Advanced Machine Learning

Bandit Problems

MEHRYAR MOHRI MOHRI@

COURANT INSTITUTE & GOOGLE RESEARCH

Multi-Armed Bandit Problem

- Problem: which arm of a K-slot machine should a gambler pull to maximize his cumulative reward over a sequence of trials?
  - stochastic setting.
  - adversarial setting.

Motivation

- Clinical trials: potential treatments for a disease to select from, new patient or category at each round (Thompson, 1933).
- Ads placement: selection of ad to display out of a finite set (which could vary with time though) for each new web page visitor.
- Adaptive routing: alternative paths for routing packets through a “series of tubes” or alternative roads for driving from a source to a destination.
- Games: different moves at each round of a game such as chess, or Go.

Key Problem

- Exploration vs exploitation dilemma (or trade-off):
  - inspect new arms with possibly better rewards.
  - use existing information to select best arm.

Outline

- Stochastic bandits
- Adversarial bandits

Stochastic Model

- K arms: for each arm $i \in \{1, \ldots, K\}$,
  - reward distribution P_i.
  - reward mean μ_i.
  - gap to best: $\Delta_i = \mu^* - \mu_i$, where $\mu^* = \max_{i \in [1,K]} \mu_i$.

Bandit Setting

- For $t = 1$ to T do
  - player selects action $I_t \in \{1, \ldots, K\}$ (randomized).
  - player receives reward $X_{I_t,t} \sim P_{I_t}$.
- Equivalent descriptions:
  - on-line learning with partial information ($\neq$ full).
  - one-state MDPs (Markov Decision Processes).

Objectives

- Expected regret

$$\mathrm{E}[R_T] = \mathrm{E}\left[\max_{i \in [1,K]} \sum_{t=1}^{T} X_{i,t} - \sum_{t=1}^{T} X_{I_t,t}\right].$$

- Pseudo-regret

$$\begin{aligned}\overline{R}_T &= \max_{i \in [1,K]} \mathrm{E}\left[\sum_{t=1}^{T} X_{i,t} - \sum_{t=1}^{T} X_{I_t,t}\right].\\ &= \mu^* T - \mathrm{E}\left[\sum_{t=1}^{T} X_{I_t,t}\right].\end{aligned}$$

- By Jensen's inequality, $\overline{R}_T \leq \mathrm{E}[R_T]$.

Expected Regret

- If $(X_{i,t} - \mu_i)$s take values in $[-r, +r]$, then

$$\mathrm{E}\left[\max_{i \in [1,K]} \sum_{t=1}^{T} (X_{i,t} - \mu^*)\right] \leq r\sqrt{2T \log K}.$$

- The $O(\sqrt{T})$ dependency cannot be improved;

  → better guarantees can be achieved for pseudo-regret.

Pseudo-Regret

- Expression in terms of Δ_is:

$$\overline{R}_T = \sum_{i=1}^{K} \mathrm{E}[T_i(T)]\Delta_i \ ,$$

where $T_i(t)$ denotes the number of times arm i was pulled up to time t, $T_i(t) = \sum_{s=1}^{t} 1_{I_s=i}$.

- Proof: $\overline{R}_T = \mu^* T - \mathrm{E}\left[\sum_{t=1}^{T} X_{I_t,t}\right] = \mathrm{E}\left[\sum_{t=1}^{T} (\mu^* - X_{I_t,t})\right]$

$$= \mathrm{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{K} (\mu^* - X_{i,t}) 1_{I_t=i}\right] = \sum_{t=1}^{T}\sum_{i=1}^{K} \mathrm{E}[(\mu^* - X_{i,t})]\,\mathrm{E}[1_{I_t=i}]$$

$$= \sum_{i=1}^{K} (\mu^* - \mu_i)\,\mathrm{E}\left[\sum_{t=1}^{T} 1_{I_t=i}\right] = \sum_{i=1}^{K} \mathrm{E}[T_i(T)]\Delta_i.$$

ε-Greedy Strategy

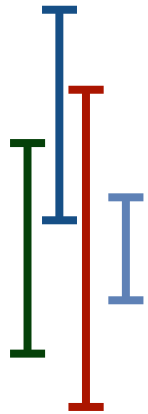
(Auer et al. 2002a)

