Thompson Sampling: Aplicaciones Economía Digital

Alvaro J. Riascos Villegas
Universidad de los Andes y Quantil

Septiembre de 2024

Contenido

Introducción

- El sector de servicios es el 80 % del PIB de los Estados Unidos. Una gran parte de este sector involucra transacciones en línea.
- El costo de experimentar lo domina el costo de oportunidad por no prestar el servicio óptimo.
- Experimetar tiene un costo marginal muy bajo (en contraposición a lo que sucede con otro tipo de servicios que requieren estudios de mercado, etc.).
- Otra carcaterística importante es la posibilidad de hacer un monitoreo continuo de cualquier experimento (e.g., UBER).
- Ilustramos el uso de la teoría de bandidos multiarmados para descubrir conocimiento de la forma más rápida y económica posible.

A/B Testing: Asignación aleatoria

- Dos versiones de una pagina web. La tasa de conversion en cada una es $p = 0{,}001$ y $p = 0{,}0011$.
- Las paginas se asignan de forma aleatoria a cada visitante.
- Para detectar esas diferencias en tasas de conversion aleatorizando 50-50, se necesitan muchas observaciones: Aproximadamente 2.5 millones de observaciones de cada arma para una confianza del 95 % o 0.5 millones cada una con 50 % de confianza.

A/B Testing: Thompson sampling

- Se hacen 100 simulaciones. Cada simulación se interrumpe cuando el arrepentimiento relativo es menor al 5 %.
- En cada simulacion (que pueden ser millones de interacciones con la página web), después de 100 observaciones (visitas a cada pagina por simulación), se actualizan los parámteros de la prior usando la regla de Bayes.
- En las simulaciones, en el 84 % se eligió la página óptima (B).

A/B Testing

- La figura muestra (panel izquierdo) el número de simulaciones necesarias para terminar el experimento. Por ejemplo, en el 70 % de los experimentos, basta con medio millón de interacciones para terminar el experimento.

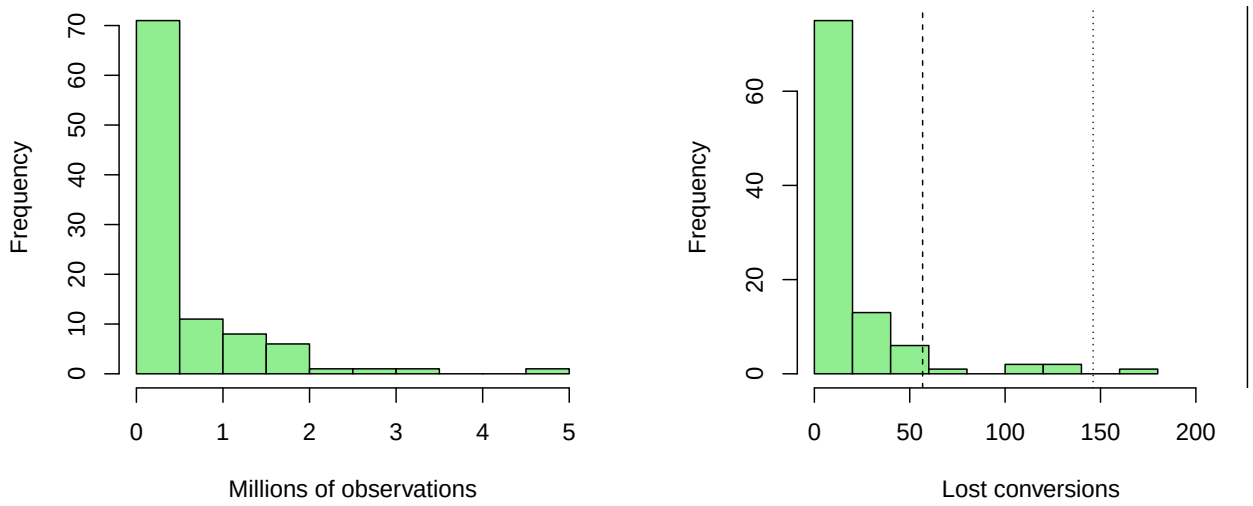

Figure 1: *(a) Histogram of the number of observations required in 100 runs of the binomial bandit described in Section 4.1. (b) The number of conversions lost during the experiment period. The vertical lines show the number of lost conversions under the traditional experiment with 95% (solid), 50% (dashed), and 84% (dotted) power.*

