Leonardo Daniel Rivera Valdez

Probabilidad discreta: un repaso

1) Ensayo

Un **ensayo**, en el contexto de probabilidad, es un intento de hacer. Por ejemplo, un ensayo al estar nosotros interesados en tirar un dado es justamente un intento de tirar los dados, si tiramos una moneda pues será tirar la moneda en una ocasión, etc.

2) Definición frecuentista de la probabilidad

Para el **frecuentista la probabilidad** de un resultado es la frecuencia con la que ocurre un evento si repetimos hubiésemos repetido dicho evento una infinidad de veces (o una cantidad muy alta de veces tal que la probabilidad de ocurrencia del evento se aproxima a la proporción de ocurrencias de este).

Por otro lado, para un **bayesiano la probabilidad** de un resultado es el grado de creencia que uno tiene sobre dicho resultado.

3) Espacio muestral

El **espacio muestral** es el conjunto de los resultados posibles de un ensayo, por ejemplo si uno va a tirar una moneda dos veces, entonces el espacio muestral es:

$$\{AA,\ AS,\ SA,\ SS\}$$

Donde A= águila y S=sol

Si tiramos un dado de seis caras entonces el espacio muestral es:

$$\{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$$

si por otro lado tiramos dos dados de 6 caras, entonces el espacio muestral es:

$$\{1\ y\ 1,\ 1\ y\ 2,\ 1\ y\ 3,\ 1\ y\ 4, 1\ y\ 5,\ 1\ y\ 6, ..., 6\ y\ 6\}$$

habría 36 combinaciones posibles (6x6=36).

4) Evento

Un **evento**, en probabilidad, es algún subconjunto del espacio muestral. Por lo general se refiere a la ocurrencia de un caso del espacio muestral, pero puede ser una combinación de N elementos del espacio muestral, incluyendo el espacio muestral mismo (aunque no es lo común).

Por lo general los eventos se denotan como A, B, C, etc. Por ejemplo, uno puede decir:

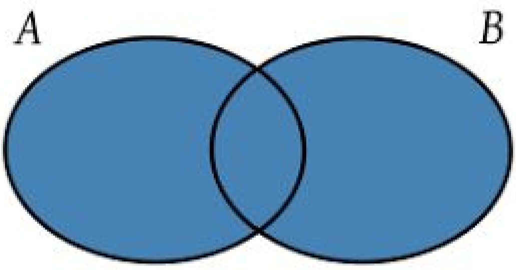
$$\textit{Deja que } A \textit{ sea el evento "comer un taquito de pastor"}$$

$$\textit{Deja que } B \textit{ sea el evento "ponerle salsita al taco de pastor"}$$

5) Unión e intersección

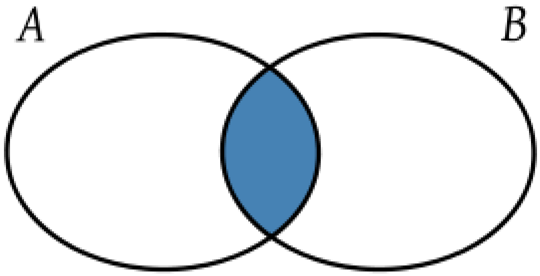
La **unión** de dos eventos A y B contiene los resultados que están en A o B. Supón que hablamos del tiro de una moneda 2 veces, entonces usando el símbolo ∪ diríamos que:

$$A \cup B = \{AA,\ AS,\ SA,\ SS\}$$



En cambio si hablamos de una **intersección** de A y B, entonces contiene los resultados que están en ambos A y B. Usando el símbolo ∩
tendríamos que:

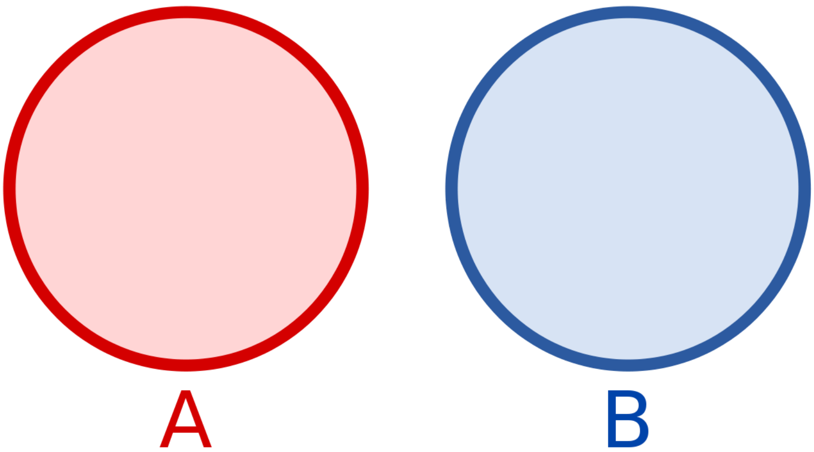
$$A \cap B = \{AS,\ SA\}$$



6) Eventos disjunto

Dos eventos A y B son **disjuntos** si no tienen ningún resultado en común tal que:

