Macau: Why talk about regret when betting is so much more fun?

Dean Foster

Amazon.com, NYC

My message in one slide

- Setting: On-line decision making
  (*aka adversarial data or robust time series*)
- Goal: Use economic forecasts for decision making

My message in one slide

- Setting: On-line decision making
  (*aka adversarial data or robust time series*)
- Goal: Use economic forecasts for decision making
- Problem: Accuracy doesn't guarantee good decisions
  (*We'll take "accuracy" = "low regret." Regret compares actual decisions to "20/20 hindsight." 100s of papers say how to get low regret.*)

My message in one slide

- Setting: On-line decision making
  (*aka adversarial data or robust time series*)
- Goal: Use economic forecasts for decision making
- Problem: Accuracy doesn't guarantee good decisions
  (*We'll take "accuracy" = "low regret." Regret compares actual decisions to "20/20 hindsight." 100s of papers say how to get low regret.*)
- Solution: Falsifiable is better.
  - A falsifiable forecast can be bet against
  - The amount lost to the adversary is its *macau*.

My message in one slide

- Setting: On-line decision making
  (*aka adversarial data or robust time series*)
- Goal: Use economic forecasts for decision making
- Problem: Accuracy doesn't guarantee good decisions
  (*We'll take "accuracy" = "low regret." Regret compares actual decisions to "20/20 hindsight." 100s of papers say how to get low regret.*)
- Solution: Falsifiable is better.
  - A falsifiable forecast can be bet against
  - The amount lost to the adversary is its *macau*.

Take Aways

crazy-Calibration + low-regret $\Longrightarrow$ *low-macau* $\Longrightarrow$ *good decisions*

Prove the Earth is round!

- Fun question: What personal evidence do you have that the earth is round?

Prove the Earth is round!

- Fun question: What personal evidence do you have that the earth is round?
- Can you prove it is round? NO!
- But, you can make claims that could easily be shown wrong.
- Called falsifiability

Operationalizing falsifiability

- We will falsify someone's claim by winning bets placed against them
- Claim: $\hat{Y} \approx EY$
  - Prove it wrong by winning lots of money:

$$\text{Winnings} = E\left(B\,(Y - \hat{Y})\right)$$

  - $(Y - \hat{Y})$ is a "fair" bet
  - B is amount bet

Operationalizing falsifiability

- We will falsify someone's claim by winning bets placed against them
- Claim: $\hat{Y} \approx EY$
  - Prove it wrong by winning lots of money:

$$\text{Winnings} = E\left(B\,(Y - \hat{Y})\right)$$

  - $(Y - \hat{Y})$ is a "fair" bet
  - B is amount bet
- How to avoid being proven wrong by:

$$E\left(B\,(Y - \hat{Y})\right)$$

(*Start with bet B*)

Operationalizing falsifiability

- We will falsify someone's claim by winning bets placed against them
- Claim: $\hat{Y} \approx EY$
  - Prove it wrong by winning lots of money:

$$\text{Winnings} = E\left(B\,(Y - \hat{Y})\right)$$

  - $(Y - \hat{Y})$ is a "fair" bet
  - B is amount bet
- How to avoid being proven wrong by:

$$\text{Macau} \equiv \max_{|B| \leq 1} E\left(B\,(Y - \hat{Y})\right)$$

(*worry about worst bet*)

Operationalizing falsifiability

- We will falsify someone's claim by winning bets placed against them
- Claim: $\hat{Y} \approx EY$
  - Prove it wrong by winning lots of money:

$$\text{Winnings} = E\left(B\,(Y - \hat{Y})\right)$$

  - $(Y - \hat{Y})$ is a "fair" bet
  - B is amount bet
- How to avoid being proven wrong by:

$$\min_{\hat{Y}} \max_{|B| \leq 1} E\left(B\,(Y - \hat{Y})\right)$$

(*mini-max*)

Crazy calibration variable

Y	X_1	X_2	X_3	X_4
Y_1	X_{11}	X_{12}	X_{13}	X_{14}
Y_2	X_{21}	X_{22}	X_{23}	X_{24}
Y_3	X_{31}	X_{32}	X_{33}	X_{34}
Y_4	X_{41}	X_{42}	X_{43}	X_{44}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_t	X_{t1}	X_{t2}	X_{t3}	X_{t4}

Starting with our data that we observed up to time t

Crazy calibration variable

Y	X_1	X_2	X_3	X_4
Y_1	X_{11}	X_{12}	X_{13}	X_{14}
Y_2	X_{21}	X_{22}	X_{23}	X_{24}
Y_3	X_{31}	X_{32}	X_{33}	X_{34}
Y_4	X_{41}	X_{42}	X_{43}	X_{44}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_t	X_{t1}	X_{t2}	X_{t3}	X_{t4}

$$\hat{\beta}_t = \arg\min_\beta \sum_{i=1}^{t} (Y_i - \beta' X_i)^2$$

We can fit $\hat{\beta}_t$ on everything up to time t

Crazy calibration variable

$$\begin{array}{c|cccc|l}
Y & X_1 & X_2 & X_3 & X_4 & \\
\hline
Y_1 & X_{11} & X_{12} & X_{13} & X_{14} & \\
Y_2 & X_{21} & X_{22} & X_{23} & X_{24} & \\
Y_3 & X_{31} & X_{32} & X_{33} & X_{34} & \\
Y_4 & X_{41} & X_{42} & X_{43} & X_{44} & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \\
Y_t & X_{t1} & X_{t2} & X_{t3} & X_{t4} & \\
\hline
 & X_{t+1,1} & X_{t+1,2} & X_{t+1,3} & X_{t+1,4} & \hat{\beta}_t
\end{array}$$

$$\hat{Y}_{t+1} = \hat{\beta}_t' X_{t+1}$$

From a new X_{t+1}'s we can compute $\hat{Y}_{t+1}$

Crazy calibration variable

Y	X_1	X_2	X_3	X_4	$\hat{\beta}$
Y_1	X_{11}	X_{12}	X_{13}	X_{14}	0
Y_2	X_{21}	X_{22}	X_{23}	X_{24}	$\hat{\beta}_1$
Y_3	X_{31}	X_{32}	X_{33}	X_{34}	$\hat{\beta}_2$
Y_4	X_{41}	X_{42}	X_{43}	X_{44}	$\hat{\beta}_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_t	X_{t1}	X_{t2}	X_{t3}	X_{t4}	$\hat{\beta}_{t-1}$

Looking at only the first part of the data, we can generate:

$$\hat{\beta}_0, \quad \hat{\beta}_1, \quad \hat{\beta}_2, \quad \hat{\beta}_3, \quad \hat{\beta}_4, \quad \ldots, \quad \hat{\beta}_{t-1}$$

