Clase 3: Tests de permutaciones I

Matías Carrasco

6 de octubre de 2019

Índice

1. La Catadora de té

La Catadora de té es un experimento aleatorio que Ronald Fisher presentó en su libro *The Design of Experiments* en 1935. En una típica reunión de té inglesa, Fisher cuenta que:

> "Una señora declara que al probar una taza de té con leche se puede discriminar si la leche o la infusión de té se añadió primero en la taza. Vamos a considerar el problema de diseñar un experimento mediante el cual esta afirmación pueda ser probada. [...]
>
> Consiste en mezclar ocho tazas de té con leche, cuatro de una manera y cuatro de la otra, y su presentación a la señora en un orden aleatorio. A la señora se le ha dicho en qué consistirá la prueba, a saber, se le pedirá probar las ocho tazas, de las cuales habrá cuatro de cada tipo [...]."

¿Es capaz la señora de discriminar correctamente las maneras de servir el té?

Tenemos que hacer un modelo del experimento en el cual podamos poner números, hacer cuentas, y tomar una decisión. El experimento no tiene un único desarrollo posible, por lo que empezamos por enumerar los resultados posibles.

Supongamos que a la Sra. se le presentan las tazas alternadamente como se muestra en la Figura 1. La Sra. debe decirnos cuáles son las cuatro tazas en las cuales el té se sirvió primero, y cuáles son las tazas en las que la leche se sirvió primero.

Digamos que ella decide que las tazas $1,6,4$ y 7 son aquellas en las cuales el té se sirvió primero. Las tazas $6,4,3$ y 5 son correctas, pero las $1,7,2$ y 8 son incorrectas. En total tiene 2 aciertos para "el té se sirvió primero" y 2 aciertos para "la leche se sirvió primero".

Está claro que es suficiente decir cuáles son las tazas en las que el té se sirvió primero, pues el resto queda determinado. De este modo, lo que la Sra. debe hacer es darnos una lista

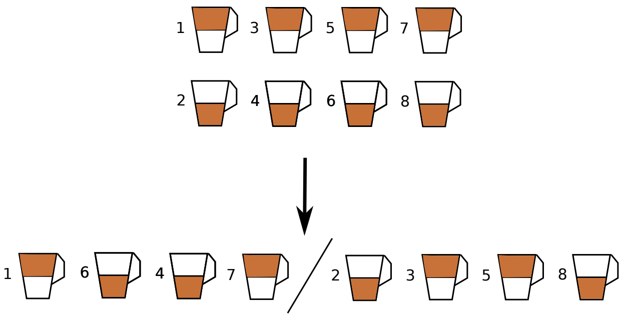


Figura 1: Un posible resultado del experimento. A la Sra. se le presentan las tazas alternadamente y ella decide que las tazas $1,6,4$ y 7 son aquellas en las cuales el té se sirvió primero.

de cuatro números que corresponden a las tazas en las cuales ella cree que el té se sirvió primero:

$$\boxed{1\ 2\ 3\ 4\ 5\ 6\ 7\ 8} \longrightarrow \{*\ *\ *\ *\}$$

Para simplificar la discusión, supongamos que a la Sra. se le presentan las tazas alternadamente como en el caso anterior. En la Tabla 1, se listan todas las posibles decisiones que ella puede tomar.

Este es el primer paso para construir nuestro modelo, acabamos de definir el espacio muestral. Los elementos de Ω son los resultados posibles del experimento, en nuestro caso, cada $\omega \in \Omega$ es una lista de 4 números, sin ordenar, que representan la elección de tazas hecha por la Sra.

Tabla 1: Todas las posibles decisiones para las tazas en las cuales el té se sirvió primero. En total hay 70 posibilidades.

1,2,3,4	1,2,3,5	1,2,3,6	1,2,3,7	1,2,3,8	1,2,4,5	1,2,4,6
1,2,4,7	1,2,4,8	1,2,5,6	1,2,5,7	1,2,5,8	1,2,6,7	1,2,6,8
1,2,7,8	1,3,4,5	1,3,4,6	1,3,4,7	1,3,4,8	1,3,5,6	1,3,5,7
1,3,5,8	1,3,6,7	1,3,6,8	1,3,7,8	1,4,5,6	1,4,5,7	1,4,5,8
1,4,6,7	1,4,6,8	1,4,7,8	1,5,6,7	1,5,6,8	1,5,7,8	1,6,7,8
2,3,4,5	2,3,4,6	2,3,4,7	2,3,4,8	2,3,5,6	2,3,5,7	2,3,5,8
2,3,6,7	2,3,6,8	2,3,7,8	2,4,5,6	2,4,5,7	2,4,5,8	2,4,6,7
2,4,6,8	2,4,7,8	2,5,6,7	2,5,6,8	2,5,7,8	2,6,7,8	3,4,5,6
3,4,5,7	3,4,5,8	3,4,6,7	3,4,6,8	3,4,7,8	3,5,6,7	3,5,6,8
3,5,7,8	3,6,7,8	4,5,6,7	4,5,6,8	4,5,7,8	4,6,7,8	5,6,7,8

Nuestro segundo paso consiste en asignar probabilidades. Vamos a actuar como abogados del diablo y suponer que la Sra. está tirando a embocar el orden en que el té fue servido en las tazas.