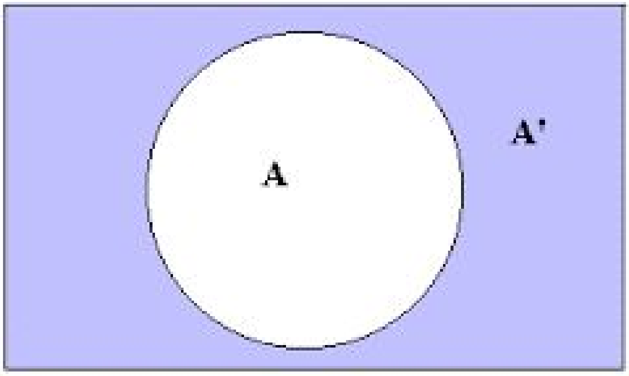
$$A \cap B = \emptyset$$



7) Eventos complementarios

Los **eventos complementarios** de A son todos los resultados que no se encuentran en A. Estos eventos son disjuntos y la unión de estos es el espacio muestral completo:

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



8) Notación de probabilidad

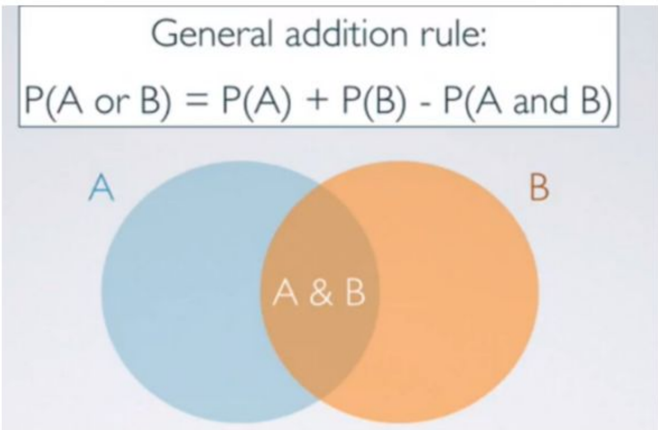
Probabilidad de un evento a se escribe de la siguiente forma:

$$P(A)$$

Donde P() es la probabilidad de un evento cualquiera, y A es el evento al cual nos referimos.

9) Regla de la adición

Esta se expresa en un diagrama de Venn como:



Por esto vemos que la **regla general de la adición** es:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

La fórmula nos dice que $P(A \; or \; B)$ es igual a la suma de las probabilidades de A y B menos la probabilidad de $P(A \; and \; B)$; esto último es para quitar los repetidos que ya están en A y B.

Cuando nuestros eventos A y B son **disjuntos** las fórmula se simplifica a:

$$P(A \; o \; B) \; = \; P(A) \; + \; P(B)$$

Esto debido a que no hay ninguna intersección de A y B. Así mismo si uno quiere la **complementaria** de un evento entonces uno hace lo siguiente:

$$P(A \; o \; A^c) = P(A) + P(A^c) - P(A \cap B)$$

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

10) Distribución binomial

Primero notamos que un **ensayo de Bernoulli** es un ensayo el cual toma solamente dos posibles resultados. Estos se dividen en **éxito** con un 1, y **fallo** con un 0. De tal manera que tenemos que el éxito tiene una probabilidad p de ocurrencia mientras que el fallo tiene una probabilidad de ocurrencia $1-p$.

Con esto en mente, una **distribución binomial** da la probabilidad de tener x éxitos en n ensayos independientes de Bernoulli con una probabilidad de ocurrencia p.

La fórmula de la distribución binomial es:

$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

Donde $\binom{n}{x}$ significa $\frac{n!}{(n-x)!\,x!}$, o sea, el factorial de n entre el factorial de $n-x$ por el factorial de x. El símbolo $!$ denota el factorial de un número, es decir la multiplicación de todos los números de uno en uno de el número en cuestión desde el uno. Así mismo, p denota la probabilidad del éxito y q la probabilidad del fallo

$1-p$. Finalmente, $P(x)$ se lee como la probabilidad de que ocurran q éxitos en n ensayos independientes de Bernoulli con una probabilidad p.

Un ejemplo de esto es si tenemos una bolsa de M&M´s con cinco lunetas de colores (rojo, verde, amarillo, azul, y naranja), si la probabilidad de obetener un M&M rojo es de $1/5$, entonces cuál es la probabilidad de obtener 1 o más M&M´s amarillos en 5 bolsas.

Para poder hacer esto en un planteamiento binomial debemos hacer uso de la regla general de adición cuando los eventos son disjuntos, tal que:

$$P(1 \text{ o más } M\&M\text{'s amarillos}) =$$
$$= P(1\ M\&M \text{ amarillo}) + P(2\ M\&M\text{'s amarillos}) + P(3\ M\&M\text{'s amarillos}) + P(4\ M\&M\text{'samarillos}) + P(5\ M\&M\text{'s amarillos})$$

Para cada una de estas habría que calcular una distribución binomial! Lo cual es mucho trabajo, así que podemos hacer uso de la complementaria la cual en nuestro ejemplo es obtener cero M&M 's. Por lo que su cálculo sería:

$$P(0 \text{ yellow M\&M}) = \binom{5}{0}(0.1)^0(1-0.1)^{5-0} = (0.9)^5$$

Luego haríamos

$$P(1 \text{ or more yellow M\&M's}) = 1 - P(0 \text{ yellow M\&M}) = 1 - (0.9)^5 \approx 0.41.$$

Lo cual es igual a sacar la distribución binomial de todas las posibilidades (cuando M&M’s son 1, 2, 3, 4, y 5)!

