Notes from Bandits

Divij Khaitan

May 2024

Basic Probability

Probability Spaces

A σ**-algebra** is a subset $\mathcal{F} \subseteq 2^{\Omega}$ of the power set of some outcome space Ω such that

- $\Omega \in \mathcal{F}$
- $A^c \in \mathcal{F} \; \forall \; A \in \mathcal{F}$
- $\bigcup_i A_i \in \mathcal{F} \; \forall \; \{A_i\} \in \mathcal{F}$

It is simply a collection of subsets of Ω which contains Ω, is closed under complementation and closed under countable union

A function $\mathbb{P} : \mathcal{F} \to R$ is called a **probability measure** or **probability distribution** or **distribution** if

- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}(A) \geq 0$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i) \forall$ disjoint and countable sequences of sets A_i.

If $\mathcal{G} \subset F$ is also a σ algebra, then it is called a **sub** σ **algebra**. The restriction of a measure $\mathbb{P}$ to $\mathcal{G}$ is denoted by $\mathbb{P}_{|\mathcal{G}}$

The elements of a σ-algebra are called **measurable sets**, and they are measurable in the sense that $\mathbb{P}$ can measure them.

$(\Omega, \mathcal{F})$: Measurable space
$(\Omega, \mathcal{F}, \mathbb{P})$: Probability space
Probability measure without the restriction that $\mathbb{P}(\Omega) = 1$ is just a **measure**
Probability measure without the restrictions that $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(A) \geq 0$ is a **signed measure**

Random variables lead to probability measures. $\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A))$. For a random variable $X, \mathbb{P}_X$ is called the **law** of X or **push-forward** measure of $\mathbb{P}$ under X.

$(\Omega, \mathcal{F})$ is a measurable space, $\mathcal{G} \subseteq 2^{\mathcal{X}}$ for $\mathcal{X}$ an arbitrary set, $X : \Omega \to \mathcal{X}$ is called a $\mathcal{F}/\mathcal{G}$ **measurable map** if $X^{-1}(A) \in \mathcal{F}\ \forall A \in \mathcal{G}$, $\mathcal{G}$ need not be a σ-algebra.

For a real valued map, $\mathcal{G}$ is typically the set of all open intervals $\{(a, b)|a < b\}$. If a map is $\mathcal{F}/\mathcal{G}$ measurable, it is also $\mathcal{F}/\sigma(\mathcal{G})$ **measurable**, where $\sigma(\mathcal{G}$ is the smallest σ algebra containing $\mathcal{G}$. It contains all of the sets A that are the intersection of every σ algebra containing $\mathcal{G}$.

The **Borel σ-algebra** is the smallest σ algebra on open rectangles of the form $\prod_{i=1}^{k}(a_i, b_i)$. On $\mathbb{R}$, these are just the open intervals (a, b).

A **random vector** on a measurable space $(\Omega, \mathcal{F})$ is a $\mathcal{F}/\mathfrak{B}(R^k)$ measurable function. A random element bewteen measurable spaces $(\Omega, \mathcal{F})$ and $(\mathcal{X}, \mathcal{G})$ is a $\mathcal{F}/\mathcal{G}$ measurable function $X : \Omega \to \mathcal{X}$.

$\sigma(X) = \{X^{-1}(A) \in \Omega : A \in \mathcal{G}\}$ is the **σ-algebra generated by** X. This is $\mathcal{F}/\mathcal{G}$ measurable iff $\sigma(X) \subseteq \mathcal{F}$. This is a sub-$\sigma$-algebra of $\mathcal{F}$, and the smallest σ algebra X is measurable over.

Borel functions are measurable functions over real spaces.
Caratheodory's extension theorem: $(\Omega_1, \mathcal{F}_1) \ldots (\Omega_n, \mathcal{F}_n)$ are a sequence of measurable spaces and $\overline{\mu} : \mathcal{F}_1 \times \ldots \mathcal{F}_n \to [0, 1]$ such that

- $\overline{\mu}(\Omega_1 \times \ldots \Omega_n) = 1$
- $\overline{\mu}(\bigcup_i A_i) = \sum_i \overline{\mu}(A_i)$ for sequences $A_i \in \mathcal{F}_1 \times \ldots \mathcal{F}_n$ of disjoint sets

Let $\Omega = \Omega_1 \times \ldots \Omega_n$ and $\mathcal{F} = \mathcal{F}_1 \times \ldots \mathcal{F}_n$, $\exists$ a unique probability measure μ which is identical to $\overline{\mu}$. The elements of $\mathcal{F}$ are called measurable rectangles.

