2), (4, 1)\}) = \frac{4}{36} = \frac{1}{9}$$

Es importante observar que es posible considerar los dados indistinguibles, por lo cual el espacio muestral resulta

$$\tilde{\Omega} = \{\{x_1, x_2\} \text{ con } x_i = 1, 2, 3, 4, 5, 6, i = 1, 2\}$$

Este espacio **no es equiprobable** por lo cual para calcular la probabilidad del evento  $A$  no basta saber cuantos puntos contiene. Sin embargo, la probabilidad del evento es la misma

$$P(A) = P(\{\{1, 4\}, \{2, 3\}\}) = \frac{2}{18} = \frac{1}{9}$$

pues la asignación de probabilidad de este espacio se calcula viendo cada punto **como un evento del espacio anterior**. Esto es,

$$P(\{x_1, x_2\}) = P(\{\text{salió un } x_1 \text{ y un } x_2\}) = P(x_1, x_2) + P(x_2, x_1) = 2/36$$

salvo que  $x_1 = x_2$ , en ese caso

$$P(\{x_1, x_1\}) = P(\{\text{salieron dos } x_1\}) = P(x_1, x_1) = 1/36$$

Un error muy común es considerar el espacio  $\tilde{\Omega}$  equiprobable. ¿Porqué no es esto razonable? En el siguiente ejemplo mostraremos si consideramos el espacio  $\Omega$  como equiprobable, dos eventos supuestamente con igual probabilidad (por tener igual cantidad de puntos), resultan tener diferente frecuencia relativa. Esto es, al realizar la experiencia muchas veces, un evento era observado más veces que el otro.

### Ejemplo 3.2.14

Un juego muy popular en Italia en el siglo XVII consistía en tirar tres dados y apostar a la suma a obtener. En esa época se creía que los sucesos

$$A = \{\text{la suma de los tres dados es } 9\} \quad B = \{\text{la suma de los tres dados es } 10\}$$

era la misma, ya que se podía obtener respectivamente como

suma 9	126	135	144	234	225	333
suma 10	145	136	226	235	244	334

Los dados eran iguales y balanceados, y la persona que los arrojaba no los manipulaba así que los resultados eran producto del azar. ¿Porqué un resultado iba a ser más probable que otro?. Todos los resultados tenían igual probabilidad y eventos con la misma cantidad de puntos eran obviamente equiprobables para los jugadores de la época. Sin embargo, la experiencia demostraba que el 10 salía un poco más frecuentemente que el 9.

Le consultaron entonces a Galileo sobre esta contradicción, que razonó como sigue:

Supongamos que en vez de usar tres dados blancos usamos un dado blanco, uno negro y uno gris, entonces el espacio correspondiente es

$$\Omega = \{(x_1, x_2, x_3) \text{ tal que } x_i = 1, 2, 3, 4, 5, 6, \text{ para } i = 1, 2, 3\}$$

donde  $x_1$  es el resultado del dado blanco,  $x_2$  el del dado negro y  $x_3$  el del dado gris. Por las mismas razones que argumentaban los jugadores de la época, el espacio es equiprobable, ningún resultado debería tener mas probabilidad de salir que otro. Entonces, ¿Cual es la probabilidad de  $A$  y  $B$ ?. Para calcularlo basta contar cuantas ternas hay en  $A$  y en  $B$ . Ternas que sumen 9 son las 25 siguientes, por lo cual  $P(A) = 25/36 = 0,116$

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

Ternas que sumen 10 son las 27 siguientes por lo cual  $P(B) = 27/36 = 0,125$ .

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

Por lo cual  $P(B) > P(A)$  (pero no mucho), como se había observado en las repeticiones del experimento.

¿Cuál es el error entonces? Considerar que al ser los dados físicamente indistinguibles (todos blancos, por ejemplo), el  $\{1, 2, 3\}$  sale las mismas veces que el  $\{1, 1, 1\}$  o el  $\{1, 2, 1\}$ , lo cual no es cierto. El conjunto  $\{1, 2, 3\}$  engloba los puntos  $(1, 2, 3), (2, 1, 3), (3, 2, 1), (2, 3, 1), (1, 3, 2), (3, 1, 2)$ , pues el dado de la izquierda puede ser 1 o el del medio o el de la derecha, a pesar de ser todos blancos, la distinción siempre existe. Por eso es que la frecuencia relativa del evento  $A$  era más baja que la del evento  $B$ .

**Ejemplo 3.2.15** Un cajón tiene 3 pares de medias negras, 4 pares de medias azules y 5 pares de medias verdes. Si todas las medias están sueltas y mezcladas en el cajón, y se sacan dos medias en el oscuro:

1. Encuentre el espacio muestral resultante y asigne probabilidades a cada punto
2. Calcule la probabilidad de que ambas medias sean del mismo color.
3. Calcule la probabilidad de que al menos una media sea negra.

### Resolución

1. El espacio muestral corresponde al de bolas sacadas al mismo tiempo, son subconjuntos de medias, no importa si se saca la media verde primero y la azul después, importa que se sacó una media verde y una azul. Tampoco importa que media azul se extrajo del cajón, salvo el color, son todas iguales. Por lo cual el espacio más razonable resulta

$$\Omega = \{\{n, n\}, \{n, v\}, \{n, a\}, \{a, v\}, \{a, a\}, \{v, v\}\}$$

donde  $n, v, a$  indican el color de la media. El espacio **no es equiprobable**, pues hay mas medias de un color que de otro, pero es razonable asignar la probabilidad de Laplace, número de casos posibles sobre número de casos probables **a cada punto del espacio muestral**. Recordemos que una vez asignadas las probabilidades a los puntos, las probabilidades de los eventos son suma de estas probabilidades. Por ejemplo, si sacamos dos medias negras, los casos posibles serán todas las formas de elegir dos medias de un total de 6, (3 pares), que es el número de subconjuntos de dos elementos tomados de un total de seis. El número de casos probables es el número de subconjuntos de dos elementos de un total de 24 medias (3 pares negros, 4 pares azules y 5 pares verdes son 24 medias en total).

$$P(\{n, n\}) = \frac{\binom{6}{2}}{\binom{24}{2}}$$

Siguiendo el mismo razonamiento, la probabilidad de obtener una media negra y una verde se calcula dividiendo el numero de casos posibles sobre los probables, y los posibles son todas las formas de elegir una media negra entre 6 por todas las formas de elegir una media verde entre 10 medias.

$$P(\{n, v\}) = \frac{\binom{6}{1} \binom{10}{1}}{\binom{24}{2}}$$

Los restantes son

$$P(\{n, a\}) = \frac{\binom{6}{1} \binom{8}{1}}{\binom{24}{2}}, \quad P(\{a, v\}) = \frac{\binom{8}{1} \binom{10}{1}}{\binom{24}{2}}, \quad P(\{a, a\}) = \frac{\binom{8}{2}}{\binom{24}{2}}, \quad P(\{v, v\}) = \frac{\binom{10}{2}}{\binom{24}{2}}$$

2. Par a calcular la probabilidad de que ambas medias sean del mismo color sumo las probabilidades de los puntos muestrales del evento

$$P(A) = P(\{n, n\}, \{a, a\}, \{v, v\}) = \frac{\binom{6}{2}}{\binom{24}{2}} + \frac{\binom{8}{2}}{\binom{24}{2}} + \frac{\binom{10}{2}}{\binom{24}{2}}$$

3. Para calcular la probabilidad de que alguna media sea negra podemos hacer dos cosas, una de ellas es sumar la probabilidad de los eventos, “hay exactamente una media negra” y “hay exactamente dos medias negras”:

$$P(A) = P(\{n, n\}, \{n, a\}, \{n, v\}) = \frac{\binom{6}{2}}{\binom{24}{2}} + \frac{\binom{6}{1}\binom{8}{1}}{\binom{24}{2}} + \frac{\binom{6}{1}\binom{10}{1}}{\binom{24}{2}}$$

La otra es calcular la probabilidad del evento “ninguna media es negra” y restar ese valor a uno

$$P(A) = 1 - P(\text{ninguna media es negra}) = 1 - \left[ \frac{\binom{8}{1}\binom{10}{1}}{\binom{24}{2}} + \frac{\binom{8}{2}}{\binom{24}{2}} + \frac{\binom{10}{2}}{\binom{24}{2}} \right]$$

**Ejemplo 3.2.16** Supongamos tirar un dardo al azar en un blanco cuadrado de lado 1. ¿Cuál es la probabilidad de que caiga en el triángulo acotado por  $x = 0$ ,  $y = 0$  y  $x + y = 1$ ?

### Resolución

Observemos que el triángulo denotado por las rectas  $(x = 0, y \in \mathbb{R})$ ,  $(y = 0, x \in \mathbb{R})$  y  $(y = 1 - x, x \in \mathbb{R})$  es la mitad del cuadrado de lado 1, por lo cual su área es la mitad del área del cuadrado. Entonces

$$P(A) = \frac{1}{2}$$

### Ejemplo 3.2.17

Supongamos tirar un dardo al azar en un blanco cuadrado de lado 2, que podemos suponer es el cuadrado centrado en  $(0, 0)$  con vértices en  $(0, 1)$ ,  $(1, 0)$ ,  $(-1, 0)$ ,  $(0, -1)$ . ¿Cuál es la probabilidad de que caiga en el círculo de radio 1 centrado en el  $(0, 0)$ ?

### Resolución

El área del cuadrado de lado 2 es 4, y el área del círculo de radio 1 es  $\pi$ . Entonces

$$P(A) = \frac{\pi}{4}$$

sin necesidad de integrar.

### Ejemplo 3.2.18

Supongamos tirar un dardo al azar en un blanco redondo de radio 1, que podemos suponer centrado en el  $(0, 0)$ . ¿Cuál es la probabilidad de que caiga en sector angular de 0 a  $\pi/4$  radianes?

### Resolución

El área del sector angular es  $A = \frac{1}{2}r^2\frac{\pi}{4}$  es  $1/8$  del área del círculo de radio  $r = 1$ , y el área del círculo de radio 1 es  $\pi$ . Entonces

$$P(A) = \frac{\pi}{8} \cdot \frac{1}{\pi} = \frac{1}{8}$$

### 3.3. Probabilidad condicional

Volvamos a pensar el problema de repetir un experimento un número grande de veces denotado por  $N$ . Llamemos  $n_a, n_b, n_{a,b}$  al número de veces que se observaron los eventos  $A, B$  y  $A \cap B$  respectivamente. Si  $N$  es suficientemente grande entonces la asignación de probabilidades frecuentista debería cumplir

$$P(A) \approx \frac{n_a}{N} \quad P(B) \approx \frac{n_b}{N} \quad P(A \cap B) \approx \frac{n_{a,b}}{N}$$

Supongamos ahora que solo anotamos los experimentos donde  $A$  ocurre, entonces tenemos  $n_a$  experimentos, dentro de los cuales  $B$  ocurre  $n_{a,b}$  veces. Entonces la proporción de veces que  $B$  ocurre dentro de esos  $n_a$  experimentos es  $n_{a,b}/n_a$  y

$$\frac{n_{a,b}}{n_a} = \frac{n_{a,b}/N}{n_a/N} \approx \frac{P(A \cap B)}{P(A)}$$

por lo cual, desde el punto de vista frecuentista, la probabilidad de que ocurra  $B$  dado que ocurrió el evento  $A$  (esto es, restringiendo mi universo al conjunto  $A$ ), debería ser

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

Esta entonces es la forma de **definir** la probabilidad condicional, que es una asignación de probabilidades legítima sobre todos los eventos intersección con el evento  $A$ .

**Definición 3.3.1** Sea  $A$  un evento de un espacio de probabilidad, tal que  $P(A) \neq 0$ .

- Para todo  $B$  evento en  $\Omega$ , la probabilidad condicional del evento  $B$  dado que ocurrió el evento  $A$  es igual a

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

y  $P(B/A)$  es una asignación de probabilidades legítima sobre  $\Omega$ .

**Ejemplo 3.3.1** La siguiente tabla se obtuvo de los datos del censo poblacional de Estados Unidos en 1970 (los datos están expresados en miles de personas).

Area	Población de los EEUU en 1970		
	Blancos	Negros y otros	Total
<b>Urbana</b>	128.773	20.552	149.325
• dentro de las áreas urbanizadas	100.952	17.495	118.447
a) centro de la urbe	49.547	14.375	63.922
b) suburbios	51.405	3.120	54.525
• fuera de las áreas urbanizadas	27.821	3.057	30.878
<b>Rural</b>	48.976	4.911	53.887
<b>Total</b>	177.749	25.463	203.312

Fuente The World Almanac and Book of Facts, 1979 Edition, copyright Newspaper Enterprise Association Inc, 1978, New York, NY 10166.

Si se selecciona al azar una persona de la población de Estados Unidos, encuentre las probabilidades, basadas en los datos de 1970, de los siguientes casos.

- a) que la persona sea de raza blanca.

- b) que la persona viva en el centro de la ciudad.
- c) que la persona viva en los suburbios dado que es de raza blanca.
- d) que la persona sea de raza blanca dado que vive en los suburbios.
- e) que la persona viva fuera del área urbanizada, dado que no es de raza blanca.
- f) que la persona no sea de raza blanca y viva en el centro de una ciudad o que sea de raza blanca y viva fuera del área urbanizada.

**Resolución**

Vamos a suponer que si traducimos la tabla a frecuencias relativas (dividiendo cada entrada por el total de casos en la muestra) estamos acercándonos al verdadero valor de la probabilidad de cada uno de los casos, dado que estos valores se obtuvieron de un censo (y la cantidad de personas que no fue relevada puede considerarse no significativa).

La tabla siguiente muestra las probabilidades aproximadas por la frecuencia relativa

Area	Población de los EEUU en 1970		
	Blancos	Negros y otros	Total
<b>Urbana</b>	0.633	0.101	0.734
• dentro de las áreas urbanizadas	0.496	0.086	0.582
a) centro de la urbe	0.243	0.071	0.314
b) suburbios	0.253	0.015	0.268
• fuera de las áreas urbanizadas	0.137	0.015	0.152
<b>Rural</b>	0.242	0.024	0.266
<b>Total</b>	0.875	0.125	1

- a) Entonces

$$P(\text{la persona sea de raza blanca}) \approx \frac{177749}{203312} = 0,875$$

- b) Sea  $A = \{\text{la persona que vive en el centro de ciudad}\}$ , por lo tanto:

$$P(A) = 0,314$$

- c) Sea  $B = \{\text{la persona vive en los suburbios}\}$ ,  $C = \{\text{la persona es de raza blanca}\}$ , entonces:

$$P(B/C) = \frac{P(B \cap C)}{P(C)} = \frac{0,256}{0,875} = 0,292$$

- d)

$$P(C/B) = \frac{P(C \cap B)}{P(B)} = \frac{0,256}{0,265} = 0,94$$

- e) Sea  $D = \{\text{la persona vive fuera de las áreas urbanizadas}\}$ ,  $B^c = \{\text{la persona no es de raza blanca}\}$  entonces:

$$P(D/B^c) = \frac{P(D \cap B^c)}{P(B^c)} = \frac{0,015}{0,125} = 0,12$$

- f) Sea  $B^c \cap A = \{\text{la persona no es de raza blanca y vive en el centro de una ciudad}\}$ ,  
 $B \cap D = \{\text{la persona es de raza blanca y vive fuera de las áreas urbanizadas}\}$  entonces

$$P((B^c \cap A) \cup (B \cap D)) = 0,071 + 0,137 = 0,208$$

**Ejemplo 3.3.2** Supongamos que una caja contiene 4 bolas rojas numeradas 1, 2, 3, 4 y 3 bolas negras numeradas 1, 2, 3. Si elegimos una bola al azar, y observamos que es roja, cual es la probabilidad de que tenga el número 1?

**Resolución**

Sea  $A$  el evento saque una bola roja, y  $B$  el evento saqué una bola con el numero 1, quiero calcular  $P(B/A)$ , la probabilidad de que tenga el número 1 dado que es roja. Entonces, como la bola fue sacada al azar, la probabilidad de cualquier de ser escogida es  $1/(4+3) = 1/7$ , y resulta

$$P(A) = \frac{4}{7} \quad P(A \cap B) = \frac{1}{7} \quad P(B/A) = \frac{A \cap B}{A} = \frac{1}{4}$$

Estos resultados son bien razonables, dado que si solo miro las bolas rojas, tengo 4 bolas con igual chance de ser extraídas de la caja, por lo cual la probabilidad de sacar una bola (la número 1 o cualquier otra) debe ser  $1/4$ . Sin embargo, la probabilidad de  $B$  es  $2/7$  pues tengo dos bolas con el número 1 en la caja.

**Observación 3.3.1** El experimento general no es el mismo que el condicional, de ahí que la asignación de probabilidades cambie. También, en algunas situaciones es más fácil obtener  $P(B/A)$  mirando el experimento condicional, por lo cual  $P(A \cap B)$  puede calcularse despejando la fórmula de la probabilidad condicional.

**Lema 3.3.1** *Ley de multiplicación.*

Sea  $A$  y  $B$  eventos y asumamos que  $P(A) \neq 0$ . Entonces

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

**Demostración:**

Despejando de la definición de esperanza condicional sale el resultado.

**Ejemplo 3.3.3** Una urna contiene tres bolas rojas y una bola azul. Dos bolas son seleccionadas **sin reposición** y su color anotado. ¿Cual es el espacio muestral del experimento? ¿Es equiprobable? ¿Cuál es la probabilidad de que ambas bolas seleccionadas sean rojas?

**Resolución**

El espacio muestral es

$$\Omega = \{(r, r), (r, a), (a, r)\}$$

y no es un espacio equiprobable, dado que puedo observar la sucesión rojo y azul cuatro veces con bolas distintas, pero puedo observar la sucesión rojo y rojo seis veces con bolas distintas.

Elaboremos un poco más, una forma de calcular la probabilidad del punto  $(r, r)$  es considerar cuantos pares de bolas rojas puedo formar con 3 bolas rojas (distintas), si no repongo la bola que saqué primero. Puedo formar 6 pares. Luego considero cuantos pares puedo formar, sin reposición, con 4 bolas distintas, los cuales son 12 pares. Como los pares fueron elegidos al azar,

$$P(r, r) = \frac{6}{12} = \frac{1}{2}$$

De esta forma estoy calculando  $P(r, r)$  como un evento del espacio equiprobable que resulta de seleccionar dos bolas sin reposición de una urna con 3 bolas rojas numeradas y una azul.



Sin embargo, es más intuitivo calcular estas probabilidades usando eventos condicionales. Sean  $R_1$  y  $R_2$  eventos tales que la bola roja se selecciona en la primera extracción y en la segunda extracción, respectivamente. De la regla de multiplicación

$$P(R_1 \cap R_2) = P(R_1)P(R_2/R_1)$$

La probabilidad de  $R_1$  es claramente  $3/4$ , pues tengo cuatro bolas, tres de ellas rojas. Ahora, si una de las bolas rojas fue removida en la primera extracción, quedan dos bolas rojas y una azul en la urna, por lo cual cambio mi experimento y la probabilidad de  $R_2/R_1$ , es decir sacar una bola roja dado que cambio mi experimento es ahora  $2/3$ . Por lo cual

$$P(R_1 \cap R_2) = \frac{3}{4} \cdot \frac{2}{3} = \frac{1}{2}$$

En este ejemplo solo hicimos dos extracciones, pero si tenemos 100 bolas y hacemos 10 extracciones, la forma más simple de calcular probabilidades es usando la regla condicional.

**Ejemplo 3.3.4** Supongamos que una caja contiene 4 bolas rojas numeradas del 1, 2, 3, 4 y 3 bolas negras numeradas 1, 2, 3. Si sacamos dos bolas al azar sin reposición y observamos el color y el número de cada una ¿Cual es la probabilidad de sacar una bola roja en la segunda extracción dado que observé un 1 en la primera extracción?

**Resolución**

Sea el evento  $R_i$  es igual a observar una bola roja en la  $i$ -ésima extracción,  $N_i$  es igual a observar una bola negra en la  $i$ -ésima extracción y  $B_i$  el evento salió un uno en la  $i$ -ésima extracción. Entonces la probabilidad de sacar roja en la segunda dado que observé un uno en la primera es

$$\begin{aligned} P(R_2/B_1) &= \frac{P(R_2 \cap B_1)}{P(B_1)} \\ &= \frac{P(R_2 \cap B_1 \cap N_1) + P(R_2 \cap B_1 \cap R_1)}{P(B_1)} \\ &= \frac{P(R_2/(B_1 \cap N_1))P(B_1 \cap N_1) + P(R_2/(B_1 \cap R_1))P(B_1 \cap R_1)}{P(B_1)} \end{aligned}$$

Calculemos cada una de estas probabilidades, teniendo en cuenta que al condicionar, estoy pensando en el experimento restringido. Empecemos con  $P(R_2/(B_1 \cap N_1))$ , si observé una bola negra en la primera extracción, para la segunda extracción sé que estan todas las bolas rojas en la urna, y tengo 6 bolas en total en la urna pues ya saqué una, por lo cual

$$P(R_2/(B_1 \cap N_1)) = \frac{4}{4+3-1} = \frac{2}{3};$$

Pensemos ahora  $P(R_2/(B_1 \cap R_1))$ , si salió la bola roja con el número 1 en la primera extracción, para la segunda extracción solo tengo 3 bolas rojas, y hay seis bolas en la urna, por lo cual

$$P(R_2/(B_1 \cap R_1)) = \frac{4-1}{4+3-1} = \frac{1}{2}$$

Los restantes eventos solo se refieren a la primera extracción, por lo cual los calculo pensando solo en ese experimento

$$P(B_1 \cap N_1) = \frac{1}{4+3} = \frac{1}{7}; \quad P(B_1 \cap R_1) = \frac{1}{4+3} = \frac{1}{7}; \quad P(B_1) = \frac{2}{7}$$

Reemplazando lo que obtuve en la fórmula original

$$P(R_2/B_1) = \frac{\frac{2}{3} \frac{1}{7} + \frac{1}{2} \frac{1}{7}}{\frac{2}{7}} = \frac{7}{12}$$

En los ejemplos anteriores el espacio de probabilidad estaba especificado y todas las probabilidades de los eventos, condicionales o no podían ser calculadas aplicando la definición o el método de los puntos muestrales. En muchos problemas, sin embargo, se procede en la dirección opuesta. Lo que es conocido son las probabilidades condicionales de algunos eventos y se usa esa información para establecer el espacio de probabilidad general.

**Ejemplo 3.3.5** Supongamos que la población de una ciudad está constituida por un 40 % de hombres y un 60 % de mujeres. Suponga también que el 50 % de los hombres y el 30 % de las mujeres fuma. Si se elige una persona al azar de dicha población y se registra su género y si es fumador o no. Encuentre las probabilidades del espacio muestral.

**Resolución**

El espacio muestral del experimento es

$$\Omega = \{(H, F), (H, NF), (M, F), (M, NF)\}$$

Observemos que  $P(\text{Fumador} / \text{Hombre}) = 0,5$ ,  $P(\text{Fumador} / \text{Mujer}) = 0,3$  y  $P(\text{Mujer}) = 0,6$ ,  $P(\text{Hombre}) = 0,4$ . Entonces  $P(\text{Mujer que es fumadora}) = P(\text{Fumador} / \text{Mujer}) \cdot P(\text{Mujer}) = 0,3 \cdot 0,6 = 0,18$  y  $P(\text{Hombre que es fumador}) = P(\text{Fumador} / \text{Hombre}) \cdot P(\text{Hombre}) = 0,5 \cdot 0,4 = 0,20$

	Mujer	Hombre	Total
Fumador	0.18	0.2	<b>0.38</b>
No Fumador	<b>0.42</b>	<b>0.2</b>	<b>0.62</b>
Total	0.6	0.4	1

Las probabilidades en negrita se calcularon usando la aditividad de la ley de probabilidad si los conjuntos son disjuntos.

**Ejemplo 3.3.6** Un estudio de la conducta después del tratamiento de un gran número de drogadictos, sugiere que la probabilidad de reincidencia dentro de los dos años siguientes al tratamiento podía depender de la educación del trasgresor. Las proporciones del número total de casos que caen dentro de cuatro categorías de educación-reincidencia se presentan a continuación:

Educación	Condición dentro del período de dos años después del tratamiento		Totales
	Reincidente	No reincidente	
10 años o más	0.10	0.30	0.40
9 años o menos	0.27	0.33	0.60
Totales	0.37	0.63	1.00

Supóngase que se selecciona un único transgresor del programa de tratamiento. Encuentre las probabilidades de los eventos

A: el transgresor tiene diez años o más de educación.

B: el transgresor reincide dentro del período de los dos años siguientes al tratamiento.

C: A dado que ocurrió el evento B

D: B dado que ocurrió el evento A.

**Resolución**

Las probabilidades de los eventos A y B son las proporciones totales que se encuentran en los márgenes de las tablas,  $P(A) = 0,4$ ,  $P(B) = 0,37$ . La probabilidades condicionales se calculan aplicando la definición,  $P(A|B) = P(A \cap B) / P(B) = 0,10 / 0,37 = 0,27$ , y  $P(B|A) = P(A \cap B) / P(A) = 0,10 / 0,4 = 0,25$

**Lema 3.3.2** Ley de probabilidad total.

Sean  $A_1, \dots, A_n$  una partición de  $\Omega$ , una colección de conjuntos disjuntos cuya unión es  $\Omega$ , tales que  $P(A_i) > 0$ , para todo  $i$ . Entonces todo evento  $B$  de  $\Omega$  cumple

$$P(B) = \sum_{i=1}^n P(B/A_i)P(A_i)$$

**Demostración:**

Observemos primero que

$$P(B) = P(B \cap \Omega) = P(B \cap (\cup_{i=1}^n A_i)) = P(\cup_{i=1}^n (B \cap A_i))$$

Como los eventos  $(B \cap A_i)$  son disjuntos

$$P(\cup_{i=1}^n (B \cap A_i)) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B/A_i)P(A_i)$$

Este lema nos dice que si conocemos como dato las probabilidades de una partición del conjunto total, la probabilidad de un evento cualquiera la podemos calcular viendo como es ese evento dentro de cada elemento de la partición. vamos a usar muchas veces esta regla en los ejercicios, cuando tengamos diversas características, como color y número y escritorios y monedas.

**Ejemplo 3.3.7** En un ejemplo anterior, consideramos una urna que contiene tres bolas rojas y una bola azul (sin numerar). Dos bolas son seleccionadas **sin reposición**. ¿cuál es la probabilidad de sacar una bola roja en la segunda extracción?

**Resolución**

En este ejemplo es simple armar el espacio muestral, pero no es equiprobable. Por lo cual es más fácil calcular esta probabilidad directamente usando la regla de probabilidad total y el diseño del experimento.

En la primera extracción puedo sacar una bola roja o una azul, siendo ambos eventos excluyentes. Sea  $R_i$  el evento *bola roja en la  $i$ -ésima extracción* y  $A_i$  el evento *bola azul en la  $i$ -ésima extracción*. Entonces  $\Omega = R_1 \cup A_1$ , y como ambos son obviamente disjuntos, aplicando la regla de probabilidad total resulta

$$P(R_2) = P(R_1)P(R_2/R_1) + P(A_1)P(R_2/A_1) = \frac{3}{4} \cdot \frac{2}{3} + \frac{1}{4} \cdot \frac{3}{3} = \frac{3}{4}$$

**Teorema 3.3.1** Regla de Bayes Sea  $A_1, \dots, A_n$  una partición de  $\Omega$  con  $P(A_i) > 0$ , y  $B$  un evento cualquiera. Entonces

$$P(A_j/B) = \frac{P(B/A_j)P(A_j)}{\sum_{i=1}^n P(B/A_i)P(A_i)}$$

**Demostración:**

Observemos primero que

$$P(B/A_j)P(A_j) = P(A_j \cap B) = P(A_j/B)P(B)$$

por la definición de probabilidad condicional. Por lo cual

$$P(A_j/B) = \frac{P(B/A_j)P(A_j)}{P(B)}$$

Ahora, por la regla de probabilidad total,

$$P(B) = \sum_{i=1}^n P(B/A_i)P(A_i)$$

entonces

$$P(A_j/B) = \frac{P(B/A_j)P(A_j)}{\sum_{i=1}^n P(B/A_i)P(A_i)}$$

En los ejemplos siguientes vamos a notar que la resolución es compleja, pero el enunciado, bolas sacadas de una urna, es muy simple. Es más, las experiencias pueden realizarse varias veces para verificar que las probabilidades obtenidas son parecidas a las frecuencias relativas, verificando que hemos modelado correctamente el experimento. En cuanto empecemos a enunciar ejemplos más reales (aunque todos hayan sido simplificados), la comprensión del enunciado va a agregar una dificultad más a la ya difícil resolución de los experimentos condicionales. Los psicólogos, biólogos y médicos usan mucho los experimentos condicionales, y el diseño de éstos para que puedan resolverse de la misma forma que los ejemplos de escritorios, y bolas en urnas son estudiados cuidadosamente.

**Ejemplo 3.3.8** En una urna hay 3 bolas azules, 4 blancas y 7 rojas. Se extrae una bola. Si es blanca no se repone; si es roja se vuelve a poner en la urna y si es azul se devuelve a la urna y se agrega una azul mas. A continuación se extrae otra bola.

1. Hallar la probabilidad de sacar una bola azul en la segunda extracción.
2. Si la segunda bola es azul, que color tuvo mayor probabilidad de haber salido en la primera extracción?
3. Hallar la probabilidad de no sacar azul en ninguna de las dos extracciones.

### Resolución

En este caso es más obvio aun que conviene resolver las preguntas usando información condicional en vez de pasar por el espacio muestral.

Sea  $B_1, R_1, A_1$  los eventos se seleccionó en la primera extracción una bola blanca, una bola roja, una bola azul, respectivamente. Sean  $B_2, R_2, A_2$  los eventos se seleccionó en la segunda extracción una bola blanca, una roja y una azul, respectivamente.

Entonces,

$$\begin{aligned} P(A_2) &= P(B_1 \cap A_2) + P(R_1 \cap A_2) + P(A_1 \cap A_2) \\ &= P(A_2/B_1)P(B_1) + P(A_2/R_1)P(R_1) + P(A_2/A_1)P(A_1) \\ &= \frac{3}{13} \frac{4}{14} + \frac{3}{14} \frac{7}{14} + \frac{4}{15} \frac{3}{14} = \frac{3}{14} \left[ \frac{4}{13} + \frac{7}{14} + \frac{4}{15} \right] = 0,2302198 \end{aligned}$$

$$P(B_1/A_2) = \frac{P(A_2/B_1)P(B_1)}{P(A_2)} = \frac{\frac{3}{13} \frac{4}{14}}{0,2302198} = 0,2863963$$

$$P(R_1/A_2) = \frac{P(A_2/R_1)P(R_1)}{P(A_2)} = \frac{\frac{3}{14} \frac{7}{14}}{0,2302198} = 0,4653939$$

$$P(A_1/A_2) = \frac{P(A_2/A_1)P(A_1)}{P(A_2)} = \frac{\frac{4}{15} \frac{3}{14}}{0,2302198} = 0,2482102$$

$$P(A_1^c \cap A_2^c) = 1 - P(A_1) - P(A_2) + P(A_1 \cap A_2) = 1 - \frac{3}{14} - 0,2302198 + \frac{4}{15} \frac{3}{14} = 0,5554945 + 0,0571429 = 0,6126374$$

**Ejemplo 3.3.9** Supongamos que tenemos tres escritorios con dos cajones cada uno. El primer escritorio tienen una moneda de oro en cada cajón, el segundo escritorio tiene una moneda de oro en un cajón y una de plata en el otro, y el tercer escritorio tiene una moneda de plata en cada cajón. Si se elige al azar un escritorio y se abre uno de los cajones encontrando una moneda de oro.

1. ¿Cuál es la probabilidad de que el otro cajón tenga una moneda de oro también?
2. ¿Cuál es la probabilidad de que el otro cajón tenga una moneda de de plata?

### Resolución

Para responder estas preguntas, primero tenemos que ponerle nombre a los eventos y ordenar los datos que tenemos. Sean  $A_1, A_2$ , y  $A_3$  los eventos “elegimos el escritorio 1, el 2, el 3”, respectivamente. Estos eventos forman una partición del espacio muestral, dado que se elige un solo escritorio, y como la elección se hace al azar, la probabilidad de cada evento es  $P(A_i) = 1/3, i = 1, 2, 3$ . Sea  $B$  el evento “la moneda observada es de oro”. Entonces, recordando que monedas tiene cada escritorio es claro que

$$P(B/A_1) = 1 \quad P(B/A_2) = 1/2 \quad P(B/A_3) = 0$$

El problema es encontrar la probabilidad de que la segunda moneda del escritorio sea de oro si la primera observada lo era. Como el único de los escritorios que cumple esta propiedad es el primero, lo que queremos saber es  $P(A_1/B)$ , es decir, la probabilidad de haber elegido el primer escritorio dado que observamos una moneda de oro.

$$P(A_1/B) = \frac{P(B/A_1) \cdot P(A_1)}{P(B/A_1) \cdot P(A_1) + P(B/A_2) \cdot P(A_2) + P(B/A_3) \cdot P(A_3)} = \frac{2}{3}$$

En el segundo caso lo que queremos saber es, si eligiendo al azar un escritorio y un cajón observamos una moneda de oro, ¿cual es la probabilidad de haber escogido el segundo escritorio?

$$P(A_2/B) = \frac{P(B/A_2) \cdot P(A_2)}{P(B/A_1) \cdot P(A_1) + P(B/A_2) \cdot P(A_2) + P(B/A_3) \cdot P(A_3)} = \frac{1}{3}$$

Como dijimos, es esencial ordenar los datos que tenemos y pensar que nos están preguntando.

**Ejemplo 3.3.10** Un director de personal dispone de dos listas de solicitantes para trabajos. La lista 1 incluye los nombres de 3 hombres y de 4 mujeres, mientras que la lista 2 incluye los nombres de 7 hombres y de 3 mujeres. Se selecciona un nombre al azar de la lista 1 para incorporarla a la lista 2. A continuación se selecciona al azar un nombre de la lista 2 (ya aumentada). Dado que el nombre seleccionado es de un hombre, ¿cuál es la probabilidad de que se haya seleccionado el nombre de una mujer de la lista 1?

### Resolución

Sea  $A$  = Se seleccionó una mujer de la lista 1,  $B$  = se seleccionó un hombre de la lista 1, y  $C$  = se seleccionó un hombre de la lista 2. Usando el lema de Bayes

$$P(A|C) = \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C|B)P(B)}$$

Calculemos ahora las probabilidades que necesitamos. Si se eligió una mujer de la lista 1, y se agregó ese nombre a la lista 2,  $P(C|A) = \frac{7}{11}$ , pues no se modificó la cantidad de hombres de la lista 2, solo la cantidad final de personas de la lista. Pero  $P(C|B) = \frac{8}{11}$ , pues sí se modificó el número de hombres de la lista 2 y su número final de personas. También  $P(A) = \frac{4}{7}$  y  $P(B) = \frac{3}{7}$ , por lo cual

$$P(A|C) = \frac{\frac{7}{11} \cdot \frac{4}{7}}{\frac{7}{11} \cdot \frac{4}{7} + \frac{8}{11} \cdot \frac{3}{7}} = \frac{28}{28 + 24} = \frac{7}{13}$$

### 3.3.1. Independencia

Intuitivamente diríamos que dos eventos  $A$  y  $B$  son independientes si el conocimiento de que uno ha ocurrido no nos da ninguna información sobre la ocurrencia del otro. En muchos casos esto es más que razonable, por ejemplo, si una enfermera mide la presión arterial usando un aparato digital dos veces consecutivas, los resultados son independientes pues no el resultado obtenido en la primera medición no influye sobre el segundo. También, si alguien tira una moneda balanceada dos veces seguidas, los resultados son independientes, los tiradores no pueden influenciar el resultado. En cambio si se realiza la misma ejercitación dos veces, los resultados no son independientes porque el conocimiento sobre la prueba adquirido la primera vez influye el resultado de la segunda prueba.

Para poder incluir esto en nuestro modelo matemático, vamos a pensar que dos eventos deberían ser independientes si al condicionar uno con el otro no varía su probabilidad esto es,  $P(A/B) = P(A)$  y  $P(B/A) = P(B)$ , Ahora, si

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

resulta

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

Usaremos esta última ecuación como la definición de independencia matemática. Notemos que es simétrica con respecto a  $A$  y  $B$  y que no requiere la existencia de la probabilidad condicional, es decir  $P(B)$  puede ser cero.

**Definición 3.3.2** *Dos eventos son independientes si*

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

*Si no ocurre esto entonces los eventos son dependientes.*

Debemos distinguir en este momento dos situaciones: cuando sabemos que hay independencia (por el diseño del experimento) y cuando no lo sabemos y queremos probarlo. En el primer caso, usamos la fórmula de independencia para calcular la probabilidad de la intersección (porque la sabemos cierta). En el segundo caso, queremos verificar que la ecuación de independencia vale. En los siguientes ejemplos enfatizaremos esta diferencia.

#### Ejemplo 3.3.11

Consideremos los eventos:

A: una familia tiene niños de ambos sexos    B: una familia tiene a lo sumo un varón.

1. Supongamos que la familia tiene tres hijos, probar que  $A$  y  $B$  son independientes.
2. ¿Son dependientes si la familia tiene dos hijos?

#### Resolución

Primero debemos suponer que el sexo de los niños es elegido al azar por la cigüeña. El espacio muestral, si la familia tiene tres hijos, resulta

$$\Omega = \{(X, X, X), X = V, M\} \quad \#\Omega = 8$$

y puede considerarse equiprobable. El evento  $A$  consta de 6 puntos, pues solo los puntos  $(M, M, M)$  y  $(V, V, V)$  no están en  $A$ .  $B$  en cambio tiene 4 puntos,  $(M, M, M), (V, M, M), (M, V, M), (M, M, V)$ . La intersección consta de éstos últimos tres puntos, entonces

$$P(A) = \frac{6}{8} \quad P(B) = \frac{4}{8} \quad P(A \cap B) = \frac{3}{8} = \frac{6}{8} \cdot \frac{4}{8}$$

por lo cual se verifica la independencia. En este caso no se la supone de entrada, se la verifica, no hay ninguna razón para pensar que son independientes estos eventos. Esto es correcto pues si la familia tiene solo dos hijos,

$$P(A) = \frac{2}{4} \quad P(B) = \frac{3}{4} \quad P(A \cap B) = \frac{1}{2} \neq \frac{1}{2} \cdot \frac{3}{4}$$

por lo cual los eventos son dependientes. No es muy intuitivo saber porqué pero es interesante ver la gran diferencia que hay entre los espacios muestrales.

Hay un ejemplo muy importante de independencia por diseño, es el de los experimentos idénticos repetidos en sucesión. Si se conoce la probabilidad de los resultados del experimento simple, se puede encontrar usando la independencia la del experimento compuesto. Como ejemplo estudiaremos primero la tirada de la moneda muchas veces.

**Ejemplo 3.3.12** Consideremos el experimento que consiste en tirar una misma moneda (no necesariamente balanceada) 20 veces y registrando el símbolo que se vé al caer al piso en cada tirada. Si tiramos la moneda una sola vez, la probabilidad de obtener una cara es  $1/3$  y la de obtener número  $(1 - 1/3) = 2/3$ . Intuitivamente, es obvio que el resultado de la  $i$ -ésima tirada no debería influenciar el resultado de las otras tiradas, por lo cual la probabilidad de salir cara en una tirada particular es la misma que en cualquier otra. ¿Cual sería la probabilidad de obtener las primeras 9 caras y las siguientes 11 números?

**Resolución**

Pensemos que cada tirada es independiente de la siguiente, y que la probabilidad de obtener una cara es  $1/3$  y la de un número es  $2/3$ , la probabilidad de observar  $(c, c, c, c, c, c, c, c, c, n, n, n, n, n, n, n, n, n, n, n)$  (9 caras y después 11 números) es la probabilidad de una intersección, que resulta en el producto de las probabilidades de cada uno de los eventos, por lo cual

$$P(c, c, c, c, c, c, c, c, c, n, n, n, n, n, n, n, n, n, n, n) = 1/3 \times \cdots \times 1/3 \times 2/3 \times \cdots \times 2/3 = (1/3)^9 (2/3)^{11}$$

Si tenemos cualquier otro experimento que tenga dos resultados posibles, y lo repetimos un número fijo de veces, el cálculo de las probabilidades de los puntos resultantes se hace de igual forma. Lo describiremos en el siguiente lema:

**Proposición 3.3.1** *Supongamos que repetimos un experimento  $n$  veces en forma independiente. Cada realización del experimento tiene dos posibles resultados, éxito y fracaso, con probabilidades complementarias  $p$  y  $(1 - p)$  respectivamente. Si representamos cada punto muestral como la  $n$ -upla  $(x_1, \dots, x_n)$  donde  $x_i = 1$  o  $0$  si el  $i$ -ésimo experimento resulta en éxito o fracaso, respectivamente, y definimos los eventos  $A_i$  el evento que consta de los puntos donde el  $i$ -ésimo experimento resulta en un éxito, entonces*

$$P((x_1, \dots, x_n)) = p^k (1 - p)^{n-k}$$

es una asignación de probabilidades válida tal que

$$P(A_i) = p \quad P(A_i^c) = 1 - p$$

y los conjuntos  $A_i$  y  $A_i^c$  son todos independientes.

Esta proposición es muy importante, pues valida la intuición de que si los experimentos no están relacionados entre si, las probabilidades de los resultados del experimento compuesto realizado en sucesión se definen como el producto de los experimentos por separado.

**Ejemplo 3.3.13** Una baraja contiene 10 cartas, 6 reyes y cuatro damas y se extraen tres de ellas aleatoriamente con reposición, anotándose únicamente si se observó un rey o una dama en cada extracción.

1. Definir el espacio muestral y la probabilidad asociada.
2. Calcular la probabilidad de no obtener ningún trio de cartas con la misma figura.
3. Calcular la probabilidad de que a lo sumo un aparezca un rey en el trio.

### Resolución

Se extraen tres cartas con reposición, por lo cual el espacio muestral son 3-uplas ordenadas y el trío que muestra tres veces la misma carta está permitido. Ahora como solo me interesa la figura impresa en la carta, denotemos al rey con una r y a la dama con una d. Por lo cual

$$\Omega = \{(r, r, r), (r, r, d), (r, d, r), (d, r, r), (r, d, d), (d, r, d), (d, d, r), (d, d, d)\}$$

Entonces  $\#\Omega = 8$  y el espacio **no** es equiprobable. Calculemos la probabilidad de cada punto. La probabilidad de extraer un rey en una sola tirada es  $p = \frac{6}{10} = \frac{3}{5}$  y la de extraer una dama en una sola tirada es  $(1 - p) = \frac{4}{10} = \frac{2}{5}$ . Como las extracciones son independientes, resulta

$$P(r, r, r) = p^3; \quad P(r, r, d) = P(r, d, r) = P(d, r, r) = p^2(1 - p)$$

$$P(P(r, d, d) = P(d, r, d) = P(d, d, r) = p(1 - p)^2; \quad P(d, d, d) = (1 - p)^3$$

Si  $A$  es el evento “no hay ningún trio de cartas con la misma figura”, entonces

$$P(A) = P(\{(r, r, d), (r, d, r), (d, r, r), (r, d, d), (d, r, d), (d, d, r)\}) = 1 - P(r, r, r) - P(d, d, d) = 1 - p^3 - (1 - p)^3 = 1 - \frac{3^3}{5^3} - \frac{2^3}{5^3}$$

Si  $B$  es el evento “a lo sumo un aparece un rey en el trio”, entonces

$$P(B) = P((r, d, d), (d, r, d), (d, d, r), (d, d, d)) = 3p(1 - p)^2 + (1 - p)^3 = 3\frac{3}{5}\frac{2^2}{5^2} + \frac{2^3}{5^3}$$

**Ejemplo 3.3.14** Un sociólogo realiza el siguiente experimento con adolescentes en edad escolar. Toma un dado balanceado con 4 caras rojas y 2 caras verdes, y anuncia que va a arrojarlo muchas veces y anotar el color que obtuvo. Le pide a los jóvenes que elijan de la siguiente lista el resultado que creen que va a salir, y si aciertan ganan 25 \$.

*RGRRR   RGRRRG   GRRRRR*

El estudio dice que la segunda opción es la más elegida por los jóvenes que nunca estudiaron probabilidad. ¿Es esa la opción más probable?

### Resolución

Yo apostaría por la primera porque es más corta, una tirada menos que adivinar, pero es mejor hacer las cuentas. El lema anterior nos dice como calcular las probabilidades de observar todos estos puntos, solo debemos deducir la probabilidad de observar una cara verde y la probabilidad de observar una cara roja. Como tengo 4 verdes y dos rojas, la probabilidad de observar verde es  $4/6=2/3$  y la de observar rojo es  $2/6=1/3$ . Entonces

$$P(RGRRR) = \frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{2^4}{3^5}$$

$$P(RGRRRG) = \frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} = \frac{2^4}{3^5} \times \frac{1}{3}$$

$$P(GRRRRR) = \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{2^4}{3^5} \times \frac{2}{3}$$

El valor más probable es el primero, no el segundo.



**Ejemplo 3.3.15** La paradoja del caballero de Méré

En el siglo XVII los juegos de azar eran la principal diversión de la alta sociedad francesa. Ellos solían apostar en un juego al suceso  $A = \{ \text{en cuatro tiradas de un dado sale al menos un uno} \}$ . En otro juego apostaban al suceso  $B = \{ \text{en 24 tiradas de un par de dados sale al menos un doble uno} \}$ . Antoine Gombard, Caballero De Meré, un experto jugador, pensaba que los dos sucesos tenían igual probabilidad, pues había realizado los siguientes cálculos.

- Primer juego: Al tirar un dado tengo  $1/6$  de chance de que salga un uno. Luego, en 4 tiradas, tendré  $(1/6) + (1/6) + (1/6) + (1/6) = 2/3$  de chance de que salga un uno.
- Segundo juego: Al tirar dos dados, tengo una chance de  $1/36$  de obtener un doble uno. Luego en 24 tiradas, tendré  $24 \times 1/36 = 2/3$  de chance de que salga al menos un doble uno.

Sin embargo la experiencia demostraba que el primer suceso ocurría más a menudo que el segundo. El caballero de Méré le consulto a Blas Pascal, que resolvió el problema con ayuda de su amigo Pierre Fermat. Lo primero que observaron fue que el razonamiento de De Méré era incorrecto, pues implicaba que en seis tiradas, la probabilidad de sacar al menos un uno es 1. Y una simple experiencia muestra que no es así. El error se basaba en que De Méré estaba **sumando probabilidades de sucesos no disjuntos**, como son “saqué un 1 en la primer tirada” y “saqué un 1 en la segunda tirada”, dado que el  $(1, 1)$  está en ambos conjuntos.

La forma de resolverlo correctamente consistía en en calcular la probabilidad del complemento de  $A$ .

$$P(A) = P(\{ \text{al menos un 1 en 4 tiradas} \}) = 1 - P(\{ \text{ningún 1 en las cuatro tiradas} \}) = 1 - \left(\frac{5}{6}\right)^4 = 0,518$$

pues la probabilidad de no sacar ningún 1 en una tirada es  $5/6$  y la probabilidad de no sacar ningún 1 en cuatro tiradas independientes es  $(5/6)^4$  (en vez de  $4 \times 5/6$ ).

Para el segundo juego,

$$P(B) = P(\{ \text{al menos un doble 1 en 24 tiradas} \}) = 1 - P(\{ \text{ningún doble 1 en las 24 tiradas} \}) = 1 - \left(\frac{35}{36}\right)^{24} = 0,491$$

Los valores no son muy distintos, pero si se juega mucho, como hacia De Mere, la diferencia se llegaba a notar.

## 3.4. Variables aleatorias

Hemos estudiado hasta el momento modelos probabilísticos, espacios muestrales con una asignación de probabilidad definida. Los estudios, experimentos, experiencias, encuestas, relevamientos y otras formas de recaudar datos son modelados con estos espacios, pero a menudo la información es demasiada y se necesita focalizar en un aspecto particular de la experiencia. O simplemente se necesita ordenar la encuesta, o el relevamiento.

**Definición 3.4.1** Una variable aleatoria es el resultado de un fenómeno aleatorio.

Esta definición es la esencia de la noción de variable aleatoria. Podríamos pensar que no se diferencia mucho de espacio muestral, en realidad, es una forma de precisar los resultados que componen el espacio muestral. Por ejemplo, si tiramos un dado seis veces consecutivas, el espacio muestral consta de las sucesiones de seis números posibles de ser observados, y podemos definir muchas variables aleatorias, como “número de seis que se observaron”, o “cantidad de números impares observados”, o la variable binaria que vale 1 si se obtuvo dos o más números mayores que 4, y cero si no, etc. Las variables focalizan un aspecto determinado de la experiencia. Veamos más ejemplos

- Supongamos que una fábrica produce sistemas eléctricos. El departamento de control de calidad de la fábrica considera que un sistema eléctrico puede ser defectuoso o no defectuoso, al estudiar todos los sistemas fabricados en la planta en un mes determinado. El ingeniero tiene interés en el  $n^0$  de defectuosos producidos en esa planta ese mes. Para estimarlo elije al azar una cantidad finita de sistemas y los inspecciona; la variable aleatoria de interés es el número de defectuosos de la muestra.
- Supongamos que el gobierno realiza un reporte sobre la supervivencia de pacientes de cirugía en hospitales provinciales. Es de interés comparar dos hospitales que sirven a la misma comunidad. Las variables de interés, medidas sobre los pacientes que recibieron cirugía son variables categóricas, tipo de hospital (A o B) y si el paciente sobrevivió o no. Otra variable de interés es la condición en que el paciente llegó al hospital, buena o mala, y su relación con la variable supervivencia. Para decidir cuál hospital pierde menos pacientes después de una cirugía, las tres variables deben ser estudiadas al mismo tiempo.
- Una pantalla de computadora tiene una fina red detrás de la superficie visible. Durante el ensamblado de la pantalla, la red se estira y moldea alrededor de un marco de metal. Muy poca tensión en este punto causaría arrugas, mientras que demasiada partiría la red. La tensión se mide con un aparato electrónico que mide en milivoltios, y dos pantallas no presentan la misma tensión aunque hayan sido manufacturadas en la misma línea. Si un ingeniero de planta quiere determinar si la máquina que realiza el proceso está bien calibrada, la tensión medida en una pantalla elegida al azar es la variable aleatoria a estudiar. Es una variable continua.
- Sea  $X$  la variable aleatoria que cuenta el número de caras en cuatro tiradas de una moneda. Los posibles valores de  $X$  son 0,1,2,3,y 4. Tirando la moneda cuatro veces obtendremos uno de esos posibles valores. Tirando la moneda cuatro veces más nos dará otro valor.  $X$  se llama variable porque su valor varía cuando el experimento es repetido. esta variable se llama discreta porque toma un número finito de valores.
- Un consultor de marketing esta interesado en en cuanto gasta cada comprador de un supermercado de tamaño medio de la zona norte de la ciudad, y cuanto gastan los compradores de la sucursal de la misma cadena en zona sur. Ambas son variables aleatorias que toman valores positivos en un rango de la recta real, por lo cual se llaman continuas.

Hemos visto ejemplos de variables continuas, discretas y categóricas, que son una clase de las discretas. Existen también las variables mixtas, que toman valores específicos con probabilidad distinta de cero, y valores reales en un intervalo. Lo más importante de una variable aleatoria es su distribución de probabilidad, es decir, que probabilidad hay de que la variable tome un valor cualesquiera  $x$ , o que el valor que tome esté en cierto intervalo. Por ejemplo, nos interesa saber cual es la probabilidad de que en cuatro tiradas de la moneda salgan cuatro caras. O cual es la probabilidad de que un comprador de zona norte gaste mas de 15 pesos.

Si las variables aleatorias concentran información sobre un aspecto del espacio muestral, su distribución de probabilidad va a coleccionar la información de la probabilidad de los valores que la variable tome.

**Definición 3.4.2** *La distribución de probabilidad de una variable aleatoria está compuesta de las probabilidades de que se obtengan los diferentes valores de la variable. En el caso continuo, como los puntos tienen probabilidad cero, la distribución de probabilidad consta de todos las posibles áreas bajo la curva densidad.*

El evento  $(X \leq x)$  define el conjunto de elementos de  $\Omega$  que van a parar via  $X$  a la semirrecta  $(-\infty, x]$ .

**Proposición 3.4.1** *La función de distribución acumulada de la variable  $X$  (continua o discreta), definida por*

$$F(x) = P(X \leq x)$$

cumple las siguientes condiciones.

1.  $0 \leq F(x) \leq 1$  para todo  $x$ .
2.  $F$  es una función no decreciente en  $x$ .
3.  $\lim_{x \rightarrow \infty} F(x) = 1$  y  $\lim_{x \rightarrow -\infty} F(x) = 0$
4.  $\lim_{x \rightarrow x_0+} F(x) = F(x_0)$  para todo  $x$ .

Es más, toda función  $F$  que cumple estas propiedades es la función de distribución acumulada de alguna variable aleatoria  $X$ .

Esta función describe completamente la distribución de probabilidades de un avariable aleatoria. Con conocer la probabilidad de estos eventos, ya podemos conocer la probabilidad de cualquier otro evento. En el próximo capítulo estudiaremos variables discretas en particular y en el siguiente continuas. También estudiaremos relaciones entre variables aleatorias, como condicionalidad e independencia, que son muy importantes en el caso de las variables categóricas pues indican causalidad.

**Definición 3.4.3** Los eventos  $A$  y  $B$  son independientes si el conocer  $A$  no cambia la probabilidad de que  $B$  ocurra. Por lo cual dos variables aleatorias  $X$  e  $Y$  son independientes si todo evento que concierna solo a  $X$  es independiente de cualquier otro evento que concierna solo a  $Y$ .

Esta es una definición bastante informal, que será precisada en los próximos capítulos. Las variables  $X$  e  $Y$  que miran si salió un seis en la primera tirada y si salió un seis en la segunda tirada, respectivamente, cuando tiro un dado siete veces, son independientes, pues no influencia la primer tirada el resultado de la segunda.

**Ejemplo 3.4.1** El test ELISA se desarrolló en los años ochenta para relevar la presencia de SIDA en la sangre donada a los bancos de sangre. El test detecta los anticuerpos, sustancias que el cuerpo produce cuando el virus está presente. Cuando hay anticuerpos presentes, el ELISA es positivo con una probabilidad de 0.98 y es negativo con una probabilidad de 0.02. Cuando la sangre estudiada no tiene anticuerpos, ELISA da un resultado positivo con probabilidad 0.07 y negativo con probabilidad 0.93. Como el ELISA es un test diseñado para mantener los bancos de sangre libres de contaminación, la probabilidad de falso positivo de 0.07 (muy alta) es tolerada a cambio de la baja probabilidad, 0.02, de fallar en detectar sangre contaminada. Si se sabe que el 1 % de la población en estudio tiene el virus en sangre, y se define a  $X$  como la variable que vale uno si una persona elegida al azar tiene el virus, y cero si no lo tiene, y la variable  $Y$  como uno si el test ELISA realizado en la sangre de esa misma persona da positivo y cero si da negativo.

1. ¿Cuál es la probabilidad de que  $Y = 1$ ? Esto es, cuál es la probabilidad de que el ELISA de positivo?
2. ¿Cuál es la probabilidad de que  $X = 0$  dado que  $Y = 1$ ? Esto es, cuál es la probabilidad de que la persona no tenga el virus, dado que el ELISA fue positivo?

### Resolución

Veamos la primera de las probabilidades

$$P(Y = 1) = P(\text{"ELISA positivo"}) = P(Y = 1|X = 1)P(X = 1) + P(Y = 1|X = 0)P(X = 0) = 0,98 \times 0,01 + 0,07 \times 0,98 = 0,0784.$$

La segunda de las probabilidades pedidas es

$$P(X = 0|Y = 1) = \frac{P(Y = 1|X = 0)P(X = 0)}{P(Y = 1)} = \frac{0,07 \times 0,98}{0,0784} = 0,8750.$$

Este es un ejemplo que ilustra el hecho que cuando se consideran variables que ocurren muy pocas veces, casi siempre los positivos son falsos positivos.

### 3.4.1. Media y Varianza

Una característica importante de las variables aleatorias es que su distribución de probabilidad puede ser caracterizados por parámetros importantes, como centro y dispersión. La distribución de probabilidad de una variable continua se describe por la curva densidad, y el centro de peso de la curva puede hallarse encontrando el punto en el cual la curva estaría balanceada si fuera de un material solido y se apoyara sobre un alfiler. En la siguiente imagen vemos curvas continuas y su centro y dispersión alrededor del centro.

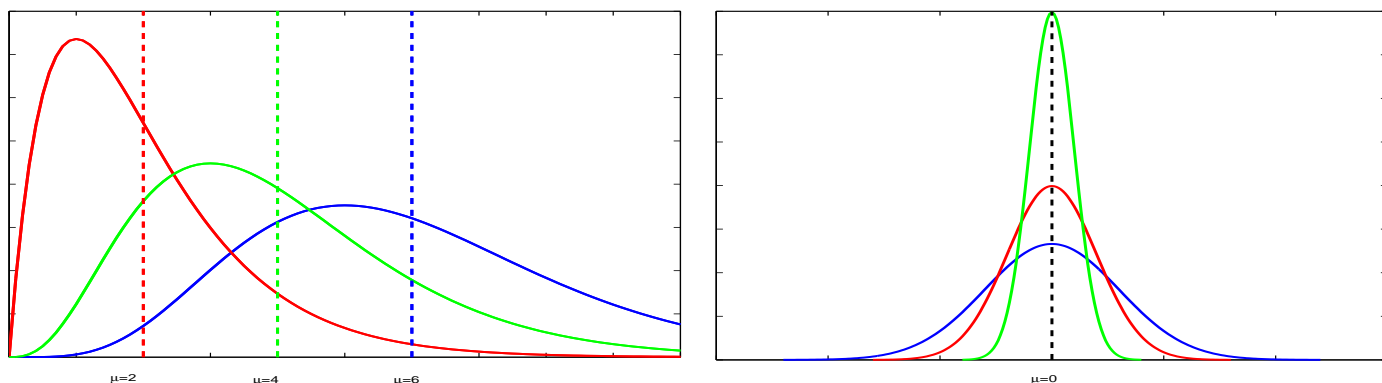


Figura 3.1: Panel izquierdo: Tres densidades con diferente asimetría. Panel derecho: Tres densidades simétricas. Las líneas verticales marcan el centro de la distribución

En cambio, en las variables discretas la esperanza o centro de la distribución es el valor calculado como un promedio pesado de todos los valores que la variable toma. Veamos esto con un ejemplo

**Ejemplo 3.4.2** Supongamos que estamos considerando el jugar un juego de azar que, para ingresar, tiene una apuesta fija de 20 pesos. El juego consiste en girar una enorme rueda, que tiene marcado 40 números en ella, y un puntero que marca un número cuando la rueda se detiene. Como resultado del juego uno recibe  $X$  pesos donde  $X$  es la variable aleatoria que muestra el número que salió elegido en la rueda, que puede tomar los valores  $1, \dots, 40$ . La pregunta es, ¿deberíamos jugar el juego?