- At time t,
  - with probability $1-\epsilon_t$, select arm i with best emp. mean.
  - with probability ϵ_t, select random arm.
- For $\epsilon_t = \min(\frac{6K}{\Delta^2 t}, 1)$, with $\Delta = \min_{i\colon \Delta_i > 0} \Delta_i$,
  - for $t \geq \frac{6K}{\Delta^2}$, $\Pr[I_t \neq i^*] \leq \frac{C}{\Delta^2 t}$ for some $C > 0$.
  - thus, $\mathrm{E}[T_i(T)] \leq \frac{C}{\Delta^2} \log T$ and $\overline{R}_T \leq \sum_{i\colon \Delta_i > 0} \frac{C\Delta_i}{\Delta^2} \log T$.
- Logarithmic regret but,
  - requires knowledge of Δ.
  - sub-optimal arms treated similarly (naive search).

UCB Strategy

(Lai and Robbins, 1985; Agrawal 1995; Auer et al. 2002a)

- Optimism in face of uncertainty:
  - at each time $t \in [1, T]$ compute upper confidence bound (UCB) on the expected reward of each arm $i \in [1, K]$.
  - select arm with largest UCB.
- Idea: wrong arm i cannot be selected for too long.
  - by definition, $\mu_i \leq \mu^* \leq \mathrm{UCB}_i$.
  - pulling i often $\longrightarrow$ UCB closer to μ_i .



Note on Concentration Ineqs

- Let X be a random variable such that for all $t \geq 0$,

$$\log \mathrm{E}\left[e^{t(X-\mathrm{E}[X])}\right] \leq \Psi(t),$$

where Ψ is a convex function. For Hoeffding's inequality and $X \in [a, b]$, $\Psi(t) = \frac{t^2(b-a)^2}{8}$.

- Then, $\mathbb{P}[X - \mathbb{E}[X] > \epsilon] = \mathbb{P}[e^{t(X-\mathbb{E}[X])} > e^{t\epsilon}]$

$$\begin{aligned}
&\leq \inf_{t>0} \left\{e^{-t\epsilon}\, \mathbb{E}[e^{t(X-\mathbb{E}[X])}]\right\} \\
&\leq \inf_{t>0} \left\{e^{-t\epsilon} e^{\Psi(t)}\right\} \\
&= e^{-\sup_{t>0}\{t\epsilon - \Psi(t)\}} \\
&= e^{-\Psi^*(\epsilon)}.
\end{aligned}$$

UCB Strategy

- Average reward estimate for arm i by time t:

$$\widehat{\mu}_{i,t} = \frac{1}{T_i(t)} \sum_{s=1}^{t} X_{i,s} 1_{I_s = i}.$$

- Concentration inequality (e.g., Hoeffding's ineq.):

$$\Pr\Big[\mu_i - \frac{1}{t} \sum_{s=1}^{t} X_{i,s} > \epsilon\Big] \leq e^{-t\psi^*(\epsilon)}.$$

- Thus, for any $\delta > 0$, with probability at least $1 - \delta$,

$$\mu_i < \frac{1}{t} \sum_{s=1}^{t} X_{i,s} + \psi^{*-1}\Big(\frac{1}{t} \log \frac{1}{\delta}\Big).$$

(α, ψ)-UCB Strategy

- Parameter $\alpha > 0$; (α, ψ)-UCB strategy consists of selecting at time t

$$I_t \in \underset{i \in [1, K]}{\operatorname{argmax}} \left[\widehat{\mu}_{i,t-1} + \psi^{*\,-1}\left(\frac{\alpha \log t}{T_i(t-1)} \right) \right].$$

(α, ψ)-UCB Guarantee

- Theorem: for $\alpha > 2$, the pseudo-regret of (α, ψ)-UCB satisfies

$$\overline{R}_T \leq \sum_{i \colon \Delta_i > 0} \left(\frac{\alpha \Delta_i}{\psi^*(\frac{\Delta_i}{2})} \log T + \frac{\alpha}{\alpha - 2} \right).$$

  - for Hoeffding's lemma, α-UCB, $\psi^*(\epsilon) = 2\epsilon^2$ (Auer et al. 2002a),

$$\overline{R}_T \leq \sum_{i \colon \Delta_i > 0} \left(\frac{2\alpha}{\Delta_i} \log T + \frac{\alpha}{\alpha - 2} \right).$$

