Stochastic Optimisation - Notes

Dom Hutchinson

October 6, 2020

1 Multi-Armed Bandit

1.1 The Problem

Example 1.1 - *Motivating Example*
Consider having a group of patients and several treatments they could be assigned to. How best do you go about determining which treatment is best? The obvious approach is to assign some of the patients randomly and then assign the rest to the best treatment, but how much evidence is sufficient? And how likely are you to choose a sub-optimal treatment?

Definition 1.1 - *Multi-Armed Bandit Problem*
An agent is faced with a choice of K actions. Each (discrete) time step the agent plays action i they receive a reward from the random real-valued distribution ν_i. Each reward is independent of the past. The distributions $\nu_1, \dots, \nu_K$ are unknown to the agent.
In the *Multi-Armed Bandit Problem* the agent seeks to maximise a measure of long-run reward.

Remark 1.1 - *Informal Definition of Multi-Armed Bandit Problem*
Given a finite set of actions and a random reward for each action, how best do we learn the reward distribution and maximise reward in the long-run.

Definition 1.2 - *Formal Definition of Multi-Armed Bandit Problem*
Consider a sequence of (unknown) mutually independent random variables $\{X_i(t)\}_{i\in[1,K]}$, with $t \in \mathbb{N}$. Consider $X_i(t)$ to be the distribution of rewards an agent would receive if they performed action i at time t. Since the rewards are independent of the past $X_i(t), X_i(t+1), \dots$ are IID random variables. The *Multi-Armed Bandit Problem* tasks us to find the greatest expected reward from all the actions.

$$\mu^* := \max_{i=1}^{K} \mu_i \quad \text{where } \mu_i = \mathbb{E}(X_i(t))$$

There are a number of ways to formalise this objective.

Remark 1.2 - *Assumptions*
For the *Multi-Armed Bandit Problem* we make the following assumptions about the set up

- When action i is played only the realisation of $X_i(t)$ is observed and none of $X_j(t),\ j \neq i$, are observed. Thus when the agent's t^{th} action is played only the rewards of actions $\{1, \dots, t-1\}$ are known to the agent.
- The agent has access to an external source of randomness which is used to choose it's next action.

Definition 1.3 - *Strategy, $I(\cdot)$*

Our agent's strategy $I : \mathbb{N} \to [1, K]$ is a function which determines which action the agent shall make at a given point in time. The strategy can use the knowledge gained from previous actions & their rewards only.

$$I(t) = I\big(t, \underbrace{\{I(s)\}_{\in[1,t)}}_{\text{Prev. Actions}}, \underbrace{\{X_{I(s)}(s)\}_{\in[1,t)}}_{\text{Prev. Rewards}}\big) \in [1, K]$$

Definition 1.4 - *Long-Run Average Reward Criterion, X_**
For a strategy $I(\cdot)$ we define the following measure for *Long-Run Average Reward*

$$X_* = \lim_{T\to\infty} \inf \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(X_{I(t)})$$

The *Infinum* is taken as there is no guarantee the limit exists (depending on the strategy), typically we will only deal with strategies where this limit exists.
Most strategies as based only on realisations of $\{X_i(s)\}_{s\in[1,t)}$, thus $\mathbb{E}(X_{I(t)}) \leq \mu^*$ and thus $X_* \leq \mu^*$. A strategy $I(\cdot)$ is *Optimal* if $X_* = \mu^*$.

Remark 1.3 - *It is not hard to find an Optimal Strategy in the (very) long run, so we are going to look at Regret Minimisation First.*

1.2 Regret Minimisation

Definition 1.5 - *Regret, R_n*
Regret is a measure of how much reward was lost during the first n time steps. The *Regret* R_n of a strategy $\{I(t)\}_{t\in\mathbb{N}}$ in the first n time steps is given by

$$\begin{aligned} R_n &= \max_{k=1}^{K} \sum_{t=1}^{n} \mathbb{E}[\underbrace{X_k(t)}_{\text{Best Pos}} - \underbrace{X_{I(t)}(t)}_{\text{Actual}}] \\ &= n\mu^* - \sum_{t=1}^{n} \mathbb{E}\big[X_{I(t)}(t)\big] \end{aligned}$$

Regret only involves expectation and thus can be learnt from observations. We want to produce a strategy where *Total Regret* grows sub-linearly.(i.e. $R_T/T \overset{T\to\infty}{\longrightarrow} 0$)

Remark 1.4 - *Minimising the growth rate of R_T with T is quite hard.*
The best achievable regret scales as $R_T \sim c \log T$ (i.e. $R_T/c\log T \overset{T\to\infty}{\longrightarrow} 1$) where c depends on the reward distributions $X_1(t), \dots, X_K(t)$.

