

Statistical Rethinking

Capítulo 2

Richard McElreath

Canicas en una bolsa

- Posibilidades

(1) [○○○○○], (2) [●○○○○], (3) [●●○○○], (4) [●●●○○], (5) [●●●●●]

- Observación ●○○●

- ¿Conclusiones sobre el sistema real?

El Jardín de Senderos que se Bifurcan

- Conjetura

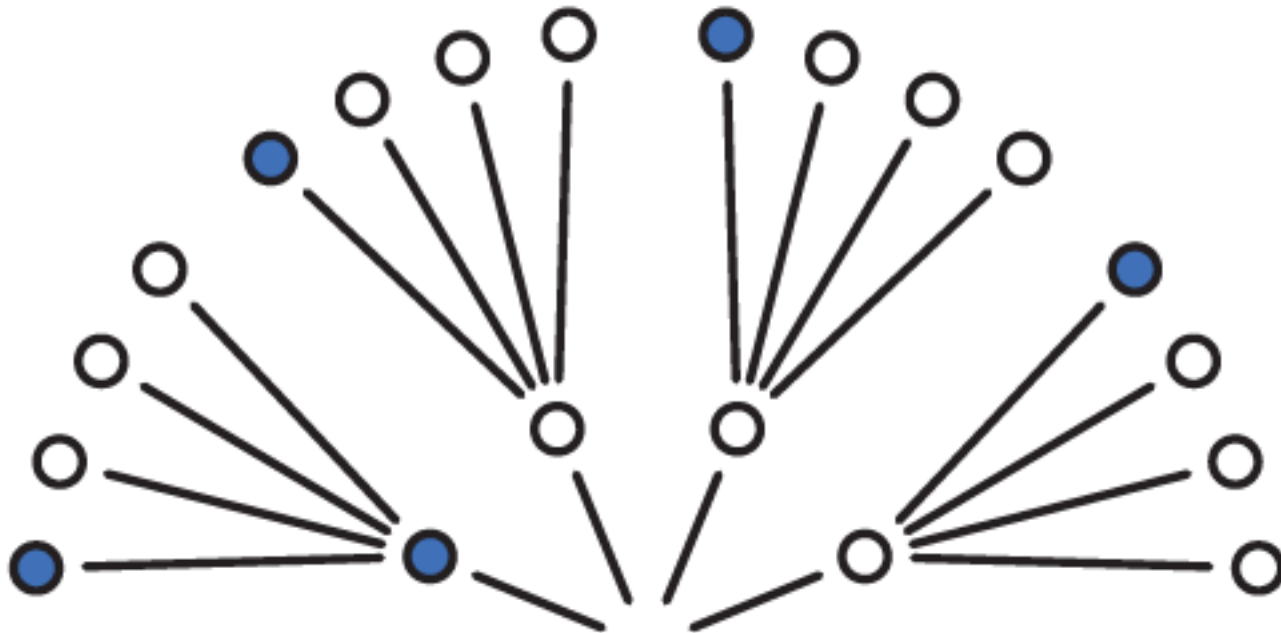
[●○○○]

- Combinaciones posibles (1era toma)