Proof

- Lemma: for any $s \geq 0$, and any $i \in [K]$,

$$\sum_{t=1}^{T} 1_{I_t=i} \leq s + \sum_{t=s+1}^{T} 1_{I_t=i}\, 1_{T_i(t-1) \geq s}.$$

- Proof: observe that

$$\sum_{t=1}^{T} 1_{I_t=i} = \sum_{t=1}^{T} 1_{I_t=i} 1_{T_i(t-1) < s} + \sum_{t=1}^{T} 1_{I_t=i} 1_{T_i(t-1) \geq s}.$$

  - Now, for $t^* = \max \{t \leq T \colon 1_{T_i(t-1) < s} \neq 0\}$,

$$\sum_{t=1}^{T} 1_{I_t=i}\, 1_{T_i(t-1) < s} = \sum_{t=1}^{t^*} 1_{I_t=i}\, 1_{T_i(t-1) < s}.$$

  - By definition of t^*, the number of non-zero terms in the sum is at most s: $T_i(t^* - 1) < s \Rightarrow \sum_{t=1}^{t^*-1} 1_{I_t=i} < s$.

Proof

- For any i and t define $\eta_{i,t-1} = \psi^{*-1}\big(\frac{\alpha \log t}{T_i(t-1)}\big)$. At time t, if i is selected, then

$$
\begin{aligned}
&(\widehat{\mu}_{i,t-1} + \eta_{i,t-1}) - (\widehat{\mu}_{i^*,t} + \eta_{i^*,t-1}) \geq 0 \\
\Leftrightarrow &[\widehat{\mu}_{i,t-1} - \mu_i - \eta_{i,t-1}] + [2\eta_{i,t-1} - \Delta_i] + [\mu^* - \widehat{\mu}_{i^*,t-1} - \eta_{i^*,t-1}] \geq 0.
\end{aligned}
$$

Thus, at least one of these three terms is non-negative. Also, if one is non-positive, at least one of the other two is non-negative.

Proof

- To bound the pseudo-regret, we bound $\mathrm{E}[T_i(T)]$. But, observe first that

$$T_i(t-1) \geq s = \left\lceil \frac{\alpha \log T}{\psi^*(\frac{\Delta_i}{2})} \right\rceil \geq \frac{\alpha \log t}{\psi^*(\frac{\Delta_i}{2})} \Rightarrow \Delta_i - 2\eta_{i,t-1} \geq 0.$$

- Thus,

$$\begin{aligned}
\mathrm{E}[T_i(T)] &= \mathrm{E}\left[\sum_{t=1}^{T} 1_{I_t=i}\right] \\
&\leq s + \mathrm{E}\left[\sum_{t=s+1}^{T} 1_{I_t=i}\, 1_{T_i(t-1)\geq s}\right] \\
&\leq s + \sum_{t=s+1}^{T} \Pr[\widehat{\mu}_{i,t-1} - \mu_{i,t-1} - \eta_{i,t-1} \geq 0] + \Pr[\mu^* - \widehat{\mu}_{i^*,t-1} - \eta_{i^*,t-1} \geq 0].
\end{aligned}$$

Proof

- Each of the two probability terms can be bounded as follows using the union bound:

$$
\begin{aligned}
&\Pr[\mu^* - \widehat{\mu}_{i^*,t-1} - \eta_{i^*,t-1} \geq 0] \\
&\leq \Pr\left[\exists s \in [1,t] \colon \mu^* - \frac{1}{s}\sum_{k=1}^{s} X_{i,k} - \psi^{*\,-1}\Big(\frac{\alpha \log t}{s}\Big) \geq 0\right] \\
&\leq \sum_{s=1}^{t} \frac{1}{t^\alpha} = \frac{1}{t^{\alpha-1}}.
\end{aligned}
$$

- Final constant of the bound obtained by further simple calculations.

