Μαθηματικές Ιδιότητες της Στοχαστικής Προσέγγισης

Καθολική προσέγγιση, στοχαστική προσέγγιση και multi-armed bandits

Νίκος Σταμάτης

22 Ιανουαρίου 2021

Περιεχόμενα

Καθολική Προσέγγιση

Ορισμός

Μια συνάρτηση $\sigma : \mathbb{R} \to \mathbb{R}$ καλείται *σιγμοειδής*, εαν

$$\sigma(t) \longrightarrow \begin{cases} 0, & \text{για } t \to -\infty, \\ 1, & \text{για } t \to +\infty. \end{cases}$$

Ορισμός

Μια συνάρτηση $\sigma : \mathbb{R} \to \mathbb{R}$ καλείται *σιγμοειδής*, εαν

$$\sigma(t) \longrightarrow \begin{cases} 0, & \text{για } t \to -\infty, \\ 1, & \text{για } t \to +\infty. \end{cases}$$

Παραδείγματα

- Σιγμοειδής: $\sigma(x) = \frac{1}{1+e^{-x}}$.
- Step function: $s(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$

Ορισμός

Μια συνάρτηση $\sigma : \mathbb{R} \to \mathbb{R}$ καλείται *σιγμοειδής*, εαν

$$\sigma(t) \longrightarrow \begin{cases} 0, & \text{για } t \to -\infty, \\ 1, & \text{για } t \to +\infty. \end{cases}$$

Ορισμός

Μια συνάρτηση $\sigma : \mathbb{R} \to \mathbb{R}$ καλείται *σιγμοειδής*, εαν

$$\sigma(t) \longrightarrow \begin{cases} 0, & \text{για } t \to -\infty, \\ 1, & \text{για } t \to +\infty. \end{cases}$$

Ορισμός

Μια συνάρτηση $A : X \to Y$ μεταξύ δύο διανυσματικών χώρων ονομάζεται *αφφινική*, εαν ισχύει ότι $A\left(\sum_{i=1}^{n} \lambda_i x_i\right) = \sum_{i=1}^{n} \lambda_i A(x_i)$ για κάθε $n \in \mathbb{N}$, $x_i \in X$ και κάθε $\lambda_i \in \mathbb{R}$ με $\sum_{i=1}^{n} \lambda_i = 1$.

Νευρωνικά Δίκτυα

Ορισμός

Μια συνάρτηση $\sigma : \mathbb{R} \to \mathbb{R}$ καλείται *σιγμοειδής*, εαν

$$\sigma(t) \longrightarrow \begin{cases} 0, & \text{για } t \to -\infty, \\ 1, & \text{για } t \to +\infty. \end{cases}$$

Ορισμός

Μια συνάρτηση $A : X \to Y$ μεταξύ δύο διανυσματικών χώρων ονομάζεται *αφφινική*, εαν ισχύει ότι $A\left(\sum_{i=1}^{n} \lambda_i x_i\right) = \sum_{i=1}^{n} \lambda_i A(x_i)$ για κάθε $n \in \mathbb{N}$, $x_i \in X$ και κάθε $\lambda_i \in \mathbb{R}$ με $\sum_{i=1}^{n} \lambda_i = 1$.

Νευρωνικό δίκτυο, ονομάζουμε κάθε συνάρτηση s της μορφής $s(x) = Tx$, όπου

$$T = A_{m+1} S_m A_m \cdots A_2 S_1 A_1$$

είναι ένας τελεστής ο οποίος ορίζεται απο τις διαδοχικές συνθέσεις αφφινικών μετασχηματισμών $A_i : \mathbb{R}^{d_{i-1}} \to \mathbb{R}^{d_i}$ με σιγμοειδείς συναρτήσεις S_i. Ο αριθμός m μετρά το πλήθος των στρωμάτων του δικτύου, ενώ οι αριθμοί d_i το πλήθος των κόμβων που εμφανίζονται σε κάθε στρώμα.

Καθολική Προσέγγιση

Ερώτημα

Για ποιες συναρτήσεις ενεργοποίησης σ ισχύει ότι το σύνολο

$$\Sigma_n(\sigma) = \left\{ f : I_n \to \mathbb{R} : f(x) = \sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j) : \right.$$
$$\left. N \in \mathbb{N}, a_j \in \mathbb{R}, w_j \in \mathbb{R}^n, \theta_j \in \mathbb{R} \right\}$$
$$= \text{span}\left\{ f : f(x) = \sigma\left(w^\intercal x + \theta\right) \quad \text{για } w \in \mathbb{R}^n, \theta \in \mathbb{R} \right\}$$

είναι πυκνό στον $C(I_n)$;

Καθολική Προσέγγιση

Θεώρημα (Cybenko, 1989)

Για κάθε συνεχή σιγμοειδή συνάρτηση ενεργοποίησης σ, το σύνολο $\Sigma_n(\sigma)$ είναι πυκνό στον $C(I_n)$.

$$\Sigma_n(\sigma) = \left\{ f : I_n \to \mathbb{R} : f(x) = \sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j) : \right.$$
$$\left. N \in \mathbb{N}, a_j \in \mathbb{R}, w_j \in \mathbb{R}^n, \theta_j \in \mathbb{R} \right\}$$
$$= \text{span} \left\{ f : f(x) = \sigma\left(w^\intercal x + \theta\right) \quad \text{για } w \in \mathbb{R}^n, \theta \in \mathbb{R} \right\}.$$

Καθολική Προσέγγιση

Θεώρημα (Cybenko, 1989)

Για κάθε συνεχή σιγμοειδή συνάρτηση ενεργοποίησης σ, το σύνολο $\Sigma_n(\sigma)$ είναι πυκνό στον $C(I_n)$.

Θεώρημα (Chen, Chen, Liu, 1991)

Κατασκευαστική απόδειξη του Θεωρήματος Cybenko.

$$\Sigma_n(\sigma) = \left\{ f : I_n \to \mathbb{R} : f(x) = \sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j) : \right.$$
$$\left. N \in \mathbb{N}, a_j \in \mathbb{R}, w_j \in \mathbb{R}^n, \theta_j \in \mathbb{R} \right\}$$
$$= \text{span} \left\{ f : f(x) = \sigma\left(w^\intercal x + \theta\right) \quad \text{για } w \in \mathbb{R}^n, \theta \in \mathbb{R} \right\}.$$

Καθολική Προσέγγιση

Θεώρημα (Cybenko, 1989)

Για κάθε συνεχή σιγμοειδή συνάρτηση ενεργοποίησης σ, το σύνολο $\Sigma_n(\sigma)$ είναι πυκνό στον $C(I_n)$.

Θεώρημα (Chen, Chen, Liu, 1991)

Κατασκευαστική απόδειξη του Θεωρήματος Cybenko.