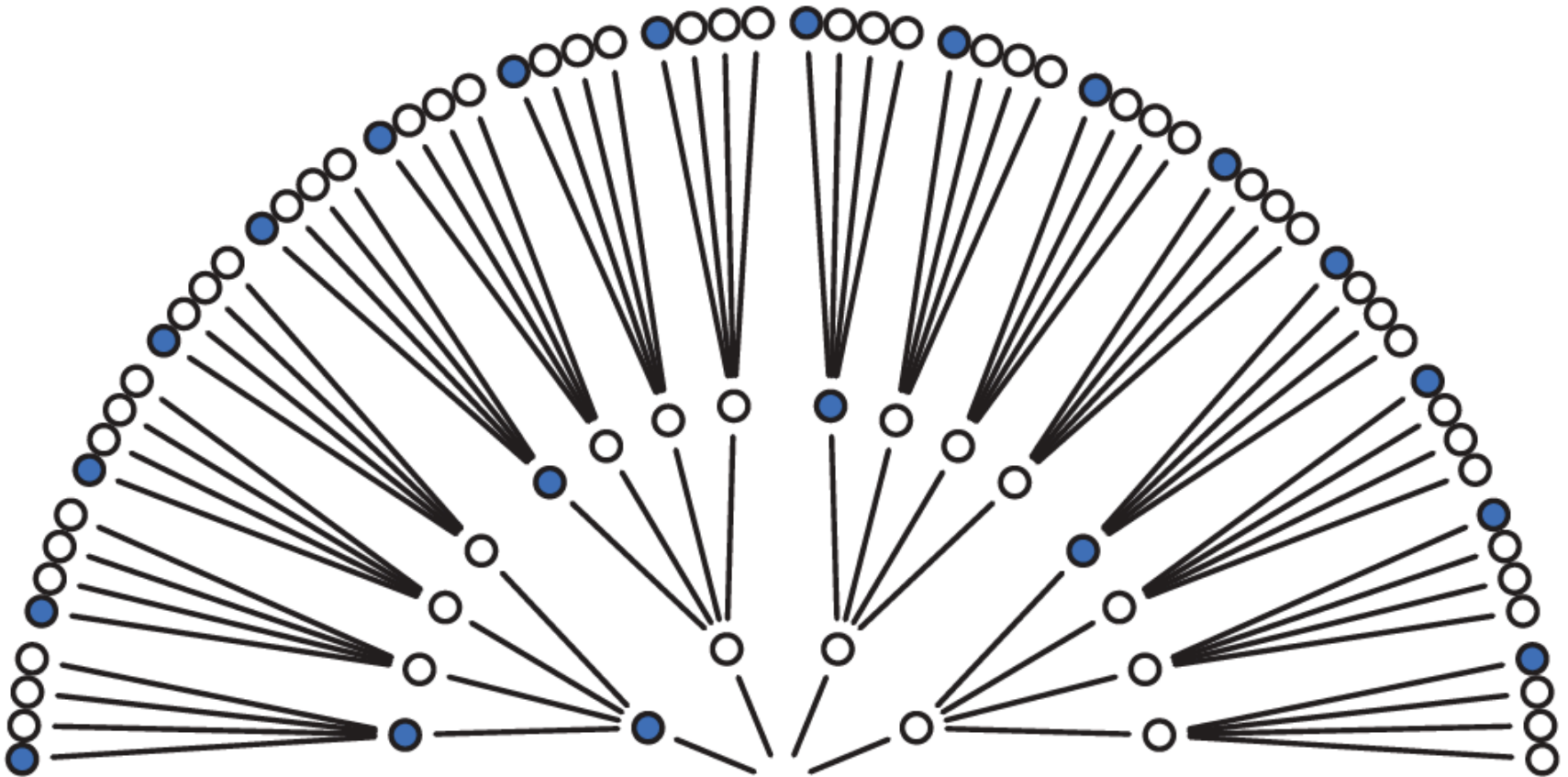
El Jardín de Senderos que se Bifurcan

- Combinaciones posibles (2da toma)



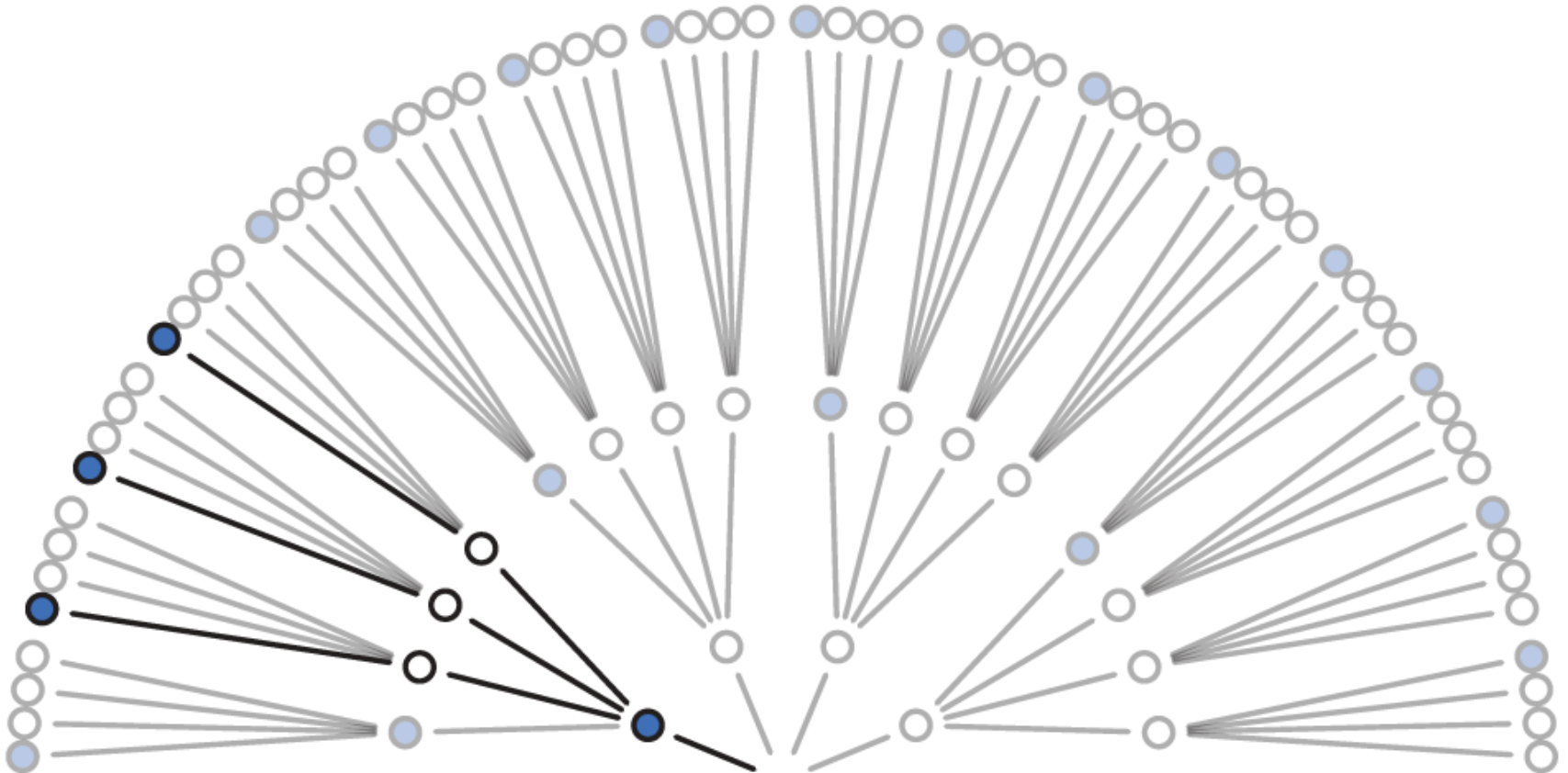
El Jardín de Senderos que se Bifurcan

- Combinaciones posibles (3ra toma)