Definition 1.6 - *Pseudo-Regret, $\tilde{R}_n$*
Pseudo-Regret $\tilde{R}_n$ is a less popular alternative to *Regret* R_n. The *Pseudo-Regret* R_n of a strategy $\{I(t)\}_{t\in\mathbb{N}}$ in the first n time steps is given by

$$\tilde{R}_n = \max_{k=1}^{K} \sum_{t=1}^{n} \big(X_k(t) - X_{I(t)}(t)\big)$$

Pseudo-Regret includes intrinsic randomness (which is independent of the past) and thus cannot be learnt from observations.

1.3 Best Arm Identification for Bernoulli Distribution

Example 1.2 - *Best Arm Identification for Bernoulli Bandits*
Consider a bandit with two *Bernoulli* arms: $\{X_1(t)\}_{t\in\mathbb{N}}$ IID RVs with distribution Bern(μ_1); and, $\{X_2(t)\}_{t\in\mathbb{N}}$ IID RVs with distribution Bern(μ_2).
Suppose $\mu_1 > \mu_2$ (i.e. arm 1 is better). Let the player play each arm n times and declare the arm with the greatest empirical mean to be the better arm. *What is the probability of choosing the wrong arm (Arm 2)?.*

An error occurs if $\sum_{t=1}^n X_2(t) \geq \sum_{t=1}^n X_1(t)$ and thus we want to calculate the probability of this event.
Define $\{Y(t)\}_{t\in\mathbb{N}}$ st $Y(t) := \{X_2(t) - X_1(t)$. This means $Y(t) \in \{-1, 0, 1\} \subset [-1, 1]$.
To use *Hoeffding's inequality* we need to scale Y to be in $[0, 1]$, so we define $Z(t) := \frac{1}{2}(Y(t) + 1)$.
We have $\mathbb{E}(Z(t)) = \frac{1}{2}(1 + \mu_2 - \mu_1)$ and an error occurs if $\sum_{t=1}^n Y(t) > 0 \iff \sum_{t=1}^n Z(t) \geq \frac{n}{2}$.
By *Hoeffding's Inequality*

$$\begin{array}{rcll}
\mathbb{P}(\text{error}) & = & \mathbb{P}\left(\sum_{i=1}^n Z(t) \geq \frac{n}{2}\right) & \\
& = & \mathbb{P}\left(\left(\sum_{i=1}^n Z(t)\right) - \frac{n}{2}(1+\mu_2-\mu_1) \geq \frac{n}{2}(\mu_1-\mu_2)\right) & \text{subtracting } \mu \text{ from both sides} \\
& = & \mathbb{P}\left(\sum_{i=1}^n \left(X_i - \underbrace{\frac{1}{2}(1+\mu_2-\mu_1)}_{\mu}\right) \geq n\underbrace{\frac{1}{2}(\mu_1-\mu_2)}_{t}\right) & \text{arranging for Hoeffding's} \\
& \leq & \exp\left(-2n\cdot\frac{1}{4}(\mu_1-\mu_2)^2\right) & \text{by Hoeffding's Inequality} \\
& = & \exp\left(-\frac{n}{2}(\mu_1-\mu_2)^2\right) &
\end{array}$$

1.4 Heuristic

2 Probability Inequalities

Remark 2.1 - *We can use the moments of a random variable to determine bounds on the probability of it taking values in a certain set.*
Theorem 2.1 - *Markov's Inequality*
Let X be a non-negative random variable. Then

$$\forall\, c > 0 \quad \mathbb{P}(X \geq c) \leq \frac{\mathbb{E}(X)}{c}$$

Proof

Consider an event A and define its indicator $\mathbb{1}(A)(\omega) := \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$. Fix $c > 0$, then

$$\begin{array}{rcl}
\mathbb{E}(X) & \geq & \mathbb{E}[X\mathbb{1}(X \geq c)] \\
& \geq & \mathbb{E}[c\mathbb{1}(X \geq c)] \\
& = & c\mathbb{P}(X \geq c) \\
\implies \mathbb{P}(X \geq c) & \leq & \frac{1}{c}\mathbb{E}(X)
\end{array}$$

Theorem 2.2 - *Chebyshev's Inequality*

Let X be a random-variable with finite mean and variance. Then

$$\forall\ c>0\quad \mathbb{P}(|X-\mathbb{E}(X)|\geq c)\leq\frac{\text{Var}(X)}{c^2}$$

Proof
Note that the events $|X-\mathbb{E}(X)|\geq c$ and $(X-\mathbb{E}(X))^2\geq c^2)$ are equivalent. Note that $\text{Var}([X-\mathbb{E}(X)]^2)=\text{Var}(X)$. Then the result follows by *Markov's Inequality.*

Theorem 2.3 - *Chebyshev's Inequality for Sum of IIDs*
Let $X_1,\dots,X_n$ be IID random variables with finite mean μ and finite variance σ^2.

$$\forall\ c>0\quad \mathbb{P}\left(\left|\left(\sum_{i=1}^n X_i\right)-n\mu\right|\geq nc\right)\leq\frac{\sigma^2}{nc^2}$$

Proof
This is proved by extending the proof of `Theorem 2.2` and noting that the variance of a sum of IIDs is the sum of the individual variances.

