Lecture 3: Bandit models

Initial question:

Two probability distributions ν_1, ν_2

At time t, you choose $a_t \in \{1, 2\}$

and you observe y_t

if $a_t = j$, y_t is an independent sample of ν_j

Example: $\nu_j = \mathcal{N}(\mu_j, \sigma^2 = 1)$, $j = 1, 2$

Question: how do you find $j^* = \underset{j \in \{1,2\}}{\operatorname{argmax}}\ \mu_j$?

First approach:

total budget T

policy: assign n_1 times ν_1
n_2 times ν_2 with $n_1 + n_2 = T$

- pick $\hat{j} = \underset{j \in \{1,2\}}{\operatorname{argmax}} \overline{X^1_{n_1}}, \overline{X^2_{n_2}}$

where X^j_k = the k-th outcome when using distribution ν_j

and $\overline{X^j_n} = \frac{1}{n}(X^j_1 + \dots + X^j_n)$

We do the computation in the gaussian case.

$$\overline{X^1_{n_1}} = \mathcal{N}\left(\mu_1, \frac{1}{n_1}\right), \quad \overline{X^2_{n_2}} = \mathcal{N}\left(\mu_2, \frac{1}{n_2}\right)$$

Probability of mistake (assume $j^* = 1$ wlog)

$$\mathbb{P}\left(\hat{j}_T \neq j^*\right) = \mathbb{P}\left(\overline{X^1_{n_1}} < \overline{X^2_{n_2}}\right)$$

$$= \mathbb{P}\left(\frac{\overline{X^1_{n_1}} - \overline{X^2_{n_2}} - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -\frac{\mu_1 - \mu_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right) = \Phi\left(-\frac{\mu_1 - \mu_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right)$$

10)

where $\Phi(x) = \int_{-\infty}^{x} \frac{e^{-u^2/2}}{\sqrt{2\pi}} du$

Let $\rho_1 = \frac{n_1}{T}$ and $\rho_2 = \frac{n_2}{T}$

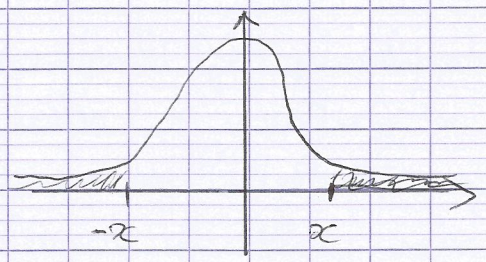
$$\frac{1}{n_1} + \frac{1}{n_2} = \frac{n_1 + n_2}{n_1 n_2} = \frac{1}{T\rho_1\rho_2}$$

$$\mathbb{P}\left(\hat{j}_T \neq j^*\right) = \Phi\left(-(\mu_1 - \mu_2)\sqrt{T\underbrace{\rho_1\rho_2}_{\leq \frac{1}{4}}}\right)$$

decreasing with T

minimale if $\rho_1 = \rho_2 = \frac{1}{2}$ ie $n_1 = n_2 = \frac{T}{2}$

I want a more explicit formula.



$x > 0$

$$\Phi(-x) = \int_{x}^{+\infty} \frac{e^{-u^2/2}}{\sqrt{2\pi}} du$$

$$\leq \int_{x}^{+\infty} \frac{u}{x} \frac{e^{-u^2/2}}{\sqrt{2\pi}} du$$

$$= \frac{1}{x\sqrt{2\pi}} \left[-u e^{-u^2/2} \right]_{x}^{+\infty}$$

$$= \frac{e^{-x^2/2}}{x\sqrt{2\pi}} = \frac{\phi(x)}{x}$$

where $\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \left(= \Phi'(x)\right)$

$\left(\Phi(x) \geq \left(\frac{1}{x} - \frac{1}{3x^2}\right)\phi(x) \right)$, in fact $\Phi(-x) \underset{x\to+\infty}{\sim} \frac{\phi(x)}{x}$