A la gente no le gusta jugar un juego en donde solo va perder plata. Si solo quisiéramos jugar una vez, entonces la respuesta sería muy difícil. Por ahí tenemos suerte y ganamos, o quizás no. Sin embargo, si uno considera jugar un número grande de veces, digamos 1200 veces, uno pagaría  $1200 \times 20 = 24000$  pesos y recibiría  $X_1 + \dots + X_{1200}$  pesos, donde  $X_i$  pueden ser consideradas independientes con la misma densidad  $p_X$  (pues son repeticiones independientes de un mismo experimento). Sean  $N_{1200}(x_i)$  la cantidad de juegos en que resultó el valor  $x_i$ , entonces podemos escribir

$$X_1 + \dots + X_{1200} = \sum_{i=1}^{40} x_i N_{1200}(x_i) = 1N_{1200}(1) + 2N_{1200}(2) + \dots + 40N_{1200}(40)$$

El monto promedio recibido es

$$\frac{X_1 + \dots + X_{1200}}{1200} = \frac{1N_{1200}(1) + 2N_{1200}(2) + \dots + 40N_{1200}(40)}{1200} = \sum_{i=1}^{40} x_i \left[ \frac{N_{1200}(x_i)}{1200} \right]$$

Si suponemos que los números salieron la misma cantidad de veces, unas 300 cada una, el monto promedio recibido sería

$$\frac{X_1 + \dots + X_{1200}}{1200} = \frac{1 \times 300 + 2 \times 300 + \dots + 40 \times 300}{1200} = (1 + \dots + 40) \frac{300}{1200} = 820 \frac{1}{40} = 20,5$$

lo cual nos dice que si jugamos esa cantidad de veces (y salen esos números) no perdemos, nos devuelven la plata.

¿Y si las frecuencias relativas no son las mismas, perdemos mucho o ganamos mucho?

De acuerdo a la interpretación frecuentista de las probabilidades, si  $n$  es grande, (en nuestro caso es 1200), los números  $N_n(x_i)/n$  deberían ser aproximadamente iguales a  $P(X = x_i)$ , y entonces la suma de la derecha debería ser aproximadamente igual a

$$\mu = \sum_{i=1}^{40} x_i P(X = x_i).$$

Por lo cual parece razonable anticipar la existencia de ganancias si  $\mu > 20$ , y pérdidas si  $\mu < 20$ , o decir que se terminará “a mano” si  $\mu = 20$ . La cantidad  $\mu$  se llama valor esperado o esperanza de la variable aleatoria  $X$ , y no depende de un caso particular observado.

En este ejemplo es fácil calcular las probabilidades de que salgan los números del 1 al 40, pues son todos igualmente probables, y es  $1/40$ , por lo cual  $\mu = (1 + \dots + 40) \frac{1}{40} = 20,5$  el mismo que obtuvimos con el promedio cuando supusimos las frecuencias relativas balanceadas e iguales. Por lo cual, en este juego, solo del diseño del juego podemos obtener la información de la ganancia promedio  $Y = X - 20$ , que va a ser  $E(Y) = 0,5$  pesos.

Hemos comentado la utilidad de conocer la media de una distribución, el punto de centro, de equilibrio, llamado también la esperanza de la variable aleatoria. El centro de la distribución cobra mucha más importancia en las variables que no cambian muy mucho, que toman valores parecidos, pues la esperanza realmente da un “sumario” de los valores de la distribución. Un ejemplo de esto es la medida de un objeto físico, el diámetro interno de una tuerca, el ancho de un tornillo, o la tensión con que se tensa la malla que forma una pantalla de monitor. La medida debería ser única, pero los aparatos de medida y las líneas de ensamblaje o las máquinas que los tornean no hacen dos objetos iguales exactamente. Esas medidas son variables aleatorias cuyo centro es la medida exacta que “debería” ser. En algunos casos, cuando se quiere saber si la máquina que tornean está bien calibrada, se quiere calcular el centro de la distribución de los tornillos para ver cuán diferente es de lo que debería ser. Un parámetro importante para decidir si los tornillos están siendo bien torneados es el “ancho” de la distribución, la dispersión de los valores medidos alrededor de la media. En la figura 3.2 observamos tres distribuciones con el mismo centro y diferente dispersión. Intuitivamente, la distribución más pegada al centro, con dispersión más chica, va a ser la que produzca valores más cercanos al centro. Es importante en control de calidad calcular la proporción de tornillos de un lote que no alcanza el estándar de calidad, es decir, que su diámetro es más grande que un valor fijo (no entra en la tuerca) o más grande que otro valor (no ajusta en la tuerca).

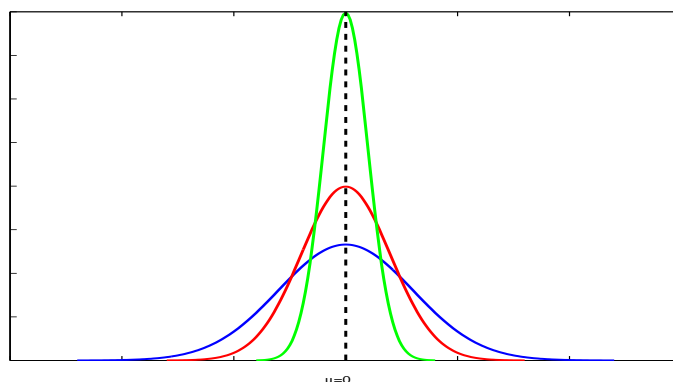


Figura 3.2: Tres densidades con igual centro y diferente dispersión

La dispersión es también un valor inherente a la distribución de probabilidad. La varianza y la desviación estandar son medidas de dispersión que acompañan la elección de la esperanza como medida de centro. La varianza es el cuadrado promediado de los desvíos de los datos de la media.

**Definición 3.4.4** Si  $Y^2$  tiene esperanza finita, entonces la **varianza** de  $Y$ ,  $V(Y)$ , se define como

$$V(Y) = E[(Y - E(Y))^2] = E(Y^2) - E(Y)^2$$

El **desvío standard**  $\sigma$  es la raíz cuadrada positiva de la varianza, y mide la variabilidad de la distribución de la variable alrededor de la media. Cuanto mayor sea el desvío más dispersa será la distribución.

### 3.4.2. Correlación

Ya dijimos que podemos definir más de una variable en un espacio de probabilidad, y que podemos ver sus relaciones, si son independientes o si están **correlacionadas**. Dos variables están correlacionadas cuando la distribución de probabilidad conjunta de ambas variables no es el producto de las densidades de cada variable. La asociación (o falta de asociación) entre variables es uno de los problemas más estudiados de la estadística causal usada en ciencias médicas y sociales.

La varianza de una variable aleatoria es una medida de su variabilidad alrededor de la media, y la covarianza es una medida de su variabilidad conjunta o su grado de asociación, por lo cual juega un papel importante en el cálculo de la varianza de sumas de variables.

**Definición 3.4.5** Sean  $X$  e  $Y$  variables aleatorias con distribución conjunta  $p_{X,Y}$ , y  $E(X^2)$  y  $E(Y^2)$  finita. Entonces

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

En casos más simples, la correlación se presenta en forma de funciones que involucran a ambas variables, la más estudiada de todas las relaciones es la lineal. En la figura 3.3 podemos ver gráficos de variables independientes y correlacionadas.

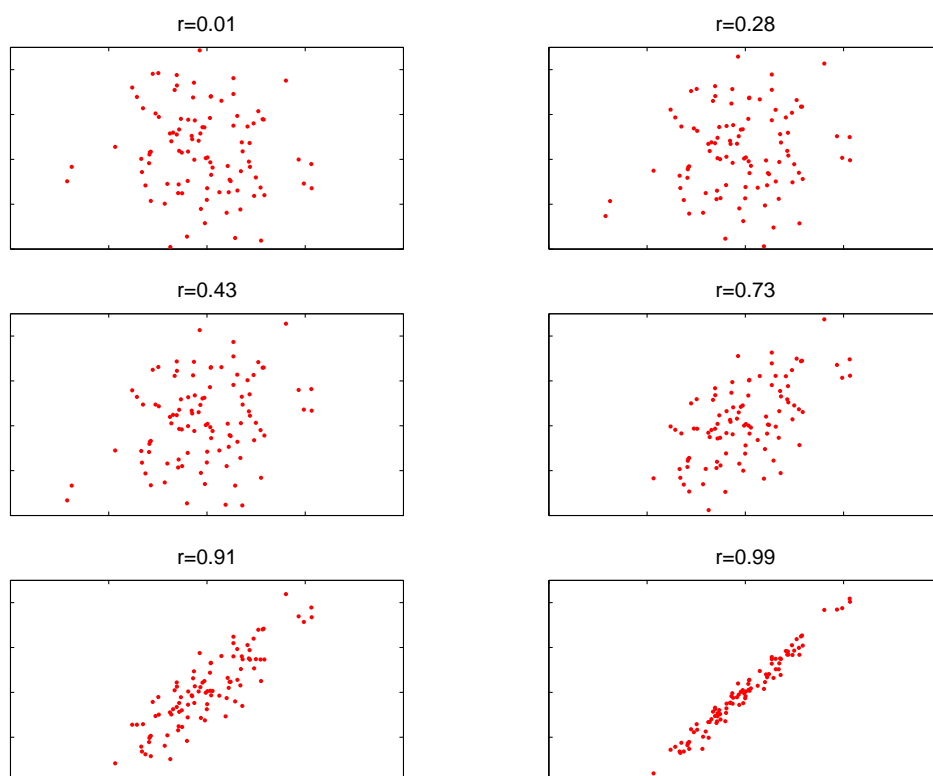


Figura 3.3: Variables normales con media cero y varianza 1, y distintos valores de correlación positiva

En este diagrama una variable ha sido graficada en el eje de las abscisas y otra en el de las ordenadas. En el cuadro donde  $r = 0,99$ , la figura formada por los puntos es rectilínea, sugiriendo que una de las variables es el producto de una constante por la otra variable, mas un valor fijo. Es un tipo de asociación entre variables muy estudiado, pero no el único posible. El cuadro con  $r = 0,01$  muestra a los puntos formando un círculo, la típica figura que anuncia independencia.

La covarianza tiene una definición muy matemática, que es muy útil para realizar cálculos. Ya dijimos, para conocer la varianza de sumas de variables u otras funciones de variables aleatorias, es necesario saber la relación que estas tienen entre si. Pero aparte de la fórmula, que depende de la distribución, está la idea intuitiva de correlación que es muy interesante. Veamos algunos ejemplos de esto.

#### Ejemplo 3.4.3

Un artículo en una revista para señoras reportaba que las mujeres que dan de mamar a sus bebés se sienten mas comunicadas con sus niños que las que eligen dar la mamadera. El autor concluía que el amamantar lleva a una mejor actitud hacia el bebé. ¿Es esto correcto? ¿No puede explicarse la relación observada de otra forma?

Las mujeres que eligen no amamantar a sus bebés pueden reflejar actitudes pre existentes hacia sus niños. Esto es, las madres que ya se sienten mejor predispuestas hacia sus niños eligen amamantar mas a menudo mientras que las que los consideran una carga eligen más seguido dar la mamadera. Los efectos de la variable “tipo de nutrición” en la variable “actitud de la madre”, se confunden con los efectos de la variable “actitud de la madre anterior al nacimiento”. Por lo cual no se puede establecer realmente si hay o no correlación causal entre las variables en estudio.

#### Ejemplo 3.4.4

En muchos estudios sobre niños en edad escolar se ha observado que el número de calzado de los niños esta directamente relacionado con su habilidad para leer. Es más, la correlación entre ambas variables es casi lineal, comprobada por el coeficiente de correlación (covarianza normalizada) cercano a uno. Quiere decir esto que mientras más grandes sean los pies de los niños, mejor leen?. Bueno, la asociación entre las variables es indudable, pero eso no quiere decir que la causa de una lectura adecuada sea el tamaño del pie del alumno, sino la edad que tiene. Los niños en edad escolar que tienen los pies más grandes son más grandes de edad, por lo cual usualmente leen mejor. Asociación y causalidad son cosas muy diferentes, y los coeficientes de correlación solo miden asociación.

La covarianza es una buena medida de asociación que debe ser usada con precaución, en el caso en que se intente observar causalidad. El modelo de regresión o el modelo exponencial, en donde una variable es una función lineal, o el logaritmo de una variable es una función lineal de la otra, son modelos de relación explícita entre dos variables, que permiten predecir una variable conociendo la otra. Si produzco una acción química, obtengo una reacción química proporcional a la acción. La causa de la reacción es la acción. En algunos modelos es claro esto, en otros, sobre todo los que involucran emociones personales, o enfermedades, la causalidad puede ser mucho más difícil de detectar pues existen, demasiado a menudo, factores de confusión. En cambio en las reacciones químicas se puede controlar mejor el experimento para poder evitar el ingreso de factores de confusión en el modelo.

Correlación, causalidad y asociación son conceptos muy complicados, que han sido estudiados en detalle y siguen siendo estudiados principalmente en las escuelas de Medicina y Salud Pública. Establecer la asociación o la no asociación entre variables es un tema muy importante. Por ejemplo, el uso del chupete afecta la lactancia materna? Las empresas de venta de chupetes estan muy interesadas en decir que no tienen asociación alguna. Los niños no dejan de tomar el pecho porque usen chupete, si lo abandonan es por otras causas, principalmente porque la madre no insiste lo suficiente. Otro ejemplo, que es mejor, fumigar

la ciudad en forma preventiva para evitar los mosquitos, (transmisores del dengue u otras enfermedades) o fumigar las zonas donde se detectan casos de la enfermedad?. En algunas zonas de USA, como California, no hay mosquitos, ni cucarachas ni otros insectos, es muy difícil verlos en las zonas urbanas, pero lo que si se nota es la gran cantidad de gente con problemas de piel. Las fumigaciones han terminado con las plagas, pero las dermatitis, son la consecuencia de éstas?

Ahora veamos la relación entre la correlación y la independencia. Es de suponer que dos variables independientes no estan correlacionadas. Por lo cual el indicador de la correlación, la covarianza, debería ser cero. La otra propiedad interesante de las variables no correlacionadas, las independientes, es que la varianza de una suma de variables independientes es la suma de las varianzas. Y las sumas son variables muy importantes en estadística.

### Proposición 3.4.2 Aplicaciones

1.  $Var(X) = Cov(X, X)$
2.  $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
3. *Más generalmente*

$$Var\left(\sum_{i=1}^n b_i X_i\right) = \sum_{i=1}^n b_i^2 Var(X_i) + \sum_{i \neq j} b_i b_j Cov(X_i, X_j)$$

4. Si  $X$  e  $Y$  variables independientes, entonces  $Cov(X, Y) = 0$ .
5. Si  $X_1, \dots, X_n$  son variables aleatorias independientes,

$$Var\left(\sum_{i=1}^n b_i X_i\right) = \sum_{i=1}^n b_i^2 Var(X_i)$$

### 3.4.3. Coeficiente de Correlación lineal

En la figura 3.4 una variable ha sido graficada en el eje de las abscisas y otra en el de las ordenadas. En el cuadro donde  $r = -0,99$ , la figura formada por los puntos es rectilínea, sugiriendo que una de las variables es el producto de una constante (negativa en este caso) por la otra variable, mas un valor fijo. Es un tipo de asociación entre variables muy estudiado, pero no el único posible. El cuadro con  $r = -0,01$  muestra a los puntos formando un círculo, la típica figura que anuncia independencia. Este coeficiente que mide la asociación lineal es el coeficiente de correlación o covarianza normalizada.

**Definición 3.4.6** Sean  $X$  e  $Y$  variables aleatorias con  $0 < Var(X) < \infty$ ,  $0 < Var(Y) < \infty$ . Definimos el coeficiente de correlación lineal

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

El coeficiente de correlación lineal es independiente de posición y escala, pues

$$\rho(X + a, Y + a) = \rho(X, Y) \quad \rho(aX, aY) = \rho(X, Y), a > 0$$

El coeficiente de correlación lineal mide el grado de dependencia lineal de dos variables en el siguiente sentido.



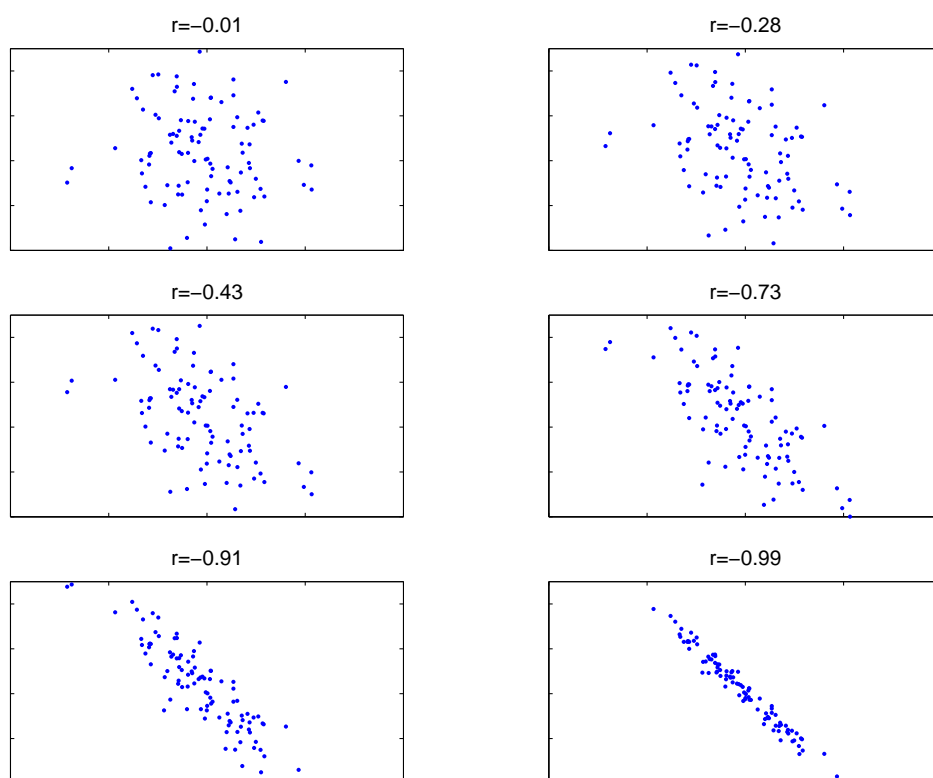


Figura 3.4: Variables normales con media cero y varianza 1, y distintos valores de correlación negativa

**Proposición 3.4.3** *Propiedades.*

1.  $-1 \leq \rho(X, Y) \leq 1$
2.  $\rho(X, Y) = 1$  si y solamente si  $P(Y = aX + b) = 1$ , con  $a > 0$ .  
 $\rho(X, Y) = -1$  si y solamente si  $P(Y = aX + b) = 1$ , con  $a < 0$ .

Los gráficos de las figuras 3.3 y 3.4 ilustran el cambio en la relación lineal de las variables cuando  $\rho$  se acerca a 1 o -1. El signo de  $\rho$  indica solo la dirección de la asociación, por lo cual  $\rho = -0,7$  y  $\rho = 0,7$  indican asociaciones lineales de la misma magnitud pero en direcciones opuestas.

El valor del coeficiente de correlación solo está relacionado con la correlación lineal, puede haber correlación muy fuerte de otra clase y el coeficiente de correlación no mostrará nada. Por ejemplo, si  $(X, Y)$  son las coordenadas de puntos elegidos al azar sobre un círculo, están fuertemente correlacionados pero  $\rho$  será cercano a cero, simplemente pues están muy lejos de una recta.

Esta sección ha intentado dar definiciones intuitivas de los temas que se van a estudiar en detalle en los siguientes capítulos. Es necesario hacer distinciones entre variables aleatorias discretas y continuas; las primeras tienen un tratamiento en el cual las técnicas de conteo que se estudiaron en el capítulo uno van a seguir teniendo un gran protagonismo, mientras que el estudio de las variables continuas necesita de matemática más avanzada. Algunos cálculos se realizarán con la ayuda de las nociones generales de integración, pues es necesario calcular áreas bajo curvas. En otros casos, no es posible integrar, así que debe manejarse o bien una tabla de integrales aproximadas, o algún programa de computación que tenga incorporado esas tablas.



# Capítulo 4

## Variables aleatorias discretas

### 4.1. Introducción

**Definición 4.1.1** Una **variable aleatoria discreta** es una variable aleatoria definida en un espacio muestral discreto. Esta función cumple la condición de que el conjunto  $(X = x) = \{\omega : X(\omega) = x\}$  de los puntos del espacio muestral que van a para a un valor  $x$  es un evento y puede asignarsele probabilidades.

**Ejemplo 4.1.1** Son ejemplos de variables discretas:

- $Y$  número de elementos defectuosos producidos en la fábrica por mes,
- $Z$  el número de incendios declarados por kilómetro cuadrado en un mes determinado.
- $X$  hospital en el que se le realizó la cirugía (A o B),  $Y$ , estado del paciente a los 15 días, (sobrevivio, o no),  $V$ , estado del paciente al ingresar a cirugía, (grave o bueno).
- $X$  es el número de horas que pasan hasta registrar una falla en el motor de un avión.

**Definición 4.1.2** Sea  $Y$  una variable aleatoria discreta. La función  $p_Y$  definida por

$$p_Y(y) = P(Y = y) = P(\{\omega \in \Omega : Y(\omega) = y\})$$

se denomina **densidad discreta** (o **función de frecuencia**) de la variable  $Y$ . Como  $\Omega$  es un espacio finito o numerable,  $P(Y = y)$  es la suma de las probabilidades de todos los puntos muestrales de  $\Omega$  que tienen asignado el valor  $y$ .

$$p_Y(y) = \sum_{\omega \in (Y=y)} P(\{\omega\})$$

Esta función puede ser presentada en forma de tabla, de una gráfica o de una fórmula.

**Ejemplo 4.1.2** Un capataz en una fábrica tiene 3 hombres y 3 mujeres trabajando para él. Desea elegir 2 trabajadores para una labor especial y decide seleccionarlos al azar para no crear problemas entre ellos. Sea  $Y$  el número de mujeres en su elección. Encuentre la de densidad discreta (o función de frecuencia) de  $Y$ .

#### Resolución

El capataz puede escoger 2 trabajadores entre 6 de 15 maneras. Por lo tanto  $\Omega$  tiene 15 puntos muestrales con igual probabilidad, considerando que utilizó una muestra aleatoria. Por lo cual la probabilidad de cada punto es  $\frac{1}{15}$ .

La variable  $Y$  toma los valores 0,1,2, la función densidad de frecuencia en esos valores es la suma de la probabilidad de los puntos que van a para a esos valores,

$$\begin{aligned} p(0) &= P(Y = 0) = \frac{\binom{3}{0}\binom{3}{2}}{15} = 0,2 \\ p(1) &= P(Y = 1) = \frac{\binom{3}{1}\binom{3}{1}}{15} = 0,6 \\ p(2) &= P(Y = 2) = \frac{\binom{3}{2}\binom{3}{1}}{15} = 0,2 \end{aligned}$$

La función de frecuencia puede también presentarse en forma de tabla

$Y$	0	1	2
$p_Y$	0.2	0.6	0.2

Obsérvese que  $Y = 1$  es el resultado más probable, lo cual parece adecuado ya que el número de mujeres es igual al número de hombres en el grupo original.

**Ejemplo 4.1.3** Consideremos el siguiente juego de azar. Un blanco circular de radio 1 tiene marcados 7 anillos concéntricos del mismo ancho. Un dardo se arroja al azar al blanco y si cae en el anillo denotado por los círculos con radio  $k/7$  y  $(k+1)/7$  se ganan  $(7-k)10$  pesos, con  $k = 0, 1, 2, \dots, 6$ . Esto es, si cae en la diana ( $k = 0$ ), gana 70 pesos, si cae en el anillo que le sigue a la diana (o círculo central),  $k = 1$  y gana 60 pesos. Sea  $X$  la variable aleatoria que denota la cantidad de dinero ganado en el juego. ¿Cuál es la densidad discreta de  $X$ ?

### Resolución

El espacio de probabilidad para este experimento es el espacio uniforme del disco con radio 1. Si  $A$  es un subconjunto del disco

$$P(A) = \frac{\text{área}(A)}{\pi}$$

Claramente  $X$  es una variable aleatoria discreta que toma los valores 10, 20, 30, 40, 50, 60 y 70. El evento  $A = (X = (7-k)10)$  ocurre si y solamente si el dardo cae en una región acotada por los círculos de radio  $k/7$  y  $(k+1)/7$ . Entonces

$$P(X = (7-k)10) = P(A) = \frac{\pi[(\frac{k+1}{7})^2 - (\frac{k}{7})^2]}{\pi} = \frac{2k+1}{7^2}$$

Esta es la fórmula general, la tabla de valores es

$X$	70	60	50	40	30	20	10
$p_X$	0.0204	0.0612	0.1020	0.1429	0.1837	0.2245	0.2653

**Ejemplo 4.1.4** Un trompo puede detenerse de cuatro modos A, B, C y D, con la misma probabilidad. Se usa el trompo dos veces, y se anota cada vez su posición;  $Y$  denota el número de posiciones que el trompo no adoptó. Calcule las probabilidades para cada valor de  $Y$ .

### Resolución

El espacio de probabilidad del experimento es el espacio equiprobable  $(\Omega, P)$  con

$$\Omega = \{(x, y), \quad x, y = A, B, C, D\}$$

y  $P(\omega) = \frac{1}{\#\Omega} = \frac{1}{16}$ . La variable aleatoria  $Y$  puede tomar los valores 2 y 3 con probabilidad no nula. La función densidad  $p_Y$  se calcula

$$p_Y(2) = P(Y = 2) = P(\{(A, B), (A, C), (A, D), (B, A), (B, D), (B, C), (C, A), (C, B), (C, D), (D, A), (D, B), (D, C)\}) = \frac{12}{16}$$

$$p_Y(3) = P(Y = 3) = P(\{(A, A), (B, B), (C, C), (D, D)\}) = \frac{4}{16}$$

La tabla de valores resulta

$Y$	2	3
$p_Y$	0.75	0.25

## 4.2. Cómputos con densidades

Hasta ahora solo hemos calculado  $P(X = x)$ , pero a menudo estaremos interesados en calcular la probabilidad de

$$(X \in A) = \{\omega : X(\omega) \in A\}$$

donde  $A$  es más grande que un subconjunto puntual. Si la variable es discreta, entonces

$$P(X \in A) = \sum_{x_i \in A} P(X = x_i) = \sum_{x_i \in A} p_X(x_i)$$

Esta notación abreviada se usa también para las probabilidades condicionales. Si  $A$  y  $B$  son subconjuntos del rango de la variable entonces

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

**Definición 4.2.1** La función  $F_X : \mathbb{R} \rightarrow [0, 1]$  definida por

$$F(t) = P(X \leq t) = \sum_{x \leq t} p_X(x) \quad t \in \mathbb{R}$$

se llama función de distribución acumulada de la variable  $X$  con densidad discreta  $f_X$ . Se sigue que

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

**Ejemplo 4.2.1** En un ítem de una prueba aplicada a niños pequeños, se les pide que hagan corresponder cada uno de los tres dibujos de animales con la palabra que identifica a ese animal. Si un niño asigna aleatoriamente las tres palabras a los tres dibujos, encuentre la función densidad discreta y la función de distribución acumulada de  $Y$ , el número de aciertos. Encuentre  $P(1 < X \leq 6)$ .

### Resolución

Si la posición en el vector  $(x, y, z)$  denota el dibujo y  $d_1, d_2, d_3$  denotan los nombres de los animales, el espacio muestral del experimento es

$$\Omega = \{(d_1, d_2, d_3), (d_2, d_1, d_3), (d_3, d_2, d_1), (d_1, d_3, d_1), (d_3, d_1, d_2), (d_2, d_3, d_1)\}$$

y es equiprobable, con  $P(\{\omega\}) = \frac{1}{6}$ .

La variable  $Y$  toma 3 valores con probabilidad diferente de cero por lo cual  $\mathbf{R}_Y = \{0, 1, 3\}$ . La función  $p_Y$  toma los valores

$$p_Y(0) = P(Y = 0) = P(\{\text{ningún acierto}\}) = P(\{(d_2, d_3, d_1), (d_3, d_1, d_2)\}) = \frac{2}{6}$$

$$p_Y(1) = P(Y = 1) = P(\{\text{exactamente un acierto}\}) = P(\{((d_1, d_3, d_2), (d_3, d_2, d_1), (d_2, d_1, d_3))\}) = \frac{3}{6}$$

$$p_Y(2) = P(Y = 2) = P(\{\text{exactamente dos aciertos}\}) = 0$$

$$p_Y(3) = P(Y = 3) = P(\{\text{exactamente tres aciertos}\}) = P(\{(d_1, d_2, d_3)\}) = \frac{1}{6}$$

y  $F_Y$  resulta

$$F_Y(t) = P(Y \leq t) = 0 \quad t < 0$$

$$F_Y(t) = P(Y \leq t) = p_Y(0) = \frac{2}{6} \quad 0 \leq t < 1$$

$$F_Y(t) = P(Y \leq t) = p_Y(0) + p_Y(1) = \frac{2}{6} + \frac{3}{6} \quad 1 \leq t < 2$$

$$F_Y(t) = P(Y \leq t) = p_Y(0) + p_Y(1) + p_Y(2) = \frac{2}{6} + \frac{3}{6} + 0 \quad 2 \leq t < 3$$

$$F_Y(t) = P(Y \leq t) = p_Y(0) + p_Y(1) + p_Y(2) + p_Y(3) = P(Y = 3) = \frac{2}{6} + \frac{3}{6} + 0 + \frac{1}{6} = 1 \quad 3 \leq t$$

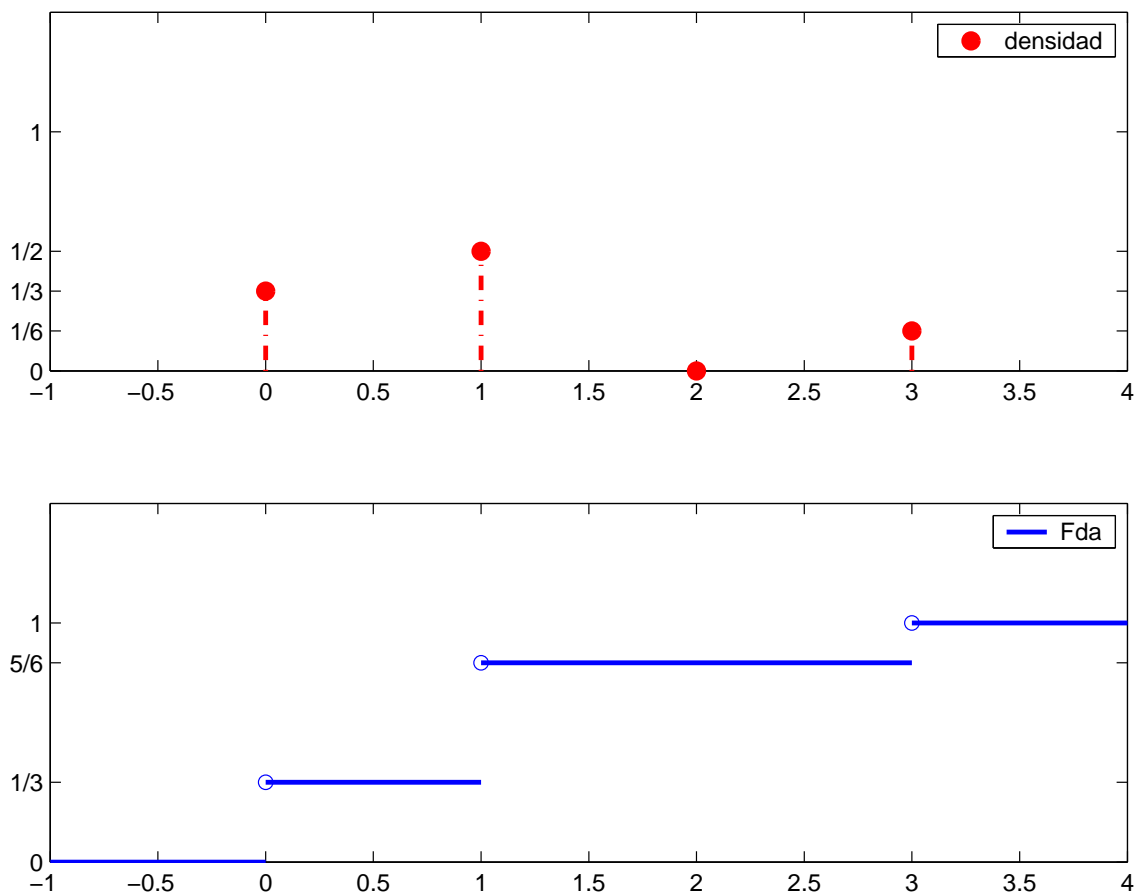


Figura 4.1: Densidad Discreta y Función de Distribución Acumulada de la variable  $Y$  número de aciertos de los niños.

$$P(1 < X \leq 6) = F_X(6) - F_X(1) = 1 - 5/6 = 1/6$$

En la figura 4.1 podemos ver gráficos de la función densidad y la función de distribución acumulada de  $Y$ .

**Ejemplo 4.2.2** Una moneda honesta se tira una vez. Si sale cara se tira una segunda vez y si sale número se tira 2 veces más. Hallar la función densidad discreta de  $X$  el número total de caras obtenidas.

**Resolución**

La variable  $X$  toma los valores 0,1,2, no puede tomar el valor 3 pues nunca se tira tres veces si salió cara la primera vez. Definamos  $Y$  la variable que vale 1 si salió cara la primera vez, y 0 si no. Entonces, si  $Y = 0$  tenemos el experimento condicional con espacio muestral equiprobable

$$\Omega_0 = \{(s, s, s), (s, c, s), (s, s, c), (s, c, c)\}$$

y si  $Y = 1$ , tenemos el experimento condicional con espacio muestral equiprobable

$$\Omega_1 = \{(c, s), (c, c)\}$$

por lo tanto

$$\begin{aligned}
p_X(0) &= P(X=0/Y=0)P(Y=0) + P(X=0/Y=1)P(Y=1) \\
&= P(\{(s,s,s)\})P(\text{seca en la primera}) + P(\emptyset)P(\text{cara en la primera}) \\
&= \frac{1}{4} \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{8} \\
p_X(1) &= P(X=1/Y=0)P(Y=0) + P(X=1/Y=1)P(Y=1) \\
&= P(\{(s,c,s), (s,s,c)\})P(\text{seca en la primera}) + P(\{(c,s)\})P(\text{cara en la primera}) \\
&= \frac{2}{4} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2} \\
p_X(2) &= P(X=2/Y=0)P(Y=0) + P(X=2/Y=1)P(Y=1) \\
&= P(\{(s,c,c)\})P(\text{seca en la primera}) + P(\{(c,c)\})P(\text{cara en la primera}) \\
&= \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{8}
\end{aligned}$$

### 4.3. Vectores aleatorios

En muchos experimentos hay más de una cantidad o aspecto cualitativo de interés, que pueden ser modelados como variables aleatorias sobre el mismo espacio muestral. Cada variable tiene su distribución de probabilidad intrínseca, pero para completar el modelo del experimento es necesario observar sus interacciones, caracterizadas por la estructura de probabilidad conjunta de las variables.

**Definición 4.3.1** Si  $X$  e  $Y$  son variables aleatorias sobre el mismo espacio de probabilidad, se llama *densidad marginal de  $X$*  (respectivamente de  $Y$ ) a la densidad de probabilidad  $p_X$  definida como  $p_X(x) = P(X=x)$ . Se llama *densidad conjunta de  $X$  e  $Y$*  a la función

$$p_{X,Y}(x,y) = P(X=x, Y=y) = P((X=x) \cap (Y=y)) = P(\{\omega \mid X(\omega)=x, Y(\omega)=y\})$$

Observemos que  $p_X(x) = \sum_{y \in \mathbf{R}_Y} p_{X,Y}(x,y)$ , y  $p_Y(y) = \sum_{x \in \mathbf{R}_X} p_{X,Y}(x,y)$ . También ocurre que  $p_{X,Y}$  es no negativa, discreta y la suma sobre todos los valores donde es no nula es 1.

**Ejemplo 4.3.1** Consideremos el experimento que consiste en tirar un dado y observar el número que salió. Sea  $X$  la variable aleatoria que representa dicho número, e  $Y$  la que representa la cualidad, “el número que salió es par”. ¿Cuál sería  $p_X$ ,  $p_Y$ , y  $p_{X,Y}$ ?

#### Resolución

$X$  tiene densidad  $p_X(k) = \frac{1}{6}$  si  $k \in \mathbf{R}_X = \{1, 2, 3, 4, 5, 6\}$ . En cambio  $Y$  es una variable Bernoulli, esto es  $Y=0$  cuando el resultado no es par, e  $Y=1$  cuando lo es, por lo cual  $p_Y(1) = p_Y(0) = \frac{1}{2}$ . Entonces

$p_{X,Y}$	1	2	3	4	5	6
1	0	$\frac{1}{6}$	0	$\frac{1}{6}$	0	$\frac{1}{6}$
0	$\frac{1}{6}$	0	$\frac{1}{6}$	0	$\frac{1}{6}$	0

**Ejemplo 4.3.2** Supongamos tirar una moneda honesta tres veces. Definimos las siguientes variables aleatorias,  $X_1$  representa el número de caras en la primer tirada,  $X_2$  representa el número de caras en las primeras dos tiradas y  $X_3$  es el total de caras en las tres tiradas. Encuentre la densidad discreta conjunta de  $X_1$  y  $X_3$  y las densidades marginales  $p_{X_1}$  y  $p_{X_3}$ . Encuentre la densidad discreta conjunta de las tres variables.

**Resolución**

Si llamamos cara por  $c$  y número por  $s$ , el espacio muestral  $\Omega$  es

$$\Omega = \{(ccc), (ccs), (csc), (scc), (ssc), (scs), (css), (sss)\}$$

La densidad conjunta entre  $X_1$  y  $X_3$  se calcula de la siguiente forma:

$$\begin{aligned} p_{X_1, X_3}(0, 0) &= P(X_1 = 0, X_3 = 0) = P((sss)) = \frac{1}{8} \\ p_{X_1, X_3}(0, 1) &= P(X_1 = 0, X_3 = 1) = P((ssc), (scs)) = \frac{2}{8} \\ p_{X_1, X_3}(0, 2) &= P(X_1 = 0, X_3 = 2) = P((scc)) = \frac{1}{8} \end{aligned}$$

los demás valores se calculan de forma similar. La siguiente tabla muestra la densidad conjunta completa, y las marginales fueron añadidas en el borde derecho y en la base de la tabla. La marginal  $p_{X_1}$  resulta de sumar las filas de la tabla y la marginal  $p_{X_3}$  resulta de sumar las columnas de la tabla.

$p_{X_1, X_3}$	0	1	2	3	$p_{X_1}$
0	1/8	2/8	1/8	0	1/2
1	0	1/8	2/8	1/8	1/2
$p_{X_3}$	1/8	3/8	3/8	1/8	1

Para calcular la densidad conjunta de las tres variables operamos de la misma forma, solo que tenemos que armar una tabla más complicada

$$\begin{aligned} p_{X_1, X_2, X_3}(0, 0, 0) &= P(X_1 = 0, X_2 = 0, X_3 = 0) = P((sss)) = \frac{1}{8} \\ p_{X_1, X_2, X_3}(1, 1, 1) &= P(X_1 = 1, X_2 = 1, X_3 = 1) = P((css)) = \frac{1}{8} \\ p_{X_1, X_2, X_3}(1, 2, 2) &= P(X_1 = 1, X_2 = 2, X_3 = 2) = P((ccs)) = \frac{1}{8} \end{aligned}$$

	$X_3 = 0$		$X_3 = 1$		$X_3 = 2$		$X_3 = 3$	
$X_2 \backslash X_1$	0	1	0	1	0	1	0	1
0	1/8	0	1/8	0	0	0	0	0
1	0	0	1/8	1/8	1/8	1/8	0	0
2	0	0	0	0	0	1/8	0	1/8

Si quisiéramos calcular  $P(X_1 + X_3 = 2)$  o  $P(X_1 + 1 \leq X_3)$ , basta con ver cuantos duplas de valores satisfacen estos eventos y buscar su probabilidad en la tabla de  $p_{X_1, X_3}$ .

$$P(X_1 + X_3 = 2) = P(X_1 = 0, X_3 = 2) = P(X_1 = 1, X_3 = 1) = p_{X_1, X_3}(0, 2) + p_{X_1, X_3}(1, 1) = 1/8 + 1/8 = 1/4$$

$$\begin{aligned} P(X_1 + 1 \leq X_3) &= P(X_1 = 0, X_3 \geq 1) + P(X_1 = 1, X_3 \geq 2) \\ &= p_{X_1, X_3}(0, 1) + p_{X_1, X_3}(0, 2) + p_{X_1, X_3}(0, 3) + p_{X_1, X_3}(1, 2) + p_{X_1, X_3}(1, 3) \\ &= 2/8 + 1/8 + 0 + 2/8 + 1/8 = 1/4 \end{aligned}$$



**Ejemplo 4.3.3** Supongamos que tiramos un dado 10 veces y registramos los valores obtenidos. Sea  $X_i$  el número observado en la  $i$  esima tirada. Entonces  $\mathbf{X} = (X_1, \dots, X_n)$  es un vector aleatorio que representa los resultados del experimento. ¿Cuál es su densidad conjunta?

**Resolución**

Sabemos como deberían ser sus marginales, pues conocemos bien el experimento simple "tirar un dado y registrar que número salió". Ahora, el conocimiento de las densidades marginales no es suficiente para caracterizar la densidad conjunta, se necesita más información, por ejemplo, alguna información condicional. Observemos que

$$P(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1, \dots, X_{10} = x_{10}) = P(X_1 = x_1, \dots, X_9 = x_9 | X_{10} = x_{10})P(X_{10} = x_{10})$$

aplicando la definición de probabilidad condicional. Ahora, en el ejemplo que estamos considerando, conocer el resultado de la última tirada no arroja ninguna luz sobre las ocurrencias en las primeras nueve tiradas. Esto es

$$P(X_1 = x_1, \dots, X_9 = x_9 | X_{10} = x_{10}) = P(X_1 = x_1, \dots, X_9 = x_9)$$

simplemente porque construimos el espacio de probabilidad para que esto sea cierto, los eventos  $(X_i = x_i)$  son todos mutuamente independientes. Entonces

$$P(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1, \dots, X_9 = x_9)P(X_{10} = x_{10})$$

Repitiendo el razonamiento con todas las variables resulta

$$P(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1) \dots P(X_{10} = x_{10})$$

por lo cual

$$p_{\mathbf{X}}(\mathbf{x}) = p_{X_1}(x_1) \dots p_{X_{10}}(x_{10})$$

y en este ejemplo la densidad conjunta es el producto de las marginales.

## 4.4. Variables aleatorias independientes

Consideremos el experimento de tirar un dado y una moneda honesta. Intuitivamente, sabemos que cualquiera sea el resultado de tirar la moneda no puede influir en el resultado del dado, y viceversa. Por lo cual el espacio de probabilidad conjunto que queremos construir debería reflejar esto. La variable aleatoria que representa la tirada de la moneda es una variable Bernoulli que vale 1 cuando sale cara y 0 cuando sale número, por lo cual  $P(Y = 1) = P(Y = 0) = \frac{1}{2}$ . La variable  $X$  que gobierna la tirada del dado toma los valores 1, 2, 3, 4, 5, 6 con probabilidad  $P(X = k) = \frac{1}{6}$ ,  $k = 1, \dots, 6$ . Ahora el resultado del experimento combinado es el vector aleatorio  $(X, Y)$  cuyas marginales acabamos de describir. ¿Cuál debería ser su estructura de probabilidad conjunta? Nuestra noción intuitiva de que el resultado del dado no influye el de la moneda, se traduce en el hecho que los eventos  $(X = k)$  y  $(Y = j)$  deberían ser independientes. Por lo cual el vector aleatorio  $(X, Y)$  debería tener función densidad conjunta  $p_{X,Y}$  dada por

$$p_{X,Y}(x, y) = \begin{cases} P(X = x)P(Y = y) & x = 1, 2, 3, 4, 5, 6, \quad y = 0, 1 \\ 0 & \text{en otro caso} \end{cases}$$

En otras palabras, la densidad conjunta de  $(X, Y)$  debería ser

$$p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

**Definición 4.4.1** Sea  $\mathbf{X} = (X_1, \dots, X_n)$  un vector aleatorio discreto con densidades marginales  $p_{X_1}, \dots, p_{X_n}$  y densidad conjunta  $p_{\mathbf{X}}$ . Entonces las variables aleatorias  $X_1, \dots, X_n$  se dicen mutuamente independientes si su función densidad conjunta cumple

$$p_{\mathbf{X}}(x_1, \dots, x_n) = p_{X_1}(x_1) \dots p_{X_n}(x_n)$$

**Lema 4.4.1** Si  $(X, Y)$  es un vector aleatorio, tal que  $X$  e  $Y$  son mutuamente independientes, entonces para cada par de eventos  $A$  y  $B$  de  $\Omega$ , se cumple

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

**Ejemplo 4.4.1** Consideremos un equipo formado por 3 personas. Cada una de ellas tira un dado en forma independiente. El puntaje del equipo, al cual llamaremos  $W$ , es el máximo de los tres resultados. Hallar  $P(W \leq 4)$  y  $P(W = 4)$ .

**Resolución**

Sean  $X_1, X_2, X_3$  los puntajes de los tres integrantes del equipo, entonces  $W = \max(X_1, X_2, X_3)$  y podemos escribir

$$\begin{aligned} P(W \leq 4) &= P(\max(X_1, X_2, X_3) \leq 4) \\ &= P(X_1 \leq 4, X_2 \leq 4, X_3 \leq 4) \\ &= P(X_1 \leq 4)P(X_2 \leq 4)P(X_3 \leq 4) \\ &= \left(\frac{2}{3}\right)^3 = \frac{8}{27} \sim 0,296 \end{aligned}$$

También

$$P(W = 4) = P(W \leq 4) - P(W \leq 3) = \left(\frac{2}{3}\right)^3 - \left(\frac{1}{2}\right)^3 \sim 0,17$$

#### 4.4.1. Esperanza de una variable aleatoria discreta

Supongamos que estamos considerando el jugar un juego de azar que, para ingresar, tiene una apuesta fija de  $a$  pesos. Como resultado del juego uno recibe  $X$  pesos donde  $X$  es una variable aleatoria que puede tomar  $x_1, \dots, x_k$  valores. La pregunta es, deberíamos jugar el juego? Si solo quisiéramos jugar una vez, entonces la respuesta sería muy difícil. Sin embargo, si uno considera jugar un número  $n$  de veces, uno pagaría  $na$  pesos y recibiría  $X_1 + \dots + X_n$  pesos, donde  $X_i$  pueden ser consideradas independientes con la misma densidad  $p_X$  (pues son repeticiones independientes de un mismo experimento). Sean  $N_n(x_i)$  la cantidad de juegos en que resultó el valor  $x_i$ , entonces podemos escribir

$$X_1 + \dots + X_n = \sum_{i=1}^k x_i N_n(x_i)$$

El monto promedio recibido es

$$\frac{X_1 + \dots + X_n}{n} = \sum_{i=1}^k x_i \left[ \frac{N_n(x_i)}{n} \right]$$

De acuerdo a la interpretación frecuentista de las probabilidades, si  $n$  es grande, los números  $N_n(x_i)/n$  deberían ser aproximadamente iguales a  $p_X(x_i)$ , y entonces la suma de la derecha debería ser aproximadamente igual a  $\mu = \sum_{i=1}^k x_i p_X(x_i)$ . Por lo cual parece razonable anticipar la existencia de ganancias si  $\mu > a$ , y pérdidas si  $\mu < a$ , o decir que se terminará “a mano” si  $\mu = a$ . La cantidad  $\mu$  se llama valor esperado o esperanza de la variable aleatoria  $X$ .

**Definición 4.4.2** Sea  $Y$  una variable aleatoria discreta con función distribución de probabilidad  $p_Y$ . Si  $\sum_y |y| p_Y(y) < \infty$ , entonces la **esperanza** de  $Y$ ,  $E(Y)$ , está definida por:

$$E(Y) = \sum_y y p_Y(y)$$

donde la suma se realiza sobre todos los valores del rango de la variable  $Y$ .

**Observación 4.4.1** La esperanza es una medida de tendencia central. Cada valor que toma la variable aleatoria contribuye al valor de la esperanza en función de su peso probabilístico. Por ello es que muchos valores con poco peso pueden equipararse a un solo valor con mucho peso. También, un valor muy grande con poco peso puede contribuir mucho a la esperanza, a pesar de ser poco representativo. Por ello, debemos tener cuidado con la noción de “centro”. El valor de la esperanza es el **centro de equilibrio** de la distribución, que **no necesariamente se encuentra en el punto medio de los valores de  $X$** . Si la distribución es simétrica, entonces el punto de equilibrio será el centro de la distribución. En la figura 3.1 hemos graficado distribuciones asimétricas y sus esperanzas y distribuciones simétricas con la misma esperanza.

**Ejemplo 4.4.2** Consideremos los siguientes tres juegos:

1. Tiro una moneda. Gano 1\$ si sale cara y pierdo 1\$ si sale número.
2. Apuesto al rojo en la ruleta. Gano 1\$ si sale rojo y pierdo 1\$ si no sale rojo.
3. Gano 1\$ si saco una bolilla roja de una caja con 5 bolas, solo una roja, las demás negras. Pierdo 0.1\$ si no saco la bola roja.

¿A cuál juego conviene jugar?

#### Resolución

Sea  $X$  la ganancia neta (toma en cuenta lo que me paga la banca, lo que cuesta jugar y lo que apuesto). En el primer juego, la esperanza de  $X$  es

$$E(X) = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$$

En el segundo juego,

$$E(X) = 1 \cdot \frac{18}{37} + (-1) \cdot \frac{19}{37} = -\frac{1}{37}$$

En el tercer juego,

$$E(X) = 1,0,2 + (-0,1),0,8 = 0,2 - 0,008 = 0,12$$

Nos conviene jugar al tercer juego en el cual la ganancia neta tiene esperanza positiva.

**Ejemplo 4.4.3** Keno es un juego del tipo de lotería muy popular en los casinos de Nevada y Atlantic City. Ochenta bolas numeradas se mezclan en un tambor mientras las apuestas se realizan, y luego se retiran 20 bolas al azar. Los jugadores seleccionan los números marcando tarjetas. La mas simple de las apuestas realizables es "Marque un número". La ganancia de una apuesta de 1 U\$D es 3 U\$D si el número elegido es seleccionado. Como se seleccionan 20 de 80 números la probabilidad de ganar es  $20/80$  o  $0.25$ . La densidad discreta de la ganancia (bruta)  $X$  de un único juego es

Ganancia	3 U\$D	0 U\$D
Probabilidad	0.25	0.75

Esto es, la ganancia bruta promedio de una apuesta de un dólar es 75 centavos. La banca se queda con los otros 25 centavos. Como cientos de jugadores están apostando, el casino no arriesga nada, a la larga, se queda con el 25 % de cada dólar apostado.

**Ejemplo 4.4.4** ¿Cuál es el tamaño promedio de una familia argentina? Aquí presentamos una tabla de la distribución de las familias argentinas de acuerdo a uno de los institutos de estadística provinciales

Personas en la familia	2	3	4	5	6	7	8
Proporción de familias	0.231	0.397	0.212	0.097	0.038	0.017	0.008

Si imaginamos seleccionar una sola familia al azar el tamaño de la familia seleccionada es la variable aleatoria  $X$  con distribución de probabilidad dada por la tabla. La esperanza  $E(X)$  es el tamaño medio de las familias de la población. Esta media es

$$E(X) = 2 \times 0,231 + 3 \times 0,397 + 4 \times 0,212 + 5 \times 0,097 + 6 \times 0,038 + 7 \times 0,017 + 8 \times 0,008 = 3,396$$

En este ejemplo hemos ignorado las familias con 9 miembros o más, en realidad el tamaño real de las familias argentinas es más cercano a cuatro que lo que muestra este ejemplo.

#### 4.4.2. Propiedades de la esperanza

A menudo conocemos una variable  $Y$  pero nos interesa obtener información sobre una función de dicha variable, como  $Y^2$  o  $1/Y$ . Más aun, si tenemos un vector  $\mathbf{Y}$ , y una función  $g$ , nos gustaría poder expresar la esperanza de  $g(\mathbf{Y})$  en términos de la función  $g$  y la densidad del vector  $p_{\mathbf{Y}}$ , sin necesidad de calcular la densidad de  $g(\mathbf{Y})$ , que puede ser muy difícil de calcular o sin fórmula cerrada. Por ejemplo, si  $Y_1$  representa el número de veces que salió cara en 5 tiradas de una moneda, e  $Y_2$  el cuadrado del número de secas en las tres primeras tiradas, ¿cuál es la esperanza de  $Y_1 + Y_2$ ?. Para aplicar la definición debo calcular la densidad discreta de la variable  $Y_1 + Y_2$  y su rango, y si bien no es difícil, puede ser engorroso. Pero lo que es fácil de calcular es la densidad de  $Y_1$  y la densidad de  $X$ , el número de veces que sale número en las primeras tres tiradas. Veremos que es lo único que necesito para calcular la esperanza de  $Y_1 + X^2$ .

**Teorema 4.4.1** Sea  $\mathbf{Y}$  un vector aleatorio con densidad discreta  $p_{\mathbf{Y}}$  y sea  $g(\mathbf{y})$  una función a valores reales sobre  $\mathbf{R}_{\mathbf{Y}}$ . Entonces:

$$E(g(\mathbf{Y})) = \sum_{\mathbf{y}} g(\mathbf{y}) p_{\mathbf{Y}}(\mathbf{y})$$

Este teorema es muy importante pues nos permite deducir fórmulas para calcular más fácilmente otras cantidades que son necesarias para describir las distribuciones. La siguiente proposición enumera algunas de las fórmulas de uso más común.

**Corolario 4.4.1** Sean  $X$  e  $Y$  variables aleatorias con densidad discreta  $p_X$  y  $p_Y$  respectivamente y  $c$  una constante. Entonces:

1. Si  $P(Y = c) = 1$ , entonces  $E(Y) = c$ .
2.  $E(cY) = cE(Y)$ .
3.  $E(X + Y) = E(X) + E(Y)$ .
4. Si  $X$  e  $Y$  son independientes, entonces

$$E(XY) = E(X)E(Y)$$

**Ejemplo 4.4.5** Sea  $Y$  el número de veces que salió cara en 5 tiradas de una moneda, e  $Y_1$  el cuadrado del número de números en las tres primeras tiradas, ¿cuál es la esperanza de  $Y + Y_1$ ?.  
**Resolución**

Calculemos la densidad de  $Y$  y la densidad de  $X$ , el número de veces que sale número en las primeras tres tiradas. Esto es lo único que necesito para calcular la esperanza de  $Y + Y_1 = Y + X^2$ .

Las densidades de  $Y$  y  $X$  son

$Y$	0	1	2	3	4	5
$p_Y$	$(1/2)^5$	$5(1/2)^5$	$10(1/2)^5$	$10(1/2)^5$	$5(1/2)^5$	$(1/2)^5$

$X$	0	1	2	3
$p_X$	$(1/2)^3$	$3(1/2)^3$	$3(1/2)^3$	$(1/2)^3$

Entonces

$$\begin{aligned}
 E(Y + X^2) &= E(Y) + E(X^2) = \sum_y y p_Y(y) + \sum_x x^2 p_X(x) \\
 &= 0(1/2)^5 + 1,5 \cdot (1/2)^5 + 2,10 \cdot (1/2)^5 + 3,10 \cdot (1/2)^5 + 4,5 \cdot (1/2)^5 + 5 \cdot (1/2)^5 \\
 &\quad + 0^2 \cdot (1/2)^3 + 1^2 \cdot 3 \cdot (1/2)^3 + 2^2 \cdot 3 \cdot (1/2)^3 + 3^2 \cdot (1/2)^3 \\
 &= 2,5 + 3 \\
 &= 5,5
 \end{aligned}$$

Este ejemplo puede parecer muy matemático, pero usualmente la esperanza del cuadrado de la variable en estudio es importante, así como otras potencias de una variable, llamadas momentos de la distribución. Es porque toman parte en las fórmulas que nos dicen como es la forma de la distribución, dispersa alrededor de la media o concentrada, simétrica o asimétrica.

### 4.4.3. Varianza de una variable discreta

Además de las medidas de centro de una distribución, vamos a estudiar medidas de dispersión que describen la variabilidad de una variable aleatoria. La varianza y la desviación estándar son medidas de dispersión que acompañan la elección de la esperanza como medida de centro. La varianza es el cuadrado promediado de los desvíos de los datos de la media.

**Definición 4.4.3** Si  $Y^2$  tiene esperanza finita, entonces la **varianza** de  $Y$ ,  $V(Y)$ , se define como

$$V(Y) = E[(Y - E(Y))^2]$$

El desvío standard  $\sigma$  es la raíz cuadrada positiva de la varianza, y mide la variabilidad de la distribución de la variable alrededor de la media. Cuanto mayor sea el desvío más dispersa será la distribución.

**Proposición 4.4.1** Sea  $Y$  una variable aleatoria con varianza finita  $\sigma^2$ , entonces

1.  $Var(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2$
2.  $Var(aX) = a^2 Var(X)$
3.  $Var(X) = 0$  implica  $P(X = c) = 1$  para algún  $c$ .
4.  $Var(X + Y) = Var(X) + Var(Y) + 2[E(XY) - E(X)E(Y)]$
5. Si  $X$  e  $Y$  son independientes,  $Var(X + Y) = Var(X) + Var(Y)$

**Ejemplo 4.4.6** Jorge y Mario juegan al golf en el mismo club. El puntaje de Mario, al cual llamaremos  $X$ , varía de ronda en ronda pero tiene

$$E(X) = 110 \quad \sigma_X = 10$$

El puntaje de Jorge, al cual llamaremos  $Y$  tiene

$$E(X) = 100 \quad \sigma_X = 8$$

Jorge y Mario están jugando la primera ronda del torneo de su club. Como no juegan juntos, podemos asumir que sus puntajes varían independientemente uno del otro. La diferencia entre sus puntajes en la primera ronda tiene media

$$E(X - Y) = 110 - 100 = 10$$

La varianza de la diferencia de los puntajes es

$$\sigma_{X-Y}^2 = 10^2 + 8^2 = 164$$

La desviación standard de la varianza es

$$\sigma_{X-Y} = \sqrt{10^2 + 8^2} = \sqrt{164} = 12,8$$

La variación de la diferencia entre los puntajes es mayor que la variación de cada puntaje individual. Esto podría ser explicado pensando que hay dos fuentes de variación en este caso en vez de una. También hace el torneo interesante pues Jorge no terminará siempre arriba de Mario a pesar de tener el menor puntaje medio. No olvidemos que en golf, el que usa menos golpes gana.

La siguiente proposición es muy importante para los cálculos, pero es necesario recordar que la media de una suma es siempre la suma de las medias, mientras que en las varianzas la independencia es necesaria para que esta regla sea válida.

**Proposición 4.4.2** Si  $X_1, \dots, X_n$  son variables aleatorias independientes, entonces

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$$

**Ejemplo 4.4.7** Si  $X_1, \dots, X_n$  son variables aleatorias independientes con media  $\mu$  y varianza  $\sigma^2$ , encuentre la media y la varianza de  $X = \frac{1}{n} \sum_{i=1}^n X_i$ .

**Resolución**

Aplicando la linealidad de esperanza y varianza (al ser las variables independientes)

$$E(X) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu$$

$$\text{Var}(X) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Este es un ejemplo teórico pero de mucha aplicación, no importa como sea la distribución de un grupo de variables discretas, siempre que sean independientes e idénticamente distribuidas, la variable promedio va a tener la misma esperanza y su varianza va a disminuir, va a ser la fracción  $1/n$  de la varianza original. Esto nos dice que repetir muchas veces el mismo experimento y promediar los resultados es bueno, la variable resultante tiene una distribución más concentrada alrededor de la media.