Θεώρημα (Leshno, Lin, Pinkus, Schocken, 1993)

Το σύνολο $\Sigma_n(\sigma)$ είναι πυκνό στον $C(\mathbb{R}^n)$, εαν και μόνο εαν η συνάρτηση ενεργοποίησης σ δεν είναι πολυώνυμο.

$$\Sigma_n(\sigma) = \left\{ f : I_n \to \mathbb{R} : f(x) = \sum_{j=1}^{N} a_j \sigma(w_j^\intercal x + \theta_j) : \right.$$
$$\left. N \in \mathbb{N}, a_j \in \mathbb{R}, w_j \in \mathbb{R}^n, \theta_j \in \mathbb{R} \right\}$$
$$= \text{span} \left\{ f : f(x) = \sigma\left(w^\intercal x + \theta\right) \quad \text{για } w \in \mathbb{R}^n, \theta \in \mathbb{R} \right\}.$$

Το Θεώρημα Kolmogorov-Arnold

Θεώρημα (Maiorov, Pinkus, 1999)

Υπάρχει λεία, σιγμοειδής συνάρτηση ενεργοποίησης σ, τέτοια ώστε για κάθε $d \in \mathbb{N}$, κάθε συμπαγές $K \subseteq \mathbb{R}^d$, κάθε $f \in C(K)$ και $\varepsilon > 0$, να υπάρχουν πραγματικές σταθερές d_i, c_{ij}, θ_{ij}, γ_i και διανύσματα $w_{ij} \in \mathbb{R}^d$, με

$$\left| f(x) - \sum_{i=1}^{6d+3} d_i \sigma \left(\sum_{j=1}^{3d} c_{ij} \sigma \left(w_{ij}^{\mathsf{T}} x + \theta_{ij} \right) + \gamma_i \right) \right| < \varepsilon$$

για κάθε $x \in K$.

Το Θεώρημα Kolmogorov-Arnold

Θεώρημα (Maiorov, Pinkus, 1999)

Υπάρχει λεία, σιγμοειδής συνάρτηση ενεργοποίησης σ, τέτοια ώστε για κάθε $d \in \mathbb{N}$, κάθε συμπαγές $K \subseteq \mathbb{R}^d$, κάθε $f \in C(K)$ και $\varepsilon > 0$, να υπάρχουν πραγματικές σταθερές d_i, c_{ij}, θ_{ij}, γ_i και διανύσματα $w_{ij} \in \mathbb{R}^d$, με

$$\left| f(x) - \sum_{i=1}^{6d+3} d_i \sigma \left(\sum_{j=1}^{3d} c_{ij} \sigma \left(w_{ij}^{\mathsf{T}} x + \theta_{ij} \right) + \gamma_i \right) \right| < \varepsilon$$

για κάθε $x \in K$.

Θεώρημα (Kolmogorov-Arnold, 1957)

Υπάρχουν σταθερές $\lambda_1, \ldots, \lambda_d$ τέτοιες ώστε $\sum_{j=1}^{d} \lambda_j \leq 1$, και συνεχείς συναρτήσεις $\phi_1, \ldots, \phi_{2d+1}$ από το $[0, 1]$ στον εαυτό του, με την ιδιότητα ότι κάθε $f \in C[0, 1]^d$ μπορεί να γραφτεί ως

$$f(x_1, \ldots, x_d) = \sum_{i=1}^{2d+1} g \left(\sum_{j=1}^{d} \lambda_j \phi_i(x_j) \right),$$

όπου $g \in C[0, 1]$ μια συνάρτηση που εξαρτάται από την f.

Στοχαστική Προσέγγιση

Αλγόριθμος Robbins-Monro

Πλαίσιο:

- Μεταβλητή απόκρισης Y, επεξηγηματική μεταβλητή X.

Αλγόριθμος Robbins-Monro

Πλαίσιο:

- Μεταβλητή απόκρισης Y, επεξηγηματική μεταβλητή X.
- Δοθέντος ότι $X = x$, η κατανομή της $Y(x)$ είναι $P[Y(x) \leq y] = H(y \mid x)$.

Αλγόριθμος Robbins-Monro

Πλαίσιο:

- Μεταβλητή απόκρισης Y, επεξηγηματική μεταβλητή X.
- Δοθέντος ότι $X = x$, η κατανομή της $Y(x)$ είναι $P[Y(x) \leq y] = H(y \mid x)$.
- Έστω $M(x)$ η μέση τιμή της Y δοθέντος του x, δηλ. $M(x) = \mathbb{E}[Y \mid X = x]$.

Αλγόριθμος Robbins-Monro

Πλαίσιο:

- Μεταβλητή απόκρισης Y, επεξηγηματική μεταβλητή X.
- Δοθέντος ότι $X = x$, η κατανομή της $Y(x)$ είναι $P[Y(x) \leq y] = H(y \mid x)$.
- Έστω $M(x)$ η μέση τιμή της Y δοθέντος του x, δηλ. $M(x) = \mathbb{E}[Y \mid X = x]$.
  πχ. στο γραμμικό μοντέλο $Y = Xb + \varepsilon$, έχουμε $E[Y \mid X = x] = x'b$.

Αλγόριθμος Robbins-Monro

Πλαίσιο:

- Μεταβλητή απόκρισης Y, επεξηγηματική μεταβλητή X.
- Δοθέντος ότι $X = x$, η κατανομή της $Y(x)$ είναι $P[Y(x) \leq y] = H(y \mid x)$.
- Έστω $M(x)$ η μέση τιμή της Y δοθέντος του x, δηλ. $M(x) = \mathbb{E}[Y \mid X = x]$.
  πχ. στο γραμμικό μοντέλο $Y = Xb + \varepsilon$, έχουμε $E[Y \mid X = x] = x'b$.
- Να βρεθεί θ με $M(\theta) = a$.

Αλγόριθμος Robbins-Monro

Πλαίσιο:

- Μεταβλητή απόκρισης Y, επεξηγηματική μεταβλητή X.
- Δοθέντος ότι $X = x$, η κατανομή της $Y(x)$ είναι $P[Y(x) \leq y] = H(y \mid x)$.
- Έστω $M(x)$ η μέση τιμή της Y δοθέντος του x, δηλ. $M(x) = \mathbb{E}[Y \mid X = x]$.
  πχ. στο γραμμικό μοντέλο $Y = Xb + \varepsilon$, έχουμε $E[Y \mid X = x] = x'b$.
- Να βρεθεί θ με $M(\theta) = a$.

Εργαλεία:

Αλγόριθμος Robbins-Monro

Πλαίσιο:

- Μεταβλητή απόκρισης Y, επεξηγηματική μεταβλητή X.
- Δοθέντος ότι $X = x$, η κατανομή της $Y(x)$ είναι $P[Y(x) \leq y] = H(y \mid x)$.
- Έστω $M(x)$ η μέση τιμή της Y δοθέντος του x, δηλ. $M(x) = \mathbb{E}[Y \mid X = x]$.
  πχ. στο γραμμικό μοντέλο $Y = Xb + \varepsilon$, έχουμε $E[Y \mid X = x] = x'b$.
- Να βρεθεί θ με $M(\theta) = a$.

Εργαλεία:

- Μπορούμε να προσομοιώνουμε από την $H(y \mid x)$ για κάθε τιμή του x.

Αλγόριθμος Robbins-Monro

Πλαίσιο:

- Μεταβλητή απόκρισης Y, επεξηγηματική μεταβλητή X.
- Δοθέντος ότι $X = x$, η κατανομή της $Y(x)$ είναι $P[Y(x) \leq y] = H(y \mid x)$.
- Έστω $M(x)$ η μέση τιμή της Y δοθέντος του x, δηλ. $M(x) = \mathbb{E}[Y \mid X = x]$.
  πχ. στο γραμμικό μοντέλο $Y = Xb + \varepsilon$, έχουμε $E[Y \mid X = x] = x'b$.
- Να βρεθεί θ με $M(\theta) = a$.