Note: $\mathcal{F}_1 \times \mathcal{F}_2 \neq \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$. Let $\mathcal{F}_1 = \mathcal{F}_2 = 2^{\{1,2\}} = \{\{\}, \{1\}, \{2\}, \{1, 2\}\}$. $\{(1, 1), (2, 2)\} \in 2^{\{1,2\} \times \{1,2\}}$, which is not present in $\mathcal{F}_1 \times \mathcal{F}_2$

$\sigma(F_1 \times \ldots F_n)$ is called the **product σ-algebra**, also denoted by $F_1 \bigotimes \cdots \bigotimes F_n$. It's associative, so no brackets are needed. The n-fold product σ-algebra of $\mathcal{F}$

can be written as $\mathcal{F}^{\bigotimes n}$.

σ-algebras and Knowledge

$(\Omega, \mathcal{F})$, $(\mathcal{X}, \mathcal{G})$, $(\mathcal{Y}, \mathcal{H})$ are 3 measurable spaces. $X : \Omega \to \mathcal{X}$, $Y : \Omega \to \mathcal{Y}$ are 2 random elements. The question what does information on X tell us about Y can be answered using the σ-algebras they generate.
Factorisation Lemma
Given a Borel space $(\mathcal{Y}, \mathcal{H})$, $\mathcal{Y}$ is $\sigma(X)$ measurable iff $\exists$ a $\mathcal{G}/\mathcal{H}$ measurable map $f : \mathcal{X} \to \mathcal{Y}$.

Note: This is not the same as saying the random element Y can be deduced from the random element X, because the space of maps from $\mathcal{X} \to \mathcal{Y}$ may be much larger than the $\mathcal{G}/\mathcal{H}$ measurable maps. This is because $\sigma(X)$ also depends on $\mathcal{G}$ in addition to X, and if it is coarse such as the extreme case $\mathcal{G} = \{\mathcal{X}, \phi\}$ $\sigma(X)$ does not capture everything about X. It is also possible that f is not measurable, in which case the information in X cannot be measurably extracted.

Consider a sequence of random variables $X_1, \ldots, X_n$ on some common measurable space $(\Omega, \mathcal{F})$. Every response after observing the first t of the random variables, $X_{1:t} = (X_1, \ldots, X_t)$ is of the form $f \circ X_{1:t}$. These are all the $\sigma(X_{1:t})/\mathfrak{B}(\mathbb{R})$ measurable maps by the statements above. Thus, we can reason about these maps using just $\sigma(X_{1:t})$. $\mathcal{F}$ is independent of the response space, and also hides the range space of $X_{1:t}$ because it is a subset of $\mathcal{F}$. $F_0 = \{\Omega, \phi\}$ and $\mathcal{F}_0$ measurable maps are all constant functions on Ω. $F_0 \subseteq \mathcal{F}_1 \cdots \subseteq \mathcal{F}_n \subseteq \mathcal{F}$. **Filtration** is a sequence of σ-algebras $\{\mathcal{F}_t\}_{t=0}^n$ such that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$, and $\mathcal{F}_\infty = \sigma(\bigcup_{t=0}^\infty \mathcal{F}_t)$.

A sequence of random variables $X_t{}_{t=1}^n$ is **adapted** to a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t=1}^n$. We also say it is $\mathbb{F}$-adapted. X_t is **$\mathbb{F}$-predictable** if X_t is $\mathcal{F}_{t-1}$ measurable.

A process if $\mathbb{F}$-predictable if X_n can be predicted using $\mathcal{F}_{t-1}$, and $\mathbb{F}$-adaptable if X_n can be predicted using $\mathcal{F}_t$. Since $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$, all adapted processes are predictable, but the converse is not true.

A **filtered probability space** is a probability space $(\sigma, \mathcal{F}, \mathbb{P})$ with a filtration $\mathbb{F}$.

$\mathbb{P}(A|B)$ is the called the **a posteriori** probability, while the **a priori** probability is $\mathbb{P}(A)$. The mapping from $A \mapsto \mathbb{P}(A|B)$ is called the posterior probability measure given B, and exists $\forall$ B such that $\mathbb{P}(B) > 0$.
Independence: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, or $\mathbb{P}(A|B) = \mathbb{P}(A)$, i.e. the occurence of B carries no additional information on the likelihood of A occurring.

$\mathcal{G} \subseteq \mathcal{F}$ is **pairwise independent** if any 2 distinct elements of $\mathcal{G}$ are independent. $\mathcal{G}$ is **mutually independent** if $\forall A_i \in \mathcal{G}$, $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^n \mathbb{P}(A_i)$.

