On Directed Information and Gambling

Haim Permuter
Stanford

Young-Han Kim
UCSD

Tsachy Weissman
Stanford/ Technion

ISIT
July 2008

The horse race gambling problem

The horse race gambling problem

- m horses

The horse race gambling problem

- m horses
- X_i denotes the winning horse at time i. $\mathcal{X} = \{1, 2, .., m\}$

The horse race gambling problem

- m horses
- X_i denotes the winning horse at time i. $\mathcal{X} = \{1, 2, .., m\}$
- $\{X_i\}_{i \geq 1}$, i.i.d. $\sim p(x)$

The horse race gambling problem

- m horses
- X_i denotes the winning horse at time i. $\mathcal{X} = \{1, 2, .., m\}$
- $\{X_i\}_{i \geq 1}$, i.i.d. $\sim p(x)$
- $b(X_i)$ is the fraction of wealth that we invest at time i on horse X_i

The horse race gambling problem

- m horses
- X_i denotes the winning horse at time i. $\mathcal{X} = \{1, 2, .., m\}$
- $\{X_i\}_{i \geq 1}$, i.i.d. $\sim p(x)$
- $b(X_i)$ is the fraction of wealth that we invest at time i on horse X_i
- uniform odds, i.e., $1 : m$

The horse race gambling problem

- m horses
- X_i denotes the winning horse at time i. $\mathcal{X} = \{1, 2, .., m\}$
- $\{X_i\}_{i \geq 1}$, i.i.d. $\sim p(x)$
- $b(X_i)$ is the fraction of wealth that we invest at time i on horse X_i
- uniform odds, i.e., $1 : m$
- our wealth after investing n times is

$$\prod_{i=1}^{n} b(X_i) \cdot m$$

The horse race gambling problem

- m horses
- X_i denotes the winning horse at time i. $\mathcal{X} = \{1, 2, .., m\}$
- $\{X_i\}_{i \geq 1}$, i.i.d. $\sim p(x)$
- $b(X_i)$ is the fraction of wealth that we invest at time i on horse X_i
- uniform odds, i.e., $1 : m$
- our wealth after investing n times is

$$\prod_{i=1}^{n} b(X_i) \cdot m$$

- our goal is to maximize the growth rate, i.e.,

$$\max_{b(X_i)} \mathsf{E}\left[\log \prod_{i=1}^{n} b(X_i) m\right]$$

The Horse Race

- X_i the horse that wins at time i
- $b(X_i)$ investment at time i
- goal: $W^*(X^n) = \max_{b(X_i)} \mathsf{E}[\log \prod_{i=1}^{n} b(X_i)m]$

Note: if we invest all our money on one horse, we will eventually go broke .

The Horse Race

- X_i the horse that wins at time i
- $b(X_i)$ investment at time i
- goal: $W^*(X^n) = \max_{b(X_i)} \mathrm{E}[\log \prod_{i=1}^{n} b(X_i)m]$

Note: if we invest all our money on one horse, we will eventually go broke .

Kelly (1956)

The optimal strategy is to invest the capital according to $p(x)$, i.e., $b(x) = p(x)$. The optimal growth is

$$\frac{1}{n} W^*(X^n) = \log m - H(X)$$

Gambling with side information

Summary of the problem:

- X_i the horse that wins at time i
- Y_i side information
- (X_i, Y_i), i.i.d. $\sim p(x, y)$.
- $b(X_i|Y_i)$ investment at time i
- goal: $W^*(X^n|Y^n) = \max_{b(X_i|Y_i)} \mathsf{E}[\log \prod_{i=1}^{n} b(X_i|Y_i)m]$

"A new interpretation of information rate" [Kelly56]

The optimal strategy is to invest the capital proportional to $p(x|y)$, i.e., $b(x|y) = p(x|y)$. The increase in the growth rate due to side information Y is

$$\Delta W = I(X;Y).$$

Directed Information

- Massey introduced it in 1990:

$$I(X^n \to Y^n) \triangleq \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}).$$

Directed Information

- Massey introduced it in 1990:

$$I(X^n \to Y^n) \triangleq \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}).$$

- feedback capacity: [Massey90] [Kramer98] [Tatikonda00] [Yang/Kavcic/Tatikonda05] [Chen/Berger05] [Kim07] [P./Cuff/Van Roy/Weissman07] [Yuksel/Tatikonda07] [Shrader/P.07] [P./Weissman/Chen08] [Dabora/Goldsmith08]
- rate distortion: [Zamir/Kochman/Erez06] [Venkataramanan/Pradhan07]