Esto anterior en R se logra mediante la función dbinom(), usando el siguiente código para la manera que sacamos cada una de las distribuciones binomiales posibles:

```
prob = dbinom(1, 5, 0.1) + dbinom(2, 5, 0.1) + dbinom(3, 5, 0.1) + dbinom(4, 5, 0.1)
+ dbinom(5, 5, 0.1)
 prob
```

Siendo la salida:
[1] 0.40951

Nótese que el primer parámetro de la función es el número de éxitos, luego nos pide el número de ensayos, y la probabilidad asociada al ensayo de Bernoulli.

La ruta por la complementaria es un código más simple:

```
prob = 1 - dbinom(0, 5, 0.1)
prob
## [1] 0.40951
```

11) Variables aleatorias

La siguiente fórmula nos permite determinar el número de éxitos por evento ya que $X = k$, al número de éxitos.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Esa notación, aunque abstracta, permite modelar cualquier evento de Bernoulli con distribución binomial donde se indica $k \,\&\, n$. Así mismo nos permite modelar para N cantidad de eventos disjuntos, ya que $X = k$ indica que nunca tendrán resultados compartidos. FInalmente si $X = k$ cubre todo el espacio muestral, entonces tenemos que la regla de adición nos dice:

$$\sum_{todos\ los\ k} P(X = k) = 1$$

Es decir que dan una probabilidad perfecta o de uno debido a que conforman la sumatoria de las probabilidades de todos los resultados del espacio muestral.

12) Más notaciones de probabilidad y parámetros

En los problemas de distribución binomial a veces se puede denotar esta mediante uso de la siguiente notación debido a que la fórmula de esta depende de los parámetros n, k, y p:

$$B(n,\ p)$$

O en su defecto puede hacerse el uso de:

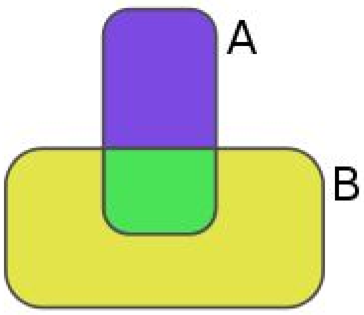
$$p(k)$$

13) Probabilidad condicional

La ecuación de Bayes trata de responder a la pregunta **¿dados los datos observados cuál es la probabilidad de ocurrencia de un evento de interés?** Esto en probabilidad se llama probabilidad condicional y se escribe con la siguiente notación que significa la probabilidad de A dado que B:

$$P(A \mid B)$$

Visualmente esto se ve así:



Como se ve esta nos muestra la probabilidad de un evento A dado que un evento B ocurre, lo verde. Así mismo no es igual a P(A). La fórmula general de la fórmula es:

$$P(A \mid B) \ = \frac{P(A\ and\ B)}{P(B)}$$

Esto es evidente si se observa la región de color verde que es el punto de intersección entre los eventos, o sea la ocurrencia de A dado B. Es por ello que tenemos la probabilidad de ocurrencia de ambos eventos entre la probabilidad de B.

14)Regla general del producto

La anterior fórmula, de la probabilidad condicional, es derivable de la **regla general del producto**, siendo esta:

$$P(A\ y\ B)\ =\ P(A\mid B)\times P(B)$$

Esta nos dice que la probabilidad de ocurrencia de A y B es la multiplicación de la probabilidad de ocurrencia de B por la ocurrencia de A dado que B. Visualmente sería la superficie completa de A y B, los rectángulos amarillo y morado de la imagen anterior.

15)Ley de la probabilidad total

Esta está hecha para describir el posible caso donde tenemos N eventos disjuntos B, si ese es el caso entonces podemos la probabilidad de A está dada como:

$$P(A) = \sum_{todos\ los\ B_i} P(A\mid Bi)P(Bi)$$

O sea:

$$P(A)\ =\ P(A\mid B_1)P(B_1)\ +\ P(A\mid B_2)P(B_2)\ +\ \cdots P(A\mid B_n)P(B_n)$$

Otra forma de representar esta fórmula es por medio de la complementaria de B, esto queda así:

$$P(A) = P(A\mid B)\,P(B) +\ P(A\mid B^{\,c})\,P(B^{\,c})$$

Un ejemplo de esto es cuando queremos la probabilidad de un evento A y lo partimos entre P(A | B) y P(A | noB). Esto se ve así:

16) Independencia

La **independencia** de un evento es cuando un determinado evento A no da ninguna información relevante sobre otro evento B. Esto se escribe:

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

Lo cual es diferente de los eventos **disjuntos**, en donde solamente no ocurren al mismo tiempo pero sí dicen información relevante (el cima de ayer y el de hoy). Por otro lado, los eventos independientes sí pueden ocurrir al mismo tiempo (que se caiga una pluma en tu casa y que alguien coma un bolillo). Así mismo esto modifica la regla general del producto (la independencia) que es:

$$P(A\ y\ B) = P(A \mid B) \times P(B) = P(A) \times P(B).$$

Es decir que no hay probabilidad de A dado que B, o que B no afecta a A. Esta independencia se mantiene a ambas direcciones:

$$P(A \mid B) = P(A) \Longrightarrow P(B \mid A) = P(B)$$

O sea que si A es independiente de B por necesidad B lo es de A.