This is a stronger condition than pairwise independence, because it means knowing the occurence of any combination of finitely many events in $\mathcal{G}$ does not influence our prediction. Collections $\mathcal{G}_1$ and $\mathcal{G}_2$ are independent if $A \in \mathcal{G}_1$ and $B \in \mathcal{G}_2$ are independent $\forall\ A, B$. This can be extended to random variables, X and Y being independent when $\sigma(X)$ and $\sigma(Y)$ are independent.

Integrals and Expectation

Expectation is defined with respect to a particular random variable as well as a particular probability measure. Whenever the measure is not obvious, the expectation is denoted $\mathbb{E}_{\mathbb{P}}[X]$. It is the Lebesgue integral of X with respect to $\mathbb{P}$.

$$\int_\Omega X(\omega) d\mathbb{P}(\omega)$$

$$\int_\Omega \mathbb{I}\{\omega \in A\} d\mathbb{P} = \mathbb{P}(A)$$

$$\int_\Omega (\alpha_1 X_1 + \alpha_2 X_2) d\mathbb{P} = \alpha_1 \int_\Omega X_1 d\mathbb{P} + \alpha_2 \int_\Omega X_2 d\mathbb{P}$$

This property can be extended to the following

$$\int_\Omega X d\mathbb{P} = \sum_i \alpha_i \mathbb{P}(A_i)$$

For $X = \sum_i \alpha_i \mathbb{I}(A_i)$, $A_i \in \mathcal{F}$
When this is defined for finite sequences A_i, X is called a **simple function**. The above is the definition of the lebesgue integral of a simple function.

The main idea behind the **lebesgue integral** is approximating an arbitrary X from below using simple functions
For a non-negative random variable, this is
$\int_\omega X d\mathbb{P} = \sup\{\int_\Omega h d\mathbb{P} : h \text{ simple}, 0 \leq h \leq X\}$
For an arbitrary random variable, we can split it up into
$X^+(\omega) = X(\omega)\mathbb{I}\{X(\omega) > 0\}$
$X^-(\omega) = -X(\omega)\mathbb{I}\{X(\omega) > 0\}$
$X(\omega) = X^+(\omega) - X^-(\omega)$
This is the sum of 2 non-negative random variables, so by linearity we can split up the difference and integrate each component.
If $|f|$ is borel measurable, the Lebesgue and Riemann integrals are equal.

X_i is a sequence of random variables, such that $\mathbb{E}[X_i]$ exists, $X = \sum X_i$ and $\mathbb{E}[\sum X_i]$ exists, then
$\mathbb{E}[X] = \sum \mathbb{E}[X_i]$

If X and Y are independent and the expectations of $|X|, |Y|$ or $|XY|$ are finite, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

For nonnegative X, $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq x)dx$
$\mathbb{P}(X \geq x)$ is called the **Tail probability function**
The **Cumulative Distribution Function** is the complement of this, $\mathbb{P}(X \leq x)$

CDFs can be generalised to random vectors, where $F_X(x) = \mathbb{P}(X \leq x)$, where $X \leq x$ when every component of X is less than or equal to x.

The push-forward of a random element X is also a useful summary, where for a measurable $f(x)$,
$\mathbb{E}[f(X)] = \int_{\mathcal{X}} f(x)d\mathbb{P}_X(x)$

The **conditional expectation** is the random variable $\mathbb{E}[X|Y](\omega) = \mathbb{E}[X|Y = Y(\omega)]$
$\mathbb{E}[X|Y = y] = \sum_{\mathcal{X}} x\mathbb{P}(X = x|Y = y) = \sum_{\mathcal{X}} \frac{x\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}$
This is naturally only defined for events such that $\mathbb{P}(y) > 0$, and also does not generalise to continuous random variables

The conditional expectation for a continuous random variable is defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{H} \subseteq F$ a sub-σ-algebra of $\mathcal{F}$. The conditional expectation of X given $\mathcal{H}$ is $\mathbb{E}[X|\mathcal{H}]$, and is any $\mathcal{H}$ measurable random variable such that $\forall\ H \in \mathcal{H}$
$\int_H \mathbb{E}[X|\mathcal{H}]d\mathbb{P} = \int_H X d\mathbb{P}$
For a random variable Y, Such a random variable exists and is unique almost surely.

$\mathbb{P}$**-almost surely equal** ($\mathbb{P}$-a.s.) is defined as $\mathbb{P}(X = Y) = 1$. A $\mathbb{P}$**-null set** is any $U \in \mathcal{F}$ such that $\mathbb{P}(U) = 0$. $X = Y\mathbb{P}$-a.s. iff they disagree only on $\mathbb{P}$-null sets.
Properties of Conditional Expectations
$(\Omega, \mathcal{F}, \mathbb{P})$ measurable space, $\mathcal{G}, \mathcal{G}_1, \mathcal{G}_2$ are sub-σ-algebras of $\mathcal{F}$, X, Y integrable random variables on $\mathcal{F}$.

