

# A/B Testing bayesiano

Mario González Olmedo<sup>1</sup>  
mario.gonzalezolmedo@mercadolibre.com  
1. Mercado Libre

**Motivación.** Buscando mejorar la experiencia de nuestros usuarios, analizamos grandes cantidades de datos a fin de tomar decisiones basadas en la evidencia. Para analizar el impacto en las métricas de interés que tendrá un cambio propuesto en alguno de nuestros productos, contamos con una plataforma que permite realizar **A/B Testing** a escala sin necesidad de contar con conocimientos estadísticos para implementar dichos experimentos (ver Figura 1).

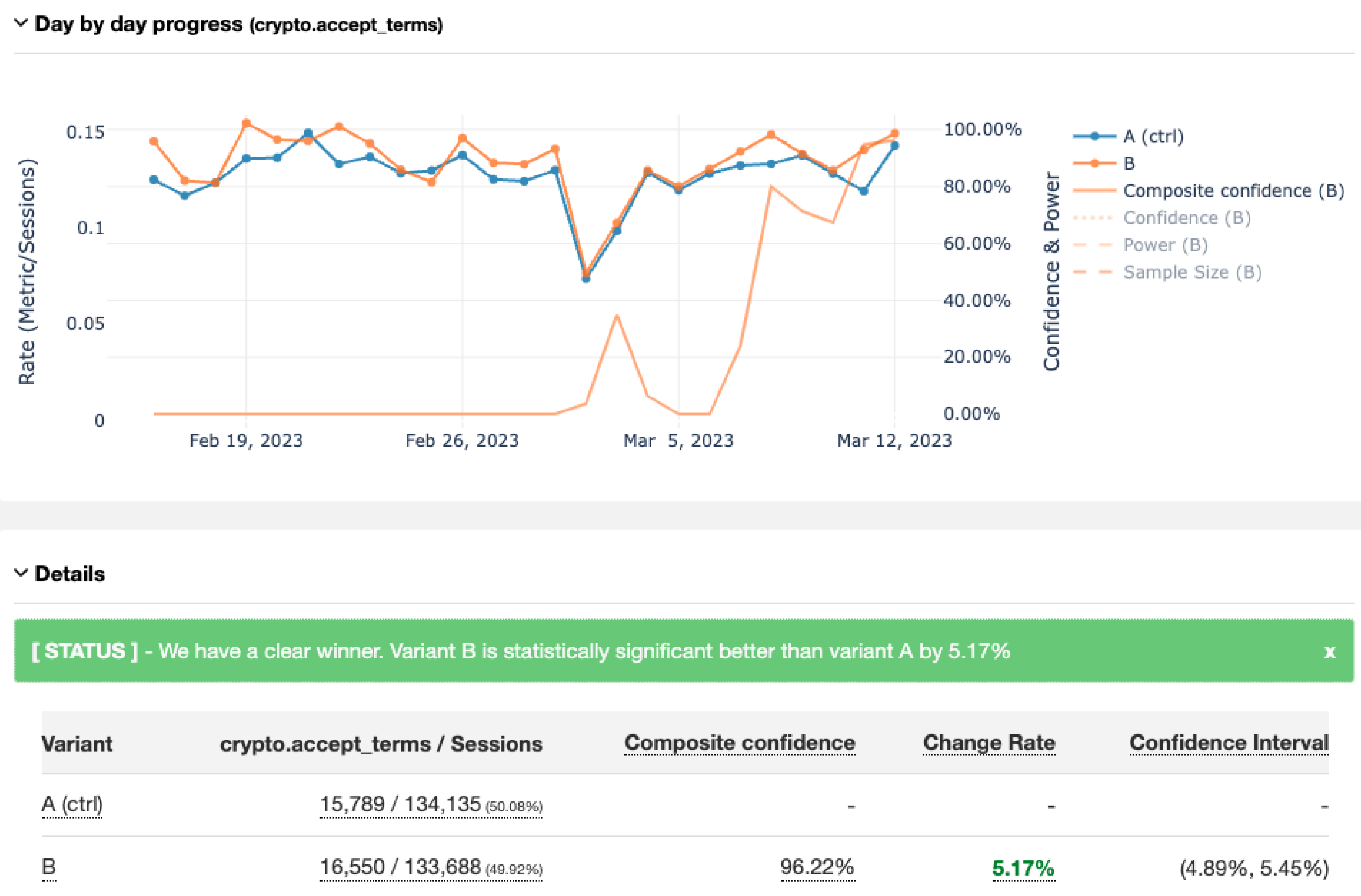


Figure 1: Plataforma de A/B Testing

**Métodos.** En un A/B Test entre dos variantes de una misma componente del sitio, se puede plantear un **test de hipótesis clásico**, en el cual la *hipótesis nula*  $H_0$  consiste en afirmar que *no hay diferencia entre las variantes*, y la *hipótesis alternativa*  $H_A$  afirma que *sí la hay*. Al monitorear los datos que el experimento va generando, es usual no esperar a alcanzar el número de observaciones para el cual fue desarrollado el mismo sino detenerlo cuando el  $p$ -valor llegue al umbral elegido, lo que se conoce como el **peaking problem**.