## 4.5. Distribuciones discretas clásicas

Es importante saber que toda función  $p$  real que cumple las propiedades

1. el conjunto donde toma valores distintos de cero,  $\mathbf{R}_p$ , es un conjunto discreto de la recta real,
2.  $0 \leq p(x) \leq 1 \quad \forall x \in \mathbb{R}$
3.  $\sum_{x \in \mathbf{R}_p} p(x) = 1$

es la función densidad discreta de alguna variable  $X$ , por lo cual, podemos definir densidades discretas sin especificar el espacio muestral o el experimento del cual provienen. Usualmente, las densidades llamadas clásicas, Poisson, Binomial, Hipergeométrica, Geométrica y Binomial Negativa, entre otras, surgen de experimentos específicos, por lo cual puede reconocerse el tipo de experimento y deducir que distribución tiene, o puede nombrarse la distribución directamente sin necesidad de reconocerla.

### 4.5.1. Variable aleatoria Bernoulli

Una variable aleatoria Bernoulli toma solo 2 valores, 0 y 1, con probabilidades  $(1 - p)$  y  $p$  respectivamente. Su densidad discreta es

$$p(1) = p \quad p(0) = 1 - p \quad p(x) = 0, \quad \text{si } x \neq 0 \text{ y } x \neq 1$$

Una representación alternativa de esta densidad es

$$p(x) = \begin{cases} p^x(1-p)^{1-x} & \text{si } x = 0, 1 \\ 0 & \text{en otro caso} \end{cases}$$

**Ejemplo 4.5.1** Sea  $A$  un evento, entonces la variable aleatoria indicadora del evento  $I_A$ , vale 1 si  $A$  ocurre y 0 si no.

$$I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{en otro caso} \end{cases}$$

Entonces  $I_A$  es una variable Bernoulli de parámetro  $p = P(A)$ .

**Teorema 4.5.1** La esperanza de una variable Bernoulli es  $p$  y su varianza es  $p(1 - p)$ .

**Demostración:**

Por definición de esperanza y varianza, y por el teorema 4.1,

$$E(X) = 1P(X = 1) + 0P(X = 0) = p \quad E(X^2) = 1^2P(X = 1) + 0^2P(X = 0) = p$$

$$\text{Var}(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p)$$

### 4.5.2. Distribución Binomial

Algunos experimentos consisten en la observación de una serie de pruebas idénticas e independientes, las cuales pueden generar uno o dos resultados. Cada artículo que sale de una línea de producción puede ser defectuoso o no defectuoso. Cada tiro de una serie de tiros al blanco puede ser éxito o fracaso y cada una de las personas interrogadas ante una elección local puede estar a favor o en contra del candidato oficialista.

**Definición 4.5.1** Un experimento binomial es aquel que tiene las siguientes características:

1. Se consideran una cantidad finita  $n$  de pruebas idénticas e independientes.
2. Cada prueba tiene 2 resultados posibles llamados éxito o fracaso.
3. La probabilidad de tener éxito en una prueba es  $p$  y permanece constante de prueba en prueba. La probabilidad de obtener fracaso en una prueba es  $q = (1 - p)$ .
4. La variable bajo estudio es  $Y$ , el número de éxitos observados en las  $n$  pruebas. Por lo cual  $\mathbf{R}_Y = \{1, \dots, n\}$ .

Se puede obtener la distribución de  $Y$ ,  $p_Y(y)$ , aplicando las técnicas de enumeración de puntos muestrales. Cada punto muestral se puede denotar mediante una upla ordenada de 1 correspondientes a cada éxito y 0 correspondiente a los fracasos. Si consideramos el evento  $Y = k$ , un punto posible es  $(1, 1, 1, \dots, 0, 0, 0)$  donde hay  $k$  éxitos y  $(n-k)$  fracasos. Como las  $n$  pruebas son independientes con probabilidad de éxito  $p$  y probabilidad de fracaso  $(1 - p)$ , la probabilidad del punto listado anteriormente es  $p^k(1 - p)^{n-k}$  por la independencia de las pruebas. El evento  $Y = k$  consta de tantos puntos como formas hay de elegir subconjuntos de  $k$  elementos entre  $n$ , (los lugares donde asignar los 1). También, todos los puntos de  $(Y = k)$  son equiprobables, por lo tanto

$$p_Y(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 1, \dots, n$$

**Definición 4.5.2** Una variable aleatoria  $Y$  tiene una **distribución binomial** basada en  $n$  pruebas, con probabilidad de éxito  $p$  sii

$$p_Y(y) = \binom{n}{y} p^y (1 - p)^{n-y} \quad y : 1 \dots n, \quad 0 \leq p \leq 1$$

Observemos que los puntos muestrales de un experimento binomial no son equiprobables salvo que  $p = 0,5$ .

**Ejemplo 4.5.2** Supongamos que un lote de 300 fusibles eléctricos contiene un 5% de defectuosos. Determine la probabilidad de que se puede encontrar el menos un fusible defectuoso en una muestra de cinco fusibles.

**Resolución**

Si seleccionamos 5 fusibles de una manera aleatoria de la población, tenemos cinco pruebas casi idénticas donde éxito ocurre si encuentro un fusible defectuoso, fracaso si no. La variable de interés es  $Y$ =número de éxitos=número de defectuosos. La probabilidad de encontrar un defectuoso en la primera prueba es 0.05, como la población de fusibles es grande y la muestra chica, la probabilidad de éxito en la segunda extracción y las siguientes puede seguir considerándose 0.05 y las pruebas son independientes. Por lo cual el experimento es binomial y

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - p(0) = 1 - \binom{5}{0} 0,05^0 (1 - 0,05)^5 = 1 - 0,95^5 = 0,226.$$

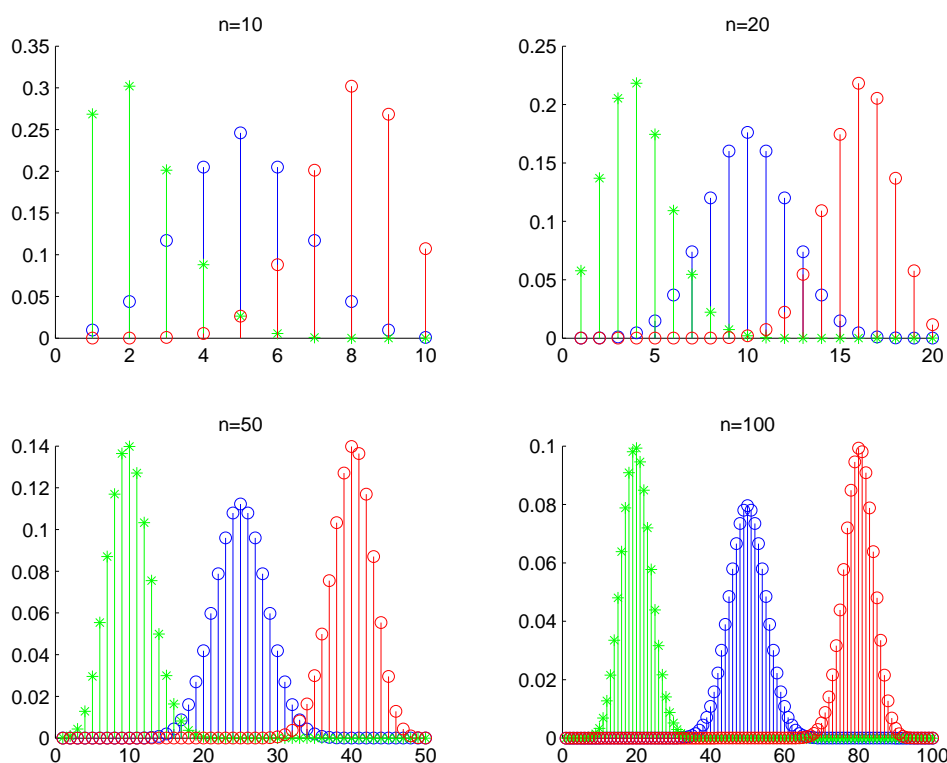
Si la muestra fuera del 20% del total del lote o más grande, el experimento ya no sería binomial pues la probabilidad de sacar el décimo fusible sería distinta a la del primero. En este caso la distribución se llama hipergeométrica.

Una variable aleatoria  $X$  con distribución binomial de parámetros  $n$  y  $p$  puede ser expresada en términos de  $n$  variables aleatorias independientes Bernoulli. Específicamente, sean  $X_1, \dots, X_n$   $n$  variables independientes Bernoulli de parámetro  $p$ , entonces

$$X = X_1 + \dots + X_n$$

**Teorema 4.5.2** Una variable aleatoria con distribución binomial basada en  $n$  pruebas con probabilidad de éxito  $p$  tiene  $E(Y) = np$  y  $V(Y) = np(1 - p)$ .



Figura 4.2: Variables binomiales con diferentes parámetros  $p = 0,2, 0,5, 0,8$ **Demostración:**

Si escribimos a  $X$  como suma de variables Bernoulli independientes, y aplicamos la linealidad de la esperanza, resulta

$$E(X) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = p + \cdots + p = n \cdot p$$

Como las variables  $X_i$  son independientes,

$$Var(X) = Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n) = np(1 - p)$$

**Ejemplo 4.5.3** En un negocio de electrodomésticos, el 25 % de los clientes que compran a crédito incurrir en atrasos en el pago de las cuotas, según los registros de la empresa. Si esta situación se mantiene en el tiempo, ¿cuál es la probabilidad de que entre los próximos 10 clientes que compren a crédito, más de 3 incurran en mora?

**Resolución**

Podemos considerar que los próximos 10 clientes que compren a crédito es una muestra aleatoria de la población de personas que compran a crédito ( lo cual es bastante arriesgado de suponer). Si esto vale, entonces el número de clientes que entran en mora de los 10 que considero es una variable binomial de parámetro  $p = 0,25$ . La probabilidad de que más de tres entren en mora de los 10 es

$$P(X \geq 3) = 1 - P(X < 3) = 1 - p(0) - p(1) - p(2) = 1 - \binom{10}{0} 0,75^{10} - \binom{10}{1} 0,25 \cdot 0,75^9 - \binom{10}{2} 0,25^2 \cdot 0,75^8$$

**4.5.3. Distribución Hipergeométrica**

**Definición 4.5.3** Sea  $X$  una variable aleatoria con densidad

$$p_X(x) = \frac{\binom{r_1}{x} \binom{r - r_1}{n - x}}{\binom{r}{n}} \quad 0 \leq x \leq n, \quad n \leq r_1 \leq r$$

Se dice entonces que  $X$  tiene densidad Hipergeométrica de parámetros  $(r_1, r, n)$ .

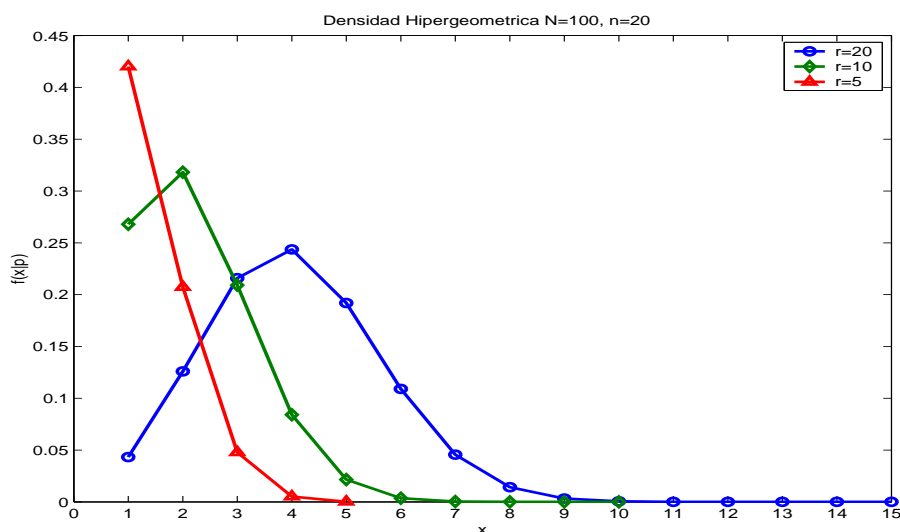


Figura 4.3: Distribución hipergeométrica con parámetros  $r = 100$ ,  $r_1 = 20$  y  $n = 10$ .

Esta distribución aparece en muestreos sin reposición de poblaciones de elementos con dos características diferentes, defectuosos y no defectuosos, personas a favor o en contra de un candidato, etc, en donde se observa solo el número de elementos en cada clase. El siguiente ejemplo teórico describe este caso en detalle.

**Ejemplo 4.5.4** Consideremos una población de  $r$  objetos, de los cuales  $r_1$  son de un tipo y  $r_2 = r - r_1$  son de otro tipo. Supongamos que una muestra aleatoria de tamaño  $n \leq r_1$  se extrae de la población (sin reposición). Sea  $X$  el número de los objetos del primer tipo en la muestra. Entonces  $X$  es una variable aleatoria cuyos valores posibles son  $0, 1, \dots, n$ . Calculemos su función de densidad discreta. Todas las posibles muestras son equiprobables, pues son elegidas al azar, así que la probabilidad de elegir una muestra particular es

$$P(\{x_1, \dots, x_n\}) = \frac{1}{\binom{r}{n}}$$

El evento  $X = x$  consta de todas las muestras (subconjuntos) que constan de  $x$  elementos del tipo 1, por lo cual hay  $\binom{r_1}{x}$  formas de elegir los elementos del tipo 1 y por cada elección hay  $\binom{r - r_1}{n - x}$  formas de elegir los elementos del tipo 2. Entonces

$$p_X(x) = P(X = x) = \frac{\binom{r_1}{x} \binom{r - r_1}{n - x}}{\binom{r}{n}}$$

**Proposición 4.5.1** Si  $X$  tiene distribución hipergeométrica de parámetros  $(r_1, r, n)$  entonces

$$E(X) = \frac{nr_1}{r} \quad \text{Var}(X) = \frac{r_1 n (r - r_1) (r - n)}{r^2 (r - 1)}$$

**Ejemplo 4.5.5** La cámara legislativa de cierta ciudad está compuesta por 30 concejales de los cuales 18 pertenecen al partido A y los otros 12 al partido B. Se elige al azar una comisión de 5 concejales.

1. ¿Cuál es la probabilidad de que todos los miembros de la comisión elegida sean del mismo partido?
2. ¿Cuál es la probabilidad de que la mayoría de los miembros de la comisión pertenezcan al partido A?

**Resolución**

Consideremos la variable aleatoria  $X$  número de concejales del partido A de la comisión, es una variable hipergeométrica de parámetros  $r_1 = 18$ ,  $r = 30$  y  $n = 5$ . La probabilidad de que todos los concejales elegidos sean de un mismo partido es  $P(X = 5) + P(X = 0)$  pues ocurre que los 5 pertenecen al partido A o bien los 5 no pertenecen al partido A.

$$P(X = 5) + P(X = 0) = \frac{\binom{18}{5}\binom{12}{0}}{\binom{30}{5}} + \frac{\binom{18}{0}\binom{12}{5}}{\binom{30}{5}}$$

La probabilidad de que la mayoría de los concejales sea del partido A es  $P(X \geq 3)$ , y se calcula

$$P(X \geq 3) = \frac{\binom{18}{3}\binom{12}{2}}{\binom{30}{5}} + \frac{\binom{18}{4}\binom{12}{1}}{\binom{30}{5}} + \frac{\binom{18}{5}\binom{12}{0}}{\binom{30}{5}}$$

**4.5.4. Distribución Geométrica**

**Definición 4.5.4** Una variable aleatoria  $X$  tiene distribución geométrica de parámetro  $p$  si su densidad de probabilidad está dada por

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

Esta es una distribución bien definida pues  $p_X(k) \geq 0$  y

$$\sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty} (1 - p)^{k-1} p = p + p \sum_{n=1}^{\infty} (1 - p)^n = p + \frac{(1 - p)}{1 - (1 - p)} p = p + (1 - p) = 1$$

usando el límite de la serie geométrica de razón  $(1 - p)$ .

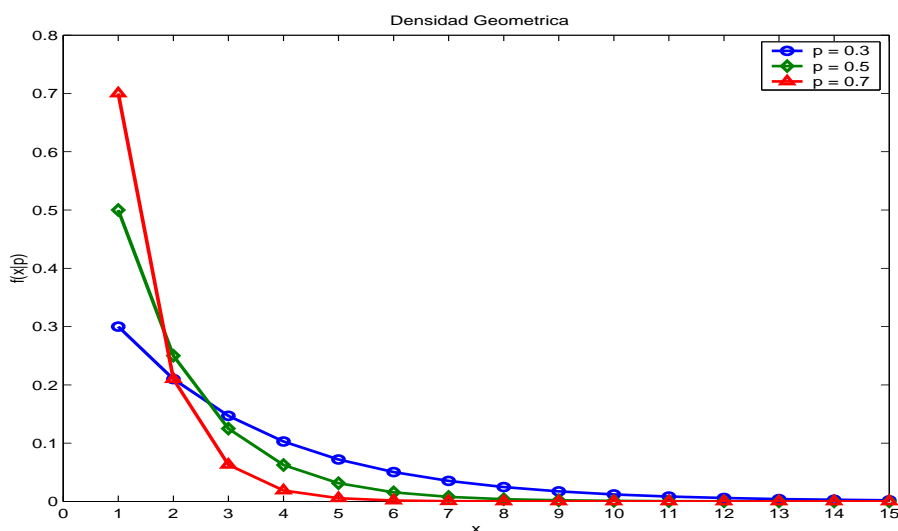


Figura 4.4: Distribución geométrica con parámetro  $p = 0,3$ .

**Ejemplo 4.5.6** Una experiencia tiene como resultados posibles éxito con probabilidad  $p$  y fracaso con probabilidad  $(1 - p)$ . Supongamos que realizamos la experiencia hasta obtener el primer éxito. Registramos  $X$  el número de repeticiones necesarias. Encuentre la distribución de probabilidad de  $X$ .

**Resolución**

$X$  es una variable aleatoria discreta cuya densidad de probabilidad está dada por

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

dado que las realizaciones son todas independientes. Por lo cual  $X$  tiene distribución geométrica de parámetro  $p$ .

**Ejemplo 4.5.7** Supongamos que arrojamamos un dado hasta observar el primer as. La probabilidad de observar un as en una sola tirada es  $1/6$ , por lo cual si  $X$  es el número de tiradas necesarias hasta observar el primer as,  $X$  tiene distribución geométrica de parámetro  $p = 1/6$ . La variable  $X$  no puede tomar el valor 0, pues eso implicaría que el dado nunca se arrojó. Sin embargo, en otros casos, el valor cero implicaría que el proceso no empezó. Por ejemplo, si consideramos  $Y$  el número de horas (completas) hasta que falla el tubo de un televisor, podemos considerar que  $Y = 0$  representa el evento "el televisor falló antes de la primera hora", lo cual incluye el hecho de que el televisor puede no encender. En ese caso, la distribución de la variable es

$$p_Y(k) = P(Y = k) = p(1 - p)^k \quad k = 0, 1, 2, 3, \dots$$

y se sigue llamando geométrica de parámetro  $p$ . Esto se explica pensando que la variable  $X$  mide el número de intentos hasta obtener el primer éxito, por lo cual

$$p_X(k) = p(1 - p)^{k-1} \quad k = 1, 2, \dots$$

e  $Y$  mide el número de fracasos hasta obtener el primer éxito, por lo cual

$$p_Y(k) = p(1 - p)^k \quad k = 0, 1, \dots$$

**Ejemplo 4.5.8** Suponga que la probabilidad de una falla de funcionamiento de un motor de avión, durante un período de una hora es  $p = 0,02$ . Encuentre la probabilidad de que un motor no tenga fallas en dos horas.

**Resolución**

Sea  $Y$  es la variable que mide durante qué hora posterior al encendido se registró la primera falla, es decir,  $Y = 3$  si el motor falló durante la tercer hora. Entonces

$$P(\text{sobrevivir dos horas completas}) = P(\text{fallar durante la tercera hora o después}) = P(Y \geq 3) = \sum_{y=3}^{\infty} p_Y(y)$$

donde  $p_Y$  es la densidad geométrica. Como  $\sum_{y=1}^{\infty} p_Y(y) = 1$ ,

$$P(\text{sobrevivir 2 horas}) = 1 - \sum_{y=1}^2 p_Y(y) = 1 - p - qp = 1 - 0,02 - (0,98)(0,02) = 0,9602$$

**Proposición 4.5.2** Si  $X$  tiene distribución geométrica de parámetro  $p$ , entonces

$$E(X) = \frac{1}{p} \quad \text{Var}(X) = \frac{1-p}{p^2}$$

**Ejemplo 4.5.9** Sea  $X$  una variable con distribución geométrica de parámetro  $p$ . Entonces se cumple que:

1. La función de distribución de  $X$  es

$$F_X(t) = 1 - (1 - p)^{[t]} \quad t \geq 1, t \in \mathbb{R}$$

2.  $P(X \geq k) = (1 - p)^{k-1}$  para todo  $k \in \mathbb{N}$ .
3. La distribución geométrica sufre de amnesia, esto es

$$P(X \geq s + t | X > t) = P(X \geq s)$$

**Ejemplo 4.5.10** Sean  $X$  e  $Y$  variables independientes con distribución geométrica de parámetro  $p$ . Entonces

1. La función de distribución del  $\min(X, Y)$  es una variable geométrica con parámetro  $2p - p^2$ .
2.  $P(\min(X, Y) = X) = P(Y \geq X) = p/(2p - p^2)$ .

### 4.5.5. Distribución Binomial negativa

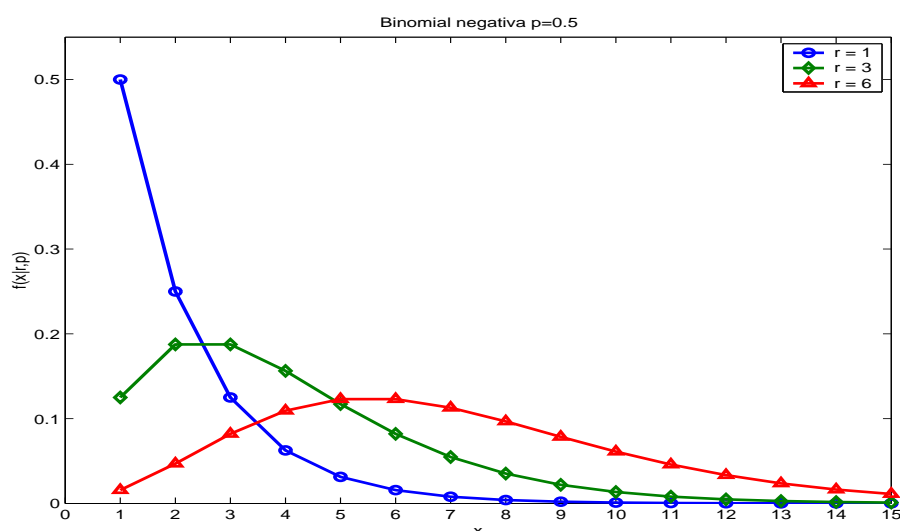


Figura 4.5: Distribución Binomial negativa con parámetros  $r = 1, 3, 6$  y  $p = 0,5$ .

La distribución binomial negativa surge de una generalización de la distribución geométrica. Suponga que se realiza una sucesión de experimentos independientes, cada uno de ellos con probabilidad de éxito  $p$ , hasta que se observen  $r$  éxitos en total. Sea  $X$  el número de experimentos necesarios. Para encontrar  $P(X = k)$  se puede realizar el siguiente argumento: cada sucesión de resultados producto del experimento es de la forma

$$(X = k) = \{(x_1, \dots, x_{k-1}, 1) \mid x_i = 0, 1, \quad r \text{ símbolos } 1 \text{ y } k - r \text{ símbolos } 0\}$$

todos con probabilidad  $p^r(1 - p)^{k-r}$ , dado que los experimentos son independientes. El cardinal del conjunto  $(X = k)$  es  $\binom{k-1}{r-1}$  pues el último experimento es siempre un éxito, y los restantes  $(r - 1)$  éxitos pueden ser asignados a los restantes  $(k - 1)$  experimentos en  $(k - 1)! / (r - 1)!(k - r)!$  formas. Por lo cual

$$P(X = k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r} \quad k = r, r + 1, \dots$$

**Definición 4.5.5** Sea  $X$  una variable aleatoria que tiene densidad

$$p_X(k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r} \quad k = r, r + 1, \dots$$

entonces  $X$  se dice que tiene distribución binomial negativa de parámetros  $r, p$ .

Es útil saber que una variable  $X$  con distribución binomial negativa puede ser vista como suma de  $r$  variables independientes con distribución geométrica de parámetro  $p$ :  $X_1$  el número de experimentos necesarios para obtener el primer éxito más  $X_2$  el número de experimentos necesarios desde que se obtuvo el primer éxito hasta el segundo éxito,..., más  $X_r$  el número de experimentos necesario desde que se obtuvo el  $(r-1)$  éxito para observar el  $r$ -ésimo éxito.

**Proposición 4.5.3** Si  $X$  tiene distribución binomial negativa con parámetros  $r, p$ , entonces

$$E(X) = \frac{r}{p} \quad \text{Var}(X) = r \left( \frac{1-p}{p^2} \right)$$

**Demostración:**

La variable  $X$  se escribe como

$$X = X_1 + \cdots + X_r$$

donde cada  $X_i$  representa el número de repeticiones necesarias hasta obtener el primer éxito (después de haber obtenido el  $i-1$ ). Son todas independientes pues las repeticiones del experimento son independientes y son geométricas de parámetro  $p$ . Por la linealidad de la esperanza y la independencia de las variables

$$\begin{aligned} E(X) &= E(X_1 + \cdots + X_r) = E(X_1) + \cdots + E(X_r) = \frac{1}{p} + \cdots + \frac{1}{p} = \frac{r}{p} \\ \text{Var}(X) &= \text{Var}(X_1 + \cdots + X_r) = \text{Var}(X_1) + \cdots + \text{Var}(X_r) = \frac{(1-p)}{p^2} + \cdots + \frac{1-p}{p^2} = \frac{r(1-p)}{p^2} \end{aligned}$$

**Ejemplo 4.5.11** La probabilidad de que el lanzamiento de un satélite sea exitoso es 0.8. Supóngase que se hacen ensayos hasta que ocurren 3 lanzamientos exitosos.

1. ¿Cuál es la probabilidad de que sean necesarios 6 intentos?
2. ¿Cuál es la probabilidad de que sean necesarios menos de 6 intentos?
3. Si cada ensayo de lanzamiento cuesta 5000 y cada ensayo que falla tiene un costo adicional de 500. Calcular el costo esperado de todo el procedimiento.

**Resolución**

La variable aleatoria que mide el número de intentos necesarios hasta alcanzar tres lanzamientos exitosos del satélite es una variable binomial negativa con parámetros  $r = 3$  y  $p = 0,8$ . Entonces

$$P(X = 6) = \binom{5}{2} (0,8)^3 (0,2)^3$$

Si queremos calcular la probabilidad de que se necesiten menos de seis intentos, entonces debo calcular

$$P(X < 6) = P(X = 5) + P(X = 4) + P(X = 3) = \binom{4}{2} (0,8)^3 (0,2)^2 + \binom{3}{2} (0,8)^3 (0,2) + (0,8)^3$$

Ahora necesitamos calcular la esperanza de una nueva variable, el costo del procedimiento

$$C = 5000X + 500(X - 3)$$

Conociendo cual es la fórmula de la esperanza de una variable binomial negativa, y usando el hecho que la esperanza es lineal, resulta

$$E(C) = 5000E(X) + 500(E(X) - 3) = 5000 \frac{3}{0,8} + 500 \left( \frac{3}{0,8} - 3 \right)$$

### 4.5.6. Distribución Poisson

Supongamos que queremos encontrar la distribución de probabilidad del número de accidentes de automóvil en una intersección particular durante un período de tiempo de una semana. A primera vista, esta variable aleatoria, el número de accidentes, puede no estar ni remotamente relacionada con una variable aleatoria binomial, pero veremos que existe una relación interesante.

Pensemos en el período de tiempo que fijamos, una semana, y dividamos-lo en  $n$  sub-intervalos, cada uno de ellos tan pequeño que a lo sumo un accidente puede ocurrir con probabilidad diferente de cero. Denotemos a la probabilidad de tener un accidente en un sub-intervalo cualquiera por  $p$ , entonces, para propósitos prácticos, tenemos

$$\begin{aligned} P(\text{ningún accidente en un sub-intervalo}) &= 1 - p \\ P(\text{exactamente un accidente en un sub-intervalo}) &= p \\ P(\text{más de un accidente en un sub-intervalo}) &= 0 \end{aligned}$$

Entonces el número total de accidentes en una semana es el número de sub-intervalos que contienen un accidente. Si la ocurrencia de un accidente en un intervalo de tiempo puede ser considerada independiente de la ocurrencia de otro en otro intervalo de tiempo, el número total de accidentes tienen una distribución binomial.

A pesar de que no hay una única forma de elegir los sub-intervalos, y por lo tanto no conocemos  $n$  ni  $p$ , parece razonable que si dividimos a la semana en un número mayor de sub-intervalos, la probabilidad  $p$  de tener un accidente en ese sub-intervalo menor va a decrecer. Supongamos que  $p_n$  decrece de tal forma que  $np_n = \lambda$ , una constante, para todo  $n$ , entonces, haciendo crecer  $n$  hacia el infinito, tenemos

$$\begin{aligned} \lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} &= \lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1) \dots (n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \end{aligned}$$

Observando que

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

y que los otros términos a la derecha del límite tienen límite 1, (no nos olvidemos que  $n$  se está haciendo cada vez más grande!!), se obtiene

$$p_Y(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

donde  $\lambda$ , esa constante que supusimos existe, es la tasa de accidentes en la semana. Vamos a llamar a toda variable aleatoria que tiene esa densidad de probabilidad variable Poisson, y a pesar de ser un poquito difícil, con exponenciales y factoriales, es extremadamente útil.

**Definición 4.5.6** Sea  $Y$  una variable aleatoria con densidad

$$p_Y(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k \geq 0$$

Entonces  $Y$  se dice que tiene distribución de Poisson de parámetro  $\lambda$ .

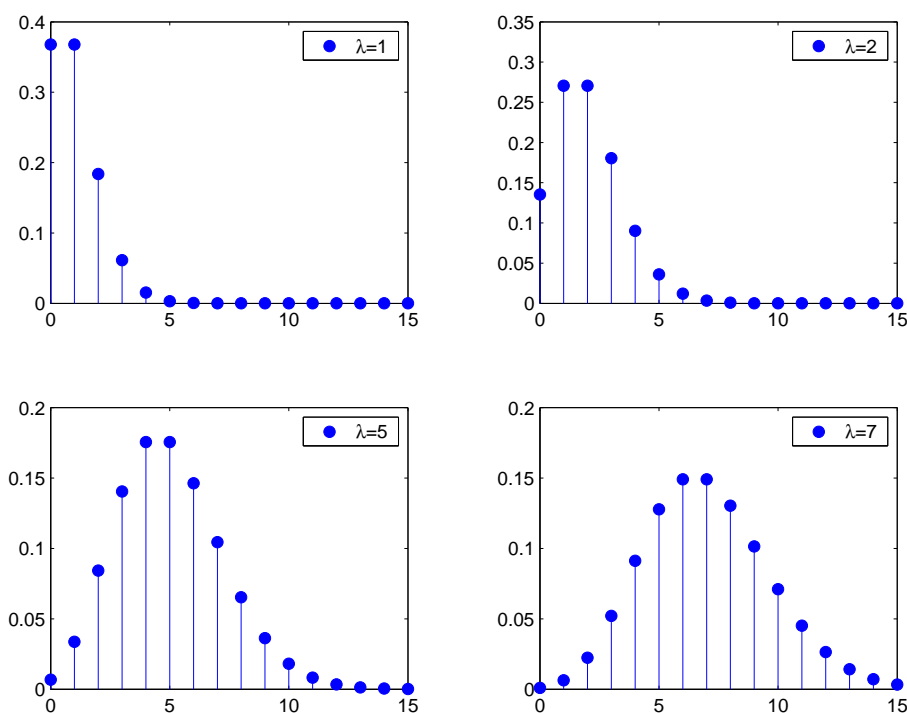


Figura 4.6: Distribución Poisson con parámetros  $\lambda = 1, 2, 5, 7$ .

La distribución de Poisson es usualmente un buen modelo para el número  $Y$  de eventos raros ocurridos en espacio, tiempo o volumen, donde  $\lambda$  es el promedio de los valores de  $Y$ . Algunos ejemplos son el número de partículas radioactivas que decaen en un intervalo de tiempo, el número de errores que un mecanógrafo hace por página, y el número de accidentes de tráfico, o incendios declarados, por unidad de tiempo.

**Ejemplo 4.5.12** Al formar números binarios con 25 dígitos, la probabilidad de que aparezca un dígito incorrecto es 0.00002. Si los errores son independientes, ¿cuál es la probabilidad de encontrar cero, uno o más de un dígito incorrecto en un número binario de 25 dígitos?. Si la computadora genera  $10^6$  de tales números de 25 dígitos por segundo, ¿cuál es la probabilidad de que se forme al menos número incorrecto durante cualquier periodo de un segundo?

#### Resolución

El número de dígitos incorrectos en una serie de 25 dígitos, suponiendo que los errores son independientes entre sí, es una variable binomial,  $X$ , de parámetros  $n = 25$  y  $p = 0,002$ . Entonces, las probabilidades de encontrar cero, uno o más de un dígito incorrecto en un número binario de 25 dígitos son

$$P(X = 0) = (0,99998)^{25} = 0,9995 \quad P(X = 1) = 25(0,99998)^{24}(0,00002) = 0,0004997$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0,99998)^{25} = 0,0005$$

Ahora, si consideramos cada número de 25 dígitos como correcto o incorrecto, y estudiamos todos los números generados en el lapso de un segundo, tenemos una nueva variable binomial  $Y$ , el número de números binarios incorrectos realizados en el periodo de un segundo. Esta variable  $Y$  tiene parámetros  $n = 10^6$  y  $p = P(X \geq 1) = 0,0005$ . Observemos que  $n$  es muy grande, por lo cual podemos considerar el problema Poisson en vez de binomial, con parámetro  $\lambda = np = 500$ . Con esa tasa de error por segundo, la probabilidad de que se forme por lo menos un número incorrecto es

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-\lambda} = 1 - e^{-500} = 1 - 7,124576406741286 \times 10^{-218} \sim 1$$

La probabilidad de cometer un error en un dígito es muy baja, sin embargo al observar tantas instancias, los errores son inevitables. Por ello es que los códigos computacionales se crean con alta tolerancia a errores, así como nosotros



podemos leer un libro y entenderlo a pesar de las faltas de ortografía. Las faltas de gramática introducen problemas más graves que las de ortografía en la lectura de un libro, y producen problemas insalvables en la estructura de un programa de computación.

**Proposición 4.5.4** Si  $Y$  es una variable con distribución Poisson de parámetro  $\lambda$ , entonces

$$E(Y) = \lambda \quad \text{Var}(Y) = \lambda$$

**Ejemplo 4.5.13** El gerente de una planta industrial está planeando comprar una nueva máquina y tiene dos opciones, una máquina del tipo  $A$  y una del tipo  $B$ . El número de ajustes diarios que requiere la manutención de la máquina  $A$  es una variable Poisson con media  $0,10t$  donde  $t$  denota el número de horas de operación diaria. El número de reparaciones diarias  $Y_2$  de la máquina  $B$  es una variable aleatoria Poisson con media  $0,12t$ . El costo diario de operar la máquina  $A$  es  $C_A(t) = 10t + 30Y_1^2$ , para  $B$  es  $C_B(t) = 8t + 30Y_2^2$ . Asumamos que las reparaciones toman un tiempo insignificante y que cada noche las máquinas se ajustan de tal forma que operan esencialmente como nuevas al comienzo de cada día. ¿Qué máquina minimiza el costo diario esperado si un día de trabajo consta de 10 horas, y si consta de 20 horas?.

#### Resolución

El costo diario esperado para  $A$  es

$$\begin{aligned} E(C_A(t)) &= E(10t + 30Y_1^2) = 10t + 30E(Y_1^2) \\ &= 10t + 30[\text{Var}(Y_1) + [E(Y_1^2)]] = 10t + 30[0,10t + (0,10t)^2] \\ &= 13t + 0,3t^2 \end{aligned}$$

En forma similar

$$\begin{aligned} E(C_B(t)) &= E(8t + 30Y_2^2) = 8t + 30E(Y_2^2) \\ &= 8t + 30[\text{Var}(Y_2) + [E(Y_2^2)]] = 8t + 30[0,12t + (0,12t)^2] \\ &= 11,6t + 0,432t^2 \end{aligned}$$

Por lo cual, si  $t = 10$  horas

$$E(C_A(10)) = 160 \quad \text{y} \quad E(C_B(10)) = 159,2$$

y la máquina  $B$  minimiza el costo esperado.

Si  $t = 20$  horas,

$$E(C_A(20)) = 380 \quad \text{y} \quad E(C_B(20)) = 404,8$$

En conclusión,  $B$  es más económica para períodos cortos de tiempo dado su menor costo de operación, pero para períodos más largos  $A$  es más económica pues tiende a ser reparada menos frecuentemente.

### 4.5.7. Distribución Multinomial

La distribución multinomial es una importante generalización de la distribución binomial al caso de más de dos posibles resultados del experimento, y corresponde a la siguiente descripción.

Supongamos que repetimos en forma independiente  $n$  veces un experimento que tiene  $r$  posibles resultados  $A_1, \dots, A_r$ , y que forman una partición del espacio  $\Omega$ . Por ejemplo, en el juego de ruleta, podemos definir  $A_1$ =sale par,  $A_2$ = sale impar y  $A_3$ = sale el cero.

Sean  $p_1 = P(A_1), \dots, p_r = P(A_r)$ , entonces como los  $A_i$  son una partición de  $\Omega$ , tengo que  $p_1 + \dots + p_r = 1$ . Defino las siguientes  $r$  variables,  $X_i$ =número de veces que ocurrió el evento  $A_i$  en las  $n$  repeticiones. La distribución conjunta del vector  $(X_1, \dots, X_r)$  se llama distribución multinomial. Calculemos la densidad conjunta  $p_{\mathbf{X}}$  del vector

$$p_{\mathbf{X}}(x_1, \dots, x_r) = P(X_1 = x_1, \dots, X_r = x_r)$$

Comencemos por ver qué puntos integran el evento

$$C = (X_1 = x_1) \cap \cdots \cap (X_r = x_r)$$

Si observamos  $(x_1, \dots, x_r)$  entonces  $A_1$  ocurrió  $x_1$  veces,  $A_2$  ocurrió  $x_2$  veces, ...,  $A_r$  ocurrió  $x_r$  veces en la sucesión de repeticiones. Denotemos cada punto observado como una sucesión de valores  $i = 1, 2, \dots, r$ , cada uno de estos repetido  $x_i$  veces. Entonces la probabilidad de cada punto

$$11111\dots 2222\dots rrrr$$

y sus permutaciones es  $p_1^{x_1} \dots p_r^{x_r}$ . La cantidad de puntos distintos con éstas características son menos que las permutaciones de  $n$  símbolos pues si intercambio dos símbolos iguales de lugar tengo la misma sucesión. Por lo cual debo dividir  $n!$  por las permutaciones de los símbolos repetidos de la sucesión. Esto es, el número de sucesiones distintas con  $x_i$  símbolos  $i$ ,  $i = 1, \dots, r$  es

$$\frac{n!}{x_1!x_2!\dots x_n!}$$

por lo cual

$$p_{\mathbf{X}}(x_1, \dots, x_r) = P(X_1 = x_1, \dots, X_r = x_r) = \frac{n!}{x_1!x_2!\dots x_n!} p_1^{x_1} \dots p_r^{x_r}$$

**Definición 4.5.7** Se dice que el vector  $\mathbf{X} = (X_1, \dots, X_r)$  tiene distribución multinomial de parámetros  $(p_1, \dots, p_r)$  y  $n$  si su densidad discreta está dada por

$$p_{\mathbf{X}}(x_1, \dots, x_r) = P(X_1 = x_1, \dots, X_r = x_r) = \frac{n!}{x_1!x_2!\dots x_n!} p_1^{x_1} \dots p_r^{x_r}$$

**Proposición 4.5.5** Sea  $\mathbf{X} = (X_1, \dots, X_r)$  con distribución multinomial de parámetros  $(p_1, \dots, p_r)$  y  $n$ . Entonces cada  $X_i$  tiene distribución binomial de parámetros  $p_i$  y  $n$ .

**Demostración:** Sale directamente del hecho de que podemos considerar éxito observar el resultado  $A_i$  y fracaso todo lo demás. Por lo cual el experimento compuesto es binomial de parámetros  $p_i$  y  $n$  y el número de éxitos observados en las  $n$  repeticiones es  $X_i$ .

**Ejemplo 4.5.14** Suponga que la probabilidad de una falla de funcionamiento de un motor de avioneta con autonomía de vuelo de una hora, durante un período de 100 horas, es  $p = 0,02$ . Si se detecta una falla antes de que se cumplan 500 horas de operación se dice que el motor es defectuoso, si se detecta una falla después de las 500 horas pero antes de las 800 horas de operación se dice que el motor tiene fallas leves y si se detecta una falla después de las 800 horas de operación se dice que el motor tiene el rango de fallas usual.

Si se elige una muestra aleatoria de 15 motores, ¿cuál es la probabilidad de que 5 sean defectuosos, 5 tengan fallas leves y 5 tengan el rango de fallas usual?, ¿cuál es la probabilidad de que ninguno sea defectuoso, 2 tengan fallas leves y 12 tengan el rango de fallas usual?

**Resolución**

Sea  $Y$  es la variable que mide (en cientos de horas) el tiempo de encendido hasta la primera falla, es decir,  $Y = 3$  si el

motor falló durante el tercer período de 100 horas. Entonces, se tiene

$$\begin{aligned}
 p_1 &= P(\text{fallar antes de las 500 horas}) = P(Y \leq 5) = \sum_{k=1}^5 p_Y(k) \\
 &= (0,02) + 0,02(0,98) + 0,02(0,98)^2 + 0,02(0,98)^3 + 0,02(0,98)^4 = 0,0960792 \\
 p_2 &= P(\text{fallar después de las 500 horas y a lo sumo en las 800 horas}) = P(6 \leq Y \leq 8) \\
 &= \sum_{k=6}^8 p_Y(k) = 0,02(0,98)^5 + 0,02(0,98)^6 + 0,02(0,98)^7 = 0,0531578 \\
 p_3 &= P(\text{fallar después de las 800 horas}) = P(Y \geq 9) = 1 - \sum_{k=1}^8 p_Y(k) \\
 &= 1 - [(0,02) + 0,02(0,98) + 0,02(0,98)^2 + 0,02(0,98)^3 + 0,02(0,98)^4 + 0,02(0,98)^5 + 0,02(0,98)^6 + 0,02(0,98)^7] \\
 &= 0,850763
 \end{aligned}$$

donde  $p_Y$  es la densidad geométrica.

Entonces, si  $X_1$ =número de motores defectuosos de la muestra,  $X_2$  = número de motores con fallas leves y  $X_3$  = número de motores con rango de fallas usual, el vector  $(X_1, X_2, X_3)$  es multinomial con probabilidades  $p_1, p_2, p_3$ ,  $n = 15$ . Por lo cual

$$\begin{aligned}
 P(5 \text{ sean defectuosos, } 5 \text{ tengan fallas leves y } 5 \text{ tengan el rango de fallas usual}) &= P(X_1 = 5, X_2 = 5, X_3 = 5) = \frac{15!}{5!5!5!} p_1^5 p_2^5 p_3^5 \\
 &= 4,9133e-007 \\
 P(0 \text{ sean defectuosos, } 2 \text{ tengan fallas leves y } 13 \text{ tengan el rango de fallas usual}) &= P(X_1 = 0, X_2 = 2, X_3 = 13) \\
 &= \frac{15!}{0!2!13!} p_1^0 p_2^2 p_3^{13} = 0,7355
 \end{aligned}$$

#### 4.5.8. Suma de variables aleatorias independientes

Veamos ahora que ocurre cuando sumamos variables independientes del tipo de las vistas en la sección anterior.

##### Proposición 4.5.6

1. Sean  $X_1, \dots, X_r$  variables aleatorias independientes geométricamente distribuidas con parámetro  $p$ . Entonces  $X_1 + \dots + X_r$  tiene distribución binomial negativa de parámetros  $r, p$ .
2. Sean  $X_1, \dots, X_r$  variables aleatorias independientes con distribución binomial con parámetro  $n_i, p$ . Entonces  $X_1 + \dots + X_r$  tiene distribución binomial de parámetros  $n_1 + \dots + n_r, p$ .
3. Sean  $X_1, \dots, X_r$  variables aleatorias independientes con distribución binomial negativa con parámetro  $n_i, p$ . Entonces  $X_1 + \dots + X_r$  tiene distribución binomial negativa de parámetros  $n_1 + \dots + n_r, p$ .
4. Sean  $X_1, \dots, X_r$  variables aleatorias independientes con distribución Poisson con parámetro  $\lambda_i$ . Entonces  $X_1 + \dots + X_r$  tiene distribución Poisson con parámetro  $\lambda_1 + \dots + \lambda_r$ .

En el caso de la binomial, y la binomial negativa, sumar variables independientes e idénticamente distribuidas es como considerar un experimento del mismo tipo con mayor cantidad de repeticiones. El caso particular de la distribución Poisson es muy interesante, pues suele usarse mucho en ejercicios, se dice solamente la tasa por unidad (segundo, centímetro, página, etc) y luego se considera un mayor de unidades. Veamos ejemplos de esto.

**Ejemplo 4.5.15** Supongamos que un libro de 585 páginas contiene 43 errores tipográficos. Si los errores están distribuidos aleatoriamente a lo largo del libro, ¿cuál es la probabilidad de que 10 páginas seleccionadas al azar estén libres de errores?

**Resolución**

Para resolver este ejercicio debemos realizar muchas suposiciones razonables, que constituyen el modelo del experimento. La primera suposición es que el número de errores por página es una variable aleatoria Poisson con media  $\lambda$ . La segunda suposición es que las 585 páginas de este libro generan una muestra aleatoria  $(X_1, \dots, X_{585})$  de la misma variable Poisson con parámetro  $\lambda$ . La tercera suposición es que el número de errores en las 585 páginas del libro es la esperanza de la variable  $Y = \sum_{i=1}^{585} X_i$ , la cual es una variable Poisson, por ser suma de Poisson independientes. Por lo cual  $43 = E(Y) = 585\lambda$ , y resulta

$$\lambda = \frac{43}{585} = 0,073$$

Entonces, cada  $X_i$  tiene distribución Poisson de parámetro  $\lambda = 0,073$ . Si consideramos 10 páginas elegidas al azar, el número de errores en las 10 páginas elegidas es  $Z = \sum_{k=1}^{10} X_{i_k}$  y como son todas Poisson independientes con el mismo  $\lambda$ , resulta  $Z$  una Poisson con parámetro  $10\lambda = 0,73$ . Entonces

$$P(Z = 0) = \frac{e^{-0,73} \lambda^0}{0!} = 0,4794$$

**Ejemplo 4.5.16** El número de imperfecciones que tiene una placa fotográfica sigue la distribución de Poisson de parámetro 0,1 imperfecciones por  $cm^2$ .

1. Si de tal placa se toma una muestra de  $30 cm^2$ , ¿cuál es la probabilidad de que esa muestra contenga exactamente tres irregularidades?
2. Si ahora se toma en forma independiente 5 muestras de  $30 cm^2$ , ¿cuál es la probabilidad de que exactamente dos de ellas contengan a lo sumo una irregularidad?

**Resolución**

Siguiendo el espíritu del ejemplo anterior, si  $Y$  es el número imperfecciones que tiene una placa fotográfica sigue la distribución de Poisson de parámetro 0,1 imperfecciones por  $cm^2$ . Si tenemos  $30 cm^2$ , entonces el número de imperfecciones  $X$  en toda la placa es la suma de las imperfecciones en cada centímetro cuadrado,  $X = Y_1 + \dots + Y_{30}$ , una suma de variables aleatorias independientes Poisson con el mismo parámetro 0,1. Por lo cual  $X$  también es Poisson, con parámetro  $\lambda = 0,1 + \dots + 0,1 = 30 \times 0,1 = 3$ . Esto quiere decir que podemos esperar una media de 3 imperfecciones por placa de  $30 cm^2$ . Ahora la probabilidad de que la placa contenga **exactamente** 3 irregularidades es

$$P(X = 3) = \frac{\lambda^3}{3!} e^{-\lambda} = \frac{27}{6} e^{-3} = 0,2240$$

Si ahora tomamos 5 muestras independientes de  $30 cm^2$  cada una, la variable  $Z$  que mide el número de placas con a lo sumo una irregularidad es una variable binomial, con parámetro  $n = 5$  y probabilidad  $p = P(X \leq 1)$ , donde  $X$  es la variable Poisson que mide las irregularidades por placa.

$$p = P(X \leq 1) = e^{-\lambda} + \frac{\lambda^2}{2!} e^{-\lambda} = e^{-3} \left(1 + \frac{9}{2}\right) = 0,2738$$

Por lo cual la probabilidad de que exactamente dos tengan a lo sumo una irregularidad es

$$P(Z = 2) = \binom{5}{2} 0,2738^2 (1 - 0,2738)^{5-2}$$

# Capítulo 5

## Variables aleatorias continuas

### 5.1. Introducción

Hasta el momento hemos considerado solo variables aleatorias discretas y sus densidades. Estas variables denotan usualmente el número de objetos de un cierto tipo, tales como el número de bolas rojas de una muestra de  $n$  bolas extraídas al azar, con o sin reposición, o el número de llamadas ingresadas a una central telefónica en un intervalo de tiempo.

Hay otras situaciones en las cuales es mas natural considerar variables aleatorias del tipo “continuo” en vez de discreto. Por ejemplo, en vez de estudiar el número de llamadas ingresadas a una central telefónica, lo importante es la duración en el tiempo de cada llamada.

En forma tentativa, se puede definir una variable continua en un espacio de probabilidad  $\Omega$  como una función  $X$  que toma valores en la recta real, y que toma valores puntuales con probabilidad cero, esto es

$$P(\{\omega \mid X(\omega) = x\}) = 0 \quad x \in \mathbb{R}.$$

Ya comentamos cuando definimos probabilidades que los espacios que tienen infinitos puntos y que no se pueden numerar, no pueden tener definidas probabilidades positivas para cada punto. La suma de todos esos puntos no daría 1. Por eso es que las variables continuas no otorgan probabilidad a un valor particular sino a un subconjunto denso de valores, como un intervalo. Esto implica mayor dificultad a la hora de caracterizar estas distribuciones de probabilidad, pues es necesaria matemática más compleja. Sin embargo, en este texto trabajaremos solo con variables cuya distribución de probabilidad pueda expresarse como áreas bajo curvas, por lo cual solo necesitaremos recordar conceptos de cálculo diferencial e integral.

Para estudiar variables discretas necesitamos aprender a contar elementos de un conjunto finito. Para ello, las técnicas de combinatoria fueron de gran ayuda. Para estudiar variables continuas necesitamos ayuda del cálculo diferencial e integral. Para aquellos que no hayan visto cursos de cálculo, se puede restringir la atención a la distribución normal y la uniforme, y quizás sea de ayuda el Apéndice A, en el cual hemos realizado un breve reseña de las ideas motivadoras de integración y diferenciación. La distribución normal tiene tabulada las áreas bajo la curva, necesarias para calcular probabilidades, y la uniforme tiene muy fácil su cálculo. Sin embargo, ejemplos de modelización estocástica más ajustados a la realidad necesitan distribuciones más precisas. Daremos a continuación dos ejemplos de experimentos que introducen la distribución uniforme.

**Ejemplo 5.1.1** Supongamos que una llamada telefónica entra a una central en cualquier momento del intervalo  $[0, T]$ , y sea  $X$  la variable que dice cual es el valor del tiempo observado. Todo tiempo  $t$  tiene igual probabilidad de ser observado, y esta probabilidad es cero, pues como dijimos anteriormente, hay un número infinito de valores posibles en el intervalo. Pero también, todo conjunto de igual longitud debería tener igual probabilidad de **contener** el tiempo

observado. Por eso la distribución de probabilidad de este experimento es la que otorga a cada segmento su largo, dividido  $T$ , como probabilidad.

**Ejemplo 5.1.2** Otro ejemplo que vimos anteriormente es el experimento de arrojar un dardo a un blanco circular de diámetro 1. Sean  $(X, Y)$  las coordenadas espaciales del punto observado. Estas son dos variables diferentes,  $X$  es la posición en el eje de las abscisas e  $Y$  en el de las ordenadas. Sabemos que este experimento tiene una asignación de probabilidades que solo distingue áreas, por lo cual la probabilidad de cada punto  $(x, y)$  observado es cero. Es mas,

$$P(X = x) = P(X = x, -\sqrt{1-x^2} \leq Y \leq \sqrt{1-x^2}) = 0$$

pues el conjunto es una línea en el círculo y por lo tanto tiene área cero.

También existen las variables mezcla, cuyo espacio de probabilidad consiste en valores aislados e intervalos. Como ejemplo consideremos una variable que mide el tiempo hasta la primera falla de un equipo electrónico. El espacio muestral es la semirrecta de los reales positivos, y es razonable modelar la variable con una distribución continua como la exponencial. Pero sin embargo hay una proporción de equipos que no encienden nunca, por lo cual el valor  $x = 0$  de la variable debería tener una probabilidad positiva. Estas variables se denominan mixtas.

## 5.2. Distribución de probabilidad continua

**Definición 5.2.1** Sea  $X$  una función de un espacio de probabilidad que toma valores en un subconjunto de la recta real que llamaremos  $\mathbf{R}_X$ , rango de  $X$ . Entonces  $X$  es una variable aleatoria continua si

1. el conjunto de los puntos muestrales que va a parar a todos los valores menores o iguales a  $x$ ,

$$(X \leq x) = \{\omega \in \Omega \text{ tal que } X(\omega) \leq x\},$$

es un evento para todo  $x \in \mathbf{R}_X$ ,

2.  $(X = x)$ , el conjunto de todos los puntos muestrales que van a parar al valor específico  $x$  tiene probabilidad cero para todo  $x \in \mathbf{R}_X$

En general, variables aleatorias que expresan mediciones de cantidades físicas, como voltaje, peso, coordenadas en tiempo y espacio, temperatura se describen mejor mediante variables continuas, mientras que variables aleatorias que cuentan objetos o eventos son ejemplos claros de variables discretas.

En algunos casos no es tan clara la distinción, como puede ser el ingreso per capita, que si se lo redondea a sumas enteras se lo considera una variable discreta, y si se lo considera variando en un intervalo de la recta real es una variable continua.

Las variables aleatorias expresan resultados específicos de un experimento, o una experiencia, y su distribución de probabilidad focaliza la información de probabilidad de la situación general en los valores particulares de la variable. Por ejemplo, si pensamos en el problema de elegir una central telefónica que pueda manejar el ingreso y salida de llamadas de una empresa, el número de llamadas entrantes es una variable aleatoria Poisson, y la duración en el tiempo de cada llamada es una variable aleatoria exponencial. Para elegir el tipo de central, hay que tener en cuenta los parámetros de ambas distribuciones, la media de llamadas entrantes y el tiempo medio de llamado. La distribución exponencial es una de las distribuciones continuas más importantes, y es la que surge en muchos casos donde se mide el tiempo de duración de situaciones.

### 5.2.1. Densidad de probabilidad

Ya hemos observado con las variables continuas que hay situaciones que generan distribuciones de probabilidad, y hay modelos de distribuciones que pueden “ajustarse” a experiencias particulares. El ajuste de un modelo es simplemente una idealización que permite calcular una probabilidad, que puede ser muy distante de la realidad si el modelo no es el adecuado. Pero la idealización es necesaria para poder realizar ese cálculo. La función de distribución es esta idealización, contiene toda la información sobre la probabilidad de los eventos que pueden ocurrir, y junto al rango de valores de la variable, caracteriza al “modelo” que ajustamos al problema. Hablaremos de ajuste cuando lleguemos a la parte de inferencia estadística, cuando miremos los datos de nuestras experiencias e intentemos decidir cual es el mejor modelo para esos datos. En algunos casos, cuando la cantidad de datos es suficientemente grande, los teoremas asintóticos, los que calculan distribuciones límites, van a dar un resultado teórico que usualmente en la práctica comete un error muy pequeño.

La siguiente proposición caracteriza las distribuciones de probabilidad de variables continuas. En la figura 5.1 vemos la función de distribución acumulada de una variable discreta Poisson y de una continua exponencial. Vemos que la discreta tiene una distribución acumulada continua a trozos, mientras que la distribución acumulada de la variable continua es una función continua, su gráfica no se corta en ningún momento.

**Proposición 5.2.1** *La función de distribución acumulada de la variable  $X$ , definida por*

$$F(x) = P(X \leq x)$$

*es una función continua si y solamente si  $X$  es una variable continua.*

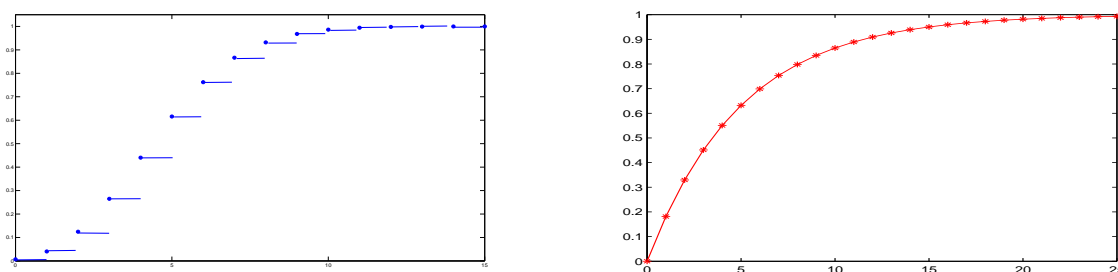


Figura 5.1: Distribución acumulada Poisson con parámetro  $\lambda = 5$  y distribución acumulada exponencial de parámetro  $\lambda = 5$ .

**Definición 5.2.2** *La variable  $X$  es una variable con distribución mixta si*

$$F(x) = P(X \leq x) = c_1 F_{X_1}(x) + (1 - c_1) F_{X_2}(x)$$

*con  $F_{X_1}$  la función de distribución de una variable discreta  $X_1$ ,  $F_{X_2}$  la función de distribución de una variable continua  $X_2$  y  $c_1$  es la probabilidad acumulada de la parte discreta.*

**Ejemplo 5.2.1** Sea  $X$  la vida útil de un componente electrónico, medida en cientos de horas. Estos componentes fallan frecuentemente en el momento en que se introducen en el sistema. Se ha observado que la probabilidad de falla inmediata es  $1/4$ . Si el componente no falla inmediatamente, entonces la distribución acumulada de su vida útil es

$$F(y) = 1 - e^{-y} \quad y > 0$$

Encuentre la distribución de  $X$  y evalúe  $P(X > 10)$ .

**Resolución**

Sabemos que  $c_1 = P(X = 0) = 1/4$ , entonces la variable  $X_1$  está concentrada en  $x = 0$ , con

$$F_{X_1}(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

y la variable  $X_2$  tiene

$$F_{X_2}(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geq 0 \end{cases}$$

Por lo cual

$$F(x) = (1/4)F_{X_1}(x) + (3/4)F_{X_2}(x) = \begin{cases} 0 & x < 0 \\ (1/4) + (3/4)(1 - e^{-x}) & x \geq 0 \end{cases}$$

y

$$P(X > 10) = 1 - P(X \leq 10) = 1 - F(10) = 1 - [(1/4) + (3/4)(1 - e^{-10})] = (3/4)e^{-10}$$

**Definición 5.2.3** Supongamos que  $X$  es una variable aleatoria con distribución  $F_X$  tal que existe una función  $f_X$  no negativa que cumple

$$\int_{-\infty}^{\infty} f(x)dx = 1 \quad F_X(x) = \int_{-\infty}^x f_X(y)dy \quad -\infty < x < \infty$$

entonces  $f_X$  se llama función densidad de probabilidad de la variable  $X$ . Es importante notar que  $f$  no es única.