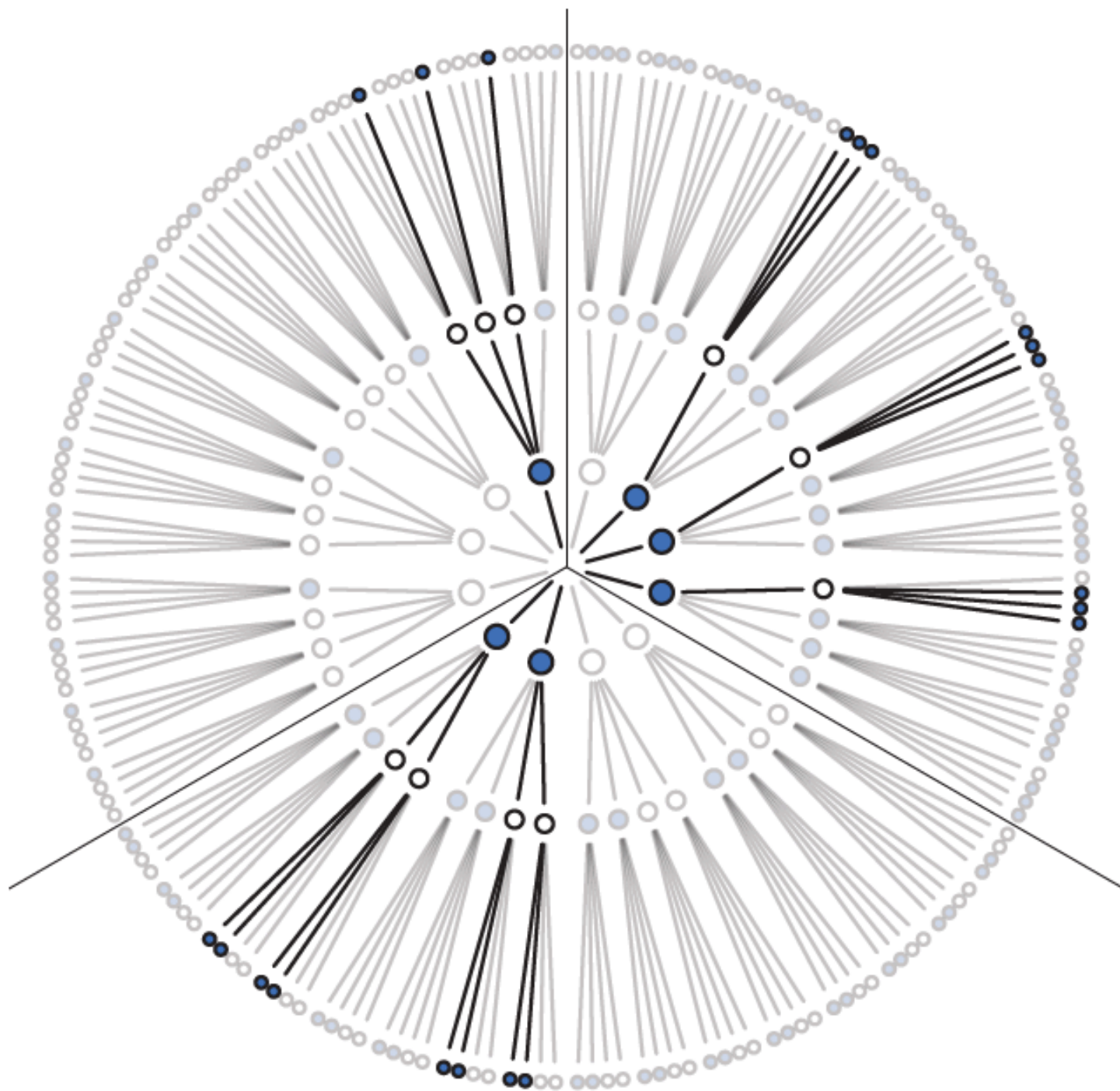
El Jardín de Senderos que se Bifurcan

- Combinaciones posibles **consistentes con la observación** ● ○ ●



Conteos

Conjetura	Maneras de producir ●○○●
[○○○○○]	$0 \times 4 \times 0 = 0$
[●○○○○]	$1 \times 3 \times 1 = 3$
[●●○○○]	$2 \times 2 \times 2 = 8$
[●●●○○]	$3 \times 1 \times 3 = 9$
[●●●●○]	$4 \times 0 \times 4 = 0$



Nueva observación: ●

- Datos actualizados ●○●●
- Opciones:
 - Contar de nuevo
 - Actualizar conteos

Conteo actualizado

Conjetura	Maneras de producir ●	Conteo anterior	Conteo nuevo
[○○○○○]	0	0	$0 \times 0 = 0$
[●○○○○]	1	3	$3 \times 1 = 3$
[●●○○○]	2	8	$8 \times 2 = 16$
[●●●○○]	3	9	$9 \times 3 = 27$
[●●●●○]	4	0	$0 \times 4 = 0$

Actualizando conteos

- Si para cada conjetura:
 - (1) Hay M_{prior} maneras de producir observación D_{prior}
 - (2) Una observación nueva D_{nuevo} tiene M_{nuevo} maneras de ser producida
- Entonces:
 - (3) Las maneras de producir D_{nuevo} habiendo observado D_{prior} anteriormente es $M_{\text{prior}} \times M_{\text{nuevo}}$

Información (más) prior

- Puede venir antes de las observaciones

Conjetura	Conteo de fábrica
[○○○○○]	0
[●○○○○]	3
[●●○○○]	2
[●●●○○]	1
[●●●●○]	0

Actualizando conteos (de nuevo)

Conjetura	Conteo prior	Conteo de	
		fábrica	Conteo nuevo
[○○○○○]	0	0	$0 \times 0 = 0$
[●○○○○]	3	3	$3 \times 3 = 9$
[●●○○○]	16	2	$16 \times 2 = 32$
[●●●○○]	27	1	$27 \times 1 = 27$
[●●●●○]	0	0	$0 \times 0 = 0$

De posibilidades a probabilidades

- Los valores de los conteos son importantes sólo cuando se comparan entre sí

$$\begin{array}{c} \text{plausibilidad de } [\bullet \circ \circ \circ] \text{ después de observar } \bullet \circ \bullet \\ \propto \\ \text{maneras de que } [\bullet \circ \circ \circ] \text{ produzca } \bullet \circ \bullet \\ \times \\ \text{plausibilidad prior de } [\bullet \circ \circ \circ] \end{array}$$

De posibilidades a probabilidades

- Proporción de canicas azules: p

$$\begin{array}{c} \text{plausibilidad de } p \text{ después de observar } D_{\text{nuevo}} \\ \propto \\ \text{maneras de que } p \text{ produzca } D_{\text{nuevo}} \\ \times \\ \text{plausibilidad prior de } p \end{array}$$

De posibilidades a probabilidades

- Normalizando, obtenemos probabilidades

plausibilidad de p después de observar $D_{\text{nuevo}} =$

$$\frac{\text{maneras de que } p \text{ produzca } D_{\text{nuevo}} \times \text{plausibilidad prior de } p}{\text{suma de productos}}$$

De posibilidades a probabilidades

- Normalizando, obtenemos probabilidades

Composición posible	p	Maneras de producir los datos	Plausibilidad
[○○○○○]	0	0	0
[●○○○○]	0.25	3	0.15
[●●○○○]	0.5	8	0.40
[●●●○○]	0.75	9	0.45
[●●●●○]	1	0	0

Formalizando

- Parámetro: p (posible explicación de los datos)
- Verosimilitud: Número relativo de maneras en las que algún valor de p puede explicar los datos
- Plausibilidad prior: probabilidad prior
- Plausibilidad actualizada: probabilidad posterior

Construyendo un Modelo Bayesiano

- (1) Historia para los datos: Motivar el modelo narrando cómo se generan los datos
- (2) Actualizar: Educar el modelo alimentándolo con los datos
- (3) Evaluar: Todos los modelos estadísticos requieren supervisión, lo que puede llevar a una revisión del modelo

Agua/Tierra



Agua/Tierra

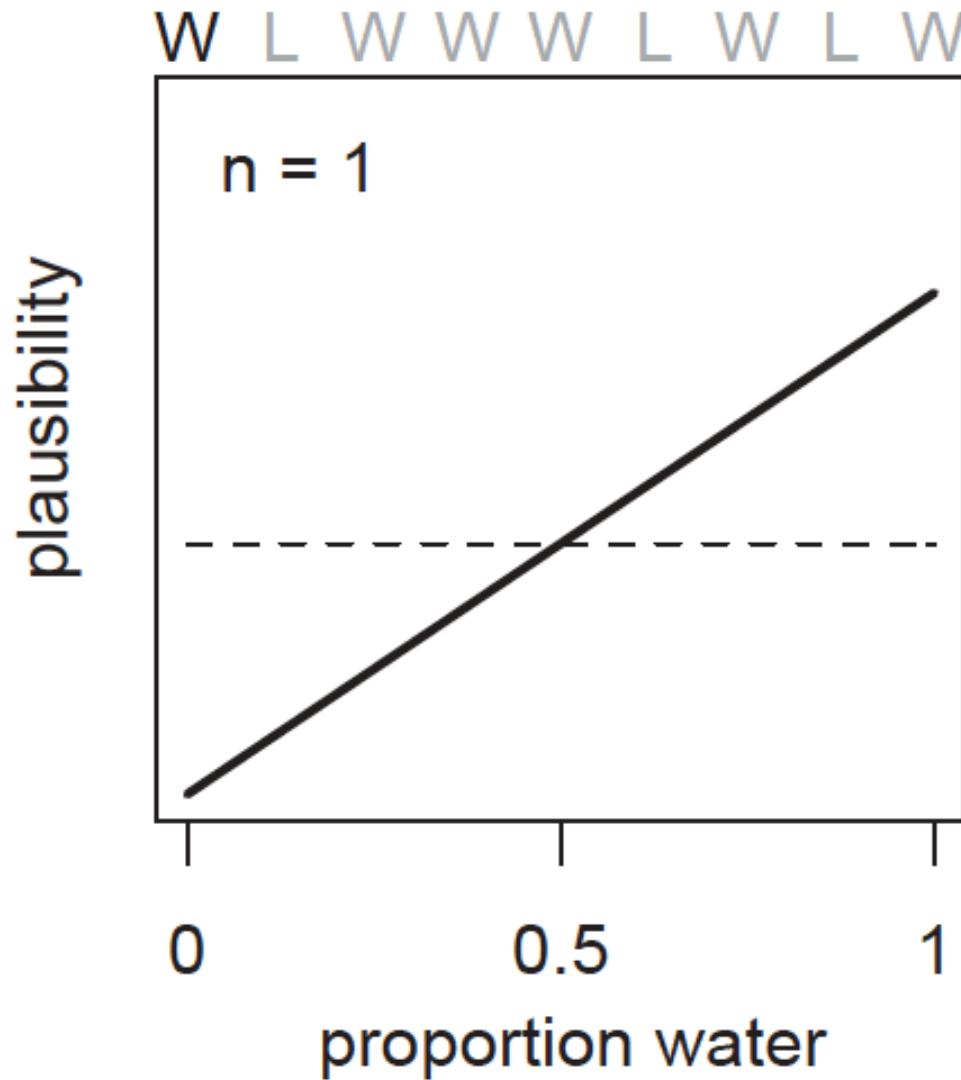


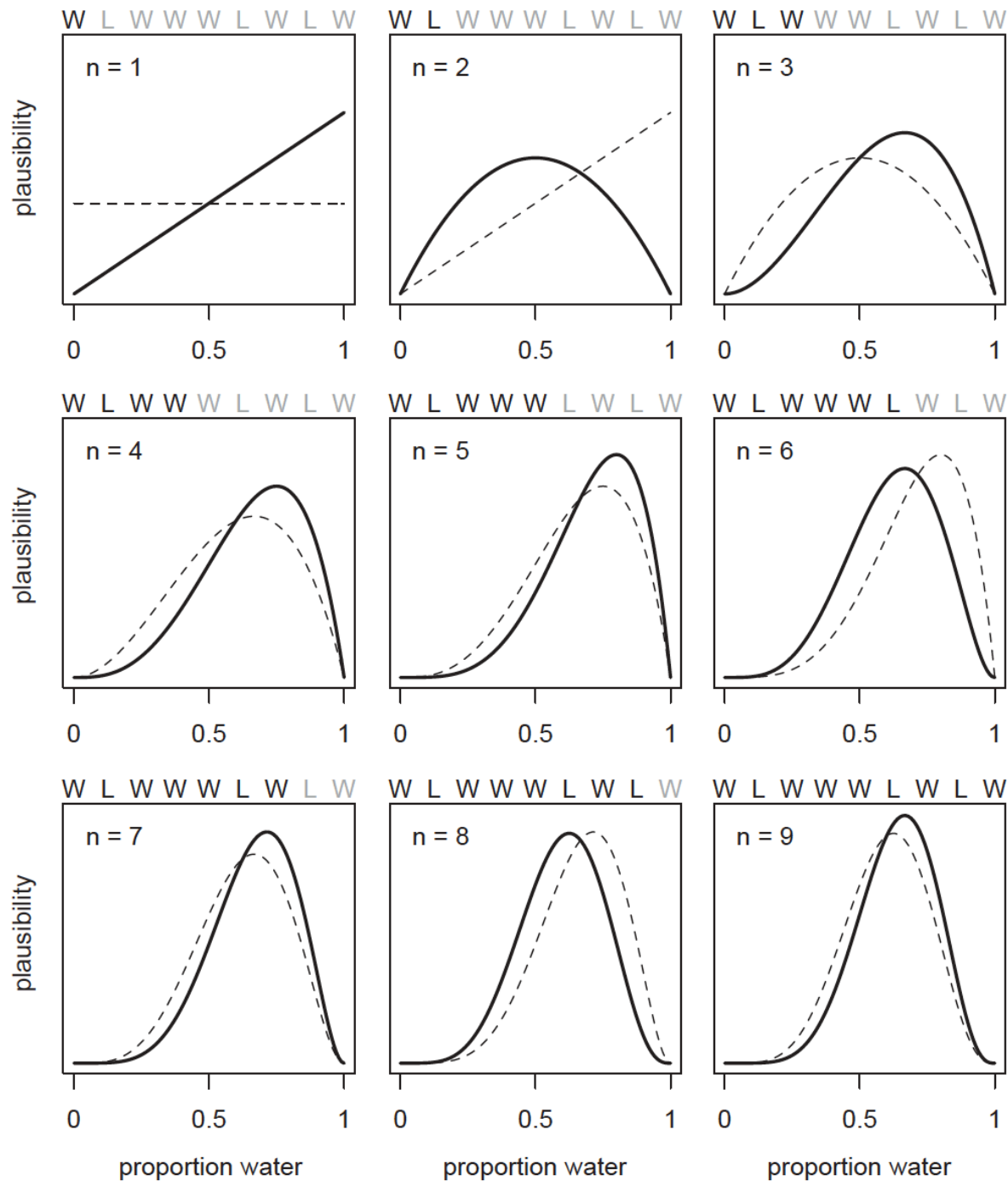
Datos: W L W W W L W L W

Historia para los datos

- (1) La verdadera proporción de superficie cubierta por agua es p
- (2) Cada lanzamiento del globo tiene una probabilidad p de producir una observación de agua (W) y una probabilidad $1-p$ de producir una observación de tierra (L)
- (3) Cada lanzamiento de globo es independiente de los otros

Actualizar





El proceso de inferencia

- (1) El número de maneras en que cada conjetura puede producir una observación
- (2) El número acumulado de maneras en que cada conjetura puede producir todos los datos
- (3) La plausibilidad inicial para cada conjetura

Verosimilitud

- Escogemos una expresión matemática que pueda explicar (generar) las observaciones
- En este caso hay dos opciones para cada dato
- Cada lanzamiento es independiente de los otros
- La probabilidad p de observar W es la misma en todos los lanzamientos
- -> Distribución binomial

Verosimilitud

- Probabilidad de que dado un valor de p , haya un número w de observaciones de W en n lanzamientos

$$\Pr(w|n, p) = \frac{n!}{w!(n-w)!} p^w (1-p)^{n-w}$$

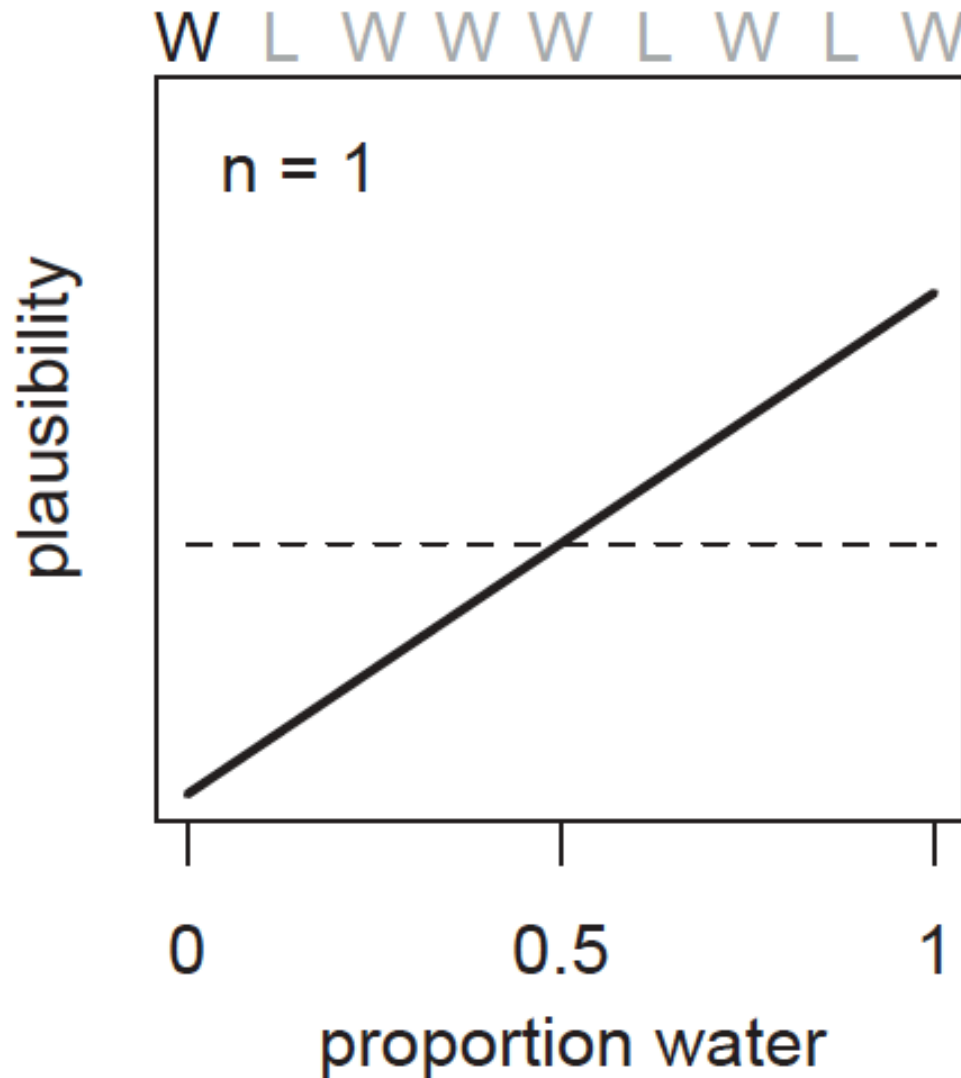
Prior

- Todos los valores en el rango $[0,1]$ son igualmente probables

$$\Pr(p) = \frac{1}{1 - 0} = 1$$

- *Prior débilmente informativo/a*
- Al actualizar, la *posterior* se vuelve la *prior* de la estimación siguiente

Prior: Línea punteada



Posterior

- Objetivo: Dados los datos (!) ¿cuál es la probabilidad de que el parámetro tenga cierto valor?

$$\Pr(p|n, w)$$

- Regla de Bayes (obviando n):

$$\Pr(p|w) = \frac{\Pr(w|p) \Pr(p)}{\Pr(w)}$$

Posterior

$$\text{Posterior} = \frac{\text{Verosimilitud} \times \text{Prior}}{\text{Verosimilitud promedio}}$$

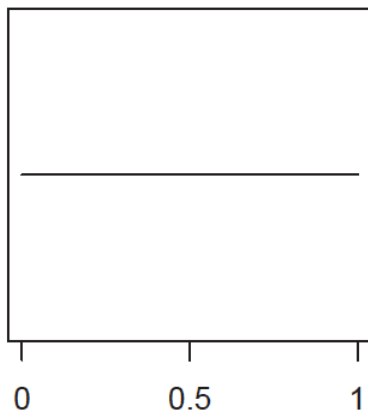
$$\Pr(w) = E(\Pr(w|p)) = \int \Pr(w|p) \Pr(p) dp$$

Sirve para que la probabilidad posterior sume 1

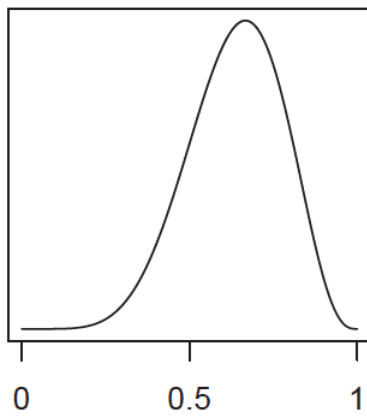
prior

likelihood

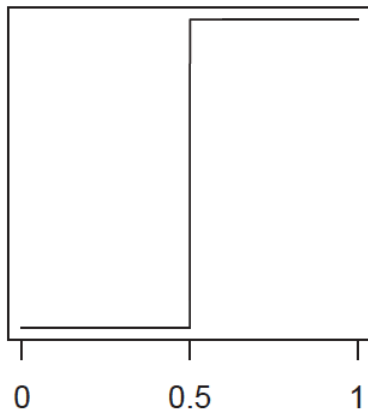
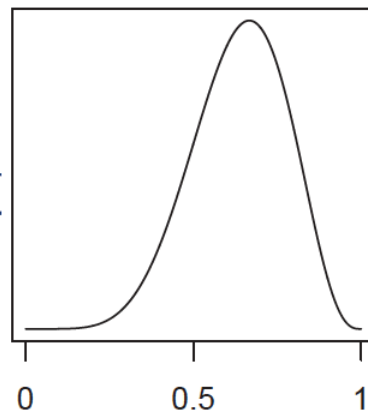
posterior