Εργαλεία:

- Μπορούμε να προσομοιώνουμε από την $H(y \mid x)$ για κάθε τιμή του x.

Αλγόριθμος Robbins-Monro

Πλαίσιο:

- Μεταβλητή απόκρισης Y, επεξηγηματική μεταβλητή X.
- Δοθέντος ότι $X = x$, η κατανομή της $Y(x)$ είναι $P[Y(x) \leq y] = H(y \mid x)$.
- Έστω $M(x)$ η μέση τιμή της Y δοθέντος του x, δηλ. $M(x) = \mathbb{E}[Y \mid X = x]$.
  πχ. στο γραμμικό μοντέλο $Y = Xb + \varepsilon$, έχουμε $E[Y \mid X = x] = x'b$.
- Να βρεθεί θ με $M(\theta) = a$.

Εργαλεία:

- Μπορούμε να προσομοιώνουμε από την $H(y \mid x)$ για κάθε τιμή του x.

Λύση:

Αλγόριθμος Robbins-Monro

Πλαίσιο:

- Μεταβλητή απόκρισης Y, επεξηγηματική μεταβλητή X.
- Δοθέντος ότι $X = x$, η κατανομή της $Y(x)$ είναι $P[Y(x) \leq y] = H(y \mid x)$.
- Έστω $M(x)$ η μέση τιμή της Y δοθέντος του x, δηλ. $M(x) = \mathbb{E}[Y \mid X = x]$.
  πχ. στο γραμμικό μοντέλο $Y = Xb + \varepsilon$, έχουμε $E[Y \mid X = x] = x'b$.
- Να βρεθεί θ με $M(\theta) = a$.

Εργαλεία:

- Μπορούμε να προσομοιώνουμε από την $H(y \mid x)$ για κάθε τιμή του x.

Λύση:

- Μέσω της αναδρομικής ακολουθίας

$$x_{n+1} = x_n + a_n(a - y_n),$$

  όπου η παρατήρηση y_n έχει προσομοιωθεί από την $H(y \mid x_n)$.

Αλγόριθμος Robbins-Monro

Θεώρημα (Robbins-Monro 1951)

Αν υπάρχει $C > 0$ τ.ω. $x \in \mathbb{R}$, $P[|Y(x)| \leq C] = 1$, αν η $M(x) = \mathbb{E}[Y \mid X = x]$ είναι αύξουσα με $M(\theta) = a$ και $M'(\theta) > 0$, τότε για κάθε $(a_n)_n \in \ell_2^+ \setminus \ell_1$ η $(x_n)_n$ που κατασκευάζεται από τον αλγόριθμο Robbins-Monro συγκλίνει κατά πιθανότητα στο θ.

Αλγόριθμος Robbins-Monro

Θεώρημα (Robbins-Monro 1951)

Αν υπάρχει $C > 0$ τ.ω. $x \in \mathbb{R}$, $P[|Y(x)| \leq C] = 1$, αν η $M(x) = \mathbb{E}[Y \mid X = x]$ είναι αύξουσα με $M(\theta) = a$ και $M'(\theta) > 0$, τότε για κάθε $(a_n)_n \in \ell_2^+ \setminus \ell_1$ η $(x_n)_n$ που κατασκευάζεται από τον αλγόριθμο Robbins-Monro συγκλίνει κατά πιθανότητα στο θ.

Αλγοριθμος Robbins-Monro	
Βημα 1	Επιλέγουμε $(a_n)_n \in \ell_2^+ \setminus \ell_1$ και $x_1 \in \mathbb{R}$.
Βημα 2	Υποθέτουμε ότι έχουμε κατασκευάσει το $x_1, \ldots, x_{n-1}$. Προσομοιώνουμε μια παρατήρηση y_n από την κατανομή $H(y \mid x_n)$ και θέτουμε $x_{n+1} = x_n + a_n(a - y_n)$.
Βημα 3	Επιστρέφουμε στο Βήμα 2.

Σταθερά σημεία

Ορισμός

Έστω X σύνολο και $H : X \to X$ συνάρτηση. Ένα σημείο $x_0 \in X$ καλείται *σταθερό σημείο* της H, εαν $H(x_0) = x_0$.

Σταθερά σημεία

Ορισμός

Έστω X σύνολο και $H : X \to X$ συνάρτηση. Ένα σημείο $x_0 \in X$ καλείται *σταθερό σημείο* της H, εαν $H(x_0) = x_0$.

Ορισμός

Μια συνάρτηση $H : (X, \rho) \to (X, \rho)$ ορισμένη σε έναν μετρικό χώρο (X, ρ), καλείται β-*συστολή*, όπου $\beta > 0$ σταθερά, εαν

$$\rho(Hx, Hy) \leq \beta\rho(x, y) \text{ για κάθε } x, y \in X.$$

Σταθερά σημεία

Ορισμός

Έστω X σύνολο και $H : X \to X$ συνάρτηση. Ένα σημείο $x_0 \in X$ καλείται *σταθερό σημείο* της H, εαν $H(x_0) = x_0$.

Ορισμός

Μια συνάρτηση $H : (X, \rho) \to (X, \rho)$ ορισμένη σε έναν μετρικό χώρο (X, ρ), καλείται β-*συστολή*, όπου $\beta > 0$ σταθερά, εαν

$$\rho(Hx, Hy) \leq \beta\rho(x, y) \text{ για κάθε } x, y \in X.$$

Θεώρημα σταθερού σημείου του Banach (1922)

Έστω (X, ρ) πλήρης μετρικός χώρος και $H : X \to X$ συνάρτηση η οποία είναι β-συστολή για κάποιο $\beta \in (0, 1)$. Τότε η H έχει μοναδικό σταθερό σημείο x_0. Επιπλέον, η ακολουθία $(x_n)_n$ που ορίζεται αναδρομικά ως $x_1 = y$ και $x_{n+1} = H(x_n)$ για κάθε $n \in \mathbb{N}$, συγκλίνει στο x_0 για οποιαδήποτε αρχική επιλογή $y \in X$.

Σταθερά σημεία υπό την παρουσία θορύβου

Ερώτημα

Έστω X σύνολο και $H : X \to X$ συνάρτηση με μοναδικό σταθερό σημείο το $x_0 \in X$. Πώς μπορούμε να προσδιορίσουμε το σταθερό της σημείο, όταν δε γνωρίζουμε τις ακριβείς τιμές της H;

Σταθερά σημεία υπό την παρουσία θορύβου

Ερώτημα

Έστω X σύνολο και $H : X \to X$ συνάρτηση με μοναδικό σταθερό σημείο το $x_0 \in X$. Πώς μπορούμε να προσδιορίσουμε το σταθερό της σημείο, όταν δε γνωρίζουμε τις ακριβείς τιμές της H;

Αλγόριθμος

Κατασκευάζουμε αναδρομικά την ακολουθία:

$$x_{n+1} = (1 - \gamma_n)x_n + \gamma_n(Hx_n + w_n),$$

όπου $(\gamma_n)_n$ είναι κατάλληλη ακολουθία στο $(0, 1]$, και οι όροι w_n αντιστοιχούν στο "θόρυβο" των τιμών Hx_n.