- If $X \geq 0$, $\mathbb{E}[X|\mathcal{G}]$ almost surely
- $\mathbb{E}[1|\mathcal{G}] = 1$ almost surely
- $\mathbb{E}[X + Y|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}]$
- $\mathbb{E}[XY|G] = Y\mathbb{E}[X|\mathcal{G}]$ almost surely if $\mathbb{E}[XY]$ exists and Y is $\mathcal{G}$ measurable.
- If $\mathcal{G}_1 \subseteq \mathcal{G}_2$ then $\mathbb{E}[X|\mathcal{G}_1] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1]$
- If $\sigma(X)$ is independent of $\mathcal{G}_2$ given $\mathcal{G}_1$, then $\mathbb{E}[X|\sigma(\mathcal{G}_1 \cup \mathcal{G}_2)] = \mathbb{E}[X|\mathcal{G}_1]$
- If $\mathcal{G}$ is the trivial σ-algebra $\{\Omega, \phi\}$, then $\mathbb{E}[X|G] = \mathbb{E}[X]$ almost surely.

Stochastic Processes and Markov Chains

Measurable Spaces $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ are called isomorphic if $\exists$ a bijective map which is $\mathcal{F}/\mathcal{G}$ measurable, with $\mathcal{G}/\mathcal{F}$ measurable inverse. A **Borel Space** is a space that is isomorphic to $(A, \mathfrak{B}(A))$ for any Borel measurable subset A of the reals.

For a borel space $\mathcal{S}$ with measure μ, $\exists$ a sequence of random elements on $([0,1], \mathfrak{B}([0,1]), \lambda)$ such that $\lambda_{X_t} = \mu \ \forall t$

Stochastic Processes

A **Stochastic Process** is a collection of random variables on some defined probability space, $\{X_t : t \in \mathcal{T}\}$.

$n \in \mathbb{N}^+$, $(\Omega_n, \mathcal{F}_n)$ is a measurable space. Let μ_n be a recursively defined measure on $(\Omega_1 \times \cdots \times \Omega_n), (F_1 \bigotimes \cdots \bigotimes F_n)$ as
$\mu_n(A \times \Omega_n) = \mu_{n-1}(A) \ \forall \ A \in \Omega_1 \bigotimes \cdots \bigotimes \Omega_{n-1}$
$\exists$ a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random elements $X_t : \Omega \to \Omega_t$ such that $\mathbb{P}_{X_1,\ldots,X_n} = \mu_n$

A **Markov Chain** is a sequence of random elements which has the property of $\mathbb{P}(X_n|X_1, \ldots, X_{n-1}) = \mathbb{P}(X_n|X_{n-1})$
A **probability kernel** or **Markov kernel** from $(\mathcal{X}, \mathcal{F})$ to $(\mathcal{Y}, \mathcal{G})$ is a map $K : (\mathcal{X}, \mathcal{G}) \to [0,1]$ such that

- $K(x, \dot{)}$ is a probability measure on $(\mathcal{Y}, \mathcal{G}) \ \forall \ x \in \mathcal{X}$
- $K(\dot{;}A)$ is an $\mathcal{F}$ measurable map $\forall A \in \mathcal{G}$

This can be related to a stochastic transition over the kernel space, where having arrived at state x the next state is sampled from $K(x, \dot{)}$. A can be thought of as being the set of possible states from x.
A **homogenous markov chain** is a sequence of random elements X_t taking values in a state space $\mathcal{S} = (\mathcal{X}, \mathcal{F})$ with $\mathbb{P}(X_{t+1} \in A|X_1, \ldots, X_t) = \mathbb{P}(X_{t+1} \in A|X_t) = \mu(X_t, A)$ almost surely.
μ is a probability kernel from $(\mathcal{X}, \mathcal{F})$ to $(\mathcal{X}, \mathcal{F})$, and $\mathbb{P}(X_1 \in A) = \mu_0(A)$ for some μ_0.
For $\mu(x, A) = \mu_0$, X_t is an i.i.d sequence of variables distributed as μ_0.