Figura: AB Testing

Contenido

Recomendación de Artículos de Noticias

- Una pagina web interactua con una sucesion de usuarios: $t = 1, 2, 3..$
- En cada ronda, el administrador de la página web, observa un vector de características del usuario t, $z_t \in R^d$ (e.g., visitas anteriores, características demográficas, geográficas, día de la visita, etc). Este elige un artículo para mostrarle $\Xi = \{1, ..., k\}$ y se observa una recompensa $r_i \in \{0, 1\}$ (i.e., le gusto o no el artículo).
- Suponemos que (i.e., bandidos con contexto):

$$P(r_i = 1 \mid x_t = i, \theta_i, z_i) = g(z_t^T \theta_i) = \frac{1}{1 + e^{-z_t^T \theta_i}}$$

- Definimos el arrepentimiento (i.e., error) como:

Recomendación de Artículos de Noticias

- Definimos el arrepentimiento (i.e., error) como:

$$regret_t(\theta_1, ..., \theta_k) = \max_i g(z_t^T \theta_i) - g(z_t^T \theta_i)$$

- Para más detalles véase: A Contextual-Bandit Approach to Personalized News Articles Recommendation. Li, Langford, Schapire [2012].

Recomendación de Artículos de Noticias

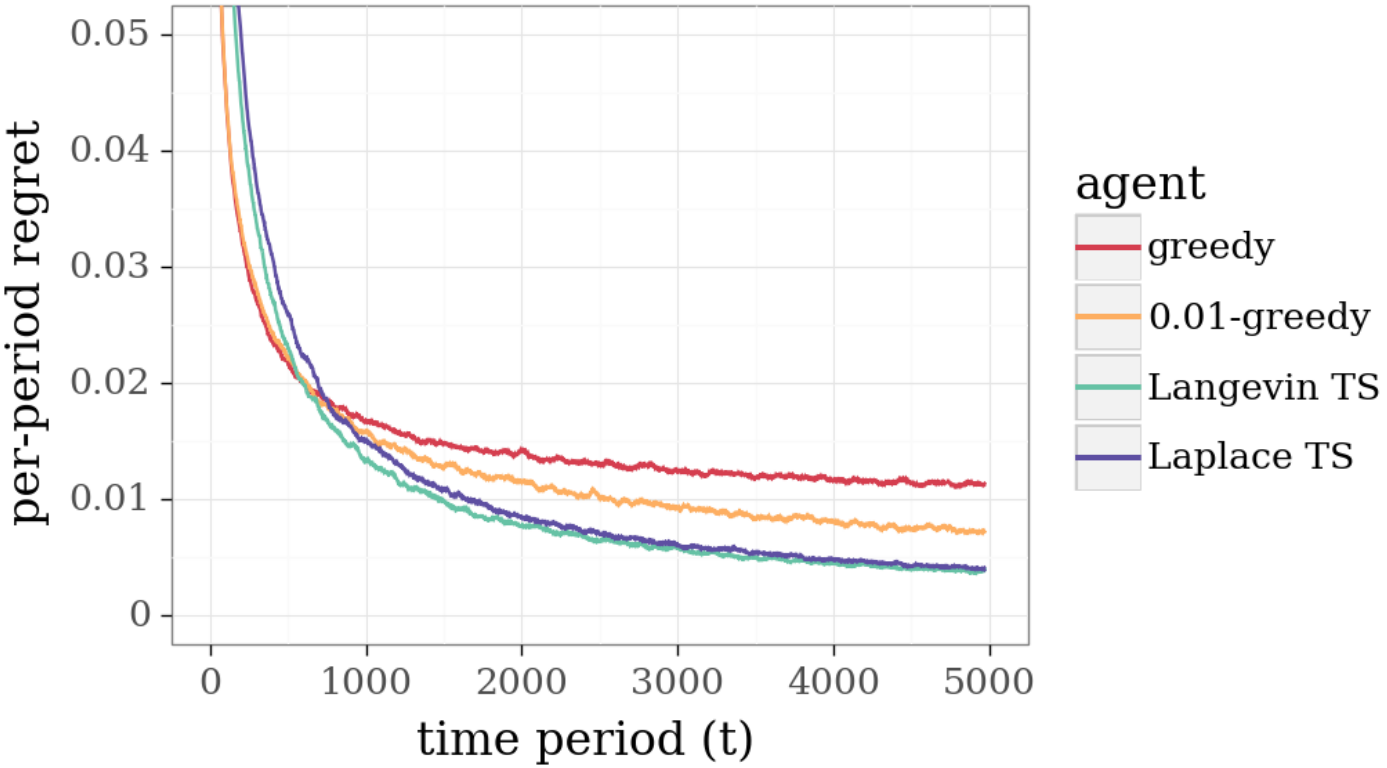

Figure 7.1: Performance of different algorithms applied to the news article recommendation problem.