Si tira a embocar, es razonable suponer que cada decisión que tome tiene la misma chance de ocurrir. Esto es, cada una de las posibilidades listadas en la Tabla 1 tiene la misma probabilidad de ser elegida.

El total de listas sin orden es

$$\binom{8}{4} = \frac{8!}{4!4!} = 70.$$

Suponiendo entonces que la Sra. elige al azar las supuestas tazas que se sirvieron con té primero, definimos la probabilidad de que una elección particular de tazas ocurra como siendo igual a $1/70$.

Veamos que con esta elección de probabilidades, es como si la Sra. tirara una moneda para decidir si el té fue servido primero o no. Por ejemplo, consideremos el evento "la Sra. elige la taza número 1". Para calcular la probabilidad de este evento basta ver cuáles son las listas de la Tabla 1 que contienen al número 1.

Vemos que hay 35 listas que tienen la taza número 1, de modo que

$$\mathbf{P}(\text{“la Sra. elige la taza número 1”}) = \frac{35}{70} = \frac{1}{2}.$$

Así, la taza número 1 tiene 50% de chances de ser elegida. Lo mismo pasa con las otras tazas, lo cual confirma que nuestro modelo es razonable ya que ninguna taza tiene tendencia, o un sesgo, a ser elegida o no.

Nuestro último paso para terminar el modelo es definir lo que se conoce como un *estadístico*. El estadístico debe ser una función del resultado del experimento[1] que nos permita decidir a favor o en contra de la Sra. En este caso parece razonable considerar la variable X que cuenta cuántos aciertos hay en la lista elegida por la Sra. Claramente X puede ser cualquier valor entre 0 y 4.

Ahora sí, ya podemos hacer cuentas. ¿Con qué probabilidad toma X cada uno de los diferentes valores? Para $k \in \{0,1,2,3,4\}$, queremos calcular la probabilidad de que la variable X sea igual a k. Este evento consiste simplemente en aquellas listas para las cuales se obtendrían k aciertos. Recordar que las tazas en las que se sirvió el té primero son las tazas pares: $2,4,6$ y 8. En la Tabla 2 se muestra con diferentes colores los eventos correspondientes para cada valor de k.

Tabla 2: Los eventos $\{X = k\}$ en colores según el valor de k: $k = 0$ en rojo, $k = 1$ en azul, $k = 2$ en negro, $k = 3$ en verde, y $k = 4$ en naranja.

1,2,3,4	1,2,3,5	1,2,3,6	1,2,3,7	1,2,3,8	1,2,4,5	1,2,4,6
1,2,4,7	1,2,4,8	1,2,5,6	1,2,5,7	1,2,5,8	1,2,6,7	1,2,6,8
1,2,7,8	1,3,4,5	1,3,4,6	1,3,4,7	1,3,4,8	1,3,5,6	1,3,5,7
1,3,5,8	1,3,6,7	1,3,6,8	1,3,7,8	1,4,5,6	1,4,5,7	1,4,5,8
1,4,6,7	1,4,6,8	1,4,7,8	1,5,6,7	1,5,6,8	1,5,7,8	1,6,7,8
2,3,4,5	2,3,4,6	2,3,4,7	2,3,4,8	2,3,5,6	2,3,5,7	2,3,5,8
2,3,6,7	2,3,6,8	2,3,7,8	2,4,5,6	2,4,5,7	2,4,5,8	2,4,6,7
2,4,6,8	2,4,7,8	2,5,6,7	2,5,6,8	2,5,7,8	2,6,7,8	3,4,5,6
3,4,5,7	3,4,5,8	3,4,6,7	3,4,6,8	3,4,7,8	3,5,6,7	3,5,6,8
3,5,7,8	3,6,7,8	4,5,6,7	4,5,6,8	4,5,7,8	4,6,7,8	5,6,7,8

Contando las listas en cada evento, vemos que

$$\mathbf{P}(X=0) = \frac{1}{70},\ \mathbf{P}(X=1) = \frac{16}{70},\ \mathbf{P}(X=2) = \frac{36}{70},\ \mathbf{P}(X=3) = \frac{16}{70},\ \mathbf{P}(X=4) = \frac{1}{70}.$$

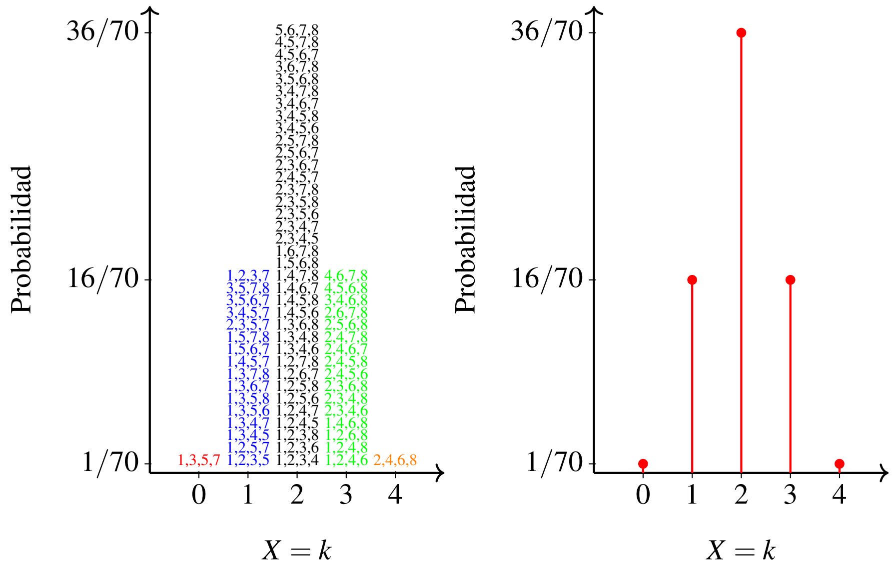
[1] Es decir, una variable aleatoria.