17) Independencia en la probabilidad condicional

Supón que un evento A y B son independientes de sí pero que están condicionados por un evento C. Si ese es el caso entonces:

$$P(A\ y\ B \mid C) = P(A \mid C) \times P(B \mid C)$$

Lo que pasa es que no hay P(A) ni P(B) alaber independencia, sin embargo dado a que están mediados por C la independencia no está garantizada!

Los básicos de la estadística bayesiana

La estadística bayesiana se basa principalmente en el empleo de la probabilidad condicional, es decir, la probabilidad de ocurrencia de un evento A dado un evento B. A parte de esto esta rama suele establecer una probabilidad previa, luego una posterior a partir de la previa y el “likelihood” (no encontré una buena traducción).

1) Probabilidad condicional y la regla de Bayes

Para poder ejemplificar esto el texto nos brinda los siguientes datos:

Table 1.1: Results from a 2015 Gallup poll on the use of online dating sites by age group

	18-29	30-49	50-64	65+	Total
Used online dating site	60	86	58	21	225
Did not use online dating site	255	426	450	382	1513
Total	315	512	508	403	1738

Por ejemplo, calculemos primero la P(de que un adulto use el sitio de citas):

$$= \frac{Número\ total\ de\ adultos\ que\ usan\ sitios\ de\ citas}{El\ total\ de\ las\ personas\ de\ la\ encuesta} = \frac{225}{1738} \approx 0.13$$

Si por otro lado uno quisiera sacar la P(probabilidad de usar sitios de citas | que uno tiene de 30-49 años):

$$= \frac{Número\ de\ personas\ que\ usan\ sitios\ entre\ 30{-}49}{Total\ de\ personas\ entre\ 30{-}49} = \frac{86}{512} \approx 0.17$$

Si dividimos el numerador y el denominador entre el número total de personas obtenemos esto pero en sentido de probabilidad regular y no condicional:

$$= \frac{\frac{Número\ de\ personas\ que\ usan\ sitios\ entre\ 30{-}49}{Número\ total\ de\ personas}}{\frac{Número\ de\ personas\ entre\ 30{-}49}{Número\ total\ de\ personas}}$$

$$= \frac{P(Usar\ un\ sitio\ de\ sitas\ \&\ tener\ 30{-}49\ años)}{P(caer\ en\ el\ grupo\ de\ 30{-}49)}$$

Ésta en esencia es la **regla de Bayes**! Veámoslo de manera más abstracta:

$$(\ |\ ) \ = \ \frac{(\ \&\ )}{()}$$

Para notar la igualdad nótese que en el numerador tenemos la **ley general del producto**:

$$P(A\ y\ B)\ =\ P(A\ |\ B) \times P(B)$$

además de que en el numerador tenemos una **probabilidad total**:

$$P(A) = \sum_{todos\ los\ B_i} P(A\ |\ Bi)P(Bi)$$

2) Regla de Bayes y test diagnóstico

Con el fin de que sea más claro, se nos propone un caso de diagnóstico de VIH. En este caso se nos dice que hay 2 pruebas que se solían aplicar, una llamada ELISA (enzyme-linked immunosorbent assay) con un índice de sensibilidad (verdaderos positivos) de 0.93. Esto se expresa como:

$$P(ELISA\ \ positivo\ |\ que\ se\ tiene\ VIH) = 0.93$$

Por otro lado, su verdadero negativo o especificidad es de 0.99 y se expresa:

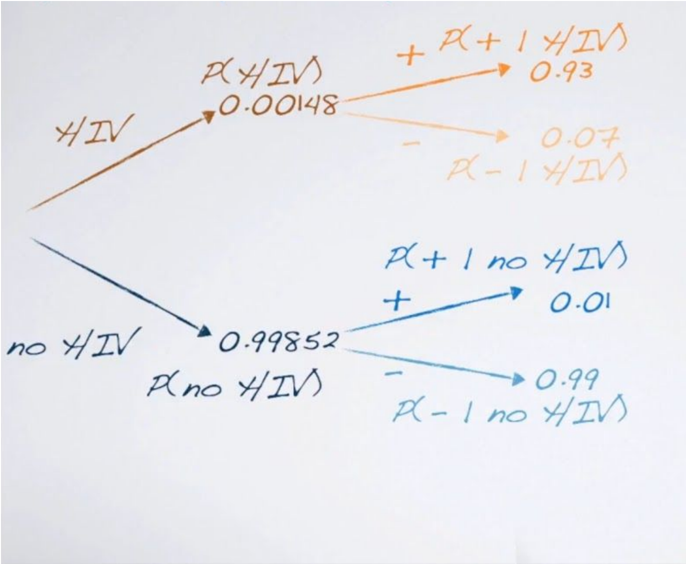
$$P(ELISA\ negativo\ |\ que\ no\ se\ tiene\ VIH) = 0.99$$