Figura: $k = 3, d = 7, z_1 = 1$ y las otras seis características son binarias y se generan aleatoriamente con una distribución de Bernoulli con probabilidad de acierto $\frac{1}{6}$. Promedios de 2,000 simulaciones. $\theta_i \sim N(0, I)$.

Contenido

Modelo

- Un agente tiene para la venta $i = 1, ..., n$ productos. El beneficio neto de vender i es p_i. Dado que los productos pueden ser sustitutos o complementos, el objetivo es armar canastas de productos que maximizen el beneficio neto.
- $x_t \in \{0, 1\}^n$. Variable indicador de que productos ofrece en t.
- Una vez se ofrecen los productos se observa una demanda d_i.
- Sea θ una matriz $k \times k$.
- Suponemos que: $log(d_i) \mid \theta, x \sim N((\theta x)_i, \sigma^2)$ donde la varianza es conocida.
- Obsérvese que si el producto i es ofrecido: $(\theta x)_i = \theta_{ii} + \sum_{j \neq i} x_j \theta_{ij}$.
- Obsérvese que la demanda de i depende de todos los productos ofrecidos.

Modelo

- Cuando se ofrece x la recompensa esperada es:

$$E\left[\sum_{i=1}^{n} p_i x_i d_i \mid \theta, x\right] = \sum_{i=1}^{n} p_i x_i e^{(\theta x)_i + \frac{\sigma^2}{2}}. \qquad (1)$$

  y el obejetivo del agente es maximizar este pagado esperado. Para hacer esto tiene que aprender θ.
- Supongamos que $p(\theta)$ es Normal multivariada (prior conjugada).