Si somos perezosos y no queremos contar, no es difícil probar que

$$\mathbf{P}(X=k)=\frac{\binom{4}{k}\binom{4}{4-k}}{\binom{8}{4}},$$

pues debemos elegir k de las cuatro tazas pares, $4-k$ de las tazas impares, y dividir por el total de listas.

Esta fórmula nos dice qué valores toma X y con qué probabilidad lo hace. En otras palabras, determina la f.p.p. de la variable X. Podemos graficar la función $k \mapsto \mathbf{P}(X=k)$ poniendo en el eje x los valores de k y en el eje y la probabilidad de que X sea igual a k.



A partir de esta función podemos calcular cualquier probabilidad que nos interese respecto de la variable X. En este caso particular, la distribución de la variable X se llama *hipergeo-métrica*.

Si la Sra. no acierta ninguna taza, naturalmente estamos inclinados a concluir que no es capaz de discriminar correctamente si el té o la leche se sirvieron primero. Sin embargo, según nuestros cálculos la probabilidad de no acertar ninguna taza suponiendo que la Sra. tira a embocar es $1/70$, aproximadamente $1.4\,\%$. De hecho, es igual de probable que acertar las cuatro tazas. Pero si acierta las cuatro tazas, estaríamos inclinados a decir que sí es capaz de discriminar correctamente.

La aparente confusión viene de que estamos mirando los números equivocados. Lo que realmente es relevante para nosotros es cuál es la probabilidad de que acierte k *o más* tazas.

El evento $\{X \geq k\}$ se descompone en una unión disjunta de eventos de la forma $\{X=i\}$

para $i \geq k$. De este modo, obtenemos

$$\begin{aligned}
\mathbf{P}(X \geq 0) &= \mathbf{P}(X=0)+\mathbf{P}(X=1)+\mathbf{P}(X=2)+\mathbf{P}(X=3)+\mathbf{P}(X=4) = 1\\
\mathbf{P}(X \geq 1) &= \mathbf{P}(X=1)+\mathbf{P}(X=2)+\mathbf{P}(X=3)+\mathbf{P}(X=4) = \frac{69}{70} \approx 0.99\\
\mathbf{P}(X \geq 2) &= \mathbf{P}(X=2)+\mathbf{P}(X=3)+\mathbf{P}(X=4) = \frac{53}{70} \approx 0.76\\
\mathbf{P}(X \geq 3) &= \mathbf{P}(X=3)+\mathbf{P}(X=4) = \frac{17}{70} \approx 0.24\\
\mathbf{P}(X \geq 4) &= \mathbf{P}(X=4) = \frac{1}{70} \approx 0.014
\end{aligned}$$

Estos cálculos muestran que es mucho más probable acertar al menos 0 tazas, cosa que ocurre siempre, que al menos 4, lo cual ocurre con probabilidad $1/70$.

Es importante notar que para construir nuestro modelo probabílistico tuvimos que hacer una *hipótesis* determinante. Supusimos que si la Sra. está tirando a embocar a la hora de clasificar las tazas, entonces cada una de las listas de la Tabla 1 tiene la misma probabilidad de ocurrir. Una vez supuesto esto, pudimos hacer cuentas precisas para calcular las probabilidades de los eventos que nos interesaban.

Dicho de otro modo, tradujimos

$$\boxed{\begin{array}{c}\text{(Hipótesis)}\\ \text{La Sra. tira a embocar}\end{array} \xrightarrow{\text{por}} \begin{array}{c}\text{Equiprobabilidad}\\ \text{en las listas}\end{array}}$$

Para terminar, he aquí entonces lo que nosotros le diríamos a la Sra. sobre su capacidad de discriminar correctamente las tazas:

- Si no acierta ninguna, $X = 0$, le diríamos que es una embustera!
- Si acierta sólo una, $X = 1$, le diríamos:

  > Señora, si usted estuviera tirando a embocar, el 99 % de las veces acertaría en al menos una taza. No le creo.

- Si acierta en dos tazas, $X = 2$, le diríamos:

  > Señora, si usted estuviera tirando a embocar, el 76 % de las veces acertaría en al menos dos tazas. Sigo sin creerle.

- Si acierta en tres tazas, $X = 3$, le diríamos:

  > Señora, si usted estuviera tirando a embocar, el 24 % de de las veces acertaría en al menos tres tazas. No me convence...

- Si acierta todas, $X = 4$, le diríamos: ok, usted gana.

