Talk 3:
Macau: Betting against Aaditya

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 definition of error
  - you falsify a forecast by betting against it
  - The amount it loses 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 definition of error
  - you falsify a forecast by betting against it
  - The amount it loses 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:

$$\text{expected winnings} = E\left(B\left(Y - \hat{Y}\right)\right)$$

  - $(Y - \hat{Y})$ 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:

$$\text{expected winnings} = E\left(B\left(Y - \hat{Y}\right)\right)$$

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

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

(*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:

$$\text{expected winnings} = E\left(B\left(Y - \hat{Y}\right)\right)$$

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

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

(*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:

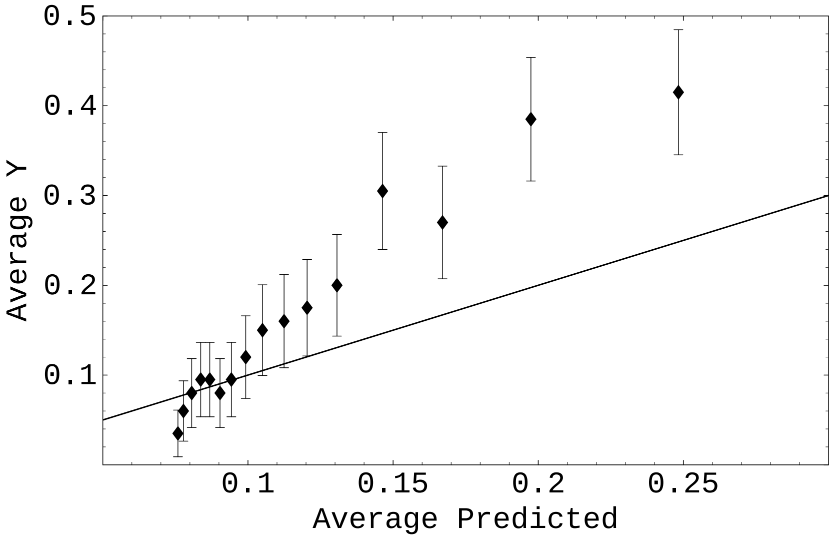
$$\text{expected winnings} = E\left(B\left(Y - \hat{Y}\right)\right)$$

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

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

(*mini-max*)

On to calibration



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

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}		
	$X_{t+1,1}$	$X_{t+1,2}$	$X_{t+1,3}$	$X_{t+1,4}$	$\hat{\beta}_t$	$\hat{Y}_{t+1} = \hat{\beta}_t' X_{t+1}$

From a new X_{t+1} 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

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

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	$\hat Y_1 = 0$
Y_2	X_{21}	X_{22}	X_{23}	X_{24}	$\hat\beta_1$	$\hat Y_2 = \hat\beta_1' X_2$
Y_3	X_{31}	X_{32}	X_{33}	X_{34}	$\hat\beta_2$	$\hat Y_3 = \hat\beta_2' X_3$
Y_4	X_{41}	X_{42}	X_{43}	X_{44}	$\hat\beta_3$	$\hat Y_4 = \hat\beta_3' X_4$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_t	X_{t1}	X_{t2}	X_{t3}	X_{t4}	$\hat\beta_{t-1}$	$\hat Y_t = \hat\beta_{t-1}' X_t$

Theorem (F 1991, Forster 1999,F and Hart (soon))

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

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

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

Works no matter what the X's are.