Canastas de Productos

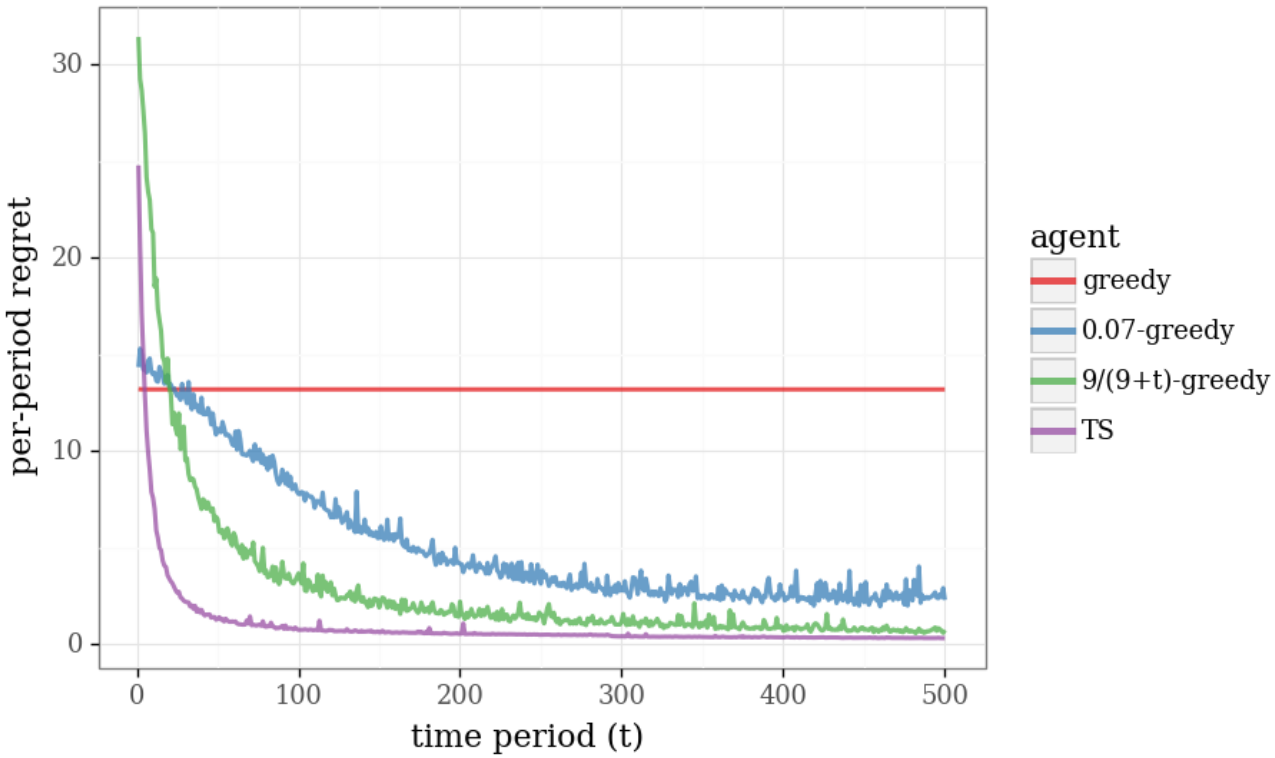

Figure 7.2: Regret experienced by different learning algorithms applied to product assortment problem.

Figura: Canastas de Productos

Contenido

Modelo de cascadas

- Supongamos que queremos recomendar $J \leq K$ items (i.e., páginas web) de un máximo de K.
- Sean $\theta \in [0,1]^K$ un vector de probabilidad de atracción de la página.
- Al uduario se le muestra en orden los items $x_t \in \{1, ..., K\}^J$.
- El usuario los examina en orden, la atractividad el item j es $\theta_{x_{t,j}}$.
- El recomendador observa $y_t = j$ si el usuario selecciona $x_{t,j}$ y $y_t = \infty$ si no selecciona nada.

Modelo de cascadas

- La recompensa es $r_t = r(y_t) = \mathbf{1}\{y_t \leq J\}$
- Para cada lista $x = (x_1, \ldots, x_J)$ y $\theta' \in [0,1]^K$, sea:

$$h(x, \theta') = 1 - \prod_{j=1}^{J} \left[1 - \theta'_{x_j}\right].$$

  luego la recompensa esperada en t es $E[r_t | x_t, \theta] = h(x_t, \theta)$.
- El objetivo es:

$$x^* \in max_{x:|x|=J} h(x, \theta) \tag{2}$$

  seleccionar los J items con las probabilidades de atracción más altas.
- El arrepentimiento por periodo es:

$$\text{regret}_t(\theta) = h(x^*, \theta) - h(x_t, \theta) \tag{3}$$

Modelo de cascadas

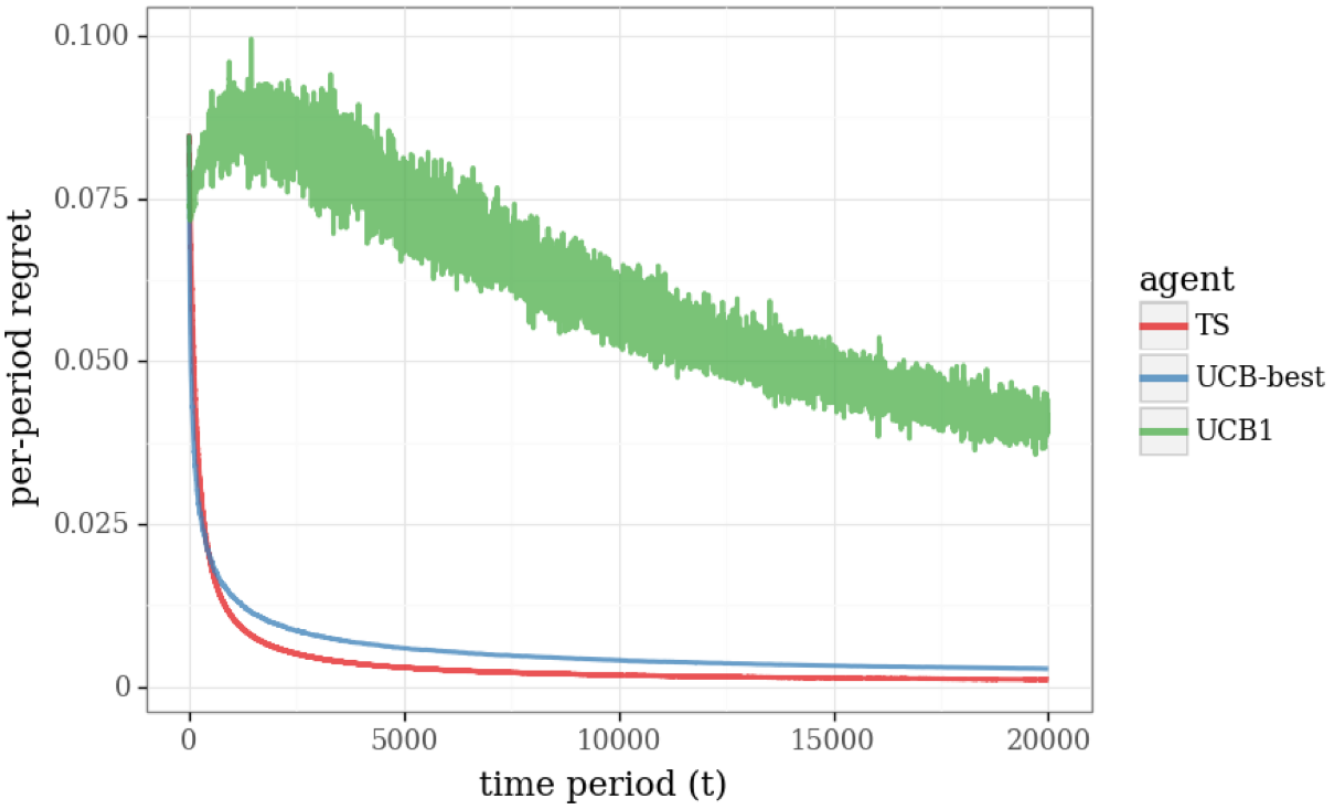

Figure 7.3: Comparison of CascadeTS and CascadeUCB with $K = 1000$ items and $J = 100$ recommendations per period.

Figura: Cascadas