Martingales

An $\mathbb{F}$-adapted process X_t is called an **$\mathbb{F}$-adapted martingale** if

- X_t is integrable

- $\mathbb{E}[X_t|F_{t-1}] = X_{t-1}$

Can be thought of a process for which the current information doesn't provide indication for expecting a gain or loss on the next sample.
Super/sub martingale is the same, with the equality in 2 is replaces with $\leq$ / $\geq$
A stopping time with respect to $\mathbb{F}$ is a random variable for which $\mathbb{I}\{\tau \leq t\}$ is $\mathcal{F}_t$ measurable for all t. The σ algebra $\mathcal{F}_\tau$ is defined as

$$\{A \in F_\infty : A \cap \{\tau \leq t\} \in F_t\}$$

Doob's optional stopping
For a filtration $\mathbb{F}$, an $\mathbb{F}$-adapted martingale X_t, and τ an $\mathbb{F}$-stopping time anad least one of the following is true

- $\exists n \in \mathbb{N}$ such that $\mathbb{P}(\tau \geq n) = 0$ (stopping time has bounded support)
- $\mathbb{E}[\tau] \leq \infty$, and $\exists\ c$ such that $\mathbb{E}[|X_{t+1} - X_t| \mid \mathcal{F}_t] \leq c$ almost surely
- $\exists\ c$ such that $\mathbb{E}[X_{\min(t,\tau)}] \leq c$ almost surely (the variable is bounded before the process has stopped)

Then X_τ is almost surely well defined, with $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$. This is replaced with $\leq$ / $\geq$ for a super/sub martingale.
The second condition says the martingale is expected to stop, and is expected to make bounded jumps between consecutive values.

Maximal Inequality
$X_t \geq 0$ almost surely is a supermartingale, then $\forall\ \varepsilon \geq 0$

$$\mathbb{P}(\sup_t X_t \geq \varepsilon) \leq \frac{\mathbb{E}[X_0]}{\varepsilon}$$

Define $A_n = \{\sup_{t \leq n} X_t \geq \varepsilon\}$, and $\tau = \min((n+1), \arg\min_t X_t \geq \varepsilon)$
τ is a stopping time, which is bound above by $n+1$
$\mathbb{E}[X_0] \geq \mathbb{E}[X_\tau] \geq \mathbb{E}[X_\tau \mathbb{I}\{\tau \leq n\}] \geq \mathbb{E}[\varepsilon \mathbb{I}\{\tau \leq n\}] = \varepsilon \mathbb{P}(\tau \leq n) = \varepsilon \mathbb{P}(A_n)$
Rearranging, we get $\mathbb{P}(A_n) \leq \frac{\mathbb{E}[X_0]}{\varepsilon}$
Now, $\sup_t X_t \geq \varepsilon = \bigcup A_n$
$\mathbb{P}(\bigcup(A_n)) \leq \frac{\mathbb{E}[X_0]}{\varepsilon}$

Stochastic Bandits

Setup

A **bandit** is a collection of probability distributions $\nu = (P_a : a \in \mathcal{A})$, where $\mathcal{A}$ is a predifined set of actions.

The game proceeds as follows.

- In round t, $A_t \in \mathcal{A}$ is chosen
- Challenger observes a reward of $X_t \sim P_{A_t}$
- Return to step 1 and continue

The interaction between the learner and the policy induces a measure on the space $A_1, X_1, \ldots, A_n, X_n$, which is the sequence of outcomes. This sequence must meet the following criteria

1. The conditional distribution of the reward given the interaction so far is $\mathbb{P}(X_t | A_1, X_1, \ldots, A_n) = P_{A_n}$
2. The conditional law of action given $A_1, X_1, \ldots, A_n, X_n$ is $\pi_{n+1}(\dot{|} A_1, X_1, \ldots, A_n, X_n)$ where the π_ts are probability kernels.

Objective of the Game

The goal of the game is to maximise $S_t = \sum_t X_t$. This isn't an optimisation problem because the time horizon is unknown, the reward is random, a utility function needs to be assigned to distributions and the distributions are unknown.

Bandit Environments

An environment class ε is a collection of bandits with some common characteristics.
An environment class is unstructured if $\mathcal{A}$ is finite and $\exists$ sets of distributions $\mathcal{M}_a$ corresponding to each action such that
$\varepsilon = \{\nu = (P_a : a \in \mathcal{A}), P_a \in \mathcal{M}_a\} \forall a \in \mathcal{A}$
$\varepsilon = \times_{a \in \mathcal{A}} \mathcal{M}_a$
Finite degrees of freedom in defining a class of bandits makes it parametric, and infinite degrees of freedom makes it non parametric.

All bandits that are not unstructured are called structured.