**Observación 5.2.1** Existen variables aleatorias continuas que no tienen densidad de probabilidad. Las variables que tienen función de distribución absolutamente continua si tienen densidad. Esta función es muy útil, pues podemos encontrar la probabilidad de cualquier evento como la integral sobre un conjunto de la densidad. Por ejemplo,

$$P(a \leq X \leq B) = F(b) - F(a) = \int_{-\infty}^b f(y)dy - \int_{-\infty}^a f(y)dy = \int_a^b f(y)dy$$

o en forma más general,

$$P(X \in A) = \int_A f(y)dy$$

para todo  $A$  tal que  $(X \in A)$  es un evento.

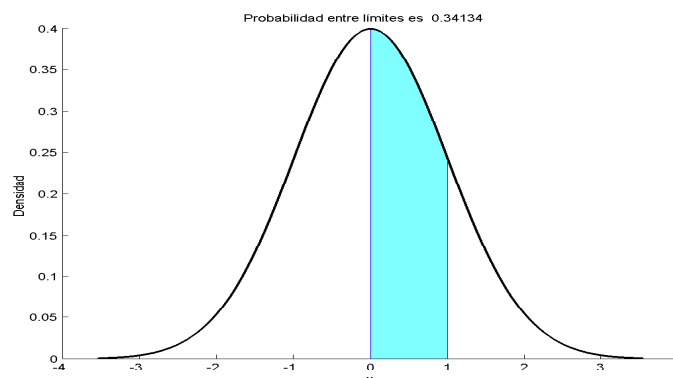


Figura 5.2: Probabilidad calculada como área bajo la curva normal



En la figura 5.2 observamos el área bajo la curva normal que corresponde a

$$P(0 \leq X \leq 1) = \int_0^1 f_X(x)dx = 0,34134$$

si  $X \sim N(0, 1)$ .

También, si  $A$  y  $B$  son subconjuntos de  $\mathbf{R}_X$  entonces

$$P(X \in A/X \in B) = \frac{P(X \in A \cap B)}{P(X \in B)} = \frac{\int_{A \cap B} f(y)dy}{\int_B f(y)dy}$$

### 5.2.2. Esperanza de una variable aleatoria continua

Recordemos que si  $X$  era una variable aleatoria discreta, tal que  $\sum_x |x|p_X(x) < \infty$  entonces  $X$  tenía esperanza y

$$E(X) = \sum_x xp_X(x)$$

En forma análoga, definimos esperanza de variables aleatorias continuas de la siguiente forma.

**Definición 5.2.4** Sea  $X$  un variable aleatoria continua con densidad  $f$ . Decimos que  $X$  tiene esperanza finita si

$$\int_{-\infty}^{\infty} |x|f(x)dx < \infty$$

y en ese caso se define la esperanza como

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx < \infty$$

y la varianza como

$$V(X) = E[(X - E(X))^2]$$

La desviación estándar es la raíz cuadrada positiva de la varianza,  $\sigma = \sqrt{V(X)}$ .

En la figura 5.3 podemos ver dos distribuciones con igual esperanza y distinta varianza, hay datos que tienen mucha más probabilidad de ocurrir bajo una distribución que bajo la otra. Recordemos que las esperanzas son los centros de equilibrio de las densidades, imaginemos que el área bajo la curva es de plomo, y queremos que haga equilibrio sobre la punta de un alfiler, el punto donde deberíamos apoyar el alfiler es la esperanza de a variable, o punto medio de la distribución. La varianza es una medida de la dispersión de los datos alrededor de la media.

En la figura 5.3 dibujamos un distribución normal y una uniforme, la media es la misma, pero ¿cuál tendrá mayor dispersión? La normal que dibujamos concentra sus valores alrededor de la media, en cambio la uniforme que dibujamos no los concentra en absoluto, solo los restringe al intervalo donde la densidad es distinta de cero. La dispersion de la uniforme dibujada es mucho mayor que la de la normal dibujada. Ambas distribuciones corresponden a modelos de problemas muy distintos. Si medimos una cantidad física de forma muy precisa, por ejemplo, el peso de paquete de papas fritas de (supuestamente) 200 gramos, podemos decir que la distribución de probabilidad de los paquetes de papas fritas fabricados por una máquina particular siguen una distribución normal con media 200 gramos. Si la máquina esá bien calibrada la desviación estándar de dicha normal debería ser muy chica, del orden del peso de una papa frita. En cambio un generador de números aleatorio va a dispersar sus valores uniformemente en el intervalo  $[0, 1]$ , y su media será el punto central del intervalo.

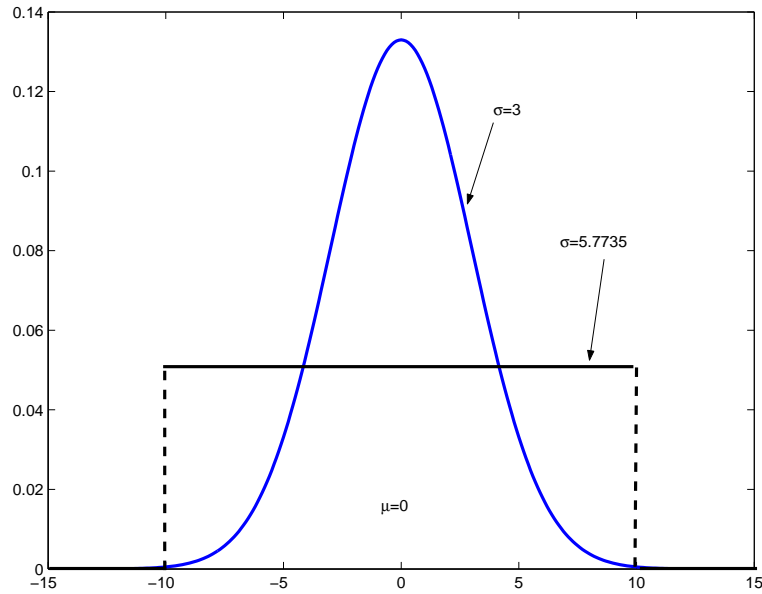


Figura 5.3: Dos distribuciones con igual esperanza y diferente desviación standard.

### 5.2.3. Cómputo de densidades

En muchas aplicaciones la forma más fácil de calcular densidades de variables aleatorias continuas es diferenciar a la función de distribución acumulada  $F$  en los puntos en los que ésta puede ser derivada, siempre que la derivada sea una función continua salvo en un número discreto de puntos, (recordemos que  $F$  es siempre continua). En esos puntos  $f(x) = F'(x)$ . Si  $F$  no es derivable en un conjunto numerable de puntos, aún así podemos definir  $f$  usando la expresión encontrada en los puntos donde  $F$  sí es derivable (y su derivada continua) y cero en los otros puntos donde no es derivable. Veamos un ejemplo de esto.

**Ejemplo 5.2.2** Supongamos que tiramos un dardo en un blanco circular de radio  $R$  totalmente al azar, y sea  $X$  la variable que denota la distancia entre el punto observado y el centro del disco.

1. Encuentre la función de distribución acumulada de  $X$  y su densidad de probabilidad.
2. Encuentre la densidad de  $X^2$ .
3. Encuentre la esperanza de  $X$ .

#### Resolución

Como el dardo fue arrojado al azar, toda región del disco  $A$  tiene probabilidad

$$P(A) = \frac{\text{área}(A)}{\text{área}(\text{Disco})} = \frac{\text{área}(A)}{\pi R^2}$$

Entonces  $(X \leq x)$  ocurre si se observa cualquier punto del círculo con radio  $x$ , por lo cual

$$P(X \leq x) = \frac{\pi x^2}{\pi R^2} = \frac{x^2}{R^2} \quad 0 \leq x \leq R$$

Si  $x < 0$  entonces  $P(X \leq x) = 0$ , pues el evento es absurdo, y si  $x > R$ , se puede observar cualquier punto del blanco, por lo cual  $P(X \leq x) = 1$ . Por lo cual

$$F(x) = \begin{cases} 0 & x < 0 \\ x^2/R^2 & 0 \leq x \leq R \\ 1 & x > R \end{cases}$$

La figura 5.4 muestra la gráfica de  $F$  y su densidad integrable.

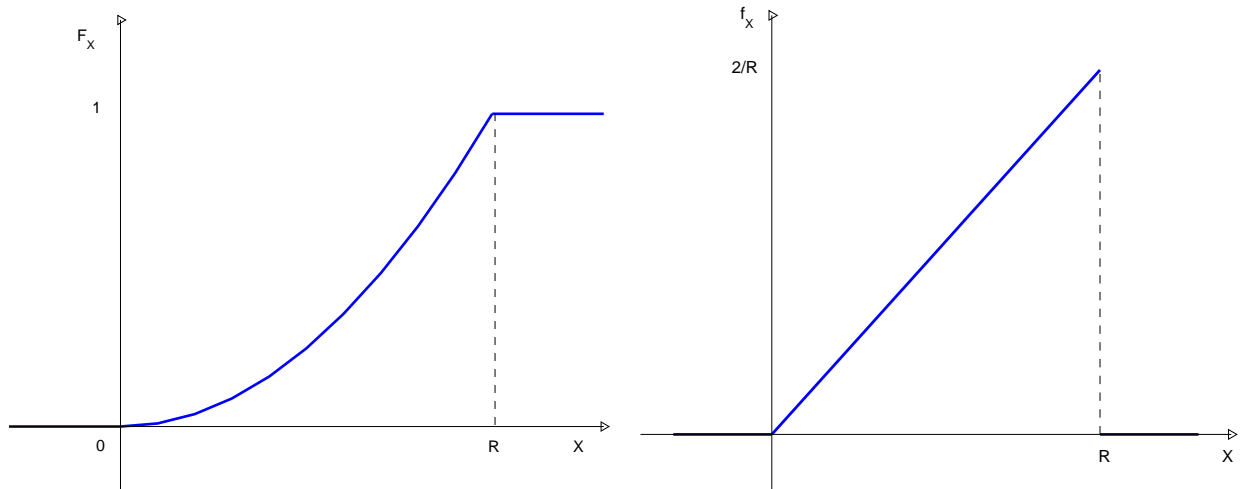


Figura 5.4: Panel izquierdo: Gráfico de la función de distribución de la variable  $X$ , distancia al centro del disco. Panel derecho, Gráfico de la función densidad de  $X$

La densidad de  $X$  se puede calcular diferenciando  $F$  en todo punto salvo en  $x = 0$  y  $x = R$ .

$$F'(x) = \begin{cases} 0 & x < 0 \\ 2x/R^2 & 0 < x < R \\ 1 & x > R \end{cases}$$

En  $x = 0$ , el límite de  $F'(x)$  por la derecha existe y es cero, al igual que el límite por la izquierda, por lo cual  $F'(0) = 0$ . Pero en  $x = R$ , la función  $F$  no es derivable, como podemos notar en la figura 5.4. Sin embargo, la función

$$f(x) = \begin{cases} F'(x) & x \neq R \\ 0 & x = R \end{cases} = \begin{cases} 2x/R^2 & 0 < x < R \\ 0 & \text{en otro caso} \end{cases}$$

es una densidad de probabilidad para  $X$  pues

$$F(x) = \int_{-\infty}^x f(y) dy$$

¿Como encontramos la densidad de  $Y = X^2$ ? Observemos primero que si  $F_Y$  es la función de distribución de  $X^2$ , entonces  $F_Y(y) = 0$  para todo  $y \leq 0$ . Para  $y > 0$ ,

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

Derivando obtenemos

$$F'_Y(y) = \frac{1}{2\sqrt{y}} (F'_X(\sqrt{y}) + F'_X(-\sqrt{y})) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) = \begin{cases} 0 & \sqrt{y} < 0 \\ \frac{(\sqrt{y})^2}{R^2} = \frac{y}{R^2} & 0 < \sqrt{y} < R \\ 1 & \sqrt{y} > R \end{cases} = \begin{cases} 0 & y < 0 \\ \frac{y}{R^2} & 0 < y < R^2 \\ 1 & y > R^2 \end{cases}$$

por lo cual  $Y = X^2$  tiene densidad

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) & y > 0 \\ 0 & y \leq 0 \end{cases} = \begin{cases} \frac{1}{2\sqrt{y}} \frac{2\sqrt{y}}{R^2} = \frac{1}{R^2} & 0 < \sqrt{y} < R \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{1}{R^2} & 0 < y < R^2 \\ 0 & \text{en otro caso} \end{cases}$$

Hemos graficado ambas funciones en la figura 5.5

La esperanza de  $X$  es

$$\int_0^R x f_X(x) dx = \int_0^R x \frac{2x}{R^2} dx = \left( \frac{2x^3}{3R^2} \right)_0^R = \frac{2R}{3}$$

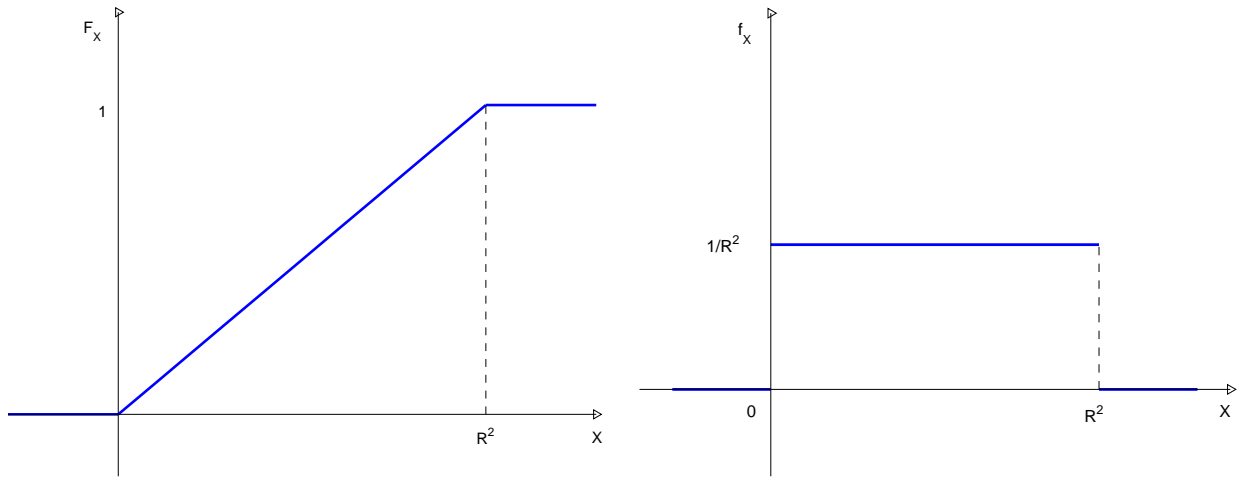


Figura 5.5: Panel izquierdo: Gráfico de la función de distribución de la variable  $X^2$ . Panel derecho, Gráfico de la función densidad de  $X^2$

**Ejemplo 5.2.3** Sea  $g(x) = x(1 - x)$ ,  $0 \leq x \leq 1$ , con  $g(x) = 0$  en otro caso. Encuentre una función densidad  $f_X$  que coincida con  $g$  salvo una constante. Encuentre la esperanza de  $X$  con densidad  $f_X$ .

**Resolución**

La función densidad debe no negativa y tener integral igual a 1, por lo cual una candidata es

$$f(x) = \frac{1}{c}g(x) \quad c = \int_{-\infty}^{\infty} g(x)dx$$

Entonces

$$c = \int_{-\infty}^{\infty} g(x)dx = \int_0^1 x(1-x)dx = \left( \frac{x^2}{2} - \frac{x^3}{3} \right)_0^1 = \frac{1}{6}$$

y la densidad  $f_X$  resulta

$$f_X(x) = \begin{cases} 6x(1-x) & 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

La esperanza de  $X$  es

$$\int_0^1 x \cdot 6x(1-x)dx = 6 \left[ \int_0^1 x^2 dx - \int_0^1 x^3 dx \right] = 6 \left( \frac{x^3}{3} \right)_0^1 + 6 \left( \frac{x^4}{4} \right)_0^1 = \frac{1}{2}$$

**Ejemplo 5.2.4** Sea  $X$  una variable aleatoria continua con densidad  $f$  y sean  $a$  y  $b \neq 0$  números reales, entonces la variable  $Y = a + bX$  tiene densidad

$$f_Y(y) = \frac{1}{|b|} f\left(\frac{y-a}{b}\right)$$

y  $f_Y(y) = 0$  en otro lado.

**Resolución**

Supongamos que  $b > 0$ , entonces  $g(y) = \frac{y-a}{b}$  es una función creciente y

$$P(Y \leq y) = P(X \leq \frac{y-a}{b}) = F_X\left(\frac{y-a}{b}\right)$$

por lo cual

$$f_Y(y) = \frac{1}{b} f\left(\frac{y-a}{b}\right)$$

Si  $b < 0$ , entonces  $g(y) = \frac{y-a}{b}$  es una función decreciente y

$$P(Y \leq y) = P(X \geq \frac{y-a}{b}) = 1 - F_X\left(\frac{y-a}{b}\right)$$

por lo cual

$$f_Y(y) = -\frac{1}{b} f\left(\frac{y-a}{b}\right)$$

Resumiendo, si  $b \neq 0$

$$f_Y(y) = \frac{1}{|b|} f\left(\frac{y-a}{b}\right)$$

## 5.3. Distribuciones continuas clásicas

### 5.3.1. Distribución Uniforme

Sean  $a$  y  $b$  constantes con  $a < b$ . La densidad uniforme sobre el intervalo  $(a, b)$  es la densidad  $f$  definida por

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{en otro caso} \end{cases}$$

La distribución correspondiente a  $f$  está dada por

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

La esperanza y varianza de  $X$  con distribución uniforme es

$$E(X) = \frac{a+b}{2} \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

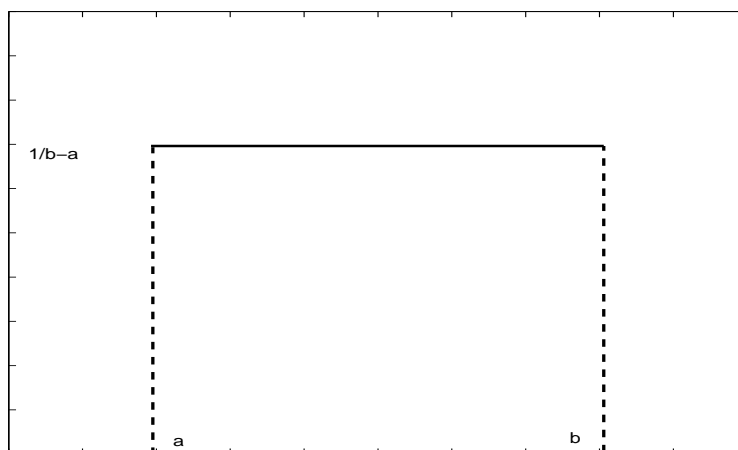


Figura 5.6: Distribución uniforme en el intervalo  $[a, b]$ .

**Ejemplo 5.3.1** Supongamos que Mario usualmente va a visitar a María entre las 8 y 8:30 de la noche. Si el horario de llegada es aleatorio, ¿cuál es la probabilidad de que Mario llegue después de las 8:25?

### Resolución

En este caso, queremos calcular la probabilidad de que  $X$ , la variable que modela el tiempo de llegada de Mario, tome valores entre las 8:25 y las 8:30 hs. Si  $X$  tiene distribución uniforme entre las 8 y las 8:30 hs, su densidad puede escribirse como

$$f(x) = \begin{cases} \frac{1}{30} & 0 < x < 30 \\ 0 & \text{en otro caso} \end{cases}$$

donde  $x$  son los minutos después de las 8 en los que Mario llega. Por lo tanto

$$P(25 \leq X \leq 30) = \int_{25}^{30} \frac{1}{30} dx = \frac{30-25}{30} = \frac{5}{30} = \frac{1}{6} = 0,166$$

**Ejemplo 5.3.2** Sea  $X$  con distribución uniforme en el  $(0, 1)$ . Encuentre la densidad de  $Y = -\frac{1}{\lambda} \log(1 - X)$  para todo  $\lambda > 0$ .

### Resolución

Observemos primero que  $Y$  es una variable positiva y por lo tanto  $F_Y(y) = 0$  para todo  $y \leq 0$ . Para  $y > 0$  se tiene

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P\left(-\frac{1}{\lambda} \log(1 - X) \leq y\right) \\ &= P(\log(1 - X) \geq -\lambda y) \\ &= P(1 - X \geq e^{-\lambda y}) \\ &= 1 - e^{-\lambda y} \end{aligned}$$

por lo cual  $f_Y(y) = F'_Y(y) = \lambda e^{-\lambda y}$  para  $y > 0$ . La densidad de  $Y$  está dada por

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

**Proposición 5.3.1** Sea  $X$  una variable aleatoria continua con función densidad  $f_X$  y función de distribución acumulada  $F_X$ , tal que  $F_X$  es estrictamente creciente en un intervalo  $I$  (que puede ser no acotado), nula a la izquierda de  $I$  y vale 1 a la derecha de  $I$ . Entonces  $F^{-1}(x)$  está bien definido para cada valor  $x \in I$ .

1. Sea  $Z = F(X)$ , entonces  $Z$  tiene distribución uniforme en  $[0, 1]$ .
2. Si  $U$  tiene distribución uniforme en el  $[0, 1]$ , entonces  $Y = F^{-1}(U)$  tiene como distribución acumulada a la función  $F$ .

### Demostración:

Observemos que

$$P(Z \leq z) = P(F(X) \leq z) = P(X \leq F^{-1}(z)) = F(F^{-1}(z)) = z$$

y como ésta es la función de distribución acumulada de una variable uniforme, sale el resultado. Observemos ahora que

$$P(Y \leq y) = P(F^{-1}(U) \leq y) = P(U \leq F(y)) = F(y)$$

por lo cual  $X$  e  $Y$  tienen la misma función de distribución acumulada.

**Observación 5.3.1** Dos variables que tienen la misma función de distribución, no son necesariamente la misma cosa. El tiempo de duración de las llamadas telefónicas que entran a un call center puede tener la misma distribución que el tiempo de espera de un consumidor en la cola de un cajero de banco, pero obviamente ambas variables no modelan el mismo fenómeno. Los problemas modelados son completamente diferentes, pero aun así coinciden en la distribución probabilística del tiempo. La probabilidad de durar más de 15 minutos va a ser la misma en ambos casos, pero las implicaciones de dicha probabilidad va a ser diferente. Para estudiar la matemática del modelo, se considera solamente la distribución, sin importar de que problema surge. Para aplicar dicho modelo, es importante saber como es el problema al cual se aplica, y las consecuencias que tiene dicha aplicación. Por ejemplo, si el tiempo de espera tiende a ser muy largo, se puede incorporar otro cajero para atender a los clientes, en cambio si la central no puede atender todas las llamadas salientes, simplemente se pide que se llame en otro momento. Los eventos tienen la misma probabilidad, la decisión es diferente.

### 5.3.2. Distribución Exponencial

**Definición 5.3.1** Sea  $\lambda$  un valor real positivo. La densidad exponencial es la densidad  $f$  definida por

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

La esperanza y varianza de  $X$  con distribución exponencial de parámetro  $\lambda$  es

$$E(X) = \frac{1}{\lambda} \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

Como vimos en el ejemplo anterior, es la densidad de la variable  $Y = -\frac{1}{\lambda} \log(1 - X)$ , cuando  $X$  es uniforme en  $(0, 1)$ .

La función de distribución acumulada correspondiente a  $f$  es

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

**Ejemplo 5.3.3** Sea  $X$  la duración en horas de las lámparas almacenadas en un depósito, y supongamos que pueden modelarse con la distribución exponencial con parámetro 0,01.

1. Hallar la función de distribución acumulada de  $X$ , la duración de una lámpara elegida al azar del lote.
2. ¿Cuál es la probabilidad de que una lámpara dure más de 120 horas?  
item ¿Cuántas horas se espera que dure una lámpara?
3. Si se tiene un lote de 10000 lámparas ¿Cuál es la probabilidad de que 10 lámparas elegidas al azar del lote duren menos de 70 horas?

#### Resolución

La distribución acumulada de  $X$  es la distribución de una exponencial, y se calcula aplicando directamente la definición:

$$P(X \leq x) = \int_0^x \lambda \exp(-\lambda t) dt = \lambda \frac{\exp(-\lambda t)}{-\lambda} \Big|_0^x = 1 - \exp(-\lambda x) \quad x > 0$$

En el caso de  $\lambda = 0,01$  resulta

$$P(X \leq x) = 1 - \exp(-0,01x)$$

Usando esta fórmula podemos calcular la probabilidad de que una lámpara elegida al azar del depósito dure más de 120 hs.

$$P(X > 120) = 1 - P(X \leq 120) = 1 - (1 - \exp(-0,01120)) = \exp(-0,01120) = \exp(-1,2) = 0,3011$$

La duración esperada de las lámparas, es la esperanza de la variable aleatoria, y como tiene distribución exponencial, sabemos que es  $1/\lambda$ . Entonces se espera que dure 100 horas.

Si ahora sabemos que el lote es realmente grande, comparado con la muestra es decir, con el conjunto de lámparas extraídas para testear, podemos suponer que estamos eligiendo las lámparas al azar con reposición, (pues la probabilidad de elegir dos veces la misma lámpara es casi nula). Entonces podemos suponer que tenemos un experimento binomial, donde en cada extracción tenemos éxito si la lámpara dura menos de 70 horas. La variable  $Y$ , número de lámparas que duran menos de 70 horas, es una variable binomial de parámetros  $n = 10$  y  $p = P(X < 70)$ . Calculemos esta probabilidad,

$$P(X < 70) = 1 - \exp(-0,0170) = 1 - \exp(-0,7) = 0,5034$$

Entonces, la probabilidad de que las 10 lámparas elegidas al azar del lote duren menos de 70 horas es

$$P(Y = 10) = \binom{10}{10} 0,5034^{10} (1 - 0,5034)^{10-10} = 0,5034^{10} = 0,00104$$

**Ejemplo 5.3.4** Sea  $X$  una variable que tiene densidad exponencial con parámetro  $\lambda$ . Encuentre la densidad de  $Y = X^{1/\beta}$ , con  $\beta \neq 0$ .

**Resolución**

Si  $X$  tiene distribución exponencial de parámetro  $\lambda$  entonces su densidad es  $f_X(x) = \lambda e^{-\lambda x}$ , si  $x > 0$ , y cero si no. Supongamos  $\beta > 0$ , entonces  $g(x) = x^\beta$  es una función creciente y si  $y > 0$

$$P(Y \leq y) = P(X^{1/\beta} \leq y) = P(X \leq y^\beta) = F_X(y^\beta) = 1 - e^{-\lambda y^\beta}$$

entonces

$$f_Y(y) = F'_X(y) = \beta \lambda y^{\beta-1} e^{-\lambda y^\beta}$$

Ahora, si  $\beta < 0$ , entonces  $g(x) = x^\beta$  es una función decreciente y si  $y > 0$

$$P(Y \leq y) = P(X^{1/\beta} \leq y) = P(X \geq y^\beta) = 1 - F_X(y^\beta) = e^{-\lambda y^\beta}$$

entonces

$$f_Y(y) = F'_X(y) = -\beta \lambda y^{\beta-1} e^{-\lambda y^\beta}$$

por lo cual

$$f_Y(y) = \begin{cases} |\beta| \lambda y^{\beta-1} e^{-\lambda y^\beta} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

**Ejemplo 5.3.5** Supongamos que un conjunto de luces navideñas tiene 1000 luces que se encienden al mismo tiempo, cuya duración puede ser considerada variables aleatorias independientes con distribución exponencial de parámetro 0.01. Encuentre la probabilidad de que exactamente 10 de ellas duren más de 10 horas.

**Resolución**

Sea  $X$  la variable aleatoria que mide la duración de una lámpara. En este caso, tengo 1000 experimentos independientes, donde éxito corresponde al evento “la lámpara duró más de 10 horas”. Este evento tiene probabilidad

$$p = P(X > 10) = 1 - P(X \leq 10) = 1 - (1 - \exp(-0,0110)) = \exp(-0,1) = 0,904$$



Por lo cual tengo una variable binomial  $Y$  que mide el número de lámparas que duran más de 10 horas, que tiene tamaño 1000 y probabilidad de éxito 0.9. No es simple realizar el cálculo de la probabilidad de que  $Y$  sea exactamente igual a 10, por lo cual es razonable usar la aproximación Poisson a la binomial, reemplazando  $Y$  por  $Z$  con distribución Poisson de parámetro  $np = 904$ .

$$P(Z = 10) = \frac{\exp(-np)(np)^{10}}{10!} = \frac{\exp(-904)(904)^{10}}{10!}$$

Este cálculo tampoco es simple, pero como quiero calcular una probabilidad discreta exacta, debo usar un programa de computación para calcularlo.

### 5.3.3. Distribución Normal

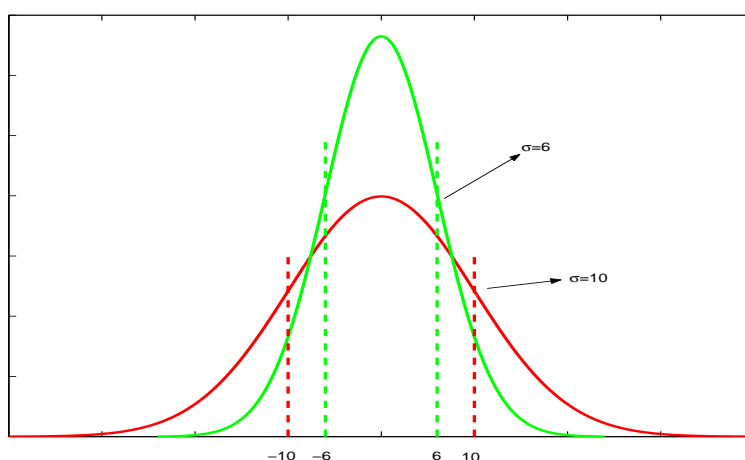


Figura 5.7: Dos densidades normales con igual media y distinta varianza. Ambos parámetros pueden localizarse fácilmente en las figuras.

La curva normal fue descubierta alrededor del 1720 por Abraham de Moivre, mientras desarrollaba sus estudios sobre la teoría de probabilidad. Aparece por primera vez en la segunda edición de *The Doctrine of Chances* como una aproximación a la distribución binomial. Más tarde la curva fue usada como un histograma ideal, contra la cual histogramas de otros datos pudieran ser comparados. El teorema central del límite (en sus diferentes versiones) justifica el uso de la distribución normal como aproximación en una gran variedad de casos.

**Definición 5.3.2** La densidad normal standard  $\phi$  está definida como

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad -\infty < x < \infty$$

en donde aparecen tres de los números más famosos de la historia de la matemática,  $\sqrt{2}$ ,  $\pi$  y  $e$ . Para completar su apariencia intimidante, la distribución acumulada correspondiente a  $\phi$ , denotada por  $\Phi$  no tiene fórmula cerrada,

$$\Phi(x) = \int_{-\infty}^x \phi(y) dy = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

por lo cual no es posible calcular probabilidades integrando la curva. Sin embargo, existen aproximaciones tabuladas de diferentes maneras.

Esta curva tiene muchas cualidades importantes. Primero, es infinitamente diferenciable, simétrica alrededor de cero, donde tiene su valor máximo, sus límites cuando  $x$  crece en forma positiva y negativa son iguales a 0 (la curva se pega al eje de las abscisas) y su cambio de concavidad a convexidad (puntos de inflexión) se encuentran en 1 y -1.

La distribución normal aparece en muchos modelos estadísticos clásicos, y por definición, es el área bajo la curva desde el extremo infinito negativo hasta el valor  $x$ ,

$$\int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

y este valor se obtiene de tablas como la de la figura 5.9. Existen también numerosos programas de computación que calculan estos valores, desde hojas de cálculo del tipo de Excel [?] hasta softwares estadísticos avanzados programables de uso libre, como R [?]. El software InfoStat fue diseñado por docentes de la Universidad Nacional de Córdoba para el uso en aulas y para la aplicación en consultoría [?]. Los gráficos de este libro fueron realizados en Matlab 7.1 [?], usando el Statistics Toolbox. La mayoría de los libros

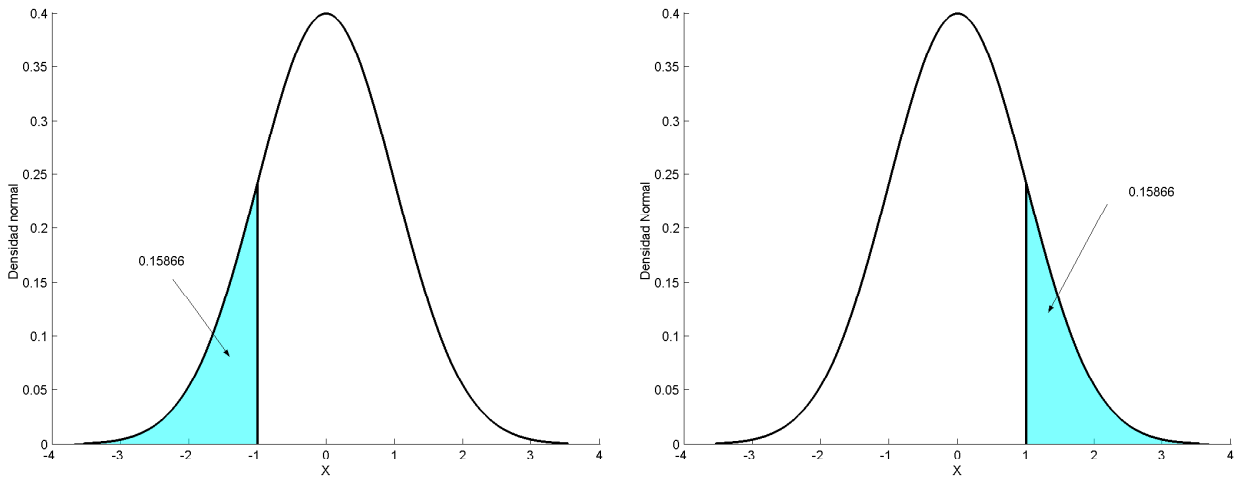


Figura 5.8: Panel izquierdo: El área sombreada es la cola a izquierda,  $\Phi(-1) = P(Z \leq -1)$ . Panel derecho, el área sombreada es la cola a derecha,  $1 - \Phi(1) = P(Z \geq 1)$ . Ambas áreas son iguales, dado que  $\Phi(-x) = 1 - \Phi(x)$

de estadística tienen en un apéndice una tabla de probabilidades generadas por una distribución normal patrón o standard. Para poder leer estas tablas es necesario tener en cuenta las siguientes propiedades de la distribución normal. Si  $Z$  tiene distribución normal standard y  $z > 0$ ,

1.  $P(Z \geq z) = 1 - P(Z \leq z) = 1 - \Phi(z)$
2.  $P(Z \leq -z) = P(Z \geq z) = 1 - P(Z \leq z) = 1 - \Phi(z)$
3.  $P(-z \leq Z \leq z) = P(Z \leq z) - P(Z \leq -z) = P(Z \leq z) - [1 - P(Z \leq z)] = 1 - 2P(Z \leq z) = 1 - 2\Phi(z)$ .
4.  $P(z_1 \leq Z \leq z_2) = P(Z \leq z_1) - P(Z \leq z_2)$ , para cualquier  $z_1, z_2$  en  $R$ .

Estas propiedades indican que con solo conocer la “cola a izquierda” de la distribución para valores  $z$  mayores que cero, puedo obtener todos los restantes valores, usando propiedades de las integrales y de la densidad normal en particular, como la simetría. En la figura 5.8 podemos ver la cola a izquierda y la cola a derecha de la normal patrón, calculadas en los puntos  $z$  y  $-z$ . Los valores de las áreas bajo la curva son iguales.

**Definición 5.3.3** Sea  $Y$  una variable aleatoria con densidad dada por

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2} = \frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \quad -\infty < y < \infty$$

La distribución de  $Y$  se llama normal con parámetros  $\mu$  y  $\sigma^2$  y se denota  $Y \sim N(\mu, \sigma^2)$ .

La curva normal patrón no es la única curva llamada normal, es simplemente la que se encuentra tabulada. Las distribuciones normales forman una familia, indexada por parámetros que denotan el centro de simetría de la curva  $\mu$  y la dispersión de la curva  $\sigma$ . Las normales toman todos los valores entre infinito positivo e infinito negativo, por lo cual la dispersión debería ser la misma, si una piensa en el “rango” de valores observados, pero en realidad nos referimos a la concentración alrededor de la media. Si  $\sigma$  es chiquito, la distribución hace un pico alto alrededor de la media, y si  $\sigma$  es grande, la curva es chata y aplastada. La siguiente proposición nos dice como calcular probabilidades de cualquier distribución normal, usando solo la tabla de la normal patrón.

**Proposición 5.3.2** Si  $Y$  tiene distribución normal con parámetros  $\mu$  y  $\sigma^2$  entonces  $Z = \frac{Y-\mu}{\sigma}$  tiene distribución normal standard. También, si  $Z$  es una variable normal standard  $Y = \mu + \sigma Z$  tiene distribución normal con parámetros  $\mu$  y  $\sigma^2$ .

**Demostración:**

Ya vimos anteriormente que si  $V = a + bW$  entonces

$$f_V(y) = \frac{1}{|b|} f_W\left(\frac{y-a}{b}\right)$$

Aplicando este resultado a  $Z = \frac{Y-\mu}{\sigma}$  resulta

$$f_Z(y) = \frac{1}{\frac{1}{\sigma}} f_Y\left(\frac{y+\frac{\mu}{\sigma}}{\frac{1}{\sigma}}\right) = \sigma \frac{1}{\sigma} \phi\left(\frac{\frac{y+\frac{\mu}{\sigma}}{\frac{1}{\sigma}} - \mu}{\sigma}\right) = \phi(y)$$

por lo cual  $Z$  es una variable con distribución normal standard.

Aplicando el mismo resultado a  $Y = \mu + \sigma Z$  resulta

$$f_Y(y) = \frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right)$$

por lo cual  $Y$  es una variable con distribución normal con parámetros  $\mu$  y  $\sigma^2$ .

**Observación 5.3.2** Esta proposición es muy importante, pues solo tenemos tabuladas normales patrón, para poder calcular probabilidades de normales generales, debemos transformar la variable en un patrón y luego usar la tabla.

**Proposición 5.3.3** Si  $Z$  es una variable con distribución normal standard entonces  $E(Z) = 0$  y  $\text{Var}(Z) = 1$ . Si  $Y$  es una variable con distribución normal con parámetros  $\mu$  y  $\sigma^2$ , entonces  $E(Y) = \mu$  y  $\text{Var}(Y) = \sigma^2$ . También, la densidad de  $Y$  está centrada en  $\mu$ , es decir, es simétrica con respecto a  $\mu$  y sus puntos de inflexión están localizados en  $\sigma$  y  $-\sigma$ .

En la figura 5.7 podemos ver dos densidades normales con igual media y distinta varianza. Se puede observar como las curvas están centradas en  $\mu$  y se puede localizar la desviación standard  $\sigma$  de cada curva localizando el punto en donde la curva cambia de concavidad a convexidad. No hay muchas distribuciones en donde la media y la varianza sea tan fácil de localizar. Usualmente son valores relacionados con la forma analítica de la curva. En algunas familias, están relacionados con los parámetros de posición y escala que gobiernan la forma de la curva.

### 5.3.4. Uso de la tabla

Hay muchas tablas de valores de probabilidades normales. Las mas usuales son las tablas de cola a derecha y cola a izquierda. La tabla de cola a derecha tabula los valores de

$$\beta = 1 - \Phi(z_\beta) = P(Z \geq z_\beta)$$

con  $z_\beta > 0$  y  $Z \sim N(0, 1)$ . La tabla de acumulación a cola a izquierda tabula los valores de

$$\alpha = \Phi(z_\alpha) = P(Z \leq z_\alpha)$$

$z_\alpha > 0$  y  $Z \sim N(0, 1)$ . Por lo cual, para un  $z > 0$  dado, el valor de  $\beta = 1 - \alpha$ . Por eso es importante estudiar cual es el tipo de tabla que se va a usar.

Supongamos tener una tabla acumulada de cola a izquierda, como la de la figure 5.9. En el centro de la tabla están las columnas con probabilidades  $\alpha$  y en los márgenes de la tabla los valores  $z_\alpha > 0$ . El margen izquierdo tiene los valores entero y primer decimal de  $z_\alpha$  y el margen superior el segundo decimal de  $z_\alpha$ . Para encontrar

$$\alpha = P(Z < z_\alpha) = P(Z < 2,44)$$

se busca la fila del  $z_\alpha = 2,4$  y luego se sigue esa fila hasta la columna del segundo decimal 0.04. La intersección de fila y columna da el valor  $\alpha = 0,9927$ .

**Ejemplo 5.3.6** Supongamos que el diámetro de un cable está distribuido normalmente con media 0,8 y varianza 0,0004

1. ¿Cuál es la probabilidad de que el diámetro sobrepase las 0.81 pulgadas?
2. Supongamos que el cable se considera defectuoso si el diámetro dista del diámetro medio esperado en más de 0.025
  - I) ¿Cuál es la probabilidad de obtener un cable defectuoso?
  - II) ¿Cuál es la probabilidad de que en un lote de 10 cables se encuentren más de 3 defectuosos?.

#### Resolución

Sea  $Y$  la variable que representa el diámetro del cable. Entonces la probabilidad de que el diámetro sobrepase las 0.81 pulgadas es

$$P(Y \geq 0,81) = P\left(\frac{Y - \mu}{\sigma} \geq \frac{0,81 - \mu}{\sigma}\right) = P\left(Z \geq \frac{0,81 - 0,8}{0,02}\right) = P(Z \geq 0,5) = 1 - P(Z \leq 0,500) = 1 - 0,6915 = 0,3085$$

El diámetro de un cable dista de su diámetro medio en más de 0.025 si  $|Y - \mu| \geq 0,025$ , entonces

$$\begin{aligned} P(|Y - \mu| \geq 0,025) &= P\left(\left|\frac{Y - \mu}{\sigma}\right| \geq \frac{0,025}{\sigma}\right) \\ &= P\left(|Z| \geq \frac{0,025}{\sigma}\right) \\ &= P\left(Z \geq \frac{0,025}{\sigma}\right) + P\left(-Z \geq \frac{0,025}{\sigma}\right) \\ &= P\left(Z \geq \frac{0,025}{\sigma}\right) + P\left(Z \leq -\frac{0,025}{\sigma}\right) \\ &= 2P\left(Z \leq -\frac{0,025}{0,02}\right) \\ &= 2\left(1 - P\left(Z \leq \frac{0,025}{0,02}\right)\right) \\ &= 2 \times 0,1056 = 0,2112 \end{aligned}$$

Ahora queremos saber cuál es la probabilidad de hallar más de tres cables defectuosos de un lote de 10 cables elegidos al azar. Si consideramos la población de cables muy grande, entonces podemos suponer que la diferencia en elegir al azar con reposición y sin reposición es despreciable, por lo cual la variable  $X$  que mira el número de cables defectuosos en la muestra es una variable binomial con parámetros  $n = 10$  y  $p = P(\text{cable defectuosos}) = P(|Y - \mu| \geq 0,025) = 0,2112$ . Por lo cual

$$P(X > 3) = 1 - P(X \leq 2) = 1 - (1 - 0,2112)^{10} - 10 \times 0,2112 \times (1 - 0,2112)^9 - 45 \times 0,2112^2 \times (1 - 0,2112)^8$$

**Ejemplo 5.3.7** La altura de cierta población se distribuye normalmente con promedio 170 cm y desvío estándar 5 cm.

1. Si elegimos una persona al azar de esa población, ¿qué altura se espera que tenga?
2. Hallar  $a$  tal que  $P(Y > a) = 0,4$ .

#### Resolución

Sea  $Y$  la altura de una persona elegida al azar de la población, entonces la altura esperada de la persona es la esperanza de la variable  $Y$ , 1.7 metros.

Si quiero calcular  $a$  tal que  $P(Y > a) = 0,4$ ,

$$0,4 = P(Y > a) = P\left(\frac{Y - \mu}{\sigma} > \frac{a - \mu}{\sigma}\right) = P\left(Z > \frac{a - 170}{5}\right) = 1 - P\left(Z \leq \frac{a - 170}{5}\right)$$

Por lo cual,

$$P\left(Z \leq \frac{a - 170}{5}\right) = 0,6$$

La tabla dice que  $z_{0,6} = \frac{a - 170}{5}$  se encuentra entre 0.25 y 0.26, podemos asignarle el punto medio entre ambos valores, 0.255, entonces

$$0,255 = \frac{a - 170}{5} \quad \text{por lo cual} \quad a = 0,255 \times 5 + 170 = 171,275$$

## 5.4. Distribución conjunta

Sean  $X_1, X_2, X_3$  variables aleatorias definidas sobre el mismo espacio de probabilidad, se define la función de distribución del vector  $(X_1, X_2, X_3)$  como

$$F(x_1, x_2, x_3) = P(X_1 \leq x_1, X_2 \leq x_2, X_3 \leq x_3)$$

Ya vimos esta definición cuando consideramos variables discretas y vimos su relación con las densidades discretas. Ahora consideraremos el caso de vectores continuos.

**Definición 5.4.1** Sean  $X_1, X_2, X_3$  variables aleatorias continuas definidas sobre el mismo espacio de probabilidad, diremos que tienen densidad integrable conjunta si existe una función no negativa  $f_X$  tal que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_X(x_1, x_2, x_3) dx_1 dx_2 dx_3 = 1$$

y para todo conjunto  $A$  "razonable"

$$P((X_1, X_2, X_3) \in A) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_A(x_1, x_2, x_3) f_X(x_1, x_2, x_3) dx_1 dx_2 dx_3$$

En particular

$$F(x_1, x_2, x_3) = P(X_1 \leq x_1, X_2 \leq x_2, X_3 \leq x_3) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \int_{-\infty}^{x_3} f_X(x_1, x_2, x_3) dx_1 dx_2 dx_3$$

**Proposición 5.4.1** Sean  $X_1, X_2, X_3$  son variables aleatorias continuas con densidad conjunta  $f_X$ , entonces cada variable  $X_i$  tiene densidad marginal conjunta  $f_{X_i}$  definida por

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_X(x_1, x_2, x_3) dx_2 dx_3 \\ f_{X_2}(x_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_X(x_1, x_2, x_3) dx_1 dx_3 \\ f_{X_3}(x_3) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_X(x_1, x_2, x_3) dx_1 dx_2 \end{aligned}$$

esto es, se integra sobre todas las variables salvo la  $i$  esima. En el caso del vector  $(X, Y)$

$$f_X(x) = \int_{-\infty}^{\infty} f_X(x, y) dy \quad f_Y(y) = \int_{-\infty}^{\infty} f_X(x, y) dx$$

Estas son las densidades marginales, cuando veamos independencia veremos que las marginales caracterizan la distribución conjunta. Si las variables no son independientes, entonces puede haber diferentes distribuciones conjuntas con las mismas distribuciones marginales.

**Ejemplo 5.4.1** Una estación de servicio realiza una carga de nafta en un tanque general al comienzo de cada semana. Sea  $Y_1$  la proporción de la capacidad del tanque que está llena después que se vierte la carga. A causa de la demanda general, no siempre se provee la misma cantidad, además el tanque rara vez se vacía con las ventas de la semana, por lo cual  $Y_1$  varía de semana en semana. Sea  $Y_2$  la proporción de la capacidad del tanque que se vende durante la semana. Es evidente que  $Y_2$  no puede sobrepasar a  $Y_1$  y que ambos valores varían entre 0 y 1 al ser proporciones. Supongamos que la relación entre ambas variables puede modelarse usando la densidad conjunta

$$f(y_1, y_2) = \begin{cases} 3y_1 & 0 \leq y_2 \leq y_1 \leq 1 \\ 0 & \text{en otro punto } (y_1, y_2) \end{cases}$$

Encuentre la probabilidad de que menos de la mitad del tanque sea llenado al principio de la semana y que más de un cuarto del tanque sea vendido durante esa semana.

**Resolución**

Queremos encontrar la probabilidad

$$P(0 \leq Y_1 \leq 0,5, Y_2 > 0,25)$$

Para cualquier vector aleatorio continuo, esta probabilidad se calcula como el volumen bajo la densidad conjunta del vector en la región de interés. Esta región es

$$A = \{(y_1, y_2), 0 \leq y_1 \leq 0,5, 0,25 < y_2 \leq y_1\}$$

por lo cual

$$\begin{aligned} P(0 \leq Y_1 \leq 0,5, Y_2 > 0,25) &= \int_{0,25}^{0,5} \int_{0,25}^{y_1} 3y_1 dy_2 dy_1 \\ &= \int_{0,25}^{0,5} \left[ \int_{0,25}^{y_1} dy_2 \right] 3y_1 dy_1 \\ &= \int_{0,25}^{0,5} [y_1 - 0,25] 3y_1 dy_1 \\ &= \int_{0,25}^{0,5} 3y_1^2 dy_1 - \int_{0,25}^{0,5} \frac{3}{4} y_1 dy_1 \\ &= [y_1^3]_{0,25}^{0,5} - [(3/8)y_1^2]_{0,25}^{0,5} \\ &= [1/8 - (3/8)(1/4)] - [(1/64) - (3/8)(1/16)] \\ &= 5/128 \end{aligned}$$

### 5.4.1. Independencia

**Definición 5.4.2** Dos variables aleatorias se dicen independientes si para todo par de eventos  $A$  y  $B$

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Es equivalente a decir que

$$F(x, y) = F_X(x)F_Y(y) \quad -\infty < x, y < \infty$$

Si las variables aleatorias son continuas basta con caracterizar la función densidad conjunta.

**Definición 5.4.3** Sean  $X_1, X_2, X_n$  las variables aleatorias continuas con densidad conjunta  $f_X$ . Diremos que son independientes si

$$f_X(x_1, x_2, x_3) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3)$$

**Ejemplo 5.4.2** Martín, Pedro y José llegan juntos al sector de atención al cliente de un banco que tiene solo dos personas atendiendo al público. Martín y Pedro comienzan a ser atendidos simultáneamente, y José será atendido en cuanto uno de los empleados esté libre. Si la duración de la consulta de cada cliente se modela como una variable aleatoria independiente con distribución exponencial de parámetro  $\lambda$ . ¿Cuál es la probabilidad de que José no sea el último en irse?

**Resolución**

Sean  $X, Y, Z$  la duración de la consulta de Martín, Pedro y José respectivamente. Como son variables independientes con distribución exponencial, la densidad conjunta es

$$f(x, y, z) = \lambda^3 e^{-\lambda(x+y+z)} \quad x > 0, y > 0, z > 0$$

Ahora, si José no es el último en irse, entonces alguno de los dos, Martín o Pedro, se retira para que José sea atendido, José termina y se retira y alguno de los dos, Martín o Pedro, todavía está realizando su consulta. Por lo cual la probabilidad que queremos calcular resulta

$$P(Z + \min(X, Y) \leq \max(X, Y))$$

Entonces, considerando que los eventos  $(X \leq Y)$  y  $(X > Y)$  particionan el espacio,

$$\begin{aligned} P(Z + \min(X, Y) \leq \max(X, Y)) &= P(Z + \min(X, Y) \leq \max(X, Y), X \leq Y) + P(Z + \min(X, Y) \leq \max(X, Y), X > Y) \\ &= P(Z + X \leq Y, X \leq Y) + P(Z + Y \leq X, X > Y) \end{aligned}$$

Ahora,  $(Z + X \leq Y) \cap (X \leq Y) = (Z + X \leq Y)$  y  $(Z + Y \leq X) \cap (X > Y) = (Z + Y \leq X)$ , y como  $X, Y$  y  $Z$  tienen igual distribución, resulta

$$P(Z + \min(X, Y) \leq \max(X, Y)) = P(Z + X \leq Y) + P(Z + Y \leq X) = 2P(Z + X \leq Y)$$

Para cualquier vector aleatorio continuo, esta probabilidad se calcula como el volumen bajo la densidad conjunta del vector en la región de interés. Esta región es

$$A = \{(x, y, z), 0 \leq x < \infty, 0 \leq z < \infty, x + z \leq y\}$$

por lo cual

$$\begin{aligned}
 P(Z + \min(X, Y) \leq \max(X, Y)) &= 2P(Z + X \leq Y) \\
 &= 2 \int_0^\infty \int_0^\infty \int_{x+z}^\infty f(x, y, z) dy dx dz \\
 &= 2 \int_0^\infty \int_0^\infty \int_{x+z}^\infty \lambda^3 e^{-\lambda(x+y+z)} dy dx dz \\
 &= 2 \int_0^\infty \int_0^\infty \lambda^2 e^{-\lambda(x+z)} \left[ \int_{x+z}^\infty \lambda e^{-\lambda(y)} dy \right] dx dz \\
 &= 2 \int_0^\infty \int_0^\infty \lambda^2 e^{-\lambda(x+z)} [1 - F_Y(x+z)] dx dz \\
 &= 2 \int_0^\infty \int_0^\infty \lambda^2 e^{-\lambda(x+z)} [e^{-\lambda(x+z)}] dx dz \\
 &= 2 \left[ \int_0^\infty \lambda e^{-2\lambda x} dx \right] \left[ \int_0^\infty \lambda e^{-2\lambda z} dz \right] \\
 &= (1/2) \left[ \int_0^\infty 2\lambda e^{-2\lambda x} dx \right] \left[ \int_0^\infty 2\lambda e^{-2\lambda z} dz \right] \\
 &= 1/2
 \end{aligned}$$

**Ejemplo 5.4.3** Sea  $X$  la vida útil (en horas) de un componente electrónico, que puede ser modelada como una variable que tiene densidad exponencial con parámetro  $\lambda = 1/100$ .

a) Calcule la probabilidad de que un componente electrónico dure más de 200 horas.

b) Si tres componentes operan independientemente en un sistema electrónico que falla solo si al menos dos de los componentes falla, encuentre la probabilidad de que el equipo opere al menos 200 horas sin fallas.

**Resolución**

Calculemos primero la probabilidad de que un componente electrónico dure más de 200 horas, si la vida útil tiene distribución exponencial con  $\lambda = 1/100$ .

$$p = P(X > 200) = 1 - P(X \leq 200) = 1 - F_X(200) = 1 - [1 - e^{-200/100}] = e^{-200/100} = 0,1353$$

Sea  $Y$  la variable que mide el número de componentes que operan más de 200 horas. Como la vida de los componentes es independiente una de otra,  $Y$  es una variable binomial con parámetros  $n = 3$  y  $p = 0,1353$ . Entonces

$$P(Y \geq 2) = P(Y = 2) + P(Y = 3) = 3p^2(1 - p) + p^3 = 0,05$$

**Proposición 5.4.2** Si  $X$  e  $Y$  son variables aleatorias continuas entonces  $X + Y$

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(x, z-x) dx \quad -\infty < z < \infty$$

Si  $X$  e  $Y$  son variables independientes entonces  $X + Y$  tiene densidad

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \quad -\infty < z < \infty$$

Si  $X$  e  $Y$  son independientes y no negativas entonces  $X + Y$  tiene densidad

$$f_{X+Y}(z) = \int_0^z f_X(x) f_Y(z-x) dx \quad -\infty < z < \infty$$

**Ejemplo 5.4.4** La vida útil de un cierto tipo de fusibles puede suponerse con distribución exponencial con  $\lambda = 1/3$ , donde las mediciones son en cientos de horas. Si un sistema tiene dos fusibles con vida útil independiente ensamblados de tal forma que uno se activa cuando el otro falla, por lo cual el sistema falla cuando ambos lo hacen, encuentra la probabilidad de que el sistema dure menos de 1 (ciento de hora).



**Resolución**

Sea  $V$  la vida útil del sistema,  $X$  e  $Y$  la vida útil de cada fusible, entonces  $V = X + Y$ , la suma de dos variables continuas no negativas independientes con distribución exponencial. Por lo tanto

$$f_V(z) = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx = \lambda^2 e^{-\lambda z} \int_0^z dx = \lambda^2 z e^{-\lambda z} \quad 0 < z < \infty$$

y

$$\begin{aligned} P(V < 1) &= \int_0^1 \lambda^2 z e^{-\lambda z} dz \\ &= \lambda^2 \left[ \left[ z \frac{1}{-\lambda} e^{-\lambda z} \right]_0^1 - \frac{1}{-\lambda} \int_0^1 e^{-\lambda z} dz \right] \\ &= \lambda^2 \left[ -\frac{1}{\lambda} e^{-\lambda} - \frac{1}{\lambda^2} e^{-\lambda} + \frac{1}{\lambda^2} \right] \\ &= 1 - e^{-\lambda}(\lambda + 1) \\ &= 1 - \frac{4}{3} e^{-\frac{1}{3}} \end{aligned}$$

**5.4.2. Propiedades de la esperanza**

**Teorema 5.4.1** Sea  $X_1, X_2, X_3$  variables aleatorias continuas con densidad conjunta  $f_{X_1, X_2, X_3}$  y sea  $Z$  una variable definida como  $Z = g(X_1, X_2, X_3)$ . Entonces  $Z$  tiene esperanza finita si y solamente si

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x_1, x_2, x_3)| f_{X_1, X_2, X_3}(x_1, x_2, x_3) dx_1 dx_2 dx_3 < \infty$$

y en ese caso

$$E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2, x_3) f_{X_1, X_2, X_3}(x_1, x_2, x_3) dx_1 dx_2 dx_3$$

Este teorema puede extenderse a muchas más variables que tres, vamos a usar esa versión extendida en la sección siguiente, cuando hablemos de muestra aleatoria.

**Ejemplo 5.4.5** Sea  $(X, Y)$  un vector aleatorio con densidad conjunta

$$f_{XY}(x, y) = e^{-y} \quad 0 \leq x \leq y$$

Hallar la esperanza de  $Z = XY$ .

**Resolución**

Aplicando el teorema anterior

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x, y) dx dy = \int_0^{\infty} y e^{-y} \left[ \int_0^y x dx \right] dy = \int_0^{\infty} y e^{-y} \left[ \frac{y^2}{2} \right] dy \\ &= \int_0^{\infty} \frac{y^3}{2} e^{-y} dy = \frac{\Gamma(4)}{2} \int_0^{\infty} \frac{1^4}{\Gamma(4)} y^{4-1} e^{-y} dy = \frac{\Gamma(4)}{2} = \frac{(4-1)!}{2} = 6/2 = 3 \end{aligned}$$

usando el hecho de que

$$1 = \int_0^{\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} dy$$

para todo  $\alpha$  y  $\lambda$  positivos.

**Corolario 5.4.1** Sean  $X_1, X_2, X_3$  variables aleatorias con densidad continua  $f_{X_i}$  y  $c$  una constante. Entonces:

1. Si  $P(X_1 = c) = 1$ , entonces  $E(X_1) = c$ .
2.  $E(cX_1) = cE(X_1)$ .
3.  $E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3)$ .
4. Si  $X_1, X_2, X_3$  son independientes, entonces

$$E(X_1 X_2 X_3) = E(X_1)E(X_2)E(X_3)$$

*Este corolario también vale para más de tres variables. Vamos a usar la versión extendida cuando hablemos de muestra en los próximos capítulos.*

**Ejemplo 5.4.6** El tiempo  $X$  que un técnico requiere para realizar una operación de mantenimiento de un aparato de aire acondicionado puede considerarse con distribución exponencial de parámetro  $\lambda = 1$  hora. Si un edificio de oficinas tiene 30 equipos de aire acondicionado, ¿cual sería el tiempo medio de trabajo del técnico asignado a dicho edificio? Si la tarea debe terminarse en una sola jornada de 8 horas, cuantos técnicos deben asignarse al edificio?