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

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

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

Adding the crazy calibration variable generates low macau:

$$(\forall i) \quad \sum_{t=1}^{T} X_{t,i}(Y_t - \hat{Y}_t) = O(\sqrt{T \log(T)})$$

Macau as the "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 as the "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 as the “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	$\min_f E((Y - \underbrace{f(X)}_{\text{aka } E(Y\|X)})^2)$	$(\forall g)\ E(g(X)\ (Y - f(X))) = 0$

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 as the “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	$\min_{f} E((Y - \underbrace{f(X)}_{aka\ E(Y\|X)})^2)$	$\min_{f} \max_{g} E\Big(g(X) \; (Y - f(X))\Big)$

Macau as the “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	$\min_f E((Y - \underbrace{f(X)}_{\text{aka } E(Y\|X)})^2)$	$\min_f \max_g E\Big(g(X) \; (Y - f(X))\Big)$
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 as the "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	$\min_{f} E((Y - \underbrace{f(X)}_{\text{aka } E(Y\|X)})^2)$	$\min_{f} \max_{g} E\Big(g(X) \; (Y - f(X))\Big)$
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 as the “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	$\min_f E((Y - \underbrace{f(X)}_{\text{aka } E(Y\|X)})^2)$	$\min_f \max_g E\Big(g(X)\ (Y - f(X))\Big)$
online	low regret	low macau

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

Macau as the "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	$\min_f E((Y - \underbrace{f(X)}_{\text{aka } E(Y\|X)})^2)$	$\min_f \max_g E\Big(g(X) \ (Y - f(X))\Big)$
online	low regret	low macau

Take Aways

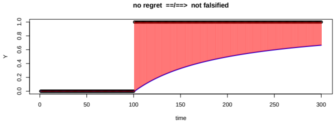
on-line low regret $\not\Leftrightarrow$ *on-line low macau*

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)

Short break

ASIDE: 4rd proof of calibration

- Yesterday morning we proved existance of calibration by a flow condition and using any bandit algorithm
- Yesterday afternoon we proved calibration by the minimax theorem.
- Yesterday we also proved calibration by calibeating oneself
- Today we prove it via least squares (So we'll have to prove on-line least squares first.)

Goal:

$$\sum_{t=1}^{T}(Y_t - \hat{\beta}_{t-1}^{\top} X_t)^2 \leq \sum_{t=1}^{T}(Y_t - \hat{\beta}_{T}^{\top} X_t)^2 + o(T)$$

$$\min_{\beta} \sum_{t=1}^{T} (Y_t - \beta^\top X_t)^2 \quad = \quad \sum_{t=1}^{T} (Y_t - \hat{\beta}_T^\top X_t)^2$$

$$
\begin{aligned}
\min_{\beta} \sum_{t=1}^{T} (Y_t - \beta^\top X_t)^2 &= \sum_{t=1}^{T} (Y_t - \hat{\beta}_T^\top X_t)^2 \\
&= \sum_{t=1}^{T-1} (Y_t - \hat{\beta}_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2
\end{aligned}
$$

$$
\begin{aligned}
\min_{\beta} \sum_{t=1}^{T} (Y_t - \beta^\top X_t)^2 &= \sum_{t=1}^{T} (Y_t - \hat{\beta}_T^\top X_t)^2 \\
&= \sum_{t=1}^{T-1} (Y_t - \hat{\beta}_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&\geq \min_{\beta} \sum_{t=1}^{T-1} (Y_t - \beta_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2
\end{aligned}
$$

$$
\begin{aligned}
\min_{\beta} \sum_{t=1}^{T} (Y_t - \beta^\top X_t)^2 &= \sum_{t=1}^{T} (Y_t - \hat{\beta}_T^\top X_t)^2 \\
&= \sum_{t=1}^{T-1} (Y_t - \hat{\beta}_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&\geq \min_{\beta} \sum_{t=1}^{T-1} (Y_t - \beta_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&= \sum_{t=1}^{T-1} (Y_t - \hat{\beta}_{T-1}^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2
\end{aligned}
$$

$$
\begin{aligned}
\min_{\beta} \sum_{t=1}^{T} (Y_t - \beta^\top X_t)^2 &= \sum_{t=1}^{T} (Y_t - \hat{\beta}_T^\top X_t)^2 \\
&= \sum_{t=1}^{T-1} (Y_t - \hat{\beta}_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&\geq \min_{\beta} \sum_{t=1}^{T-1} (Y_t - \beta_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&= \sum_{t=1}^{T-1} (Y_t - \hat{\beta}_{T-1}^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&\vdots
\end{aligned}
$$