Lower Bound

(Lai and Robbins, 1985)

- Theorem: for any strategy such that $\mathrm{E}[T_i(T)] = o(T^\beta)$ for any arm i and any $\beta > 0$ for any set of Bernoulli reward distributions, the following holds for all Bernoulli reward distributions:

$$\liminf_{T \to +\infty} \frac{\overline{R}_T}{\log T} \geq \sum_{i \colon \Delta_i > 0} \frac{\Delta_i}{D(\mu_i \parallel \mu^*)}.$$

  - a more general result holds for general distributions.

Notes

- Observe that

$$\sum_{i\colon \Delta_i > 0} \frac{\Delta_i}{D(\mu_i \parallel \mu^*)} \geq \mu^*(1-\mu^*) \sum_{i\colon \Delta_i > 0} \frac{1}{\Delta_i},$$

since $D(\mu_i \parallel \mu^*) = \mu_i \log \frac{\mu_i}{\mu^*} + (1-\mu_i) \log \frac{1-\mu_i}{1-\mu^*}$

$$\begin{aligned} &\leq \mu_i \frac{\mu_i - \mu^*}{\mu^*} + (1-\mu_i) \frac{\mu^* - \mu_i}{1-\mu^*} \\ &= \frac{(\mu_i - \mu^*)^2}{\mu^*(1-\mu^*)} = \frac{\Delta_i^2}{\mu^*(1-\mu^*)}. \end{aligned}$$

Outline

- Stochastic bandits
- Adversarial bandits

Adversarial Model

- K arms: for each arm $i \in \{1, \ldots, K\}$,
  - no stochastic assumption.
  - rewards in $[0, 1]$.

Bandit Setting

- For $t = 1$ to T do
  - player selects action $I_t \in \{1, \ldots, K\}$ (randomized).
  - player receives reward $x_{I_t, t}$.
- Notes:
  - rewards $x_{i,t}$ for all arms determined by adversary simultaneously with the selection I_t of an arm by player.
  - adversary oblivious or nonoblivious (or adaptive).
  - strategies: deterministic, regret of at least $\frac{T}{2}$ for some (bad) sequences, thus must consider randomization.

Scenarios

- Oblivious case:
  - adversary rewards selected independently of the player's actions; thus, reward vector at time t only a function of t.
- Non-oblivious case:
  - adversary rewards at time t function of the player's past actions $I_1, \ldots, I_{t-1}$.
  - notion of regret problematic: cumulative reward compared to a quantity that depends on the player's actions! (single best action in hindsight function of actions $I_1, \ldots, I_T$ played; playing that single "best" action could have resulted in different rewards.)

Objectives

- Minimize regret ($\ell_{i,t} = 1 - x_{i,t}$), expectation or high prob.:

$$R_T = \max_{i \in [1,K]} \sum_{t=1}^{T} x_{i,t} - \sum_{t=1}^{T} x_{I_t,t} = \sum_{t=1}^{T} \ell_{I_t,t} - \min_{i \in [1,K]} \sum_{t=1}^{T} \ell_{i,t}.$$

- Pseudo-regret:

$$\overline{R}_T = \mathrm{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \min_{i \in [1,K]} \mathrm{E}\left[\sum_{t=1}^{T} \ell_{i,t}\right].$$

- By Jensen's inequality, $\overline{R}_T \leq \mathrm{E}[R_T]$.

Importance Weighting

- In the bandit setting, the cumulative loss of each arm is not observed, so how should we update the probabilities?
- Estimates via surrogate loss:
$$\widetilde{\ell}_{i,t} = \frac{\ell_{i,t}}{p_{i,t}} 1_{I_t = i} \ ,$$
where $\mathsf{p}_t = (p_{1,t}, \ldots, p_{K,t})$ is the probability distribution the player uses at time t to draw an arm ($p_{i,t} > 0$).
- Unbiased estimate: for any i,
$$\mathop{\mathrm{E}}_{I_t \sim \mathsf{p}_t}[\widetilde{\ell}_{i,t}] = \sum_{j=1}^{K} p_{j,t} \frac{\ell_{i,t}}{p_{i,t}} 1_{j=i} = \ell_{i,t}.$$

EXP3

(Auer et al. 2002b)

$\text{EXP3}(K)$

$\mathsf{p}_1 \leftarrow (\frac{1}{K}, \ldots, \frac{1}{K})$
$(\widetilde{L}_{1,0}, \ldots, \widetilde{L}_{K,0}) \leftarrow (0, \ldots, 0)$
for $t \leftarrow 1$ **to** T **do**
 $\quad$ $\text{SAMPLE}(I_t \sim \mathsf{p}_t)$
 $\quad$ $\text{RECEIVE}(\ell_{I_t,t})$
 $\quad$ **for** $i \leftarrow 1$ **to** K **do**
 $\quad\quad$ $\widetilde{\ell}_{i,t} \leftarrow \frac{\ell_{i,t}}{p_{i,t}} 1_{I_t=i}$
 $\quad\quad$ $\widetilde{L}_{i,t} \leftarrow \widetilde{L}_{i,t-1} + \widetilde{\ell}_{i,s}$
 $\quad$ **for** $i \leftarrow 1$ **to** K **do**
 $\quad\quad$ $p_{i,t+1} \leftarrow \dfrac{e^{-\eta \widetilde{L}_{i,t}}}{\sum_{j=1}^{K} e^{-\eta \widetilde{L}_{j,t}}}$
return p_{T+1}

EXP3 (Exponential weights for Exploration and Exploitation)

EXP3 Guarantee

- Theorem: the pseudo-regret of EXP3 can be bounded as follows:

$$\overline{R}_T \leq \frac{\log K}{\eta} + \frac{\eta K T}{2}.$$

Choosing η to minimize the bound gives

$$\boxed{\overline{R}_T \leq \sqrt{2KT \log K}.}$$

- Proof: similar to that of EG, but we cannot use Hoeffding's inequality since $\widetilde{\ell}_{i,t}$ is unbounded.

Proof

- Potential: $\Phi_t = \log \sum_{i=1}^K e^{-\eta \widetilde{L}_{i,t}}$.
- Upper bound:

$$\Phi_t - \Phi_{t-1} = \log \frac{\sum_{i=1}^K e^{-\eta \widetilde{L}_{i,t}}}{\sum_{i=1}^N e^{-\eta \widetilde{L}_{i,t-1}}} = \log \frac{\sum_{i=1}^K e^{-\eta \widetilde{L}_{i,t-1}} e^{-\eta \widetilde{\ell}_{i,t}}}{\sum_{i=1}^N e^{-\eta \widetilde{L}_{i,t-1}}}$$

$$= \log \Big[\mathop{\mathrm{E}}_{i \sim \mathsf{p}_t} \big[e^{-\eta \widetilde{\ell}_{i,t}} \big] \Big]$$

$$\leq \mathop{\mathrm{E}}_{i \sim \mathsf{p}_t} \big[e^{-\eta \widetilde{\ell}_{i,t}} \big] - 1 \qquad (\log x \leq x - 1)$$

$$\leq \mathop{\mathrm{E}}_{i \sim \mathsf{p}_t} \Big[-\eta \widetilde{\ell}_{i,t} + \frac{\eta^2}{2} \widetilde{\ell}^2_{i,t} \Big] \quad (e^{-x} \leq 1 - x + \tfrac{x^2}{2})$$

$$= -\eta \mathop{\mathrm{E}}_{i \sim \mathsf{p}_t} [\widetilde{\ell}_{i,t}] + \frac{\eta^2}{2} \mathop{\mathrm{E}}_{i \sim \mathsf{p}_t} \Big[\frac{l^2_{i,t} 1_{I_t = i}}{p^2_{i,t}} \Big]$$

$$= -\eta \ell_{I_t,t} + \frac{\eta^2}{2} \frac{l^2_{I_t,t}}{p_{I_t,t}} \leq -\eta \ell_{I_t,t} + \frac{\eta^2}{2} \frac{1}{p_{I_t,t}}.$$

Proof

- **Upper bound**: summing up the inequalities yields

$$\mathrm{E}[\Phi_T - \Phi_0] \leq -\eta \underset{I_t \sim \mathsf{p}_t}{\mathrm{E}} \left[\sum_{t=1}^{T} \ell_{I_t,t}\right] + \underset{I_t \sim \mathsf{p}_t}{\mathrm{E}} \left[\sum_{t=1}^{T} \frac{\eta^2}{2p_{I_t,t}}\right] = -\eta\, \mathrm{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] + \frac{\eta^2 KT}{2}.$$

