

### Binomial vs Hipergeométrica

La idea es tratar de diferenciar cuando elegir la binomial o la hipergeométrica para resolver un ejercicio.

Supongamos que tenemos una bolsa que tiene 6 bolas blancas y 4 rojas. En este caso, la probabilidad de sacar una bola blanca es 0.6, y de sacar una bola roja es 0.4. Si el ejercicio nos habla de hacer varias extracciones con reposición, es claro que, en cada extracción habrá la misma cantidad de bolas rojas, y la misma cantidad de bolas blancas, por lo que las probabilidades siempre serán las mismas, y estaremos en presencia de eventos independientes. En este caso, si queremos saber la probabilidad de sacar 2 bolas blancas en 3 extracciones independientes, debemos calcular  $P[X=2]$ , donde  $X$  distribuye binomial con parámetros 3 y 0.6.

Si las extracciones se hacen sin reposición, es claro que al querer efectuar la segunda extracción, la proporción de bolas blancas sobre las rojas habrá cambiado, y por ende la probabilidad de sacar una bola blanca, será distinta. En este caso no podemos modelar el ejercicio con una distribución binomial, ya que la probabilidad de éxito paso a paso irá cambiando. Para esta situación, se trabaja con la distribución hipergeométrica con parámetros 10, 6 y 3. Para entender más esta situación ver las notas de Probabilidad de Yohai.

Algo parecido sucede cuando agregamos bolas de otros colores al experimento, en tal caso debemos elegir entre la distribución Multinomial y la Hipergeométrica Multivariada. Esto también lo pueden ampliar con las Notas de Probabilidad de Victor Yohai.

Ahora bien, luego de esta introducción al problema, hilemos mas fino en el problema. Si ustedes miran el ejercicio 4 de la práctica 3, uno a primera vista ve que se trata de un ejercicio típico para modelar con la distribución binomial, sin embargo, uno puede correctamente decir que estamos en presencia de extracciones sin reposición, por lo que deberíamos modelar el problema con la distribución Hipergeométrica. En tal caso, nos faltarían parámetros, debido a que no conocemos  $N$ , ni tampoco  $D$ . En caso de tener estos datos, uno podría tranquilamente modelar la situación con tal distribución, pero en este caso no contamos con dichos datos. En realidad lo que sucede es que estamos ante una situación conocida en estadística como población infinita (muy grande a comparación de la muestra). Para entender mejor, supongamos que tenemos 1000 bolas blancas y 300 rojas, y deseamos extraer 3 de ellas sin reposición. La probabilidad de extraer (R,B,R) es  $(300/1300) \cdot (1000/1299) \cdot (299/1298) = 0.0409\dots$ . Si ahora realizamos el mismo cálculo pero para extracciones con reposición, tenemos que la probabilidad de extraer (R,B,R) es  $(300/1300) \cdot (1000/1300) \cdot (300/1300) = 0.0409\dots$ . Observamos que la diferencia es insignificante, y el análisis para problemas mas complejos se simplifica considerablemente si trabajamos con extracciones con reposición.

Resumen: Usar Binomial si no se conoce el tamaño de la población y se puede suponer que el mismo es significativamente grande con respecto a la muestra.

En el ejercicio 8 de la práctica 3 tenemos una situación parecida. En este caso la muestra es de 20 peces, y seguramente la población es significativamente mas grande que 20.