Así mismo nos interesa saber la prevalencia de VIH en la población, es decir:

$$P(ser\ positivo\ para\ VIH) = \frac{1.48}{1000} = 0.00148$$

Lo que nosotros queremos conocer es P(VIH | ELISA positivo), esto se desglosa como:

$$P(VIH\ |\ ELISA\ positivo) = \frac{P(personas\ que\ tienen\ VIH\ \&\ su\ ELISA\ es\ positivo)}{P(ELISA\ positivo)}$$



Como se ve en nuestro diagrama de árbol tenemos todo lo que necesitamos para obtener las probabilidades necesarias y sus complementarias. La probabilidad de tener VIH es:

$$P(ser\ positivo\ para\ VIH) = \frac{1.48}{1000} = 0.00148$$

Si este es el caso, entonces la probabilidad de no tener VIH es la complementaria de la anterior (1-P):

$$P(ser\ negativo\ para\ VIH) = 0.99852$$

Además sabemos que:

$$P(ELISA\ \ positivo \mid que\ se\ tiene\ VIH) = 0.93$$

por lo que podemos sacar su complementaria, es decir, la P(ELISA negativo | se tiene VIH):

$$P(ELISA\ negativo \mid se\ tiene\ VIH) = 0.07$$

Lo mismo podemos hacer con:

$$P(ELISA\ negativo \mid que\ no\ se\ tiene\ VIH) = 0.99$$

siendo su complementaria la P(ELISA positivo | no se tiene VIH):

$$P(ELISA\ positivo \mid no\ se\ tiene\ VIH) = 0.01$$

Ya que tenemos lo anterior queremos conocer $P(personas\ que\ tienen\ VIH\ \&\ su\ ELISA\ es\ positivo)$, y para poder conocer esto necesitamos aplicar la regla general del producto:

$$P(A\ \&\ B) = P(B) \times P(A|B)$$

por lo tanto multiplicamos nuestros resultados y tenemos que:

$$P(tener\ VIH\ \&\ ELISA\ positivo) = P(ser\ positivo\ para\ VIH) \times P(ELISA\ positivo\ |\ que\ se\ tiene\ VIH) =$$

$$= 0.00148 \times 0.93 = 0.0013764$$

Después necesitamos sacar la P(ELISA positivo), esto es mediante:

$$P(ELISA\ positivo) = 0.0013764 + P(no\ tener\ VIH\ \&\ ELISA\ positivo)$$

Luego

$$P(no\ tener\ VIH\ \&\ ELISA\ positivo) = 0.99852 \times 0.01 = 0.0099852$$

$$P(ELISA\ positivo) = 0.0013764 + 0.0099852 = 0.01136$$

Ahora sí podemos aplicar la regla de Bayes para conocer la probabilidad posterior:

$$P(VIH\ |\ ELISA\ positivo) = \frac{P(personas\ que\ tienen\ VIH\ \&\ su\ ELISA\ es\ positivo)}{P(ELISA\ positivo)}$$

$$= \frac{0.0013764}{0.01136} = 0.12$$

Es decir que la probabilidad de obtener VIH dado que nuestro ELISA es positivo es del 12% aproximadamente. Es decir que es una probabilidad baja debido a que la prevalencia es baja en general, por ello es muy probable que si uno se toma una prueba ELISA uno tenga un falso positivo (88% de probabilidad).

Otra pregunta que podemos hacer es la

$$P(VIH\ |\ ELISA\ negativo) = \frac{P(personas\ que\ tienen\ VIH\ \&\ su\ ELISA\ es\ negativo)}{P(ELISA\ negativo)}$$

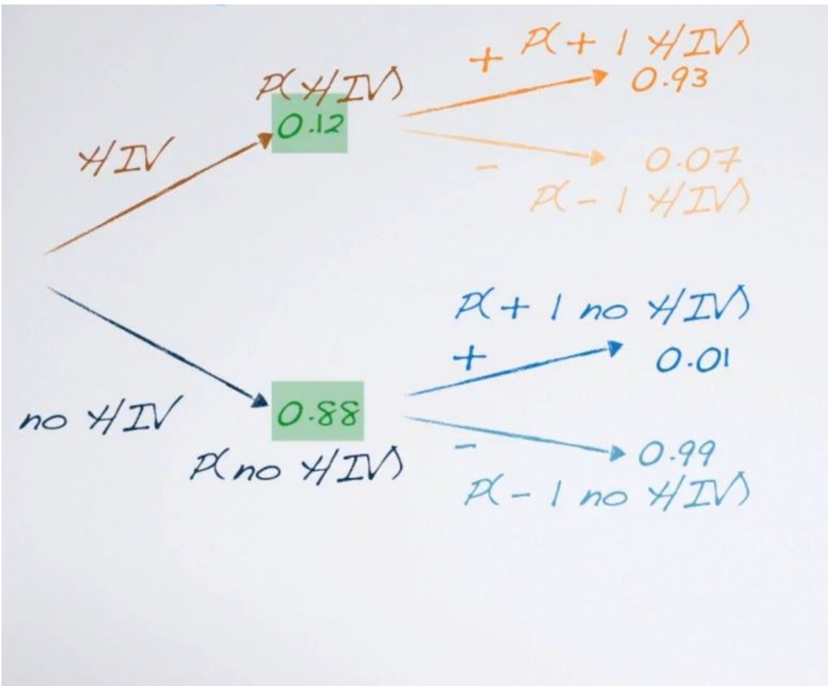
Esto es justamente la complementaria de:

$$P(VIH\ |\ ELISA\ positivo) = \frac{P(personas\ que\ tienen\ VIH\ \&\ su\ ELISA\ es\ positivo)}{P(ELISA\ positivo)}$$

Es decir, que la respuesta es 88%, como ya habíamos mencionado arriba.