Directed Information

- Massey introduced it in 1990:

$$I(X^n \to Y^n) \triangleq \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}).$$

- feedback capacity: [Massey90] [Kramer98] [Tatikonda00] [Yang/Kavcic/Tatikonda05] [Chen/Berger05] [Kim07] [P./Cuff/Van Roy/Weissman07] [Yuksel/Tatikonda07] [Shrader/P.07] [P./Weissman/Chen08] [Dabora/Goldsmith08]
- rate distortion: [Zamir/Kochman/Erez06] [Venkataramanan/Pradhan07]

Does it have an interpretation in gambling?

Definitions

$$I(X^n;Y^n) \triangleq H(Y^n) - H(Y^n|X^n)$$

$$H(Y^n|X^n) \triangleq E[-\log P(Y^n|X^n)]$$

$$P(y^n|x^n) = \prod_{i=1}^{n} P(y_i|x^n, y^{i-1})$$

Definitions

Directed Information [Massey90]

$$
\begin{aligned}
I(X^n \rightarrow Y^n) &\triangleq H(Y^n) - H(Y^n || X^n) \\
I(X^n ; Y^n) &\triangleq H(Y^n) - H(Y^n | X^n)
\end{aligned}
$$

$$H(Y^n|X^n) \triangleq E[-\log P(Y^n|X^n)]$$

$$P(y^n|x^n) = \prod_{i=1}^{n} P(y_i|x^n, y^{i-1})$$

Definitions

Directed Information [Massey90]

$$
\begin{aligned}
I(X^n \rightarrow Y^n) &\triangleq H(Y^n) - H(Y^n || X^n) \\
I(X^n ; Y^n) &\triangleq H(Y^n) - H(Y^n | X^n)
\end{aligned}
$$

Causal Conditioning [Kramer98]

$$
\begin{aligned}
H(Y^n || X^n) &\triangleq E[-\log P(Y^n || X^n)] \\
H(Y^n | X^n) &\triangleq E[-\log P(Y^n | X^n)]
\end{aligned}
$$

$$
P(y^n | x^n) = \prod_{i=1}^{n} P(y_i | x^n, y^{i-1})
$$

Definitions

Directed Information [Massey90]

$$I(X^n \rightarrow Y^n) \triangleq H(Y^n) - H(Y^n || X^n)$$
$$I(X^n; Y^n) \triangleq H(Y^n) - H(Y^n | X^n)$$

Causal Conditioning [Kramer98]

$$H(Y^n || X^n) \triangleq E[-\log P(Y^n || X^n)]$$
$$H(Y^n | X^n) \triangleq E[-\log P(Y^n | X^n)]$$

$$P(y^n || x^n) \triangleq \prod_{i=1}^{n} P(y_i | x^i, y^{i-1})$$
$$P(y^n | x^n) = \prod_{i=1}^{n} P(y_i | x^n, y^{i-1})$$

Definitions

Directed Information [Massey90]

$$I(X^n \rightarrow Y^n) \triangleq H(Y^n) - H(Y^n || X^n)$$
$$I(X^n; Y^n) \triangleq H(Y^n) - H(Y^n | X^n)$$

Causal Conditioning [Kramer98]

$$H(Y^n || X^n) \triangleq E[-\log P(Y^n || X^n)]$$
$$H(Y^n | X^n) \triangleq E[-\log P(Y^n | X^n)]$$

$$P(y^n || x^n) \triangleq \prod_{i=1}^{n} P(y_i | x^i, y^{i-1})$$
$$P(y^n || x^{n-1}) \triangleq \prod_{i=1}^{n} P(y_i | x^{i-1}, y^{i-1})$$

Chain rule

causal conditioning

$$p(x^n||y^n) \triangleq \prod_{i=1}^{n} p(x_i|x^{i-1}, y^i),$$
$$p(y^n||x^{n-1}) \triangleq \prod_{i=1}^{n} p(y_i|y^{i-1}, x^{i-1})$$

chain rule

$$p(x^n, y^n) = p(x^n||y^n)p(y^n||x^{n-1})$$

Gambling with *causal* side information

- X_i the horse that wins at time i
- Y_i side information that is known causally
- $(X^n, Y^n) \sim p(x^n, y^n)$.
- $b(X_i|X^{i-1}, Y^i)$ investment at time i

$$W^*(X^n||Y^n) = \max_{\{b(X_i|X^{i-1},Y^i)\}} \mathsf{E}[\log \prod_{i=1}^{n} b(X_i|X^{i-1}, Y^i)m]$$

Gambling with *causal* side information

- X_i the horse that wins at time i
- Y_i side information that is known causally
- $(X^n, Y^n) \sim p(x^n, y^n)$.
- $b(X_i|X^{i-1}, Y^i)$ investment at time i

$$\begin{aligned} W^*(X^n||Y^n) &= \max_{\{b(X_i|X^{i-1},Y^i)\}} \mathrm{E}[\log \prod_{i=1}^{n} b(X_i|X^{i-1}, Y^i)m] \\ &= \max_{b(X^n||Y^n)} \mathrm{E}[\log b(X^n||Y^n)m^n] \end{aligned}$$

Gambling with *causal* side information

- X_i the horse that wins at time i
- Y_i side information that is known causally
- $(X^n, Y^n) \sim p(x^n, y^n)$.
- $b(X_i|X^{i-1}, Y^i)$ investment at time i

$$\begin{aligned} W^*(X^n||Y^n) &= \max_{\{b(X_i|X^{i-1},Y^i)\}} \mathsf{E}[\log \prod_{i=1}^{n} b(X_i|X^{i-1}, Y^i)m] \\ &= \max_{b(X^n||Y^n)} \mathsf{E}[\log b(X^n||Y^n)m^n] \end{aligned}$$

Theorem

The optimal strategy is $b(x^n||y^n) = p(x^n||y^n)$. The increase in the growth rate due to causal side information Y_i is

$$\Delta W = \frac{1}{n} I(Y^n \to X^n).$$

causal side information: the i.i.d. case

If (X^n, Y^n) are i.i.d. $\sim p(x, y)$, we simply obtain

$$\frac{1}{n} I(Y^n \to X^n) \;=\; I(X; Y)$$

The proof is simple

$$\mathsf{E}[\log b(X^n||Y^n)]$$
$$= \sum_{x^n,y^n} p(x^n,y^n)\left[\log p(x^n||y^n) + \log \frac{b(x^n||y^n)}{p(x^n||y^n)}\right]$$

The proof is simple

$$
\begin{aligned}
&\mathsf{E}[\,\log b(X^n||Y^n)] \\
&\quad = \sum_{x^n,y^n} p(x^n,y^n)\left[\log p(x^n||y^n) + \log \frac{b(x^n||y^n)}{p(x^n||y^n)}\right] \\
&\quad = -H(X^n||Y^n) + \sum_{x^n,y^n} p(x^n,y^n) \log \frac{b(x^n||y^n)}{p(x^n||y^n)}
\end{aligned}
$$

The proof is simple

$$\begin{aligned}
&\mathsf{E}[\,\log b(X^n||Y^n)] \\
&\quad = \sum_{x^n,y^n} p(x^n,y^n)\left[\log p(x^n||y^n) + \log \frac{b(x^n||y^n)}{p(x^n||y^n)}\right] \\
&\quad = -H(X^n||Y^n) + \sum_{x^n,y^n} p(x^n,y^n) \log \frac{b(x^n||y^n)}{p(x^n||y^n)} \\
&\quad \leq -H(X^n||Y^n) + \log \sum_{x^n,y^n} p(x^n,y^n) \frac{b(x^n||y^n)}{p(x^n||y^n)}
\end{aligned}$$

The proof is simple

$$
\begin{aligned}
&\mathsf{E}[\log b(X^n||Y^n)] \\
&= \sum_{x^n,y^n} p(x^n,y^n)\left[\log p(x^n||y^n) + \log \frac{b(x^n||y^n)}{p(x^n||y^n)}\right] \\
&= -H(X^n||Y^n) + \sum_{x^n,y^n} p(x^n,y^n) \log \frac{b(x^n||y^n)}{p(x^n||y^n)} \\
&\leq -H(X^n||Y^n) + \log \sum_{x^n,y^n} p(x^n,y^n) \frac{b(x^n||y^n)}{p(x^n||y^n)}
\end{aligned}
$$

Recall the chain rule $p(x^n,y^n) = p(x^n||y^n)p(y^n||x^{n-1})$.