Fisher no cuenta como terminó la historia, aunque algunos historiadores afirman que la Sra. embocó las cuatro tazas.

Razonando por improbable: la noción de p-valor

Veamos en detalle el razonamiento que hemos utilizado para tomar una decisión en el problema de las tazas. Todos deben conocer el extraño, aunque poderoso, método de demostración por absurdo o *razonamiento por absurdo*.

Para los que no se acuerdan, lo recordamos: si queremos probar que una afirmación o una cierta hipótesis es falsa, la suponemos verdadera, y a partir de esto intentamos encontrar una contradicción (cualquiera sea).

En estadística existe un razonamiento similar, el *razonamiento por improbable*. El mismo consiste en suponer que una cierta hipótesis es verdadera, y a partir de esta suposición calcular la probabilidad de observar algo tanto o más extremo que lo observado. Si esta probabilidad es muy baja, concluímos que la hipótesis es falsa.

Por ejemplo, en el caso de las tazas, supongamos que la Sra. acierta en $X_{\text{obs}} = 4$ tazas[2]. Nuestra hipótesis es que la Sra. está tirando a embocar, y bajo esta hipótesis la probabilidad de observar algo tanto o más extremo que lo observado, en nuestro caso

$$\mathbf{P}(X \geq X_{\text{obs}}) = \mathbf{P}(X \geq 4) = 1/70 \approx 0.014,$$

es muy pequeña. Concluímos entonces que la Sra. no está tirando a embocar.

Este tipo de razonamiento es muy frecuente en estadística. El primero en ponerlo en evidencia fue Fisher, quien llamo a la probabilidad $\mathbf{P}(X \geq X_{\text{obs}})$ con el nombre de p-valor.

> En palabras, el p-valor es la probabilidad, suponiendo la hipótesis verdadera, de observar algo tanto o más extremo que lo observado.

Notar que el p-valor depende del resultado del experimento, esto es, depende de los datos.

2. El gurú de los exámenes

Al aplicar el razonamiento por improbable tenemos siempre el riesgo de equivocarnos. El principio se basa en no creer demasiado en las coincidencias. Pero que algo sea muy poco probable no quiere decir que sea imposible, de hecho existe gente que ha ganado la lotería. A veces podemos formar parte de ese selecto grupo de afortunados que compraron el boleto ganador.

Imaginemos que nuestro curso consiste en una cantidad enorme de estudiantes, digamos más de mil[3]. Supongan que al final de cada semana hay un examen, que consiste en una pregunta elegida por el profesor entre dos posibles. Por supuesto, como el profesor tiene cierta piedad de sus alumnos, les avisa al principio de cada semana cuáles son las dos preguntas posibles numeradas 1 y 2.

Pero no todo son rosas, el estudiante que no sepa la respuesta a la pregunta es automáticamente sancionado. Aunque puede seguir cursando sin problemas, debe realizar tareas comunitarias en la Facultad, como limpiar la harina y los huevos que dejan los recién recibidos, arreglar los bancos rotos de los salones, y un gran etc[4].

[2] La notación X_{obs} quiere decir simplemente el valor observado de X.
[3] Basta imaginarse un curso de Cálculo 1 de la FIng.
[4] Pocos recursos, es lo que tiene la masividad.

He aquí lo que le pasó a Ana, estudiante del curso. A mitad de la primer semana, sin que ella haya hecho ningún pedido, recibe una carta en su casa indicando la pregunta que será elegida para el próximo examen. Digamos que la carta dice la número 2. Sin darle mucha importancia Ana asiste al examen y efectivamente la pregunta elegida es la número 2.

Ana no le da mucha importancia al asunto. Sin embargo, a mitad de la segunda semana recibe otra vez sin haber pedido nada, otra carta indicando la pregunta que será elegida por el profesor para el próximo examen. Digamos la número 1. Aún escéptica, Ana asiste al examen y efectivamente la pregunta elegida es la número 1. Coincidencias se dice Ana a sí misma.

Esto mismo le ocurre a Ana durante diez semanas consecutivas, recibiendo siempre una carta que no pidió, que indica de forma implacable la pregunta que será elegida por el profesor para el próximo examen.

Sin embargo, a la onceava semana Ana recibe una carta diferente. La misma dice que para conocer la predicción sobre la pregunta que será elegida por el profesor en el próximo examen se debe pagar una cierta suma, algo así como \$1000.

Ana, que algo aprendió en el curso, se decide a construir un modelo, poner números y hacer las cuentas. La hipótesis: el gurú que envía las cartas es un impostor.

Los resultados posibles son para Ana secuencias de tamaño diez de 1's y 2's. De modo que el espacio muestral es

$$\Omega = \{\omega = (a_1, a_2, \dots, a_{10}) : a_i \in \{1,2\}\}.$$

Si el gurú está tirando a embocar, cada secuencia ω debe tener la misma probabilidad de ocurrir. ¿Cuántas secuencias hay en Ω? Como hay dos posibilidades para cada a_i, tenemos que

$$|\Omega| = \underbrace{2 \times 2 \times \cdots \times 2}_{10 \text{ veces}} = 2^{10} = 1024.$$

Para chequear que su modelo es razonable, Ana calcula la probabilidad de que el i-ésimo pronóstico sea 1: el total de secuencias con $a_i = 1$ es 2^9, por lo que

$$\mathbf{P}(a_i = 1) = \frac{2^9}{2^{10}} = \frac{1}{2}.$$

Es decir, según el modelo el gurú está tirando a embocar $50/50$ el pronóstico de cada pregunta. Hasta aquí todo parece OK.