3) Actualización de Bayes

Imagina que hacemos un segundo ELISA, en ese caso ya no empezamos con una $P(VIH) = 0.00148$ sino con $P(VIH) = 0.12$! Esto es debido a que ya hubo un primer ensayo. Si ese es el caso entonces nuestro árbol de probabilidad se ve:



Como se ve las otras probabilidades se mantienen iguales y solamente cambia P(VIH) y P(no VIH).

Entonces tenemos que hacer lo mismo que antes pero ahora considerando el cambio de las probabilidades. Por lo que nuestro objetivo sigue siendo obtener:

$$P(VIH \mid ELISA\ positivo) = \frac{P(personas\ que\ tienen\ VIH\ \&\ su\ ELISA\ es\ positivo)}{P(ELISA\ positivo)}$$

Por lo que:

$$P(personas\ que\ tienen\ VIH\ \&\ su\ ELISA\ es\ positivo) = P(ELISA\ positivo \mid VIH) \times P(ELISA\ positivo)$$

$$P(personas\ que\ tienen\ VIH\ \&\ su\ ELISA\ es\ positivo) = 0.12 \times 0.93 = 0.1116$$

Posteriormente tenemos que obtener:

$$P(ELISA\ positivo) = P(personas\ que\ tienen\ VIH\ \&\ su\ ELISA\ es\ positivo) + P(personas\ que\ no\ tienen\ VIH\ \&\ ELISA\ positivo)$$

$$P(ELISA\ positivo) = 0.1116 + P(no\ VIH) \times P(ELISA\ positivo \mid no\ VIH)$$

$$P(ELISA\ positivo) = 0.1116 + 0.88 \times 0.01 = 0.1204$$

Finalmente hacemos uso de la ecuación de Bayes:

$$P(VIH \mid ELISA\, positivo) = \frac{0.1116}{0.1204} = .9269$$

Es decir que si dos ELISA salen positivos tenemos una probabilidad del 93% de tener VIH. O sea, mejora la certeza al aumentar el número de pruebas tomadas.

4) Definición de probabilidad bayesiana v.s frecuentista

Empecemos por los **frecuentistas** que definen la probabilidad como:

$$() = \lim_{n \to \infty} \frac{n_E}{n}$$

donde la probabilidad de E está dada por la proporción de éxitos cuando el número de ensayos se aproxima al infinito. Es decir que la probabilidad de un evento está definida como la proporción de ocurrencia suponiendo que pudiésemos repetir dicho evento una infinidad de veces.

Por otro lado los **bayesianos** definen la probabilidad de un evento como:

$$()$$

donde se define como cualquier probabilidad con cualquier distribución, solamente que debe ser congruente con nuestras creencias (i.e. no puedes tener probabilidades que su suma den más de uno o inconsistencias por el estilo).

Estas diferencias generan diferencias en la inferencia. Como se recordará un **intervalo de confianza** del 95% nos dice que si tomáramos muestras aleatorias, supon 100 y de X tamaño, de la población entonces el 95% (si de 99% entonces el 99% de estas) de esas muestras contendrán el parámetro real de la población. Por lo que la muestra contiene o no el parámetro de la población pero no nos dice nada de estimaciones sobre el parámetro, ni la probabilidad de que el intervalo contenga el parámetro.

En vez de un intervalo de confianza el bayesiano usa un **intervalo de credibilidad**. **Esto nos dice realmente la probabilidad (supón 95%) de que el intervalo tenga**

o no el parámetro de la población. Por ejemplo, si dice que hay un 95% de probabilidad de que el parámetro esté entre X & Y, realmente está diciendo eso y no que el 95% de las muestras de X tamaño lo tendrán! Así que es más simple de interpretación y es directo de intervalo y no de las muestras.

5) Inferencia de una proporción

Se nos propone un problema donde hay 2 grupos de mujeres que reciben diferentes tratamientos contraceptivos de emergencia. Un grupo recibe el tradicional y se embarazan 12 de 20, y otro grupo recibe uno nuevo y se embaraza 4 de 20. La pregunta es qué tan fuertemente indican los datos que el tratamiento es mejor que el tradicional. Para poder hacer esto un problema de una proporción se nos invita a pensarlo de la siguiente manera (debido a que la n es igual):

$$H_0 : \ p = 0.5$$

$$H_{a:} : p < 0.5$$

Es decir que si la proporción es significativamente menor a 0.5 entonces entonces los embarazos son menores con el tratamiento que sin el tratamiento (el tradicional). Por ello la división en la mitad.