Regret

Define the mean of action to be
$\mu_a(\nu) = \int_{-\infty}^{\infty} x dP_a(x)$
$\mu^\star(\nu) = \max_a \mu_a(\nu)$
The regret is defined as

$$R_n(\nu, \pi) = n\mu^\star(\nu) - \mathbb{E}[\sum_{t=1}^{n} X_t]$$

There are certain conditions every bandit problem must follow

1. $R(\nu, \pi) \geq 0$
2. A policy π choosing $\arg\max_a \mu_a(\nu)$ at every round has zero regret
3. If a policy has zero regret, $P(\mu_{A_t} = \mu^\star) = 1 \forall t$

Possible goals for the problem would be

- Make the regret sublinear, i.e. $\lim_{n\to\infty} \frac{R_n(\nu,\pi)}{n} = 0$
- A particular flavour of sublinearity, i.e. $R_n(\nu, \pi) \leq Cn^p$ for $C \geq 0$ and $0 \leq p \leq 1$
- For particular functions, find a policy satisfying $R_n(\nu, \pi) \leq C(\nu)f(n)$
- **Bayesian Regret** needs a distribution $\mathcal{Q}$ over the class of bandits, and is defined as $\int_\varepsilon R_n(\nu, \pi) d\mathcal{Q}(\nu)$

Regret Decomposition

Define $T_a(t) = \sum_{i=1}^{t} \mathbb{I}_{A_i=t}$
This is the number of times arm a has been pulled up until time t
$\Delta_a = \mu^\star - \mu_a$
This is the increase in reward which would have been incurred by choosing the optimal arm
Using these quantities, we can decompose the regret as follows
$R_n(\nu, \pi) = n\mu^\star(\nu) - \mathbb{E}[\sum_{t=1}^{n} X_t]$
$= \sum_a \sum_t \mathbb{E}[(\mu^\star - X_t)\mathbb{I}\{A_t = a\}] = \sum_a \sum_t \mathbb{I}\{A_t = a\}\mathbb{E}[(\mu^\star - X_t)] = \sum_a \sum_t \mathbb{I}\{A_t = a\}(\mu^\star - \mu_a)$ Simplifying, we get
$= \sum_a \mathbb{E}[T_a(n)](\mu^\star - \mu_a) = \sum_a \mathbb{E}[T_a(n)]\Delta_a$

This can be extended to a space with uncountably many actions, defining
$G(U) = \mathbb{E}[\sum_t \mathbb{I}\{A_t \in U\}]$
$R_n = \mathbb{E}[\sum_t \Delta_{A_t}] = \int_{\mathcal{A}} \Delta_a dG(a)$

Concentration Inequalities

Tail Probabilities

$X_1, X_2, \ldots, X_n$ i.i.d, with $\mathbb{E}[X] = \mu$ and $\mathbb{V}[X] = \sigma^2$
$\hat{\mu} = \frac{1}{n}\sum X_i$ is the most natural estimator and also happens to be unbiased. If we want to check how far it may be from the actual value μ, we can use the variance of the random variable $\hat{\mu}$
$\mathbb{V}[\hat{\mu}] = \mathbb{V}[\frac{1}{n}\sum X_i]$
$= \frac{1}{n^2}\mathbb{V}[\sum X_i]$
$= \frac{1}{n^2}\sum \mathbb{V}[X_i]$
$= \frac{1}{n^2} n\sigma^2$
$= \frac{\sigma^2}{n}$
We are particularly interested in the tail probabilities

$$\mathbb{P}[\hat{\mu} - \mu \leq \varepsilon] \text{ and } \mathbb{P}[\hat{\mu} - \mu \geq -\varepsilon]$$

Markov's inequality
For any nonnegative random variable X
$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}$
Proof: $\mathbb{E}[X] = \int_{-\infty}^{\infty} E[X]dx$
$= \int_0^{\infty} xf(x)dx$
$\geq \int_a^{\infty} xf(x)dx$
$\geq a\int_a^{\infty} f(x)dx$
$\geq a\mathbb{P}(|X| \geq a)$
$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}$
This can be applied to tail probabilities by looking at $|X|$

Chebyshev's inequality
$\mathbb{P}((X - \mathbb{E}[X])^2 \geq \varepsilon^2) \leq \frac{\mathbb{E}[(X-\mathbb{E}[X])^2]}{\varepsilon^2}$
$\mathbb{P}((X - \mathbb{E}[X]) \geq \varepsilon) \leq \frac{\mathbb{V}[X]}{\varepsilon^2}$

Using Chebyshev's inequality on the sample mean, we get
$\mathbb{P}(|\hat{\mu} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$