Theorem 2.4 - *Chernoff Bounds*
Let X be a random variable whose moment-generating function $\mathbb{E}[e^{\theta X}]$ is finite $\forall\theta$. Then

$$\forall\ c\in\mathbb{R}\quad \mathbb{P}(X\geq c)\leq\inf_{\theta>0}e^{-\theta c}\mathbb{E}(e^{\theta X})\quad\text{and}\quad\mathbb{P}(X\leq c)\leq\inf_{\theta<0}e^{-\theta c}\mathbb{E}(e^{\theta X})$$

Proof
Note that the events $X\geq c$ and $e^{\theta X}\geq e^{\theta c}$ are equivalent for all $\theta>0$. The result follows by applying *Markov's Inequality* to $r^{\theta X}$ and taking the best bound over all possible θ.

$$\begin{array}{rcl}\mathbb{P}(X\geq c)&=&\mathbb{P}(e^{\theta X}\geq e^{\theta c})\\&\leq&e^{-\theta c}\mathbb{E}(e^{\theta X})\\&\leq&\inf_{\theta<0}e^{-\theta c}\mathbb{E}(e^{\theta X})\end{array}$$

Theorem 2.5 - *Chernoff Bounds for Sum of IIDs*
Let $X_1,\dots,X_n$ be IID random variables. Then $\forall\ c\in\mathbb{R}$

$$\begin{array}{rl}\mathbb{P}\left(\sum_{i=1}^n X_i\geq nc\right)&\leq\inf_{\theta>0}e^{-n\theta c}\left(\mathbb{E}\left[e^{\theta X}\right]\right)^n\\\mathbb{P}\left(\sum_{i=1}^n X_i\leq nc\right)&\leq\inf_{\theta<0}e^{-n\theta c}\left(\mathbb{E}\left[e^{\theta X}\right]\right)^n\end{array}$$

Theorem 2.6 - *Jensen's Inequality*
Let f be a *Convex Function* and X be a random variable. Then

$$\mathbb{E}[f(X)]\geq f(\mathbb{E}[X])$$

Theorem 2.7 - *Bound on Moment Generating Function*
Let X be a random variable taking values in $[0,1]$ with finite expected value μ. Then we can bound the MGF of the centred random variable with

$$\forall\ \theta\in\mathbb{R}\quad\mathbb{E}\left[e^{\theta(X-\mu)}\right]\leq e^{\theta^2/8}$$

Proof (of weaker version)
Let X_1 be an independent copy of X, so both have mean μ. We can easily verify that $f(x)=e^{\theta x}$ is a convex function for all $\theta\in\mathbb{R}$. By *Jensen's Inequality* to $f(\cdot)$ and X_1

$$\mathbb{E}[e^{-\theta X_1}]\geq e^{-\theta\mathbb{E}[X_1]}=e^{-\theta\mu}\quad(1)$$

Consequently

$$\begin{array}{rrcll} & \mathbb{E}[e^{\theta(X-X_1)}] & = & \mathbb{E}[e^{\theta X}]\cdot\mathbb{E}[e^{-\theta X_1}] & \text{by independence}\\ & & \geq & \mathbb{E}[e^{\theta X}]\cdot e^{-\theta\mu} & \text{by (1)}\\ & & = & \mathbb{E}[e^{\theta(X-\mu)}] & \\ \implies & \mathbb{E}[e^{\theta(X-X_1)}] & \geq & \mathbb{E}[e^{\theta(X-\mu)}] & \end{array}$$

Since $X,X_1\in[0,1]$ then $(X-X_1)\in[-1,1]$. As X,X_1 have the same distribution $\mathbb{E}(X-X_1)=0$ and the distribution is symmetric around the mean.
Define random variable S which is independent of X,X_1 and takes values $\{-1,1\}$, each with probability $p=\frac{1}{2}$. $S(X-X_1)$ has the same distribution as $(X-X_1)$ due to independence of S and symmetry of $(X-X_1)$. Hence

$$\begin{array}{rrcll} & \mathbb{E}[e^{\theta(X-X_1)}] & = & \mathbb{E}[e^{\theta S(X-X_1)}] & \text{by identical distribution}\\ & & \leq & \mathbb{E}[e^{\theta S}] & \text{(2) since (X-X}_1) \in[-1,1]\\ & & = & \frac{1}{2}(e^{\theta}+e^{-\theta}) & \text{by def. of expectation}\\ \implies & \mathbb{E}[e^{\theta(X-X_1)}] & \leq & \frac{1}{2}(e^{\theta}+e^{-\theta}) & \end{array}$$