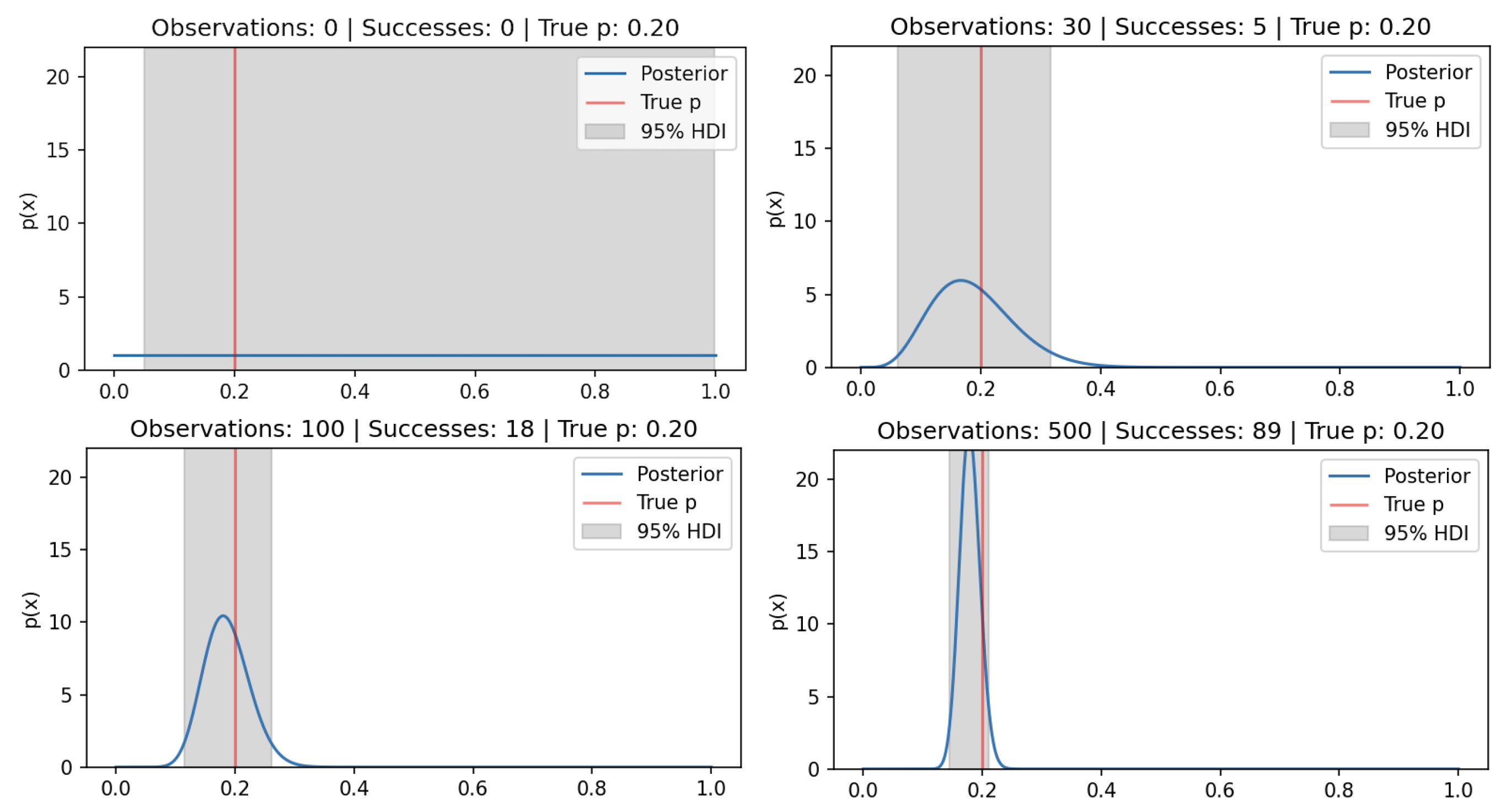


Figure 2: Método bayesiano en A/B Test

Un **test de hipótesis secuencial** es válido en todo momento del experimento pero su interpretación es compleja de transmitir a equipos que no cuentan con conocimientos estadísticos. El **A/B Testing Bayesiano** asocia una distribución de probabilidad a cada tasa de conversión, y la actualiza a medida que recibimos más observaciones (ver Figura 2), obteniendo una estimación de la diferencia entre la conversión de cada variante y la incertidumbre asociada a la misma (ver Figura 3).

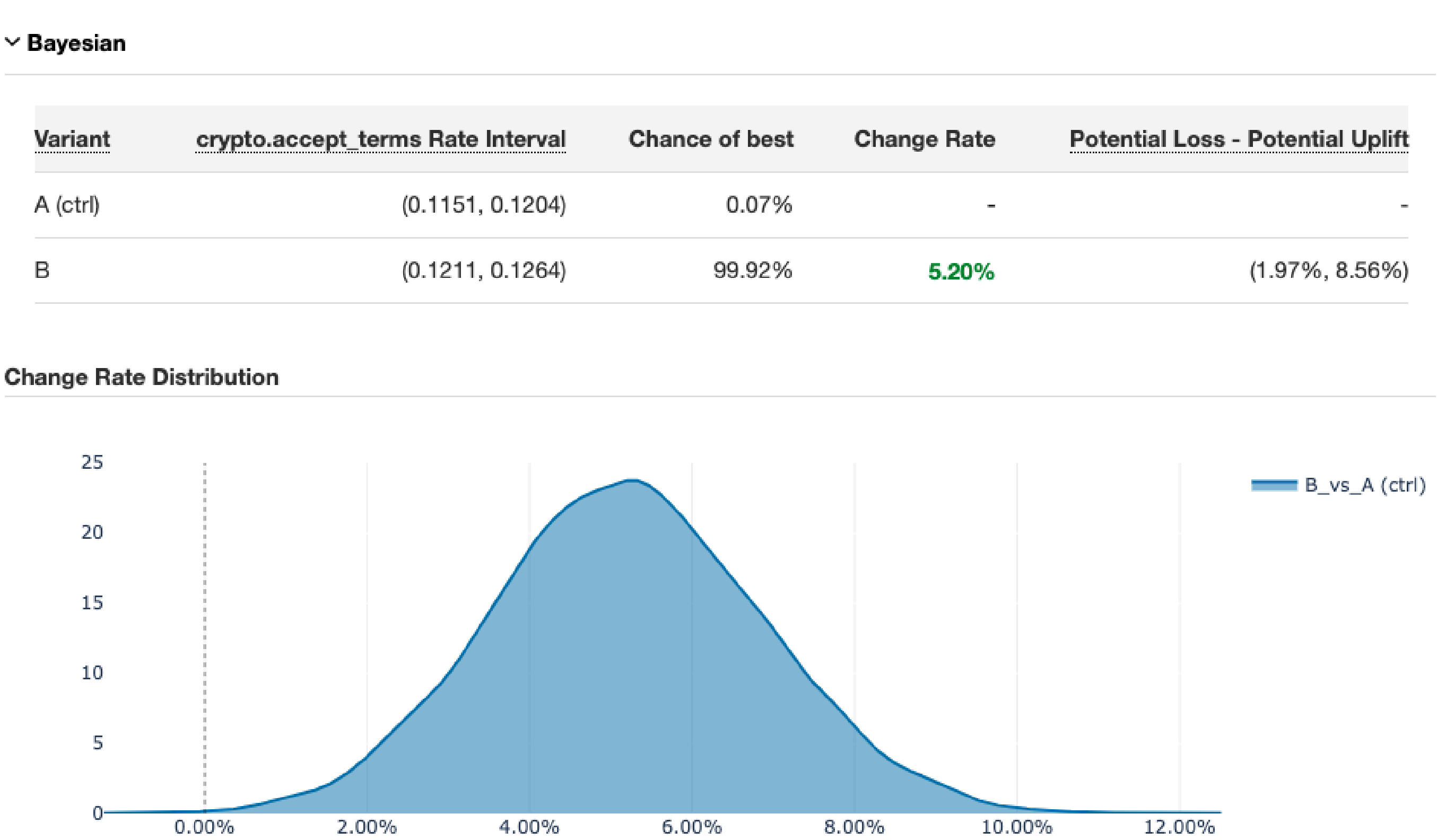


Figure 3: Impacto en la tasa de conversión (posterior)

**Resultados.** Utilizamos A/B Testing bayesiano en tres problemas:

- Se deseaba estimar el impacto en la tasa de conversión de varios descuentos en un producto de garantía extendida. Hay una cola muy larga de productos con pocas compras mensuales que requieren varios meses de A/B Testing para llegar a resultados concluyentes con los métodos frecuentistas (ver Figura 4).
- Se desarrolló un modelo de *machine learning* para excluir reclamos sobre un vendedor que impactan en su reputación, a fin de reducir el trabajo manual del equipo de Customer Experience, y se quería saber si se logra una reducción del número diarios de reclamos.
- Para aumentar la tasa de conversión entre los que llegan a los Términos y Condiciones (T&C) para operar criptomonedas en Brasil, el equipo implementó un A/B Test para estimar el impacto de un nuevo wording de los T&C.