$$
\begin{aligned}
\min_{\beta} \sum_{t=1}^{T} (Y_t - \beta^\top X_t)^2 &= \sum_{t=1}^{T} (Y_t - \hat{\beta}_T^\top X_t)^2 \\
&= \sum_{t=1}^{T-1} (Y_t - \hat{\beta}_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&\geq \min_{\beta} \sum_{t=1}^{T-1} (Y_t - \beta_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&= \sum_{t=1}^{T-1} (Y_t - \hat{\beta}_{T-1}^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&\vdots \\
&\geq \sum_{t=1}^{T} (Y_t - \hat{\beta}_t^\top X_t)^2
\end{aligned}
$$

$$
\begin{aligned}
\min_{\beta} \sum_{t=1}^{T} (Y_t - \beta^\top X_t)^2 &= \sum_{t=1}^{T} (Y_t - \hat{\beta}_T^\top X_t)^2 \\
&= \sum_{t=1}^{T-1} (Y_t - \hat{\beta}_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&\geq \min_{\beta} \sum_{t=1}^{T-1} (Y_t - \beta_T^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&= \sum_{t=1}^{T-1} (Y_t - \hat{\beta}_{T-1}^\top X_t)^2 + (Y_T - \hat{\beta}_T^\top X_T)^2 \\
&\vdots \\
&\geq \sum_{t=1}^{T} (Y_t - \hat{\beta}_t^\top X_t)^2 \approx \sum_{t=1}^{T} (Y_t - \hat{\beta}_{t-1}^\top X_t)^2
\end{aligned}
$$

It is all in the last term

Win using: $\hat{\beta}_t = \min_\beta \sum_{t=1}^{T-1}(Y - \beta^\top X_t)^2 + (Y_T - \beta^\top X_t)^2$

Minimax: $\tilde{\beta}_t = \min_\beta \sum_{t=1}^{T-1}(Y - \beta^\top X_t)^2 + (.5 - \beta^\top X_t)^2$
(called a forward model)

traditional: $\hat{\beta}_{t-1} = \min_\beta \sum_{t=1}^{T-1}(Y - \beta^\top X_t)^2 + (\hat{\beta}_{t-1} X_t - \beta^\top X_t)^2$

New: $\hat{\beta}_t = \min_\beta \sum_{t=1}^{T-1}(Y - \beta^\top X_t)^2 + (\tilde{Y}_{t-1} - \beta^\top X_t)^2$
where $\tilde{Y}$ calibeats $\hat{y}$.

It is all in the last term

Win using: $\hat{\beta}_t = \min_\beta \sum_{t=1}^{T-1} (Y - \beta^\top X_t)^2 + (Y_T - \beta^\top X_t)^2$

- Regret ≤ 0

Minimax: $\tilde{\beta}_t = \min_\beta \sum_{t=1}^{T-1} (Y - \beta^\top X_t)^2 + (.5 - \beta^\top X_t)^2$

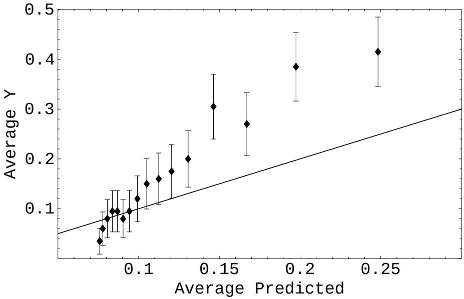
- Regret $\leq \frac{1}{4} d \log(T)$

traditional: $\hat{\beta}_{t-1} = \min_\beta \sum_{t=1}^{T-1} (Y - \beta^\top X_t)^2 + (\hat{\beta}_{t-1} X_t - \beta^\top X_t)^2$

- Regret $\leq d \log(T)$

New: $\hat{\beta}_t = \min_\beta \sum_{t=1}^{T-1} (Y - \beta^\top X_t)^2 + (\tilde{Y}_{t-1} - \beta^\top X_t)^2$
where $\tilde{Y}$ calibeats $\hat{y}$.