Note that $f(x)=e^x+e^{-x}$ is increasing for $x\in(0,\infty)$; decreasing for $x\in(-\infty,0)$; and symmetric around 0.
Using a *Taylor Series* we an observed that

$$\begin{array}{rrcll} & \frac{1}{2}(e^{\theta}-e^{-\theta}) & = & \sum\limits_{n=0}^{\infty}\frac{\theta^{2n}}{(2n)!} & \text{by Taylor expansion of} e^x\\ & & \leq & \sum\limits_{n=0}^{\infty}\frac{(\theta^2/2)^n}{n!} & \\ & & \overset{\text{def.}}{=} & e^{\theta^2/2} & \\ \implies & \frac{1}{2}(e^{\theta}+e^{-\theta}) & \leq & e^{\theta^2/2} & \end{array}$$

Combining all these results we get

$$\begin{array}{rl} & \mathbb{E}[e^{\theta(X-\mu)}]\leq\mathbb{E}[e^{\theta(X-X_1)}]\leq\frac{1}{2}(e^{\theta}+e^{-\theta})\leq e^{\theta^2/2}\\ \implies & \mathbb{E}[e^{\theta(X-\mu)}]\leq e^{\theta^2/2} \end{array}$$

□

Theorem 2.8 - *Hoeffding's Theorem*
Let $X_1,\dots,X_n$ be IID random variables taking values in $[0,1]$ and with finite expected value μ. Then

$$\forall\ t>0\quad\mathbb{P}\left(\sum_{i=1}^n(X_i-\mu)>nt\right)\leq e^{-2nt^2}$$

Proof
From *Chernoff's Bound* we have that

$$\forall\ \theta>0\quad\mathbb{P}\left(\sum_{i=1}^n(X_i-\mu)>nt\right)\leq e^{-\theta nt}\left(\mathbb{E}[e^{\theta(X-\mu)}]\right)^n$$

Using `Theorem 2.7` to bound the moment generating function, we get

$$\forall\ \theta > 0 \quad \mathbb{P}\left(\sum_{i=1}^n (X_i - \mu) > nt\right) \leq e^{-\theta nt} \cdot e^{n\frac{\theta^2}{8}} = e^{n\left(-\theta t + \frac{1}{8}\theta^2\right)}$$

Thus, by taking logs and rearranging, we get

$$\forall\ \theta > 0 \quad \frac{1}{n}\log\mathbb{P}\left(\sum_{i=1}^n (X_i - \mu) > nt\right) \leq -\theta t + \frac{\theta^2}{8}$$

We have that $-\theta t + \frac{\theta^2}{8}$ is minimised at $\theta = 4t$ which is positive if t is positive. Thus, by applying this bound and substituting $\theta = 4t$ we get

$$\forall\ \theta > 0 \quad \mathbb{P}\left(\sum_{i=1}^n (X_i - \mu) > nt\right) \leq e^{n\left(-4t^2 + \frac{1}{8}(16t^2)\right)} = e^{n(-4t^2 + 2t^2)} = e^{-2nt^2}$$

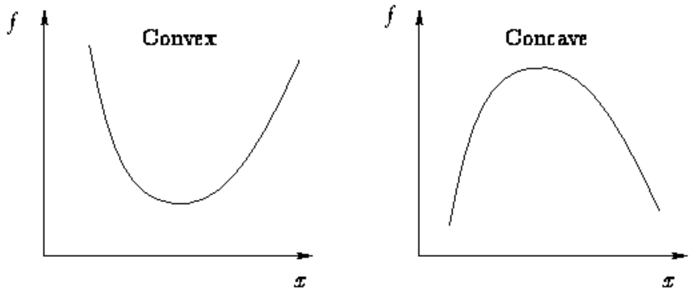
□

0 Reference

Definition 0.1 - *Convex Function*
A function $f : \mathbb{R} \to (\mathbb{R} \cup \{+\infty\})$ is *Convex* if, $\forall\ x, y \in \mathbb{R},\ \alpha \in [0, 1]$, we have

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

A smooth function f is convex iff f is twice differentiable and $f''(x) \geq 0\ \forall\ x \in \mathbb{R}$.
Visually, a function is convex if you can draw a line between any two points on the function and the function lies below the line.


f
Convex
x
f
Concave
x