Crazy calibration variable

$$
\begin{array}{c|cccc|ccc}
Y & X_1 & X_2 & X_3 & X_4 & \hat{\beta} & & \hat{Y} \\
\hline
Y_1 & X_{11} & X_{12} & X_{13} & X_{14} & 0 & \rightarrow & \hat{Y}_1 = 0 \\
Y_2 & X_{21} & X_{22} & X_{23} & X_{24} & \hat{\beta}_1 & \rightarrow & \hat{Y}_2 = \hat{\beta}_1' X_2 \\
Y_3 & X_{31} & X_{32} & X_{33} & X_{34} & \hat{\beta}_2 & \rightarrow & \hat{Y}_3 = \hat{\beta}_2' X_3 \\
Y_4 & X_{41} & X_{42} & X_{43} & X_{44} & \hat{\beta}_3 & \rightarrow & \hat{Y}_4 = \hat{\beta}_3' X_4 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \rightarrow & \vdots \\
Y_t & X_{t1} & X_{t2} & X_{t3} & X_{t4} & \hat{\beta}_{t-1} & \rightarrow & \hat{Y}_t = \hat{\beta}_{t-1}' X_t \\
\hline
\end{array}
$$

Each of these leads to a next round

$$\hat{Y}_1, \quad \hat{Y}_2, \quad \hat{Y}_3, \quad \hat{Y}_4, \quad \ldots, \quad \hat{Y}_t$$

Crazy calibration variable

Y	X_1	X_2	X_3	X_4	$\hat{\beta}$		$\hat{Y}$
Y_1	X_{11}	X_{12}	X_{13}	X_{14}	0	$\rightarrow$	$\hat{Y}_1 = 0$
Y_2	X_{21}	X_{22}	X_{23}	X_{24}	$\hat{\beta}_1$	$\rightarrow$	$\hat{Y}_2 = \hat{\beta}_1' X_2$
Y_3	X_{31}	X_{32}	X_{33}	X_{34}	$\hat{\beta}_2$	$\rightarrow$	$\hat{Y}_3 = \hat{\beta}_2' X_3$
Y_4	X_{41}	X_{42}	X_{43}	X_{44}	$\hat{\beta}_3$	$\rightarrow$	$\hat{Y}_4 = \hat{\beta}_3' X_4$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\rightarrow$	$\vdots$
Y_t	X_{t1}	X_{t2}	X_{t3}	X_{t4}	$\hat{\beta}_{t-1}$	$\rightarrow$	$\hat{Y}_t = \hat{\beta}_{t-1}' X_t$

Theorem (Foster 1991, Forster 1999)

Such an on-line least squares forecast generates low regret:

$$\sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2 - \min_{\beta} \sum_{t=1}^{T} (Y_t - \beta' X_t)^2 \leq O(\log(T))$$

Crazy calibration variable

$$
\begin{array}{c|cccc|ccc}
Y & X_1 & X_2 & X_3 & X_4 & \hat\beta & & \hat Y \\
\hline
Y_1 & X_{11} & X_{12} & X_{13} & X_{14} & 0 & \rightarrow & \hat Y_1 = 0 \\
Y_2 & X_{21} & X_{22} & X_{23} & X_{24} & \hat\beta_1 & \rightarrow & \hat Y_2 = \hat\beta_1' X_2 \\
Y_3 & X_{31} & X_{32} & X_{33} & X_{34} & \hat\beta_2 & \rightarrow & \hat Y_3 = \hat\beta_2' X_3 \\
Y_4 & X_{41} & X_{42} & X_{43} & X_{44} & \hat\beta_3 & \rightarrow & \hat Y_4 = \hat\beta_3' X_4 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \rightarrow & \vdots \\
Y_t & X_{t1} & X_{t2} & X_{t3} & X_{t4} & \hat\beta_{t-1} & \rightarrow & \hat Y_t = \hat\beta_{t-1}' X_t \\
\hline
\end{array}
$$

Works no matter what the X's are.

Crazy calibration variable

$$
\begin{array}{c|cccc|ccc}
Y & X_1 & X_2 & X_3 & X_4 & \hat{\beta} & & \hat{Y} \\
\hline
Y_1 & X_{11} & X_{12} & \hat{Y}_1 & X_{14} & 0 & \rightarrow & \hat{Y}_1 = 0 \\
Y_2 & X_{21} & X_{22} & \hat{Y}_2 & X_{24} & \hat{\beta}_1 & \rightarrow & \hat{Y}_2 = \hat{\beta}_1' X_2 \\
Y_3 & X_{31} & X_{32} & \hat{Y}_3 & X_{34} & \hat{\beta}_2 & \rightarrow & \hat{Y}_3 = \hat{\beta}_2' X_3 \\
Y_4 & X_{41} & X_{42} & \hat{Y}_4 & X_{44} & \hat{\beta}_3 & \rightarrow & \hat{Y}_4 = \hat{\beta}_3' X_4 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \rightarrow & \vdots \\
Y_t & X_{t1} & X_{t2} & \hat{Y}_t & X_{t4} & \hat{\beta}_{t-1} & \rightarrow & \hat{Y}_t = \hat{\beta}_{t-1}' X_t \\
\hline
\end{array}
$$

Even if one of the X's were $\hat{Y}$!

Crazy calibration variable

Y	X_1	X_2	X_3	X_4	$\hat{\beta}$		$\hat{Y}$
Y_1	X_{11}	X_{12}	$\hat{Y}_1$	X_{14}	0	$\rightarrow$	$\hat{Y}_1 = 0$
Y_2	X_{21}	X_{22}	$\hat{Y}_2$	X_{24}	$\hat{\beta}_1$	$\rightarrow$	$\hat{Y}_2 = \hat{\beta}_1' X_2$
Y_3	X_{31}	X_{32}	$\hat{Y}_3$	X_{34}	$\hat{\beta}_2$	$\rightarrow$	$\hat{Y}_3 = \hat{\beta}_2' X_3$
Y_4	X_{41}	X_{42}	$\hat{Y}_4$	X_{44}	$\hat{\beta}_3$	$\rightarrow$	$\hat{Y}_4 = \hat{\beta}_3' X_4$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\rightarrow$	$\vdots$
Y_t	X_{t1}	X_{t2}	$\hat{Y}_t$	X_{t4}	$\hat{\beta}_{t-1}$	$\rightarrow$	$\hat{Y}_t = \hat{\beta}_{t-1}' X_t$

Theorem ($\implies$ Foster and Kakade 2008, Foster and Hart 2018)

