Thompson Sampling: Tópicos Avanzados

Alvaro J. Riascos Villegas
Universidad de los Andes y Quantil

Septiembre de 2024

Contenido

Introducción

- Recordemos la versión general.
- Supongamos que un agente toma una sucesión de acciones $x_1, x_2, ...,$ donde cada $x_i \in \Xi$.
- Depués de la acción i el agente observa un resultado y_i, $y_i \backsim q_{\theta_i}(. \mid x_t)$.
- θ es desconocido pero el agente cuantifica su incertidumbre usando una prior $p(\theta)$.
- El agente recibe una recompensa $r_t = r(y_t)$.
- El objetivo del agente es maximizar el valor esperado de la recompensa: $v_{x_t}(\theta) = E_{q_\theta(.|x_t)}[r_t]$
- El algoritmo general es:

Ejemplo: Bernoulli

Algorithm 3 Greedy$(\mathcal{X}, p, q, r)$

for $t = 1, 2, \ldots$ **do**
  #estimate model:
  $\hat{\theta} \leftarrow \mathbb{E}_p[\theta]$

  #select and apply action:
  $x_t \leftarrow \text{argmax}_{x \in \mathcal{X}} \mathbb{E}_{q_{\hat{\theta}}}[r(y_t)|x_t = x]$
  Apply x_t and observe y_t

  #update distribution:
  $p \leftarrow \mathbb{P}_{p,q}(\theta \in \cdot | x_t, y_t)$
end for

Algorithm 4 Thompson$(\mathcal{X}, p, q, r)$

for $t = 1, 2, \ldots$ **do**
  #sample model:
  Sample $\hat{\theta} \sim p$

  #select and apply action:
  $x_t \leftarrow \text{argmax}_{x \in \mathcal{X}} \mathbb{E}_{q_{\hat{\theta}}}[r(y_t)|x_t = x]$
  Apply x_t and observe y_t

  #update distribution:
  $p \leftarrow \mathbb{P}_{p,q}(\theta \in \cdot | x_t, y_t)$
end for

Figura: TS General

Ejemplo: Bernoulli

Example

- $\Xi = \{1, 2, ...K\}$.
- $y_t = r_t$.
- $q_\theta(1 \mid k) = \theta_k, q_\theta(0 \mid k) = 1 - \theta_k$
- $p(\theta)$ es el producto de distribuciones Beta:

$$p(\theta) = \prod_{k=1}^{K} \frac{\Gamma(\alpha_k + \beta_k)}{\Gamma(\alpha_k)\Gamma(\beta_k)} \theta_k^{\alpha_k - 1}(1 - \theta_k)^{\beta_k - 1} \tag{1}$$

- En el algoritmo greedy: $E_p[\theta_k] = \alpha_k/(\alpha_k + \beta_k)$.
- En TS $\hat{\theta}_k$ se samplea de la distribucion Beta con parámetros (α_k, β_k).

Ejemplo: Caminos más cortos (tramos independientes)

Example

- Consideremos un grafo dirigido $G = (V, E)$, $V = \{1, ...N\}$. Una persona desea ir del punto 1 al N y los tiempos de desplazamiento son en **promedio** θ_e, donde e es un enlace entre nodos.