#### Resolución

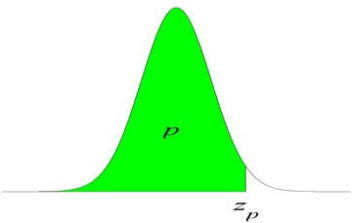
Si tenemos 30 equipos de aire acondicionado que operan en forma independiente, podemos considerar que el tiempo necesario para realizar la manutención de cada equipo es una variable aleatoria, que todas son independientes entre sí, y tienen igual distribución. Por lo cual el tiempo medio de trabajo del técnico es también una variable aleatoria  $Y = X_1 + \dots + X_{30}$ . La esperanza de dicha variable sería un indicador de cuanto tiempo necesita ser asignado al edificio el técnico para poder realizar su tarea.

Como las variables son independientes e idénticamente distribuidas,

$$E(Y) = E(X_1 + \dots + X_{30}) = 30E(X_1) = \frac{30}{\lambda} = 30 \text{ horas}$$

Si la jornada tiene 8 hs, por lo menos cuatro días de trabajo debe tener asignados a ese edificio. Ahora, si todos los técnicos trabajan de la misma forma, es decir, la distribución de probabilidad del tiempo de trabajo de cada técnico es la misma, se deben asignar cuatro técnicos para tener el trabajo terminado en una jornada laboral.

Función de distribución acumulada de una variable normal estandar.



$Z_p$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Figura 5.9: Tabla cola izquierda



# Capítulo 6

## Distribuciones de Muestreo

### 6.1. Introducción a la metodología estadística

El uso de métodos estadísticos en Ciencias Exactas data del siglo XVII. Hemos visto como los grandes matemáticos de la época estaban interesados en modelar teóricamente los juegos de azar, pero también los avances en las mediciones astronómicas, entre otras mediciones físicas, imponían el estudio de un modelo que permitiera incluir el error de medición. Era importante poder predecir, dentro de la precisión de dicho error, observaciones futuras.

Los avances en las ciencias biológicas, agronómicas y el advenimiento en ambas de las teorías genéticas, también tuvieron su impacto en las teorías estadísticas. Las propuestas de Sir Ronald Fisher, el gran matemático, biólogo y por sobre todo, estadístico, a principio del siglo XX, revolucionaron la concepción de distribución de una muestra, tarea que fue profundizada por Jerzy Neyman y Karl Pearson.

En los comienzos del siglo XX, las disciplinas de las ciencias sociales presenciaron a su vez un aumento paulatino en el uso de métodos estadísticos. El estudio de como comparar cultivos para que sus resultados sean significativos estadísticamente que realizó Sir Ronald Fisher, tienen su contraparte en los estudios de diseños de muestras para encuestas y otros estudios demográficos. Como la investigación en otras ciencias, la investigación en Ciencias Sociales se ha vuelto fuertemente orientada hacia el análisis de datos empíricos. La revolución informática ha vuelto accesible una enorme cantidad y variedad de información acercándola tanto a los investigadores como a los estudiantes y al público en general. Las computadoras también han facilitado el uso de herramientas estadísticas.

En el mundo de hoy, la comprensión de la estadística se ha vuelto esencial en muchas profesiones desde la medicina hasta los negocios. Pero la estadística es importante aún para aquellos profesionales que no usan métodos estadísticos como parte cotidiana de su trabajo. Todos los días están expuestos a una explosión de información desde la publicidad, las noticias, las campañas políticas las encuestas de opinión sobre temas controvertidos y otras comunicaciones que contienen argumentos estadísticos. La estadística nos permite ordenar y entender esta información para entender mejor la realidad.

Para introducirnos dentro de los conceptos de muestra y población, consideremos un ejemplo:

**Ejemplo 6.1.1** Supongamos que se quiere estudiar y comparar la cantidad de palabras del idioma castellano que conoce un niño que cursa 3er grado de la escuela primaria en la región noroeste del país, Salta, Jujuy, Tucumán, y La Rioja, con los de los niños de la ciudad de Buenos Aires.

Un primer paso razonable es el de diseñar una herramienta para medir esto en forma no traumática para el niño (test), y luego aplicar este test a todos los niños de las escuelas de las provincias del noroeste, y de la ciudad de Buenos Aires. Esto constituye un censo, y representa un claro problema económico, el costo de coordinación, de preparación del material, entrega, y finalmente de procesamiento de los datos es elevado.

Un modo más económico de hacerlo es buscar una muestra es decir, un subgrupo de niños en cada una de las ciudades, que sea bien representativa de la población de interés y luego comparar los datos conseguidos.

Este método va a tener dos problemas fundamentales:

- Como elegir las muestras
- Como comparar los resultados

Para poder estudiar estos problemas vamos a definir ciertos conceptos:

**Definición 6.1.1** *La clase objetivo de nuestro estudio se llama población, es el conjunto de elementos del cual queremos estudiar algunas características.*

**Definición 6.1.2** *La parte de la población que puede ser examinada se llama muestra. El objetivo del estudio estadístico es hacer generalizaciones de la parte observada en la muestra al todo, esto es, hacer inferencia de la muestra a la población.*

Dejando para más adelante la descripción de cómo realizar las muestras, pensemos como realizar esta inferencia de la muestra a la población.

**Definición 6.1.3** *Si suponemos que la población tiene una distribución de probabilidad gobernada por parámetros, valores fijos, desconocidos, entonces la inferencia se realizará intentando dar información sobre estos parámetros, basados en los valores de la muestra.*

Por ejemplo, la cantidad promedio de palabras del cuestionario que los chicos del noroeste conocen, y la cantidad de palabras promedio del cuestionario que los chicos de la Ciudad de Buenos Aires conocen. Ambos son parámetros poblacionales. La cantidad promedio de palabras que conocen los chicos de las muestras elegidas son los estadísticos muestrales, que estiman los parámetros poblacionales.

El problema crucial es como precisar el error de estimación, es decir como estimar el error cometido con el estadístico si no se conoce el valor real del parámetro. Estimar los parámetros de la población a partir de la muestra se justifica cuando la muestra representa a la población. Esto es imposible de verificar, por lo cual se debe estudiar la forma en que la muestra fue elegida.

Los siguientes ejemplos ilustran sobre los siguientes hechos:

1. El método de elección de la muestra importa mucho.
2. Los mejores métodos involucran la introducción planificada del azar.

### **Ejemplo 6.1.2** Gallup poll

En 1936 Franklin Delano Roosevelt estaba compitiendo por primera vez en las elecciones presidenciales de los Estados Unidos, por el partido Demócrata. Por el partido republicano el candidato era Alfred Landon. El país estaba tratando de recuperarse de la Gran Depresión. Había aún 9 millones de desocupados, el ingreso real había bajado drásticamente en el periodo 1929-1933 y recién comenzaba a recuperarse. La campaña de Landon estaba basada en un programa de restricción de los gastos de gobierno, Roosevelt se presentó con un programa más populista. Muchos de los observadores pensaban que Roosevelt ganaría fácilmente. Sin embargo, la revista Literary Digest predijo una victoria de Landon por 57 % a 43 %.

Esta predicción estaba basada en una muestra de 2.400.000 individuos, y estaba apoyada por el enorme prestigio de la Digest, que había predicho los ganadores de todas las elecciones presidenciales desde 1916 correctamente. Sin embargo, Roosevelt ganó la elección de 1936 por 62 % a 38 %.

George Gallup, fundador de una de las empresas de consultoría estadística más importante de los Estados Unidos, recién comenzaba su organización. Usando una muestra de 50.000 individuos, predijo correctamente la victoria de Roosevelt, aunque le asignó el 56 % de los votos, cuando el resultado correcto fue del 62 %, presentando un error del 6 %.

Para detectar las causas del error de la Digest, uno debe preguntarse como seleccionaron la muestra. Debemos notar que la muestra de Gallup fue 50 veces menor, y sin embargo acertó. En principio, no parece ser falta de datos lo que produjo la falla en la estimación de la Digest.

**Definición 6.1.4** *Un procedimiento de muestreo debe seleccionar los elementos a incluir en la muestra de una forma imparcial y obtener una muestra representativa de los datos.*

**Definición 6.1.5** *Una tendencia sistemática en el procedimiento de selección de la muestra al excluir algún tipo de elementos o sobreponderar algún grupo se llama sesgo de selección*

El procedimiento de Digest consistió en enviar por correo 10 millones de cuestionarios. Obtuvieron 2,4 millones de respuestas en las que basaron sus proyecciones. Los nombres y direcciones de las 10 millones de personas fueron obtenidos de directorios telefónicos o listados de clubes. Pero en 1936 había 11 millones de teléfonos residenciales y 9 millones de desocupados, (que podemos suponer no tenían acceso a teléfono).

El procedimiento de selección de la muestra de la Digest tenía un marcado sesgo sistemático contra los sectores de menores recursos económicos. Antes de 1936, el sesgo contra los pobres podía no haber afectado demasiado las proyecciones electorales ya que los ricos y pobres votaban de forma similar. Pero en 1936, los pobres y la clase media empobrecida, votaron mayoritariamente a Roosevelt mientras que los ricos votaron mayoritariamente a Landon. Uno de los motivos del gran error de la Digest fue el sesgo sistemático en el procedimiento de elección de la muestra.

Cuando un procedimiento de selección es sesgado, tomar una muestra más grande no ayuda. Solamente repite el error básico a una escala más grande. Pero hay un segundo problema. Una vez que se selecciona que personas integrarán la muestra hay que obtener las respuestas. Si una cantidad importante de las personas seleccionadas para la muestra no responden al cuestionario o entrevista, esto puede crear una seria distorsión llamada sesgo de no respuesta, o en general, de no adherencia al protocolo. Las personas que son parte de un estudio médico y dejan de tomar las pastillas, o lo hacen esporádicamente, también producen un sesgo de no adherencia al protocolo, médico en este caso.

En el caso de las encuestas, los que no adhieren al protocolo difieren de los que adhieren al protocolo de una forma obvia: no respondieron. La experiencia ha mostrado que en general tienden a diferir en otras formas también, por lo cual la muestra de los que quedan es muy diferente a la población real.

En la misma elección, Digest hizo un estudio en la ciudad de Chicago en que enviaron cuestionarios por correo a uno de cada tres registrados para votar, es decir, la muestra era más representativa que el listado telefónico. Alrededor del 20 % respondió y entre los que respondieron, más de la mitad lo hicieron a favor de Landon. Pero en Chicago, Roosevelt obtuvo más del 65 % de los votos.

Los no respondientes pueden ser muy distintos a los que responden. Cuando hay una tasa importante de no respuesta, se debe estudiar si hay sesgos de no-respuesta. En general, los niveles socio-económicos más bajos y más altos tienden a no responder a los cuestionarios y por lo tanto los sectores medios están sobre-representados entre los respondientes. Es por ello que son preferibles las entrevistas personales, en vez de los cuestionarios por correo. Un porcentaje de respuesta típico para encuestas personales es del 75 % contra 25 % para cuestionarios por correo.

**Ejemplo 6.1.3** Las siguientes son predicciones realizadas por 3 empresas de consultoría estadística sobre los resultados en las elecciones presidenciales de 1948 en los Estados Unidos.

Candidato	Predicción de Crossley	Predicción de Gallup	Predicción de Roper	Resultado de la elección
TRUMAN	45 %	44 %	38 %	50 %
DEWEY	50 %	50 %	53 %	45 %
THURMOND	2 %	2 %	5 %	3 %
WALLACE	3 %	4 %	4 %	2 %

Para poder averiguar porqué ninguna de las empresas dió por ganador al candidato demócrata Truman, es necesario averiguar como eligieron las muestras.

El método que usaron las tres empresas fue lo que se llama “ muestreo en cuotas ”. En este procedimiento a cada encuestador se le asigna una cuota fija de individuos a entrevistar: la cantidad de individuos por ciertas categorías (en general lugar de residencia, sexo, edad, nivel socioeconómico) también se fija. Salvo estas restricciones el encuestador puede seleccionar a quién considere. Por ejemplo, un encuestador de Gallup en Saint Louis debía elegir 13 individuos de los cuales

- 6 debían vivir en los suburbios, y 7 en la parte céntrica.
- 7 debían ser hombres y 6 mujeres
- de los 7 hombres (había cuotas similares para mujeres también)
  - 3 debían tener menos de 40 años y 4 mas de 40.
  - 6 debían ser blancos y 1 negro
- el alquiler mensual que pagaban los hombres blancos también estaba especificada
  - 1 debía pagar más de 44,01 dólares.
  - 3 debían pagar entre 18.01 y 44 dólares,
  - y dos debían pagar menos de 18 dólares.

A primera vista el muestreo por cuotas asírealizado era razonable. Parece garantizar que la muestra será como la población respecto a muchas características importantes que pueden afectar el comportamiento electoral, pues hay datos censales previos sobre la variables demográficas. Sin embargo este procedimiento funcionó mal en 1948.

El deseo de las empresas era diseñar una muestra que represente equitativamnete las opiniones políticas de la población. Sin embargo no se pueden asignar cuotas a los votantes republicanos y demócratas porque esas cuotas es lo que queremos estudiar. Las cuotas en las otra variables reflejan un intento indirecto para lograr que la muestra refleje las opiniones políticas de la Nación. Pero hay muchos factores que inciden sobre el comportamiento de los votantes, además de los que elegimos controlar, que pueden hacer que el comportamiento de la muestra y la población sean distintos.

La principal característica de este diseño que actuó en contra de los deseos de los encuestadores fue el hecho de que dentro de las cuotas asignadas, el encuestador puede elegir a quien le parezca. Esto deja una importante parte de la elección en manos humanas, y la elección humana está siempre sujeta a sesgos. En 1948 los encuestadores eligieron demasiados republicanos. Globalmente los republicanos son más saludables y mejor educados que los demócratas. Tienden a tener teléfono, dirección permanente y viven en mejores edificios. En cada grupo, los republicanos son un poco más fáciles de entrevistar. Por lo cual los encuestadores terminaron entrevistando a demasiados republicanos. La siguiente tabla muestra como en cada elección presidencial desde 1936 a 1948 los encuestadores entrevistaron a más republicanos de lo que la proporción de votantes real hubiera sugerido. También, antes de 1948, la diferencia a favor de los demócratas era tan grande que sobrepasaba el sesgo a favor de los republicanos de la muestra y Gallup predecía correctamente el ganador.

Año	Proyección de votos a republicanos	Resultado electoral	error a favor de los republicanos
1936	44 %	38 %	6 %
1940	48 %	45 %	3 %
1944	48 %	46 %	2 %
1948	50 %	45 %	5 %

En 1948 el error cometido fue tan grande que, agregado al crecimiento real republicano, hizo pasar al candidato republicano la barrera de la mayoría y dió como ganador al perdedor.

**Definición 6.1.6** *En el muestreo por cuotas, la muestra es seleccionada para replicar la población respecto de algunas características consideradas claves. El método parece razonable, pero no funciona bien pues incorpora sesgos no-intencionales, al permitir al encuestador elegir arbitrariamente la muestra. La alternativa es usar un mecanismo objetivo e imparcial de azar para seleccionar la muestra. Ello dará origen al muestreo aleatorio simple, y al muestreo aleatorio estratificado.*

Existen casos en los que no se puede evitar el uso de estudios observacionales. No se puede realizar una muestra aleatoria de la población y realizarle un bypass coronario para ver como evoluciona despues de la operación. Tampoco se puede elegir una muestra de la población con gripe y pedirle a la mitad que no tome



ni una aspirina, para ver como se compara con la otra mitad que toma antigripales. La ética en estos casos juega en contra de la completa aleatorización, pero aún así los estadísticos especialistas en salud pública estudian métodos de diseño de experimentos que permitan obtener información válida, contrarrestando lo más posible los factores de confusión.

**Definición 6.1.7** *Se denomina experimento a un estudio planeado, con una muestra diseñada para contrarrestar los efectos de factores de confusión. Se llaman estudios observacionales a aquellos estudios en donde no se puede elegir a los intervinientes en el estudio, ya sea por una cuestión ética como por escasez de elementos de la población.*

**Ejemplo 6.1.4** La prueba de campo de la vacuna Salk (antipoliomelítica).

¿Como se estudian los efectos de las drogas medicinales?. En forma simplificada, podemos decir que el método básico para testear una droga nueva es la comparación. La droga es suministrada a sujetos que forman el grupo tratamiento, y se elige otro grupo de pacientes que forman el grupo de control, y que no son tratados. Luego se comparan las respuestas de los dos grupos. La experiencia muestra que los individuos deberían ser asignados a los grupos tratamiento y control al azar, y que el experimento debe ser a doble ciego, es decir, que ni los individuos ni los médicos deberían saber quién ha tomado la droga y quién no.

Vamos a comentar un estudio realizado en la mitad del siglo pasado para verificar la eficiencia de la droga antipoliomelítica. La primera epidemia de poliomielitis en Estados Unidos se observó en 1916, y durante los siguientes 40 años causó miles de víctimas, en especial niños. En la década del 50 se descubrieron varias vacunas contra la enfermedad. La vacuna desarrollada por Jonas Salk parecía muy prometedora. En ensayos de laboratorio sobre animales se mostró segura, sin efectos colaterales y provocó la producción de anticuerpos para la polio. Un estudio de campo a gran escala era necesario para ver si la vacuna podría proteger a los chicos contra la polio fuera del laboratorio.

En 1954, el Servicio de Salud Pública Estadounidense decidió organizar este tipo de experiencia. La prueba fue llevada a cabo en niños en los grupos de todas las edades más vulnerables (1er, 2do y 3er grado). La prueba fue realizada en distritos escolares seleccionados a través del país, aquellos en los cuales se pensaba que el riesgo de contraer polio era peor. La experiencia involucró a 2 millones de niños y se vacunaron medio millón. Cerca de un millón de niños fueron deliberadamente dejados sin vacunar y otro medio millón rechazó la vacuna. Esto ilustra el método de comparación, solo los sujetos que pertenecen al grupo tratamiento fueron vacunados: los otros individuos no recibieron tratamiento y se usaron como controles. Las respuestas de ambos grupos pueden entonces ser comparadas para ver si el tratamiento produjo alguna diferencia. En el ensayo de la vacuna Salk los grupos tratamiento y control no fueron del mismo tamaño, pero eso no tiene importancia. Los investigadores compararon las tasa, (proporciones), de niños que contrajeron polio en los dos grupos, y se muestran como cantidad de casos cada 100.000 niños. Mirar las proporciones en lugar de los números absolutos ajusta las diferencias de tamaño existentes.

Nos encontramos aquí ante nuestro primer dilema ético. ¿no debería haberse dado la vacuna a todos los chicos? Una respuesta posible es que en el caso de las drogas nuevas, aun con buenos resultados de laboratorio con animales, es poco claro en cuánto el beneficio compensa el riesgo. Un ensayo de campo es necesario para saber cuales son los efectos del tratamiento cuando se lo usa en el mundo real. Se puede argumentar también que darle la vacuna a un gran número de chicos puede producir evidencia decisiva, aún sin controles. Por ejemplo, si la incidencia de la polio baja con respecto a la del año anterior, podríamos decir que la vacuna fue efectiva. Pero en realidad, la polio es una enfermedad epidémica, y no mantiene las mismas tasas de infectados todos los años. Por ejemplo, antes de la vacunación, en 1952 hubo cerca de 60 000 casos y en 1953 solo la mitad. La reducción de 1954 podría haber sido efecto de una epidemia menos severa y no de un efecto preventivo debido a la vacuna.

La única manera de saber si la vacuna funcionaba era dejar algunos chicos sin vacunar. Por supuesto, los chicos solo podían ser vacunados con el permiso de los padres, luego un diseño posible sería dejar sin vacunar a los niños sin permiso, y vacunar a todo el resto. Pero se sabía que los padres con mayor nivel adquisitivo dan su consentimiento mas fácilmente que los de bajos ingresos y eso produciría un sesgo en contra de la vacuna. ¿Porque? Pues los niños de hogares con mejor nivel socio económico eran más vulnerables a la polio.

Paradójicamente, la polio es una enfermedad de higiene. Los niños que viven en lugares menos higiénicos tienden a contraer casos leves de polio en los primeros años de vida, cuando aun tienen los anticuerpos de las madres. Una vez infectados, desarrollan sus propios anticuerpos que los protegen de infecciones posteriores más severas. Los niños

que viven en ambientes más higiénicos no desarrollaban estos anticuerpos y la primera infección se desarrollaba con todo su poder.

Los diseñadores sabían que debían evitar los sesgos, y que los grupos tratamiento y control debían ser lo más parecidos posible, salvo por el tratamiento. Esto permitiría concluir que cualquier diferencia entre los grupos era dada a la vacunación y no a factores ajenos al tratamiento, llamados factores de confusión.

Respecto al ensayo de campeo de la vacuna Salk, se propusieron varios diseños. La National Foundation for Infantile Paralysis (NFIP) de EEUU propuso vacunar a todos los niños de segundo grado cuyos padres dieran el consentimiento dejando los de 1ro y 3ro como controles. Este diseño fue aceptado en muchos distritos escolares. Sin embargo tenía dos inconvenientes serios:

1. La polio es una enfermedad que se contagia por contacto. Luego si la incidencia de la polio es mayor en 2do grado que en 1ro y 3ero esto produciría un sesgo en contra de la vacuna. En cambio si la incidencia fuera mucho menor en 2do grado que en 1ro y 3ero, el sesgo sería a favor de la vacuna.
2. Los niños para los cuales el consentimiento de los padres era requerido, podrían tener un entorno familiar diferente de los del grupo control a los cuales no se les requería el consentimiento paterno. Con el diseño NFIP, el grupo tratamiento incluiría demasiados chicos de nivel socio económico alto, haciendo que este grupo fuese mas vulnerable a la polio que el grupo control, generando un claro sesgo en contra de la vacuna.

Muchos distritos escolares vieron las falencias en el diseño NFIP y adoptaron otros diseños. El grupo tratamiento y control debían pertenecer a la misma población, la de los niños con consentimiento para la vacunación. De otra forma el efecto socio económico hubiera confundido el efecto de la vacuna. El otro problema a resolver era como asignar efectivamente los niños a cada grupo. Parecía necesario realizar una especie de distribución contemplando factores como edad, estado de salud del chico, personalidad, nivel socio económico, etc. Pero la experiencia muestra que las decisiones humanas tienden a incluir sesgos grandes. Por lo cual el procedimiento elegido en el ensayo Salk fue equivalente a tirar la moneda. Cada chico tenía un 50% de chances de entrar a uno u otro grupo. Tal procedimiento era imparcial y objetivo. Y las leyes de la probabilidad aseguran que si hay suficientes sujetos asignados, ambos grupos se equilibrarán lo suficiente con respecto a todas las variables importantes, hayan sido identificadas estas o no. Cuando un procedimiento imparcial para asignar los sujetos al grupo control y al tratamiento es utilizado en un experimento, este se llama aleatorizado controlado.

Otra precaución básica en el ensayo Salk fue la del uso del placebo. A los niños del grupo control se les dio una inyección de sal disuelta en agua de modo que durante la experiencia no supieran si ellos eran el grupo tratamiento o control, (o mejor, sus padres no supieran en que grupos estaban sus hijos). Esto asegura que la respuesta a la vacuna es más que la idea de que se les ha procurado un tratamiento. Parece poco verosímil, pero esta comprobada la existencia de un efecto placebo, pacientes que experimentaban grandes dolores después de una operación se les dio una pastilla de azúcar y mas de un tercio dijo haber experimentado un rápido alivio.

El efecto padres es otro sesgo controlado por el placebo. Como los familiares no sabían si los niños habían sido vacunados o no, los niños del grupo control no tendrían un comportamiento distinto a los del otro grupo. Como la polio se contagia por contacto, los padres de los niños no vacunados podrían reducir las salidas de sus hijos, asistir a menos eventos sociales y no dejar que los niños se relacionen con otros niños. Esto hubiera introducido diferencias de comportamiento entre los niños de ambos grupos, que era necesario evitar.

Una precaución mas fue tomada. Los médicos que debían diagnosticar la polio a los pacientes del estudio no fueron informados de la etiqueta de cada participante. Como algunas formas de poliometitis son difíciles de diagnosticar, esta precaución garantizaba que, en casos limites, el diagnostico del médico no estuviera influenciado por su vision personal de la efectividad de la vacuna. Este tipo de experiencia se denomina doble ciego: ni los sujetos ni los que evalúan la respuesta al tratamiento conocen quien ha sido tratado y quien no. Los distritos que eligieron realizar este experimento controlado, aleatorizado a doble ciego, eligieron el mejor diseño que existe.

Contaremos ahora cuales fueron los resultados del experimento.

sc Estudio aleatorizado controlado doble ciego

	Tamaño de la muestra	Tasa de polio cada 100000
Tratamiento	200 000	28
Control	200 000	71
Sin consentimiento	350 000	46

sc Estudio NFIP

	Tamaño	Tasa de polio cada 100 000
2do grado (vacunados)	225000	25
1ro y 3er grado	725000	54
2do grado (sin consentimiento)	125000	44

Fuente: Thomas Francis Jr, American Journal of Public Health. vol. 45 (1955), p1–63. Los números fueron redondeados para esta ilustración.

Si comparamos ambas tablas observamos que la proporción de infectados fue mucho menor para el grupo tratamiento en el experimento aleatorizado, lo que constituye una prueba decisiva de la efectividad de la vacuna Salk. También podemos notar que el experimento con el diseño de NFIP estuvo sesgado en contra de la vacuna, pues el descenso en el número de infectados de polio de 54 a 25 cada 100 000 niños es mucho menor que en el diseño aleatorizado.

El diseño aleatorizado controlado a doble ciego reduce el sesgo al mínimo y esa es la razón para usarlo cuando sea posible. Más aún, este diseño tiene una importante ventaja técnica. Para observar esto, pensemos lo siguiente. Supongamos que la vacuna no hace realmente ningún efecto, en este caso las diferencias que se ven entre las tasas de infectados en ambos grupos son solo debidas al azar. En el diseño NFIP, los resultados se ven afectados por sesgos que los modifican, y los investigadores no tienen como calcular cual es realmente el peso de esos factores. En cambio en el diseño completamente aleatorizado, el azar entra en juego de una forma planificada y simple, cuando se decide quien asignar a control y quien a tratamiento. Por lo cual, si la vacuna realmente no hace efecto, hay chicos predestinados a contraer la polio y chicos que no la van a contraer, y al tirar la moneda, estamos asignando mas o menos la misma proporción de predestinados a enfermarse en un grupo y en el otro. Bajo este supuesto, cualquier diferencia entre las tasas es debida a la variabilidad aleatoria de la tirada de la moneda. Bajo este modelo, se puede calcular la probabilidad de observar una diferencia tan grande como la que se obtuvo. Esta probabilidad no es cero, pero es una en un billon. Y eso es muy chico. Por eso es que se declaró la vacuna efectiva y se la aplica a todos los niños al cumplir un año de edad (con autorización paterna). Gracias a la vacunación ya no existen epidemias.

Los experimentos controlados son distintos a los estudios observacionales. En un experimento controlado el investigador decide quien estará en el grupo tratamiento y quién estará en el grupo control. En cambio en un estudio observacional son os sujetos por símismos quienes se asignan a los distintos grupos, los investigadores solo observan lo que sucede.

**Ejemplo 6.1.5** Los estudios sobre el efecto de fumar, por ejemplo, son claramente observacionales: nadie va a fumar para complacer a un estadístico. Sin embargo la idea de control y tratamiento se usa de todos modos. Los investigadores comparan fumadores (el grupo tratado o “expuesto”) con no fumadores (el grupo control) para determinar el efecto del hábito de fumar. Estudios muy cuidadosos sobre fumadores y no fumadores, controlados por rango de edad, género, sedentarismo y otras variables muestran evidencia de que fumar es perjudicial para la salud. En otros casos, la evidencia obtenida a partir de estudios observacionales no es tan clara.

**Ejemplo 6.1.6** La primera vez que se observó la enfermedad llamada pellagra o escorbuto alpino fue en Europa, durante el siglo XVIII. Fue detectada por un medico español, Gaspar Casal, que encontró que era una causa importante de la mala salud, discapacidad mental y muerte prematura de los habitantes pobres de Asturias. En los años subsiguientes varios autores describieron las mismas condiciones de salud de campesinos de Italia del Norte, en particular de la región de Lombardia. A principios del siglo XIX el escorbuto alpino se había desparramado por toda Europa, como un cinturón, causando el progresivo deterioro físico y mental de miles de personas en el suroeste de Francia, Austria y Rumania y en los dominios del Imperio Turco. Fuera de Europa, el escorbuto alpino fue detectado en Egipto y Sud Africa, Y durante la primera década del siglo XX en EEUU, especialmente en los estados del Sur.

El escorbuto alpino parecía afectar a algunos pueblos mucho más que a otros. Aún dentro de los pueblos afectados, muchos hogares no mostraron ningún miembro contagiado, mientras que otros tenían enfermos año tras año. Las condiciones sanitarias de las casas infectadas eran primitivas, había moscas por todas partes. Una mosca chupa-sangre, de la variedad *simulium* tenia la misma distribución geográfica que el escorbuto alpino, al menos en Europa,

y la mosca estaba más activa en la primavera, que es cuando la mayor parte de los casos de escorbuto alpino se desarrollaba. Varios epidemiólogos concluyeron que la enfermedad era infecciosa, como la malaria, la fiebre amarilla, o el tifus, y era transmitida de una persona a otra por insectos.

El epidemiólogo estadounidense Joseph Goldberger mostró, a través de una serie de estudios observacionales y experimentos que empezaron en 1914, que el escorbuto alpino era causado por una mala dieta, y que no es infeccioso. La enfermedad puede ser prevenida y curada con una alimentación rica en lo que Goldberger llamo el factor PP (pellagra preventive). Desde 1940 la mayor parte de la harina que se vende en los EEUU está enriquecida con el factor PP además de algunas vitaminas. En Argentina, el factor PP se llama niacina y puede verse en la etiqueta de muchas galletas y alimentos para niños.

La niacina está naturalmente contenida en la leche, carne, huevos y algunos vegetales. El maíz, sin embargo contiene poca niacina. En las zonas de escorbuto alpino los pobres comían maíz, y quizás nada más. Algunos pueblos y algunos hogares eran mas pobres que otros, y tenían dietas aún más restringidas. Por esa razón, la enfermedad los afectaba más. Las moscas eran una marca de la pobreza, no la causa del pellagra.

## 6.2. Distribuciones del muestreo

### 6.2.1. Tipos de Muestreo

En esta sección vamos a empezar a estudiar la matemática necesaria para poder realizar las inferencias que comentamos en los estudios reales de la sección anterior. Vamos a comenzar describiendo los métodos para seleccionar muestras de poblaciones, y vamos a continuar describiendo lo que son las distribuciones del muestreo:

- Muestreo no aleatorio o de juicio: Se emplea el conocimiento y la opinión personal para identificar aquellos elementos de la población que deben incluirse en la muestra.
- Muestreo aleatorio o de probabilidad: En el cual todos los elementos de la población tienen la oportunidad de ser escogidos para la muestra. Dentro de este tipo de muestreo se encuentran:
  - Muestreo aleatorio simple: Es la forma más elemental de muestrear. Cada muestra en particular de tamaño  $n$  tiene la misma probabilidad de ser observada, esto es, cada uno de los  $\binom{N}{n}$  posibles subconjuntos de tamaño  $n$  de la población de  $N$  objetos, tiene la misma probabilidad de ser elegido. Es equivalente a elegir al azar de una urna con  $N$  elementos, sin devolver los elementos extraídos a ella antes de realizar la siguiente extracción.
  - Muestreo sistemático: método en el cual los elementos de la muestra se seleccionan de la población en un intervalo uniforme que se mide con respecto al tiempo, al orden o al espacio. Por ejemplo, si tenemos una lista de objetos ordenada en forma aleatoria, si elegimos uno de cada  $N/n$  elementos de la lista, el resultado es equivalente al muestreo aleatorio simple.
  - Muestreo estratificado: método en el que la población se divide en grupos homogéneos, o estratos, y después se toma una muestra aleatoria simple de cada estrato. Aquí la variabilidad dentro de cada grupo es pequeña y entre los grupos es grande.
  - Muestreo de racimo: método en el que la población se divide en grupos o racimos de elementos, y luego se selecciona una muestra aleatoria de estos racimos. La variabilidad dentro de cada grupo es grande y entre los grupos es pequeña; es como si cada racimo fuese una pequeña representación de la población en si misma.

En la terminología estadística, la distribución que se obtendría al tomar todas las muestras de un tamaño dado constituye una distribución teórica de muestreo. En la práctica, el tamaño y el carácter de la mayor parte de las poblaciones impiden que los responsables de las decisiones tomen todas las muestras posibles

de una distribución de población, sin embargo, se han desarrollado fórmulas para estimar las características de estas distribuciones teóricas de muestreo, haciendo innecesario que se recolecten grandes números de muestras. En casi todos los casos, los responsables de las decisiones sólo toman una muestra de la población, calculan estadísticas para esa muestra y de esas estadísticas hacen inferencias sobre los parámetros de toda la población.

En las próximas secciones nos dedicaremos a estudiar el comportamiento estadístico de una población. El nombre “población” se utiliza en Inferencia estadística por motivos históricos, que tienen que ver con el estudio de características vinculadas a las poblaciones humanas, aunque muchas veces el objeto de estudio sea de una naturaleza completamente distinta.

**Ejemplo 6.2.1** Por ejemplo, si queremos conocer cuál es la proporción de “floggers” entre los alumnos de nuestra escuela (la población), podemos hacer una investigación exhaustiva (censo) entre todos los alumnos matriculados y preguntarles si se consideran “floggers” o no, o bien, seleccionar una parte representativa de la población de alumnos para extraer conclusiones que serán extrapoladas a toda la población. Aunque en este caso la población en estudio es una población humana, nos conviene más decir que la población es la característica que se está estudiando.

En concreto, definimos la variable aleatoria  $X$  que vale 1 si un alumno elegido al azar es fumador, y cero si no lo es. Obsérvese que  $X$  tiene una distribución de Bernoulli,  $b(p)$ , donde  $p$  es la probabilidad de que el alumno elegido al azar sea un “flogger”, esto es, la proporción de “floggers” del colegio. Decimos entonces que tenemos una población  $X \sim b(p)$ , es decir, denominamos población a la variable aleatoria que representa su comportamiento.

**Ejemplo 6.2.2** Supongamos que deseamos conocer si una moneda es balanceada o no. El comportamiento de esa moneda puede describirse como una variable aleatoria,  $Y$ , que vale uno si sale cara, y cero si sale número. En este caso, la población sería la variable aleatoria  $Y \sim b(p)$ , donde  $p$  sería la probabilidad de obtener cara en un lanzamiento de la moneda. Obsérvese que, aunque el modelo es el mismo, ahora la población no es humana, ni siquiera de objetos.

**Ejemplo 6.2.3** Por ejemplo: si en una Universidad hay 3000 alumnos de los cuales 1000 son fumadores, si escogemos una muestra representativa formada por 30 alumnos, más o menos 10 de ellos deberían ser fumadores. Esta cifra no se alcanza habitualmente con exactitud, dadas las posibles desviaciones provocadas por el azar. Esto implica que, aunque el proceso de selección de la muestra es muy sensato, el azar hace que sea posible, por ejemplo, que los 30 seleccionados sean fumadores (aunque sea muy raro que ello ocurra).

Cada alumno a encuestar se representa mediante una variable que se distribuye igual que  $X$ . Si, en general, encuestamos a  $n$  alumnos, les representaremos mediante un vector aleatorio  $n$ -dimensional,  $\mathbf{X}$ , que denominaremos muestra. Una vez encuestados, los resultados serán  $n$  números,  $\mathbf{x}$ , que denominaremos realización de la muestra.

**Definición 6.2.1** *La distribución del muestreo es entonces la distribución conjunta del vector de la muestra.*

En el caso del muestreo aleatorio simple cuando el tamaño de la población es mucho mayor que el tamaño de la muestra, las variables que representan a cada elemento de la muestra pueden suponerse independientes entre sí y la distribución del muestreo toma una forma simple, es el producto de todas las distribuciones marginales. Esto no es así para todos los tipos de muestreo, ni siquiera para el muestreo aleatorio simple cuando la población es pequeña, pero es el caso que podemos estudiar en detalle con la matemática que hemos desarrollado en los capítulos anteriores. Por lo cual, de ahora en más llamaremos muestra aleatoria simple a un conjunto de variables idénticamente distribuidas e independientes, es decir, réplicas independientes de las características que estamos observando. En los ejemplos haremos notar cuando estas variables provienen de un muestreo aleatorio simple producido sobre una población muy grande.

## 6.2.2. Muestreo de una distribución teórica

En la sección anterior, comentamos distribuciones de probabilidad obtenidas mediante aleatorización deliberada. Si los datos son obtenidos mediante la repetición deliberada de un mismo experimento, es razonable

considerar a estos como una muestra independiente de variables con una misma distribución. Ahora cuando estos datos provienen de mediciones realizadas sobre objetos o sujetos elegidos mediante algún método de muestreo, la independencia de las mediciones no es muy clara, si no terminantemente falsa. Sin embargo, conociendo el método de muestreo se puede describir la distribución conjunta de la muestra, estimar medias, varianzas, dar estimación de errores cometidos, entre otras cosas, pero estos métodos estadísticos son más complicados que los que vamos a estudiar en este libro. De ahora en más nos vamos a dedicar al caso más simple, el de datos independientes.

**Definición 6.2.2** Una variable aleatoria  $n$ -dimensional  $X_1, \dots, X_n$  es una **muestra aleatoria** si las  $n$  variables aleatorias son independientes y tienen la misma distribución  $f_X$ , luego la distribución conjunta de la muestra es

$$f = \prod_{i=1}^n f_{X_i}.$$

El valor  $n$  se llama tamaño de la muestra. Los valores  $x_1, \dots, x_n$  observados se llaman realizaciones de la muestra.

**Proposición 6.2.1** Sea  $X_1, \dots, X_n$  una muestra aleatoria de variables con  $E(X) = \mu$  y  $V(X) = \sigma^2$ . La media muestral es la variable aleatoria

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

y la varianza muestral es

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Entonces,  $E(\bar{X}) = \mu$  y  $E(S^2) = \sigma^2$ .

**Demostración:**

Observemos primero que al ser las  $X_i$  independientes

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu = \mu$$

y que

$$Var(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

También

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - 2 \sum_{i=1}^n X_i\bar{X} + n\bar{X}^2 \right) = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \right) \\ &= \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \end{aligned}$$

Por lo cual

$$E(S^2) = \frac{1}{n-1} \left( \sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \right)$$

Como

$$E(X_i^2) = \text{Var}(X_i) + E(X_i)^2 = \sigma^2 + \mu^2$$

y

$$E(\bar{X}^2) = \text{Var}(\bar{X}) + E(\bar{X})^2 = \frac{\sigma^2}{n} + \mu^2$$

Entonces

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} \left( \sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \right) = \frac{1}{n-1} \left( \sum_{i=1}^n (\sigma^2 + \mu^2) - n \left( \frac{\sigma^2}{n} + \mu^2 \right) \right) \\ &= \frac{1}{n-1} (n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2) = \frac{(n-1)}{(n-1)} \sigma^2 = \sigma^2 \end{aligned}$$

**Proposición 6.2.2** Sean  $X_1, \dots, X_n$  variables independientes tales que  $X_k$  tiene distribución normal con parámetros  $\mu_k, \sigma_k^2$ . Entonces  $X_1 + \dots + X_n$  tiene distribución normal con parámetros  $\mu = \mu_1 + \dots + \mu_n$  y  $\sigma^2 = \sigma_1^2 + \dots + \sigma_n^2$ . La desviación estándar resulta  $\sigma = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$ .

**Corolario 6.2.1** Si  $X_1, \dots, X_n$  son una muestra aleatoria de variables normales con media  $\mu$  y varianza  $\sigma^2$ , el promedio muestral o media aritmética  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  es una variable normal con media  $\mu$  y varianza  $\sigma_{\bar{X}}^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n} \sigma^2$ . La desviación estándar resulta  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ .

Esta proposición es muy importante, pues me dice que la distribución de la suma, y de allí de la media muestral (conocida también como promedio o media aritmética) de una muestra, es también normal. La media muestral de los datos es un estimador de la tendencia central de los datos, una forma de resumir la información, que es crucial cuando los datos obtenidos son pequeñas variaciones alrededor de un valor desconocido. En el caso en que la distribución tome un rango de valores muy grande con alta probabilidad, la media poblacional ya no es el único parámetro representativo, la varianza o la desviación estándar son necesarios para describir la densidad de probabilidad. Esta proposición dice que si tengo una muestra de variables normales no necesariamente idénticas pero si independientes, la media y la varianza de la suma de variables van a ser la suma de medias y varianzas.

En el caso del promedio de una muestra, también es una variable normal, y su media es la misma que la de la muestra, pero la varianza queda dividida por  $n$ , el tamaño de muestra. El promedio de una muestra entonces tiene la misma media pero una dispersión más chica, es decir, se pega más la curva al centro de la distribución, la media. Esto es muy importante, pues nos dice que la distribución del muestreo es cada vez más concentrada a medida que aumentamos el tamaño de la muestra.

**Ejemplo 6.2.4** Una compañía embotelladora usa una máquina para llenar botellas plásticas con su gaseosa sabor naranja. Las botellas deberían contener 500 mililitros, pero en realidad los contenidos varían de acuerdo con una distribución normal con media  $\mu = 498 \text{ ml}$  y una desviación standard de  $3 \text{ ml}$ .

1. ¿Cuál es la probabilidad de que una botella individual contenga menos de 495 ml?
2. ¿Cuál es la probabilidad de que el contenido medio de las botellas de un pack de 6 sea menos de 495?

### Resolución

Si elijo una botella rellena por la máquina, su contenido es una variable aleatoria  $X$  con distribución normal de media  $\mu = 498 \text{ ml}$  y una desviación standard de  $3 \text{ ml}$ . Por lo cual

$$P(X < 495) = P\left(\frac{X - \mu}{\sigma} < \frac{495 - 498}{3}\right) = P(Z < -1) = 1 - P(Z < 1) = 1 - 0,8413 = 0,1581$$

Si ahora considero un pack de seis botellas, y supongo que el llenado de las botellas es independiente, el contenido medio de las botellas es una variable aleatoria normal también,  $\bar{X} = \frac{1}{6} \sum_{i=1}^6 X_i$ , con media  $\mu_{\bar{X}} = \frac{1}{6} \sum_{i=1}^6 498 = 498$  y desviación estándar  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{6}}$ . Entonces

$$P(\bar{X} < 498) = P\left(\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} < \frac{495 - 498}{\frac{3}{\sqrt{6}}}\right) = P(Z < -\sqrt{6}) = 1 - P(Z < \sqrt{6}) = 1 - P(Z < 2,45) = 1 - 0,9929 = 0,0071$$

**Ejemplo 6.2.5** La altura de cierta población se distribuye normalmente con media 170 cm y desvío estándar 5 cm.

1. Si elegimos una persona al azar de esa población, ¿qué altura se espera que tenga?
2. Si tomamos una muestra aleatoria simple de 7 personas ¿Cuál es la probabilidad de que la mayoría mida menos de 174 cm?
3. Si seleccionamos personas hasta encontrar la primera persona que mida más de 185 cm ¿Cuántas personas se esperan seleccionar?

### Resolución

La altura esperada de una persona perteneciente a una población con distribución normal es el parámetro de centralidad de la normal,  $\mu$ , la media. En este caso es  $\mu = 170$  cm.

Supongamos primero que la población es muy grande, por lo cual al elegir solo 7 personas, la diferencia entre muestrear con reposición (que implica independencia entre elecciones) y muestrear sin reposición es casi nula, y tenemos una muestra aleatoria simple de 7 variables normales independientes con los mismos parámetros de la población.

Por otra parte, la noción de mayoría en 7 valores implica contar el número de éxitos mayor o igual a 4, donde éxito es la probabilidad de que una persona elegida mida menos de 174 cm. Calculemos esta probabilidad usando la curva normal

$$p = P(X < 174) = P\left(\frac{X - \mu}{\sigma} < \frac{174 - 170}{5}\right) = P(Z < 0,8) = 0,7881$$

Entonces, si  $Y$  es una variable binomial que mide el número de éxitos en 7 repeticiones del experimento independientes (la independencia ya la discutimos antes), resulta

$$P(Y \geq 4) = \binom{7}{0} 0,2119^7 + \binom{7}{1} 0,7881 0,2119^6 + \binom{7}{2} 0,7881^2 0,2119^5 + \binom{7}{3} 0,7881^3 0,2119^4 + \binom{7}{4} 0,7881^4 0,2119^3$$

Si queremos saber cuantas personas seleccionar hasta obtener alguien con altura mayor a 185 cm, debemos considerar una variable auxiliar geométrica que mida el número de repeticiones independientes hasta llegar al primer éxito. Aquí éxito será que la persona mida más de 185cm, y la probabilidad de éxito se calcula como área bajo la curva normal.

$$P(X > 185) = 1 - P(X \leq 185) = 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{185 - 170}{5}\right) = 1 - P(Z < 3) = 1 - 0,9987 = 0,0013$$

Una vez más aquí es crucial que el muestreo aleatorio simple produzca variables independientes, y eso es razonable suponerlo cuando la diferencia entre el tamaño de la población y el de la muestra es muy grande. Si esto es así, sea  $Z$  la variable geométrica auxiliar que mide el número de repeticiones independientes hasta llegar al primer éxito, la esperanza de  $Z$  es el número esperado de repeticiones hasta elegir una persona que mida mas de 185cm,

$$E(Z) = \frac{1}{0,0013} = 768$$

más de 768 personas.



**Proposición 6.2.3** Sea  $X_1, \dots, X_n$  variables aleatorias continuas independientes e idénticamente distribuidas. Definimos las siguientes variables aleatorias:

1.  $K = \min(X_1, \dots, X_n)$  el mínimo de la muestra.
2.  $M = \max(X_1, \dots, X_n)$  el máximo de la muestra.

Entonces

1. la función densidad de  $M$  es

$$f_M(x) = n(F(x))^{n-1}f(x)$$

2. la función densidad de  $K$  está dada por

$$f_K(x) = n(1 - F(x))^{n-1}f(x)$$

### Demostración:

Observemos que si el máximo de la muestra es menor que  $z$ , entonces todos los valores de la muestra deben ser menores a  $z$ , por lo cual

$$P(M \leq z) = P(X_1 \leq z, \dots, X_n \leq z) = P(X_1 \leq z) \dots P(X_n \leq z) = F(z)^n$$

Diferenciando esta ecuación obtenemos que

$$f_M(z) = n(F(z))^{n-1}F'(z) = n(F(z))^{n-1}f(z)$$

En el caso del mínimo, si el mínimo de la muestra es mayor a  $z$ , entonces todos los valores deben ser mayores a  $z$ , por lo cual

$$P(K \leq z) = 1 - P(K > z) = 1 - P(X_1 > z, \dots, X_n > z) = 1 - P(X_1 > z) \dots P(X_n > z) = 1 - [1 - F(z)]^n$$

Diferenciando esta ecuación obtenemos que

$$f_K(z) = n[1 - F(z)]^{n-1}F'(z) = n[1 - F(z)]^{n-1}f(z)$$

**Ejemplo 6.2.6** Si tenemos 5 variables aleatorias independientes e idénticamente distribuidas con distribución  $N(12, 4)$ .

1. ¿Cuál es la probabilidad de que  $Y = X_1 + X_2 + X_3 + X_4 + X_5$  sea menor que 61?
2. ¿Cuál es la probabilidad de que el mínimo de la muestra sea menor que 10?
3. ¿Cuál es la probabilidad de que el máximo de la muestra exceda 15?

### Resolución

La variable  $Y = X_1 + X_2 + X_3 + X_4 + X_5$  tiene distribución normal con media  $\mu_Y = 5 * 12 = 60$  y varianza  $\sigma_Y^2 = 5 * 4 = 20$ , por lo cual la probabilidad de que  $Y$  sea menor a 61 se calcula estandarizando la variable.

$$P(Y < 61) = P((Y - \mu_Y)/\sigma_Y < (61 - \mu_Y)/\sigma_Y) = P(Z < (61 - 60)/\sqrt{20}) = P(Z < 0,22) = 1 - 0,4129 = 0,5871$$

La probabilidad de que  $K = \min(X_1, X_2, X_3, X_4, X_5)$  sea menor a 10 es

$$P(K < 10) = P(K \leq 10) = F_K(10) = 1 - [1 - F(10)]^5$$

Ahora, observemos que si  $F$  es la función de distribución acumulada de una normal con parámetros  $\mu$  y  $\sigma$ , entonces

$$F(x) = P(X \leq x) = P((X - \mu)/\sigma \leq (x - \mu)/\sigma) = P(Z \leq (x - \mu)/\sigma) = \Phi((x - \mu)/\sigma)$$

Por lo cual

$$P(K < 10) = 1 - [1 - F(10)]^5 = 1 - [1 - \Phi((10 - 12)/\sqrt{4})]^5 = 1 - [1 - \Phi(-0,5)]^5 = 1 - (1 - 0,3085)^5 = 0,7233$$

La probabilidad de que  $M = \max(X_1, X_2, X_3, X_4, X_5)$  sea mayor a 15 es

$$P(M > 15) = P(M \geq 15) = 1 - F_M(15) = 1 - F(15)^5 = 1 - \Phi((15 - 12)/\sqrt{4})^5 = 1 - \Phi(0,75)^5 = 1 - (0,7734)^5 = 0,7233$$

**Ejemplo 6.2.7** Se diseña un ascensor de carga cuyo límite es 1000 kg. El peso de cada caja sigue una distribución normal con un peso medio de 32 kg y un desvío estándar de 10 kg.

1. ¿Cuál es la probabilidad de que un grupo de 30 cajas exceda el límite de carga?
2. Se toma una muestra aleatoria simple de 5 cajas ¿cuál es la probabilidad de que el mínimo de la muestra sea superior a 30kg e inferior a 33kg?

### Resolución

Para que el grupo exceda el límite de carga la suma de los pesos de las cajas debe exceder 1000kg. Como el peso de cada caja es una variable normal independiente con los mismos parámetros  $\mu$  y  $\sigma^2$ , la variable  $X = \sum_{i=1}^{30} X_i$  también es normal con parámetros  $\mu_X = 30 \times \mu = 960$  y  $\sigma_X^2 = 30 \times \sigma^2 = 3000$ . Entonces estandarizando calculamos la probabilidad

$$P(X > 1000) = P((X - \mu_X)/\sigma_X > (1000 - \mu_X)/\sigma_X) = P(Z > (1000 - 960)/54,7723) = P(Z > 0,7303) = 0,2326$$

Si las variables  $X_1, X_2, X_3, X_4, X_5$  representan el peso de las cinco cajas elegidas al azar, y  $K = \min(X_1, X_2, X_3, X_4, X_5)$  entonces

$$\begin{aligned} P(30 < K < 33) &= F_K(33) - F_K(30) \\ &= 1 - [1 - F(33)]^5 - (1 - [1 - F(30)]^5) \\ &= [1 - \Phi((30 - \mu)/\sigma)]^5 - [1 - \Phi((33 - \mu)/\sigma)]^5 \\ &= (1 - \Phi(-0,2))^5 - (1 - \Phi(0,1))^5 \\ &= (1 - 0,4920)^5 - (1 - 0,5040)^5 \\ &= 0,0038 \end{aligned}$$

## 6.3. Aproximaciones

**Ejemplo 6.3.1** La luz viaja muy rápido, pero no se transmite en forma instantánea. La luz tarda casi un segundo en alcanzarnos desde la luna, y mas de 100 millones de años en alcanzarnos desde los objetos más distantes que han sido observados en el universo. Como las ondas de radio y el radar también se transportan a velocidad de la luz, un valor cierto de esa velocidad es muy importante para poder comunicarse con los satélites en órbita y las sondas enviadas al espacio exterior. También, un valor preciso de la rapidez de la luz es necesaria para el diseño de computadoras, dado que la electricidad se transporta a la velocidad de la luz.

Las primeras mediciones reales de la rapidez de la luz se hicieron hace más de 100 años atrás por A.A Michelson y Simon Newcomb. En la tabla siguiente podemos observar 66 mediciones realizadas por Newcomb entre Julio y Septiembre de 1882.

Para poder interpretar esos datos, debemos entender como realizó Newcomb esas mediciones. El midió cuanto tiempo tardó un haz de luz en recorrer la distancia entre su laboratorio cerca del rio Potomac y un espejo ubicado en la base del monumento a Washington, una distancia total de 7400 metros. El instrumento con que midió el tiempo era un complicado aparato diseñado por Newcomb, que nos es muy difícil de juzgar. Pero podemos aceptar que los físicos de la época consideraron que era apropiado para realizar la tarea y más preciso que cualquier otro instrumento conocido entonces.

Observemos que al ser estos datos producto de diferentes experiencias de medición, que podemos suponer independientes, nos encontramos frente a observaciones de una muestra aleatoria simple, es decir, cada dato es la observación de una variable aleatoria, y todas ellas son independientes y tienen la misma distribución (desconocida). Este es un modelo de medición, es decir, debería haber un solo valor posible, el del tiempo de paso de la luz. Pero en las mediciones físicas se pueden sumar tantos errores involuntarios que lo único que puede esperar el científico es

28	26	33	24	34	-44
27	16	40	-2	29	22
24	21	25	30	23	29
31	19	24	20	36	32
36	28	25	21	28	29
37	25	28	26	30	32
36	26	30	22	36	23
27	27	28	27	31	27
26	33	26	32	32	24
39	28	24	25	32	25
29	27	28	29	16	23

Cuadro 6.1: Mediciones de Newcomb del tiempo que tarda la luz en recorrer 7400 metros, medida en segundo  $\times 10^{-9}$ .

que la distribución de estos datos este centrado en el verdadero valor (desconocido) y que la dispersion sea pequeña. Miremos estos datos para ver si estamos en frente a un modelo de medición

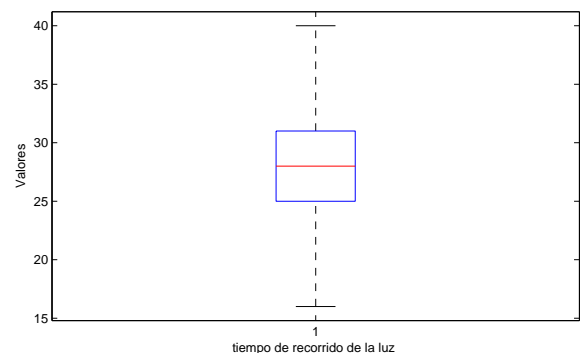
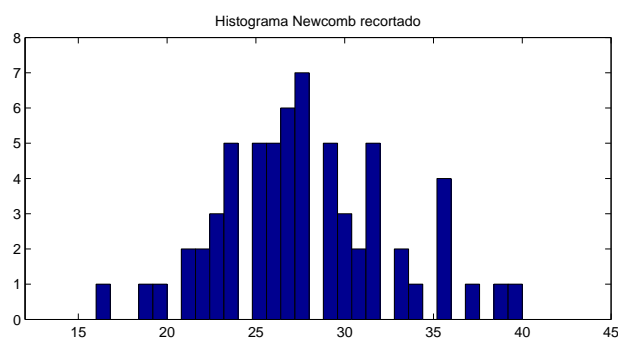
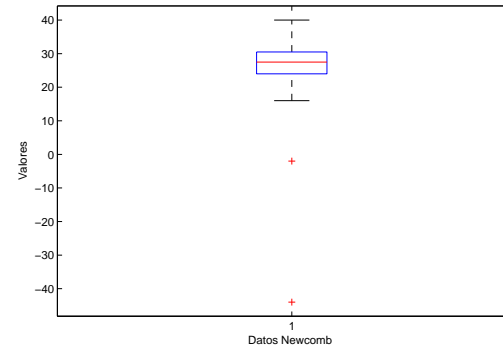
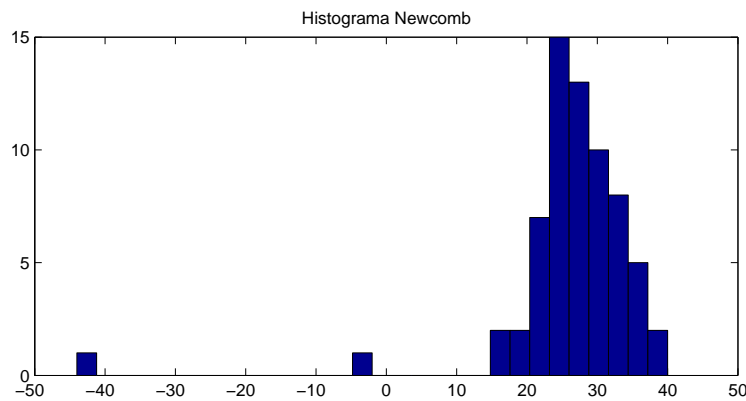
$$\mu + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Para mirar los datos vamos a realizar dos gráficos clásicos, un histograma y una diagrama de caja. El histograma cuenta los datos que caen en intervalos fijos, en este caso se dividió el rango de la variable en 30 intervalos iguales. La altura de cada rectángulo es proporcional a la cantidad de datos en el intervalo de base. El diagrama de caja muestra una caja con la mediana, (la raya central), el rango intercuartílico (los bordes de la caja) y los datos que están muy fuera de lo que debería ser el rango de la variable (las cruces que pueden verse en el gráfico en su parte más baja). La mediana es el valor del rango de la variable que deja igual cantidad de datos a la izquierda y a la derecha de él. El diagrama de caja en vez de mostrar la media de los datos muestra la mediana, que no se influencia por valores atípicos, y en este caso hay dos valores muy raros, -44 y -2, que son las cruces que se ven en el diagrama de caja. Son valores que no deberían corresponder a tiempos de paso, pues son negativos, en estos casos es de esperarse que el experimento fracasó. En ambos gráficos se ven los datos atípicos. Si quisiéramos estimar  $\mu$  con el promedio de los datos, estos datos atípicos nos alejarían del centro del “grueso” de los datos. La mediana marcada en el diagrama de caja con una raya roja, no se influencia por esos datos.

Newcomb sabía que por más cuidadoso que se sea, los experimentos, al ser independientes, varían. El entorno no es siempre el mismo, el aparato cambia un poco con la temperatura, la densidad de la atmósfera cambia todos los días, etc. También hay que considerar que Newcomb fue aprendiendo con cada medición y las últimas realizadas tenían mayor precisión que las primeras. Newcomb hizo lo posible para reducir las fuentes de variación, y es por eso que estas mediciones son mucho más precisas que las realizadas por científicos de su época menos hábiles, (o cuidadosos). Pero por sobre todo se tomaron 66 mediciones, a pesar de lo difícil que era hacerlas, porque Newcomb sabía que el promedio no depende de la temperatura del ambiente o la densidad atmosférica.

Si decidimos sacar esos datos de la muestra esos datos negativos por ser experimentos fallidos, obtenemos los siguientes histogramas y diagramas de caja. Observemos que el diagrama de caja tiene en su escala vertical el rango de los datos, no nos dice que frecuencia tienen los datos, sino cuán centrada esta la mediana, cuán simétrica es la distribución. En cambio el histograma muestra las frecuencias de los datos por intervalo.