Adding the crazy calibration variable generates low macau:

$$\sum_{t=1}^{T} X_t(Y_t - \hat{Y}_t) = O(\sqrt{T \log(T)})$$

Macau is a “normal equation”

$E(Y\|X)$	Least squares	Normal equations
Statistics	$\min_{\beta} \sum (Y_i - \beta \cdot X_i)^2$	$\sum X_i \ (Y_i - \beta \cdot X_i) = 0$

The normal equation is the same as:

$$\max_{\alpha} \sum_i \alpha' X_i (Y_i - \beta' X_i)) = 0$$

Which is solved by the β minimizer:

$$\min_{\beta} \max_{\alpha} \sum_i \alpha' X_i (Y_i - \beta' X_i)) = 0$$

Macau is a “normal equation”

$E(Y\|X)$	Least squares	Normal equations
Statistics	$\min_{\beta} \sum (Y_i - \beta \cdot X_i)^2$	$\min_{\beta} \max_{\alpha} \sum \alpha \cdot X_i \; (Y_i - \beta \cdot X_i)$

Macau is a "normal equation"

$E(Y \mid X)$	Least squares	Normal equations
Statistics	$\min\limits_{\beta} \sum (Y_i - \beta \cdot X_i)^2$	$\min\limits_{\beta} \max\limits_{\alpha} \sum \alpha \cdot X_i \; (Y_i - \beta \cdot X_i)$
Probability	$\min\limits_{f} E\left((Y - f(X))^2\right)$	$(\forall g)\; E(g(X)\;(Y - f(X))) = 0$

The normal equation is the same as:

$$\max_{g} E\left(g(X)(Y - f(X))\right) = 0$$

Which is solved by the $f(\cdot)$ minimizer:

$$\min_{f} \max_{g} E\left(g(X)(Y - f(X))\right) = 0$$

Macau is a “normal equation”

$E(Y\|X)$	Least squares	Normal equations
Statistics	$\min_\beta \sum (Y_i - \beta \cdot X_i)^2$	$\min_\beta \max_\alpha \sum \alpha \cdot X_i \ (Y_i - \beta \cdot X_i)$
Probability	$\min_f E\left((Y - f(X))^2\right)$	$\min_f \max_g E\Big(g(X) \ (Y - f(X))\Big)$

Macau is a “normal equation”

$E(Y\|X)$	Least squares	Normal equations
Statistics	$\min_\beta \sum (Y_i - \beta \cdot X_i)^2$	$\min_\beta \max_\alpha \sum \alpha \cdot X_i \ (Y_i - \beta \cdot X_i)$
Probability	$\min_f E\left((Y - f(X))^2\right)$	$\min_f \max_g E\Big(g(X) \ (Y - f(X))\Big)$
online	low regret	low macau

$$Regret \equiv \sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2 - \min_\beta \sum_{t=1}^{T} (Y_t - \beta \cdot X_t)^2$$

Macau is a "normal equation"

$E(Y\|X)$	Least squares	Normal equations
Statistics	$\min_\beta \sum (Y_i - \beta \cdot X_i)^2$	$\min_\beta \max_\alpha \sum \alpha \cdot X_i \, (Y_i - \beta \cdot X_i)$
Probability	$\min_f E\left((Y - f(X))^2\right)$	$\min_f \max_g E\Big(g(X) \, (Y - f(X))\Big)$
online	low regret	low macau

$$Macau \equiv \max_{\alpha:|\alpha|\leq 1} \sum_{t=1}^{T} \alpha \cdot X_t \left(Y_t - \hat{Y}_t\right)$$

Macau is a "normal equation"

$E(Y \mid X)$	Least squares	Normal equations
Statistics	$\min_{\beta} \sum (Y_i - \beta \cdot X_i)^2$	$\min_{\beta} \max_{\alpha} \sum \alpha \cdot X_i \; (Y_i - \beta \cdot X_i)$
Probability	$\min_{f} E\left((Y - f(X))^2\right)$	$\min_{f} \max_{g} E\Big(g(X) \; (Y - f(X))\Big)$
online	low regret	low macau

- probability: Least squares $\iff$ normal equations
- statistics: Least squares $\iff$ normal equations

Macau is a "normal equation"

$E(Y \mid X)$	Least squares	Normal equations
Statistics	$\min_{\beta} \sum (Y_i - \beta \cdot X_i)^2$	$\min_{\beta} \max_{\alpha} \sum \alpha \cdot X_i \ (Y_i - \beta \cdot X_i)$
Probability	$\min_{f} E\left((Y - f(X))^2\right)$	$\min_{f} \max_{g} E\Big(g(X) \ (Y - f(X))\Big)$
online	low regret	low macau

Take Aways

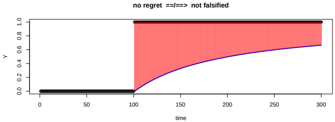
on-line low regret $\not\Leftrightarrow$ *on-line low macau*

low regret $\not\Leftrightarrow$ low macau

No regret $\not\Rightarrow$ not falsified

t	1	2	3	4	$\cdots$	T-1	T	T+1	T+2	T+3	$\cdots$	3T
Y_t	0	0	0	0	$\cdots$	0	1	1	1	1	$\cdots$	1
X_t	1	1	1	1	$\cdots$	1	1	1	1	1	$\cdots$	1
$\hat{Y}_t$	0	0	0	0	$\cdots$	0	0	$\frac{1}{T}$	$\frac{2}{T+1}$	$\frac{3}{T+2}$	$\cdots$	$\frac{2}{3}$

How about a bet?



Not falsified $\not\Rightarrow$ no regret

t	1	2	3	4	$\cdots$	T	T+1	$\cdots$
Y_t	0	1	0	1	$\cdots$	0	1	$\cdots$
X_t	1	1	1	1	$\cdots$	1	1	$\cdots$
$\hat{Y}_t$	.6	.4	.6	.4	$\cdots$	.6	.4	$\cdots$

- Macau is zero
- Regret is $T/9$
- So: low macau $\not\Rightarrow$ low regret

low regret $\not\Leftrightarrow$ low macau

No regret $\not\Rightarrow$ not falsified

t	1	2	3	4	$\cdots$	T-1	T	T+1	T+2	T+3	$\cdots$	3T
Y_t	0	0	0	0	$\cdots$	0	1	1	1	1	$\cdots$	1
X_t	1	1	1	1	$\cdots$	1	1	1	1	1	$\cdots$	1
$\hat{Y}_t$	0	0	0	0	$\cdots$	0	0	$\frac{1}{T}$	$\frac{2}{T+1}$	$\frac{3}{T+2}$	$\cdots$	$\frac{2}{3}$

How about a bet?