Central Limit Theorem
X_i i.i.d., $\frac{\lim_{n\to\infty}(\sum X_i) - n\mu}{\sigma\sqrt{n}} \sim \mathcal{N}(0,1)$
For any standard normal Z, $\mathbb{P}(Z \geq u) = \int_u^{\infty} \frac{1}{\sqrt{2\pi}} \exp(\frac{-x^2}{2})dx$
This has no closed form, but can be upper bounded as
$\int_u^{\infty} \frac{1}{\sqrt{2\pi}} \exp(\frac{-x^2}{2})dx \leq \frac{1}{\sqrt{2\pi}u}\int_u^{\infty} x\exp(\frac{-x^2}{2})dx \leq \frac{1}{\sqrt{2\pi}u}\exp(\frac{-u^2}{2})$
Note: $u \geq 0$ for this to hold, this breaks down with u and x having different signs
We can see the decay in the tail probability for the sample mean is
$\mathbb{P}(\hat{\mu} - \mu \geq \varepsilon) = \mathbb{P}(\frac{S_n}{\sigma\sqrt{n}} \geq \varepsilon\sqrt{\frac{n}{\sigma^2}}) \leq \frac{\sigma}{\sqrt{2\pi}n\varepsilon}\exp(\frac{n\varepsilon^2}{2\sigma^2})$

Cramer-Chernoff Bound

A random variable X is called **σ-subgaussian** if

$$\mathbb{E}[\exp(\lambda X)] \leq \frac{1}{2}(\lambda\sigma)^2$$

The **moment generating function** for a random variable is the function $M_X(\lambda) = \mathbb{E}[\exp(\lambda X)]$
This exists in some ε-neighbourhood around $t = 0$, and has the special property that $\frac{d^n}{d\lambda^n} M_X(\lambda)|_{\lambda=0}$ gives the nth moment of the distribution
Subgaussianity of a random variable can be expressed using the MGF, as

$$\psi_X(\lambda) = \log(M_X(\lambda)) \leq \frac{1}{2}(\lambda\sigma)^2$$

ψ_X is called the **cumulant generating function**. Both the MGF and the CGF need not exist for a distribution
All normals with mean 0 $\mathcal{N} \sim (0, \sigma^2)$ have $M_X(\lambda) = \frac{\lambda^2\sigma^2}{2}$, making them σ-subgaussian
A distribution is **heavy tailed** if it has infinite moments of all positive orders, i.e. if $M_X(\lambda) = \infty \; \forall\lambda > 0$. If any positive moment exists, it is called **light tailed**.

If X is σ-subgaussian
$\mathbb{P}(X \geq \varepsilon)$
$= \mathbb{P}(\exp(\lambda X) \geq \exp(\lambda\varepsilon)$
$\leq \mathbb{E}[\exp(\lambda\varepsilon))] \exp(-\lambda\varepsilon)$ (Markov)
$\leq \exp(\frac{(\sigma\lambda)^2}{2}) \exp(-\lambda\varepsilon)$ (σ-Subgausiian)
$= \exp(\frac{(\sigma\lambda)^2}{2} - \lambda\varepsilon)$
Setting $\lambda = \frac{\varepsilon}{\sigma^2}$
$= \exp(\frac{(\sigma\varepsilon)^2}{2\sigma^4} - \frac{\varepsilon}{\sigma^2}\varepsilon)$
$\therefore \mathbb{P}(X \geq \varepsilon) \leq \exp(-\frac{\varepsilon^2}{2\sigma^2})$
This is a useful inequality, which is proved using the **Cramer-Chernoff** method. This is the process of using the exponentials alongiside a constant that can be set later.
This is an inequality on the right tail. The inequality on the left tail is similar. Using the union bound $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ we can bound $|X|$ as

$$\mathbb{P}(|X| \geq \varepsilon) \leq 2\exp(-\frac{\varepsilon^2}{2\sigma^2})$$

The bounds can be rewritten as

$$\mathbb{P}(X \geq \sqrt{2\sigma^2\log(\frac{1}{\delta})}) \leq \delta \text{ and } \mathbb{P}(|X| \geq \sqrt{2\sigma^2\log(\frac{2}{\delta})})$$

The second tells that for delta small enough the chance of X leaving the interval $[-\sqrt{2\sigma^2 \log(\frac{2}{\delta})}, \sqrt{2\sigma^2 \log(\frac{2}{\delta})}]$ is vanishingly small

All subgaussian random variables have the following properties

1. For X σ-subgaussian, $\mathbb{E}[X] = 0$ and $\mathbb{V}[X] \leq \sigma^2$
2. cX is $|c|\sigma$-subgaussian
3. For X_1 σ_1-subgaussian and X_2 σ_2-subgaussian, $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-subgaussian

This allows us to bound the tail probabilities of $\hat{\mu} - \mu$ as
$\mathbb{E}[|\hat{\mu} - \mu| \geq \varepsilon] \leq 2\mathbb{E}(-\frac{n\varepsilon^2}{2\sigma^2})$
This is because $\hat{\mu} - \mu$ is $\frac{\sigma}{\sqrt{n}}$ subgaussian by using the above theorem for $X_i - \mu$ is σ subgaussian random variables