$Z \sim \mathcal{N}(0,1) \quad x > 0$

$$\mathbb{P}(Z > x) = \mathbb{P}(e^{\lambda Z} > e^{\lambda x})$$

$$\overset{\text{Markov}}{\leq} \frac{\mathbb{E}[e^{\lambda Z}]}{e^{\lambda x}}$$

$$= \frac{e^{\lambda^2/2}}{e^{\lambda x}} = e^{-(\lambda x - \lambda^2/2)}$$

Thus $\mathbb{P}(Z > x) \leq e^{-\sup_{\lambda > 0} \lambda x - \lambda^2/2}$

$$= e^{-x^2/2}$$

Thus $$\mathbb{P}(\hat{j}_T \neq j^*) \leq e^{-\frac{T(\mu - \mu_2)^2 p_1 p_2}{2}} \quad (1)$$

→ sample complexity of the pb: if we fix a probability of failure δ, what is the total sample size T that's we need to ensure that

$$\mathbb{P}(\hat{j}_T \neq j^*) \leq \delta \ ?$$

From (1) we have the upper-bound/approx.

$$e^{-T(\mu_1 - \mu_2)^2 p_1 p_2 / 2}$$

$$\Leftrightarrow T = \frac{2}{p_1 p_2} \times \frac{\log\left(\frac{1}{\delta}\right)}{(\mu_1 - \mu_2)^2}$$

In particular, the optimal fixed-design strategie has $p_1 = p_2 = \frac{1}{2}$ and a sample complexity

$$K_{FD}(\delta) = \frac{8 \log(1/\delta)}{(\mu_1 - \mu_2)^2}$$

→ useful only if the gap $\Delta = |\mu_1 - \mu_2|$ is known in advance

Idea 2: sequential statistics

At time t, I pick $A_t \in \{1,2\}$ according to a sampling strategy

$$(\psi_t)_{t \geq 1}, \quad \psi_t : (\{1,2\} \times \mathbb{R})^{t-1} \to \{1,2\}$$

$$A_t = \psi(A_1, Y_1, A_2, Y_2, \dots, A_{t-1}, Y_{t-1})$$

- Stopping rule: a stopping time τ with respect to $(\mathcal{F}_t)_{t \geq 1}$ where

$$\mathcal{F}_t = \sigma(A_1, Y_1, \dots, A_t, Y_t)$$

- A decision rule: $\hat{A}_\tau \in \{1,2\}$ measurable wrt $(\mathcal{F}_\tau)$

- Rules: for a failure probability δ fixed in advance, we are obliged to choose sampling strategy, a stopping rule and a decision rule such that

$$\mathbb{P}(\hat{A}_\tau \neq j^*) \leq \delta$$

- Goal: choose the strategy (= stopping rule, stopping time, decision rule) to minimize $\mathbb{E}_\nu[\tau]$, $\nu = (\nu_1, \nu_2)$.

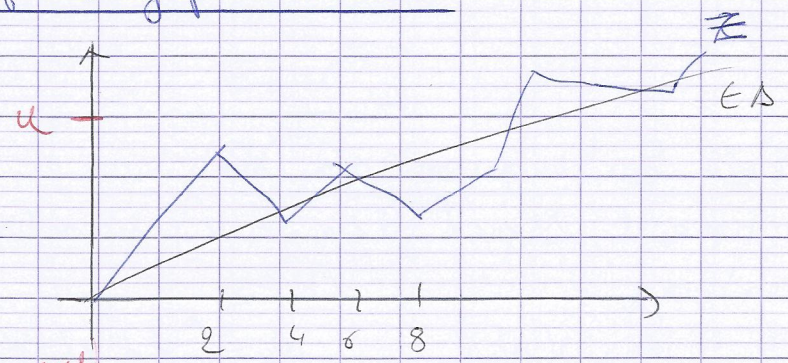
Intuitively: $\hat{A}_\tau = \underset{j \in \{1,2\}}{\operatorname{argmax}} \overline{X}_{N_j}^j(\tau)$

where $N_j(t) = \sum_{s \leq t} \mathbb{1}\{A_s = j\}$

- $A_t = 1 + (t \bmod 2)$

Decision rule ?