Not falsified $\not\Rightarrow$ no regret

t	1	2	3	4	$\cdots$	T	T+1	$\cdots$
Y_t	0	1	0	1	$\cdots$	0	1	$\cdots$
X_t	1	1	1	1	$\cdots$	1	1	$\cdots$
$\hat{Y}_t$	.6	.4	.6	.4	$\cdots$	.6	.4	$\cdots$

- Macau is zero
- Regret is $T/9$
- So: low macau $\not\Rightarrow$ low regret

(Skipping these proofs)

Economic forecasting for decision making

- Action A makes X dollars, action B makes Y dollars
  - We want forecasts that are close to X and Y
  - We want to be close on average
  - We will use least squares to estimate X and Y
- But, we want to take actions
- Will good estimates of X and Y lead to good decisions about A vs B?

Contextual Bandits

Some notation:

$$
\begin{aligned}
a &= \text{action taken} \in \Re^k \text{(eg inventory levels)} \\
X_t &= \text{Context at time } t \\
a_t^* &= \text{best action at time } t \\
r_t(a) &= \text{Reward at time } t \text{ playing } a \\
V_t^* &= \max_a E(r_t(a)|X_t) = E(r_t(a^*)|X_t) \\
\underline{q}_t(a) &\le E(r_t(a)|X_t) \le \overline{q}_t(a)
\end{aligned}
$$

Contextual Bandits

Some notation:

$$\begin{aligned}
a &= \text{action taken} \in \Re^k \text{(eg inventory levels)} \\
X_t &= \text{Context at time } t \\
a_t^* &= \text{best action at time } t \\
r_t(a) &= \text{Reward at time } t \text{ playing } a \\
V_t^* &= \max_a E(r_t(a)|X_t) = E(r_t(a^*)|X_t) \\
\underline{q}_t(a) &\leq E(r_t(a)|X_t) \ \leq \ \overline{q}_t(a)
\end{aligned}$$

What are good falsifiable claims about a^*?

Contextual Bandits

Some notation:

$$
\begin{aligned}
a &= \text{action taken} \in \Re^k \text{(eg inventory levels)} \\
X_t &= \text{Context at time } t \\
a_t^* &= \text{best action at time } t \\
r_t(a) &= \text{Reward at time } t \text{ playing } a \\
V_t^* &= \max_a E(r_t(a)|X_t) = E(r_t(a^*)|X_t) \\
\underline{q}_t(a) &\leq E(r_t(a)|X_t) \ \leq \ \overline{q}_t(a)
\end{aligned}
$$

Too precise:

- "$\underline{q}_t(a) = \overline{q}_t(a)$"

Contextual Bandits

Some notation:

$$
\begin{aligned}
a &= \text{action taken} \in \Re^k \text{(eg inventory levels)} \\
X_t &= \text{Context at time } t \\
a_t^* &= \text{best action at time } t \\
r_t(a) &= \text{Reward at time } t \text{ playing } a \\
V_t^* &= \max_a E(r_t(a)|X_t) = E(r_t(a^*)|X_t) \\
\underline{q}_t(a) &\leq E(r_t(a)|X_t) \ \leq \ \overline{q}_t(a)
\end{aligned}
$$

Too loose:

- "Here is a_t^*."

Contextual Bandits

Some notation:

$$
\begin{array}{rcl}
a & = & \text{action taken} \in \Re^k \text{(eg inventory levels)} \\
X_t & = & \text{Context at time } t \\
a_t^* & = & \text{best action at time } t \\
r_t(a) & = & \text{Reward at time } t \text{ playing } a \\
V_t^* & = & \max_a E(r_t(a)|X_t) = E(r_t(a^*)|X_t) \\
\underline{q}_t(a) & \leq & E(r_t(a)|X_t) \;\leq\; \overline{q}_t(a)
\end{array}
$$

Just right:

- $\overline{q}_t(a) = V_t^* - q||a - a_t^*||^2$
- $\overline{q}_t(a) - \underline{q}_t(a) = \Delta||a - a_t^*||^2$

Why is low macau useful?

$$C(a) = \sum_{t=1}^{T} c_t(a) \qquad a^* \equiv \arg\min_a C(a)$$

- Supposed each $c_t(\cdot)$ is convex
- We could use $\nabla c_t()$ for SGD
- Say, on-line convex optimization:

$$\begin{aligned} \text{regret} &= \sum_{t=1}^{T} (c_t(\hat{a}_t) - c_t(a^*)) \\ \text{regret} &\leq \sum_{t=1}^{T} (\hat{a}_t - a^*) \cdot \nabla c_t(\hat{a}_t) \end{aligned}$$

Why is low macau useful?

$$C(a) = \sum_{t=1}^{T} c_t(a) \qquad a^* \equiv \arg\min_a C(a)$$

The regret is bounded by the gradient:

$$\text{regret} \quad \leq \quad \sum_{t=1}^{T} (\hat{a}_t - a^*) \cdot \nabla c_t(\hat{a}_t)$$

Why is low macau useful?

$$\boxed{C(a) = \sum_{t=1}^{T} c_t(a) \qquad a^* \equiv \arg\min_a C(a)}$$

The regret is bounded by the gradient:

$$\begin{aligned}
\text{regret} &\leq \sum_{t=1}^{T} (\hat{a}_t - a^*) \cdot \nabla c_t(\hat{a}_t) \\
&= \sum_{t=1}^{T} (\hat{a}_t - a^*) \cdot \left(\nabla c_t(\hat{a}_t) - \widehat{\nabla c_t}(\hat{a}_t) \right) + (\hat{a}_t - a^*) \cdot \widehat{\nabla c_t}(\hat{a}_t)
\end{aligned}$$

Why is low macau useful?

$$C(a) = \sum_{t=1}^{T} c_t(a) \qquad a^* \equiv \arg\min_a C(a)$$

The regret is bounded by the gradient:

$$\begin{aligned}
\text{regret} &\leq \sum_{t=1}^{T} (\hat{a}_t - a^*) \cdot \nabla c_t(\hat{a}_t) \\
&= \underbrace{\sum_{t=1}^{T} (\hat{a}_t - a^*) \cdot \left(\nabla c_t(\hat{a}_t) - \widehat{\nabla c_t}(\hat{a}_t) \right)}_{(\textit{macau!})} + (\hat{a}_t - a^*) \cdot \underbrace{\widehat{\nabla c_t}(\hat{a}_t)}_{(\textit{zero @ } \hat{a}_t)}
\end{aligned}$$