A gaussian with 0 mean and σ^2 is σ-subgaussian. If X has mean 0 and $X \in [a, b]$ almost surely, it is $\frac{b-a}{2}$-subgaussian

ETC

Explore-then-commit is a simple algorithm that has an initial exploration phase of mk rounds, where each of the k arms is explored m times and then the arm with the largest explored mean is chosen.
The average reward recieved from arm i up until time t is given by
$\hat{\mu}_i(t) = \frac{1}{T_i(i)} \sum_{j=1}^{t} X_j \mathbb{I}\{A_j = t\}$
$R_n \leq m \sum_{i=1}^{k} \Delta_i + (n - mk)(\sum_{i=1}^{k} \Delta_i \exp(-\frac{m\Delta_i^2}{4}))$

$\mathbb{E}[T_i(n)] = m + (n - mk)\mathbb{P}(\hat{\mu}_i(mk) \geq \max_{j \neq i} \hat{\mu_j}(mk))$
$\mathbb{P}(\hat{\mu}_i(mk) \geq \max_{j \neq i} \hat{\mu_j}(mk))$
$\leq \mathbb{P}(\hat{\mu}_i(mk) \geq \hat{\mu_1}(mk))$
$= \mathbb{P}(\hat{\mu}_i(mk) + \mu_1 - \mu_i \geq \hat{\mu_1}(mk) + \mu_1 - \mu_i)$
$= \mathbb{P}((\hat{\mu}_i(mk) - \mu_i) - (\hat{\mu_1}(mk) - \mu_1) \geq \Delta_i)$
Now, because the first variable is the sum of the difference of the sample means of 2 subgaussian, it itself is $\sqrt{\frac{2}{m}}$ subgaussian, which gives us the inequality
$\mathbb{P}(\hat{\mu}_i(mk) \geq \max_{j \neq i} \hat{\mu_j}(mk)) \leq \mathbb{P}((\hat{\mu}_i(mk) - \mu_i) - (\hat{\mu_1}(mk) - \mu_1) \geq \Delta_i)$
$\leq \exp(-\frac{m\Delta_i^2}{4})$

Using the decomposed form of the regret, we get
$\sum_i \Delta_i \mathbb{E}[T_(n)] \leq m(\sum_i \Delta_i) + (n - mk) \sum_i \Delta_i (\exp(-\frac{m\Delta_i^2}{4}))$

In the 2 arm case, the bound can be greatly simplified to
$m\Delta_2 + (n - 2m)(\Delta_2(\exp(-\frac{m\Delta_2^2}{4})))$

$\leq m\Delta + n\Delta \exp(-\frac{m\Delta^2}{4})$
The RHS is minimised for $m = \max\{1, \lceil \frac{4}{\delta^2} \log(\frac{n\Delta^2}{4}) \rceil\}$
$R_n \leq \min\{n\Delta, \Delta + \frac{4}{\Delta}(1 + \max\{1, \lceil \frac{4}{\Delta^2} \log(\frac{n\Delta^2}{4}) \rceil\})\}$
When making the common assumption that the optimality gap is less than 1, we get $R_n \leq 1 + C\sqrt{n}$
This is a **problem independent** bound, since it doesn't depend on the horizon or the optimality gap.

Upper Confidence Bound Algorithm

UCB uses the higher value of the concentration bound as it's estimate of the mean of each arm. This can be seen as **optimism in the face of uncertainty**.

$$UCB_i(t-1, \delta) = \left\{ \begin{array}{ll} \infty & \text{for } T_i(t-1) = 0 \\ \hat{\mu}(t-1) + \sqrt{\frac{2\log(\frac{1}{\delta})}{T_i(t-1)}} & \text{otherwise} \end{array} \right\}$$

The algorithm proceeds by choosing the arm with the largest UCB$(t-1, \delta)$ value at every iteration and updating the UCB weights accordingly
Index algorithms maximise some quantity(index) at every iteration, which is commonly dependent on the time step and the observed samples from that arm.
The regret guarantee for the UCB algorithm, assuming the arms are 1-subgaussian and $\delta = \frac{1}{n^2}$ is

$$R_n \leq 3(\sum_{i=1}^{k} \Delta_i) + \sum_{i \neq i^\star} \frac{16\log(n)}{\Delta_i}$$

The samples drawn from the ith arm are X_{ti}, and $\hat{\mu}_{si} = \sum_{t=1}^{s} X_{ti}$ is the empirical mean of the first s samples