Δύο κλάσεις τελεστών

Ψευδοσυστολές

Έστω $(X, \|\cdot\|)$ χώρος με νόρμα. Μια συνάρτηση $H : X \to X$ καλείται *ψευδοσυστολή*, εαν υπάρχουν $x^* \in X$ και $\beta \in [0, 1)$ τέτοια ώστε

$$\|Hx - Hx^*\| \leq \beta \|x - x^*\|$$

για κάθε $x \in X$.

Δύο κλάσεις τελεστών

Ψευδοσυστολές

Έστω $(X, \|\cdot\|)$ χώρος με νόρμα. Μια συνάρτηση $H : X \to X$ καλείται *ψευδοσυστολή*, εαν υπάρχουν $x^* \in X$ και $\beta \in [0, 1)$ τέτοια ώστε

$$\|Hx - Hx^*\| \leq \beta \|x - x^*\|$$

για κάθε $x \in X$.

Μονότονοι τελεστές

Έστω $(X, \|\cdot\|, \leq)$ χώρος με νόρμα εφοδιασμένος με μια γραμμική διάταξη. Μια συνάρτηση $H : X \to X$ καλείται *μονότονη*, εαν $Hx \leq Hy$ για κάθε $x, y \in X$ με $x \leq y$.

Στοχαστική προσέγγιση για ψευδοσυστολές

Πρόταση

Έστω $(r_n)_n$ η ακολουθία που ορίζεται από την αναδρομική σχέση

$$r_{n+1} = (1-\gamma_n)r_n + \gamma_n(H_n r_n + w_n + u_n),$$

όπου

1. η ακολουθία $(\gamma_n)_n$ είναι τέτοια ώστε $\sum_{n=1}^{\infty} \gamma_n(i) = \infty$ και $\sum_{n=1}^{\infty} \gamma_n(i)^2 < \infty$ για κάθε $i = 1, \ldots, N$.
2. Η ακολουθία $(w_n)_n$ έχει την ιδιότητα ότι

$$\mathbb{E}\left[w_n(i) \mid \mathcal{F}_n\right] = 0 \quad \text{και} \quad \mathbb{E}\left[w_n(i)^2 \mid \mathcal{F}_n\right] \leq A + B\|r_n\|^2.$$

3. Κάθε H_n είναι ψευδοσυστολή ως προς την ίδια νόρμα $\|\cdot\|_\xi$, με το ίδιο σταθερό σημείο r^* και την ίδια σταθερά $\beta \in [0,1)$.
4. Υπάρχει ακολουθία μη-αρνητικών τυχαίων μεταβλητών $(\theta_n)_n$ η οποία συγκλίνει στο μηδέν σχεδόν παντού, τέτοια ώστε

$$\|u_n\|_\infty \leq \theta_n \left(1 + \|r_n\|_\xi\right)$$

για κάθε $n \in \mathbb{N}$.

Τότε η $(r_n)_n$ συγκλίνει στο r^* σχεδόν παντού.

Στοχαστική προσέγγιση για μονότονους τελεστές

Πρόταση

Έστω $(r_n)_n$ η ακολουθία που ορίζεται από την αναδρομική σχέση

$$r_{n+1} = (1-\gamma_n)r_n + \gamma_n(Hr_n + w_n),$$

όπου

1. η ακολουθία $(\gamma_n)_n$ είναι τέτοια ώστε $\sum_{n=1}^{\infty} \gamma_n(i) = \infty$ και $\sum_{n=1}^{\infty} \gamma_n(i)^2 < \infty$ για κάθε $i = 1, \ldots, N$.
2. Η ακολουθία $(w_n)_n$ έχει την ιδιότητα ότι

$$\mathbb{E}\left[w_n(i) \mid \mathcal{F}_n\right] = 0 \quad \text{και} \quad \mathbb{E}\left[w_n(i)^2 \mid \mathcal{F}_n\right] \leq A + B\|r_n\|^2.$$

3. Για τον τελεστή H ισχύει ότι
   - 3.1 είναι μονότονος, δηλαδή $Hx \leq Hy$ για κάθε $x \leq y$.
   - 3.2 Για κάθε $\lambda > 0$ και $r \in \mathbb{R}^N$, ισχύει ότι: $Hr - \lambda e \leq H(r - \lambda e) \leq H(r + \lambda e) \leq Hr + \lambda e$, όπου $e = (1, \ldots, 1)$.
   - 3.3 Έχει μοναδικό σταθερό σημείο, $Hr^* = r^*$.

Εαν η $(r_n)_n$ είναι φραγμένη σχεδόν παντού, τότε συγκλίνει στο r^* σχεδόν παντού.

Δυναμικός προγραμματισμός

- MDP με σύνολο καταστάσεων S.

Δυναμικός προγραμματισμός

- MDP με σύνολο καταστάσεων S.
- Σε κάθε στάδιο n του προβλήματος, παρατηρούμε την τρέχουσα κατάσταση $i \in S$ της διαδικασίας και επιλέγουμε μια απόφαση $a \in A(i)$.

Δυναμικός προγραμματισμός

- MDP με σύνολο καταστάσεων S.
- Σε κάθε στάδιο n του προβλήματος, παρατηρούμε την τρέχουσα κατάσταση $i \in S$ της διαδικασίας και επιλέγουμε μια απόφαση $a \in A(i)$.
- Η διαδικασία μεταβαίνει στην κατάσταση j με πιθανότητα $p_{ij}(a)$. Οι πιθανότητες μετάβασης εξαρτώνται από την τρέχουσα κατάσταση i και την απόφαση a που λήφθηκε.

Δυναμικός προγραμματισμός

- MDP με σύνολο καταστάσεων S.
- Σε κάθε στάδιο n του προβλήματος, παρατηρούμε την τρέχουσα κατάσταση $i \in S$ της διαδικασίας και επιλέγουμε μια απόφαση $a \in A(i)$.
- Η διαδικασία μεταβαίνει στην κατάσταση j με πιθανότητα $p_{ij}(a)$. Οι πιθανότητες μετάβασης εξαρτώνται από την τρέχουσα κατάσταση i και την απόφαση a που λήφθηκε.
- Κάθε μετάβαση επιφέρει κόστος $c(i, a, j)$.

Δυναμικός προγραμματισμός

- MDP με σύνολο καταστάσεων S.
- Σε κάθε στάδιο n του προβλήματος, παρατηρούμε την τρέχουσα κατάσταση $i \in S$ της διαδικασίας και επιλέγουμε μια απόφαση $a \in A(i)$.
- Η διαδικασία μεταβαίνει στην κατάσταση j με πιθανότητα $p_{ij}(a)$. Οι πιθανότητες μετάβασης εξαρτώνται από την τρέχουσα κατάσταση i και την απόφαση a που λήφθηκε.
- Κάθε μετάβαση επιφέρει κόστος $c(i, a, j)$.
- *Αξία πολιτικής* π: Το αναμενόμενο κόστος $J^\pi(i)$ όταν η διαδικασία ξεκινάει από την κατάσταση $i_0 = i$, και ακολουθείται η πολιτική $\pi = (\mu_0, \mu_1, \ldots)$,

$$J^\pi(i) = \lim_{N\to\infty} \mathbb{E}\left[\sum_{m=0}^{N} c\left(i_m, \mu_m(i_m), i_{m+1}\right) \mid i_0 = i\right].$$