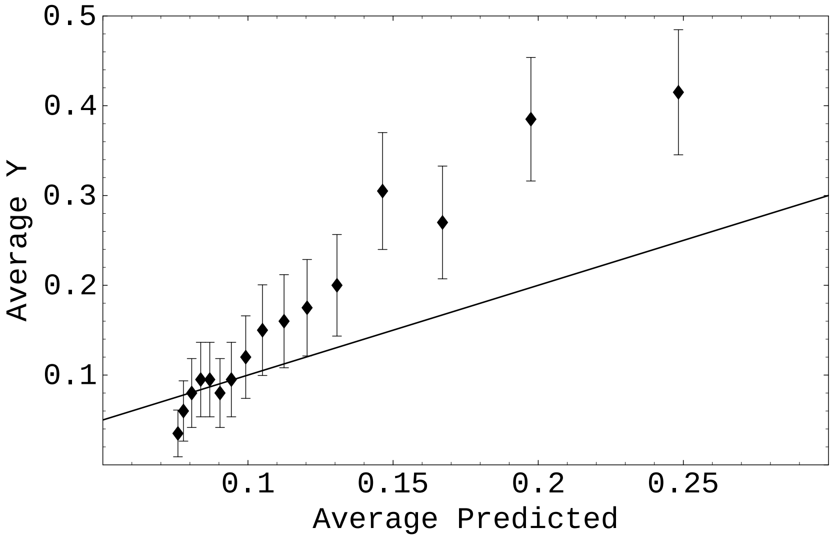
Why is low macau useful?

$$C(a) = \sum_{t=1}^{T} c_t(a) \qquad a^* \equiv \arg\min_a C(a)$$

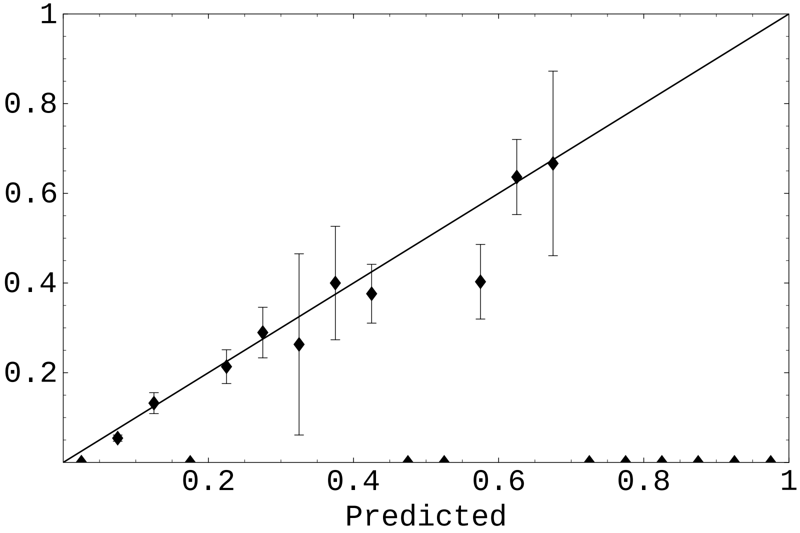
The regret is bounded by the gradient:

$$\begin{aligned}
\text{regret} &\leq \sum_{t=1}^{T} (\hat{a}_t - a^*) \cdot \nabla c_t(\hat{a}_t) \\
&= \sum_{t=1}^{T} (\hat{a}_t - a^*) \cdot \left(\nabla c_t(\hat{a}_t) - \widehat{\nabla c_t}(\hat{a}_t)\right) + (\hat{a}_t - a^*) \cdot \widehat{\nabla c_t}(\hat{a}_t) \\
\text{regret} &\leq \text{macau}
\end{aligned}$$

without crazy-calibration variable



Using the crazy-calibration variable



Calibration Theorem

Theorem ($\Longrightarrow$ F. and Kakade 2008, $\Longleftarrow$ new)

Let R be the quadratic regret of a forecast $\hat{Y}_t$ against a linear regression on X_t. Let M be the Macau of $\hat{Y}_t$ using linear functions of X_t to create falsifying bets. Then if $\hat{Y}_t = [X_t]_0$, we have $R = o(T)$ iff $M = o(T)$.

Calibration Theorem

Theorem ($\Longrightarrow$ F. and Kakade 2008, $\Longleftarrow$ new)

Let R be the quadratic regret of a forecast $\hat{Y}_t$ against a linear regression on X_t. Let M be the Macau of $\hat{Y}_t$ using linear functions of X_t to create falsifying bets. Then if $\hat{Y}_t = [X_t]_0$, we have $R = o(T)$ iff $M = o(T)$.

Note: Typically, $R = O(\log(T))$ iff $M = \tilde{O}(\sqrt{T})$ for the actual algorithms I know.

(*Sasha Rakhlin (an Amazon scholar!) and D. Foster seem to have a proof of something very close to this for IID.*)

Calibration Theorem

Theorem ($\Longrightarrow$ F. and Kakade 2008, $\Longleftarrow$ new)

Let R be the quadratic regret of a forecast $\hat{Y}_t$ against a linear regression on X_t. Let M be the Macau of $\hat{Y}_t$ using linear functions of X_t to create falsifying bets. Then if $\hat{Y}_t = [X_t]_0$, we have $R = o(T)$ iff $M = o(T)$.

Proof sketch: Consider the forecasts $w\hat{Y}_t + (1 - w)\beta \cdot X_t$ for the *best* β. Let $Q(w)$ be the total quadratic error of this forecast. Best β says, the minimum is at $Q(1)$. The following three are equivalent:

- $Q(0) \leq Q(1)$ (No regret condition)
- $Q'(0)$ is zero. (No macau condition)
- $Q(w) = 0$ for all w.

Recipe for good decisions

- List bets that you would make to show $\hat{a}_t$ is not optimial
- Convert these to regression variables
- Add the crazy-calibration variable
- Run a low regret least squares algorithm
- Make decision based on this forecast

What bets to place?

	Bet
convex	$[\hat{a}_t - a^*]_i$
experts	$e_{a*} - e_{\hat{a}_t}$
internal regret	$(e_a - e_b) I_{\hat{a}_t = b}$
bandits	$\frac{I_{a_t=a}}{P(a_t=a)} - \frac{I_{a_t=\hat{a}_t}}{P(a_t=\hat{a}_t)}$
contextual	$X_t \times \left(\frac{I_{a_t=a}}{P(a_t=a)} - \frac{I_{a_t=\hat{a}_t}}{P(a_t=\hat{a}_t)}\right)$
continuous	$(a_t - Mx_t)^2$
LQR	$(a_t - \sum_{i=1}^{\log T} M_i x_{t-i})^2$
reinforcement Learning	TD learn

What bets to place?