- Regret $\leq \overline{\tilde{\sigma}}^2 d \log(T)$
- Where $\tilde{\sigma}_t^2 = \tilde{Y}_t(1 - \tilde{Y}_t)$ and $\overline{\tilde{\sigma}}^2 = (1/T) \sum \tilde{\sigma}_t^2$.



If you saw this pattern in a regression, you might try fitting a polynomial to this variable. That is exactly what we will do!

Calibration via regression

Goal: $E(Y - \hat{Y} | \hat{Y} = c) = 0$

- Polynomial regression in $\hat{Y}$
- Add Regression variables: $\hat{Y}, \hat{Y}^2, \hat{Y}^3, \ldots, \hat{Y}^p$
- Bob Stine like $p = 5$, why? Looks pretty.

Calibration via regression

Goal: $E(Y - \hat{Y} | \hat{Y} = c) = 0$

- Polynomial regression in $\hat{Y}$
- Add Regression variables: $\hat{Y}, \hat{Y}^2, \hat{Y}^3, \ldots, \hat{Y}^p$
- Bob Stine like $p = 5$, why? Looks pretty.
- Computing $\hat{Y}$ now entails finding a full fixed point rather than just a linear equation.
- Equivalently it is finding a zero of a polynomial
- Leads to a weakly calibrated forecast
- Random rounding leads to clasic calibration

Back to Macau

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*?

Some notation:

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

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:
"Here are two bounding functions $\underline{q}$ and $\overline{q}$:

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

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:

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

Just right:
"Here is a target V^* and approximating quadratics around a^*:

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

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
- Goal: play a to minimize $C(a)$
- Eg: We could use SGD on $\nabla c_t()$
- called "on-line convex optimization"
- regret definition for this setting:

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

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:

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

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:

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

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:

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

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:

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

Calibration Theorem

Theorem ($\implies$ F. and Kakade 2008, $\impliedby$ 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 we have the crazy calibration variable (i.e. $[X_t]_0 = \hat{Y}_t$), then

$$R = o(T) \quad \textit{iff} \quad M = o(T).$$

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 we have the crazy calibration variable (i.e. $[X_t]_0 = \hat{Y}_t$), then

$$R = o(T) \quad \textit{iff} \quad M = o(T).$$

Proof sketch: Consider the forecasts $(1-w)\hat{Y}_t + w\alpha \cdot X_t$ for the *any* α. Let $Q(w)$ be the total quadratic error of this family of forecast. The following are equivalent:

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

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 we have the crazy calibration variable (i.e. $[X_t]_0 = \hat{Y}_t$), then

$$R = o(T) \quad \textit{iff} \quad M = o(T).$$

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

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

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.

- Current favorite paper: Foster and Rakhlin (2021), “Beyond UCB: Optimal and Efficient Contextual Bandits with Regression Oracles”
- Rakhlin and I have worked on calibration, optimization and contextual bandits other topics over the years

- Current favorite paper: Foster and Rakhlin (2021), “Beyond UCB: Optimal and Efficient Contextual Bandits with Regression Oracles”
- Rakhlin and I have worked on calibration, optimization and contextual bandits other topics over the years
- It isn’t by me–but by Dylan Foster

- They assume the model is true (so not individual sequence)
- Under this assumption the following algorithm does enough exploration:
  - Compute the expected value of each action using least squares
  - Pick the best action
  - Every now and then pick some other action:
    - But, make sure you don't expect to pay very much
    - Probability = ϵ/gap works well!
    - Called inverse gap weighting

- They assume the model is true (so not individual sequence)
- Under this assumption the following algorithm does enough exploration:
  - Compute the expected value of each action using least squares
  - Pick the best action
  - Every now and then pick some other action:
    - But, make sure you don't expect to pay very much
    - Probability = ϵ/gap works well!
    - Called inverse gap weighting
- $O(\sqrt{T})$ regret
- Rakhlin and I have extended it to work for:
  - Search (additive model)
  - Selecting items to sell (submodular)

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!