Δυναμικός προγραμματισμός

- MDP με σύνολο καταστάσεων S.
- Σε κάθε στάδιο n του προβλήματος, παρατηρούμε την τρέχουσα κατάσταση $i \in S$ της διαδικασίας και επιλέγουμε μια απόφαση $a \in A(i)$.
- Η διαδικασία μεταβαίνει στην κατάσταση j με πιθανότητα $p_{ij}(a)$. Οι πιθανότητες μετάβασης εξαρτώνται από την τρέχουσα κατάσταση i και την απόφαση a που λήφθηκε.
- Κάθε μετάβαση επιφέρει κόστος $c(i, a, j)$.
- *Αξία πολιτικής* π: Το αναμενόμενο κόστος $J^\pi(i)$ όταν η διαδικασία ξεκινάει από την κατάσταση $i_0 = i$, και ακολουθείται η πολιτική $\pi = (\mu_0, \mu_1, \ldots)$,

$$J^\pi(i) = \lim_{N\to\infty} \mathbb{E}\left[\sum_{m=0}^{N} c\left(i_m, \mu_m(i_m), i_{m+1}\right) \mid i_0 = i\right].$$

- *Συνάρτηση βέλτιστης τιμής* $J^*(i)$*:* Το ελάχιστο δυνατό κόστος κάτω από οποιαδήποτε πολιτική, όταν η διαδικασία ξεκινάει από την κατάσταση i,

$$J^*(i) = \min_\pi J^\pi(i).$$

Η εξίσωση Bellman

Εξίσωση Bellman

$$J^*(i) = \min_{a \in A(i)} \left\{ \sum_{j \in S} p_{ij}(a) \left(c(i,a,j) + J^*(j) \right) \right\}.$$

Εξίσωση Bellman

$$J^*(i) = \min_{a \in A(i)} \left\{ \sum_{j \in S} p_{ij}(a) \left(c(i,a,j) + J^*(j) \right) \right\}.$$

Μέθοδος διαδοχικών προσεγγίσεων

Ξεκινάμε από κάποια αυθαίρετη συνάρτηση J_0, και εν συνεχεία ορίζουμε αναδρομικά

$$J_{n+1}(i) = \min_{a \in A(i)} \left\{ \sum_{j \in S} p_{ij}(a) \left(c(i,a,j) + J_n(j) \right) \right\}.$$

Εξίσωση Bellman

$$J^*(i) = \min_{a \in A(i)} \left\{ \sum_{j \in S} p_{ij}(a) \left(c(i,a,j) + J^*(j) \right) \right\}.$$

Μέθοδος διαδοχικών προσεγγίσεων

Ξεκινάμε από κάποια αυθαίρετη συνάρτηση J_0, και εν συνεχεία ορίζουμε αναδρομικά

$$J_{n+1}(i) = \min_{a \in A(i)} \left\{ \sum_{j \in S} p_{ij}(a) \left(c(i,a,j) + J_n(j) \right) \right\}.$$

Η σύγκλιση της ακολουθίας $(J_n)_n$ στην J^* εξασφαλίζεται μέσω του θεωρήματος σταθερού σημείου του Banach, καθώς ο τελεστής $T : C(S) \to C(S)$, που ορίζεται ώς

$$(Tf)(i) = \min_{a \in A(i)} \left\{ \sum_{j \in S} p_{ij}(a) \left(c(i,a,j) + f(j) \right) \right\}, \; f \in C(S), \; i \in S,$$

αποτελεί συστολή.

Q-Μάθηση

Ερώτημα

Μπορούμε να λύσουμε την εξίσωση Bellman, όταν δε γνωρίζουμε τις πιθανότητες μετάβασης $p_{ij}(a)$ και τα κόστη $c(i, a, j)$;

Q-Μάθηση

Ερώτημα

Μπορούμε να λύσουμε την εξίσωση Bellman, όταν δε γνωρίζουμε τις πιθανότητες μετάβασης $p_{ij}(a)$ και τα κόστη $c(i, a, j)$;

Q-Παράγοντες

Για $(i, a) \in S \times A(i)$, ορίζουμε τον βέλτιστο Q-παράγοντα $Q^*(i, a)$ ως

$$Q^*(0, a) = 0 \text{ and } Q^*(i, a) = \sum_{j=0}^{N} p_{ij}(a) \left(c(i, a, j) + J^*(j)\right) \text{ για } i = 1, \ldots, N,$$

όπου J^* η συνάρτηση βέλτιστης τιμής.

Q-Μάθηση

Ερώτημα

Μπορούμε να λύσουμε την εξίσωση Bellman, όταν δε γνωρίζουμε τις πιθανότητες μετάβασης $p_{ij}(a)$ και τα κόστη $c(i,a,j)$;

Q-Παράγοντες

Για $(i,a) \in S \times A(i)$, ορίζουμε τον βέλτιστο Q-παράγοντα $Q^*(i,a)$ ως

$$Q^*(0,a) = 0 \text{ and } Q^*(i,a) = \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + J^*(j)\right) \text{ για } i = 1,\dots,N,$$

όπου J^* η συνάρτηση βέλτιστης τιμής.

Ο Q^* ικανοποιεί την εξίσωση

$$Q^*(i,a) = \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b \in A(j)} Q^*(j,b)\right) \text{ for } i = 1,\dots,N,$$

Q-Μάθηση

- Εξίσωση Bellman για την J^*:

$$J^*(i) = \min_{a \in A(i)} \left\{ \sum_{j \in S} p_{ij}(a) \left(c(i,a,j) + J^*(j) \right) \right\}.$$

- Εξίσωση Bellman για τους Q-παράγοντες:

$$Q^*(i,a) = \sum_{j=0}^{N} p_{ij}(a) \left(c(i,a,j) + \min_{b \in A(j)} Q^*(j,b) \right).$$

Q-Μάθηση

Ο Αλγοριθμος Q-Learning	
Βημα 1	Επιλέγουμε συναρτήσεις $(\gamma_n)_n$ ορισμένες στο σύνολο $\tilde{S} = \{(i,a) : i = 1,\dots,N, a \in A(i)\}$, τ.ω. $\sum_{n=0}^{\infty} \gamma_n(i,a) = \infty$ και $\sum_{n=0}^{\infty} \gamma_n(i,a)^2 < \infty$ για κάθε $i = 1,\dots,N$ και $a \in A(i)$.
Βημα 2	Αρχικοποιούμε με τη συνάρτηση $Q_0(i,a)$ για $i = 1,\dots,N$ και $a \in A(i)$.
Βημα 3	Υποθέτουμε πως έχουμε δημιουργήσει τις τιμές $Q_n(i,a)$ για κάθε ζεύγος (i,a) για κάποιο $n \geq 0$. Για κάθε ζεύγος (i,a), προσομοιώνουμε μια παρατήρηση $j_{i,a}$ από την κατανομή $p_{i,\cdot}(a)$ και θέτουμε $Q_{n+1}(i,a) = (1-\gamma_n(i,a))Q_n(i,a) + \gamma_n(i,a)\left(c(i,a,j_{i,a}) + \min_{b\in A(j_{i,a})} Q_n(j_{i,a},b)\right)$.
Βημα 4	Επιστρέφουμε στο Βημα 3.