	Bet	dimension
convex	$[\hat{a}_t - a^*]_i$	$\in \Re^d$
experts	$e_{a*} - e_{\hat{a}_t}$	$\in \Re^k$
internal regret	$(e_a - e_b) I_{\hat{a}_t = b}$	$\in \Re^{k^2}$
bandits	$\frac{I_{a_t=a}}{P(a_t=a)} - \frac{I_{a_t=\hat{a}_t}}{P(a_t=\hat{a}_t)}$	$\in \Re^k$
contextual	$X_t \times \left(\frac{I_{a_t=a}}{P(a_t=a)} - \frac{I_{a_t=\hat{a}_t}}{P(a_t=\hat{a}_t)}\right)$	$\in \Re^{dk}$
continuous	$(a_t - Mx_t)^2$	$\in \Re^{dk}$
LQR	$(a_t - \sum_{i=1}^{\log T} M_i x_{t-i})^2$	$\in \Re^{dk \log(T)}$
reinforcement Learning	TD learn	

RL: Falsifiability value estimation

Theorem (Dicker 2019)

Least squares plus the calibration variable generates an estimate of the RL value function with low Macau.

Theorem (Dicker 2019)

A tweaked version of TD learning with 1/sqrt(T) rates generates an estimate of the RL value function with low Macau.

RL: Falsifiability value estimation

Theorem (Dicker 2019)

Least squares plus the calibration variable generates an estimate of the RL value function with low Macau.

Proof: Follows from F. and Kakade 2008.

Theorem (Dicker 2019)

A tweaked version of TD learning with 1/sqrt(T) rates generates an estimate of the RL value function with low Macau.

Proof: Similar to Dicker and F. 2018.

Conclusions

Take Aways

crazy-Calibration + low-regret $\iff$ *low-macau* $\implies$ *good decisions*

Conclusions

Take Aways

crazy-Calibration + low-regret $\iff$ *low-macau* $\implies$ *good decisions*

Thanks!

Appendix slides

Proofs by example:

- low Regret $\not\Rightarrow$ low Macau
- low Regret $\not\Leftarrow$ low Macau

Bets:

- Experts
- No Internal Regret
- Bandits, (scalar version), (exploration).
- Contextual Bandits
- Continuous action contextual Bandits
- Convex optimization, (one point), ($1/T$ with smooth)
- Reinforcement Learning
- LQR

Proofs: Regret $\not\Leftrightarrow$ macau

No regret $\not\Rightarrow$ not falsified

t	1	2	3	4	$\cdots$	T-1	T	T+1	T+2	T+3	$\cdots$	3T
Y_t	0	0	0	0	$\cdots$	0	1	1	1	1	$\cdots$	1
X_t	1	1	1	1	$\cdots$	1	1	1	1	1	$\cdots$	1
$\hat{Y}_t$	0	0	0	0	$\cdots$	0	0	$\frac{1}{T}$	$\frac{2}{T+1}$	$\frac{3}{T+2}$	$\cdots$	$\frac{2}{3}$

No regret $\not\Rightarrow$ not falsified

t	1	2	3	4	$\cdots$	T-1	T	T+1	T+2	T+3	$\cdots$	3T
Y_t	0	0	0	0	$\cdots$	0	1	1	1	1	$\cdots$	1
X_t	1	1	1	1	$\cdots$	1	1	1	1	1	$\cdots$	1
$\hat{Y}_t$	0	0	0	0	$\cdots$	0	0	$\frac{1}{T}$	$\frac{2}{T+1}$	$\frac{3}{T+2}$	$\cdots$	$\frac{2}{3}$

On-line least squares suffers no-regret:

- β_t minimizes $\sum_{i=1}^{t}(Y_i - \beta \cdot X_t)^2$
- $\hat{Y}_t = \beta_{t-1} \cdot X_t$
- Total error: $\sum(Y_t - \hat{Y}_t)^2 = \min_\beta \sum(Y_t - \beta X_t)^2 + 4/9$
- In general, on-line least squares has $\log(T)$ total regret
- In this case, it actually wins by about $O(1)$.

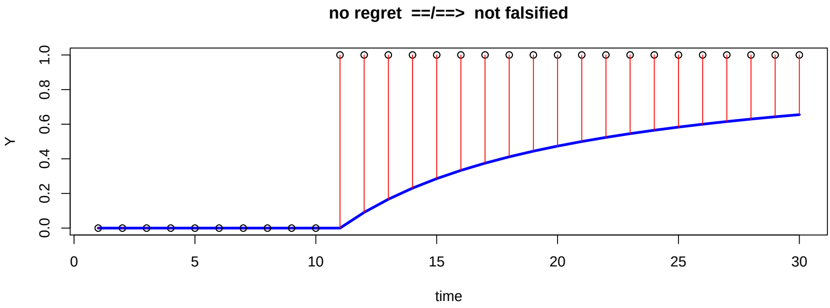
No regret $\not\Rightarrow$ not falsified

t	1	2	3	4	$\cdots$	T-1	T	T+1	T+2	T+3	$\cdots$	3T
Y_t	0	0	0	0	$\cdots$	0	1	1	1	1	$\cdots$	1
X_t	1	1	1	1	$\cdots$	1	1	1	1	1	$\cdots$	1
$\hat{Y}_t$	0	0	0	0	$\cdots$	0	0	$\frac{1}{T}$	$\frac{2}{T+1}$	$\frac{3}{T+2}$	$\cdots$	$\frac{2}{3}$

How about a bet?

t	1	2	3	4	$\cdots$	T-1	T	T+1	T+2	T+3	$\cdots$	3T
Y_t	0	0	0	0	$\cdots$	0	1	1	1	1	$\cdots$	1
X_t	1	1	1	1	$\cdots$	1	1	1	1	1	$\cdots$	1
$\hat{Y}_t$	0	0	0	0	$\cdots$	0	0	$\frac{1}{T}$	$\frac{2}{T+1}$	$\frac{3}{T+2}$	$\cdots$	$\frac{2}{3}$

How about a bet?



No regret $\not\Rightarrow$ not falsified

t	1	2	3	4	$\cdots$	T-1	T	T+1	T+2	T+3	$\cdots$	3T
Y_t	0	0	0	0	$\cdots$	0	1	1	1	1	$\cdots$	1
X_t	1	1	1	1	$\cdots$	1	1	1	1	1	$\cdots$	1
$\hat{Y}_t$	0	0	0	0	$\cdots$	0	0	$\frac{1}{T}$	$\frac{2}{T+1}$	$\frac{3}{T+2}$	$\cdots$	$\frac{2}{3}$

How about a bet?