Ⓐ If the gap Δ is known.



$$S_k = Z_{2k} = X_1^1 + \dots + X_k^1 - X_1^2 - \dots - X_k^2$$
$$= \underbrace{(X_1^1 - X_1^2)}_{D_1} + \dots + (X_k^1 - X_k^2)$$

$D_k \overset{iid}{\sim} \mathcal{N}(\mu_1 - \mu_2, 2)$, w.l.o.g. $\mu_1 > \mu_2$

Stopping rule $\{\tau \leq 2k\} = \{|S_k| > u\}$

We need to choose u so that

$$\mathbb{P}\left(\exists k \geq 1 \quad S_k < -u \right) \leq \delta$$
~~and $\forall l < k$ $S_k < u$~~

Lemma: $\mathbb{P}(\exists k \geq 1 , S_k < -u) \leq e^{-\Delta u}$

Proof: $M_k = e^{-\Delta(D_1 + \dots + D_k)}$

$$\mathbb{E}\left[e^{-\Delta D_1}\right] = \mathbb{E}\left[e^{-\Delta(\sqrt{2}Z + \Delta)}\right] \quad // Z \sim \mathcal{N}(0,1)$$

$$= e^{-\Delta^2} \mathbb{E}\left[e^{-(\Delta\sqrt{2})Z}\right] = e^{-\Delta^2} e^{\frac{2\Delta^2}{2}} = 1$$

Then M_k is a martingale. Let $\tau_m = \tau' \wedge m$ (defined after)

But for every $m \geq 1$, $\tau' \wedge m$ is a bounded stopping

$\rightarrow$ From the Doob's stopping theorem, we conclude that

$$\mathbb{E}\left[M_{\tau_m}\right] = 1$$

for the stopping time $\tau' = \inf\{k, D_k \leq -u\}$

$\forall n > \tau'$, $S_{\tau' \wedge n} \leq -u$

Δu

and $M_{\tau' \wedge n} = e$

$$1 = \mathbb{E}[M_{\tau' \wedge n}] \geq \mathbb{E}[M_{\tau'_n} \mathbb{1}(\tau'_n \leq n)]$$

$$\geq e^{\Delta u} \mathbb{P}(\tau'_n < n)$$

$$= e^{\Delta u} \mathbb{P}(\tau' < n)$$

and $\mathbb{P}(\tau' \leq n) \leq e^{-\Delta u}$ for all $n \geq 1$.

Thus

$$\mathbb{P}(\tau' < \infty) \leq e^{-\Delta u}.$$

□

Consequence: choosing u s.t. $e^{-\Delta u} \leq \delta$

i.e. $u \geq \frac{\ln 1/\delta}{\Delta}$

yields an admissible stoping rule:

$$\mathbb{P}(\hat{A}_\tau \neq j^*) \leq \delta$$

To summarize:

$$A_t = \delta + (t+2 \bmod 2)$$

$$\tau = \inf\left\{ 2k : |X_1^1 + \dots X_k^1 - X_1^2 \dots - X_k^2| > \frac{\ln 1/\delta}{\Delta} \right\}$$

$$\hat{A}_\tau = \arg\max_j \overline{X}^j_{\tau/2}$$

We need to upper-bound its sample complexity

$$K_\delta(\Delta) = \mathbb{E}_\nu[\tau_\delta]$$

Lemma: Let H be a $\mathbb{N}$-valued r.v.

Then $\mathbb{E}[H] = \sum_{k=1}^{+\infty} \mathbb{P}(H \geq k)$

Proof:

$$\sum_{k=1}^{\infty} \mathbb{P}(H \geq k) = \sum_{k=1}^{+\infty} \sum_{j=k}^{+\infty} \mathbb{P}(H = j)$$

$$= \sum_{j=1}^{+\infty} \sum_{k=1}^{j} \mathbb{P}(H = j)$$

$$= \sum_{j=1}^{\infty} j\,\mathbb{P}(H=j) = \mathbb{E}[H]$$

□

$$\mathbb{E}[\tau_\delta] \leq \sum_{s=1}^{+\infty} \mathbb{P}(\tau_\delta \geq s)$$