Q-Μάθηση

Θεώρημα (Watkins 1989, Tsitsiklis 1994)

Θεωρούμε την ακολουθία $(Q_n)_n$ που ορίζεται αναδρομικά ως

$$Q_{n+1}(i,a) = (1-\gamma_n(i,a))\, Q_n(i,a) + \gamma_n(i,a)\left(c(i,a,j) + \min_{b\in A(j)} Q_n(j,b)\right),$$

όπου η κατάσταση j έχει προσομοιωθεί από την κατανομή $p_{i,\cdot}(a)$ και $(\gamma_n)_n$ είναι ακολουθία με την ιδιότητα ότι $\sum_{n=0}^{\infty} \gamma_n(i,a) = \infty$ και $\sum_{n=0}^{\infty} \gamma_n(i,a)^2 < \infty$ για κάθε $i = 1,\ldots,N$ και $a \in A(i)$. Αν όλες οι πολιτικές είναι γνήσιες, τότε $Q_n(i,a) \to Q^*(i,a)$ για κάθε i, $a \in A(i)$ σχεδόν παντού.

Q-Μάθηση

Θεώρημα (Watkins 1989, Tsitsiklis 1994)

Θεωρούμε την ακολουθία $(Q_n)_n$ που ορίζεται αναδρομικά ως

$$Q_{n+1}(i,a) = (1-\gamma_n(i,a))\,Q_n(i,a) + \gamma_n(i,a)\left(c(i,a,j) + \min_{b\in A(j)} Q_n(j,b)\right),$$

όπου η κατάσταση j έχει προσομοιωθεί από την κατανομή $p_{i,\cdot}(a)$ και $(\gamma_n)_n$ είναι ακολουθία με την ιδιότητα ότι $\sum_{n=0}^{\infty}\gamma_n(i,a) = \infty$ και $\sum_{n=0}^{\infty}\gamma_n(i,a)^2 < \infty$ για κάθε $i = 1,\ldots,N$ και $a \in A(i)$. Αν όλες οι πολιτικές είναι γνήσιες, τότε $Q_n(i,a) \to Q^*(i,a)$ για κάθε i, $a \in A(i)$ σχεδόν παντού.

Ο τελεστής $H : C(\tilde{S}) \to C(\tilde{S})$ που ορίζεται ως

$$(HQ)(i,a) = \sum_{j=0}^{N} p_{ij}(a)\left(c(i,a,j) + \min_{b\in A(j)} Q(j,b)\right).$$

για Q στο $C(\tilde{S})$, προκύπτει είτε ότι είναι συστολή ως προς κάποια κατάλληλη νόρμα, είτε ότι είναι μονότονος.

Multi-armed Bandits

Το πλαίσιο των multi-armed bandit

- Δύο διαφορετικοί πληθυσμοί με κατανομές κερδών p_A, p_B και αναμενόμενες τιμές $\mu_A < \mu_B$ αντίστοιχα.

Το πλαίσιο των multi-armed bandit

- Δύο διαφορετικοί πληθυσμοί με κατανομές κερδών p_A, p_B και αναμενόμενες τιμές $\mu_A < \mu_B$ αντίστοιχα.
- Σε κάθε γύρο επιλέγουμε έναν πληθυσμό και τραβάμε μια τιμή από αυτόν, εισπράττοντάς την ως κέρδος.

Το πλαίσιο των multi-armed bandit

- Δύο διαφορετικοί πληθυσμοί με κατανομές κερδών p_A, p_B και αναμενόμενες τιμές $\mu_A < \mu_B$ αντίστοιχα.
- Σε κάθε γύρο επιλέγουμε έναν πληθυσμό και τραβάμε μια τιμή από αυτόν, εισπράττοντάς την ως κέρδος.
- *Ερώτημα:* Μπορούμε να βρούμε μια στρατηγική που μακροπρόθεσμα θα μας δίνει αναμενόμενο κέρδος όσο πιο κοντά γίνεται στο μ_B;

Ο Κανόνας του Robbins

Θεώρημα Robbins (1952)

Ασυμπτωτικά γίνεται να επιτύγχουμε αναμενόμενο κέρδος ακριβώς ίσο με μ_B, σχεδόν βεβαίως.

Ο Κανόνας του Robbins

Θεώρημα Robbins (1952)

Ασυμπτωτικά γίνεται να επιτύγχουμε αναμενόμενο κέρδος ακριβώς ίσο με μ_B, σχεδόν βεβαίως.

Ο Κανονας του Robbins	
Βημα 1	Επιλέγουμε $J_A, J_B \subseteq \mathbb{N}$ άπειρα με μηδενική πυκνότητα.
Βημα 2	Υποθέτουμε ότι έχουμε τραβήξει δείγμα $x_1, \ldots, x_{n-1}$ στους πρώτους $n-1$ γύρους. Αν $n \in J_A$, τραβάμε την X_n από τον πληθυσμό A. Αν $n \in J_B$, τραβάμε την X_n από τον B.
Βημα 3	Αν $n \notin J_A \cup J_B$, θέτουμε $a_n = \frac{\sum_{\{i:X_i \sim F_A\}} x_i}{\#\{i:X_i \sim F_A\}}$ και $b_n = \frac{\sum_{\{i:X_i \sim F_B\}} x_i}{\#\{i:X_i \sim F_B\}}$. Αν $a_n \geq b_n$, τραβάμε την X_n από τον πληθυσμό A, διαφορετικά από τον B.

Ο Κανόνας του Robbins

Θεώρημα Robbins (1952)

Ασυμπτωτικά γίνεται να επιτύγχουμε αναμενόμενο κέρδος ακριβώς ίσο με μ_B, σχεδόν βεβαίως.

Ο Κανονας του Robbins	
Βημα 1	Επιλέγουμε $J_A, J_B \subseteq \mathbb{N}$ άπειρα με μηδενική πυκνότητα.
Βημα 2	Υποθέτουμε ότι έχουμε τραβήξει δείγμα $x_1, \ldots, x_{n-1}$ στους πρώτους $n-1$ γύρους. Αν $n \in J_A$, τραβάμε την X_n από τον πληθυσμό A. Αν $n \in J_B$, τραβάμε την X_n από τον B.
Βημα 3	Αν $n \notin J_A \cup J_B$, θέτουμε $a_n = \frac{\sum_{\{i:X_i \sim F_A\}} x_i}{\#\{i:X_i \sim F_A\}}$ και $b_n = \frac{\sum_{\{i:X_i \sim F_B\}} x_i}{\#\{i:X_i \sim F_B\}}$. Αν $a_n \geq b_n$, τραβάμε την X_n από τον πληθυσμό A, διαφορετικά από τον B.

Μηδενική πυκνότητα:

$$d(J_A) = \lim_{n \to \infty} \frac{\#J_A \cap \{1, \ldots, n\}}{n} = 0.$$

Φράγματα Lai-Robbins

- Έχουμε K το πλήθος bandits με κατανομές $f(x, \theta_1), \ldots, f(x, \theta_K)$ και μέσες τιμές $\mu_1, \ldots, \mu_K$. Έστω $\mu^* = \max\{\mu_i\}_i$.

Φράγματα Lai-Robbins

- Έχουμε K το πλήθος bandits με κατανομές $f(x, \theta_1), \ldots, f(x, \theta_K)$ και μέσες τιμές $\mu_1, \ldots, \mu_K$. Έστω $\mu^* = \max\{\mu_i\}_i$.
- Συνολικό κέρδος S_n μετά απο n γύρους.