- $Y_t > \hat{Y}_t$, so that is a safe bet!
- Construct this bet only using X_t

$$\sum_{i=1}^{T} X_t(Y - \hat{Y}_t) \approx T\frac{\log_e(3)}{2}$$

- Betting loses $\Omega(T)$

No regret $\not\Rightarrow$ not falsified

t	1	2	3	4	$\cdots$	T-1	T	T+1	T+2	T+3	$\cdots$	3T
Y_t	0	0	0	0	$\cdots$	0	1	1	1	1	$\cdots$	1
X_t	1	1	1	1	$\cdots$	1	1	1	1	1	$\cdots$	1
$\hat{Y}_t$	0	0	0	0	$\cdots$	0	0	$\frac{1}{T}$	$\frac{2}{T+1}$	$\frac{3}{T+2}$	$\cdots$	$\frac{2}{3}$

- Regret is $O(1)$
- Macau is $T/2$
- So: low regret $\not\Rightarrow$ low macau

Not falsified $\not\Rightarrow$ no regret

t	1	2	3	4	$\cdots$	T	T+1	$\cdots$
Y_t	0	1	0	1	$\cdots$	0	1	$\cdots$
X_t	1	1	1	1	$\cdots$	1	1	$\cdots$
$\hat{Y}_t$	.6	.4	.6	.4	$\cdots$	.6	.4	$\cdots$

Not falsified $\not\Rightarrow$ no regret

t	1	2	3	4	$\cdots$	T	T+1	$\cdots$
Y_t	0	1	0	1	$\cdots$	0	1	$\cdots$
X_t	1	1	1	1	$\cdots$	1	1	$\cdots$
$\hat{Y}_t$	.6	.4	.6	.4	$\cdots$	.6	.4	$\cdots$

Betting

- No bet based on X_t will win anything
- In other words,

$$\max_{\alpha} \sum_{i=1}^{T} \alpha \cdot X_t \,(Y - \hat{Y}_t) = 0$$

- This forecast is not falsified using linear functions of X_t

Not falsified $\not\Rightarrow$ no regret

t	1	2	3	4	$\cdots$	T	T+1	$\cdots$
Y_t	0	1	0	1	$\cdots$	0	1	$\cdots$
X_t	1	1	1	1	$\cdots$	1	1	$\cdots$
$\hat{Y}_t$	.6	.4	.6	.4	$\cdots$	.6	.4	$\cdots$

But, a better forecast exists

- $\sum(Y_t - \hat{Y}_t)^2 = .36T$
- $\min_\beta(Y_t - \beta X_t)^2 = .25T$
- Regret is $.11T$
- So, regret is $\Omega(T)$

Not falsified $\not\Longrightarrow$ no regret

t	1	2	3	4	$\cdots$	T	T+1	$\cdots$
Y_t	0	1	0	1	$\cdots$	0	1	$\cdots$
X_t	1	1	1	1	$\cdots$	1	1	$\cdots$
$\hat{Y}_t$	.6	.4	.6	.4	$\cdots$	.6	.4	$\cdots$

- Macau is zero
- Regret is $T/9$
- So: low macau $\not\Longrightarrow$ low regret

Bet: Convex optimization (with gradients)

In the convex optimization problem, we observe a sequence of convex functions $c_t(\cdot)$. Or goal is to figure out a action $\hat{x}_t^*$ to take at each point in time t to minimize $\sum_t c_t(\hat{x}_t^*)$.

- Forecast: Gradient of c_t at each point in time t $(g_t(x) \equiv \nabla c_t(x))$
- Strategy: Pick a $\hat{x}_t^*$ such that $\hat{g}_t(\hat{x}_t^*) = 0$.
- Worry: "The real optimum x^* would generate better performance."
- Macau bets: $[x^* - \hat{x}_t^*]_i$ bet against $[g_t]_i - [\hat{g}_t]_i$

$$\text{Macau}_i = \sum_{t=1}^{T} [x^* - \hat{x}_t^*]_i([g_t]_i - [\hat{g}_t]_i)$$

$$\boxed{\text{Bet:} \quad [x^* - \hat{x}_t^*]_i}$$

Bet: Convex optimization (no gradients)

In the convex optimization problem, we observe a sequence of convex functions $c_t(\cdot)$. Or goal is to figure out a action $\hat{x}_t^*$ to take at each point in time t to minimize $\sum_t c_t(\hat{x}_t^*)$.

- Forecast: $c_t(x)$ at points near $\hat{x}_t^*$, for example $x_t - \hat{x}_t^* \sim N(0, \sigma^2 I)$
- Strategy: Pick a $\hat{x}_t^*$ to minimize $\hat{c}(\cdot)$
- Worry: "The real optimum x^* would generate better performance."
- Macau bets: $(x^* - \hat{x}_t^*) \cdot (x_t - \hat{x}_t^*)$

$$\text{Macau} = \sum_{t=1}^{T} (x^* - \hat{x}_t^*) \cdot (x_t - \hat{x}_t^*) c(x)$$

$$\boxed{\text{Bet:} \quad [x^* - \hat{x}_t^*]_i}$$

Bet: Optimizing continuous convex functions (with gradient)

Also assume each c_t is smooth, say $c_t \in \mathcal{C}_2$. We'll keep all else the same.

- We can use the macau to look at bets for how for $\hat{\beta}$ is from the best after the fact β
- Thus we know the optimum point is close to the best hind sight deciosion point (say $1/\sqrt{T}$ accuracy)
- This means the error in payoff space is $1/T$
- So it doesn't require a new algorithm or even new features

In the experts problem, we observe the payoff of k different experts. Our goal is to generate as much value as the best expert.

- Forecast: one value for each arm ($Y_t \in \Re^k$, so $\hat{Y}_t \in \Re^k$ also)
- Strategy: Pick arm with highest forecast ($\hat{a}_t = \arg\max_i [\hat{Y}_t]_i$)
- Worry: "Always playing arm b would generate more"
- Macau bet: $e_b = [0, 0, 0, \ldots, 1, \ldots, 0]'$

$$\text{Macau} = \max_{b \in \{1,\ldots,k\}} \sum_t (e_b - e_{\hat{a}_t}) \cdot (Y_t - \hat{Y}_t)$$

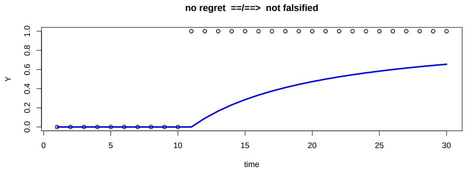
$$\boxed{\text{Bet:} \quad e_b - e_{\hat{a}_t}}$$

Example: Experts need macau

{**dpf:** *I'm not sure this actually works. Needs work!*}

Will least squares do well enough? Consider two arms

- First arm gets a known sequence of payoffs (so forecast is trivial, $r_t = V_t$
- Second arm forecast has no-regret but lots of Macau:



- We might play first arm–but we could bet the second arm would be better and win!