References

Note the three different "Fosters":

- Dean Foster (1991) "Prediction in the worst case."
- — and S. Kakade "Deterministic calibration and Nash." (Introduces most of the mathematics behind Macau.)
- — and S. Hart (2021) Easier version than above of many of the ideas of Macau.
- Dylan Foster and Sasha Rakhlin (2021) SquareCB paper. (Assumes IID data to get results much stronger than I have here. By far the best contextual bandit paper out there at the moment.)
- J. Forster (1999) "On Relative Loss Bounds in Generalized Linear Regression."

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	$(a_t - M x_t)^2$
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	

Appendix slides

Proofs by example:

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

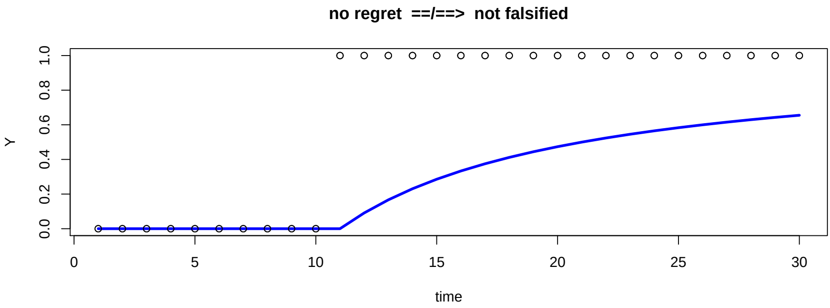
Bets:

- 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

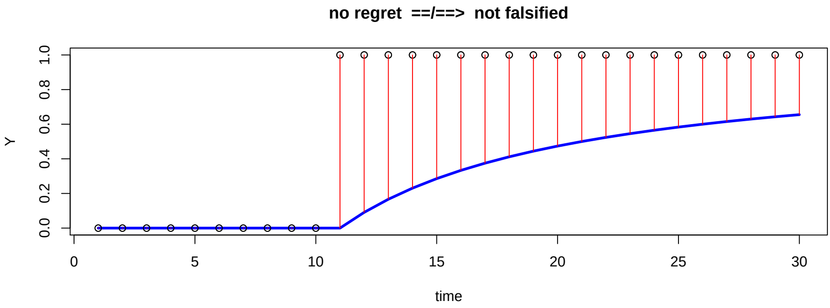
t	1	2	3	4	$\cdots$	T-1	T	T+1	T+2	T+3	$\cdots$	3T
Y_t	0	0	0	0	$\cdots$	0	1	1	1	1	$\cdots$	1
X_t	1	1	1	1	$\cdots$	1	1	1	1	1	$\cdots$	1
$\hat{Y}_t$	0	0	0	0	$\cdots$	0	0	$\frac{1}{T}$	$\frac{2}{T+1}$	$\frac{3}{T+2}$	$\cdots$	$\frac{2}{3}$

How about a bet?

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?



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 $\not\Rightarrow$ 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

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 $[g_t]_i - [\hat{g}_t]_i$

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

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

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: $(x^* - \hat{x}_t^*) \cdot (x_t - \hat{x}_t^*)$

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

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

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

Bet: Experts

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: $e_b = [0, 0, 0, \ldots, 1, \ldots, 0]'$

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

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

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:

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

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

The rest isn't done yet!

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:

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

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

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:

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

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

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.

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

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 ($\hat{a}_t \to (1-\epsilon)\hat{a}_t + \epsilon X_t^\top \beta^*$):

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

Bet in direction $X_t^\top \beta^*$: (*fillin*)

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!