En los tres casos, observamos que el **método bayesiano reduce considerablemente el tiempo de experimentación para alcanzar resultados significativos** ( $\sim 1/3$  que el test de hipótesis secuencial).

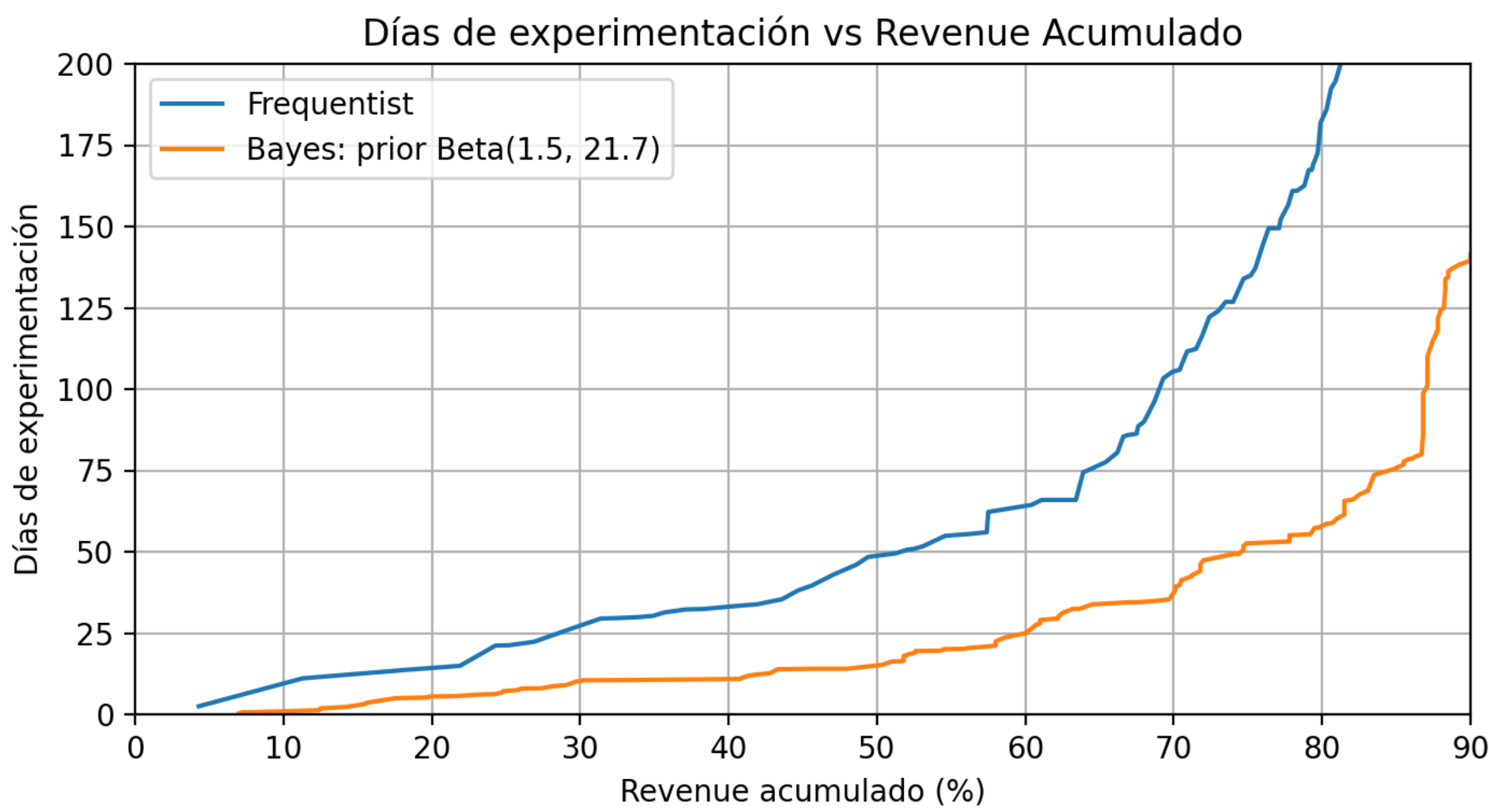


Figure 4: Comparación de métodos en garantía extendida.

**Conclusión.** Para analizar los datos generados por un A/B Test, observamos que el método bayesiano brinda información fácil de interpretar, y permite extraer resultados confiables para tomar decisiones de negocio basadas en evidencia en menos tiempo.

## References

[1] Kohavi R, Tang D, Xu Y. Trustworthy online controlled experiments: A practical guide to a/b testing. Cambridge University Press; 2020.