- **Lower bound**: for all $j \in [1, K]$,

$$\begin{aligned}\mathrm{E}[\Phi_T - \Phi_0] &= \underset{I_t \sim \mathsf{p}_t}{\mathrm{E}} \left[\log\left[\sum_{i=1}^{K} e^{-\eta \widetilde{L}_{i,T}}\right] - \log K\right] \\ &\geq -\eta \underset{I_t \sim \mathsf{p}_t}{\mathrm{E}} [\widetilde{L}_{j,T}] - \log K = -\eta \underset{I_t \sim \mathsf{p}_t}{\mathrm{E}} [L_{j,T}] - \log K.\end{aligned}$$

- **Comparison:**

$$\forall j \in [1, K], \quad \eta\, \mathrm{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \eta\, \mathrm{E}[L_{j,T}] \leq \log K + \frac{\eta^2}{2} KT$$

$$\Rightarrow \overline{R}_T \leq \frac{\log K}{\eta} + \frac{\eta KT}{2}.$$

Notes

- When T is not known:
  - standard doubling trick.
  - or, use $\eta_t = \sqrt{\frac{\log K}{Kt}}$, then $\overline{R}_T \leq 2\sqrt{KT \log K}$.
- High probability bounds:
  - importance weighting problem: unbounded second moment (see (Cortes, Mansour, MM, 2010)), $\mathrm{E}_{i \sim \mathsf{p}_t}[\widetilde{\ell^2_{i,t}}] = \frac{\ell^2_{I_t,t}}{p_{I_t,t}}$.
  - (Auer et al., 2002b): mixing probability with a uniform distribution to ensure a lower bound on $p_{i,t}$; but not sufficient for high probability bound.
  - solution: biased estimate $\widetilde{\ell}_{i,t} = \frac{\ell_{i,t} 1_{I_t=i} + \beta}{p_{i,t}}$ with $\beta > 0$ a parameter to tune.

Lower Bound

(Bubek and Cesa-Bianchi, 2012)

- Sufficient lower bound in a stochastic setting for the pseudo-regret (and therefore for the expected regret).

- Theorem: for any $T \geq 1$ and any player strategy, there exists a distribution of losses in $\{0, 1\}$ for which

$$\overline{R}_T \geq \frac{1}{20}\sqrt{KT}.$$

Notes

- Bound of EXP3 matching lower bound modulo Log term.
- Log-free bound: $p_{i,t+1} = \psi(C_t - \widetilde{L}_{i,t})$ where C_t is a constant ensuring $\sum_{i=1}^{K} p_{i,t+1} = 1$ and ψ increasing, convex, twice differentiable over $\mathbb{R}^*$ (Audibert and Bubeck, 2010).
  - EXP3 coincides with $\psi(x) = e^{\eta x}$.
  - log-free bound with $\psi(x) = (-\eta x)^{-q}$ and $q = 2$.
  - formulation as mirror descent.
  - only in oblivious case.

References

- R. Agrawal. Sample mean based index policies with O(log n) regret for the multi-armed bandit problem. *Advances in Applied Mathematics*, vol. 27, pp. 1054–1078, 1995.

- Jean-Yves Audibert and Sebastian Bubeck. Regret bounds and minimax policies under partial monitoring. Journal of Machine Learning Research, vol. 11, pp. 2635– 2686, 2010.

- Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multi- armed bandit problem, *Machine Learning Journal*, vol. 47, no. 2–3, pp. 235– 256, 2002a.

- Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert Schapire. The non-stochastic multi-armed bandit problem. *SIAM Journal on Computing*, vol. 32, no. 1, pp. 48–77, 2002b.

- Sébastian Bubeck, Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning* 5, 1-122, 2012.

References

- Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.
- Corinna Cortes, Yishay Mansour, and Mehryar Mohri. Learning bounds for importance weighting. In NIPS, 2010.
- T.L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, vol. 6, pp. 4–22, 1985.
- Gilles Stoltz. Incomplete information and internal regret in prediction of individual sequences. *Ph.D. thesis, Universite Paris-Sud*, 2005.
- R. Tyrrell Rockafellar. Convex Analysis. *Princeton University Press*, 1970.
- W. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Bulletin of the American Mathematics Society*, vol. 25, pp. 285–294, 1933.