But $\{\tau_\delta > 2k\} \subset \{S_k < u\}$ and $S_k \sim \mathcal{N}(k\Delta, 2k)$

So $\frac{S_k - k\Delta}{\sqrt{2k}} \sim \mathcal{N}(0,1)$

Thus, if $k \geq \frac{\ln \frac{1}{\delta}}{\Delta^2}$

$$\mathbb{P}(S_k < u) = \mathbb{P}\left(\frac{S_k - k\Delta}{\sqrt{2k}} < \frac{0 - k\Delta}{\sqrt{2k}}\right)$$

$$= \mathbb{P}\left(\frac{S_k - k\Delta}{\sqrt{2k}} < \frac{\left(\ln \frac{1}{\delta}\right)/\Delta - k\Delta}{\sqrt{2k}}\right)$$

$$\leq \exp\left(-\frac{1}{2}\left(\left(\frac{\ln 1/\delta}{\Delta} - k\Delta\right)/\sqrt{2k}\right)^2\right)$$

$$= \exp\left(-\frac{1}{4k\Delta^2}\left(k\Delta^2 - \ln\frac{1}{\delta}\right)^2\right)$$

We have seen that $\mathbb{P}\left(S_n < \frac{\ln 1/\delta}{\Delta}\right) \le \exp\left(-\frac{(n\Delta^2 - \ln 1/\delta)^2}{4n\Delta^2}\right)$

Thus, $\mathbb{E}[\tau_\delta] = \sum_{n=1}^{\infty} \mathbb{P}(\tau_\delta \ge n)$

$$\le 2 + 2\sum_{n=1}^{\infty} \mathbb{P}(\tau_\delta > 2n)$$

$$\le 2 + 2\left(\frac{\ln 1/\delta}{\Delta^2} + \frac{\sqrt{\ln 1/\delta}}{\Delta^2} + 1\right) + 2\sum_{n=\left\lceil \frac{\ln 1/\delta}{\Delta^2} + \frac{\sqrt{\ln 1/\delta}}{\Delta^2} \right\rceil}^{\infty} \mathbb{P}(\tau_\delta > 2n)$$

and $\sum_{n=\left\lceil \frac{\ln 1/\delta}{\Delta^2} + \frac{\sqrt{\ln 1/\delta}}{\Delta^2} \right\rceil}^{\infty} \mathbb{P}(\tau_\delta > 2n) \le \sum_{n=\left\lceil \frac{\ln 1/\delta}{\Delta^2} + \frac{\sqrt{\ln 1/\delta}}{\Delta^2} \right\rceil}^{\infty} \mathbb{P}\left(S_n < \frac{\ln 1/\delta}{\Delta}\right) \le \sum_{n=\left\lceil \frac{\ln 1/\delta}{\Delta^2} + \frac{\sqrt{\ln 1/\delta}}{\Delta^2} \right\rceil}^{\infty} \exp\left(-\frac{(n\Delta^2 - \ln 1/\delta)^2}{4n\Delta^2}\right)$

$$\le \sum_{u=\left\lceil \frac{\sqrt{\ln 1/\delta}}{\Delta^2} \right\rceil}^{\infty} \exp\left(-\frac{(u\Delta^2)^2}{4\Delta^2\left(\frac{\ln 1/\delta}{\Delta^2} + u\right)}\right) = \sum_{u=\left\lceil \frac{\sqrt{\ln 1/\delta}}{\Delta^2} \right\rceil}^{\infty} \exp\left(-\frac{u\Delta^2}{4\left(\frac{\ln 1/\delta}{u\Delta^2} + 1\right)}\right)$$