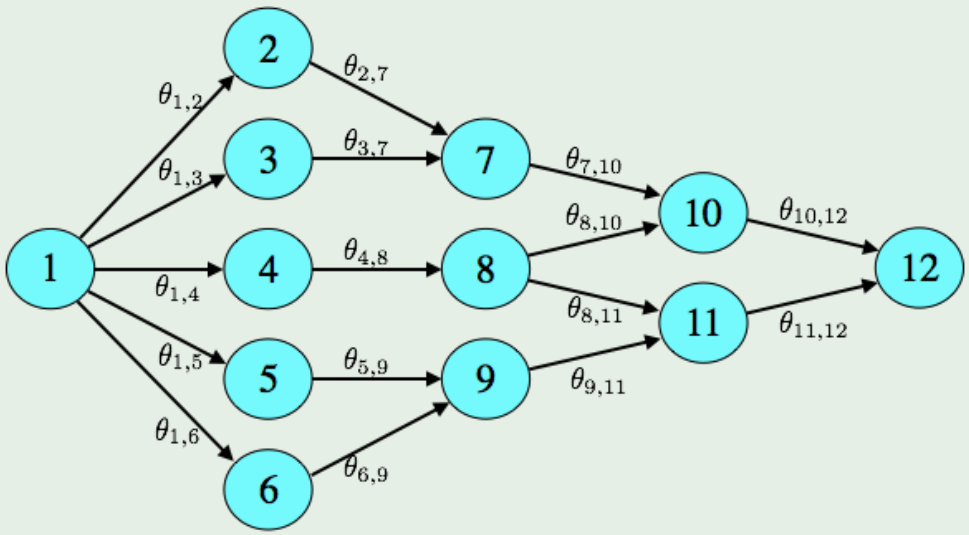


Figura: Camino más corto

- Las acciones son caminos en el grafo entre 1 y N. Un camino a_t es una sucesion de enlaces $a = (e_1, ..., e_k)$.

Ejemplo: Caminos más cortos (tramos independientes)

Example (continuación)

- Cada enlace $e_i \in a_t$ genera un tiempo de recorrido $y_{t,e}$, una relizacion independiente de una distribución log-Gaussian con parámetros, $\ln(\theta_e) - \tilde{\sigma}^2/2$ y $\tilde{\sigma}^2$, luego $E[y_{t,e}|\theta_e] = \theta_e$..
- El objetivo es minimizar el valor esperado del tiempo de recorrido: $\sum_{e \in a_t} y_{t,e}$
- Consideremos como prior de θ_e una distribución log-gaussiana con parámetros μ_e and σ_e^2. Es decir $\ln(\theta_e) \sim N(\mu_e, \sigma_e^2)$ es Gaussiana.
- Luego $E[\theta_e] = e^{\mu_e + \sigma_e^2/2}$.

Ejemplo: Caminos más cortos (tramos independientes)

Example (continuación)

- Como las distribuciones normales son conjugadas: θ_e condicional a $y_{t,e}$ es normal con parámetros:

$$(\mu_e, \sigma_e^2) \leftarrow \left(\frac{\frac{1}{\sigma_e^2}\mu_e + \frac{1}{\tilde{\sigma}^2}\left(\ln(y_{t,e}) + \frac{\tilde{\sigma}^2}{2}\right)}{\frac{1}{\sigma_e^2} + \frac{1}{\tilde{\sigma}^2}}, \frac{1}{\frac{1}{\sigma_e^2} + \frac{1}{\tilde{\sigma}^2}} \right). \quad (2)$$

Ejemplo: Caminos más cortos (tramos independientes)

Example (Continuación)

Los siguientes resultados son para el grafo con $186,000$ caminos entre la fuente (s) y el destino (d).