Ahora Ana considera el estadístico. Ella sabe que las preguntas elegidas por el profesor en las primeras diez semanas fueron

$$(e_1, \dots, e_{10}) = (2,1,1,1,2,1,2,2,1,2).$$

Define diez variables $X_i : \Omega \to \mathbb{R}$, para $i = 1, \dots, 10$, que para cada secuencia $\omega \in \Omega$ indican si el i-ésimo pronóstico fue correcto. En símbolos

$$X_i(\omega) = \mathbb{1}_{\{a_i = e_i\}}(\omega) = \begin{cases} 1 & \text{si } a_i = e_i, \\ 0 & \text{si } a_i \neq e_i. \end{cases}$$

A Ana le interesa calcular la probabilidad de observar algo tanto o más extremo que lo observado. La suma de las variables, $S = X_1 + \cdots + X_{10}$, cuenta el número de pronósticos correctos en una secuencia ω, de modo que Ana quiere $\mathbf{P}(S \geq S_{\text{obs}})$. El S observado por Ana es $S_{\text{obs}} = 10$. Es decir, que sólo una secuencia posible puede contribuir al evento $\{S \geq S_{\text{obs}}\}$, es la secuencia $\omega = (2,1,1,1,2,1,2,2,1,2)$. Dicho en números

$$\mathbf{P}(S \geq S_{\text{obs}}) = \mathbf{P}(\omega = (2,1,1,1,2,1,2,2,1,2)) = \frac{1}{1024} \approx 0.00098.$$

Para Ana esta es una probabilidad muy pequeña.

Cansada de estudiar tanto, decide comprar el pronóstico para el próximo examen. Abre la carta, la pregunta elegída es la 2. Va al examen, lee la letra, pregunta número 1. Se quiere matar. Frustrada, entrega en blanco y se prepara para una semana de trabajos comunitarios.

¿Qué salió mal? ¿Será el modelo, las cuentas? El modelo parece razonable, y los números no mienten... Al cabo de un tiempo se cruza con su amigo Beto y le cuenta lo que le pasó. Beto le cuenta que a él también le habían llegado cartas, pero en la tercer semana el pronóstico era incorrecto y nunca más volvió a tener noticias del gurú.

Ana enojada se pone a investigar. Resulta que otro estudiante, de nombre Carlos, brillante pero pésima persona, había ideado el plan. Todo era un fraude. Carlos se dió cuenta que muchos de sus compañeros iban a razonar por improbable y vió una hermosa oportunidad para hacer plata fácil.

Lo que Carlos hizo fue repartir 1024 cartas, 512 pronosticando la pregunta 1 y 512 pronosticando la pregunta 2. Aquellos que recibieron el pronóstico equivocado, nunca más volvieron a saber del gurú. A los 512 que recibieron el pronóstico correcto, Carlos les mandó una segunda carta, a 256 pronosticando la pregunta 1, y a 256 la pregunta 2. Al cabo de diez semanas Carlos sabía que un estudiantes habría recibido diez pronósticos correctos. Con un poco de suerte, sería un fiel seguidor del razonamiento por improbable. Mil pesos en el bolsillo[5].

Ana tenía el boleto ganador. Siempre hay una probabilidad de error cuando tomamos decisiones con estadística. Por eso es importante que otros hagan el experimento además de nosotros. Si Ana hubiera hablado con Beto antes de tomar la decisión, tal vez no se hubiera arriesgado.

La moraleja es: no creer en los pulpos Paul, ¿cuántos pulpos quedaron por el camino? No creer en las cábalas, siempre habrá alguien que cada vez que se ponga la camiseta de su cuadro preferido para el examen lo salvará. En definitiva, siempre hay un boleto ganador.

3. La distribución hipergeométrica

La distribución hipergeométrica aparecerá mucho en los ejemplos de esta y la siguiente clase. He aquí la definición general.

Consideremos una urna que contiene N bolillas, de las cuales K son rojas y $N-K$ son verdes. El experimento consiste en extraer, sin reposición, n bolillas de la urna. Entonces X es el número de bolillas rojas en la muestra extraída.

Calculemos la f.p.p. de X. Primero, notar que $n \leq N$ pues la extracción es sin reposición. Comencemos por el recorrido de X:

[5] Es importante agregar que Carlos siempre se lamentó de que no hubieran más estudiantes en el curso.

1. Para que no haya ninguna roja entre las n debe haber al menos n verdes en la urna, es decir, $n \leq N-K$. Si no es así, el número máximo de verdes entre las n extraídas es $N-K$, por lo que el mínimo de rojas es $n-(N-K)$. En ambos casos podemos decir que $X \geq \max\{0, n-(N-K)\}$.
2. Para que las n extraídas sean todas rojas debemos tener $n \leq K$. De lo contrario el número máximo de rojas entre las n extraídas es K. En ambos casos $X \leq \min\{n, K\}$.