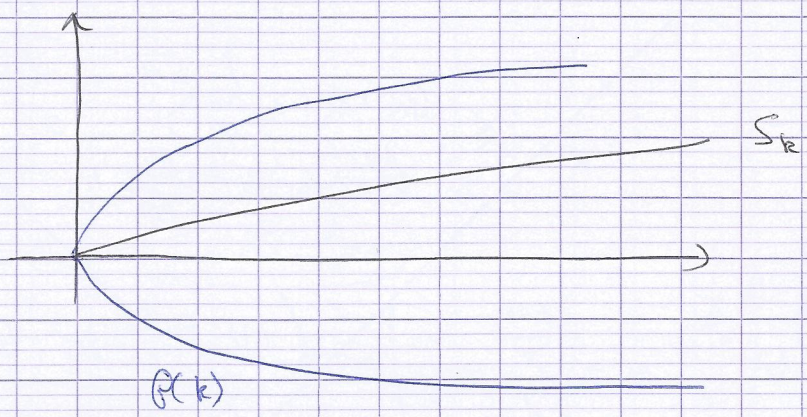
But $u\Delta^2 > \sqrt{\ln 1/\delta} \Rightarrow \frac{\sqrt{\ln 1/\delta}}{u\Delta^2} \times \sqrt{\ln 1/\delta} \le \sqrt{\ln 1/\delta}$ and thus

$$\sum_{n=\left\lceil \frac{\ln 1/\delta}{\Delta^2} + \frac{\sqrt{\ln 1/\delta}}{\Delta^2} \right\rceil}^{\infty} \mathbb{P}(\tau_\delta > 2n) \le \sum_{u=\left\lceil \frac{\sqrt{\ln 1/\delta}}{\Delta^2} \right\rceil}^{\infty} \exp\left(-\frac{u\Delta^2}{4\left(\sqrt{\ln 1/\delta} + 1\right)}\right)$$

$$\le 1 + \int_0^{+\infty} \exp\left(-\frac{u\Delta^2}{4\left(\sqrt{\ln 1/\delta} + 1\right)}\right) du = 1 + \frac{4\left(\sqrt{\ln 1/\delta} + 1\right)}{\Delta^2}$$

Thus, $\mathbb{E}[\tau_\delta] \le \frac{2\ln 1/\delta}{\Delta^2} + 4 + \frac{10\sqrt{\ln 1/\delta} + 8}{\Delta^2} \underset{\delta \to 0}{\sim} \frac{2\ln 1/\delta}{\Delta^2}$

(B) Δ is unknown



Remark: $\frac{S_k}{\sqrt{2k}} \sim \mathcal{N}(0,1)$

$\limsup_{k\to\infty} \frac{S_k}{\sqrt{2k}} = +\infty$; $\limsup_{k\to\infty} \frac{S_k}{\sqrt{2k\log k}} = 1$

Law of iterated logarithm

II Lower bounds in bandit model

Let k be a positive integer

Let $\nu_1, \nu_2, \dots, \nu_k$ be probability distributions on $\mathbb{R}$

We say that $\nu = (\nu_1, \dots, \nu_k)$ is a bandit model

For every $j \in \{1, \dots, k\}$, let $X_1^j, X_2^j, X_3^j, \dots$ be indep. samples of ν_j

A bandit strategy is a sampling strategy $(\psi_i)_{i \geq 1}$ where $\psi_t : (\{1, \dots, k\} \times \mathbb{R})^{t-1} \to \{1, \dots, k\}$

$\mathcal{S}_\nu$ defines a sequence of r.v.

- $A_1 = \psi_1(\emptyset)$, $y_1 = X_1^{A_1}$, $N_j(1) = \mathbb{1}\{A_1 = j\}$
- $A_2 = \psi_2(A_1, y_1)$, $y_2 = X^{A_2}_{N_{A_2}(2)}$, $N_j(2) = \sum_{s=1}^{2} \mathbb{1}\{A_s = j\}$

$\vdots$

- $A_t = \psi_t(A_1, y_1, \dots, A_{t-1}, y_{t-1})$, $y_t = X^{A_t}_{N_{A_t}(t)}$

$$N_j(t) = \sum_{s \le t} \mathbb{1}\{A_s = j\}$$

2 possible goals:

- Best arm identification
  → find $j^* \in \operatorname{argmax}_{1 \le j \le K} \mathbb{E}(X_1^j)$ (previously seen)

- Regret minimization

An oracle is someone who knows j^* in advance, on average, the oracle gathers

$$\mathbb{E}\left[\sum_{t=1}^{T} X_t^{j^*}\right] = T\mu^*, \text{ where } \mu^* = \max_j \mathbb{E}[X_1^j]$$

from the time 1 to T. A strategy ψ gathers

$$\mathbb{E}\left[\sum_{t=1}^{T} X^{A_t}_{N_{A_t}(t)}\right] \text{ during the same time.}$$

(This is the espected sum of rewards)

The regret is

$$R_T = T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} X^{A_t}_{N_{A_t}(t)}\right]$$

Regret minimization aims to choosing the strategy ψ so as to minimize R_T.