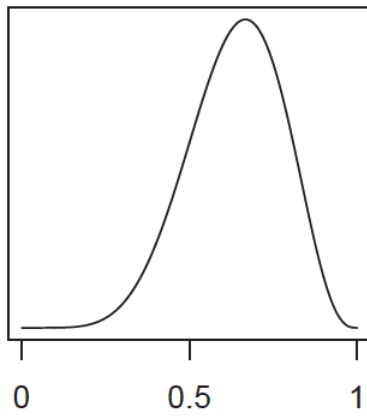
\times



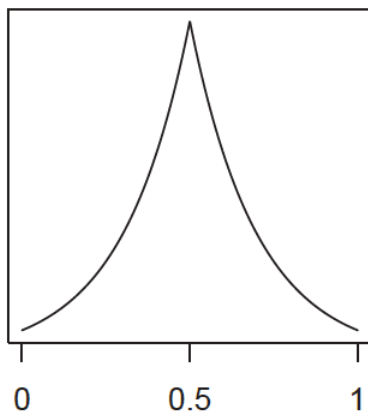
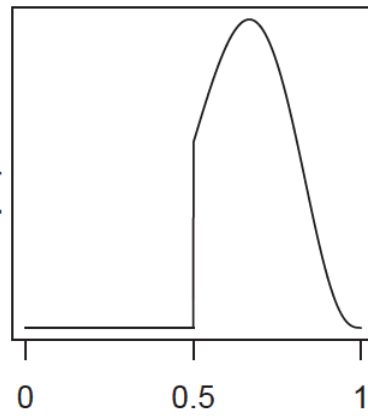
\propto