Φράγματα Lai-Robbins

- Έχουμε K το πλήθος bandits με κατανομές $f(x, \theta_1), \ldots, f(x, \theta_K)$ και μέσες τιμές $\mu_1, \ldots, \mu_K$. Έστω $\mu^* = \max\{\mu_i\}_i$.
- Συνολικό κέρδος S_n μετά απο n γύρους.
- Regret $R_n(\theta) = n\mu^* - \mathbb{E}[S_n]$.

Φράγματα Lai-Robbins

- Έχουμε K το πλήθος bandits με κατανομές $f(x, \theta_1), \ldots, f(x, \theta_K)$ και μέσες τιμές $\mu_1, \ldots, \mu_K$. Έστω $\mu^* = \max\{\mu_i\}_i$.
- Συνολικό κέρδος S_n μετά απο n γύρους.
- Regret $R_n(\theta) = n\mu^* - \mathbb{E}[S_n]$.
- Robbins: $\frac{R_n(\theta)}{n} = \mu^* - \frac{\mathbb{E}[S_n]}{n} \to 0$.

Φράγματα Lai-Robbins

- Έχουμε K το πλήθος bandits με κατανομές $f(x, \theta_1), \ldots, f(x, \theta_K)$ και μέσες τιμές $\mu_1, \ldots, \mu_K$. Έστω $\mu^* = \max\{\mu_i\}_i$.
- Συνολικό κέρδος S_n μετά απο n γύρους.
- Regret $R_n(\theta) = n\mu^* - \mathbb{E}[S_n]$.
- Robbins: $\frac{R_n(\theta)}{n} = \mu^* - \frac{\mathbb{E}[S_n]}{n} \to 0$.

Το Θεώρημα του Robbins εγγυάται ότι ναι μεν $R_n(\theta) \to \infty$, αλλά η σύγκλιση είναι πιο αργή από την ακολουθία n, δηλ. $R_n(\theta) = o(n)$.

Φράγματα Lai-Robbins

- Έχουμε K το πλήθος bandits με κατανομές $f(x, \theta_1), \ldots, f(x, \theta_K)$ και μέσες τιμές $\mu_1, \ldots, \mu_K$. Έστω $\mu^* = \max\{\mu_i\}_i$.
- Συνολικό κέρδος S_n μετά απο n γύρους.
- Regret $R_n(\theta) = n\mu^* - \mathbb{E}[S_n]$.
- Robbins: $\frac{R_n(\theta)}{n} = \mu^* - \frac{\mathbb{E}[S_n]}{n} \to 0$.

Το Θεώρημα του Robbins εγγυάται ότι ναι μεν $R_n(\theta) \to \infty$, αλλά η σύγκλιση είναι πιο αργή από την ακολουθία n, δηλ. $R_n(\theta) = o(n)$.

Μπορούμε να επιτύχουμε ακόμα πιο αργή σύγκλιση για το regret;

Ορισμός (Kullback-Leibler divergence)

Για τις κατανομές $f(x;\lambda)$ and $f(x;\mu)$ ορίζουμε την απόσταση Kullback - Leibler $I(\lambda,\mu)$ ως

$$I(\lambda,\mu) = \int f(x;\lambda)\ln\frac{f(x;\lambda)}{f(x;\mu)}dx.$$

Φράγματα Lai-Robbins

Ορισμός (Kullback-Leibler divergence)

Για τις κατανομές $f(x;\lambda)$ and $f(x;\mu)$ ορίζουμε την απόσταση Kullback - Leibler $I(\lambda,\mu)$ ως

$$I(\lambda,\mu) = \int f(x;\lambda)\ln\frac{f(x;\lambda)}{f(x;\mu)}dx.$$

Υποθέσεις (Υ1-Υ3):

Φράγματα Lai-Robbins

Ορισμός (Kullback-Leibler divergence)

Για τις κατανομές $f(x;\lambda)$ and $f(x;\mu)$ ορίζουμε την απόσταση Kullback - Leibler $I(\lambda,\mu)$ ως

$$I(\lambda,\mu) = \int f(x;\lambda) \ln \frac{f(x;\lambda)}{f(x;\mu)} dx.$$

Υποθέσεις (Υ1-Υ3):

- *Συνθήκη Κατανομών:* Μονοπαραμετρική οικογένεια κατανομών για τα rewards, $(f(x;\theta))_{\theta\in\Theta}$ με $\Theta \subseteq \mathbb{R}$.

Φράγματα Lai-Robbins

Ορισμός (Kullback-Leibler divergence)

Για τις κατανομές $f(x;\lambda)$ and $f(x;\mu)$ ορίζουμε την απόσταση Kullback - Leibler $I(\lambda,\mu)$ ως

$$I(\lambda,\mu) = \int f(x;\lambda) \ln \frac{f(x;\lambda)}{f(x;\mu)} dx.$$

Υποθέσεις (Υ1-Υ3):

- *Συνθήκη Κατανομών:* Μονοπαραμετρική οικογένεια κατανομών για τα rewards, $(f(x;\theta))_{\theta\in\Theta}$ με $\Theta \subseteq \mathbb{R}$.
- *Συνθήκη Συνέχειας:* Για κάθε $\lambda, \theta \in \Theta$, αν $(\lambda_n)_n$ τ.ω. $\mu(\lambda_n) \downarrow \mu(\lambda)$, τότε $I(\theta,\lambda_n) \to I(\theta,\lambda)$.

Ορισμός (Kullback-Leibler divergence)

Για τις κατανομές $f(x;\lambda)$ and $f(x;\mu)$ ορίζουμε την απόσταση Kullback - Leibler $I(\lambda,\mu)$ ως

$$I(\lambda,\mu) = \int f(x;\lambda) \ln \frac{f(x;\lambda)}{f(x;\mu)} dx.$$

Υποθέσεις (Υ1-Υ3):

- *Συνθήκη Κατανομών:* Μονοπαραμετρική οικογένεια κατανομών για τα rewards, $(f(x;\theta))_{\theta\in\Theta}$ με $\Theta \subseteq \mathbb{R}$.
- *Συνθήκη Συνέχειας:* Για κάθε $\lambda, \theta \in \Theta$, αν $(\lambda_n)_n$ τ.ω. $\mu(\lambda_n) \downarrow \mu(\lambda)$, τότε $I(\theta,\lambda_n) \to I(\theta,\lambda)$.
- *Συνθήκη Πυκνότητας:* Για κάθε $\lambda \in \Theta$, υπάρχει $(\lambda_n)_n$ in Θ τ.ω. $(\mu(\lambda_n))_n$ γνησίως φθίνουσα με $\mu(\lambda_n) \downarrow \mu(\lambda)$.

Φράγματα Lai-Robbins

Θεώρημα Lai-Robbins (1985)

Οι κατανομές των χεριών ικανοποιούν τις υποθέσεις Υ1-Υ3. Για έναν αλγόριθμο που ικανοποιεί την ιδιότητα ότι

$$R_n(\theta) = o(n^a)$$

για κάθε $\theta = (\theta_1, \ldots, \theta_K) \in \Theta^K$ και $a > 0$, ισχύει ότι

$$\liminf_n \frac{R_n(\theta)}{\ln n} \geq \sum_{i:\mu(\theta_i)<\mu^*} \frac{\mu^* - \mu(\theta_i)}{I(\theta_i, \theta^*)} > 0,$$

για τα θ για τα οποία τα $\mu(\theta_i)$ δεν είναι όλα ίσα.