Resumiendo, X puede tomar cualquier valor entero entre $m_n = \max\{0, n-(N-K)\}$ y $M_n = \min\{n, K\}$.

Podemos modelar el experimento con el espacio muestral

$$\Omega = \Big\{\{\omega_1, \ldots, \omega_n\} \text{ subconjuntos de tamaño } n\Big\},$$

formado por los subconjuntos de tamaño n de bolillas de la urna. Cada subconjunto es igualmente probable.

De este modo, para cada k en el recorrido de X, la probabilidad de que X sea igual a k está dada por

$$\mathbf{P}(X=k) = \overbrace{\binom{K}{k}}^{\substack{\text{elegimos}\\ \text{las } k \text{ rojas}}} \cdot \overbrace{\binom{N-K}{n-k}}^{\substack{\text{elegimos las}\\ n-k \text{ verdes}}} \Big/ \underbrace{\binom{N}{n}}_{\substack{\text{total de}\\ \text{muestras}}}$$

La distribución de X se llama *hipergeométrica* de parámetros N, K y n. La función de probabilidad puntual está dada por

$$\mathbf{P}(X=k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

para todo $m_n \leq k \leq M_n$. Escribimos $X \sim H(N, K, n)$ para indicar que X tiene distribución hipergeométrica.

En la Figura 2 se muestran cuatro gráficos de la función de probabilidad puntual de la distribución hipergeométrica. En todos los casos $N = 100$, es decir hay 100 bolas en la urna, y $n = 10$, es decir extraemos 10. La cantidad de bolas rojas varía de $K = 20$ a $K = 80$.

Notar que en cada uno de los casos, la forma de la distribución es acampanada. Sin embargo no es simétrica. En el primer caso en que sólo hay $K = 20$ bolas rojas en la urna, la variable X se concentra en valores chicos de k, teniendo un máximo para $k = 2$.

Cuando $K = 40$ la distribución es bastante más simétrica, aunque no del todo. En este caso el máximo se da en $k = 4$. A medida que K aumenta, la distribución se va corriendo hacia la derecha, en donde para $K = 60$ el máximo se da en $k = 6$, y para $K = 80$ el máximo se da en $k = 8$.

En la Figura 3 se muestran tres gráficos más. En este caso $N = 1000$, es decir hay 1000 bolas en la urna y se extrae una muestra de tamaño $n = 100$. En el primer caso la cantidad de bolas rojas es $K = 250$, por lo que la distribución está concentrada en valores pequeños de k, con un máximo en $k = 25$. Lo opuesto ocurre en el tercer caso, en donde $K = 750$.

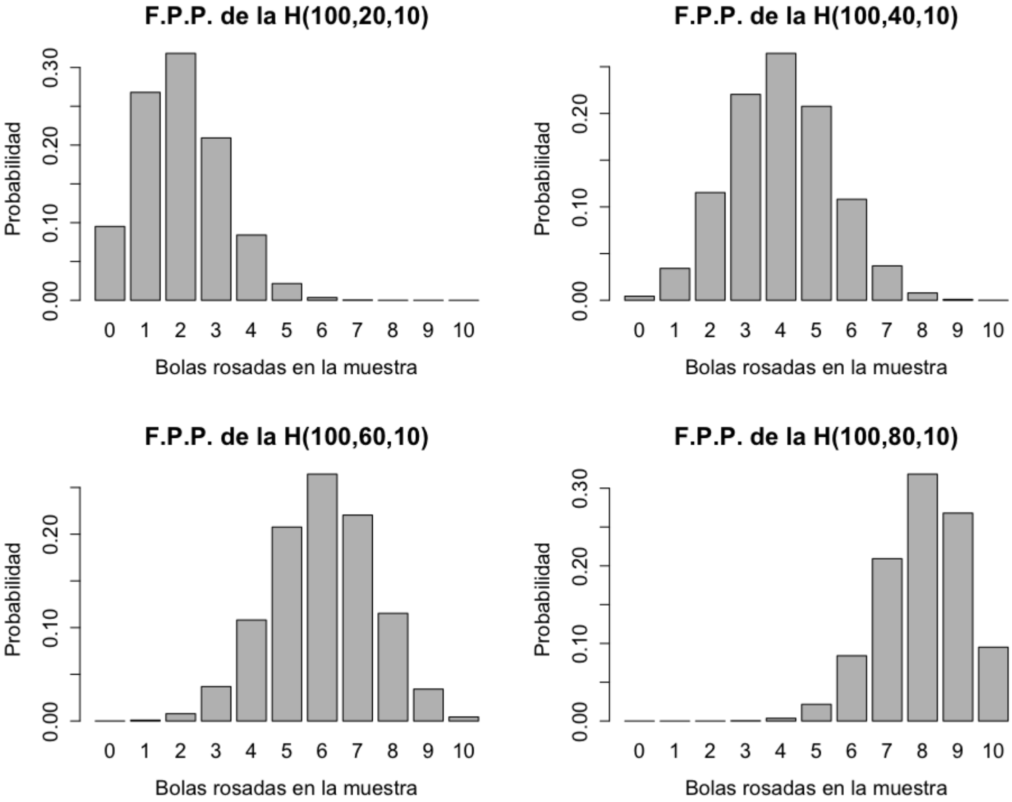


Figura 2: Distribución hipergeométrica.