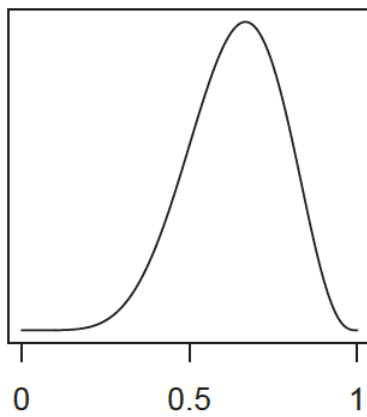
\times



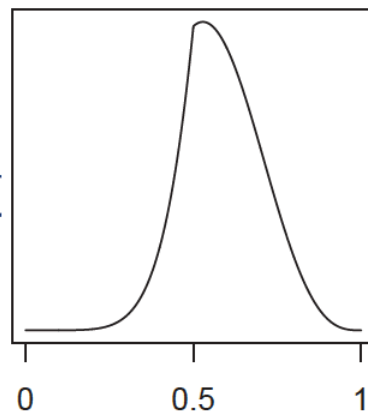
\propto



\times

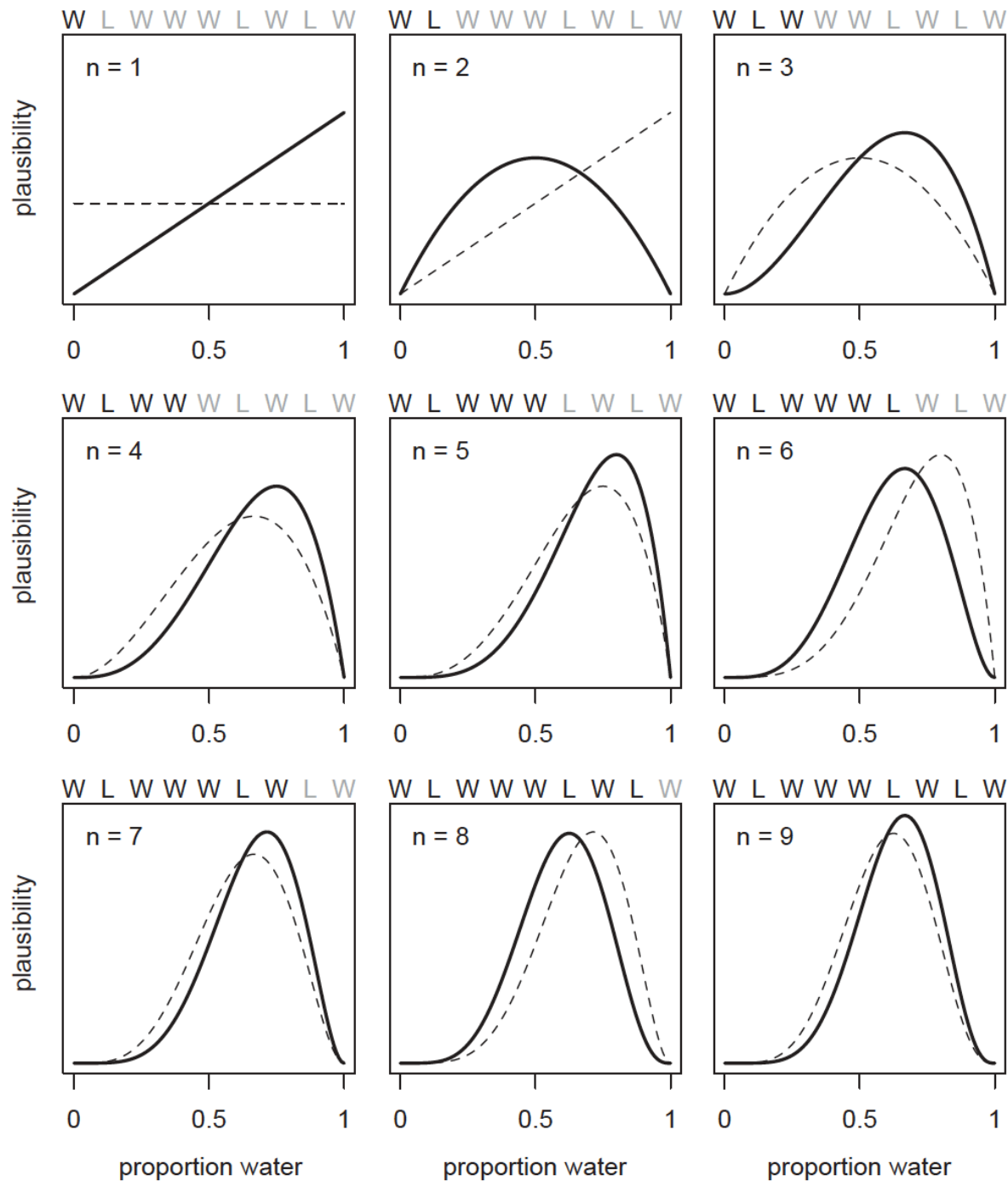


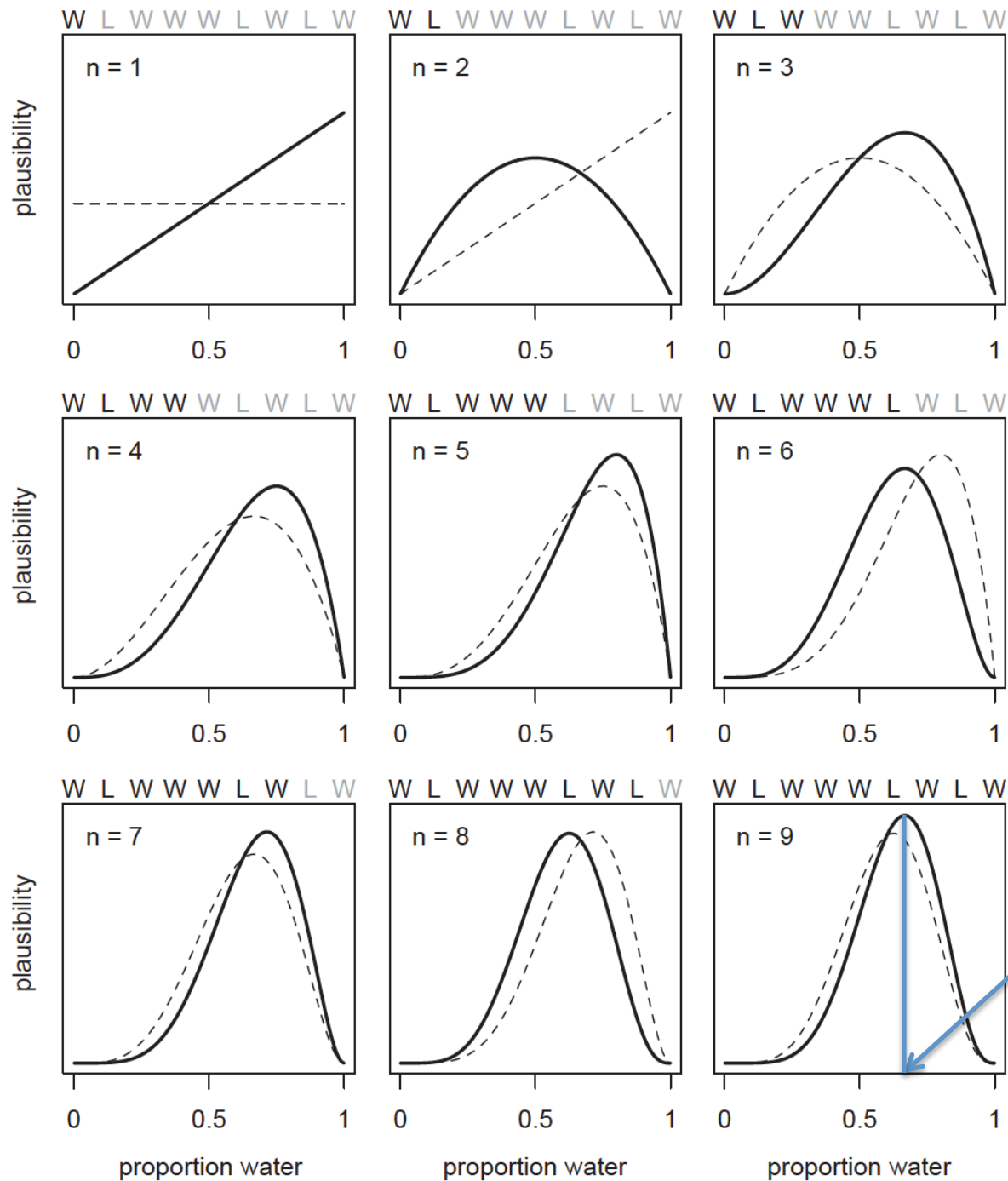
\propto



Estimación de Posterior

- Grid (fuerza bruta)
- Maximum Likelihood Estimation (frecuentista)
- Monte Carlo (eficiente)
- Markov Chain Monte Carlo (muy eficiente)





Máxima
probabilidad

Máxima Verosimilitud

- Aproximación no bayesiana (frecuentista)
- Optimización (hallar el máx/min de la posterior)
- Valor óptimo $p = 2/3$
- ¿Por qué?