- To fix ideas, we will assume that the $(\nu_j)_j$ belongs to an exponential family.

$\nu^j = \mathcal{N}(\mu_j, 1)$

or $\nu^j = \mathcal{B}(\mu_j)$

or $\nu^j = \mathcal{P}(\mu_j)$

In particular, we will assume that we can compute the KL divergence between distribut° $KL(\nu^j, \nu^k)$

Regret minimizat° strategies

(greedy, ε-greedy, randomized)

Lower bound

Let ν and ν' be two bandit model.

Then, for all $T \geq 1$

$$\sum_{j=1}^{k} \mathbb{E}_\nu[N_j(T)]\, KL(\nu_j, \nu_j') \geq kl\left(\mathbb{E}_\nu[Z], \mathbb{E}_{\nu'}[Z]\right)$$

for every r.v. s.t. $0 \leq Z \leq 1$

Proof: Let $H_t = (A_1, g_1, A_2, g_2, \dots, A_t, g_t)$ and let ν^{H_t} be the distribut° of the H_t under the model ν.

Step 1: $KL\left(\nu^{H_t}, \nu'^{H_t}\right) \underset{\text{claim}}{=} \sum_{j=1}^{k} \mathbb{E}[N_j(t)]\, KL(\nu_j, \nu_j')$

Proof by induct°: $t = 0$ obvious

$t = 1$ $H^1 = (A_1, g_{A_1}^1)$, $\nu^{H_1} = (\delta_{a_1}, \nu_{a_1})$

$\nu'^{H_1} = (\delta_{a_1}, \nu'_{a_1})$

$$KL(\nu^{H_1}, \nu'^{H_1}) = \int d\nu^{H_1}(h) \ln \frac{d\nu^{H_1}(h)}{d\nu'^{H_1}(h)}$$

$$= \int d\nu_{a_1}(g) \ln \frac{d\nu_{a_1}(h)}{d\nu'_{a_1}(h)} = KL(\nu_{a_1}, \nu'_{a_1})$$

$N_j(h) = \#\{j = a_i\} = \mathbb{E}[N_j^n(h)]$

$$\sum_{j=1}^{K} \mathbb{E}[N_j^n(h)] K(\nu_j, \nu'_j) = KL(\nu_{a_1}, \nu'_{a_1})$$

Induction: $KL(\nu^{H_{t+1}}, \nu'^{H_{t+1}})$

$$= KL(\nu^{H_t}, \nu'^{H_t}) + \mathbb{E}\left[KL(\nu^{(A_{t+1}, g_{t+1})}, \nu'^{(A_{t+1}, g_{t+1})})\right]$$

$$= \sum_{j=1}^{K} \mathbb{E}[N_j^n(h) KL(\nu_j, \nu'_j)] + \mathbb{E}\Big[\sum_{j=1}^{K} \mathbb{1}\{A_{t+1} = j\} \times KL(\nu_{A_{t+1}=a}^{g_{t+1}}, \nu_{A_{t+1}=a}'^{g_{t+1}})\Big]$$

$$= \sum_{j=1}^{K} \mathbb{E}\Big[\sum_{s=1}^{t} \mathbb{1}\{A_s = j\}\Big] KL(\nu_j, \nu'_j) + \mathbb{E}\Big[\sum_{j=1}^{K} \mathbb{1}\{A_{t+1} = j\}\Big] KL(\nu_j, \nu'_j)$$

$$= \sum_{j=1}^{K} \mathbb{E}\Big[\sum_{s=1}^{t+1} \mathbb{1}\{A_s = j\}\Big] KL(\nu_j, \nu'_j)$$

Step 2: Contraction of entropy (data processing ineq.)

