The two opposing forces

- #1 collect data (exploration)
- #2 select choice with highest win rate (exploitation)

Algorithm

- Epsilon-greedy
- optimistic initial values
- UCB1
- thompson sampling

CTRの例

- 2つの広告, どっちに投?
  - → 100サンプル, 1000サンプルでのCTRで比較.
  - 100, 1000が妥当か不明 (infinite numberのサンプルが必要)

Epsilon-Greedy

- greedy-method
  - picking the bandit with highest MLE win rate with no regard to confidence in prediction or amount of data.

ex) bandits with 90% and 80%

$$E(R) = (1-\varepsilon)\,0.9 + \varepsilon\left(\frac{0.8+0.9}{2}\right)$$

Sample mean

- X列は有限なので, $\bar{x}_n = \frac{1}{N}\sum_{i=1}^{N} x_i$ から改良が必要
  - → $O(N)$ を $O(1)$ にしたい

$$\bar{X}_N = \frac{1}{N}\left(\sum_{i=1}^{N-1} X_i + X_N\right)$$

$$\bar{X}_{N-1} = \frac{1}{N-1}\sum_{i=1}^{N-1} X_i \quad \leftrightarrow \quad (N-1)\bar{X}_{N-1} = \sum_{i=1}^{N-1} X_i$$

$$\bar{X}_N = \frac{1}{N}\left((N-1)\bar{X}_{N-1} + X_N\right)$$

$$= \frac{N-1}{N}\bar{X}_{N-1} + \frac{1}{N}X_N$$

$$= \left(1-\frac{1}{N}\right)\bar{X}_{N-1} + \frac{1}{N}X_N$$

$$= X_{N-1} + \frac{1}{N}\left(X_N - \bar{X}_{N-1}\right)$$

Optimistic Initial Values

算術平均

estimationの初期値を大きく設定する

→ arithmetic average はデータを集めれば集めるだけ小さくなる

→ もしあるbanditのestimated meanがmax estimated meanより小さくなったら，探索しない

Role of Initial Value

→ High initial value : 推定値が真の値に下がるまで時間がより必要

↕

Low initial value : 推定値が真の値に下がるまでは時間があまりかからない

UCB 1

- upper confidence bound
  - so far

    Epsilon greedy : small prob of random exploration

    Optimistic : naturally start at large value (each bandit will be chosen often)

- applying probability

  $p(\text{sample mean} - \text{true mean} \geq \text{error}) <= f(\text{error})$

  ex) $p(\text{sample mean} - \text{true mean} \geq t) \leq 1/t$

  - 大きな誤差よりも大きくなる prob は小さく
  - 小さな 〃 大きくなる

  Markov inequality / Chebyshev inequality / Hoeffding

  $$p(\bar{X}_n - E(X) \geq \epsilon) \leq e^{-2n\epsilon^2}$$

  n が大きくなると、右辺は小さくなる

$$\hat{j} = \underset{j}{\operatorname{argmax}}\left(\bar{X}_{nj} + \sqrt{2\frac{\log N}{n_j}}\right)$$

$\bar{X}_{nj}$ → sample mean

N : total plays

n_j : plays made on bandit j

Thompson Sampling

CIを思い出す。
Small Dataset : Large confidence interval
Fat → Explore more , skinny → explore less

CLT : sum of RVs RVs ~= normal Distribution

What we want ?
— the distribution of mean (win rate)
→ = a distribution の parameter も 何かの distribution に従う

$p(x \mid \theta)$: 尤度
母数θが与えられた時の, データの確率

なぜ evidence を考えずに, ∝で例りとしていいか?
→ evidence は積分であり, 計算が難しい
(Monte Carlo は計算がへビー)

事後分布 ∝ 尤度 · 事前分布

共役事前分布を使える
ex) Gaussian ∝ Gaussian × Gaussian

例. 尤度はベルヌーイ分布
→ これの共役事前分布 (conjugate Prior) は Beta 分布。

$$p(\theta|X) \propto \left(\prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{(1-x_i)}\right)\left(\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right)$$

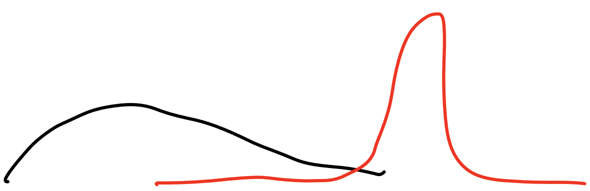
$$\propto \left(\prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{(1-x_i)}\right)\left(\theta^{\alpha-1}(1-\theta)^{\beta-1}\right)$$

$$= \left(\theta^{\sum_{i=1}^{N} x_i}(1-\theta)^{\sum_{i=1}^{N}(1-x_i)}\right)\left(\theta^{\alpha-1}(1-\theta)^{\beta-1}\right)$$

$$= \left(\theta^{\sum_{i=1}^{N} x_i+\alpha-1}(1-\theta)^{\sum_{i=1}^{N}(1-x_i)+\beta-1}\right)$$

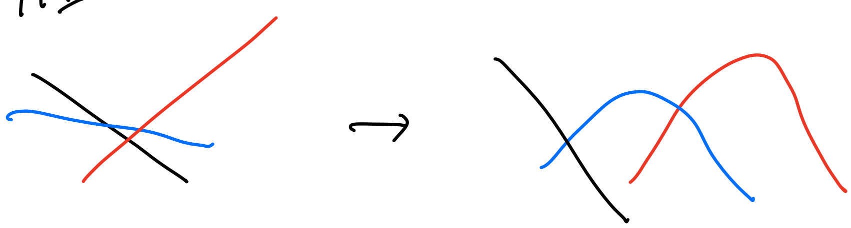
どのように prior を設定する?
→ Beta(1, 1) は一様分布
もしドメイン知識があればそれを使用する.