Sin estos datos fallidos, es razonable pensar que el modelo de medición normal ajusta. Hay dos gráficos que se realizan para observar esto, uno de ellos es el qqplot, un diagrama que muestra como se comparan los datos por cuartos de distribución con la normal, y el histograma que tiene la curva normal sobre escrita. La curva normal que esta ajustada es la que tiene a la media y la varianza estimadas por la media muestral  $\bar{X}$ , y la varianza muestral  $S^2$  que definimos en la proposición 6.2.1. Los cuantiles de una distribución son los puntos que evaluados en la función de distribución acumulada dan valores claves, como 0.25, 0.5 0.75. Estos cuantiles llevan en particular el nombre de cuantiles (dividen el area bajo la curva en cuartos). El rango intercuartílico está marcado en el diagrama de caja por los bordes externos de la caja. Adentro de la caja están el 50 % de los datos. El qq plot muestra un dibujo de los cuantiles



de las dos distribuciones, la normal y la de los datos, una en cada eje. Si la distribución de los datos es normal, el gráfico es muy cercano a una línea.

El qq plot de las 64 observaciones de Newcomb muestra que la distribución es muy cercana a la normal. La distribución de frecuencias relativas de las mediciones repetidas, que puede ser observada en el histograma, es cercana a una normal, como muestra la curva normal estimada con la media y la varianza muestral. La desviación estándar es  $s = 4,8985$ , y la media es  $\bar{X} = 27,93$ . ¿Podemos decir entonces que el tiempo  $\mu$  que deseaba medir Newcomb es 27,93? Fue la medición hecha en 1882 correcta?. ¿En que ayuda que los datos sean aproximadamente normales?

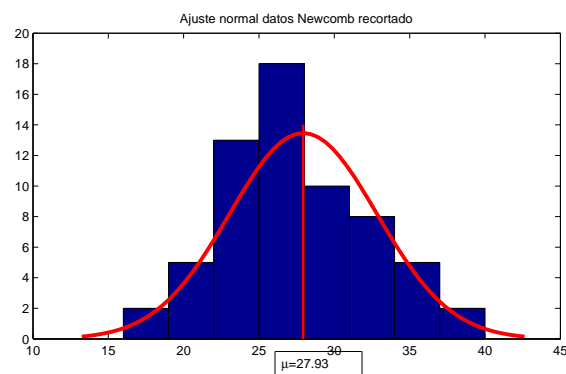
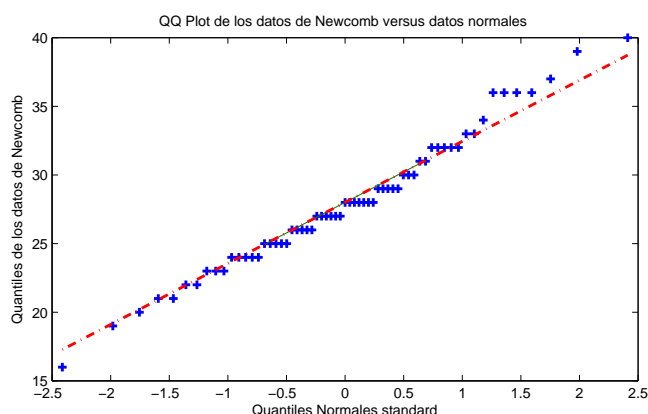
Hay tres preguntas y dos respuestas. Una, al promediar 64 datos, estoy reduciendo la variabilidad de los experimentos, pero si estoy realizando siempre el mismo error, un error sistemático, va a seguir presente, y no debemos olvidar que estas mediciones se realizaron en 1882. Mediciones actuales muestran que el verdadero valor debería rondar los 33 (medido en billonésima parte de un segundo). Ese problema no es estadístico sino físico.

También, dar como respuesta una medición sola es arriesgado, aun cuando sea un promedio de 64 mediciones diferentes. Es más razonable dar un intervalo de valores posibles, junto con la probabilidad de que el verdadero valor esté en ese intervalo. Aqué donde aparece la normal, el intervalo puede calcularse en forma simple, y depende de la media y varianza muestrales. Además, esto también dice que el modelo de medición es correcto. El modelo de medición supone que hay un centro, el verdadero valor, y errores que son suma de errores infinitesimales, que al sumarse, por el teorema central del límite, siguen la distribución normal. Por lo cual debemos “ver” la normal en los gráficos de los datos, y podemos tener confianza en el centro de los datos.

Otra observación, existen formas mucho más objetivas que mirar gráficos para decidir si la distribución normal ajusta o no los datos. Los test de hipótesis de “bondad de ajuste” realizan esta tarea por nosotros. Sin embargo, un buen estadístico siempre mira los datos, aún los que son multidimensionales.

## 6.4. Ley de los grandes números

Lo que necesitamos conocer ahora es alguna forma de decir cuán lejos está la estimación que estoy realizando del verdadero valor. Una medida del error que cometo. Esto no debería ser dependiente de la



muestra realizada sino del tipo de muestra realizada, es decir, la apreciación del error debería ser una probabilidad, y esta probabilidad debería estar basada en la distribución de todas las muestras del mismo tamaño, realizadas con el mismo procedimiento muestral.

**Definición 6.4.1** Se llama error a la diferencia entre el verdadero valor del parámetro y su estimación.

Formas más particulares de error serán definidas en el capítulo de estimación, como el error cuadrático medio, y los errores estándares. En esta sección vamos a mostrar que si  $X$  es una muestra aleatoria de tamaño  $n$  una distribución con media  $\mu$  y varianza  $\sigma^2$ , la probabilidad de que la diferencia entre  $\bar{X}$  y  $\mu$  sea mayor a  $4\sigma$  es menor o igual a  $1/n16$ . Esto es, vamos a descubrir una cota para la probabilidad de cometer un error mayor a  $4\sigma$  al reemplazar a  $\mu$  por  $\bar{X}$ . Y esta cota se va a achicar a medida que aumente el tamaño de la muestra.

$$P(|\bar{X}_n - \mu| \geq 4\sigma) \leq \frac{\text{Var}(\bar{X}_n)}{(4\sigma)^2} = \frac{\sigma^2}{n(4\sigma)^2} = \frac{1}{n16}$$

Debemos notar que no hice ninguna suposición sobre la distribución de los datos, solo pedí que la muestra sea una muestra de la misma distribución, es decir, que todas las variables que componen la muestra sean copias independientes de una misma variable. Esto ocurre cuando repito un mismo experimento tratando de no cambiar como lo hago, como el experimento de Newcomb, o bien cuando realizo una muestreo aleatorio simple de objetos de una población, y la tasa de muestreo  $n/N$ , tamaño de la muestra dividido tamaño de la población es menor a 0.05.

### 6.4.1. Desigualdad de Chebyshev

**Lema 6.4.1** Sea  $X$  una variable aleatoria positiva con media  $\mu$ , y sea  $t$  un número real positivo. Entonces

$$P(X \geq t) \leq \frac{E(X)}{t}$$

**Demostración:**

Definamos la variable aleatoria  $Y$  como

$$Y = 0 \quad \text{si } X < t \quad Y = t \quad \text{si } X \geq t$$

Entonces  $Y$  es una variable aleatoria discreta con densidad  $p_Y$  dada por

$$p_Y(0) = P(Y = 0) = P(X < t) \quad p_Y(t) = P(Y = t) = P(X \geq t)$$

por lo cual

$$E(Y) = tP(Y = t) + 0P(Y = 0) = tP(X \geq t)$$

Como  $X$  es una variable positiva, de la definición de  $Y$  resulta que  $X \geq Y$ . Entonces

$$E(X) \geq E(Y) = tP(X \geq t)$$

y resulta

$$P(X \geq t) \leq \frac{E(X)}{t}$$

#### Proposición 6.4.1 Desigualdad de Chebyshev

Sea  $X$  una variable aleatoria con media  $\mu$  y varianza  $\sigma^2$ . Entonces, para cada  $t > 0$

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

#### Demostración:

Apliquemos el lema anterior a la variable  $(X - \mu)^2$  y al número  $t^2$ . Entonces

$$P((X - \mu)^2 \geq t^2) \leq \frac{E((X - \mu)^2)}{t^2} = \frac{\sigma^2}{t^2}$$

Como  $(X - \mu)^2 \geq t^2$  si y solamente si  $|X - \mu| \geq t$  vale la proposición.

La proposición anterior dice que si  $\sigma^2$  es muy pequeño, existe una probabilidad muy alta de que  $X$  no se desvíe mucho de  $\mu$ . Podemos realizar una cota en función de la varianza.

**Corolario 6.4.1** Sea  $X$  una variable aleatoria con media  $\mu$  y varianza  $\sigma^2$ . Entonces,

$$\begin{aligned} P(|X - \mu| \geq k\sigma) &\leq \frac{1}{k^2} \\ P(|X - \mu| < k\sigma) &\geq 1 - \frac{1}{k^2} \end{aligned}$$

#### Demostración:

Directo de la desigualdad de Chebyshev, reemplazando  $t$  por  $k\sigma$ .

**Ejemplo 6.4.1** Si  $X$  una variable aleatoria con media  $\mu$  y varianza  $\sigma^2$ , la probabilidad de que la diferencia entre  $X$  y  $\mu$  sea mayor a  $4\sigma$  es menor o igual a  $1/16$ . **Y estos resultados valen para cualquier distribución con media y varianza finita.** Para distribuciones especiales, esta cota es grosera, hay cotas mucho mas ajustadas pero lo que impacta es su generalidad.

**Ejemplo 6.4.2** Un fabricante de tornillos sabe que el 5 % de su producción es defectuosa. Por ello garantiza que si una entrega de 10000 piezas tiene más de  $a$  defectuosos se reembolsará el dinero pagado por el cargamento. ¿Cuán chico puede el fabricante escoger el valor  $a$  para que no deba devolver el dinero del cargamento más del 1 % de las veces?

#### Resolución

Supongamos que el lote de 10000 tornillos es una muestra aleatoria de la población (infinita) de tornillos que produce la fábrica. Sea  $p = 0,05$  la probabilidad de elegir un tornillo defectuoso de la población. Entonces  $X$ , el número de

defectuosos del lote de 10000 tornillos, tiene distribución Binomial con media  $np$  y varianza  $np(1-p)$ , y el evento  $A = \text{devolver el cargamento}$  puede ser pensado como

$$A = (X > a) = (X - np > a - np)$$

Queremos escoger  $a$  tal que

$$P(\text{devolver el cargamento}) = P(X > a) \leq 0,01$$

Sabemos por Chebyshev que

$$P(|X - np| > t) \leq \frac{np(1-p)}{t^2} < 0,01 \text{ si } t \geq \sqrt{\frac{10000 \times 0,0475}{0,01}}$$

Entonces  $t \geq 218$  garantiza que

$$P(A) = P(X - np > a - np) \leq P(|X - np| > t) \leq 0,01$$

por lo cual  $a - np = t \geq 218$  implica que  $a = 218 + np = 218 + 500 = 718$  es un valor conservativo para  $a$ .

Un ejercicio interesante sería tratar de calcular

$$P(X > a)$$

usando la distribución binomial, o mejor, la aproximación Poisson. La ecuación resulta bastante complicada. En la próxima sección aproximaremos la distribución Binomial con una distribución Normal, y calcularemos el valor de  $a$  para compararlo.

**Ejemplo 6.4.3** Una sección de bosque de pinos tienen un número de árboles enfermos por acre,  $Y$ , que sigue una distribución de Poisson con media  $\lambda = 10$ . Los árboles enfermos son rociados con un insecticida que cuesta \$3.00 por árbol, más un costo fijo de alquiler del equipamiento, de \$50. Sea  $C$  el costo del rociado de los árboles de un acre seleccionado aleatoriamente.

1. Encuentre la esperanza y varianza de  $C$ .
2. Encuentre un intervalo que acote a  $C$  con probabilidad al menos 0.75.

#### Resolución

La esperanza y varianza de  $C = 3Y + 50$  son

$$E(C) = 3E(Y) + 50 = 3\lambda + 50 = 80 \quad \text{Var}(C) = 9\text{Var}(Y) = 9\lambda = 90$$

Aplicando la desigualdad de Chebyshev

$$P(|C - E(C)| < k\sqrt{\text{Var}(C)}) \geq 1 - \frac{1}{k^2} = 0,75$$

por lo cual si  $1 - \frac{1}{k^2} = 0,75$ , entonces  $k = 2$  y

$$P(E(C) - k\sqrt{\text{Var}(C)} \leq C \leq kE(C) + \sqrt{\text{Var}(C)}) = P(80 - 2 \times 9,48 \leq C \leq 80 + 2 \times 9,48) = P(61,04 \leq C \leq 98,96) \geq 0,75$$

El intervalo que acota a  $C$  con probabilidad al menos 0.75 es entonces  $[61,04, 98,96]$ . Esto es, de 100 veces que calcule  $C$  en forma independiente, 75 veces o más va a estar en este intervalo, que depende solo del tamaño de la muestra y de los parámetros de la distribución de  $C$ . Me dá una idea de qué gastos espero tener por acre.

Otro corolario de la desigualdad de Chebyshev es el hecho de que si la varianza es nula, la distribución de la variable está concentrada en la media.

**Corolario 6.4.2** Si  $\text{Var}(X) = 0$ , entonces  $P(X = \mu) = 1$ .

**Demostración:**

Daremos una prueba por contradicción. Supongamos que  $P(X = \mu) < 1$ . Entonces, para algún  $t > 0$   $P(|X - \mu| \geq t) > 0$ . Pero por la desigualdad de Chebyshev, para todo  $t > 0$

$$P(|X - \mu| \geq t) = 0$$

lo cual es absurdo. El absurdo vino de suponer que  $P(X = \mu) < 1$ , por lo cual

$$P(X = \mu) = 1$$

### 6.4.2. Ley de los grandes números

Hemos dicho al comienzo del curso que si una moneda honesta se tira muchas veces, la proporción de caras obtenida estará cerca de  $1/2$ . La ley de los grandes números formaliza esta afirmación empírica. El siguiente gráfico corresponde a 10000 tiradas de una moneda equilibrada. Para  $N$ =cantidad de tiradas, hemos graficado la cantidad de caras obtenidas menos la cantidad de números obtenidas en las primeras  $N$  tiradas. Como observamos, la diferencia entre la cantidad de caras y la cantidad de números no es necesariamente cero.

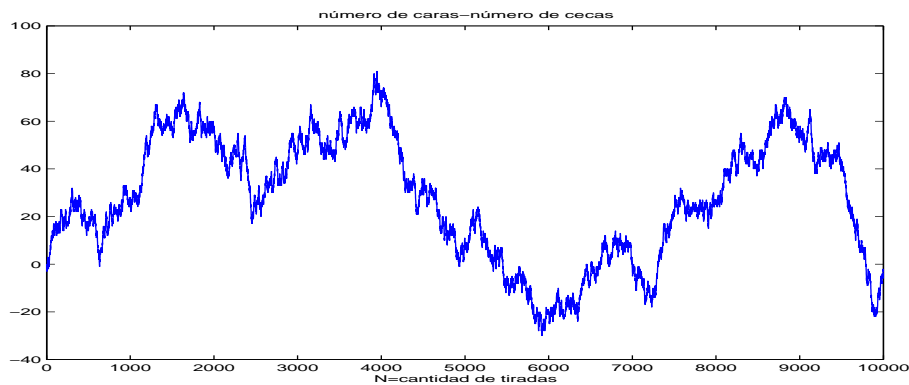


Figura 6.1: Gráfico del número de caras menos el de números en función de  $N$ , la cantidad de tiradas de la moneda

Sin embargo, si graficamos la frecuencia relativa, o sea que para cada  $N$  graficamos la proporción de caras en las primeras  $N$  tiradas tenemos que a medida que aumentamos la cantidad de tiradas, la frecuencia se acerca a  $1/2$ .

La ley de los grandes números formaliza este concepto. Los resultados de los tiros consecutivos de la moneda son variables aleatorias independientes  $X_i$  que toman valores 0 o 1 según la  $i$ -ésima repetición es cara o número, y la proporción de caras en los  $n$  tiros es

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

**Definición 6.4.2** Sea  $\{X_n\}$  una sucesión de variables aleatorias. Se dice que  $\{X_n\}$  converge en probabilidad a  $Y$  si para todo  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_n - Y| \geq \epsilon) = 0$$



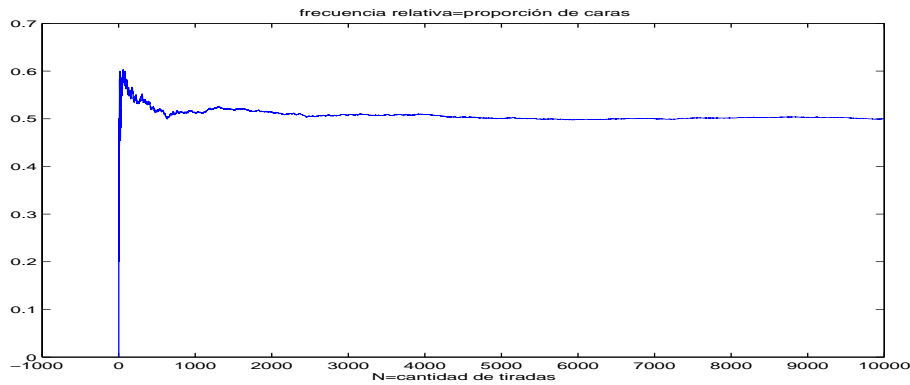


Figura 6.2: Gráfico de la proporción de caras en función de  $N$ , la cantidad de tiradas de la moneda

### Corolario 6.4.3 Ley débil de los grandes números

Sea  $X_1, \dots, X_n$  variables aleatorias independientes con  $E(X_i) = \mu$  y  $Var(X_i) = \sigma^2$ . Sea

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Entonces para cada  $\epsilon > 0$

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

$$P(|\bar{X}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$

Por lo cual

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0 \quad \lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$$

y  $\{\bar{X}_n\}$  converge en probabilidad a la variable aleatoria concentrada en  $\mu$ .

#### Demostración:

Primero encontremos  $E(\bar{X}_n)$  y  $Var(\bar{X}_n)$ .

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$$

Como los  $X_i$  son independientes

$$Var(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{\sigma^2}{n}$$

por lo cual, aplicando la desigualdad de Chebyshev, resulta

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \quad n \rightarrow \infty$$

La otra desigualdad es la cota resultante al calcular la probabilidad del complemento del evento  $|\bar{X}_n - \mu| \geq \epsilon$

**Observación 6.4.1** Es importante observar que la noción de convergencia usual en cálculo  $a_n \rightarrow a$  implica que  $a_n$  se transforma cuando  $n$  varía y ser acerca a  $a$  tanto como uno quiera, siempre que  $n$  sea suficientemente grande. En cambio la convergencia enunciada en la ley de los grandes números implica que si  $X_n \rightarrow X$  en probabilidad, entonces la probabilidad del suceso  $(|X_n - X| < \epsilon)$  puede hacerse arbitrariamente próxima a 1 tomando  $n$  suficientemente grande.

### 6.4.3. Proporciones

Sea  $A$  un suceso con  $P(A) = p$ , que es uno de los resultados posibles de un experimento. Supongamos que repetimos el experimento  $n$  veces en forma independiente y consideramos las variables aleatorias indicadoras  $X_i$  que valen 1 si el  $i$ -ésimo experimento resulta en  $A$  y cero si no. Estas variables resultan Bernoulli independientes y el promedio

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

es la frecuencia relativa del evento  $A$  en las  $n$  repeticiones. Como  $E(X_i) = p$ , y  $Var(X_i) = p(1-p)$ , por la ley de los grandes números

$$\hat{p}_n \xrightarrow[n \rightarrow \infty]{} p$$

esto es, la frecuencia relativa del evento  $A$  converge débilmente (en probabilidad) a la probabilidad del evento  $A$ . Por lo cual si no se conoce la probabilidad de un evento particular, estudiar la frecuencia relativa para tamaños grandes de  $n$  permite inferir este valor, esto es, se puede reemplazar  $p$  por  $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , sabiendo que el error  $\epsilon$  que se comete al hacerlo es del orden de  $\tau = 1/4n\epsilon^2$

$$P(|\hat{p}_n - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2} = \tau$$

usando que  $p(1-p) \leq 1/4$  si  $0 < p < 1$  (dado que la función  $p(1-p)$  tiene un máximo en  $p = 1/2$ ).

Supongamos que está fijo el máximo de tolerancia que podemos tener para el error, es decir que el error  $\epsilon$  y la tolerancia  $\tau$  han sido dados. ¿Cuál debería ser el mínimo número de experimentos que debo realizar para garantizar que la probabilidad de que  $\hat{p}_n$  difiera de  $p$  en más de  $\epsilon$  sea menor que  $\tau$ ?

$$P(|\hat{p}_n - p| \geq \epsilon) \leq \tau$$

Observemos que

$$\frac{1}{4n\epsilon^2} \leq \tau \Rightarrow \frac{1}{4\epsilon^2\tau} \leq n$$

Estas cotas son un poco holgadas, en realidad, en muchos casos, la mitad de esta cantidad es suficiente.

**Ejemplo 6.4.4** Se decide tomar una muestra de  $n$  personas para evaluar el porcentaje de ciudadanos dispuestos a apoyar la candidatura del Dr. Márquez. Sea  $p$  la proporción real de votantes de ese candidato en la población. Definimos la variable aleatoria proporción muestral  $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , donde

$$X_i = \begin{cases} 1 & \text{si el } i\text{-ésimo ciudadano entrevistado votará al candidato} \\ 0 & \text{en caso contrario} \end{cases}$$

- Demostrar que  $\hat{p}_n \xrightarrow[n \rightarrow \infty]{} p$  en probabilidad.
- Acotar  $P(|\hat{p}_{900} - p| \geq 0,025)$ .
- Calcular un valor de  $n$  que garantice que  $P(|\hat{p}_n - p| \geq 0,025) \leq 0,01$ .

#### Resolución

Observemos que  $\hat{p}_n$  es la frecuencia relativa del evento  $A = \text{"voto por el candidato Márquez"}$ . Por lo cual la ley de los grandes números dice que

$$\lim_{n \rightarrow \infty} P(|\hat{p}_n - p| \geq \epsilon) = 0$$

y  $\hat{p}_n \rightarrow p$  en probabilidad.

Para calcular la tolerancia máxima al error  $\epsilon = 0,025$  que se obtiene con  $n = 900$ , planteamos la cota que se obtiene en la ley de los grandes números.

$$P(|\hat{p}_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$$

acotando  $p(1-p)$  por  $1/4$ . Entonces  $\frac{1}{4n\epsilon^2} = \frac{1}{4 \times 900 \times 0,025^2} = 0,44$ .

Si la tolerancia  $\tau$  y el error  $\epsilon$  han sido dados, para que las cotas sean alcanzadas debemos calcular el número de repeticiones necesarias para ello.

$$P(|\hat{p}_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2} < \tau$$

implica que  $\frac{1}{4n\epsilon^2} < \tau$  por lo cual, si  $\epsilon = 0,025$  y  $\tau = 0,01$  resulta  $\frac{1}{4n(0,025)^2} < 0,01$  por lo cual  $n > \frac{1}{4 \times 0,01 \times (0,025)^2} = 40000$ . Esta cota es muy holgada. Con este tamaño de  $n$  nos aseguramos tener un error y una tolerancia al error chica, pero es muy costoso realizar una encuesta con este tamaño de muestra. Si conociéramos la distribución del estimador  $\hat{p}_n$  podríamos calcular la tolerancia

$$P(|\hat{p}_n - p| \geq \epsilon)$$

de forma más precisa.

**Ejemplo 6.4.5** Tiramos un dado 20 veces ¿Cuál es la probabilidad de que la frecuencia relativa  $\hat{p}_n$  de veces que aparece el cuatro difiere de la probabilidad real en menos de 0.1?

**Resolución**

La ley de los grandes números afirma que

$$P(|\hat{p}_n - p| < \epsilon) \geq 1 - \frac{p(1-p)}{n\epsilon^2}$$

En este caso sabemos que la verdadera probabilidad  $p$  es  $1/6$ , por lo cual no debemos acotar  $p(1-p)$  por  $1/4$  sino reemplazarlo por su verdadero valor,  $p(1-p) = 1/6(1-1/6) = 0,14$ . Como  $\epsilon = 0,01$  y  $n = 20$  resulta

$$P(|\hat{p}_n - (1/6)| < 0,01) \geq 1 - \frac{0,14}{20(0,01)^2} = 1 - 0,7 = 0,3$$

**Ejemplo 6.4.6** ¿Cuántas veces tenemos que tirar el dado para que estemos un 90% veces seguros de que la frecuencia relativa  $\hat{p}_n$  de veces que aparece el cuatro difiere de la probabilidad real en menos de 0.1?

**Resolución**

La ley de los grandes números afirma que

$$P(|\hat{p}_n - p| < \epsilon) \geq 1 - \frac{p(1-p)}{n\epsilon^2}$$

en este caso,

$$P(|\hat{p}_n - p| < 0,1) \geq 1 - \frac{0,14}{n(0,1)^2} = 0,9$$

por lo cual  $\frac{0,14}{n(0,1)^2} = 0,1$  y  $n = \frac{0,14}{(0,1)^3} = 140$ .

## 6.5. Teorema Central del Límite

### 6.5.1. Aproximaciones normales: Binomial

La ley de los grandes números tiene como corolario fundamental dar una estimación de la probabilidad del cometer **por lo menos un error**  $\epsilon$  al aproximar  $p$ , la probabilidad de un evento, usando  $\hat{p}_n$ , la frecuencia relativa de este. Dicha probabilidad se achica cuando el número de repeticiones independientes del experimento aumenta.

$$P(|\hat{p}_n - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2} = \tau$$

La generalidad de este resultado es impactante, no necesito saber nada acerca de la distribución de probabilidades de  $\hat{p}_n$  para calcular estas cotas. Pero si deseo calcular cualquier otra probabilidad que involucre a  $\hat{p}_n$  debo estudiar qué distribución tiene.

Por suerte, el número de veces que se observa un evento, calculado sobre  $n$  repeticiones independientes del experimento, tiene distribución binomial de parámetros  $n$  y  $p$ . Con los avances de la computación, es fácil calcular números combinatorios para  $n$  grande. Sin embargo, existe una aproximación que data del siglo XVIII que permite aproximar los valores de la distribución binomial mediante la distribución Normal.

**Proposición 6.5.1** *Teorema de De Moivre-Laplace*

*Si  $X$  tiene distribución binomial de parámetros  $n$  y  $p$  entonces*

$$Y = \frac{X - np}{\sqrt{np(1-p)}}$$

*tiene distribución normal standard, si  $n$  es suficientemente grande.*

**Ejemplo 6.5.1** Sea  $p$  la proporción real de votantes a favor del candidato Marquez en la población, y sea  $\hat{p}_n$  la proporción muestral de votantes a favor de Marquez, calculado sobre una muestra aleatoria de tamaño  $n$ .

- (a) Acotar  $P(|\hat{p}_{900} - p| \geq 0,025)$  y comparar con la cota encontrada usando la ley de los grandes números.
- (b) Calcular un valor de  $n$  que garantice que  $P(|\hat{p}_n - p| \geq 0,025) \leq 0,01$ . y comparar con el  $n$  encontrado usando la ley de los grandes números.

**Resolución**

Como la variable  $X$  que mide el número de votantes de la muestra a favor de Marquez tiene distribución binomial de parámetros  $n$  y  $p$ , usando la aproximación normal a la binomial resulta

$$\begin{aligned} P(|\hat{p}_{900} - p| \geq 0,025) &= P(|X - np| \geq n0,025) \\ &= P\left(\left|\frac{X - np}{\sqrt{np(1-p)}}\right| \geq \frac{n0,025}{\sqrt{np(1-p)}}\right) \\ &= P\left(|Z| \geq \frac{n0,025}{\sqrt{np(1-p)}}\right) \\ &= 2\Phi\left(-\frac{900 \times 0,025}{\sqrt{900 \frac{1}{4}}}\right) = 2\Phi(-1,5) = 2 \times 0,0668 = 0,1336 \end{aligned}$$

Debemos observar que la ley de los grandes números solo nos permite decir que

$$P(|\hat{p}_{900} - p| \geq 0,025) \leq 0,44$$

mientras que la verdadera probabilidad es tres veces menor.

Si queremos encontrar un valor de  $n$  que garantice que

$$P(|\hat{p}_n - p| \geq 0,025) \leq 0,01$$

entonces

$$\begin{aligned} P(|\hat{p}_n - p| \geq 0,025) &= P\left(\left|\frac{X - np}{\sqrt{np(1-p)}}\right| \geq \frac{n0,025}{\sqrt{np(1-p)}}\right) \\ &= P\left(|Z| \geq \frac{n0,025}{\sqrt{np(1-p)}}\right) \\ &= 2\Phi\left(-\frac{n \times 0,025}{\sqrt{n \frac{1}{4}}}\right) = 0,01 \end{aligned}$$

por lo cual  $-2,575 = z_{-0,005} = -\frac{n \times 0,025}{\sqrt{n \frac{1}{4}}} = -n^{1/2} 0,05$  y  $n = 2653$ . Este es un tamaño de muestra grande pero 20 veces más chico que el calculado con la ley de los grandes números.

**Ejemplo 6.5.2** Supongamos que un ingeniero experto en control de calidad debe elegir 10 sistemas de freno en forma aleatoria de un gran cargamento. El ingeniero no sabe que el 10 % de los frenos no alcanzan las especificaciones de calidad. ¿Cuál es la probabilidad de que no más de 1 de los 10 sistemas de freno en la muestra no pase la inspección?

### Resolución

¿Es esta una situación binomial? No, no lo es. Al sacar un sistema del cargamento esta modificando la proporción de sistema de freno malos que quedan en la población, por lo cual la segunda extracción no es independiente de la primera. Si el cargamento es grande, de todas formas sacar pocos elementos tiene muy poco efecto en la población restante y por lo tanto los resultados de una inspección sucesiva van a ser casi independientes. En este caso podemos actuar como si el número de sistemas de freno que no pasan la inspección  $X$  es una distribución binomial. Como el 10 % de la población son defectuosos,  $X$  es una binomial de parámetro  $p = 0,1$  y  $n = 10$ , el tamaño de la muestra. En la figura 6.3 podemos ver un histograma de esta distribución, como los valores de probabilidad son calculables al conocerse los parámetros, se graficaron como barras.

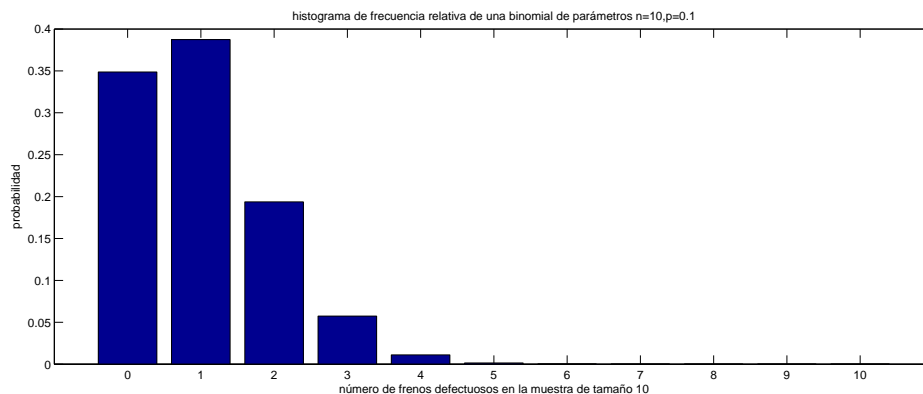


Figura 6.3: Histograma de probabilidad de una distribución binomial con  $n = 10$  y  $p = 0,1$ .

La distribución es marcadamente asimétrica. A pesar de que  $X$  puede tomar valores de 0 a 10, las probabilidades de que se observen valores mayores a 5 son muy pequeñas que casi no aparecen en el histograma.

Calculamos

$$P(X \leq 1) = P(X = 1) + P(X = 0) = 0,3874 + 0,3487 = 0,7361$$

Vemos que casi el 70 % de las muestras van a contener no más de un freno defectuoso, mientras que un 35 % no va a contener ningún freno defectuoso. Una muestra de tamaño 10 no va a alertar al ingeniero de defectos en el cargamento.

Supongamos que se realiza una muestra de tamaño 100 de ese mismo cargamento. Si la tasa  $n/N = 100/N$ , donde  $N$  es el tamaño del cargamento, es menor a 0,05, podemos seguir considerando que la el muestreo aleatorio simple produce observaciones independientes y el número de frenos defectuosos de la muestra va a seguir una distribución binomial de parámetros  $n = 100$  y  $p = 0,1$ .

El histograma de frecuencias relativas de una binomial con tamaño 10 era claramente sesgado a izquierda, es decir, tenía toda su masa concentrada en los primeros dos valores, en cambio el histograma de probabilidad de una distribución binomial de tamaño  $n = 100$  con el mismo  $p = 0,1$ , es aproximadamente normal, como podemos ver en la figura 6.4.

La probabilidad  $P(X \leq 9)$  de que no más de 9 frenos en la muestra de 100 no pasen la inspección es de 0,4513. Si utilizamos la aproximación normal  $X$  va a ser aproximadamente normal con media  $\mu = np = (100)(0,1) = 10$  y desviación estándar  $\sigma = \sqrt{np(1-p)} = \sqrt{(100)(0,1)(0,9)} = 3$ . En la figura 6.4 mostramos la curva binomial sobre escrita sobre el histograma de probabilidad. Tanto el área del histograma como el área bajo la curva normal, ambos suman uno. Notemos que si bien la binomial toma valores hasta el 100, después del 20 son tan pequeños esos valores que no se visualizan como altura en el histograma.

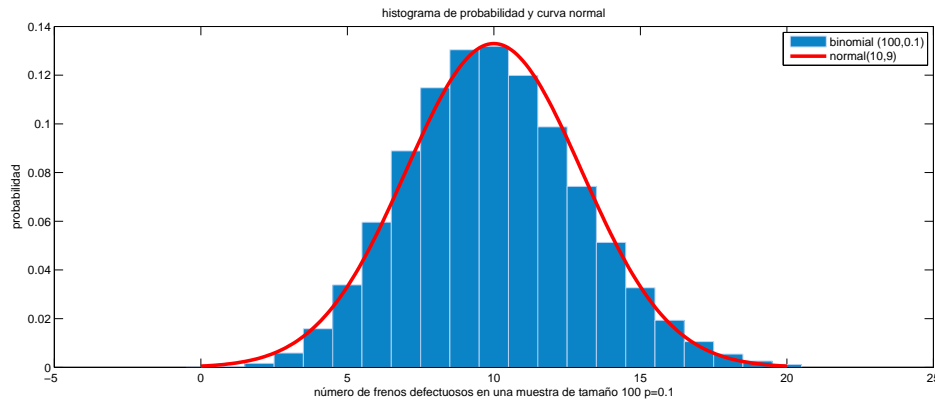


Figura 6.4: Histograma de probabilidad de una distribución binomial de tamaño 100 con  $p = 0,1$ .

La aproximación normal a la probabilidad de que haya no más de 9 frenos defectuosos es

$$P(X \leq 9) = P\left(Z \leq \frac{9 - 10}{3}\right) = P(Z \leq -0,33) = 0,3707$$

La aproximación 0.37 de la probabilidad binomial 0.45 no es muy precisa, aun cuando el los histogramas muestran que el cambio en la distribución es notable al agrandar la muestra, pero se necesita un muestra mucho más grande para generar una aproximación mejor, con un  $p$  tan lejano a 0,5.

### Ejemplo 6.5.3 Simulación

Conocer la distribución nos permite realizar experimentos de simulación. Una forma fácil de realizarlo es usando una tabla de números aleatorios, o un programa de computación que genere números aleatorios. No es trivial definir uno de estos generadores, pero suponiendo que lo tenemos, y que obtenemos números entre 0 y 1 de una forma totalmente aleatoria, entonces podemos pedir  $n = 10$  números y mirar cuántos de ellos son menores a 0.1. El valor resultante es una observación de una variable binomial con parámetros  $n = 10$ ,  $p = 0,1$ .

Si repetimos este procedimiento 1000 veces, obtenemos 1000 valores posibles de una distribución binomial con parámetro  $n = 10$  y  $p = 0,1$ . En las siguientes figuras vamos a mostrar una simulación realizada con el programa MatLab. Vamos a graficar los valores que obtuvimos al simular datos (1000 veces) de dos distribuciones binomiales con distinto tamaño de muestra. Es importante notar que hemos usado la palabra muestra dos veces. Una es la muestra que saca el ingeniero, de tamaño  $n$  y otra es la cantidad de muestras de tamaño  $n$  que simulamos para obtener el histograma de frecuencias relativas.

Estos histogramas no son iguales a las figuras 6.3 y 6.4, pues son frecuencias de muestreo, mientras que los histogramas de las figuras 6.3 y 6.4 son histogramas de probabilidad exactas.

## 6.5.2. Teorema central del límite para una muestra aleatoria simple

**Definición 6.5.1** Una muestra aleatoria de tamaño  $n$  es un vector de variables  $X_1, \dots, X_n$  independientes e idénticamente distribuidas.

### Teorema 6.5.1 Teorema central del límite

Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con media  $\mu$  y varianza  $\sigma^2$ . Entonces si  $S_n = X_1 + \dots + X_n$ , la función de distribución de la variable

$$\frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

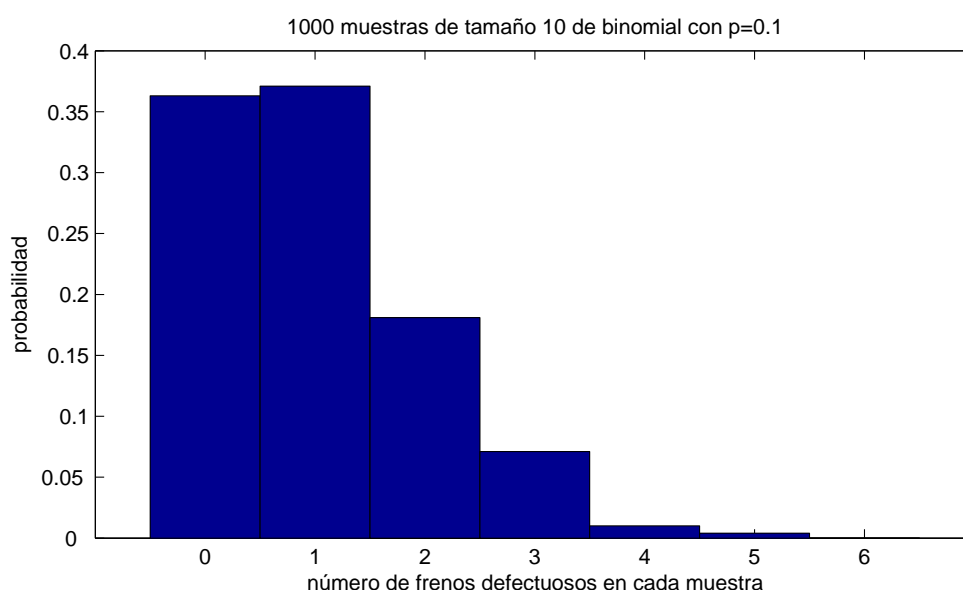


Figura 6.5: Histograma de frecuencia relativa de 1000 muestras de una distribución binomial con  $n = 10$  y  $p = 0,1$ .

converge a la la función de distribución de una variable normal standard. Esto es

$$P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq t\right) \longrightarrow \Phi(t) \quad \forall t$$

**Observación 6.5.1** Si la muestra aleatoria considerada es de variables Bernoulli independientes (indicadoras de la presencia de un éxito en el  $k$ -ésimo ensayo), la variable  $S_n$  tiene distribución binomial. Por lo cual el teorema central del límite se transforma en la aproximación normal a la binomial. El teorema de De Moivre Laplace fue uno de los primeros teoremas de convergencia en distribución enunciados. El teorema central del límite (llamado central por su importancia en la teoría probabilística) siempre se refiere al cálculo de la distribución límite de sumas de variables, bajo diversas hipótesis. La versión que hemos dado aquí es muy restrictiva. Se aplica a **muestras aleatorias** de variables que poseen primer y segundo momento finito. Existen versiones que no requieren que todas las variables tengan la misma distribución (aun admiten que no sean independientes) pero deben formar un "sistema infinitesimal de variables." en algún sentido. El error de medición es un ejemplo de tal sistema. Se supone que el error al medir sin sesgos esta formado por errores infinitesimales, independientes o no, que tomados de a uno son despreciables pero su suma produce un error significativo. Por el teorema central del limite, esa suma tiene distribución normal.

**Observación 6.5.2** Existe un grupo de distribuciones que se pueden aproximar por la distribución normal, tal como se hace con la binomial. Son aquellas distribuciones que resultan de sumar variables independientes con la misma distribución. Enunciaremos algunos casos a continuación.

- a) Si  $X_1, \dots, X_n$  son variables aleatorias independientes con distribución Poisson de parámetros  $\lambda_1, \dots, \lambda_n$  respectivamente, entonces  $X = X_1 + \dots + X_n$  tiene distribución Poisson de parámetro  $\lambda_1 + \dots + \lambda_n$ . Esta propiedad permite aplicar el TCL a una variable  $X$  con distribución Poisson de parámetro  $\lambda$ . Simplemente, pensemos a  $X$  como suma de variables Poisson independientes  $X_1, \dots, X_n$  todas con el mismo parámetro  $\lambda/n$ . Cuando  $n$  es suficientemente grande

$$P(X \leq x) \sim \Phi(x - E(X) / \sqrt{Var(X)}) = \Phi(x - \lambda / \sqrt{\lambda})$$

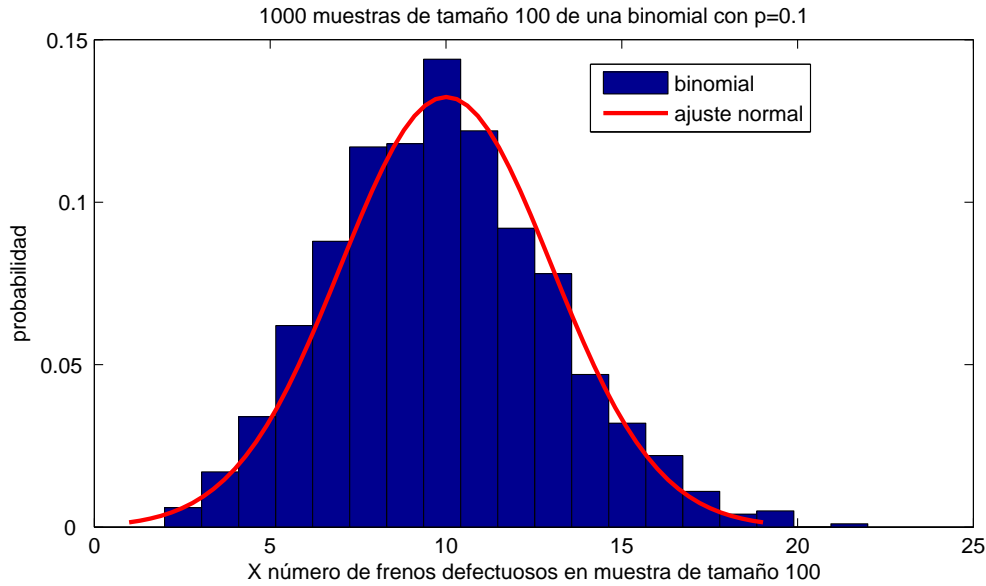


Figura 6.6: Histograma de frecuencia relativa de 1000 muestras de una distribución binomial de tamaño 100 con  $p = 0,1$ .

- b) Sea  $X$  una variable con distribución binomial negativa de parámetros  $r$  y  $p$ , entonces  $X$  puede ser vista como una suma de  $r$  variables geométricas independientes de parámetro  $p$ . Por lo tanto, cuando  $n$  es suficientemente grande

$$P(X \leq x) \sim \Phi(x - E(X)/\sqrt{\text{Var}(X)}) = \Phi(x - (r/p)/\sqrt{r(1-p)/p})$$

**Ejemplo 6.5.4** Supongamos que un programa suma números aproximando cada sumando al entero más próximo. Si todos los errores cometidos son independientes entre sí están distribuidos uniformemente entre -0.5 y 0.5 y se suman 1500 números, ¿cuál es la probabilidad de que la magnitud del error total exceda 15? ¿A lo sumo cuántos números pueden sumarse juntos para que la magnitud del error total se mantenga menor que 10 con probabilidad 0.9?

#### Resolución

Cada error cometido es una variable aleatoria  $\epsilon_k$  con distribución  $\mathcal{U}[-0,5, 0,5]$ , media  $E(\epsilon_k) = [0,5 + (-0,5)]/2 = 0$  y varianza  $\text{Var}(\epsilon_k) = (0,5 - (-0,5))^2/12 = 1/12$ .

Definamos  $S_n = \sum_{k=1}^n \epsilon_k$ , con  $n = 1500$ . Entonces por el teorema central del límite,  $[S_{1500} - nE(\epsilon)]/\sqrt{n\text{Var}(\epsilon)} \sim N(0, 1)$  y

$$\begin{aligned} P(|S_{1500}| > 15) &= 1 - P(|S_{1500}| \leq 15) = 1 - P(-15 \leq S_{1500} \leq 15) \\ &= 1 - P\left(\frac{-15 - nE(\epsilon)}{\sqrt{n\text{Var}(\epsilon)}} \leq \frac{S_{1500} - nE(\epsilon)}{\sqrt{n\text{Var}(\epsilon)}} \leq \frac{15 - nE(\epsilon)}{\sqrt{n\text{Var}(\epsilon)}}\right) \\ &= 1 - P(-1,34 \leq Z \leq 1,34) \\ &= 2\Phi(-1,34) = 0,1802 \end{aligned}$$

Ahora, deseamos encontrar el  $n$  más grande para el cual

$$0,9 = P(|S_n| < 10)$$



Usando el teorema central del límite  $[S_{1500} - nE(\epsilon)]/\sqrt{nVar(\epsilon)} \sim N(0, 1)$  y

$$\begin{aligned} P(|S_n| < 10) = P(-10 < S_n < 10) &= P\left(\frac{-10 - nE(\epsilon)}{\sqrt{nVar(\epsilon)}} \leq \frac{S_n - nE(\epsilon)}{\sqrt{nVar(\epsilon)}} \leq \frac{10 - nE(\epsilon)}{\sqrt{nVar(\epsilon)}}\right) \\ &= P\left(\frac{-10}{\sqrt{\frac{n}{12}}} \leq Z \leq \frac{10}{\sqrt{\frac{n}{12}}}\right) \\ &= 1 - 2P\left(Z \leq \frac{-10}{\sqrt{\frac{n}{12}}}\right) \end{aligned}$$

por lo cual

$$0,9 = 1 - 2P\left(Z \leq \frac{-10}{\sqrt{\frac{n}{12}}}\right) \Rightarrow P\left(Z \leq \frac{-10}{\sqrt{\frac{n}{12}}}\right) = 0,05$$

y  $\frac{-10}{\sqrt{\frac{n}{12}}} = -1,65$ . Entonces, despejando resulta

$$n = \frac{10^2 12}{1,65^2} = 440,7$$

**Ejemplo 6.5.5** Suponga que se tienen 100 lámparas de un cierto tipo, cuya duración puede modelarse como una variable exponencial de parámetro  $\lambda = 0,002$ . Si la duración de cada lámpara es independiente de la duración de las otras, encuentre la probabilidad de que el promedio muestral  $\bar{T} = (1/100)(T_1 + \dots + T_{100})$  se encuentre entre 400 y 550 horas.

**Resolución**

Como  $n$  es 100, podemos suponerlo suficientemente grande y aproximar la distribución del promedio por una normal. Entonces la esperanza y varianza de  $S_n = T_1 + \dots + T_n$  son

$$E(S_n) = E(T_1 + \dots + T_{100}) = 100.E(T_1) = \frac{100}{0,002} = 50000$$

$$Var(S_n) = Var(T_1 + \dots + T_{100}) = 100.Var(T_1) = \frac{100}{0,002^2}$$

$$\begin{aligned} P(400 \leq (1/100)T_1 + \dots + T_{100} \leq 550) &= P(40000 \leq T_1 + \dots + T_{100} \leq 55000) \\ &\sim \Phi(55000 - E(S_n)/\sqrt{Var(S_n)}) - \Phi(40000 - E(S_n)/\sqrt{Var(S_n)}) \\ &= \Phi(55000 - 50000/5000) - \Phi(40000 - 50000/5000) \\ &= \Phi(1) - \Phi(-2) = 0,8413 - 0,0228 = 0,8185 \end{aligned}$$



# Capítulo 7

## Inferencia estadística

El propósito de la inferencia estadística es obtener conclusiones de los datos, fundamentándolas con cálculo probabilístico. Un análisis exploratorio de los datos, realizado mediante la observación de gráficos, puede mostrar efectos que parecen sistemáticos pero solo ser producto de la variabilidad aleatoria. Diferentes gráficos pueden mostrar o disfrazar tendencias, pero los gráficos solos no nos deben convencer de que hay algo más que mala suerte detrás de la tendencia observada.

En éste capítulo nos preparamos para el estudio de inferencia estadística introduciendo ideas de probabilidad. La probabilidad da una descripción matemática a la aleatoriedad presente en muchas situaciones. Vamos a aprender las leyes fundamentales de la probabilidad para poder discutir luego como realizar inferencias, ya sean predicciones o toma de decisiones.

### 7.1. ¿Cómo se hacen las inferencias?

El procedimiento para hacer inferencias se puede entender mejor al analizar nuestros propios métodos intuitivos para hacer inferencias, esto es, tomar decisiones o realizar predicciones basados en resultados parciales.

**Ejemplo 7.1.1** Supongamos que hay dos candidatos para la gobernación y se desea determinar si el candidato Márquez tiene posibilidades de ganar. Entonces la población implicada es el conjunto de respuestas (1 si están a favor, 0 en contra) de todas las personas que van a votar y lo que se pretende determinar es si la fracción a favor de Márquez excede a 0.5. Para simplificar la situación supongamos que todas las personas empadronadas van a votar y se seleccionan 20 personas del Padrón Electoral Oficial. Los 20 afirman que están a favor de Márquez. ¿Qué puede concluirse acerca de las posibilidades de Márquez de ganar las elecciones ?.

La respuesta intuitiva es que Márquez va a ganar las elecciones.  
Analicemos esta respuesta,

- ¿concluimos esto pues creemos que la fracción de votantes de la muestra a favor de Márquez corresponde idénticamente a la fracción respectiva de la población?. No, pues si tomamos varias muestras, estas podrían dar fracciones diferentes, mayores o menores que 0.5.
- ¿Concluiríamos que Márquez tiene que ganar pues sería imposible tener 20 personas de 20 a favor de Márquez si tuviera menos del 50 % del electorado a su favor?. No, pues no es imposible obtener 20 personas de 20 a favor de Márquez (aun cuando menos del 50 % esté a su favor), **pero es muy poco probable** que esto ocurra si Márquez es el perdedor. Por eso concluimos que va a ganar.

Este ejemplo induce al estadístico a pensar que en cada población existe una asignación de probabilidades subyacente que hay que descubrir para poder realizar las inferencias. En este capítulo y en el siguiente nos concentraremos en 2 tipos prominentes de estadística inferencial:

1. Intervalos de confianza
2. Test de hipótesis

## 7.2. Estimar con confianza

El objetivo de la estadística es hacer inferencia sobre la población basados en información contenida en la muestra. Como muchas poblaciones pueden caracterizarse mediante medidas descriptivas llamadas parámetros el objetivo de muchas investigaciones en estadística es hacer inferencia sobre uno o varios parámetros de la población. En esta sección consideraremos la estimación de parámetros poblacionales como la media poblacional, la varianza y la desviación standard.

La estimación tiene muchas aplicaciones prácticas. Por ejemplo, un fabricante de maquinaria puede estar interesado en estimar la proporción  $p$  de máquinas que fallan antes del período de expiración de la garantía, un consultor podría estar interesado en estudiar el tiempo de espera medio  $\mu$  en la cola de la caja de un supermercado, o un ingeniero en la desviación standard del error de medición  $\sigma$  de un instrumento electrónico. Para poder obtener información en todos estos casos, se debe identificar la población, sus parámetros, realizar una muestra y por ultimo definir un estimador como función de los datos que se acerque al parámetro identificado cuando la muestra se agrande.

### 7.2.1. Estimadores puntuales

Veamos estos elementos en un ejemplo:

**Ejemplo 7.2.1** Supongamos que deseamos estimar la cantidad promedio  $\mu$  de mercurio que un proceso determinado puede remover de una onza de mineral extraído de una mina en exploración. Nuestra estimación puede tener dos formatos distintos, un valor o un intervalo. El primero de ellos es un número, por ejemplo, 0.13 onzas, y la idea general es que este número esté lo más cerca posible de la media poblacional desconocida  $\mu$ . Este tipo de estimador es conocido como un estimador puntual, dado que produce un único valor o punto como estimación del parámetro. Si en cambio decimos que  $\mu$  se encuentra entre dos valores distintos, por ejemplo, entre 0.07 y 0.19 onzas, estamos dando un intervalo. Este segundo tipo de procedimiento de estimación genera dos valores que son usados para construir el intervalo (0.08,0.19) con el cual se pretende acotar el verdadero valor  $\mu$  del parámetro poblacional. Llamaremos a este procedimiento estimación por intervalo, y al estimador, o regla que genera los valores de los extremos, intervalo de confianza.

**Definición 7.2.1** *Un procedimiento de estimación puntual utiliza la información en la muestra para generar un único número o punto que estima el parámetro poblacional de interés.*

**Definición 7.2.2** *Un procedimiento de estimación por intervalo hace uso de la información en la muestra para obtener dos números con los cuales se pretende acotar el parámetro poblacional de interés.*

En ambos casos la estimación se realiza mediante un estimador, una regla que nos dice como emplear los datos para determinar el valor, o los valores que usamos en la estimación puntual o por intervalo.

**Definición 7.2.3** *Un estimador es una regla que nos dice como calcular una estimación basado en las mediciones contenidas en la muestra*

Con frecuencia, los estimadores se expresan mediante una fórmula, como por ejemplo, la media muestral

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Esta es un estimador puntual de la media poblacional  $\mu$ . Claramente, la expresión de  $\bar{X}$  es una regla y una fórmula, que nos dice que sumemos las observaciones y dividamos por el tamaño de la muestra.

Un investigador que desea encontrar un intervalo de estimación para un parámetro debe usar los datos muestrales para obtener los dos extremos. Es importante que el intervalo formado por estos puntos tenga una probabilidad alta de contener fehacientemente al parámetro objetivo.

Existen muchos estimadores posibles para un mismo parámetro. En realidad, existen tantos estimadores como funciones de la muestra podamos construir. Eso nos lleva a considerar un punto importante, cuán bueno es un estimador determinado, en el sentido de o cuán bien estima el parámetro objetivo un estimador determinado. Para realizar esta evaluación del desempeño de nuestra regla de estimación, debemos establecer criterios de medición del desempeño en situaciones controladas.

### 7.2.2. Propiedades de un estimador

En muchos aspectos, la estimación puntual es parecida a disparar una flecha a un blanco fijo. El estimador, al generar estimaciones, es análogo al revolver, una estimación particular, a la flecha disparada, y el parámetro de interés es el centro del blanco. Escoger una muestra de la población al azar, y estimar el valor del parámetro con dicha muestra es equivalente a disparar una sola flecha al blanco.

Ahora, supongamos que una persona tira una única flecha al blanco y que la flecha pega en el centro del disco. El tiro es excelente, ¿podemos concluir que es un excelente arquero? Puede ser. ¿Sostendríamos una manzana en la cabeza para que realice el siguiente disparo? No lo creo.

Obviamente, no podemos decidir que esta persona sea un excelente arquero ante tan poca evidencia. Quizás, si se tiraran un millón de flechas en sucesión y todas pegaran en el centro del disco, podríamos adquirir suficiente confianza en el arquero para aceptar sostener la manzana durante el siguiente tiro, si nos pagaran suficientemente. Aunque Guillermo Tell debe haber fallado alguna vez también.

Nuestro argumento es claro, no podemos evaluar la bondad de un estimador basándonos en una sola estimación. Mas bien deberíamos observar muchos resultados de la estimación, y como esas estimaciones son números, construir una distribución de frecuencia del estimador y observar cuán cerca se agrupan alrededor del parámetro de interés.

Vamos a especificar esta idea. Si  $\theta$  es un parámetro poblacional de interés, y  $\hat{\theta}$  un estimador puntual de  $\theta$ , deseáramos que la distribución de probabilidad del estimador se centre en el parámetro a estimar y se concentre alrededor de este valor. Igual que las flechas alrededor del centro del blanco. En otras palabras, deseamos que

$$E(\hat{\theta}) = \theta \quad \text{Var}(\hat{\theta}) < \delta$$

con  $\delta$  chico.

**Definición 7.2.4** Sea  $\hat{\theta}$  un estimador puntual del parámetro  $\theta$ . Entonces  $\hat{\theta}$  es un estimador insesgado si  $E(\hat{\theta}) = \theta$ . Si esto no ocurre, el estimador se llama sesgado.

Ya estudiamos la noción de sesgo en muestreo, o la noción de sesgo sistemático en experimentos. Se refiere a desviaciones de lo que debería ser el centro de masa de la distribución del estimador.

**Definición 7.2.5** El sesgo  $B$  de un estimador puntual  $\hat{\theta}$  está dado por

$$B = E(\hat{\theta}) - \theta$$

Una vez que definimos a donde apuntamos con la flecha, nos vamos a concentrar en cuantas veces le erramos al centro. Es decir, si ponemos todas las flechas en la diana central, o si ponemos algunas allí y otras en el aro siguiente. En promedio, seguimos centrados, pero hay mayor dispersión en los tiros. Matemáticamente, si consideramos dos estimadores insesgados  $\hat{\theta}_1$  y  $\hat{\theta}_2$ , vamos a preferir el de menor dispersión, pues una varianza pequeña va a garantizar que si realizamos muchas muestras independientes, una mayor fracción de

valores del estimador va a estar cerca de  $\theta$ . Por lo cual, además de la propiedad de insesgado, deseamos que un estimador tenga mínima varianza. Por lo cual, no es de sorprender que se llame error cuadrático medio a la función de la muestra

$$ECM = E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + B^2$$

Esta función mide el error al estimar  $\theta$  con el estimador  $\hat{\theta}$ , y permite comparar estimadores según la noción de bondad que describimos anteriormente, varianza y sesgo. Si los dos estimadores son insesgados, el estimador con menor varianza va a ser el estimador con menor error cuadrático medio también.

Este error permite comparar estimadores con sesgo con estimadores insesgados. En algunos casos, un estimador con sesgo pequeño y varianza pequeña tiene menor error cuadrático medio que un estimador insesgado con varianza grande.

### 7.2.3. Algunos estimadores puntuales muy usados

Supongamos tener una población con una característica que puede ser medida mediante variable aleatoria  $X$ , con distribución conocida salvo un conjunto finito de parámetros poblacionales  $\Theta$ . Ejemplos claros de estas poblaciones son los siguientes

1. Los clientes que esperan en la cola de la caja de un supermercado conforman una población cuya característica de interés  $X$  es el tiempo que esperan haciendo cola para pagar. El tiempo medio que espera una persona en la cola es  $\mu$ , la esperanza de la variable  $X$ .
2. Supongamos que mediante un proceso especial se intenta remove mercurio de un mineral extraído en una exploración. Si  $X$  es la variable aleatoria que mide la cantidad de mercurio removida, el parámetro de interés es la cantidad de mercurio removida esperada, esto es, la esperanza  $\mu$  de la variable.
3. Supongamos que un instrumento de precisión mide una cantidad sin cometer errores de sesgo, es importante conocer el error de medición, esto es, la desviación standard de la distribución de la variable  $X$  que representa las posibles mediciones del instrumento.
4. La proporción  $p$  de personas a favor de construir un centro comercial en un barrio residencial es un parámetro poblacional de interés, y es la media de la variable bernoulli  $X$  que vale 1 si una persona elegida al azar contesta en forma afirmativa y cero si no.