Para obtener nuestra inferencia podemos, desde el **método frecuentista**, obtener las probabilidades de obtener k=1:4 en n=20. Es decir que hablamos de un evento de Bernoulli y por ende podemos hacer uso de la distribución binomial. Donde:

$$P(k \leq 4) == \ P(k = 0) \ + \ (\ = \ 1) \ + \ (\ = \ 2) \ + \ (\ = \ 3) \ + \ (\ = \ 4)$$

Para resolver podemos hacer uso de R y la función dbinom():

```
sum(dbinom(0:4, size=20, p=0.5))
## [1] 0.005908966
```

Dado a que p= 0.005908966, entonces rechazamos H0 y concluimos que los datos sí nos muestran que el tratamiento es superior al tradicional a la hora de evitar embarazos.

Ahora si queremos hacer uso del método bayesiano, entonces debemos proceder de manera distinta. Aquí queremos resolver qué tan probable es tener 4 embarazos en el grupo del tratamiento.

Así mismo tenemos p= 0.5 en caso de que ambos tratamientos sean igual de efectivos. Por otro lado, debemos hacer otras presuposiciones desde el método bayesiano. Primero, p es una probabilidad por lo que sus valores van de 0:1. Por esto podemos hacer nueve modelos distintos de p, donde 10%, 20%, 30%,...90% son las probabilidades. Esto quiere decir que si 20%, entonces hay 2 de 10 embarazos en el grupo del tratamiento, si 30% entonces 3 de 10, etc. Por ende, en el paradigma bayesiano vamos a hacer uso de más modelos a diferencia del modelo frecuentista que solamente toma uno.

Esto que consideramos está en la siguiente tabla:

Table 1.2: Prior, likelihood, and posterior probabilities for each of the 9 models

Model (p)	0.1000	0.2000	0.3000	0.4000	0.5000	6e-01	0.70	0.80	0.90
Prior P(model)	0.0600	0.0600	0.0600	0.0600	0.5200	6e-02	0.06	0.06	0.06
Likelihood P(data|model)	0.0898	0.2182	0.1304	0.0350	0.0046	3e-04	0.00	0.00	0.00
P(data|model) x P(model)	0.0054	0.0131	0.0078	0.0021	0.0024	0e+00	0.00	0.00	0.00
Posterior P(model|data)	0.1748	0.4248	0.2539	0.0681	0.0780	5e-04	0.00	0.00	0.00

Así mismo se puede observar que hay una probabilidad previa que captura 2 creencias, que los beneficios son simétricos y que es igualmente probable que uno mejore con el tratamiento nuevo que con el estándar. Dicen que luego se discutirá cómo hacer esto en el libro; cómo determinar la previa.

Además de esto tenemos el **likelihood** que es la probabilidad de los datos observados dado el modelo considerado. Matemáticamente:

$$(data|model) = (= x| = y,)$$

en nuestro caso

$$(data|model) = (= 4| = 20,)$$

Así que igualmente vamos a hacer uso de la distribución binomial con k=4, n=20, y p dado cada modelo diferente. Así mismo haríamos uso de la ecuación de Bayes para obtener la posterior. Todo esto viene en la tabla anterior.

El modelo que uno debe tomar es el que al final tiene una posterior mayor, en este caso 20%. Lo cual es sumamente diferente de p del frecuentista que es la probabilidad de los datos por lo menos tan extremos como los observados dado que la hipótesis nula.

Esto visualmente se ve así

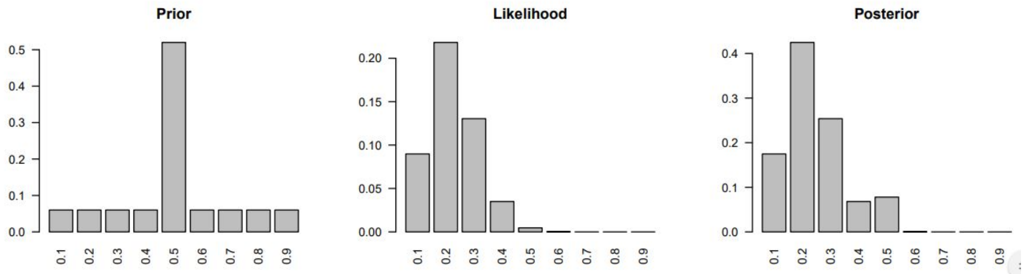


Figure 1.1: Original: sample size $n = 20$ and number of successes $k = 4$

la previa era pico en 0.5 pero tras la posterior es en 0.2, es decir que es el modelo más probable dados nuestros datos. Así mismo podemos hacer un cálculo de qué tanto es mejor el tratamiento que el estándar al sumar las posteriores inferiores a 0.5. Nos da un **92.16% de probabilidad de que el tratamiento sea mejor que el estándar (es decir que es 92.16% menos probable embarazarse con el tratamiento que con el tradicional)**; estas probabilidades del modelo no son posibles en los métodos frecuentistas que solamente son sí o no (1 o 0).