Notar que el segundo gráfico es perfectamente simétrico respecto de $k = 50$. Esto refleja el hecho de que hay la misma cantidad de bolas rojas $K = 500$ que verdes $N - K = 500$ en la urna. La probabilidad de que X tome los valores $k = 50 \pm x$ son iguales, para todo x entre -50 y 50.

Calculemos ahora la esperanza y la varianza de X. Para esto, ordenamos las bolas extraídas de forma arbitraria, y notar que podemos escribir X como una suma de variables Bernoulli

$$X = \sum_{i=1}^{n} X_i$$

en donde

$$X_i = \begin{cases} 1 & \text{si } i \text{ es roja} \\ 0 & \text{si no.} \end{cases}$$

El parámetro p de X_i es igual a K/N. Entonces

$$\mathbf{E}(X_i) = \frac{K}{N} \text{ y } \mathbf{var}(X_i) = \frac{K(N-K)}{N^2}.$$

De aquí se sigue inmediatamente que $\mathbf{E}(X) = np = nK/N$. Notar que es la misma esperanza que para la binomial de parámetros n y p.

Para calcular la varianza precisamos calcular las covarianzas. El producto $X_i X_j$ para $i < j$ es una variable Bernoulli de parámetro igual a

$$\mathbf{P}(i \text{ y } j \text{ son rosadas}) = \frac{K(K-1)}{N(N-1)}.$$

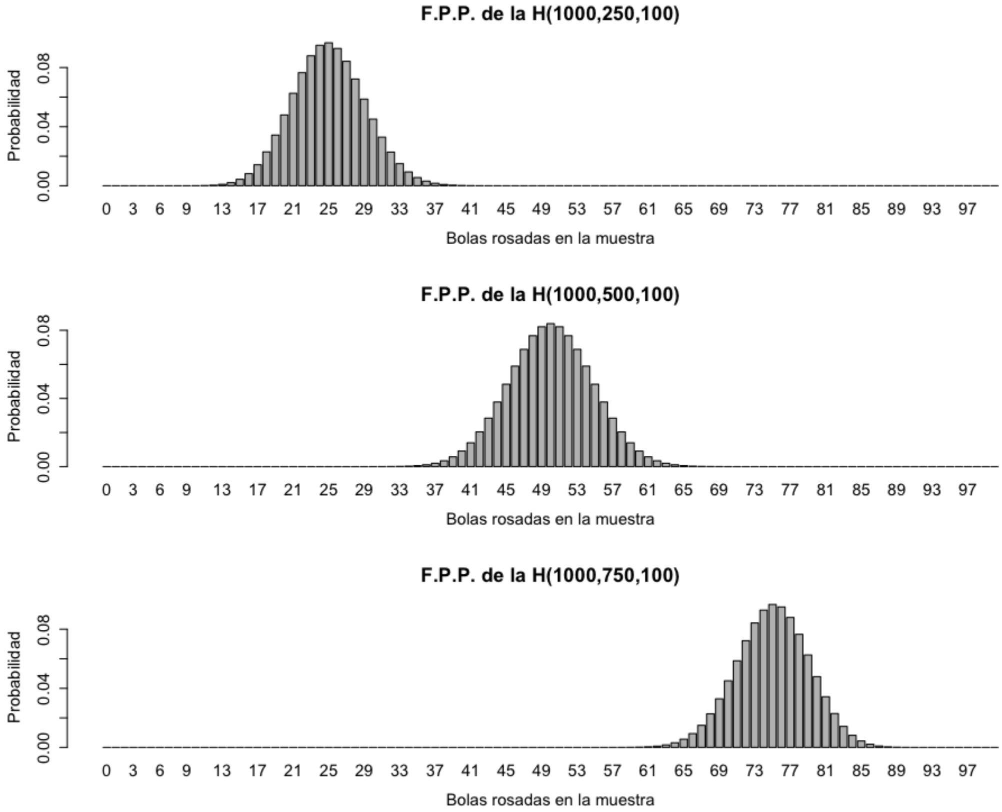


Figura 3: Distribución hipergeométrica.

Entonces

$$\operatorname{cov}(X_i, X_j) = \frac{K(K-1)}{N(N-1)} - \frac{K^2}{N^2} = \frac{K}{N}\frac{K-N}{N(N-1)} = -\frac{K(N-K)}{N^2(N-1)}.$$

Luego

$$\mathbf{var}(X) = \frac{nK(N-K)}{N^2} - n(n-1)\frac{K(N-K)}{N^2(N-1)} = \frac{nK(N-K)}{N^2}\left(1 - \frac{n-1}{N-1}\right)$$

El cálculo directo es mucho más engorroso, por no decir casi imposible. Notar que cuando N es grande, la varianza de X se aproxima a la varianza de una binomial de parámetros n y $p = K/N$.

4. Dedos sensibles

A veces se afirma que ciertas personas, muy sensibles, pueden distinguir los colores con la punta de los dedos. La afirmación generalmente dice que la persona sensible, incluso con los ojos vendados, puede identificar el color de un objeto tocándolo.

La mayoría de los científicos no le dan importancia a esta idea. Critican la noción de órganos sensibles al color en los dedos, y discuten otras explicaciones de experimentos de sensibilidad aparentemente exitosos.