Estos problemas enunciados tienen en común la estimación de la esperanza  $\mu$  o la varianza  $\sigma^2$  basados en una muestra aleatoria, esto es, variables aleatorias independientes e idénticamente distribuidas  $X_1, \dots, X_n$  con una distribución caracterizada por los parámetros  $\mu$  o la varianza  $\sigma^2$ . Los estimadores naturales de  $\mu$  y  $\sigma$  son

$$\bar{X} = (1/n) \sum X_i$$

y

$$S^2 = (1/n - 1) \sum (X_i - \bar{X})^2$$

También, si se quiere estimar la probabilidad de éxito  $p$  (la proporción poblacional de éxitos), el estimador mas razonable es la proporción muestral

$$\hat{p} = \sum X_i / n$$

Estos tres estimadores son insesgados, y tiene otras propiedades interesantes, como el hecho que la varianza se reduce cuando el tamaño de la muestra se agranda.

**Proposición 7.2.1** Sea  $X_1, \dots, X_n$  variables aleatorias independientes con distribución común. Sea  $\mu = E(X_i)$  y  $\sigma^2 = Var(X_i)$ . Entonces

1.  $E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \mu$
2.  $Var(\bar{X}) = \frac{\sigma^2}{n}$
3.  $E(S^2) = \sigma^2$

Ya probamos esta proposición anteriormente, la volvemos a escribir para tenerla en cuenta. Es muy importante observar que no decimos que las variables sean normales, sino que solo les pedimos que tengan igual distribución y sean independientes. La proporción muestral es la media muestral de una muestra de Bernoullies, variables dicotómicas (toman valores cero y uno solamente).

### 7.3. Estimación por intervalo

**Ejemplo 7.3.1** El test de aptitud escolástica americano SAT es una medida muy usada en los Estados Unidos para medir la aptitud para estudios universitarios de aspirantes al ingreso a Universidades públicas y privadas.

Hasta el 2004, el SAT consistía en 6 secciones: dos secciones de matemática (con puntajes conjuntos entre 200-800 puntos), dos secciones de habilidad verbal (con puntajes conjuntos entre 200-800 puntos), el test de Inglés escrito standard y una sección de moderación. Cuando el SAT se diseña, los puntajes de cada sección del test se ajustan de tal forma que la distribución de probabilidad del puntaje tenga una media de 500 puntos y desviación standard 100 puntos. Los puntajes medios observados pueden ser menores o mayores, en el año 2003 la media observada nacional sobre 1.4 millones de estudiantes que tomaron el test, fue de 1026 puntos sumando los puntajes de las secciones de matemática y la de habilidad verbal.

Supongamos que se quiere estimar la media de los puntajes del SAT-M (matemática) de una región de los Estados Unidos, que tiene unos 250000 estudiantes a punto de finalizar la escuela secundaria. Si usáramos los puntajes de los alumnos que se han inscripto para tomar el test, estaríamos cometiendo un grave error de sesgo muestral, siendo que estos estudiantes planean ir a la universidad y por eso se preparan en forma especial para tomar el examen. Si queremos tener una buena estimación de la media del test de matemática de todos los estudiantes que finalizan la escuela este año en esta región, debemos realizar una muestra aleatoria de éstos alumnos, tomarles el examen y promediar sus puntajes.

Supongamos que elegimos aleatoriamente 500 alumnos, y el valor obtenido promedio de los puntajes de estos alumnos en el SAT-M es de  $\bar{x}_{500} = 461$ . ¿que se puede decir sobre el puntaje medio de la población basados en la información que tenemos de la muestra?

**Ejemplo 7.3.2** Muchos valores de constantes físicas no son conocidos precisamente sino que deben ser determinados por procedimientos experimentales. Tales operaciones, que parecen poco complicadas a simple vista, como pesar un objeto, determinar un voltaje o medir un intervalo de tiempo son en realidad bastante complicadas cuando se toman en cuenta todos las posibles fuentes de error. Se suele hacer una distinción entre un error sistemático, como puede ser el producido por la mala calibración del equipamiento, y las incontrolables fluctuaciones aleatorias que siempre se observan en mediciones repetidas. Si el verdadero valor de la cantidad a medir se denota por  $\mu$ , entonces la medición se modela como

$$X = \mu + \beta + \varepsilon$$

donde  $\beta$  es una constante que representa el error sistemático, llamado sesgo del procedimiento, y  $\varepsilon$  es la componente aleatoria del error, que a menudo se supone con  $E(\varepsilon) = 0$  y  $Var(\varepsilon) = \sigma^2$ . Entonces

$$E(X) = \mu + \beta \quad Var(X) = \sigma^2$$

Una medición perfecta tendría sesgo  $\beta$  y desvío  $\sigma$  iguales a 0.

Por ejemplo, si queremos estimar la duración media  $\mu$  de una pieza específica de un equipo electrónico, y se probaron 100 piezas elegidas al azar con un procedimiento que no tiene sesgo ( $\beta = 0$ ) que dieron una duración promedio de  $\bar{x} = 501,2$  horas. Si se sabe que  $\sigma$  es de 4 horas, **¿que podemos decir sobre el verdadero valor de la constante  $\mu$  basados en la información que contiene la muestra?**

## 7.4. Confianza estadística

Supongamos que se realizan repeticiones independientes **insesgadas** de la medición de una misma cantidad  $\mu$ , las cuales llamaremos  $X_1, \dots, X_n$ . Entonces

$$E(X_i) = \mu \quad \text{Var}(X_i) = \sigma^2$$

En nuestros ejemplos anteriores, los puntajes de los alumnos de la muestra y la duración de cada uno de las piezas electrónicas son nuestras mediciones.

En nuestros estudios de probabilidad de los capítulos anteriores, descubrimos que si  $n$  es grande, la función de la muestra llamada promedio muestral,  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , tiene las siguientes propiedades

- **Es insesgado**, esto es

$$E(\bar{X}) = \mu$$

- **Es consistente**, pues por la ley de los grandes números  $\bar{X} \rightarrow \mu$  en probabilidad cuando  $n$  va a hacia infinito.

Por lo cual es razonable pensar que la versión observada de  $\bar{X}$  es una buena estimación de  $\mu$ . Pero cuán buena sea esta estimación no puede ser calculada sin conocer la distribución de probabilidad de la muestra. Pero si  $n$  es suficientemente grande, sabemos que

- **La media muestral tiene distribución asintótica normal**, pues por el teorema central del Límite,  $\bar{X} \sim N(\mu, \sigma^2/n)$ .

Entonces, usando la aproximación que nos da el TCL, podemos calcular la probabilidad de que cometamos un error a lo sumo del orden de  $z \sigma / \sqrt{n}$ .

$$\begin{aligned} 2\Phi(z) - 1 &= P(|\sqrt{n}(\bar{X} - \mu)/\sigma| < z) \\ &= P(|\bar{X} - \mu| < z \sigma / \sqrt{n}) \\ &= P(\bar{X} - z \sigma / \sqrt{n} \leq \mu \leq \bar{X} + z \sigma / \sqrt{n}) \end{aligned}$$

por lo cual podemos decir cuál es la probabilidad de que el verdadero valor  $\mu$  esté en un intervalo centrado en la media con ancho dos veces el error de estimación.

Entonces, si existen datos  $x_1, \dots, x_n$ , se suele reportar como estimación de  $\mu$  **el intervalo observado**

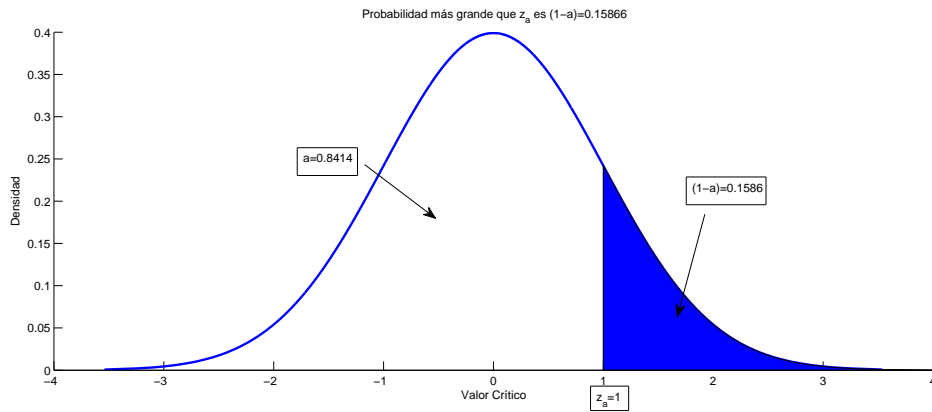
$$[\bar{x} - z \sigma / \sqrt{n}, \bar{x} + z \sigma / \sqrt{n}]$$

y su probabilidad de cobertura  $2\Phi(z) - 1$ , con  $z > 0$ . Debemos puntualizar que el verdadero valor  $\mu$  puede o no estar comprendido dentro de este intervalo observado, la confianza que tenemos en nuestra estimación se basa en el criterio de esperar que ocurra lo más probable, por lo cual se debe pedir que  $2\Phi(z) - 1$  sea cercano a 1. Para formalizar esta idea de confianza en la estimación se usa una notación especial, la de percentil. Si  $z_a > 0$  es tal que  $\Phi(z_a) = a$ , entonces  $z_a$  deja a su izquierda un área bajo la curva normal estándar igual a  $a$ , y a la derecha igual a  $(1 - a)$  y se llama percentil del  $a100\%$ . Veamos en la figura 7.1 el valor de  $a$  y  $(1 - a)$  cuando  $z_a = 1$ .

Elijamos ahora  $\alpha$  chico, y sea  $z_{(1-\alpha/2)}$  el percentil de la normal estandar que deja  $\alpha/2$  área a su derecha (y  $(1 - \alpha/2)$  a su izquierda). Observemos también que  $-z_{(1-\alpha/2)} = z_{\alpha/2}$ , el percentil negativo que deja a su izquierda  $\alpha/2$ .

$$\Phi(z_{(1-\alpha/2)}) = (1 - \alpha/2) \quad 1 - \Phi(z_{(1-\alpha/2)}) = \alpha/2$$



Figura 7.1: Curva normal estándar con  $z_a = 1$ .

Entonces, usando esta notación,

$$\begin{aligned}
 P(|\sqrt{n}(\bar{X} - \mu)/\sigma| < z_{(1-\alpha/2)}) &= P(\bar{X} - z_{(1-\alpha/2)} \sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_{(1-\alpha/2)} \sigma/\sqrt{n}) \\
 &= \Phi(z_{(1-\alpha/2)}) - (1 - \Phi(z_{(1-\alpha/2)})) \\
 &= 1 - 2\Phi(z_{(1-\alpha/2)}) \\
 &= 1 - 2(1 - \alpha/2) \\
 &= 1 - 2\alpha/2 = 1 - \alpha
 \end{aligned}$$

y el intervalo

$$[\bar{X} - z_{(1-\alpha/2)} \sigma/\sqrt{n}, \bar{X} + z_{(1-\alpha/2)} \sigma/\sqrt{n}]$$

se llama intervalo de confianza de nivel  $(1 - \alpha)$  para  $\mu$ .

En la figura 7.2 vemos que si la probabilidad de cobertura es del 95 %, los valores  $z_{(1-\alpha/2)}$  y  $z_{\alpha/2}$  deben ser 1.96 y -1.96. Debemos observar también que cada cola tiene probabilidad  $\alpha/2$ , por lo cual si queremos un intervalo con cobertura 95 %, entonces  $\alpha = 0,05$  y  $\alpha/2 = 0,025$ . Nuestra tabla solo muestra valores acumulados, es decir es una tabla de valores de  $\Phi$  para  $z$  positivos, por lo cual en la tabla debemos buscar el valor que deja probabilidad a la derecha 0.025 y a la izquierda  $1-0.025=0.975$ , (que no es 0.95!!!!). En tabla debemos buscar el valor de  $z$  que hace  $\Phi(z) = 1 - \alpha/2$ , por eso se llama percentil  $(1 - \alpha/2)$ .

**Definición 7.4.1** Sean  $X_1, \dots, X_n$  una muestra de variables aleatorias con  $E(X_i) = \mu$  desconocida y  $\text{Var}(X_i) = \sigma^2$  conocida. Entonces, si  $n$  es suficientemente grande para que valga el TCL, el intervalo aleatorio

$$[\bar{X} - z_{(1-\alpha/2)} \sigma/\sqrt{n}, \bar{X} + z_{(1-\alpha/2)} \sigma/\sqrt{n}]$$

es un intervalo de confianza para  $\mu$  de nivel  $(1 - \alpha)$ . Esto es

$$P(\bar{X} - z_{(1-\alpha/2)} \sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_{(1-\alpha/2)} \sigma/\sqrt{n}) \sim 1 - 2\Phi(-z_{(1-\alpha/2)}) = 1 - \alpha$$

Se dice precisión, o error cometido, a la cantidad  $z_{(1-\alpha/2)} \sigma/\sqrt{n}$ .

**Ejemplo 7.4.1** En el caso del examen SAT-M, la media observada es  $\bar{x}_{500} = 461$ , ¿que se puede decir sobre el puntaje medio de la población basados en la información que tenemos de la muestra?.

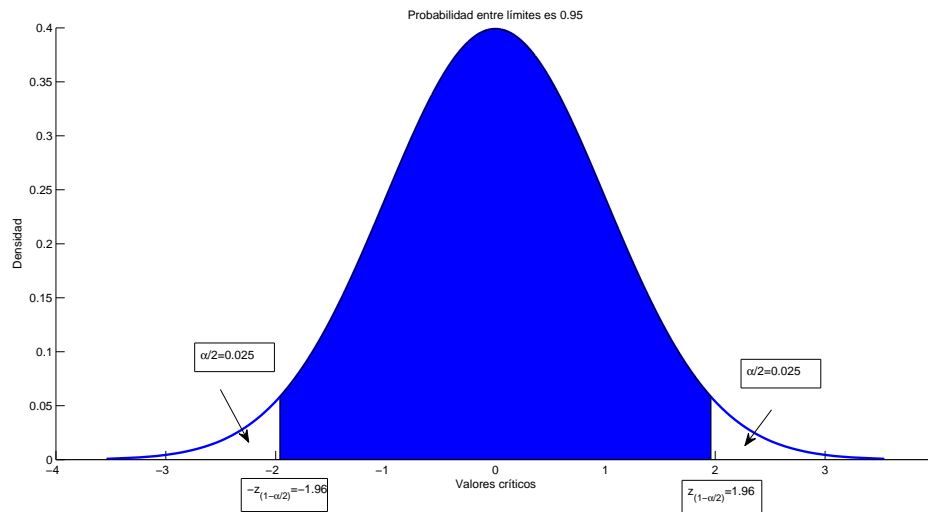


Figura 7.2: Curva normal estándar con  $z_{(1-\alpha/2)} = 1,96$ .

Si se usa la desviación estándar del diseño del test como dato, un intervalo de confianza del 95 % para  $\mu$ , el puntaje medio verdadero de los estudiantes, es

$$[\bar{x} - z_{(1-\alpha/2)}\sigma/\sqrt{n}, \bar{x} + z_{(1-\alpha/2)}\sigma/\sqrt{n}] = [461 - 1,96 \cdot 4,5461 + 1,96 \cdot 4,5] \sim [452, 470]$$

Por lo cual se suele reportar el intervalo observado [452,470] o el valor de la media muestral mas o menos el error cometido  $461 \pm 9$  y se dice que tenemos una confianza del 95 % de que la verdadera media de los puntajes se encuentra en este intervalo. El error depende de la varianza (que se supone conocida), del tamaño de muestra y de la confianza que le querríamos dar a la estimación.

La interpretación de este resultado es delicada. Pueden ocurrir dos cosas

- el intervalo observado [452,470] contiene el verdadero valor de  $\mu$
- Nuestra muestra es una de las pocas muestras para las cuales  $\bar{x}$  no está a 9 puntos del valor de  $\mu$ . Solo un 5 % de todas las muestras dan estos resultados incorrectos.

Por lo tanto, no podemos asegurar que  $\mu \in [452, 470]$ , pero tenemos un 95 % de confianza de que lo contiene, pues llegamos a estos números por un procedimiento que falla solo el 5 % de las veces.

**Ejemplo 7.4.2** Siguiendo con el SAT-M, ¿cual es la precisión con que  $\bar{X}$  estima a  $\mu$  en un 99 % de los casos? Debemos plantear el problema:

$$0,99 = P(|\bar{X} - \mu| < precision) = P(|\sqrt{n}(\bar{X} - \mu)/\sigma| < (\sqrt{n})precision/\sigma)$$

por lo cual

$$\frac{precision}{\sigma/\sqrt{n}} = 2,575 \implies precision = 2,575\sigma/\sqrt{n} = 2,575 \cdot 100/\sqrt{500} = 11,5$$

observemos que la precisión depende del diseño del experimento, esto es, del tamaño de muestra, de la desviación standard y de la confianza con la que queremos estimar, **no de los datos obtenidos**.

**Ejemplo 7.4.3** ¿Cual es el tamaño de la muestra necesario para que la precisión con que  $\bar{X}$  estima a  $\mu$  en un 99 % de los casos sea de 5 puntos?

Debemos plantear el problema:

$$0,99 = P(|\bar{X} - \mu| < 5) = P(|\sqrt{n}(\bar{X} - \mu)/\sigma| < (\sqrt{n})5/\sigma)$$

por lo cual

$$\sqrt{n} \frac{5}{\sigma} = 2,575 \implies \sqrt{n} = 2,575\sigma/5 = 2,575 \cdot 100/5 = 51,5 \implies n = 2652,25 \sim 2653 \text{ personas}$$

**Ejemplo 7.4.4** Supongamos que  $X$  representa la duración de una pieza de un equipo. Supóngase que se probaron 100 piezas que dieron una duración promedio observado de  $\bar{x} = 501,2$  horas. Si se sabe que  $\sigma$  es de 4 horas, encuentre intervalos observados del 95 % de confianza para  $\mu$  y del 99 %. ¿Cuál de los dos es el intervalo más angosto?

#### Resolución

Reemplazando los valores observados en

$$[\bar{X} - z_{(1-\alpha/2)} \sigma/\sqrt{n}, \bar{X} + z_{(1-\alpha/2)} \sigma/\sqrt{n}]$$

con  $\alpha = 0,05$  resulta  $z_{(1-0,025)} = 1,96$  y

$$[501,2 - 1,96 \cdot 4/\sqrt{100}, 501,2 + 1,96 \cdot 4/\sqrt{100}] = [500,4, 502,0]$$

En el caso que  $\alpha = 0,01$  resulta  $z_{(1-0,005)} = 2,575$  y

$$[501,2 - 2,575 \cdot 4/\sqrt{100}, 501,2 + 2,575 \cdot 4/\sqrt{100}] = [500,17, 502,23]$$

El intervalo más largo es el que tiene mayor confianza, pues para obtenerla tengo que reducir la precisión con que estoy estimando. En el caso del intervalo de confianza del 99 %, solo una muestra entre 100 que realice va a generar un intervalo que no contenga a  $\mu$ . También, observemos que el largo del intervalo es dos veces el error, que depende del diseño del experimento y no de los datos.

**Ejemplo 7.4.5** El naturalista frances Buffon tiró la moneda 4040 veces y obtuvo 2048 caras. Este es un experimento binomial con  $n = 4040$  y  $p = P(\text{observar cara})$ . Si se desea estimar  $p$  con una confianza del 99 %, ¿que puedo hacer?

#### Resolución

El estimado más natural de  $p$  es la proporción muestral  $\hat{p} = \frac{1}{n} \sum X_i$ , donde  $X_i$  son variables Bernoulli, que valen 1 cuando se observó una cara y cero si no.

Como en el caso de la media muestral,  $\hat{p}$  es consistente (converge en probabilidad a  $p$ ) por la ley de los grandes números, y por el Teorema de De Moivre-Laplace, podemos decir que  $\hat{p}$  tiene distribución aproximadamente normal con media  $p$  y varianza  $p(1-p)/n$ .

Usando la aproximación normal, un intervalo de confianza del 99 % para  $p$  se obtiene resolviendo la ecuación

$$1 - \alpha = P(|(\hat{p} - p)/\sqrt{p_0(1-p_0)/n}| < z_{(1-\alpha/2)})$$

por lo cual si  $z_{(1-\alpha/2)} = z_{(1-0,005)} = 2,575$ ,  $p_0(1-p_0) \leq 1/4$  y  $n = 4040$ , el intervalo observado es

$$[\hat{p} - z_{(1-\alpha/2)}/\sqrt{4n}, \hat{p} + z_{(1-\alpha/2)}/\sqrt{4n}] = [0,5069 - 2,575/4 \times 4040, 0,5069 + 2,575/4 \times 4040] = [0,5067, 0,5070]$$

**Definición 7.4.2** Sean  $X_1, \dots, X_n$  una muestra de variables aleatorias Bernoulli de parámetro  $p$  desconocido. Sea  $\hat{p} = (1/n) \sum_{i=1}^n X_i$  la proporción muestral. Entonces, si tengo una cota para el valor de  $p$ , digamos  $p_0$ , y  $n$  es suficientemente grande para que valga el TCL, el intervalo aleatorio

$$[\hat{p} - z_{(1-\alpha/2)} \sqrt{p_0(1-p_0/n)}, \hat{p} + z_{(1-\alpha/2)} \sqrt{p_0(1-p_0/n)}]$$

es un intervalo de confianza para  $\mu$  de nivel  $(1 - \alpha)$ . Esto es

$$P(\hat{p} - z_{(1-\alpha/2)} \sqrt{p_0(1-p_0/n)} \leq \mu \leq \hat{p} + z_{(1-\alpha/2)} \sqrt{p_0(1-p_0/n)}) \sim 1 - 2\Phi(-z_{(1-\alpha/2)}) = 1 - \alpha$$

En el caso de no tener una cota  $p_0$ , siempre podemos el hecho de que  $p(1-p) \leq 1/4$ . Por lo cual, el intervalo de confianza para  $p$  resulta

$$[\hat{p} - z_{(1-\alpha/2)} / \sqrt{4n}, \hat{p} + z_{(1-\alpha/2)} / \sqrt{4n}]$$

### 7.4.1. Varianza desconocida

El intervalo depende de la varianza del error de medición  $\sigma^2$ . Si  $\sigma^2$  es desconocido, debe ser estimado y otro tipo de intervalo de confianza para  $\mu$  se deriva en este caso.

Encontremos un estimador para  $\sigma^2$ . Primero, notemos que  $n^{-1} \sum_{i=1}^n X_i^2$  converge a  $E(X^2)$ , por la ley de los grandes números. Segundo, se puede ver que si  $Z_n$  converge en probabilidad y  $g$  es una función continua, entonces

$$g(Z_n) \rightarrow g(\alpha)$$

lo cual implica que

$$\bar{X}^2 \rightarrow [E(X)]^2$$

Finalmente, como  $n^{-1} \sum_{i=1}^n X_i^2$  converge a  $E(X^2)$  y  $\bar{X}^2$  converge a  $[E(X)]^2$ , con un pequeño argumento adicional puede ser visto que

$$\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \rightarrow E(X^2) - [E(X)]^2 = \text{Var}(X) = \sigma^2$$

en probabilidad. Por lo cual, la varianza de  $X$  puede ser estimada por una función de los datos que tiene la propiedad de consistencia, (esto es, converge en probabilidad al parámetro que se quiere estimar). El estimador

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2$$

es el **estimador de momentos** de  $\sigma^2$ . El problema de  $\hat{\sigma}$  es que no es insesgado. Cuando el tamaño de la muestra tiende a infinito, el sesgo tiende a cero y aproxima correctamente el valor del parámetro, pero con muestra finita, su esperanza no es el parámetro. Se corrige a si mismo a medida que el tamaño de muestra se agranda. Por ello es que es mucho más usada una corrección de dicho estimador, el llamado error estándar,  $S^2$ .

$$S^2 = \frac{\hat{\sigma}}{E(\hat{\sigma})} = \frac{n}{n-1} \hat{\sigma} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 =$$

El único detalle es que si sustituimos la desviación estándar  $\sigma/\sqrt{n}$  por el error estándar  $s/\sqrt{n}$ , el estadístico  $T = (\bar{X} - \mu)/(s/\sqrt{n})$  **no tiene distribución normal**, con una muestra de  $n < 120$ . ¿Como calculamos entonces el valor del nivel del intervalo? Si la muestra proviene de una distribución normal, la distribución

del estadístico  $T = (\bar{X} - \mu)/(s/\sqrt{n})$  es una distribución conocida, llamada distribución “t” de Student, y denotada por una “T” caligráfica,  $T \sim \mathcal{T}$ . Esta distribución tiene un parámetro llamado grados de libertad, igual que la distribución chi cuadrado. La distribución t, la distribución chi cuadrado y la distribución F son las llamadas distribuciones del muestreo, pues aparecen cuando uno considera una muestra de variables aleatorias independientes e idénticamente distribuidas normales, y quiere saber cual es la distribución de los estadísticos  $\bar{X}$  y  $S^2$ .

W. S. Gossett descubrió esta distribución en sus datos mientras trabajaba para la cervecería Guinness. En ese momento, la cervecería no permitía a su personal publicar trabajos, por lo cual Gossett usó el sinónimo de Student (un estudiante), al escribir a Sir R. Fisher sobre sus descubrimientos.

**Definición 7.4.3** Supongamos que  $X_1, \dots, X_n$  es una muestra de variables independientes, con la misma distribución normal  $N(\mu, \sigma^2)$ . Entonces el estadístico

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene distribución t de Student, con  $(n - 1)$  grados de libertad.

La distribución  $\mathcal{T}$  es una distribución simétrica como la normal, pero tiene las colas más pesadas. En la figura 7.3 vemos una normal estandar y sobreescrita un distribución  $\mathcal{T}$  con 5 grados de libertad.

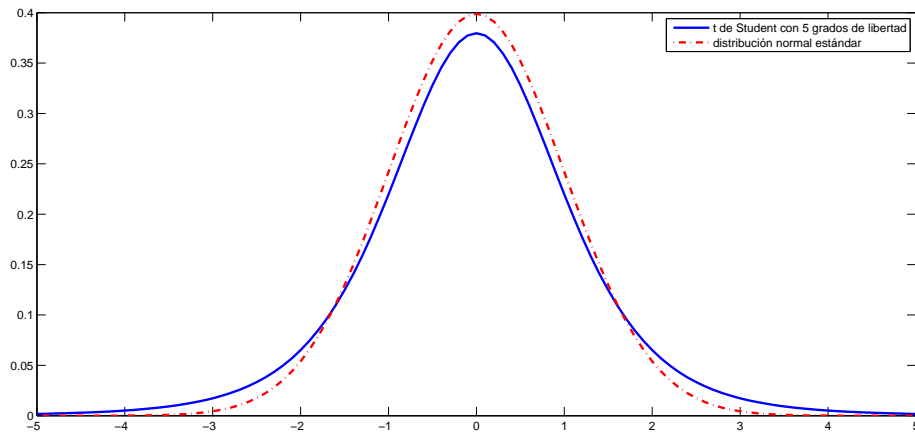


Figura 7.3: Distribución t de Student con 5 grados de libertad y distribución normal estándar.

Con esta distribución es posible entonces calcular un intervalo de confianza para la media  $\mu$  si no se conoce la varianza  $\sigma$ , ya sea que la muestra sea normal o no. El detalle es que si la muestra es normal, el nivel del intervalo es exacto y en el caso de tener que usar el TCL, el nivel del intervalo será aproximado.

**Definición 7.4.4** Supongamos que se obtiene una muestra aleatoria simple de una población, con media  $\mu$  y varianza  $\sigma^2$  desconocidas. Un intervalo de confianza para  $\mu$  de nivel  $(1 - \alpha)$  es

$$\left[ \bar{X} - t_{(1-\alpha/2), (n-1)} \frac{S}{\sqrt{n}}, \bar{X} + t_{(1-\alpha/2), (n-1)} \frac{S}{\sqrt{n}} \right]$$

donde  $t_{(1-\alpha/2), (n-1)}$  es el percentil  $(1 - \alpha/2)$  de la distribución t de student con  $(n - 1)$  grados de libertad. Este intervalo es exacto cuando la distribución de la población es normal y aproximadamente correcto para  $n$  grande en otros casos.

Cuando  $n$  es mayor a 120, la distribución  $t$  es casi indistinguible de la normal, por lo cual se suele usar directamente el intervalo

$$\left[ \bar{X} - z_{(1-\alpha/2)} \frac{S}{\sqrt{n}}, \bar{X} + z_{(1-\alpha/2)} \frac{S}{\sqrt{n}} \right]$$

con el percentil de la normal estándar.

Debemos decir que en muy pocos casos  $\sigma$  es desconocido, si  $n$  es grande se reemplaza directamente  $\sigma$  por  $S$  usando el percentil normal, y si  $n$  es chico, se usa el percentil de una  $t$ .

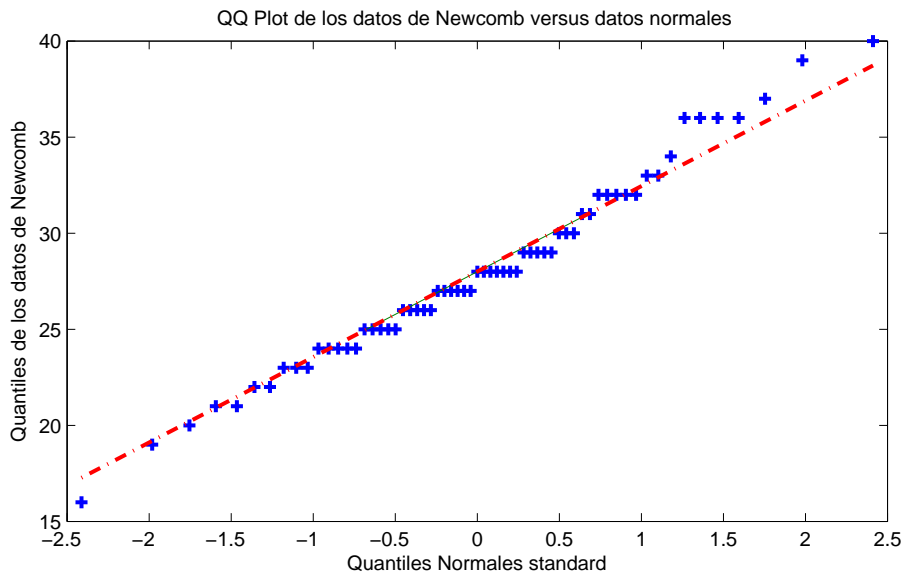


Figura 7.4: qqplot de los datos de Newcomb sobre el tiempo de paso de la luz

**Ejemplo 7.4.6** Recordemos los datos de los 66 tiempos de recorrido de un haz de luz observados por Newcomb, en el ejemplo 5.3.1 El qqplot, o gráfico de los cuantiles muestra que si retiramos las observaciones negativas por considerarlas experimentos fallidos, los datos parecen seguir una distribución normal. ¿Que resultado debió reportar Newcomb a partir de esas 64 observaciones?

Queremos estimar la media  $\mu$  de la distribución de las mediciones de las cuales las 64 mediciones de Newcomb son solo una muestra. Es decir, podemos considerar que Newcomb obtuvo una muestra aleatoria simple de la población de todas las medidas que pudo haber tomado con su instrumento. La media muestral  $\bar{X} = 27,75$  estima a  $\mu$  pero para indicar la precisión de este estimador deberíamos dar un intervalo de confianza. La desviación estandar muestral es  $S = 5,083$ . El intervalo de confianza del 99% usa un valor crítico de la distribución  $t$  de student con 63 grados de libertad, con  $\alpha = 0,005$ . Este es  $t_{(1-\alpha/2),63} = 2,656$ . El intervalo de confianza es

$$\begin{aligned} \left[ \bar{X} - t_{(1-\alpha/2), (n-1)} \frac{S}{\sqrt{n}}, \bar{X} + t_{(1-\alpha/2), (n-1)} \frac{S}{\sqrt{n}} \right] &= \left[ 27,75 - 2,656 \frac{5,083}{\sqrt{64}}, 27,75 + 2,656 \frac{5,083}{\sqrt{64}} \right] \\ &= [27,75 - 1,6876, 27,75 + 1,6876] \\ &= [26,0624, 29,4376] \end{aligned}$$

Los métodos más modernos para medir la velocidad de la luz otorgan al paso de luz de Newcomb un valor de 33.02 billonésima de segundo. Nos deberíamos preguntar si los valores obtenidos por Newcomb son significativamente distintos al valor moderno. Para responder debemos realizar un test de hipótesis pero podemos adelantar que, dado que  $\mu = 33,02$  no pertenece al intervalo de confianza del 99%, este no es un valor posible desde la óptica de Newcomb. Las mediciones fueron realizadas en 1992, y un sesgo sistemático en el experimento no debe descartarse. Ningún método estadístico puede revertir ese problema.

## 7.5. Ajuste de datos: un caso de distribución Poisson

**Ejemplo 7.5.1** Estudios sobre emisión de partículas alfa producidos por fuentes radioactivas muestran que las emisiones por unidad de tiempo no son constantes sino que fluctúan de una manera aparentemente aleatoria. Si la emisión es constante sobre un período de observación (lo cual va a ocurrir si la vida media de la partícula es mucho más grande que el período de tiempo observado), y las partículas provienen de un número grande de fuentes independientes, el modelo Poisson parece ser teóricamente apropiado. Por esa razón, la distribución Poisson es usada frecuentemente como modelo de decaimiento radioactivo. Debemos recordar que la distribución Poisson como modelo de conteo aleatorio en espacio y tiempo se basa en tres suposiciones, 1) la tasa con la que los eventos ocurren es constante en espacio o tiempo, 2) eventos en intervalos disjuntos de tiempo o espacio ocurren en forma independiente, 3) no hay eventos múltiples. Berkson (1966) realizó un análisis cuidadoso de los datos obtenidos por el “National Bureau of Standards”. La fuente de las partículas era americio 241. Los científicos observaron 10220 tiempos entre emisiones sucesivas. La tasa media de emisiones (total de emisiones dividida por el tiempo total) fue de 0.8392 emisiones por segundo. El reloj usado tenía una precisión de 0.0002 segundos, esto es, no distinguía ningún valor menor que este. La primera columna de la siguiente tabla muestra el número  $n$  de partículas observadas en los 1207 intervalos, y la segunda columna el número de intervalos de 10 segundos que vieron un determinado número de partículas. Por ejemplo, en 18 intervalos de 10 segundos contaron entre 0 y 2 partículas, en 28 intervalos se contaron 3 partículas, etc. La distribución Poisson otorga probabilidad a todos los números naturales, más el cero. Sin embargo, como esta es una muestra finita, solo se observaron algunos de estos valores. Esta es una forma de mostrar datos que de otra manera sería imposibles de visualizar, imagínense una tabla de 1207 datos. la que vemos es una tabla de frecuencia relativa, y agrupa estos datos para que en cada “celda” se observen por lo menos 5 intervalos.

$n$	Observados	Esperados	$X_i^2 = (O_i - E_i)^2 / E_i$
0-2	18	12.2	2.76
3	28	27	0.04
4	56	56.4	0.01
5	105	94.9	1.07
6	126	132.7	0.34
7	146	159.1	1.08
8	164	166.9	0.05
9	161	155.6	0.05
10	123	130.6	0.44
11	101	99.7	0.02
12	74	69.7	0.27
13	53	45	1.42
14	23	27	0.59
15	15	15.1	0.00
16	9	7.9	0.57
17+	5	7.1	0.57
	1207	1207	8.99

Al ajustar la variable Poisson al número de partículas emitidas, estamos suponiendo que los 1207 números observados son realizaciones independientes de variables Poisson con la misma tasa  $\lambda$  y función de probabilidad

$$\pi_k = P(X = k) = \frac{\lambda^k \exp(-\lambda)}{k!}$$

Como  $\lambda$  es desconocido, para ajustar la distribución debemos determinar  $\lambda$  de los datos. Como el número promedio en los intervalos de 10 segundos es 8.392, tomaremos este valor como una estimación de  $\lambda$  y la llamaremos  $\hat{\lambda}$ . Esta es una observación de una variable aleatoria,  $\hat{\lambda}$  definida como el número promedio de emisiones en períodos de 10 segundos. Esta es una variable que cambia cada vez que observo otra muestra de datos. La regla de estimación es el promedio, la observación particular en esta muestra es 8.392. Podríamos dar también un intervalo de confianza para el

parámetro  $\lambda$  basado en  $\hat{\lambda}$ , que nos diría hasta que decimal sería precisa esta estimación, pero lo que nos interesa ahora es usar este valor puntual para realizar una ponderación de cuán bueno es el ajuste del modelo Poisson a los datos.

Con esta media,  $\hat{\lambda} = 8,392$  podemos calcular exactamente cuántos intervalos de 10 segundos deberían observarse con una cantidad específica de partículas. Por ejemplo, la probabilidad de observar 0, 1 o 2 partículas tiene probabilidad

$$p_1 = \pi_0 + \pi_1 + \pi_2 = \exp(-\hat{\lambda}) + \hat{\lambda}\exp(-\hat{\lambda}) + \frac{\hat{\lambda}^2 \exp(-\hat{\lambda})}{2!}$$

La probabilidad de caer en la tercer celda (consideramos solo 16, agrupándolas para que tengan por lo menos 5 partículas cada una de ellas), es  $p_2 = \pi_3 = \frac{\hat{\lambda}^3 \exp(-\hat{\lambda})}{3!}$ . La probabilidad de que una observación caiga en la celda 16 es

$$p_{16} = \sum_{k=17}^{\infty} \pi_k$$

Bajo la suposición de que las 1207 variables son independientes con distribución Poisson, el número de observaciones de las 1207 que caen en cada celda tiene distribución binomial, y la distribución conjunta de los números de valores en cada celda es multinomial con  $n = 1207$  y  $p_1, \dots, p_{16}$ . La tercera columna de la tabla muestra los valores esperados siguiendo esta distribución. Esto es, si  $p_5 = 0,0786$ , el número esperado de intervalos que observan ese número de partículas es  $1207 \times 0,0786 = 94,9$ . Vemos que el valor observado en esta muestra es 105, esto es, observamos 105 intervalos de los 1207 con 5 partículas, y este valor no está muy lejos de lo que deberíamos esperar con una tasa  $\hat{\lambda}$ . Pero esta es comparación muy informal, para hacer un ponderación imparcial del ajuste realizado debemos realizar un test de hipótesis, usando una medida de discrepancia objetiva. Karl Pearson diseñó el siguiente estadístico para comparar observaciones multinomiales con los valores esperados en cada celda:

$$X^2 = \sum \text{todas las celdas} \frac{(O_i - E_i)^2}{E_i}$$

La distribución de probabilidad de este particular estadístico se llama chi cuadrado, y es la distribución que le corresponde a la sumas de cuadrados de variables independientes con distribución normal. Con estos datos, el valor observado de la discrepancia  $X^2$  es 8.99. Podemos adelantar que mientras más grande sea este valor, peor es el ajuste de los datos, pero una vez más debemos tener alguna forma objetiva de decir cuán grande debe ser este valor para afirmar que los datos no provienen de una distribución Poisson. Empecemos primero suponiendo que el modelo es correcto. Los valores observados generados a partir de un modelo correcto dan lugar al  $X^2$  observado. Como  $X^2$  es una función de las posibles muestra observadas, es una variable aleatoria y tiene una distribución, llamada distribución nula. Puede verse que si la distribución Poisson es correcta para estos datos,  $X^2$  tiene (aproximadamente) distribución chi cuadrado con grados de libertad dados por la fórmula:

df= número de celdas -número de parámetros independientes ajustados-1

En nuestro caso hay 16 celdas, y se estimó un solo parámetro,  $\lambda$ , por lo cual hay 14 grados de libertad. Esta aproximación nos permite dar una aproximación a la probabilidad de observar el valor 8.99 o más que el, a la cual vamos a llamar  $P$  – *valor*. Si esta probabilidad es muy pequeña, estamos ante un caso inusual, la muestra observada no es muy fácil de ser producida si las hipótesis son correctas. En cambio, en nuestro caso

$$p^* = P(X^2 > 8,99 | \text{el modelo es correcto}) = 0,83$$

Por lo cual si el modelo es correcto, veremos discrepancias de ese estilo o más grandes 8 de cada 10 veces. No hay razón para dudar del modelo basándonos en el estadístico de Pearson. En la figura 7.5 vemos la distribución chi cuadrado con 14 grados de libertad. Es una distribución asimétrica con la cola derecha muy pesada, vemos que el área bajo la curva entre la raya roja y el infinito positivo es mucho más grande que el área entre cero y la raya roja. Esto indica que la muestra observada es una muestra posible de la distribución Poisson.

Es importante observar que si bien queríamos estudiar un ajuste Poisson, una distribución discreta, transformamos los datos para tratar de estudiar un ajuste chi cuadrado, una distribución continua. La razón



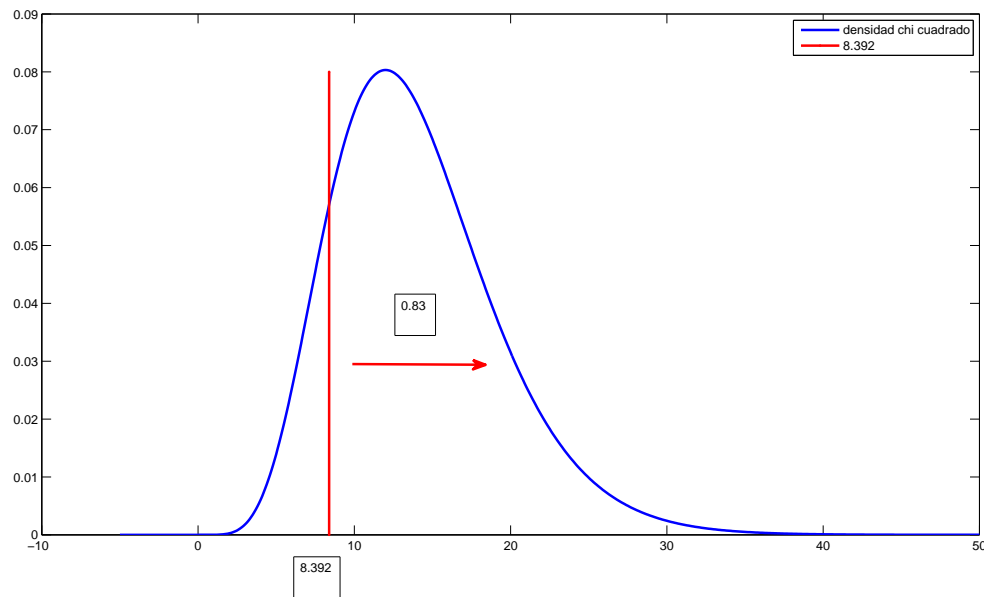


Figura 7.5: Distribución chi cuadrado con 14 grados de libertad.

de esto es que es más simple estudiar ajustes continuos que discretos, pero la equivalencia existe. La teoría dice que una muestra de la distribución Poisson genera un discrepancia  $X^2$  con distribución (aproximada) chi cuadrado. Entonces si la función de la muestra llamada discrepancia  $X^2$  tiene un buen ajuste chi cuadrado, no hay evidencia de que la muestra no sea Poisson. Si la discrepancia  $X^2$  no es chi cuadrado, entonces la muestra no puede haber sido Poisson.

En el proximo capítulo profundizaremos este método de plantear hipótesis. Pero debemos anticipar que estamos planteando test de hipótesis de acuerdo con el paradigma de Neyman-Pearson, en donde las hipótesis son planteadas de forma que la distribución nula sea conocida, y el valor de la probabilidad bajo hipótesis nula sea el que decide a favor o en contra de la hipótesis nula.



# Capítulo 8

## Nociones de test de hipótesis

### 8.1. Introducción

Estimación puntual e intervalos de confianza son procedimientos de inferencia estadística que permiten dar información sobre valores específicos de parámetros. Otro tipo de inferencia que es usada frecuentemente es el testeo de hipótesis. Los test de hipótesis intentan responder este tipo de preguntas:

**El efecto observado en la muestra, fué debido al azar o a algún otro motivo?**

Recordando el ejemplo que dimos en nuestra introducción de nuestro libro, ya estamos en condiciones de generar una familia de modelos probabilísticos que permita conocer la probabilidad de que un número  $n$  de votantes (entre  $N$  elegidos al azar) esté a favor del candidato Márquez, si la verdadera proporción de votantes a favor de Márquez es un número  $p$ .

Ahora queremos conocer como un estadístico puede inferir que Márquez tiene grandes chances de ganar las elecciones, si el porcentaje de votantes a favor de Márquez en la muestra observada es del 100 %. ¿Porque?

**El porcentaje observado en la muestra, fué debido al azar o porque Márquez tiene más del 50 % de los votantes de la población a su favor?**

Con el modelo binomial, la probabilidad de obtener esa muestra debido al azar, o sea, **si Márquez no fuera a ganar**, es muy pequeña. Por lo cual es razonable decidir a favor de que Márquez es claro ganador. Un encuestador diría que *con un test para proporciones a una cola, la hipótesis nula fue rechazada a nivel del 1 %*. La proporción a favor del candidato debe ser considerada estadísticamente significativa. Este lenguaje es muy particular, por lo cual comenzaremos por describir las ideas detrás de los test de hipótesis y familiarizarnos con el lenguaje.

**Ejemplo 8.1.1** Una nueva vacuna se testea para determinar su efectividad en prevenir la gripe. Diez personas son inyectadas con la vacuna y observadas por el período de un año. Ocho de ellas pasan el invierno sin tener gripe. Supongamos que cuando no se utilizaba la vacuna la probabilidad de no contagiarse la gripe en todo el año es de  $p = 0,5$ . Se quiere saber, basándose en la información de la muestra, si la vacuna es efectiva.

**Ejemplo 8.1.2** Una industria quesera familiar compra leche a varios proveedores como el principal ingrediente del queso. El gerente de producción sospecha que algunos proveedores están agregando agua a la leche para incrementar sus ganancias. Se sabe que el exceso de agua puede ser detectado determinando el punto de congelamiento de la leche, ya que la temperatura a la cual la leche se congela varía normalmente con media  $\mu = -0,540C^0$  y desviación standard de  $0,008C^0$  y agregar agua eleva el punto de congelamiento de la leche. En cambio un punto de congelamiento menor puede ser resultado de exceso de crema en la leche. El técnico del laboratorio de la compañía analiza 5 lotes consecutivos de leche de un mismo proveedor y encuentra que el punto de congelamiento medio de la muestra es de  $-0,538C^0$ . ¿Es ésta evidencia substancial de que el productor está agregando agua a la leche?

Estos son ejemplos muy simples de test de hipótesis.

- ¿Presentan los datos de la muestra suficiente evidencia de que la vacuna es efectiva?.
- ¿Se tiene evidencia substancial de que el proveedor está agregando agua a la leche?

Para responder estas preguntas, usualmente se plantea la posición del escéptico, que dice que las variaciones observadas son solo producto del azar, y la del investigador que argumenta a favor de un cambio real. En nuestro ejemplos, un crítico escéptico se preguntaría

- ¿cuál es la probabilidad de obtener una muestra como la observada si la vacuna no es efectiva?
- ¿cuál es la probabilidad de obtener una muestra como la observada si la leche no fue aguada?

Si estas probabilidades resultan muy chicas, podrían ocurrir dos cosas

- En ambos casos, se tuvo mucha mala suerte y se observo una muestra muy poco común, pero la leche no esta aguada y la vacuna no es efectiva.
- - La leche fue aguada y se observó un resultado esperable de la distribución del punto de congelamiento de ese tipo de leche adulterada.
  - La vacuna es efectiva y se obtuvo un resultado probable de la distribución del número de enfermos de gripe en la población de vacunados.

El escéptico debe decidir que probabilidad se considera suficientemente chica, pues la frecuencia con que se encuentran muestras inusuales depende de esa probabilidad.

## 8.2. La lógica del test de hipótesis

El razonamiento que se emplea es el siguiente, deseamos saber si la hipótesis de estudio, llamada hipótesis alternativa, es verdadera.

- La hipótesis nula dice que no hay efecto aparte del que podría ser producido por el azar.
- La hipótesis alternativa dice que hay un efecto aparte del que podría ser producido por el azar.

Para probar que hay alguna clase de efecto, hay que probar falsa la posibilidad de que los resultados puedan haber sido producidos por el azar. Hay que probar falsa la hipótesis nula para validar la alternativa. Hay que convencer al escéptico.

El planteo del test de hipótesis tiene un axioma que hay aceptar, se supone que **siempre ocurre lo más probable**, lo cual no siempre es cierto. Si uno acepta esto, entonces consideraríamos como verdadera la hipótesis más probable.

### 8.2.1. ¿Como planteamos las hipótesis?

La hipótesis nula, llamada  $H_0$ , representa la idea de "no hay efecto". La hipótesis alternativa  $H_A$  representa la idea "hay efecto". Para poder cuantificar las probabilidades de ocurrencia, debemos expresarlas a través de un parámetro poblacional estimable. La decisión se tomará a partir del valor del estadístico que estima el parámetro (al cual debemos conocerle su distribución de probabilidad exacta o aproximada si es válida la hipótesis nula).

En nuestro ejemplo de la vacuna, la hipótesis nula es **la vacuna es inefectiva** y la hipótesis alternativa **la vacuna es eficiente**. Sabemos que la la probabilidad de enfermarse es  $p = 0,5$ . Si la vacuna es ineficiente,  $p$  no varía a pesar de haber sido vacunado. Si la vacuna es eficiente,  $p$  debería ser menor. Por lo cual

$$H_0 : p = 0,5 \quad \text{vs} \quad H_A : p < 0,5$$

son las hipótesis en función de parámetros poblacionales estimables a partir de una muestra.

### 8.2.2. ¿Como tomamos la decisión?

Cuando  $H_0$  es verdadera, esperamos que el estimador tome valores cerca del parámetro representado por  $H_0$ . Si la probabilidad de observar un valor menor o igual que el observado, bajo hipótesis nula, es muy chico, (esto es, estoy lejos del parámetro verdadero) entonces consideramos que hay evidencia en contra de la hipótesis nula.

#### Ejemplo 8.2.1 Ejemplo de la vacuna.

En el ejemplo de la vacuna, intuitivamente, se selecciona el número de personas  $Y$  que se enferman como la medida de la cantidad de evidencia de la muestra. Sabemos, pues es un dato, que la probabilidad de enfermarse de la gente no vacunada es  $p = 0,5$ . Por lo cual, si tengo  $n$  personas en mi muestra aleatoria, la cantidad esperada de gente que se enferma (si la vacuna no modifica en nada a  $p$ ) es de

$$E(Y) = np = 10 \times 0,5 = 5$$

. Si en 10 personas, 6 se enferman, es fácil decidir que no hubo cambio, la vacuna no hizo nada, lo mismo que si solo una se enferma, hay evidencia a favor de la vacuna. Pero si se enferman 3 o 4, ya no es tan seguro que haya evidencia a favor de alguna de las hipótesis. La vacuna hizo bajar la tasa de enfermos o ese año la gente se cuidó más del frío?.

Si observamos  $Y = 3$ , deberíamos calcular cuál es la probabilidad de obtener ese valor o menor si la vacuna es inefectiva. Cuanto más pequeña esa probabilidad, mayor evidencia en contra de la hipótesis nula hay. Esta probabilidad es 0.1719, nos dice que veríamos un resultado como este 20 veces en 100 si la vacuna no fuese efectiva. La probabilidad es pequeña, pero no tan pequeña como para no hacer un estudio más profundo, con una muestra mayor. Siguiendo nuestra filosofía de que siempre ocurre lo más probable, no se rechazaría la hipótesis nula con un resultado de este estilo.

#### Ejemplo 8.2.2 Ejemplo de la leche.

Tratemos de definir las hipótesis de estudio en el ejemplo de la posible leche adulterada. La sospecha del gerente es la hipótesis de investigador, la hipótesis alternativa, y la hipótesis nula es la del escéptico, que no vé cambio en la leche comprada. Para precisar estas hipótesis en términos de parámetros estimables, observamos que el punto de congelamiento de la leche es una variable que puede ayudar a discriminar si la leche fue adulterada. Agregar agua eleva el punto de congelamiento, mientras que una leche con mucha crema tiene un punto de congelamiento menor a lo usual. El punto de congelamiento usual no es fijo, es una variable aleatoria con distribución normal, con media  $\mu = -0,545^\circ\text{C}$  y desviación standard de  $0,008^\circ\text{C}$ . Por lo cual un cambio a favor del adulteramiento con agua seria observar un valor correspondiente a una distribución normal con una media mayor a la usual:

$$H_0 : \mu = -0,545 \quad \text{vs} \quad H_A : \mu > -0,545$$

El técnico del laboratorio de la compañía analiza 5 lotes consecutivos de leche de un mismo proveedor y encuentra que el punto de congelamiento medio de la muestra es de  $-0,538^{\circ}\text{C}$ . ¿Es ésta evidencia substancial de que el productor está agregando agua a la leche?

Si definimos la región de rechazo como

$$RR = \{x/x \geq -0,538\},$$

el error cometido al rechazar la hipótesis nula en favor de la hipótesis de leche adulterada, cuando en realidad HN era verdadera, es el área bajo una curva normal a la derecha del valor  $-0,538$ ,

$$P(X \geq -0,538 | X \text{ es normal con media } -0,538 \text{ y desviación estandar } 0,008/\sqrt{5}) = 0,025$$

Esto quiere decir que una temperatura de congelamiento media como la observada solo ocurre 2 veces y media cada 100 mediciones de leche no adulterada, por lo cual hay evidencia de que la leche ha sido adulterada por el productor.

### 8.3. P-valores

Un test de significación evalúa la evidencia en contra de la hipótesis nula en términos de probabilidades. Si el resultado observado es poco probable bajo la suposición de que la hipótesis nula es verdadera, ese resultado es evidencia en contra de  $H_0$  en favor de  $H_a$ . Mientras menos probable sea un resultado, mayor va a ser la evidencia de que  $H_0$  es falsa. La hipótesis alternativa determina que clase de resultados son los que cuentan en contra de la hipótesis nula. En el ejemplo de la leche, valores mayores al punto de congelamiento indicaban leche adulterada con agua, valores menores indicaban concentración de crema. Mientras más lejos este el valor observado de la media del congelamiento, más evidencia en contra de la adulteración de la leche. La probabilidad que mide el peso de la evidencia de que la leche fue aguada es la probabilidad de que el estimador de la media produzca un valor de por lo menos el observado con los cinco datos

$$P(\bar{X} \geq -0,538).$$

Esta probabilidad se calcula bajo el supuesto de que la media poblacional es de  $\mu = -0,345$ , bajo hipótesis nula.

En general un test de significación encuentra la probabilidad de obtener un resultado *al menos tan extremo como el observado*. La dirección en la que se cuenta lo extremo está determinada por  $H_a$ , y puede ser a una cola (valores mayores, cola derecha, o menores, cola izquierda) o a dos colas, en los cuales la evidencia cuenta en contra en ambas direcciones. En el ejemplo de la leche realizamos un test a cola derecha y en el de la vacuna a cola izquierda.

Vamos a definir ahora el llamado **P-valor**.

**Definición 8.3.1** Se llama P-valor a la probabilidad calculada asumiendo válida la hipótesis nula, de que el estadístico tome un valor al menos tan extremo como el observado. Más chico es el P-valor, más grande es la evidencia en contra de  $H_0$  dada por los datos.

**Ejemplo 8.3.1** Volvamos al ejemplo del gerente y sospecha acerca de la leche aguada. Realizaremos ahora los cálculos que resultan un P-valor de 0.025, con el cual concluimos que hay evidencia en contra de la hipótesis nula.

Sabemos que se observó una media muestral  $\bar{x} = -0,538$ , con una muestra aleatoria simple de tamaño  $n = 5$ . La población de leche sin adulterar tiene como punto de congelamiento a una variable aleatoria normal con media  $\mu = -0,545$  y desviación estándar  $\sigma = 0,008$ . El P-valor para testear las hipótesis

$$H_0 : \mu = -0,545 \quad H_a > -0,545$$

es

$$P(\bar{X} \geq -0,538).$$

calculado asumiendo que  $H_0$  es verdadera. En ese caso, el estimador  $\bar{X}$  es una variable normal con media  $\mu = -0,545$  y desviación estándar  $\sigma_{\bar{X}} = \sigma/\sqrt{5} = 0,008/\sqrt{5}$

Estandarizando la variable como hicimos en los capítulos anteriores para usar la tabla normal patron, o usando un software de estadística, resulta

$$\begin{aligned} P(\bar{X} \geq -0,538) &= P\left[\frac{\bar{X} - (-0,545)}{0,008/\sqrt{5}} \geq \frac{-0,538 - (-0,545)}{0,008/\sqrt{5}}\right] \\ &= P(Z \geq 1,96) \\ &= 1 - 0,975 = 0,025 \end{aligned}$$

**Ejemplo 8.3.2** En el ejemplo de la vacuna,  $Y$ , el número de personas que se enferman es el estadístico del test. Como la probabilidad de enfermarse de la gente no vacunada es  $p = 0,5$  y tengo una muestra aleatoria simple de 10 personas, suponiendo válida  $H_0$ ,  $Y$  es una variable binomial con probabilidad de éxito  $p=0,5$  y número de pruebas 10. Si observamos  $Y = 3$ , el P-valor es

$$P(Y = 0, 1, 2, 3/p = 0,5) = \left[\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3}\right](0,5)^{10} = 0,1719$$

Si observamos  $Y = 2$ , el P-valor es

$$P(Y = 0, 1, 2/p = 0,5) = \left[\binom{10}{0} + \binom{10}{1} + \binom{10}{2}\right](0,5)^{10} = 0,0547$$

y este valor es suficientemente chico para rechazar la hipótesis nula.

Recordemos que como el test es a cola izquierda, solo cuentan en contra de la hipótesis nula valores chicos de  $Y$ . Si hay muchos enfermos, aun cuando sea inusual, es clara evidencia de que la vacuna no funciona.

### 8.3.1. Significado estadístico

En la sección anterior hemos estudiado casos particulares donde conocíamos las distribuciones de ciertos estadísticos, bajo hipótesis nula, lo que nos permitía calcular los  $P$ -valores y tomar decisiones. Vamos ahora a sistematizar estos procedimientos.

**Definición 8.3.2** Si el  $P$ -valor es menor o igual que un valor fijo  $\alpha$ , decimos que los datos son estadísticamente significativos con nivel  $\alpha$ .

Significativo no quiere decir importante. Quiere decir que significa algo. En estadística se lo usa para indicar que la evidencia en contra de la hipótesis nula alcanza el estándar marcado por  $\alpha$ . La interpretación de esto debe hacerse de acuerdo al problema. No todo test de hipótesis debe rechazar con nivel  $\alpha = 0,05$ . Recordemos el test de bondad de ajuste con que cerramos el capítulo anterior, allí el  $P$ -valor nos daba una indicación de que la hipótesis nula era correcta!!!

Una receta para testear el significado de la evidencia en contra (o a favor) de la hipótesis nula se llama test de significación. Recordemos lo que dijimos acerca del paradigma de Neyman-Pearson. Usualmente, la hipótesis nula es totalmente identificable, es decir, conocemos todos los parámetros de la distribución, por eso podemos calcular  $P$ -valores bajo hipótesis nula. Y en la mayoría de los test, tratamos de obtener evidencia en contra de esa hipótesis. Pero en el caso de bondad de ajuste, la hipótesis nula es que los datos siguen una distribución totalmente identificada y deseamos saber si hay evidencia a favor (o en contra) de ese ajuste. Lo importante es que la evidencia en contra se obtiene con un  $P$ -valor pequeño, y la evidencia a favor con un  $P$  valor grande. Evidencia para no dudar de la hipótesis nula debe ser obtenida también por otros medios. Si no, un  $P$ -valor mediano, ni muy chico, ni muy grande, solo es evidencia de que faltan datos.