Bet: No Internal Regret

In the no-internal regret problem, we observe the payoff of k different experts. Our goal is to avoid feeling regret about possibly switching one of our actions to some other action.

- Forecast: one value for each expert ($Y_t \in \Re^k$, so $\hat{Y}_t \in \Re^k$ also)
- Strategy: Pick arm with highest forecast ($\hat{a}_t = \arg\max_i [\hat{Y}_t]_i$)
- Worry: "Playing c when we previously played b would have been better ($R^{c \to b} > 0$)."
- Macau bet:

$$\left(I_{\hat{a}_t = c}(e_b - e_c)\right) \cdot (Y_t - \hat{Y}_t)$$

$$\boxed{\text{Bet on } c \to b: \quad I_{\hat{a}_t = c}(e_b - e_c)}$$

Bet: Bandits (vector structure)

We only see outcomes on the one of k arms we pull.

- Forecast: Each arms payoff: $[Y_t]_i = \frac{r_t I_{a_t=i}}{p(a_t=i)}$, so $\hat{Y}_t \in \Re^k$.
- Strategy: Pick arm with highest forecast ($\hat{a}_t = \arg\max_i [\hat{Y}_t]_i$) with some exploration also.
- Worry: Always playing b might have been better.
- Macau bet:

$$(e_b - e_{\hat{a}_t}) \cdot (Y_t - \hat{Y}_t)$$

$$\boxed{\text{Bet on } b\text{:} \quad (e_b - e_{\hat{a}_t})}$$

Bet: Bandits (scalar version)

Play $a_t \in \{1, \ldots, k\}$ and only see its outcome.

- Forecast: the arm actually played: $Y_t = \frac{r_t(a_t)}{p_t(a_t)}$, so $\hat{Y}_t(a_t) \in \Re$.
- Strategy: Pick arm with highest forecast ($\hat{a}_t = \arg\max_i \hat{Y}_t(i)$) with some exploration also.
- Worry: Always playing b might have been better.
- Macau bet:

$$\left(\frac{I_{a_t=b}}{p_t(b)} - \frac{I_{a_t=\hat{a}_t}}{p_t(\hat{a}_t)} \right) (Y_t - \hat{Y}_t)$$

$$\boxed{\text{Bet on } b: \quad \frac{I_{a_t=b}}{p_t(b)} - \frac{I_{a_t=\hat{a}_t}}{p_t(\hat{a}_t)}}$$

Bandits exploration

- Macau keeps the mean correct
- We would also high probability statements
- So, we need $p_t(b)$ to not be too small
  - Easy math: $p_t(b) \geq t^{-1/3}$, but not optimal rates of convergence
  - Giving up a log: $p_t(b) \geq t^{-1/2}$. But, as $\hat{Y}_t(b)$ gets closer to $\hat{Y}_t(\hat{a}_t)$ we sample more often. On a log scale, this means we need $k \log(T)$ features.
  - Note: the fixed point solution will generate some randomization above and beyond that given by the lower bounds
- Similar behavior to UCB, but a different philosophy to justify it.

Bet: Contextual Bandits (vector version)

First we observe $X_t \in \Re^d$, then we play an arm a_t and observe its outcome (vector version: $[Y_t]_i = \frac{r_t I_{a_t=i}}{p(a_t=i)}$):

- Forecast: $\hat{Y}_t = X_t \beta_{t-1}$, with $\beta \in \Re^{d \times k}$ $\hat{Y}_t \in \Re^k$.
- Strategy: Pick arm with highest forecast ($\hat{a}_t = \arg\max_i [\hat{Y}_t]_i$).
- Worry: Using some other β^* might be better.
- Naive Macau bet ($\hat{a}_t \to b$):

$$(I_{X_t(\beta^*_b - \beta^*_{\hat{a}_t})>0} - e_{\hat{a}_t}) \cdot (Y_t - \hat{Y}_t)$$

- These are hard to put in a linear space. But, given the low dimension (VC=$d+2$) hope spring eternal.

$$\boxed{\text{Bet on } b: \quad (e_b - e_{\hat{a}_t})}$$

Bet: Continuous action for contextual Bandits

First we observe $X_t \in \Re^d$, then we play an action $a_t \in \mathcal{A} \subset \Re^k$ and observe its outcome. (We'll actually penalize a quadratically and hence avoid the set $\mathcal{A}$.)

- Forecast: $\hat{Y}_t(a) = X_t^\top \beta_{t-1} a - a^\top a/2$, with $\beta \in \Re^{d\times k}$ and $\hat{Y}_t(a) \in \Re^k$.
- Strategy: Pick "best" action: $\hat{a}_t = \arg\max_{a\in\mathcal{A}} \hat{Y}_t(a) = X_t^\top \hat{\beta}_{t-1}$.
- Worry: Using some other β^* might be better.
- Naive Macau bet ($\hat{a}_t \to (1-\epsilon)\hat{a}_t + \epsilon X_t^\top \beta^*$):

$$(X_t^\top \beta^* - X_t^\top \hat{\beta}_t^*) \cdot (a_t - \hat{a}_t)(Y_t(a_t) - \hat{Y}_t(a_t))$$

$$\boxed{\text{Bet in direction } X_t^\top \beta^*: \quad (\textit{fillin})}$$

The rest isn't done yet!

The RL value function:

$$V_t^* = \max_\pi E\left(\sum_{i=t}^{\infty} \gamma^{i-t} r_i(a_i^\pi) \middle| \mathcal{F}_t\right)$$

(γ is discount rate.) Recursively:

$$V_t^* = E\left(r_t(a) + \gamma V_{t+1}^* \middle| \mathcal{F}_t\right)$$

Reinforcement Learning

The RL value function:

$$V_t^* = \max_{\pi} E\left(\sum_{i=t}^{\infty} \gamma^{i-t} r_i(a_i^{\pi}) \middle| \mathcal{F}_t\right)$$

(γ is discount rate.) Recursively:

$$V_t^* = E\left(r_t(a) + \gamma V_{t+1}^* \middle| \mathcal{F}_t\right)$$

V^* is a Y-variable and an X-variable!

Linear Quadratic Regulator: