Idea del Algoritmo del Muestreo Thompson

El problema del Bandido Multibrazo

D1

D2

D3

D4

D5

El problema del Bandido Multibrazo

- Tenemos **d** brazos. Por ejemplo, los brazos son anuncios que mostramos a los usuarios cuando se conectan a una página web.
- Cada vez que un usuario se conecta a la página web, se desencadena una ronda.
- En cada ronda, **n**, se elige uno de los anuncios a ser mostrado al usuario.
- A cada ronda **n**, el anuncio **i** da una recompensa: $r_i(n) \in \{0, 1\}$

$r_i(n) = 1$ Si el usuario hace click en el anuncio i en la ronda n
$r_i(n) = 0$ Si el usuario no hace click en el anuncio i en la ronda n

- Nuestra meta es maximizar la recompensa a través de las rondas que se lleven a cabo.

Inferencia Bayesiana

- El anuncio **i** da una recompensa **y** que sigue una distribución de Bernoulli

$$p(\boldsymbol{y}|\theta_i) \sim \mathcal{B}(\theta_i)$$

- θ_i es desconocido pero se supone que tiene una distribución uniforme $p(\theta_i) \sim \mathcal{U}([0,1])$, llamada distribución a priori
- Regla de Bayes: aproximamos $\theta_i(n)$ por la distribución a posteriori

$$\underbrace{p(\theta_i|\mathbf{y})}_{\text{posterior distribution}} = \frac{p(\boldsymbol{y}|\theta_i)p(\theta_i)}{\int p(\boldsymbol{y}|\theta_i)p(\theta_i)d\theta_i} \propto \underbrace{p(\boldsymbol{y}|\theta_i)}_{\text{likelihood function}} \times \underbrace{p(\theta_i)}_{\text{prior distribution}}$$

- Obtenemos $p(\theta_i|\mathbf{y}) \sim \beta($ número de éxitos + 1 , número de fracasos + 1 $)$
- A cada ronda, **n**, obtenemos un valor aleatorio $\theta_i(n)$ de la distribución a posteriori $p(\theta_i|\boldsymbol{y})$, para cada **i**.
- A cada ronda **n**, seleccionamos el anuncio **i** con el mayor valor $\theta_i(n)$.

Algoritmo del Muestreo Thompson

PASO 1: A cada ronda **n**, se consideran dos números para cada anuncio **i**:

- $N_i^1(n)$ El número de veces que el anuncio **i** recibe una recompensa 1 hasta la ronda **n**
- $N_i^0(n)$ El número de veces que el anuncio **i** recibe una recompensa 0 hasta la ronda **n**

PASO 2: Para cada anuncio i, se elige un valor aleatorio generado a partir de la distribución:

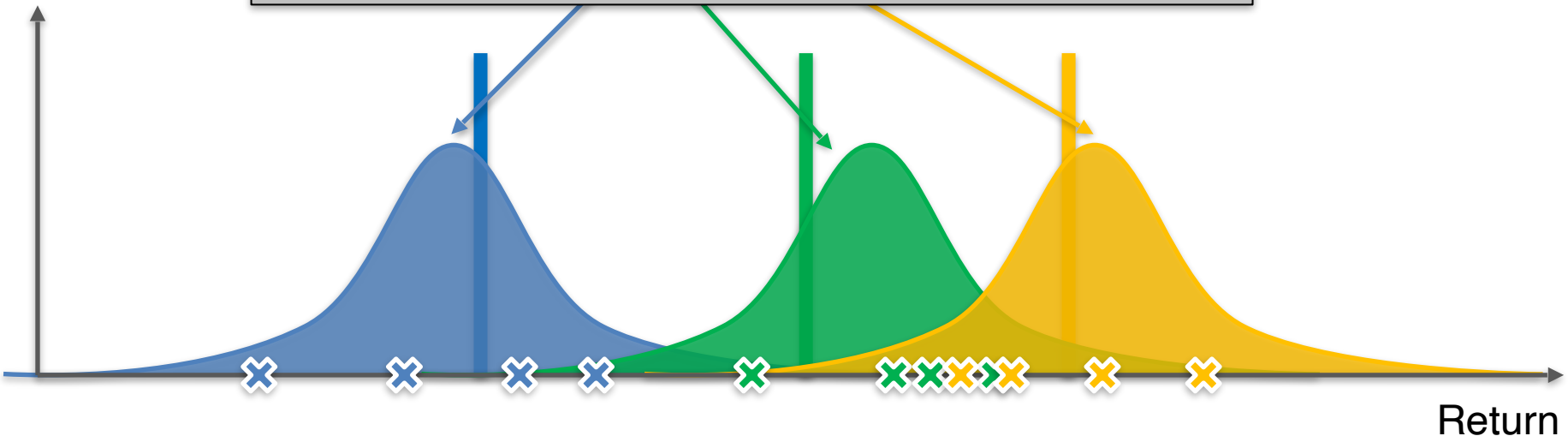
$$\theta_i(n) = \beta(N_i^1(n) + 1, N_i^0(n) + 1)$$

PASO 3: Elegimos el anuncio con mayor valor

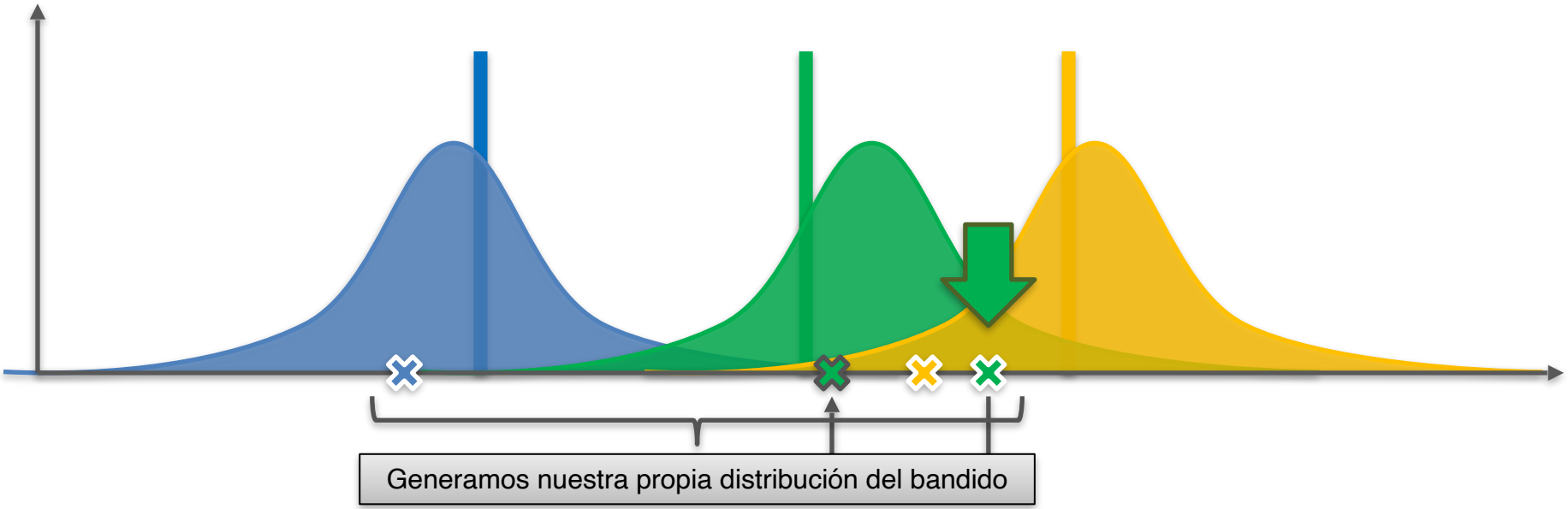
Algoritmo del Muestreo Thompson

Donde pensamos que estarán los valores μ*

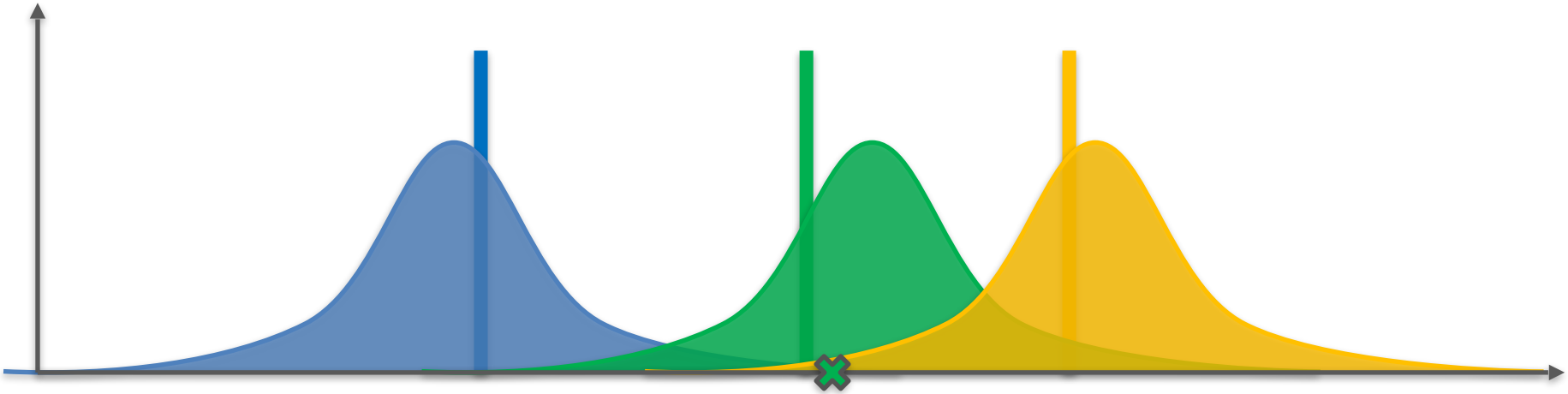
i.e. NO queremos averiguar las distribuciones de las máquinas



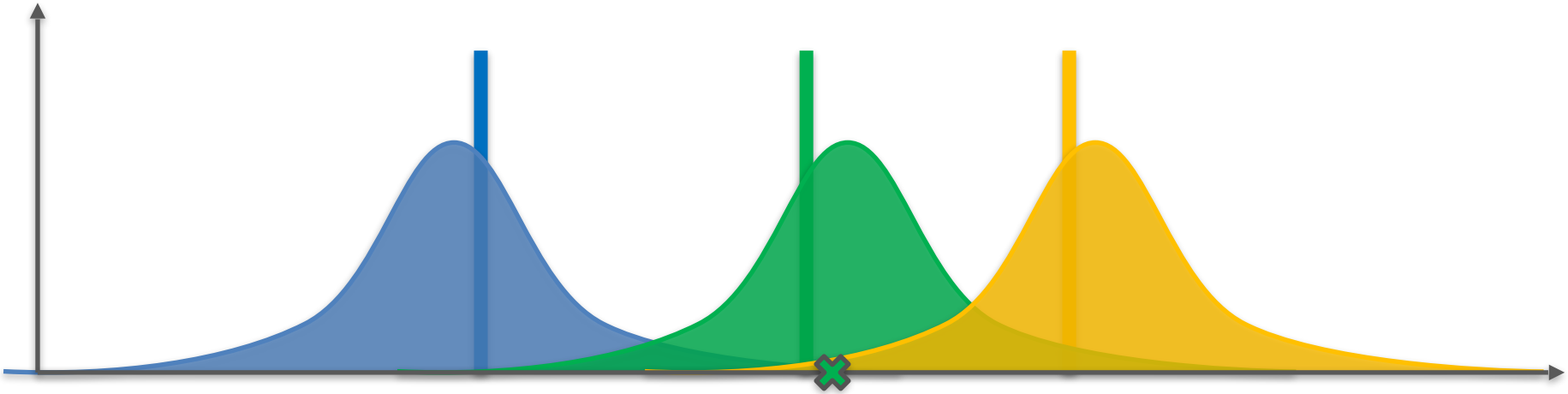
Algoritmo del Muestreo Thompson



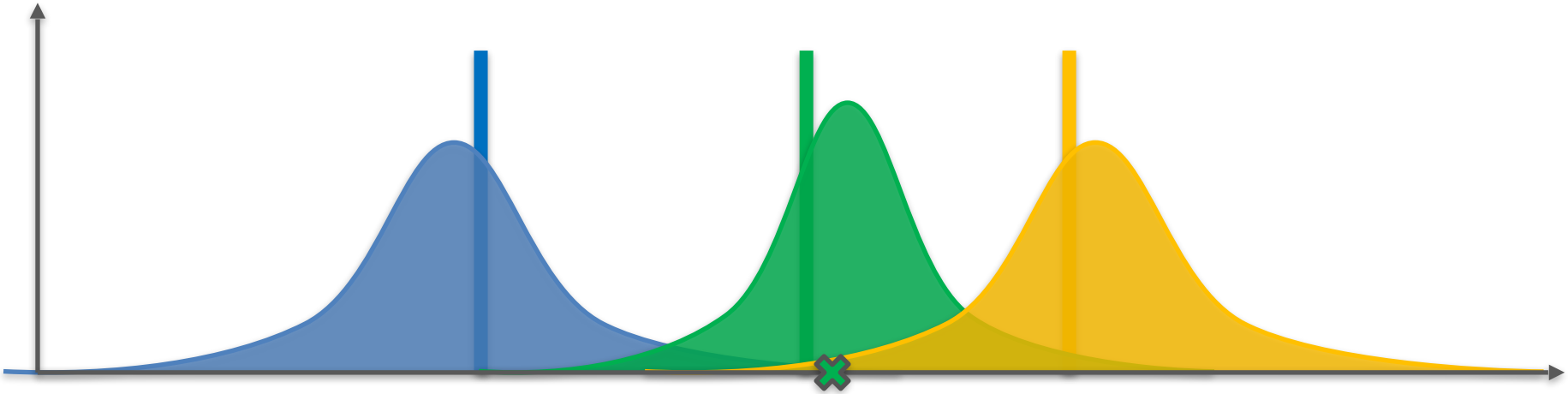
Algoritmo del Muestreo Thompson



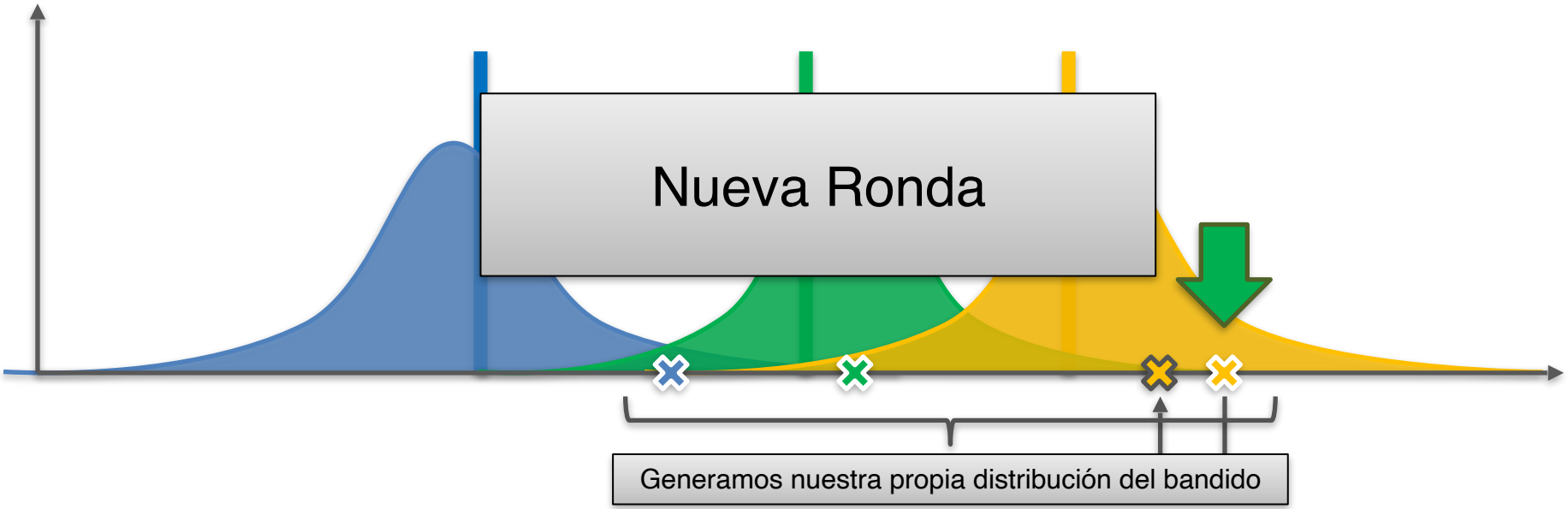
Algoritmo del Muestreo Thompson



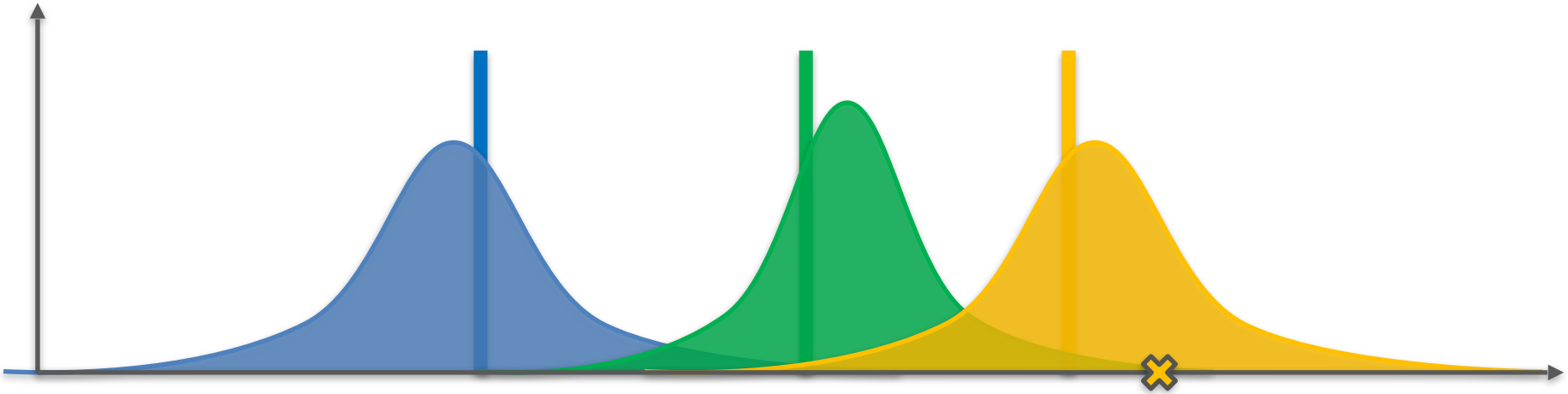
Algoritmo del Muestreo Thompson



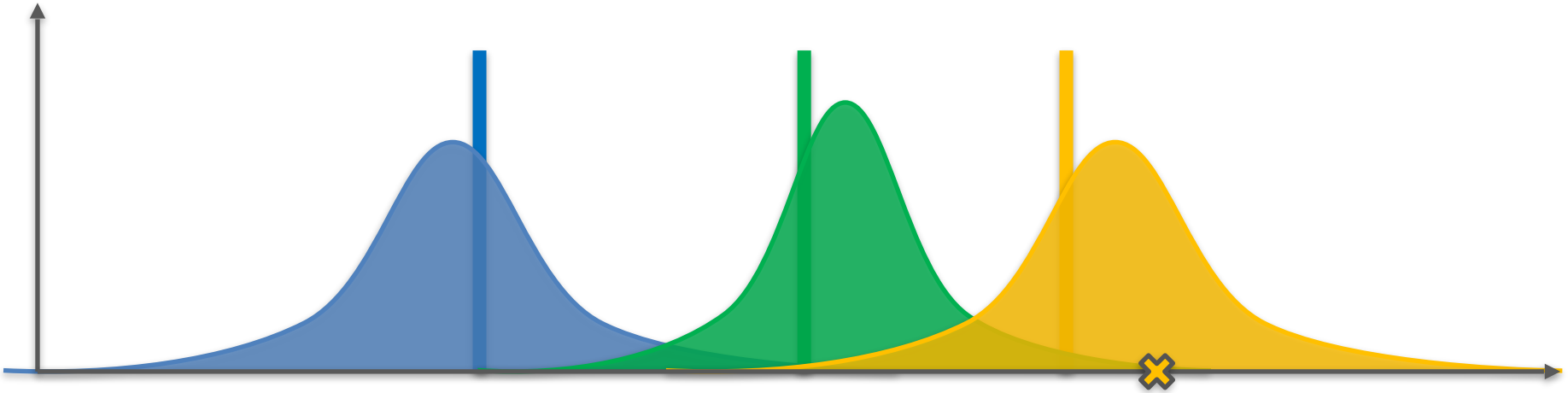
Algoritmo del Muestreo Thompson



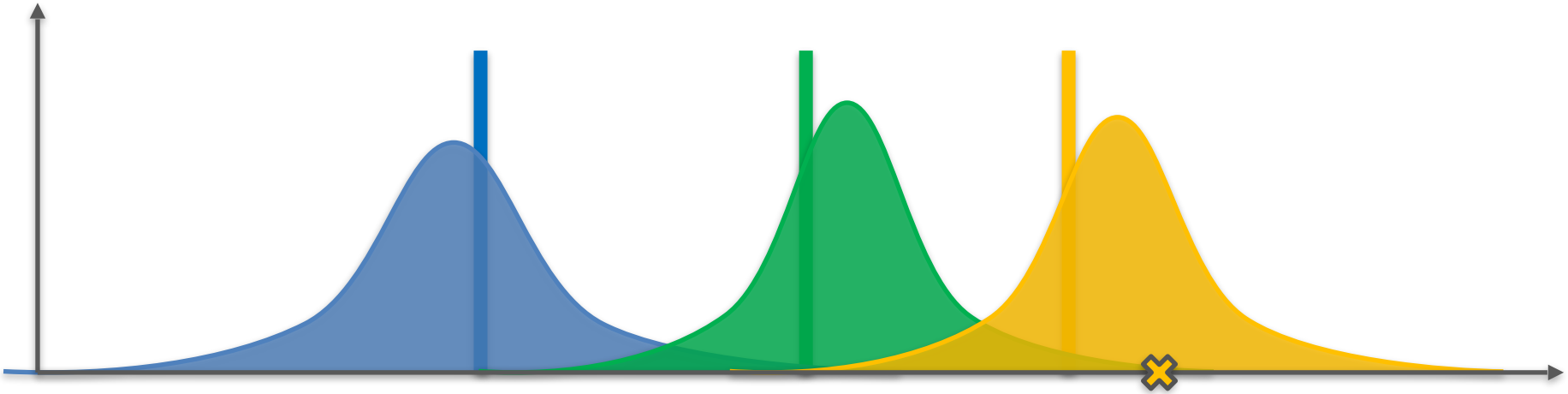
Algoritmo del Muestreo Thompson



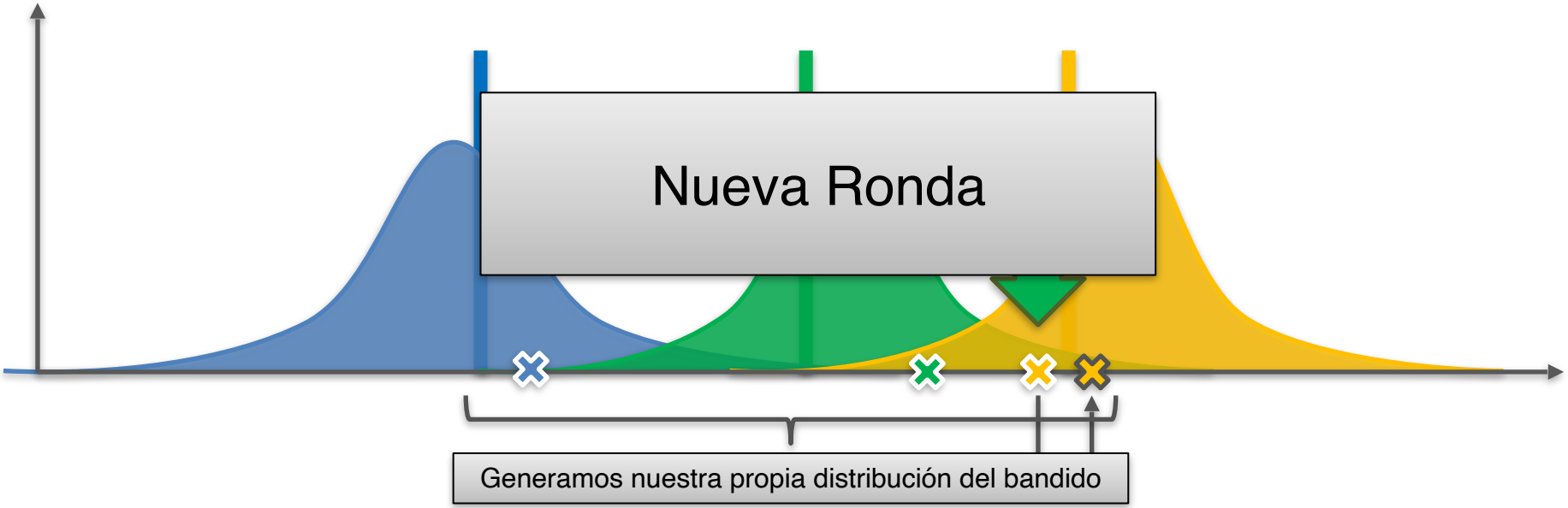
Algoritmo del Muestreo Thompson



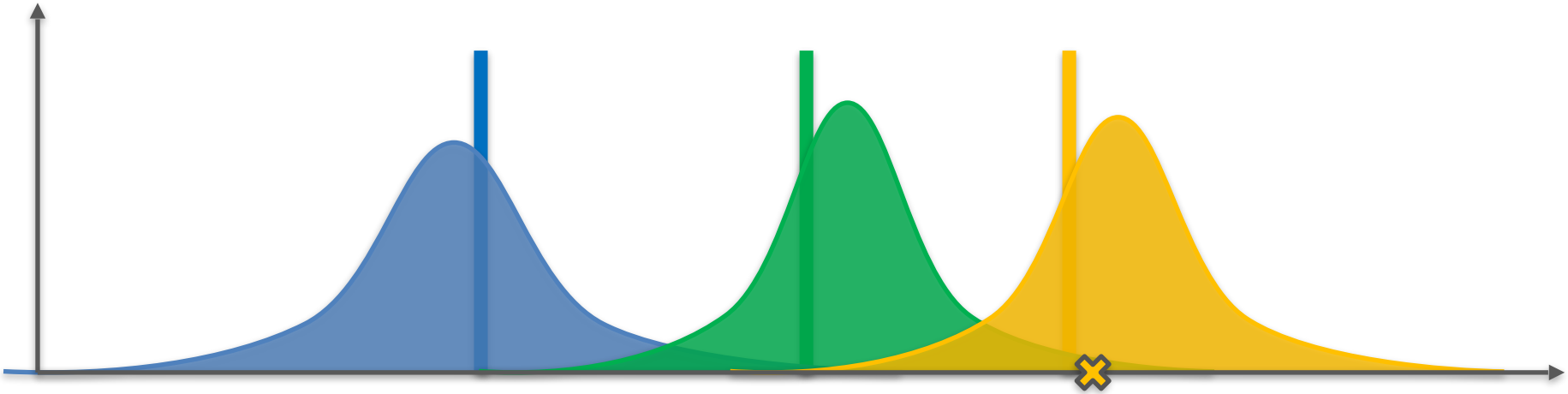
Algoritmo del Muestreo Thompson



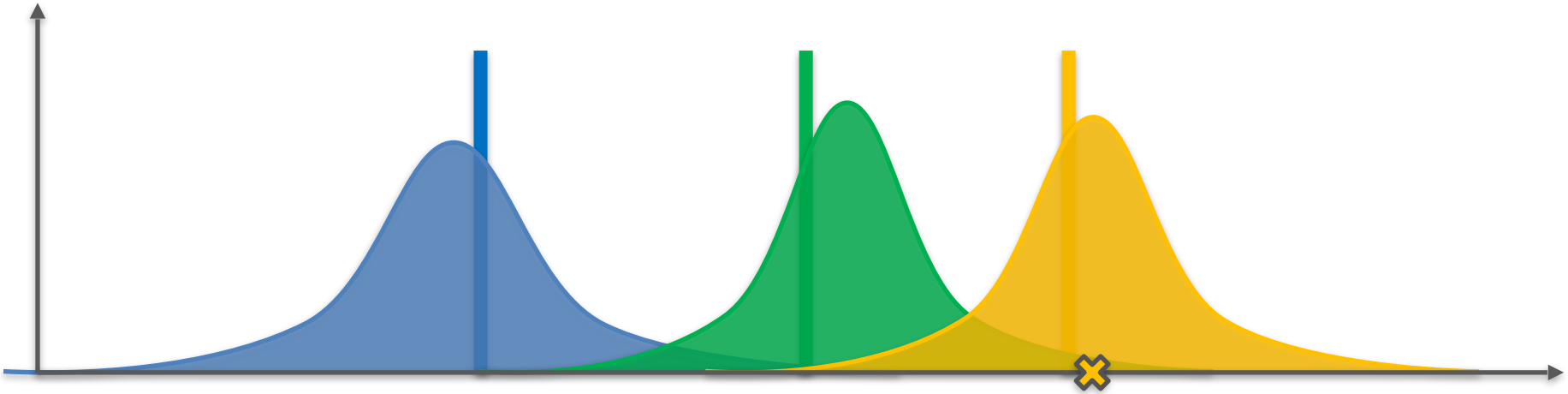
Algoritmo del Muestreo Thompson



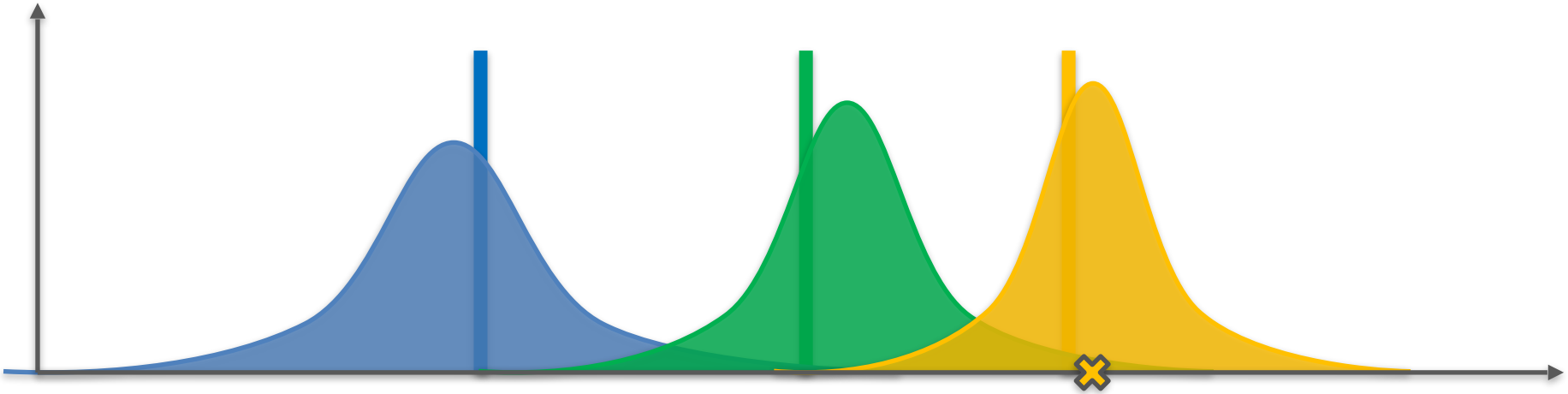
Algoritmo del Muestreo Thompson



Algoritmo del Muestreo Thompson



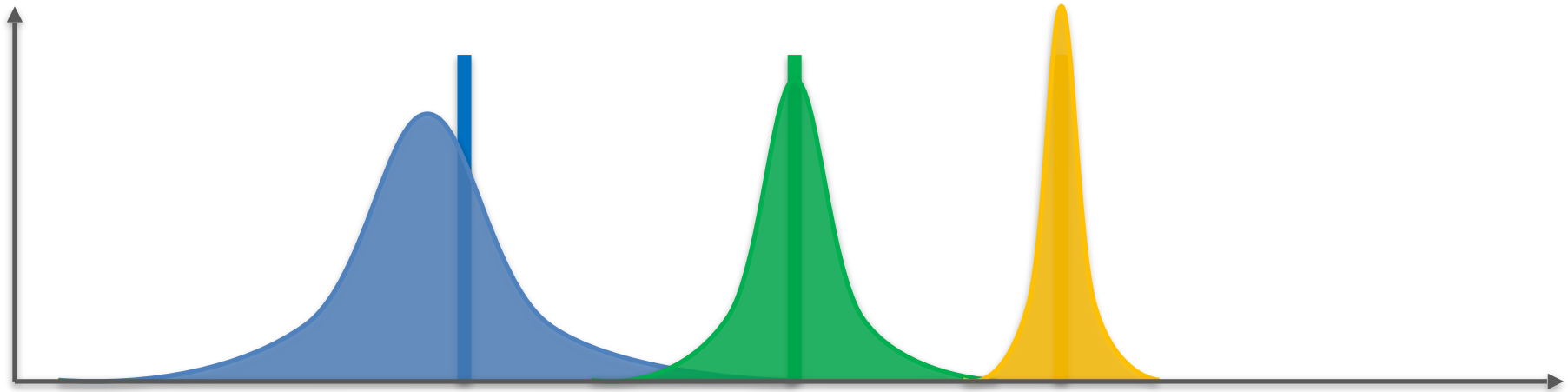
Algoritmo del Muestreo Thompson



Algoritmo del Muestreo Thompson

Y así sucesivamente…

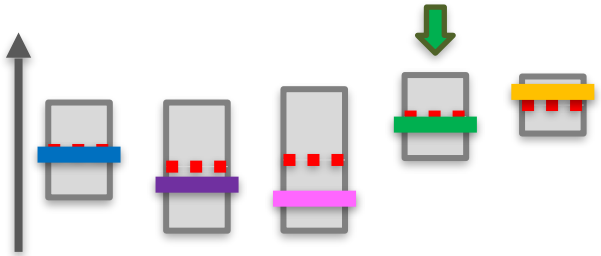
Algoritmo del Muestreo Thompson



UCB vs Muestreo Thompson

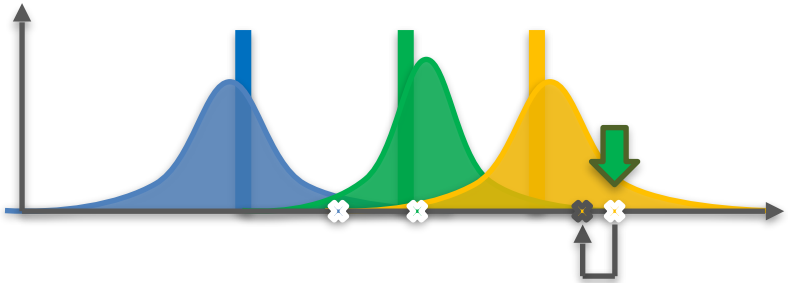
Algoritmo del Muestreo Thompson

UCB



- Determinista
- Requiere actualizar a cada ronda

Muestreo Thompson



- Probabilístico
- Se amolda gracias al feedback a posteriori
- Más evidencias empíricas