If Z is $\sigma(H_t)$-measurable

$$KL(\nu^Z, \nu'^Z) \le KL(\nu^{H_t}, \nu'^{H_t})$$

Step 3: if $0 \le Z \le 1$, then $KL(\nu^Z, \nu'^Z) \ge kl(\mathbb{E}_\nu(Z), \mathbb{E}_{\nu'}(Z))$

proof: $\mathbb{E}_\nu(Z) = \underset{U \sim \mathcal{U}}{\mathbb{E}}[U \le Z]$ where U is $\mathcal{U}([0,1])$ and $U \perp Z$.

$$P(U \le Z) = \int\int dP(z) \mathbb{1}\{u \le z\} = \int dP(z) \underbrace{\int \mathbb{1}\{u \le z\} du}_{z}$$

$$= \int z\, dP(z) = \mathbb{E}[Z]$$

Thus $KL\left((\nu\otimes\lambda)^{(\xi,a)}, (\nu'\otimes\lambda)^{(\xi,a)}\right)$

$= KL(\nu^\xi, \nu'^\xi) + KL(\lambda^a, \lambda^a)$

$\geq KL\left((\nu\otimes\lambda)^{\mathbb{1}\{U\leq\xi\}}, (\nu'\otimes\lambda)^{\mathbb{1}\{U\leq\xi\}}\right)$

$= kl\left(\mathbb{P}_{\nu\otimes\lambda}(U\leq\xi), \mathbb{P}_{\nu'\otimes\lambda}(U\leq\xi)\right)$

$= kl\left(\mathbb{E}_\nu(\xi), \mathbb{E}_{\nu'}(\xi)\right)$. □

To put everything together:

$$\sum_{j=1}^{K} \mathbb{E}[N_j(\tau)]\, KL(\nu_j, \nu'_j) = KL(\nu^{H\tau}, \nu'^{H\tau}) \geq KL(\nu^\xi, \nu'^\xi) \geq kl\left(\mathbb{E}_\nu(\xi), \mathbb{E}_{\nu'}(\xi)\right)$$

□

Example 2 arms gaussian, gap Δ unknown

$\nu = (\nu_1, \nu_2)$ $\quad \nu_j = \mathcal{N}(\mu_j, 1)$

w.l.o.g. $\mu_1 > \mu_2$, $j^* = 1$

$\nu' = (\nu_2, \nu_1)$, $\quad \xi = \mathbb{1}\{A_\tau = 1\}$

$$\mathbb{E}[N_1(\tau)]\, KL(\nu_1, \nu_2) + \mathbb{E}[N_2(\tau)]\, KL(\nu_2, \nu_1) \geq kl\left(\mathbb{E}_\nu(\mathbb{1}\{\hat{A}_\tau = 1\}), \mathbb{E}_{\nu'}[\mathbb{1}\{\hat{A}_\tau = 1\}]\right)$$

$$\Leftarrow \quad \frac{(\mu_1-\mu_2)^2}{2}\, \mathbb{E}[\underbrace{N_1(\tau)+N_2(\tau)}_{\tau}] \geq kl\Big(\underbrace{\mathbb{P}_\nu(\hat{A}_\tau = 1)}_{\geq 1-\delta}, \underbrace{\mathbb{P}_{\nu'}(\hat{A}_\tau = 1)}_{\leq \delta}\Big)$$

error at most with proba $\geq$

δ admissible strategy

$$\geq kl(1-\delta, \delta) = (1-\delta)\ln\left(\frac{1-\delta}{\delta}\right) + \delta\ln\left(\frac{\delta}{1-\delta}\right)$$

$$\geq \ln\left(\frac{1}{2,4\,\delta}\right)$$

$$\sim \ln\left(\frac{1}{\delta}\right)$$

(13)

$$\frac{(\mu_1-\mu_2)^2}{2}\,\mathbb{E}[\tau] \geq \ln\frac{1}{2.4\,\delta}$$

$$\boxed{\mathbb{E}[\tau] \geq \frac{2\ln\left(1/2.4\,\delta\right)}{\Delta^2}}$$