Los pasos comunes a la mayoría de los test de significación son los siguientes:

1. Se proponen las hipótesis  $H_0$  nula y alternativa  $H_a$ . El test se diseña para evaluar la fuerza de la evidencia en contra de  $H_0$ .  $H_a$  es la propuesta que vamos a aceptar si la evidencia nos permite rechazar  $H_0$ .
2. Se especifica el nivel de significación  $\alpha$ , que permite decir por anticipado cuanta evidencia se necesita para rechazar  $H_0$ .
3. Se calcula el estadístico del test, cuya distribución debe ser conocida suponiendo válida la hipótesis nula, y que nos permite medir cuán bien los datos soportan  $H_0$ .
4. Se calcula el  $P$ -valor de los datos observados. Esta es la probabilidad calculada asumiendo válida  $H_0$ , de que el estadístico del test vaya a pesar tanto o más en contra de  $H_0$  como lo hace con estos datos. Si el  $P$ -valor es menor o igual a  $\alpha$ , el test es significativo con nivel  $\alpha$ .

**Definición 8.3.3** *Errores De acuerdo con el paradigma de Neyman-Pearson, podemos incurrir en dos clases de errores.*

1. Uno de ellos es el llamado error de tipo I, la probabilidad de rechazar la hipótesis nula cuando era verdadera. Si observamos la definición de nivel de significación  $\alpha$ , este es el error de tipo I.
2. El error de tipo II es la probabilidad de aceptar la hipótesis nula cuando es falsa, y se llama  $\beta$ . Para ser calculado hay que proponer una alternativa completamente identificable.

#### Potencia

El valor  $(1 - \beta)$  es la probabilidad de que  $H_0$  sea rechazada cuando es falsa, y se llama potencia del test. Permite comparar dos tests diferentes para resolver el mismo problema, el que tiene potencia mayor es el mejor.

Un test ideal debería tener los errores  $\alpha$  y  $\beta$  iguales a cero. Pero eso no ocurre salvo en casos triviales. Es más, en la práctica, ocurre casi siempre que para un tamaño de muestra fija, achicar el  $\alpha$  implica agrandar el  $\beta$ , por lo cual si se quiere un nivel de significación mayor termina obteniéndose un test menos potente, un test con menos capacidad de rechazar cuando en realidad debería haber rechazado.

**Ejemplo 8.3.3** Una señora afirma que puede distinguir en su té con leche si en la taza se ha vertido antes la leche o la infusión de té. La señora no dice que sea posible clasificar en todos los casos sin error las tazas de té; afirma únicamente que le es posible distinguirlas frecuentemente por el sabor de la mezcla. Se realiza un experimento para comprobar esta hipótesis que consiste en preparar veinte tazas de té por ambos procedimientos (diez de cada forma), presentadas al azar para su clasificación. Si la señora clasifica correctamente 16 de las 20. ¿Que puede decirse de la afirmación de la señora? ¿Puede explicarse esta diferencia como debida a las fluctuaciones del azar?.

#### Resolución

Si la señora adivina totalmente al azar, la cantidad de tazas correctamente adivinadas sería la mitad de la muestra, y si tuviera un extraordinario sentido del gusto, debería ser mayor que la mitad. Transformado en proporciones, las hipótesis a testear serían

$$H_0 : p = 0,5 \quad H_A : p > 0,5$$

El estadístico del test es el número de aciertos  $X$  en la muestra, la cual es una variable binomial con parámetros  $n = 20$  y  $p = 0,5$  bajo hipótesis nula. El  $P$  valor sería el valor de la probabilidad de observar un valor tan o más grande que 16 de 20, si vale la hipótesis nula,

$$P(X \geq 16) = P(16) + P(17) + P(18) + P(19) + P(20) = 1 - 0,994 = 0,006$$



Este valor es muy chico, por lo cual hay evidencia para rechazar la hipótesis nula y darle crédito a la señora por un extraordinario sentido del gusto.

Supongamos que antes de realizar la experiencia, se decide que la decisión a favor de la señora se va a tomar si el  $P$ -valor resulta menor o igual a 0.06. ¿Cuáles serían los valores posibles para que la hipótesis se rechace?

Este rango de valores se llama región de rechazo, y en este caso,  $X$  puede ser mayor o igual que 14 aciertos.

$$RR = \{14, 15, 16, 17, 18, 19, 20\}$$

pues  $P(X \geq 14) = 0,058$  bajo hipótesis nula.

La potencia del test se puede calcular solo cuando fijamos una alternativa que identifique la distribución, es decir, si decimos que la alternativa es  $p = 0,6$ , entonces

$$1 - \beta = 1 - P(X \geq 14 | p = 0,6) = 0,75$$

Observemos que la región de rechazo la fijé bajo hipótesis nula, pero la potencia mira cuán probable es obtener valores en esa región cuando la alternativa vale.

Ya dijimos que cuando  $\alpha$  disminuye  $\beta$  crece por lo cual no podemos tener totalmente controlados los dos errores. ¿Cuál es potencialmente más peligroso? Si tomamos como ejemplo de hipótesis nula **el sujeto es inocente** contra la hipótesis alternativa **el sujeto es un asesino** el peor error es el de condenar a un inocente, el error de tipo I. ¿Qué ocurre si alternamos las hipótesis?, el peor error es el de tipo II. ¿Qué diferencia hay entonces en el planteo de las hipótesis?. Esta no es una decisión matemática, y muchas veces la elección depende de la conveniencia y costumbres de aplicación. Algunas observaciones que podemos hacer son las siguientes:

1. En el ejemplo del ajuste de modelo a la emisión de partículas alfa, elegimos como hipótesis nula que la distribución del modelo es Poisson, y como alternativa que la distribución del modelo no sea Poisson. La hipótesis nula es más simple que la alternativa, que contiene muchas más distribuciones posibles y es una convención que la hipótesis nula sea la más simple de las dos hipótesis.
2. Las consecuencias de rechazar erróneamente alguna de las hipótesis puede ser peor que la otra. En ese caso, la hipótesis nula debe ser la hipótesis cuyo error queremos controlar, pues  $\alpha$ , el error de rechazar erróneamente la hipótesis nula va a ser controlado en un test de nivel  $\alpha$ .
3. En muchas investigaciones científicas, la hipótesis nula simplemente es una aseveración que quiere ser desacreditada para demostrar la presencia de un fenómeno físico o efecto. El test de la señora y el te con leche es uno de esa clase, la hipótesis nula dice que la señora está adivinando, y la hipótesis alternativa que tiene un gran sentido del gusto. La hipótesis nula debe ser convenientemente desacreditada para poder convencer al escéptico de que en realidad hay un efecto.

**Ejemplo 8.3.4** Un investigador ha preparado el nivel de dosificación de un fármaco que afirma provocará sueño en por lo menos 80% de las personas que padecen insomnio. Después de examinar la dosificación, se considera que su afirmación acerca de la efectividad del fármaco es exagerada. En un intento por refutar su afirmación se administra la dosificación prescrita a 20 personas que padecen insomnio, y se observa  $Y$ , el número de personas que se adormecen debido al fármaco. Se desea probar la hipótesis  $H_0 : p = 0,8$  frente a la alternativa  $H_a : p < 0,8$ . Suponga que se utiliza la región de rechazo  $\{y \leq 12\}$ .

- (a) Encuentre  $\alpha$ .
- (b) Encuentre  $\beta$  para  $p = 0,6$ .
- (c) Encuentre  $\beta$  para  $p = 0,4$ .

**Resolución**

En este ejemplo el estadístico del test es  $Y$ =número de personas que se adormecen debido al fármaco, la hipótesis nula es  $p=0.8$  y la alternativa es  $p \leq 0.8$  pues se quiere probar que la afirmación es falsa. El error de tipo I,  $\alpha$  es la probabilidad de que  $Y$  esté en la región de rechazo dado que vale la hipótesis nula,  $p=0.8$ . En este caso la variable  $Y$  es binomial con probabilidad de éxito  $p=0.8$  y número de pruebas 20, por lo cual:

$$\alpha = P(Y \leq 12/p = 0.8) = \sum_{i=0}^{12} \binom{20}{i} 0.8^i 0.2^{20-i} = 0.032$$

En el punto b) se pregunta cual es el error de tipo II cometido si la alternativa es  $p = 0.6$ . Aquí suponemos que  $Y$  es binomial con probabilidad de éxito  $p = 0.6$  y número de pruebas 20, por lo cual la probabilidad de que  $Y$  esté en la región de aceptación es:

$$\beta = P(Y > 12/p = 0.6) = 1 - P(Y \leq 12/p = 0.6) = 1 - 0.584 = 0.416.$$

En el punto c) se supone que  $Y$  es binomial con probabilidad de éxito  $p = 0.4$  y número de pruebas 20, por lo cual la probabilidad de que  $Y$  esté en la región de aceptación es:

$$\beta = P(Y > 12/p = 0.4) = 1 - P(Y \leq 12/p = 0.4) = 1 - 0.976 = 0.024.$$

## 8.4. $z$ test de nivel $\alpha$

Ya hemos visto como calcular un  $P$  valor para la media si la desviación estandar es conocida. Supongamos que requerimos por anticipado un nivel de significación  $\alpha$ . En términos del  $P$ -valor, el resultado del test es significativo con nivel  $\alpha$  si  $P \leq \alpha$ . Sin embargo no es necesario calcular el  $P$ -valor para dar esta respuesta.

**Definición 8.4.1**  $z$  test para la media.

Sean  $X_1, \dots, X_n$  una muestra aleatoria de variables independientes e idénticamente distribuidas con media  $\mu$  y varianza  $\sigma^2$ , ésta última conocida. Si  $n$  es suficientemente grande para que valga el TCL, el estadístico

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

tiene distribución normal estándar. Sea  $z_{obs}$  el valor observado del estadístico en la muestra. Sea  $H_0 : \mu = \mu_0$ , la hipótesis se rechaza a favor de las alternativas

$$H_a : \mu > \mu_0 \quad \text{si} \quad z_{obs} \geq z_{(1-\alpha)}$$

$$H_a : \mu < \mu_0 \quad \text{si} \quad z_{obs} \leq z_{\alpha}$$

Se rechaza  $H_0$  a favor de la alternativa a dos colas

$$H_a : \mu \neq \mu_0 \quad \text{si} \quad |z_{obs}| \geq z_{(1-\alpha/2)}$$

**Observación 8.4.1** Por supuesto que si la muestra tiene distribución normal, no necesito tamaño de muestra grande para que la media muestral estandarizada tenga distribución normal patrón. El test  $z$  es un test exacto para muestra normal, y asintótico (vale para muestra grande) en caso de que no sepa que distribución tiene mi muestra.

**Ejemplo 8.4.1** Un fabricante de automóviles afirma que un 0km de determinado modelo tiene un rendimiento medio de, al menos, 12 km por litro de nafta. Un grupo de investigación de mercados, sospechando de tal afirmación, realiza un test. Se considera que la distancia recorrida por litro tiene distribución normal con desviación estándar  $\sigma = 2$ , y se toma una muestra aleatoria de 30 automóviles.

- (a) Hallar la región de rechazo del test para el nivel  $\alpha = 0,02$ .
- (b) Hallar la probabilidad de error de tipo II si el rendimiento real es de 11 km por litro.
- (c) Tome una decisión si el valor observado promedio es 11,32.

**Resolución**

Tenemos una muestra de variables normales con media desconocida y desviación estándar  $\sigma = 2$ . Tenemos que hacer inferencia sobre la media de la distribución, y queremos desacreditar al fabricante, por lo cual la hipótesis nula va a ser la tesis del fabricante  $H_0 : \mu = 12km$  y la alternativa es que hace menos de 12km por litro,  $H_A : \mu < 12km$ . Podemos aplicar el test  $Z$  con nivel  $\alpha = 0,02$ . La región de rechazo con ese nivel es el conjunto

$$\{\bar{x}_{obse} \mid \frac{(\bar{x}_{obse} - \mu_0)}{\sigma/\sqrt{n}} < z_\alpha\}$$

, pues

$$P(Z = \frac{(\bar{X} - \mu_0)}{\sigma/\sqrt{n}} < z_\alpha) = \alpha$$

por definición de percentil. Si  $\alpha = 0,02$ , entonces  $z_\alpha = -2,0537$  y la región de rechazo es

$$\{z \mid z < -2,0537\}$$

como vemos en la figura 8.1 Podemos traducir la misma región a valores de  $\bar{X}$ ,

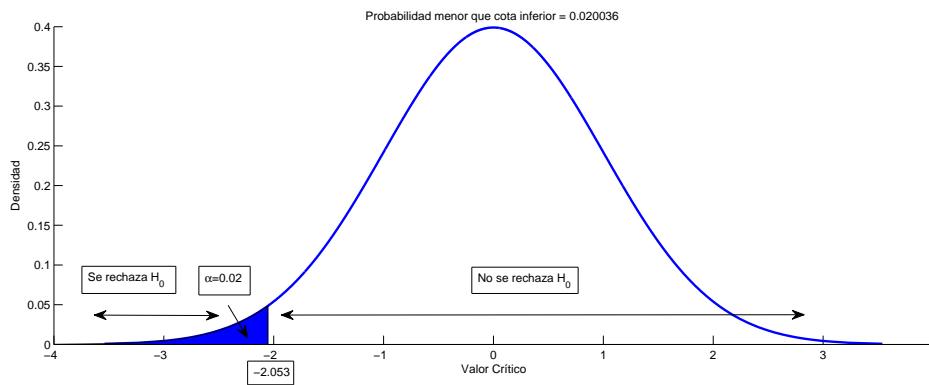


Figura 8.1: Región de rechazo en  $Z$  de un test a cola izquierda

$$\{\bar{x}_{obse} \mid \bar{x}_{obse} < -2,0537 \frac{2}{\sqrt{30}} + 12 = 11,25\}$$

Como

$$\bar{x}_{obse} = 11,32$$

no puedo rechazar la hipótesis con ese nivel de error.

El error de tipo II,  $\beta$  es la probabilidad de no rechazar cuando debía haber rechazado la hipótesis nula.

$$\begin{aligned} \beta &= P(\bar{X} > 11,25 \mid \mu = 11km) = P\left(\frac{\bar{X} - 11}{2/\sqrt{30}} > \frac{11,25 - 11}{2/\sqrt{30}}\right) \\ &= P\left(Z > \frac{0,25\sqrt{30}}{2}\right) \\ &= P(Z > 0,684) = 0,2469 \end{aligned}$$

La potencia de este test es  $1 - \beta = 1 - 0,2469 = 0,75$  bastante alta, por lo cual tenemos confianza de que este test va a rechazar la hipótesis nula cuando realmente valga la hipótesis alternativa.

**Ejemplo 8.4.2** Un contratista encarga un gran número de vigas de acero con longitud promedio de 5 metros. Se sabe que la longitud de una viga se halla normalmente distribuida con una desviación estándar de 0.02 metros. Usualmente, después de recibir un embarque, el contratista selecciona 16 vigas al azar y mide sus longitudes para decidir si acepta o rechaza el encargo.

- (a) Si la probabilidad de rechazar un embarque bueno es 0.04, ¿cuál debe ser el rango de valores de la media muestral para que el embarque sea regresado al fabricante?
- (b) Si la longitud promedio real es de 4.98 metros, ¿cuál es la potencia del test realizado?

### Resolución

Primero debemos plantear el test de hipótesis. Si el contratista desea saber si el lote que recibió corresponde con el estándar de calidad que pidió, debe corroborar que la media de la población de barras es realmente  $\mu = 5m$ . La alternativa es a dos colas, barras con media menor o mayor no pueden ser aceptadas.

$$H_0 : \mu = 5m \quad H_A : \mu \neq 5m$$

El error de quedar mal con el fabricante devolviendo algo que si alcanzaba el estándar es de  $\alpha = 0,04$ , bastante chico. El problema es que si la potencia es baja, puede no rechazarse el cargamento porque no hay evidencia suficiente en la muestra, no porque el cargamento alcance el estándar.

La región de rechazo del test es el rango de valores de la media muestral para los cuales el embarque va a ser regresado al fabricante. Como la hipótesis es a dos colas, el percentil es  $z_{(1-0,04/2)} = 2,053$

$$0,04 = P(|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}| > z_{(1-0,04/2)})$$

y

$$\begin{aligned} RR &= \{\bar{x}_{obse} \mid |\frac{\bar{x}_{obse} - \mu_0}{\sigma/\sqrt{n}}| > z_{(1-0,04/2)}\} \\ &= \{\bar{x}_{obse} \mid \frac{\bar{x}_{obse} - \mu_0}{\sigma/\sqrt{n}} > z_{(1-0,04/2)}\} \cup \{\bar{x}_{obse} \mid \frac{\bar{x}_{obse} - \mu_0}{\sigma/\sqrt{n}} < -z_{(1-0,04/2)}\} \\ &= \{\bar{x}_{obse} \mid \bar{x}_{obse} > z_{(1-0,04/2)}\sigma/\sqrt{n} + \mu_0\} \cup \{\bar{x}_{obse} \mid \bar{x}_{obse} < -z_{(1-0,04/2)}\sigma/\sqrt{n} + \mu_0\} \\ &= \{\bar{x}_{obse} \mid \bar{x}_{obse} > 5,01\} \cup \{\bar{x}_{obse} \mid \bar{x}_{obse} < 4,99\} \end{aligned}$$

Por lo cual el embarque va ser regresado al fabricante si el promedio de los largos de las vigas de la muestra extraída del lote es mayor a 5.01 o menor a 4.98. En la figura 8.2 vemos la región de rechazo sobre la normal patrón.

Si la distribución de los largos de las vigas del lote es normal pero con media  $\mu = 4,98m$  en vez de  $5m$ , el test realizado usando esta alternativa

$$H_0 : \mu = 5m \quad H_a : \mu = 4,98$$

tiene potencia

$$\begin{aligned} 1 - \beta &= 1 - P(4,99 \leq \bar{X} \leq 5,01 \mid \mu = 4,98) \\ &= 1 - P(\frac{4,99 - 4,98}{0,02/\sqrt{16}} \leq \frac{\bar{X} - 4,98}{0,02/\sqrt{16}} \leq 5,9) \\ &= 1 - P(1,2 \leq Z \leq 5,9) \\ &= 1 - 0,02275 = 0,97725 \end{aligned}$$

La potencia del test es alta, por lo cual hay alta confianza en que el test rechace si la verdadera media del lote es 4.98. La figura 8.3 muestra el error de tipo II,  $\beta$  para este test. En rojo está dibujada la densidad bajo hipótesis nula, y en azul la alternativa. El error es la intersección del área bajo la curva alternativa que entra dentro del área correspondiente a la región de aceptación de la hipótesis nula. Vemos que pocas muestras bajo la alternativa pueden ser confundidas como valederas bajo hipótesis nula, de allí el nombre de  $(1 - \beta)$  como potencia del test.

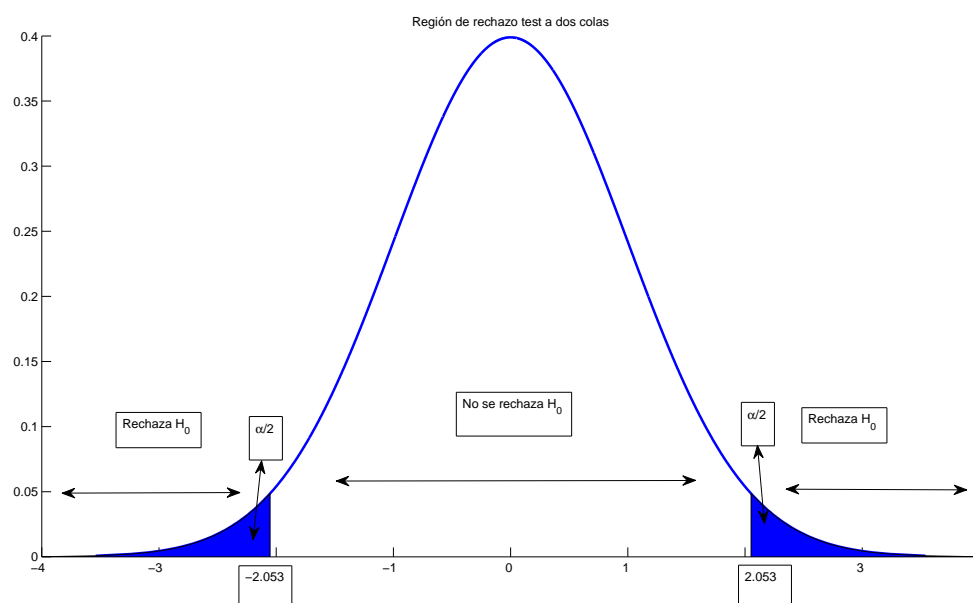


Figura 8.2: test a dos colas

## 8.5. Abuso de los test de hipótesis

Los test de hipótesis son usados en forma rutinaria como respaldo científico para rechazar o aceptar afirmaciones, y evaluar resultados de investigación en agricultura, educación ingeniería, medicina, sicología, sociología entre otros. En forma cotidiana escuchamos en la radio y televisión comentarios basados en estadística descriptiva, así como análisis y editoriales en internet y diarios. Sin embargo toda herramienta usada rutinariamente es a menudo usada sin pensar. En esta sección nos gustaría invitar a la reflexión sobre posibles abusos de esta herramienta.

### 8.5.1. Inferencia estadística no corrige faltas de diseño

Experimentos mal diseñados no producen inferencias válidas. A menudo solo producen resultados que no dicen nada sobre la hipótesis a testear. Es por ello que intentar hacer inferencia sobre un grupo de datos obtenido, por más costosa que haya sido su obtención, sin haber realizado un diseño previo acorde al problema, resulta en resultados erróneos o ningún resultado.

Demos un ejemplo simple de esto. Se dice que los pre-adolescentes que estudian inglés en academias particulares tienen muy buenas notas en castellano en sus escuelas respectivas. Pero esta aseveración sale de que el efecto de estudiar inglés en forma particular se confunde con el efecto de decidir ir a una academia o no. Puede ser que los chicos que tienen muy buenas notas en castellano elijan ir a la academia de inglés (o sus padres elijan enviarlos), y los otros no, por lo cual no es el estudio de un idioma extranjero lo que mejora su castellano, o si?. Esta evidencia solo dice que la diferencia en las notas en castellano entre grupos es más alta que la que ocurriría al azar. Pero no dice nada de que la causa. Un estudio comparativo aleatorizado dejaría aislado el efecto del estudio del inglés y daría significado estadístico a esta diferencia.

Un problema aun más complejo se observa en los estudios largos que involucran la toma de pastillas o la asistencia a centros hospitalarios. A lo largo del estudio los pacientes dejan de participar, lo cual separa a la muestra inicial en dos grupos, los que adhieren al protocolo y los que no. A pesar de que el diseño inicial fue bien realizado, no hay forma de evitar la no adherencia al protocolo, pues es una decisión de los pacientes, y en los casos de pacientes terminales, se tienen datos sobre la supervivencia de los pacientes en

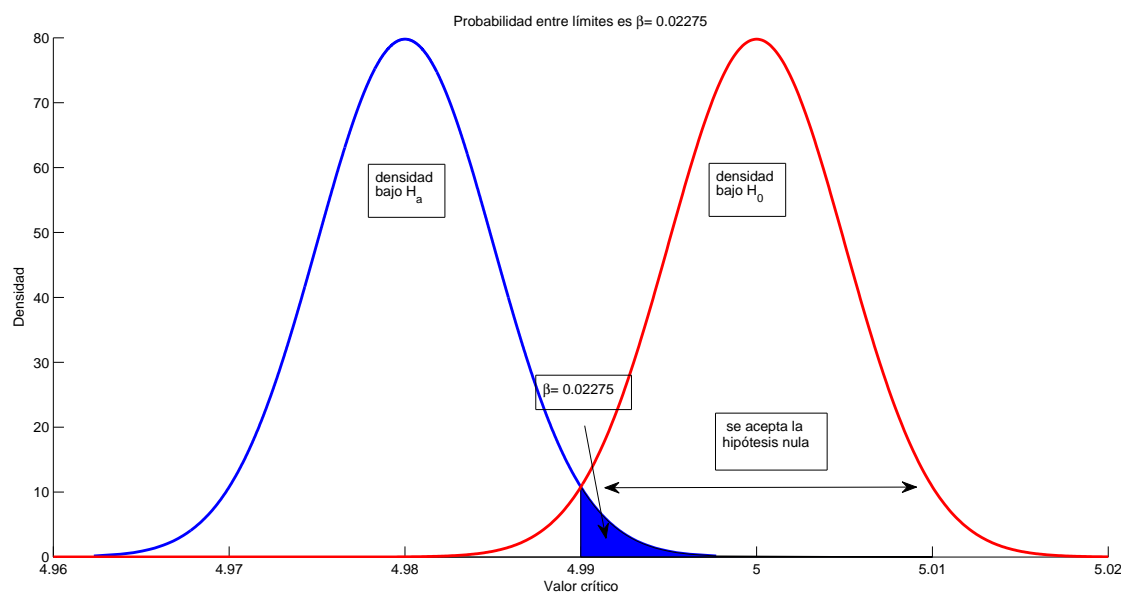


Figura 8.3: Error de tipo II,  $\beta$ , para un test de dos colas.

cada muestra. Es un error grave tratar estos estudios sin considerar el factor de confusión que representa la adherencia al protocolo, pues los pacientes que cumplen todo el tratamiento pueden ser personas que se cuidan mas y por eso tienen mejor salud, no porque la droga que tomen haga algún efecto. El análisis de supervivencia, y el manejo de factores de confusión es un área muy activa de investigación en estadística, dado que en el área de salud pública es muy difícil mantener el diseño original del estudio, y no es trivial sacar conclusiones de muestras con datos faltantes.

Los test de hipótesis y los intervalos de confianza se basan en leyes de probabilidad. La aleatorización del muestreo o la experimentación aseguran que esas leyes se aplican. Cuando esas estrategias estadísticas para producir datos no pueden ser usadas, estadística inferencia formal debe hacerse con precaución. Es usual ver test de hipótesis realizados sobre muestras no elegidas aleatoriamente, y es cierto que un resultado estadísticamente significativo al menos dice que hay un efecto mayor que el se hubiera visto solo al azar. Sin embargo, esto solo es poca evidencia contra  $H_0$  en favor de la hipótesis en investigación  $H_A$ . El sentido común no debe ser nunca reemplazado por un  $P$ -valor.

### 8.5.2. Cuidado con la búsqueda de valores significativos

Significado estadístico en un estudio es algo muy buscado por investigadores de todas las áreas. Quiere decir, o al menos debería decir, que se ha encontrado el efecto que se estaba buscando. El razonamiento detrás del significado estadístico funciona bien cuando si se decide que efecto se esta buscando, se diseña un experimento o se realiza una muestra para buscarlo, y se usa un test de hipótesis para comparar la evidencia que se obtiene. Pero como una búsqueda exitosa de evidencia sobre un fenómeno a menudo termina con un estudio de significado estadístico, mas a menudo aún lo que se busca resulta el significado estadístico. Hay muchas formas de hacer esto, algunas veces se a realizado en forma muy cándida, y otras a sido claramente por diseño.

Una táctica es hacer muchos test de hipótesis sobre el mismo grupo de datos. William Feller, en 1969, comentaba el siguiente ejemplo. Tres psiquiatras estudiaron una muestra de personas esquizofrénicas y una muestra de personas no esquizofrénicas. Midieron 77 variables por cada persona, como religión, antecedentes familiares, experiencias de la infancia, entre otras. Ellos querían descubrir como distinguir personas que

después van a convertirse en esquizofrénicas. Habiendo medido 77 variables, se hicieron 77 tests separados para las diferencias entre los dos grupos de sujetos. Los científicos encontraron dos test significativos con nivel del 5 % (de los 77 que hicieron) e inmediatamente publicaron sus resultados. El estadístico usado, basado en las 77 variables, detectaba diferencias entre los grupos!! De allí a descubrir exactamente el patrón que marcaba a los esquizofrénicos había un trecho, pero era un avance.

Sin embargo, por mas excitante que sea estar cerca de un descubrimiento de esta naturaleza, es necesario meditar sobre el procedimiento realizado. Un test de significación de nivel 0.05 implica que se espera un resultado errado 5 veces de 100, siendo verdadera la hipótesis nula, que dice que las variables no distinguen nada!! Por lo cual, hacer 77 test y encontrar 2 que rechazan la hipótesis nula no es evidencia alguna de que la hipótesis nula sea falsa. Si se hace un solo test de nivel 0.05 y se rechaza la hipótesis, entonces se encontró algo. Si hacen 77 y se rechazan 2 veces, no se encontró nada, solo se vio que el mecanismo de testeo funciona bien. Imagínense si se hacen tantos test como sean necesarios para encontrar la muestra que rechaza la hipótesis nula, y solo se reporta ese caso. Cualquier tesis seria verdadera, hasta la mas absurda.

Por eso los investigadores fundamentan sus hallazgos con algo más que un p-valor, un modelo, una simulación, un estudio a larga escala. El problema es que, dada la velocidad de las computadoras actuales, y los programas de estadística cada vez más completos, es muy fácil hacer experimentos en larga escala y hacer toda clase de test y operaciones a los datos. Los resultados no van a diferir mucho del de los 77 test realizados por los investigadores. En todo grupo de datos suficientemente grande va a haber un patrón inusual, y cuando se testee por ese patrón inusual, el resultado va a ser estadísticamente significativo. Y en realidad no va a significar nada.

No se puede pasar por encima de la lógica de la inferencia. Es convincente hacer hipótesis de que un efecto está presente, hace un estudio para buscarlo y encontrarlo con un nivel de significación bajo. No es convincente buscar algún efecto o patrón cualesquiera, y encontrar algo.

Tampoco esto quiere decir que buscar patrones está prohibido, o no es un trabajo científico apropiado. Lo es testear una hipótesis sobre la población con una muestra donde se sabe que hay un patrón. Si una muestra generó un patrón, hay que buscar otra muestra aleatoriamente, si es posible grande, y testear en ella por el patrón. Si en esta nueva muestra el patrón sigue estando, entonces se encontró algo.





## Capítulo 9

# Inferencia sobre distribuciones normales

Tanto intervalos de confianza como test de significación para la media  $\mu$  de una población normal se basan en la media muestral  $\bar{X}$  que estima el parámetro poblacional  $\mu$ . La distribución muestral de  $\bar{X}$  depende de  $\sigma$ . Esto no presenta ningún problema cuando  $\sigma$  es conocido, lo cual deriva en los llamados  $z$ -test que vimos en el capítulo anterior. Pero cuando  $\sigma$  es desconocido, se debe estimar  $\sigma$  aunque uno esté interesado en hacer inferencia sobre  $\mu$ . La desviación estándar muestral  $S$  es el estimador que será usado en la derivación de los test de hipótesis para la media con varianza desconocida.

### 9.1. Procedimientos basados en la distribución $t$

Supongamos que tenemos una muestra aleatoria simple de la distribución normal con media  $\mu$  y varianza  $\sigma$ . La media muestral  $\bar{X}$  tiene media  $\mu$  y varianza  $\sigma/\sqrt{n}$ . Cuando  $\sigma$  es desconocido la desviación estándar de  $\bar{X}$  puede estimarse por  $S/\sqrt{n}$ . Esta cantidad es llamada el error estándar de la media muestral  $\bar{X}$ .

**Definición 9.1.1** Cuando la desviación estándar de un estadístico se estima usando los datos, el resultado se llama error estándar del estadístico.

La media muestral estandarizada es

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

es la base los  $Z$  test cuando  $\sigma$  es conocida. Este estadístico tiene distribución normal patrón,  $N(0, 1)$ . Cuando sustituimos la desviación estándar  $\sigma/\sqrt{n}$  por el error estándar  $S/\sqrt{n}$ , el estadístico resultante no tiene distribución normal.

**Definición 9.1.2** Si  $X_1, \dots, X_n$  es una muestra aleatoria de la distribución normal  $N(\mu, \sigma^2)$ , entonces la distribución del estadístico

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

es la distribución  $t$  de Student, con  $n - 1$  grados de libertad.

Para estudiar hipótesis con respecto a la media de poblaciones normales con media y varianza desconocidas, vamos a usar el estadístico  $T$  en vez de  $Z$ , y usar percentiles de la distribución  $T$  en vez de la normal patrón.

**Definición 9.1.3** Para testear las hipótesis

$$H_0 : \mu = \mu_0$$

basados en una muestra de tamaño  $n$ , se usa el estadístico

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

y se computa su valor observado en la muestra  $t_{obse}$ . En término de la variable  $T$  que tiene distribución  $t$  de Student con  $n - 1$  grados de libertad, el  $P$ -valor del test de  $H_0$  versus

$$H_a : \mu > \mu_0 \quad \text{es} \quad P(T \geq t_{obse})$$

$$H_a : \mu < \mu_0 \quad \text{es} \quad P(T \leq t_{obse})$$

$$H_a : \mu \neq \mu_0 \quad \text{es} \quad P(|T| \geq |t_{obse}|)$$

Estos  $P$ -valores son exactos si la distribución es normal y aproximados para  $n$  grande en otros casos.

**Ejemplo 9.1.1** Volviendo al ejemplo de los datos de Newcomb, deseábamos saber si el valor moderno 33.02 era significativamente diferente de los datos obtenidos en 1882. Para contestar esta pregunta debemos realizar un test de hipótesis.

$$H_0: \mu = 33,02 \quad \text{vs} \quad H_A : \mu \neq 33,02$$

El valor observado de  $T$  es

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{27,75 - 33,02}{5,083/\sqrt{64}} = -8,29$$

El  $P$ -valor es

$$P(|T| \leq 8,29)$$

donde  $T$  tiene distribución  $t$  de Student con 63 grados de libertad. La distribución  $t$  es simétrica y bastante parecida a la normal, el valor de esta probabilidad como puede verse en la figura 9.1, es extremadamente bajo, casi 12 ceros después de la coma,  $1,110889158439932e - 011$ . La evidencia en contra de la hipótesis nula es importante, y en

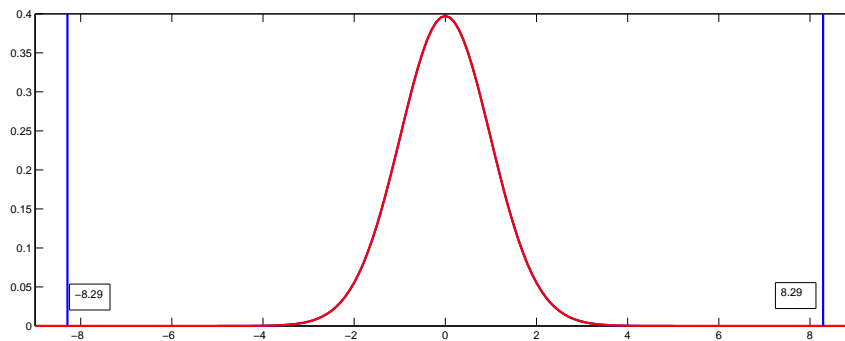


Figura 9.1:  $P$ -valor de los datos de Newcomb.

este caso, muestra aun mas que el procedimiento de Newcomb tiene un sesgo sistemático que no puede removerse estadísticamente.

### 9.1.1. Muestras Apareadas

Newcomb buscaba estimar una constante física, por lo cual se encontraba en estudio una sola población. Sin embargo en muchos otros casos es importante considerar dos poblaciones para evitar riesgos de factores de confusión. En los estudios apareados, los sujetos son asignados en pares y los resultados son comparados con su compañero en el par. El investigador puede tirar una moneda para designar que sujeto va a tratamiento y que sujeto va al grupo de control por cada par. Una situación de comparación apareada es la de mediciones antes y después del mismo sujeto.

**Ejemplo 9.1.2** El Fondo de apoyo a las humanidades, apoya cursos de verano para mejorar los conocimientos de los profesores de idiomas extranjeros. Presentamos aquí datos de un instituto que recibió 20 profesores de francés por 4 semanas. En el comienzo del período los profesores tomaron un test de comprensión de francés hablado. Después de 4 semanas de inmersión en francés dentro y fuera de clases, se les tomó un nuevo test, modificado para que el hecho de tomar el primer test no diera ventajas sobre el segundo. La tabla siguiente da los valores del primer y segundo test. El máximo valor posible del test es 36.

Para analizar estos datos, primero substraemos el primer test de el segundo test para obtener la mejoría de cada estudiante. Estas diferencias forman una sola muestra.

Docente	1er test	2do test	diferencia	docente	1er test	2do test	diferencia
1	32	34	2	11	30	36	6
2	32	31	0	12	20	26	6
3	29	35	6	13	24	27	3
4	10	16	6	14	24	24	0
5	30	33	3	15	31	32	1
6	33	36	3	16	30	31	1
7	22	24	2	17	15	15	5
8	25	28	3	18	32	34	2
9	32	26	-6	19	23	26	3
10	20	26	6	20	23	26	3

Cuadro 9.1: Tabla de diferencias en los test de idioma francés.

Para determinar si el curso de verano produjo una mejoría en los niveles de comprensión del francés hablado, se debe realizar un test de hipótesis:

$$H_0 : \mu = 0 \quad \mu_a > 0$$

El parámetro  $\mu$  representa la media que se alcanzaría si la de la población completa de profesores de Francés asistiera a los cursos de verano. La hipótesis nula dice que se produce ninguna mejoría, mientras que la hipótesis alternativa dice que los puntajes medios del segundo test son mayores en promedio. Las 20 diferencias tienen

$$\bar{x} = 2,5 \quad y \quad s = 2,89$$

El estadístico de test  $t_{obse}$  es

$$t_{obse} = \frac{\bar{x} - 0}{s/\sqrt{n}} = \frac{2,5}{2,89/\sqrt{20}} = 3,87$$

El  $P$ -valor obtenido, suponiendo una distribución  $t$  con 19 grados de libertad, es de  $P = 0,000515$ . No es muy probable observar solo por casualidad la mejoría observada en los puntajes de los profesores, por lo cual hay evidencia de que asistir a cursos de verano mejora la comprensión del francés hablado.

Veamos que intervalo de confianza podemos calcular para la media. Un intervalo de confianza del 90 % requiere un valor crítico que deje 0.05 de área en la cola derecha de la distribución  $t$  con 19 grados de libertad.  $t_{(1-\alpha/2)} = 1,729$ , por lo cual

$$\left[ \bar{x} - t_{(1-\alpha/2)} \frac{s}{\sqrt{n}}, \bar{x} + t_{(1-\alpha/2)} \frac{s}{\sqrt{n}} \right] = \left[ 2,5 - 1,729 \frac{2,89}{\sqrt{19}}, 2,5 + 1,729 \frac{2,89}{\sqrt{19}} \right] = [2,5 - 1,12, 2,5 + 1,12] = [1,38, 3,62]$$

La mejoría es significativa pero bastante chica.

## 9.2. Comparando dos medias

Los siguientes son ejemplos de comparación de poblaciones

- Un medico investigador está interesado en estudiar el efecto de agregar calcio a la dieta en la presión sanguínea. Se realiza un experimento de comparación aleatorizado en donde uno de los grupos recibe un suplemento de calcio y otro de los grupos toma un placebo.
- Un psicólogo desarrolla un test que mide la habilidad intuitiva con la que un sujeto obtiene información general de otras personas. El investigador quiere comparar dichas habilidades de percepción en hombres y mujeres, y para ello le realiza el test a dos grandes grupos de estudiantes universitarios, uno de hombres y otro de mujeres.
- Un educador considera que ciertas actividades extra-curriculares de lectura mejoran la habilidad de lectura de los niños de segundo grado. Para evaluar esta hipótesis planea realizar un estudio con dos grupos de niños, en los cuales uno de ellos va a recibir la curricula general y el otro va a tener las actividades extra-curriculares en estudio.

Estos son algunos ejemplos de problemas que involucran comparaciones entre dos poblaciones. Para realizar inferencias, las poblaciones deben ser comparables, y las muestras obtenidas de cada población deben ser representativas.

### Definición 9.2.1 Problemas de dos muestras

- *La estadística inferencial de dos poblaciones busca comparar respuestas a estímulos en dos grupos.*
- *Cada grupo debe ser considerado una muestra aleatoria de las posibles respuestas de toda la población considerada.*
- *Las respuestas de cada grupo son independientes de las respuestas del otro grupo.*

Un problema de dos muestras puede derivar de un experimento aleatorizado que divide aleatoriamente los sujetos en dos grupos y expone a cada uno de ellos a un tratamiento diferente. La comparación de dos muestras independientes seleccionadas de dos poblaciones es también un problema de dos muestras. La diferencia con las muestras apareadas es principalmente que ambas muestras son independientes. En el caso apareado no lo son, y por eso se realiza inferencia de una sola muestra.

Para realizar inferencia, hay que tener alguna idea de la distribución de la población, para poder decidir el estadístico a usar y su distribución muestral. En el caso de que la distribución de las poblaciones sean simétricas, y especialmente cuando son normales, la inferencia se realizará a través de la estimación de la media de ambas poblaciones. Las poblaciones normales son diferentes si sus medias son diferentes y/o sus varianzas son diferentes.

Vamos a establecer una notación para las muestras y estadísticos. Su media de la población 1 va a ser  $\mu_1$ , y su varianza  $\sigma_1^2$ . Su muestra aleatoria es  $X_1^1, \dots, X_{n_1}^1$ , con tamaño de muestra  $n_1$ , su media muestral  $\bar{X}_1$  y desviación estándar muestral  $S_1$ . La notación para la segunda población es media  $\mu_2$ , varianza  $\sigma_2^2$ , tamaño de muestra  $n_2$ , media muestral  $\bar{X}_2$  y desviación estándar muestral  $S_2$ .

Se quiere comparar las medias, ya sea mediante un intervalo de confianza para  $\mu_1 - \mu_2$  o testeando la hipótesis de diferencia nula.

$$H_0 : \mu_1 = \mu_2$$

la cual es equivalente a decir que

$$H_0 : \mu_1 - \mu_2 = 0$$

### 9.2.1. El z-test para dos muestras: $\sigma_1$ y $\sigma_2$ conocidos

Usualmente, para realizar inferencia necesitamos un estadístico con distribución conocida bajo hipótesis nula. El estadístico natural para estimar  $\mu_1 - \mu_2$  es  $\bar{X}_1 - \bar{X}_2$ , es un estimador insesgado y consistente, es decir, la esperanza del estimador es

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$$

y se acerca a ese valor cuando el tamaño de las muestras tienden a infinito. También, como las muestras son independientes, las variables  $\bar{X}_1$  y  $\bar{X}_2$  también lo son, por lo cual la varianza de  $\bar{X}_1 - \bar{X}_2$  es

$$Var(\bar{X}_1 - \bar{X}_2) = Var(\bar{X}_1) + Var(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Si las distribuciones poblacionales son normales, entonces la distribución de  $\bar{X}_1 - \bar{X}_2$  también lo es, con parámetros  $\mu_1 - \mu_2$  y  $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ . En el caso de que las muestras no sean normales, pero  $n_1$  y  $n_2$  sean grandes para que valga la aproximación dada por el teorema central del límite, cada uno de los promedios muestrales será aproximadamente normal e independientes entre sí, por lo cual  $\bar{X}_1 - \bar{X}_2$  tiene también distribución normal aproximada, con parámetros  $\mu_1 - \mu_2$  y  $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ .

**Definición 9.2.2** Supongamos que tenemos dos muestras aleatorias simples independientes entre sí, de dos poblaciones supuestamente distintas, con medias  $\mu_1$ ,  $\mu_2$  y su varianzas  $\sigma_1^2$  y  $\sigma_2^2$ . Entonces un intervalo de confianza para la diferencia de medias de nivel  $(1 - \alpha)$  puede obtenerse usando el hecho de que el estadístico

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - \mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

tiene distribución Normal patrón. Esta distribución es aproximada si las muestras no son normales y  $n_1$  y  $n_2$  es suficientemente grande, y exactas si las muestras son normales. El intervalo de confianza de nivel  $(1 - \alpha)$  es

$$\left[ (\bar{X}_1 - \bar{X}_2) - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} z_{(1-\alpha/2)}, (\bar{X}_1 - \bar{X}_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} z_{(1-\alpha/2)} \right]$$

**Ejemplo 9.2.1** El test de un agujero es un test que mide la facilidad de manipulación de objetos pequeños de aspirantes a un trabajo manual. El test requiere que el aspirante tome un alfiler, lo lleve a un agujero, lo inserte y vuelva por otro alfiler. El puntaje del test es el número de alfileres insertados correctamente en un intervalo de tiempo fijo. En un estudio, mujeres trabajadoras expertas de una compañía son comparadas con estudiantes varones universitarios. Supongamos que los puntajes históricos de las trabajadoras tienen media 120 cada 5 minutos, con una desviación estándar de  $\sigma_1 = 28$ , y que los jóvenes universitarios tienen un puntaje medio de 105 alfileres cada 5 minutos, con una desviación estándar de  $\sigma_2 = 35$ . Si se seleccionan 10 trabajadoras al azar, y 12 jóvenes universitarios, y se comparan los puntajes medios de cada muestra, el estadístico  $\bar{X}_1 - \bar{X}_2$  va a tener media

$$\mu_1 - \mu_2 = 120 - 105 = 15$$

y varianza

$$Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{28^2}{10} + \frac{35^2}{12} = 180,48$$

La desviación estándar de la diferencia de las medias muestrales es

$$\sigma(\bar{X}_1 - \bar{X}_2) = \sqrt{180,48} = 13,43$$

Grupo	$n$	$\bar{x}$	$s$
Trabajadoras	412	186.60	29.15
Estudiantes	750	175.50	31.55

¿Cual es la probabilidad de que 10 muchachos elegidos tengan una media mayor que las 12 mujeres elegidas? Si los puntajes varían normalmente, la diferencia es también normal y estandarizando

$$P((\bar{X}_1 - \bar{X}_2) < 0) = P\left[\frac{(\bar{X}_1 - \bar{X}_2) - 15}{13.43} < \frac{0 - 15}{13.43}\right] = P(Z < -1.12) = 0.1314$$

Supongamos que un sociólogo duda de que la diferencia de medias sea solo de 15 alfileres, y planea un estudio grande, con 750 estudiantes y 412 trabajadoras. Obtiene como resultados los siguientes valores

Como los valores difieren de los históricos, tanto en media como en desviación estándar muestral, es razonable dar un intervalo de confianza para la diferencia de medias usando las desviaciones estándar muestrales en vez de las históricas. El problema es que la distribución del estadístico ya no es normal, si alguna de las muestras es chica. Pero en este caso, la muestra más chica tiene tamaño 412, y el teorema central del límite puede aplicarse, y es razonable considerar al estadístico

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - \mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

con distribución normal patrón. El intervalo de confianza para la diferencia de medias del 99 % es

$$\begin{aligned} & \left[ (\bar{X}_1 - \bar{X}_2) - z_{(1-\alpha/2)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{X}_1 - \bar{X}_2) + z_{(1-\alpha/2)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right] = \\ & = [(186.6 - 175.5) - 2.576 \sqrt{(31.55^2/750) + (29.15^2/412)}, (186.6 - 175.5) + 2.576 \sqrt{(31.55^2/750) + (29.15^2/412)}] \\ & = [6.357, 15.842] \end{aligned}$$

Vemos que no deja afuera al valor 15, con una confianza alta, por lo cual no habría razón pensar que los valores históricos no siguen valiendo.

**Definición 9.2.3** Si se tienen muestras independientes de poblaciones con varianza conocida, y media desconocida, para testear la hipótesis  $H_0 : \mu_1 = \mu_2$ , se computa el valor del estadístico en las muestras

$$z_{obse} = \frac{\bar{x}_1 - \bar{x}_2 - \mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

. En términos del estadístico  $Z$ , con distribución normal patrón, el  $P$ -valor con respecto a

$$\begin{aligned} H_a : \mu_1 > \mu_2 & \quad \text{es} \quad P(Z \geq z_{obse}) \\ H_a : \mu_1 < \mu_2 & \quad \text{es} \quad P(Z \leq z_{obse}) \\ H_a : \mu_1 \neq \mu_2 & \quad \text{es} \quad P(|Z| \geq |z_{obse}|) \end{aligned}$$

Estos  $P$ -valores son exactos si la distribución es normal y aproximados para  $n$  grande en otros casos. Los test de nivel  $\alpha$  para testear la hipótesis  $H_0 : \mu_1 = \mu_2$  rechazan en favor de la alternativa

$$\begin{aligned} H_a : \mu_1 > \mu_2 & \quad \text{si} \quad z_{obse} \geq z_{(1-\alpha)} \\ H_a : \mu_1 < \mu_2 & \quad \text{es} \quad z_{obse} \leq z_{\alpha} \\ H_a : \mu_1 \neq \mu_2 & \quad \text{es} \quad |z_{obse}| \geq z_{(1-\alpha/2)} \end{aligned}$$

En el ejemplo anterior usamos la dualidad entre intervalo de confianza y test de dos colas para ver que  $\mu_1 - \mu_2 = 15$  es un valor posible para la diferencia de muestras, y por lo cual la hipótesis de  $\mu_1 - \mu_2 = 15$  no se puede rechazar con la evidencia de estas muestras. Las muestras eran suficientemente grandes para que pudiese substituirse las varianzas reales por las muestrales, pero esto no ocurre siempre, en las secciones siguientes vamos a discutir cuál sería una mejor aproximación para el estadístico bajo  $H_0$ .

### 9.2.2. El test $t$ para dos muestras: $\sigma_1 = \sigma_2$ desconocido

Supongamos que las poblaciones bajo estudio tienen distribución normal con media posiblemente diferente, y la misma varianza  $\sigma^2$ . Si queremos hacer inferencia sobre las poblaciones, como las normales están caracterizadas por media y varianza, la diferencia entre las poblaciones se reflejará en la diferencia de medias  $\mu_1 - \mu_2$ . Un estimador natural para esta diferencia es la diferencia entre medias muestrales  $\bar{X}_1 - \bar{X}_2$ , que también tiene distribución normal con media  $\mu_1 - \mu_2$  y varianza  $Var(\bar{X}_1 - \bar{X}_2) = \sigma^2(\frac{1}{n_1} + \frac{1}{n_2})$ . Si  $\sigma$  es conocido, entonces

$$Z = \left[ \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{(\frac{1}{n_1} + \frac{1}{n_2})}} \right]$$

tiene distribución normal patrón, como mostramos la sección anterior. Sin embargo, en muy pocos casos se conoce a  $\sigma$  por lo cual debe estimarse de las muestras. Como ambas muestras tienen la misma varianza  $\sigma^2$ , es más razonable estimarlo a partir de la muestra conjunta por

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + ((n_2 - 1)s_2^2)}{n_1 + n_2 - 2}$$

Si reemplazamos a  $\sigma$  por este estimador  $S_p$ , el estimador resultante tiene distribución  $t$  de Student con  $n_1 + n_2 - 2$  grados de libertad.

**Definición 9.2.4** Supongamos que una muestra aleatoria se obtiene de una población normal con media  $\mu_1$  y que otra muestra, independiente de la anterior, se obtiene de otra población normal con media desconocida  $\mu_2$ . Si suponemos también que las dos poblaciones tienen la misma desviación estándar  $\sigma$ , entonces un intervalo de confianza de nivel  $(1 - \alpha)$  para la diferencia de medias  $\mu_1 - \mu_2$  es

$$\left[ (\bar{X}_1 - \bar{X}_2) - t_{(1-\alpha/2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\bar{X}_1 + \bar{X}_2) - t_{(1-\alpha/2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

donde  $t_{(1-\alpha/2)}$  es el percentil  $(1 - \alpha/2)$  de la distribución  $t$  con  $n_1 + n_2 - 2$  grados de libertad.

Para testear hipótesis  $H_0 : \mu_1 = \mu_2$  se computa el valor observado  $t_{obse}$  del estadístico

$$T = \frac{(\bar{X}_1 - \bar{X}_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

que tiene distribución  $t$  con  $n_1 + n_2 - 2$  grados de libertad, y el  $P$ -valor del test de  $H_0$  versus

$$H_a : \mu_1 > \mu_2 \quad \text{es} \quad p^* = P(T \geq t_{obse})$$

$$H_a : \mu_1 < \mu_2 \quad \text{es} \quad p^* = P(T \leq t_{obse})$$

$$H_a : \mu_1 \neq \mu_2 \quad \text{es} \quad p^* = P(|T| \geq |t_{obse}|)$$

Y se rechaza la hipótesis nula con nivel  $\alpha$  a favor de

$$H_a : \mu_1 > \mu_2 \quad \text{si} \quad t_{obse} \geq t_{(1-\alpha)}$$

$$H_a : \mu_1 < \mu_2 \quad \text{si} \quad t_{obse} \leq t_\alpha$$

$$H_a : \mu_1 \neq \mu_2 \quad \text{es} \quad |t_{obse}| \geq t_{(1-\alpha/2)}$$

Templado	En agua salada	en aceite
	144.98	145.02
	145.04	144.94
	145.02	144.98
	145.04	144.97
	145.03	144.97
	145.03	145.03
	145.04	144.95
	144.97	144.97
	145.05	
	145.03	
	145.02	
	145	
	145.02	
Promedio	145.021	144.979
s	0.02396	0.03137
n	13	8

Cuadro 9.2: Dureza de metal de acuerdo al templado.

**Ejemplo 9.2.2** Se quieren comparar dos procedimientos diferentes de templado del metal, uno en agua salada y otro en aceite, con respecto a la dureza (o grado de temple) obtenido. Para ello se recogen los siguientes datos experimentales, que mostramos en la tabla 9.2.

Entonces  $\bar{X}_1 = 145,021$ ,  $\bar{X}_2 = 144,979$  y

$$s_p^2 = \frac{12 \times s_1^2 + 7 \times s_2^2}{19} = 0,007178$$

por lo cual  $s_p = 0,027$ , y  $s_p \sqrt{\frac{1}{13} + \frac{1}{8}} = 0,012$ . Un intervalo de confianza del 95 % para la diferencia de medias sería

$$\left[ (\bar{X}_1 - \bar{X}_2) - t_{(1-\alpha/2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\bar{X}_1 - \bar{X}_2) + t_{(1-\alpha/2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right] = [0,015, 0,065]$$

pues  $t_{0,975} = 2,093$  con 19 grados de libertad. Observemos que 0 no pertenece al intervalo de confianza calculado, por lo cual a nivel  $\alpha = 0,05$  se rechaza la hipótesis de igualdad de medias.

Con un nivel menor  $\alpha = 0,01$ , las hipótesis son

$$H_0 : \mu_1 = \mu_2 \quad H_a : \mu_1 \neq \mu_2.$$

El valor observado

$$|t_{obs}| = \left| \frac{(\bar{x}_1 - \bar{x}_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| = 3,33$$

y el percentil  $t_{0,995} = 2,861$ , con 19 grados de libertad, es más chico que el valor observado 3.33, por lo cual la hipótesis de igualdad de medias se rechaza aún con nivel 0.01.

Hay muy poca duda de que los métodos de templado producen diferente dureza en el metal.

### 9.2.3. El test t para dos muestras: $\sigma_1$ y $\sigma_2$ desconocidos

Supongamos ahora que las desviaciones estándar poblacionales  $\sigma_1$  y  $\sigma_2$  no son conocidas. Siguiendo el razonamiento del caso de una muestra, deberíamos substituir las desviaciones estándar poblacionales por



las muestrales, lo cual resultaría en el estadístico

$$T = \frac{\bar{T}_1 - \bar{T}_2 - \mu_1 - \mu_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Pero lamentablemente, aún bajo la suposición de normalidad de las poblaciones, este estadístico no tiene distribución  $t$  de Student. Pero puede verse que su distribución se puede aproximar por una  $t$  de Student con grados de libertad estimados de los datos.

**Definición 9.2.5** Si se tienen muestras independientes de poblaciones con varianza conocida, y media desconocida, para testear la hipótesis  $H_0 : \mu_1 = \mu_2$ , se computa el valor del estadístico en las muestras

$$t_{obse} = \frac{\bar{x}_1 - \bar{x}_2 - \mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

. En términos del estadístico  $T$ , con distribución  $t$  aproximada, el  $P$ -valor con respecto a

$$H_a : \mu_1 > \mu_2 \quad \text{es} \quad P(T \geq t_{obse})$$

$$H_a : \mu_1 < \mu_2 \quad \text{es} \quad P(T \leq t_{obse})$$

$$H_a : \mu_1 \neq \mu_2 \quad \text{es} \quad P(|T| \geq |t_{obse}|)$$

Los test de nivel  $\alpha$  para testear la hipótesis  $H_0 : \mu_1 = \mu_2$  rechazan en favor de la alternativa

$$H_a : \mu_1 > \mu_2 \quad \text{si} \quad t_{obse} \geq t_{(1-\alpha)}$$

$$H_a : \mu_1 < \mu_2 \quad \text{es} \quad t_{obse} \leq t_\alpha$$

$$H_a : \mu_1 \neq \mu_2 \quad \text{es} \quad |t_{obse}| \geq t_{(1-\alpha/2)}$$

Los grados de libertad de la distribución  $t$  se estiman con la siguiente fórmula

$$gl = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1} \left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1} \left(\frac{S_2^2}{n_2}\right)^2}$$

Si este número no es entero, se lo aproxima al entero más próximo.

**Ejemplo 9.2.3** El pesticida DDT causa temblores y convulsiones al ser ingerido por humanos y mamíferos. Investigadores buscan entender porque se producen las convulsiones. En un experimento comparativo aleatorio, 6 ratas blancas son envenenadas con DDT y comparadas con el grupo control de 6 ratas no envenenadas. Las mediciones eléctricas de la actividad nerviosa son las pistas principales para deducir como trabaja el envenenamiento por DDT. Cuando un nervio es estimulado su respuesta nerviosa muestra un pico agudo, seguido por otro pico mucho menor. En las ratas envenenadas, el segundo pico era mucho mayor que en las ratas normales. Esta observación es importante para entender como causa los temblores el DDT.

Los investigadores midieron la amplitud del segundo pico como un porcentaje del primer pico, cuando se estimulo un nervio de una pata de la rata. Para las ratas envenenadas, los valores fueron

12,207   16,869   25,050   22,429   8,456   20,589

En el grupo control, los datos fueron

11,074   9,686   12,064   9,351   8,182   6,642

La diferencia en las medias es bastante grande, pero en estas muestras tan pequeñas la media muestral es muy variable. Un test de significación ayudaría a confirmar que lo que se está viendo es un efecto real. Como el incremento en el segundo pico no era un efecto que se sospechaba, sino se observó en este experimento, es mejor realizar un test a dos colas.

$$H_0 : \mu_1 = \mu_2 \quad H_a : \mu_1 \neq \mu_2$$

Los datos parecen ser bastante normales (tanto como seis datos lo pueden ser). El valor del estadístico  $t_{obs}$  es de 2,99, y el  $P$ -valor, calculado con una distribución  $t$  con  $5,9 \sim 6$  grados de libertad, es de

$$P(|T| \geq 2,99) = 0,03$$

bastante pequeño. Por lo cual hay buena evidencia de que el tamaño medio del segundo pico difiere en el grupo envenenado del grupo control.

# Bibliografía

- [1] Mosterín, Jesús: Los lógicos. Espasa Calpe, Madrid, 2000.
- [2] Notas de Algebra. Enzo Gentile. EUDEBA.
- [3] Estadística Matemática con aplicaciones. W.Mendenhall, D.Wackerly, R. Scheaffer. Grupo Editorial Sudamericano.
- [4] Probabilidad y Aplicaciones Estadísticas. Paul Meyer. Adison Wesley Iberoamericana.
- [5] Introduction to Probability and Statistics. W. Mendenhall. Ed.Duxbury.
- [6] Introduction to Probability Theory. P. Hoel, S. Port, C. Stone. Houghton Mifflin Series in Statistics.
- [7] Mathematical Statistics and Data Analysis. John Rice. Duxbury Press.
- [8] Introduction to the Practice of Statistics. David Moore and George McCabe. Freeman and Co.
- [9] Do robust statistics work with real data? s.m stigler, annals of statistics 5 1977 pp 1055-1078.