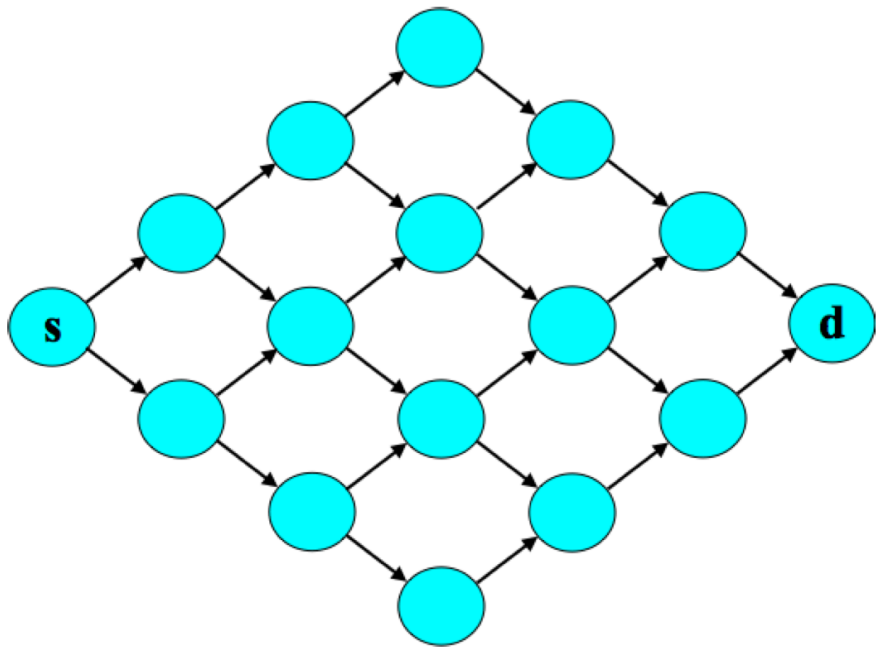
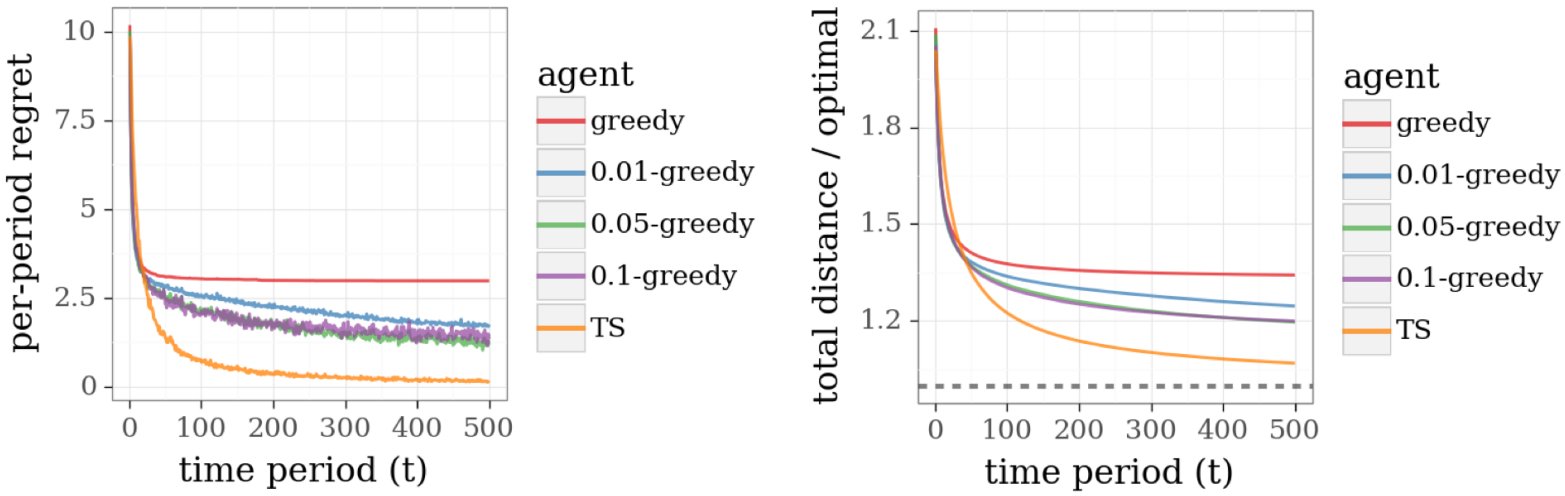


Figure 4.2: A binomial bridge with six stages.

Figura: Desempeño algoritmos

Ejemplo: Caminos más cortos (tramos independientes)

Example (Continuación)



(a) regret **(b)** cumulative travel time vs. optimal

Figure 4.1: Performance of Thompson sampling and ϵ-greedy algorithms in the shortest path problem.

Figura: Desempeño algoritmos

Contenido

Motivación

- La razon por la cuál los anteriores ejemplos han sido relativamente fáciles de simular es porque se han utilizado distibuciones conjugadas que se pueden simular eficientemente.
- En general en las aplicaciones no se tienen distribuciones conjugadas y es necesario a recurrir a métodos aprximados: Gibbs, Laplace, Langevin Monte Carlo, Bootstrapp, etc.