Ο Αλγόριθμος Lai-Robbins

Υποθέτουμε ότι υπάρχουν

- συναρτήσεις $\tilde{\mu}_n(j)$ που παίζουν το ρόλο του δειγματικού μέσου του j-πληθυσμού, μετά από n-γύρους.

Ο Αλγόριθμος Lai-Robbins

Υποθέτουμε ότι υπάρχουν

- συναρτήσεις $\tilde{\mu}_n(j)$ που παίζουν το ρόλο του δειγματικού μέσου του j-πληθυσμού, μετά από n-γύρους.
- Συναρτήσεις $U_n(j)$ που παίζουν το ρόλο του άνω άκρου ενός διαστήματος εμπιστοσύνης για το μέσο του j-πληθυσμού, μετά από n-γύρους.

Ο Αλγόριθμος Lai-Robbins

Υποθέτουμε ότι υπάρχουν

- συναρτήσεις $\tilde{\mu}_n(j)$ που παίζουν το ρόλο του δειγματικού μέσου του j-πληθυσμού, μετά από n-γύρους.
- Συναρτήσεις $U_n(j)$ που παίζουν το ρόλο του άνω άκρου ενός διαστήματος εμπιστοσύνης για το μέσο του j-πληθυσμού, μετά από n-γύρους.

Ο Αλγοριθμος Lai-Robbins	
Βημα 1	Για $m = 1, \ldots, K$ επιλέγουμε $\phi(m) = m$.
Βημα 2	Κατά τον $n+1 \equiv j \bmod K$ γύρο, θέτουμε $I_n = \{m \in \{1, \ldots, K\} : T_n(m) \geq \delta n\}$, $j_n = \arg\max\{\tilde{\mu}_n(m) : m \in I_n\}$ και $\tilde{\mu}_n(j_n) = \max\{\tilde{\mu}_n(m) : m \in I_n\}$.
Βημα 3	Αν $\tilde{\mu}_n(j_n) \leq U_n(j)$, τότε $\phi(n+1) = j$, διαφορετικά $\phi(n+1) = j_n$.
Βημα 4	Επιστρέφουμε στο Βημα 2.

Ο Αλγόριθμος Auer-Bianchi-Fischer

- Οι συναρτήσεις $\tilde{\mu}_n(j)$ είναι ακριβώς οι δειγματικοί μέσοι του j-πληθυσμού μετά από n-γύρους.

Ο Αλγόριθμος Auer-Bianchi-Fischer

- Οι συναρτήσεις $\tilde{\mu}_n(j)$ είναι ακριβώς οι δειγματικοί μέσοι του j-πληθυσμού μετά από n-γύρους.
- Οι συναρτήσεις $U_n(j)$ δίνονται από την έκφραση

$$\overline{x}_{j,n_j} + \sqrt{\frac{3 \ln n}{2n_j}},$$

όπου $\overline{x}_{j,n_j}$ είναι ο δειγματικός μέσος του j-πληθυσμού και n_j το πλήθος των φορών που επιλέχθηκε ο j-πληθυσμός κατά τους πρώτους n γύρους.

Ο Αλγοριθμος Upper Confidence Bound	
Βημα 1	Επιλέγουμε κάθε πληθυσμό μία φορά.
Βημα 2	Στον $n+1$-γύρο, επιλέγουμε τον πληθυσμό που μεγιστοποιεί την έκφραση $\overline{x}_{j,n_j} + \sqrt{\frac{3 \ln n}{2n_j}}$.

Συνοπτική σύγκριση

Αλγοριθμος Lai-Robbins	Αλγοριθμος UCB

Συνοπτική σύγκριση

Αλγοριθμος Lai-Robbins	Αλγοριθμος UCB
Συγκεκριμένες υποθέσεις για τις κατανομές των κερδών (παραμετρικές, απαιτήσεις συνέχειας ως προς τη μετρική Kullback-Leibler, κλπ.).	Μόναδική υπόθεση είναι οι κατανομές να φέρονται στο $[0, 1]$.

Συνοπτική σύγκριση

Αλγοριθμος Lai-Robbins	Αλγοριθμος UCB
Συγκεκριμένες υποθέσεις για τις κατανομές των κερδών (παραμετρικές, απαιτήσεις συνέχειας ως προς τη μετρική Kullback-Leibler, κλπ.).	Μόναδική υπόθεση είναι οι κατανομές να φέρονται στο $[0, 1]$.
Η κατασκευή των δ.ε. είναι δύσκολη υπόθεση. Ακόμα και όταν τα δ.ε. δίνονται, ο υπολογισμός τους είναι απαιτητικός.	Η κατασκευή των δ.ε. είναι εύκολη και υπολογιστικά αποδοτική.

Συνοπτική σύγκριση

Αλγοριθμος Lai-Robbins	Αλγοριθμος UCB
Συγκεκριμένες υποθέσεις για τις κατανομές των κερδών (παραμετρικές, απαιτήσεις συνέχειας ως προς τη μετρική Kullback-Leibler, κλπ.).	Μόναδική υπόθεση είναι οι κατανομές να φέρονται στο $[0, 1]$.
Η κατασκευή των δ.ε. είναι δύσκολη υπόθεση. Ακόμα και όταν τα δ.ε. δίνονται, ο υπολογισμός τους είναι απαιτητικός.	Η κατασκευή των δ.ε. είναι εύκολη και υπολογιστικά αποδοτική.
Λογαριθμική απώλεια ασυμπτωτικά.	Λογαριθμική απώλεια ομοιόμορφα.

Συνοπτική σύγκριση

Αλγοριθμος Lai-Robbins	Αλγοριθμος UCB
Συγκεκριμένες υποθέσεις για τις κατανομές των κερδών (παραμετρικές, απαιτήσεις συνέχειας ως προς τη μετρική Kullback-Leibler, κλπ.).	Μόναδική υπόθεση είναι οι κατανομές να φέρονται στο $[0, 1]$.
Η κατασκευή των δ.ε. είναι δύσκολη υπόθεση. Ακόμα και όταν τα δ.ε. δίνονται, ο υπολογισμός τους είναι απαιτητικός.	Η κατασκευή των δ.ε. είναι εύκολη και υπολογιστικά αποδοτική.
Λογαριθμική απώλεια ασυμπτωτικά.	Λογαριθμική απώλεια ομοιόμορφα.
Η λογαριθμική σταθερά ισούται με $\frac{1}{2\Delta_j}$ για τους μη βέλτιστους πληθυσμούς j.	Η αντίστοιχη λογαριθμική σταθερά ισούται με $\frac{6}{\Delta_j} > \frac{1}{2\Delta_j}$.

References i


D. Bertsekas and J. Tsitsiklis.
***Neuro-Dynamic Programming*.**
Athena Scientific, 1996.


G. Cybenko.
Approximation by superposition of a sigmoidal function.
Mathematics of Control, Signal, and Systems, 2:303–314, 1989.


T. Lai and H. Robbins.
Asymptotically efficient adaptive allocation rules.
Advances in Applied Mathematics, 2:4–22, 1985.


T. Lattimore and C. Szepesvári.
***Bandit Algorithms*.**
Cambridge University Press, 2020.


P. F. P. Auer, N. Cesa-Bianchi.
Finite-time analysis of the multiarmed bandit problem.
Machine Learning, 47:235–256, 2002.