X を得る度に分布を更新する
ex) prior = Beta(1,1), collect 1, post (1+1, 1+1-1) = Beta(2,1)
(2,1) 1 Beta(3,1)
(3,1) 0 Beta(3,2)



→ このような分布では、黒はあまり選択されない
suboptimal な分布を使用しない

序盤



1 が出ないことが判明し、
コブになる

Rewards coming from normal distribution

- precision = 1 / variance

$$p(X|\mu,\tau) = \prod_{i=1}^{N} \sqrt{\frac{\tau}{2\pi}}\, e^{-\frac{\tau}{2}(x_i-\mu)^2}$$

$$X \sim N(\mu, \tau^{-1}) \quad , \quad \mu | X \sim N(m, \lambda^{-1})$$

$$p(\mu|X) = \frac{p(\mu, X)}{p(X)} = \frac{p(X|\mu)\,p(\mu)}{p(X)}$$

$$X \sim N(\mu, \tau^{-1}) \; , \; \mu \sim N(m_0, \lambda_0^{-1}) \; , \; \mu | X \sim N(m, \lambda^{-1})$$

$$p(\mu|X) \propto p(X|\mu)\,p(\mu)$$

$$= \left(\prod_{i=1}^{N} \sqrt{\frac{\tau}{2\pi}}\, e^{-\frac{\tau}{2}(x_i-\mu)^2}\right)\left(\sqrt{\frac{\lambda_0}{2\pi}}\, e^{-\frac{\lambda_0}{2}(\mu-m_0)^2}\right)$$

$$= \left(\left[\frac{\tau}{2\pi}\right]^{N} e^{-\frac{\tau}{2}\sum_{i=1}^{N}(x_i-\mu)^2}\right)\left(\quad .. \quad \right)$$

const $\frac{\tau}{2\pi} \rightarrow$

$$\propto \left(e^{-\frac{\tau}{2}\sum(x_i-\mu)^2}\right)\left(e^{-\frac{\lambda_0}{2}(\mu-m_0)^2}\right)$$

$$= e^{-\frac{\tau}{2}\sum^{N}(x_i-\mu) - \frac{\lambda_0}{2}(\mu-m_0)^2}$$

$$= e^{-\frac{\tau}{2}\sum^{N}(\mu^2 - 2\mu x_i + x_i)^2 - \frac{\lambda_0}{2}(\mu^2 - 2\mu m_0 + m_0^2)}$$

$$= \exp\left(-\frac{\tau}{2}\sum_{i=1}^{N}(\mu^2 - 2\mu x_i + x_i^2) - \frac{\lambda_0}{2}(\mu^2 - 2\mu m_0 + m_0^2)\right)$$

$$= \exp\left(-\frac{\tau}{2}(N\mu^2 - 2\mu\sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_i^2 - \frac{\lambda_0}{2}\mu^2 - \frac{\lambda_0}{2}2\mu m_0 - \frac{\lambda_0}{2}m_0^2\right)$$

$$\propto \exp\left(-\frac{\tau}{2}(N\mu^2 - 2\mu\sum^{N} x_i) - \frac{\lambda_0}{2}(\mu^2 - 2\mu m_0\right)$$

$$= \exp\left(-\frac{\tau N + \lambda_0}{2}\mu^2 + (\tau\sum_{i=1}^{N} x_i + \lambda_0 m_0)\mu\right)$$

$$p(\mu|x) = \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda}{2}(\mu - m)^2\right)$$

$$= \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda}{2}(\mu^2 - 2m\mu + m^2)\right)$$

$$\propto \exp\left(-\frac{\lambda}{2}(\mu^2 - 2m\mu)\right)$$

$$= \exp\left(-\frac{\lambda}{2}\mu^2 + m\lambda\mu\right)$$

$$\exp\left(-\frac{\tau N + \lambda_0}{2}\mu^2 + \left(\tau \sum_{i=1}^{N} x_i + \lambda_0 m_0\right)\mu\right)$$

$$\exp\left(-\frac{\lambda}{2}\mu^2 + m\lambda\mu\right)$$

$$\lambda = \tau N + \lambda_0$$

$$m\lambda = \tau \sum_{i=1}^{N} x_i + \lambda_0 m_0 \iff m = \frac{1}{\lambda}\left(\tau \sum_{i=1}^{N} x_i + \lambda_0 m_0\right)$$

$$m = \frac{1}{\tau N + \lambda_0}\left(\tau \sum_{i=1}^{N} x_i + \lambda_0 m_0\right)$$

$$\frac{x - \mu}{\sigma} = z$$

$$x = \sigma z + \mu$$