6) Efecto del tamaño de la muestra en la posterior

Lo siguiente pasa cuando n= 40

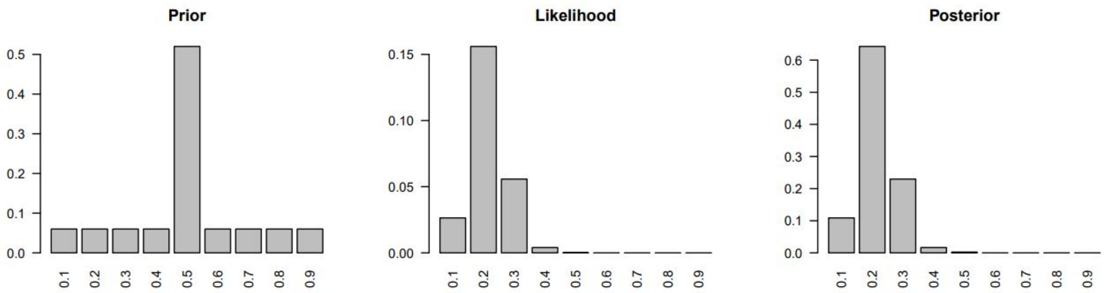


Figure 1.2: More data: sample size $n = 40$ and number of successes $k = 8$

el resultado es igual pero hay más masa en 0.2 que en las otras. Si aumentamos todavía más, n=200 pero manteniendo la proporción:

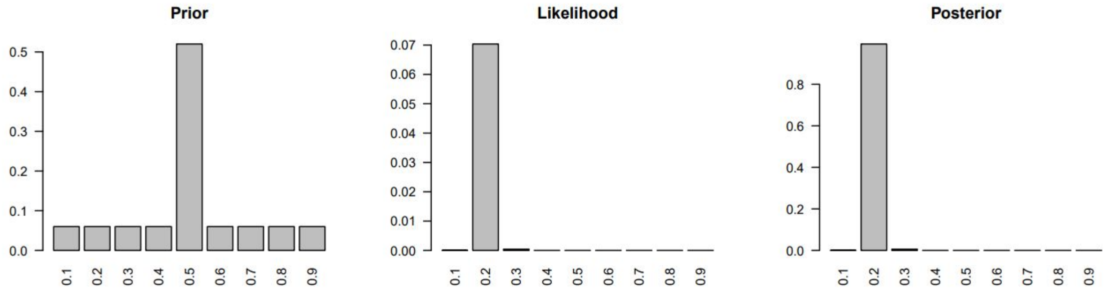


Figure 1.3: More data: sample size $n = 200$ and number of successes $k = 40$

lo mismo pasa pero cada vez es más evidente que se concentra en 0.2. Esto muestra que **conforme aumenta la muestra la likelihood (razón de verosimilitud) domina la previa**. También se ve que una previa baja se compensa con una muestra grande.

7) Inferencia frecuentista v.s. bayesiana

Finalmente consideremos un caso donde podemos comparar los dos métodos de inferencia. Considera el caso donde tenemos que conocer si la proporción de M&M´s amarillos es de 10% o 20% (hay de otrs colores: verde, rojo, etc)

Para el lado frecuentista tenemos que:

- Hypothesis: H_0 is 10% yellow M&Ms, and H_A is >10% yellow M&Ms.
- Significance level: $\alpha = 0.05$.
- Sample: red, green, **yellow**, blue, orange
- Observed data: $k = 1, n = 5$
- P-value: $P(k \geq 1|n = 5, p = 0.10) = 1 - P(k = 0|n = 5, p = 0.10) = 1 - 0.90^5 \approx 0.41$

Si este es el caso, entonces no es significativo y fallamos en rechazar la H0 concluyendo que no hay suficiente evidencia como para concluir que la proporción sea del 20% y asumimos que es del 10%.

En el caso bayesiano, por otro lado tenemos que:

- Hypotheses: H_1 is 10% yellow M&Ms, and H_2 is 20% yellow M&Ms.
- Prior: $P(H_1) = P(H_2) = 0.5$

- Sample: red, green, **yellow**, blue, orange
- Observed data: $k = 1, n = 5$
- Likelihood:

$$P(k=1|H_1) = \binom{5}{1} \times 0.10 \times 0.90^4 \approx 0.33$$
$$P(k=1|H_2) = \binom{5}{1} \times 0.20 \times 0.80^4 \approx 0.41$$

- Posterior

$$P(H_1|k=1) = \frac{P(H_1)P(k=1|H_1)}{P(k=1)} = \frac{0.5 \times 0.33}{0.5 \times 0.33 + 0.5 \times 0.41} \approx 0.45$$
$$P(H_2|k=1) = 1 - 0.45 = 0.55$$

tenemos dos modelos, uno con una proporción del 10% y otro con una del 20%, y vemos que no podemos decidirnos debido a lo muestra pequeña, sin embargo cuando n incrementa, entonces vemos que 20% es el modelo adecuado. Véase la siguiente tabla:

Table 1.3: Frequentist and Bayesian probabilities for larger sample sizes

	Frequentist	Bayesian H_1	Bayesian H_2
Observed Data	P(k or more | 10% yellow)	P(10% yellow | n, k)	P(20% yellow | n, k)
n = 5, k = 1	0.41	0.45	0.55
n = 10, k = 2	0.26	0.39	0.61
n = 15, k = 3	0.18	0.34	0.66
n = 20, k = 4	0.13	0.29	0.71

Aquí se ve el efecto del tamaño de la muestra que hace que el likelihood domine. Por otro lado vemos que los resultados son opuestos a los del método frecuentista y esto es debido a que este **se ve influenciado por la hipótesis nula** a diferencia del método bayesiano.