Un grupo decidió investigar la sensibilidad del color en los dedos por medio del siguiente experimento. Reunieron 20 naipes, 16 azules y 4 rojos. Usaron naipes de mazos con partes de atrás idénticas a excepción del color.

Los naipes se barajaron completamente, y la persona bajo prueba sintió las superficies de las tarjetas con los ojos cuidadosamente vendados. Luego, la persona seleccionó las 4 cartas que debían ser rojas. El resultado fue que la persona seleccionó correctamente 3 naipes rojos. ¿Es casualidad?

Claramente, el resultado que más confirmaría fuertemente la conjetura de la visión a través de los dedos sería la correcta selección de los 4 naipes rojos. Este resultado se puede describir en una tabla:

		Color elegido		
		Rojo	Azul	Total
Color real	Rojo	4	0	4
	Azul	0	16	16
	Total	4	16	20

El 4 de la parte superior izquierda de la tabla significa que 4 naipes realmente rojos fueron elegidos como rojos; los ceros significan que en realidad no se eligieron naipes rojos como azul, y viceversa. Observe que los totales marginales, en este caso 4 y 16 tanto para filas y columnas, son fijos. La persona a prueba color-sensible sabe que 4 de los naipes son rojos y se le indica que elija 4, ni más ni menos. El total es 20 naipes.

El siguiente resultado más confirmatorio sería la correcta selección de tres de los 4 naipes rojos; la tabla para ese resultado es:

		Color elegido		
		Rojo	Azul	Total
Color real	Rojo	3	1	4
	Azul	1	15	16
	Total	4	16	20

Por lo tanto, de los 4 elegidos como rojos, solo 3 son en realidad; de los 4 realmente rojos, solo 3 son elegidos. Quince naipes son realmente azules y elegidos como azules.

Hay otros tres resultados posibles para el experimento, que corresponden a 2, 1 y 0 opciones correctas de naipes rojos. La tabla en este último caso, en la que no se elige correctamente ninguno de los naipes rojos, parece que muestra una especie de destreza negativa o hacia atrás.

	R	A	T
R	2	2	4
A	2	14	16
T	4	16	20

	R	A	T
R	1	3	4
A	3	13	16
T	4	16	20

	R	A	T
R	0	4	4
A	4	12	16
T	4	16	20

Restringimos la atención a dos preguntas importantes: ¿Cuáles son las probabilidades de los cinco resultados anteriores si, de hecho, la persona no tiene la sensibilidad del color en la punta de los dedos, por lo que las opciones son completamente al azar? ¿Cómo puede el conocimiento de tales probabilidades ayudar a evaluar la hipótesis de la selección de azar?

Notar que todas las tablas examinadas anteriormente tienen la siguiente forma:

		Color elegido		
		Rojo	Azul	Total
Color real	Rojo	r	$4-r$	4
	Azul	$4-r$	$12+r$	16
	Total	4	16	20

Si elegimos como estadístico la variable

$$X = \text{n}^{\text{o}} \text{ de naipes rojos correctos},$$

entonces su distribución es hipergeométrica de parámetros $N = 20$, $K = 4$, y $n = 4$. De este modo, la probabilidad de que la persona haga r aciertos es

$$\mathbf{P}(X = r) = \frac{\binom{4}{r}\binom{16}{4-r}}{\binom{20}{4}}, \ r = 0, 1, 2, 3, 4.$$

Los argumentos son exactamente los mismos que en el ejemplo de la Catadora de té, pero hemos usado las tablas para mostrar otra forma de resumir las posibilidades del experimento.

En este caso, el valor del estadístico es $X_{\text{obs}} = 3$, por lo que el p-valor es por definición igual a

$$\mathbf{P}(X \geq X_{\text{obs}}) = \mathbf{P}(X \geq 3) = 0.0132 + 0.0002 = 0.0134.$$

Es una probabilidad razonablemente pequeña.

Si la persona sensible elige los 4 naipes rojos, entonces algo bastante notable ha sucedido desde el punto de vista de alguien que no cree en la visión de la punta de los dedos. Porque bajo la hipótesis de aleatoriedad, la probabilidad de elegir los 4 naipes rojos es solo de 2 en 10000.

Si se eligen 3 naipes rojos, eso todavía es notable, ya que la probabilidad de elegir 3 *o más* naipes rojos por casualidad es solo de 0.0134, o un poco menos de 1 en 70.

Por otro lado, si se eligen 2 naipes rojos, uno podría notar que *dos o más* naipes rojos aparecerían por casualidad con una probabilidad de 0.162, o aproximadamente una vez de cada seis, y concluiría que nada notable ha sucedido.

Entonces, podría ser razonable aceptar que si se eligen 2 o menos naipes rojos, decidiremos que no hay evidencia sustancial contra la hipótesis de sentido común de la aleatoriedad.

Por otro lado, si se eligen 3 o 4 naipes rojos, podríamos decidir que algún fenómeno real está ocurriendo, posiblemente la visión de la punta de los dedos, pero más probablemente diferentes texturas de pintura roja y azul, sugerencias de observadores, u otros eventos más comprensibles. Pero es poco creíble que sea por pura casualidad.