The proof is simple

$$
\begin{aligned}
&\mathsf{E}[\,\mathsf{log}\, b(X^n||Y^n)] \\
&= \sum_{x^n,y^n} p(x^n,y^n)\left[\mathsf{log}\, p(x^n||y^n) + \mathsf{log}\, \frac{b(x^n||y^n)}{p(x^n||y^n)}\right] \\
&= -H(X^n||Y^n) + \sum_{x^n,y^n} p(x^n,y^n)\,\mathsf{log}\, \frac{b(x^n||y^n)}{p(x^n||y^n)} \\
&\le -H(X^n||Y^n) + \mathsf{log} \sum_{x^n,y^n} p(x^n,y^n)\frac{b(x^n||y^n)}{p(x^n||y^n)} \\
&\le -H(X^n||Y^n) + \mathsf{log} \sum_{x^n,y^n} p(y^n||x^{n-1})b(x^n||y^n)
\end{aligned}
$$

Recall the chain rule $p(x^n,y^n) = p(x^n||y^n)p(y^n||x^{n-1})$.

The proof is simple

$$
\begin{aligned}
&\mathsf{E}[\log b(X^n||Y^n)] \\
&= \sum_{x^n,y^n} p(x^n,y^n)\left[\log p(x^n||y^n) + \log \frac{b(x^n||y^n)}{p(x^n||y^n)}\right] \\
&= -H(X^n||Y^n) + \sum_{x^n,y^n} p(x^n,y^n)\log \frac{b(x^n||y^n)}{p(x^n||y^n)} \\
&\le -H(X^n||Y^n) + \log \sum_{x^n,y^n} p(x^n,y^n)\frac{b(x^n||y^n)}{p(x^n||y^n)} \\
&\le -H(X^n||Y^n) + \log \sum_{x^n,y^n} p(y^n||x^{n-1})b(x^n||y^n) \\
&= -H(X^n||Y^n) + 0
\end{aligned}
$$

Recall the chain rule $p(x^n,y^n) = p(x^n||y^n)p(y^n||x^{n-1})$.

Proof...

Causal side information:

$$\frac{1}{n}W^*(X^n||Y^n) = \log m - \frac{1}{n}H(X^n||Y^n).$$

no side information

$$\frac{1}{n}W^*(X^n) = \log m - \frac{1}{n}H(X^n).$$

The increase in growth rate

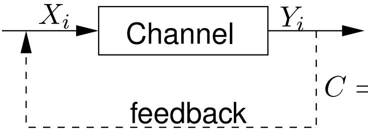
$$\begin{aligned} \frac{1}{n}W^*(X^n||Y^n) - \frac{1}{n}W^*(X^n) &= -\frac{1}{n}H(X^n||Y^n) + \frac{1}{n}H(X^n) \\ &= \frac{1}{n}I(Y^n \to X^n) \end{aligned}$$

Intuition

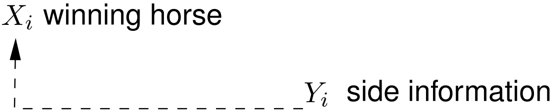
- $I(Y^n; X^n) = H(X^n) - H(X^n|Y^n)$ amount of uncertainty about X^n reduced by knowing Y^n.
- $I(Y^n \to X^n) = H(X^n) - H(X^n||Y^n)$ amount of uncertainty about X^n reduced by knowing Y^n causally.

Intuition

- $I(Y^n; X^n) = H(X^n) - H(X^n|Y^n)$ amount of uncertainty about X^n reduced by knowing Y^n.
- $I(Y^n \to X^n) = H(X^n) - H(X^n||Y^n)$ amount of uncertainty about X^n reduced by knowing Y^n causally.

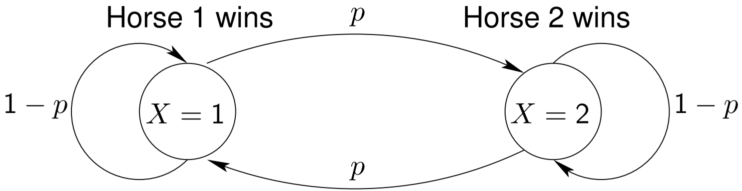


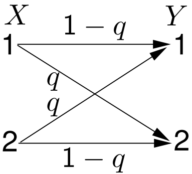
$$C = \frac{1}{n} I(X^n \to Y^n)$$



$$\Delta W = \frac{1}{n} I(Y^n \to X^n)$$

An example





$$\frac{1}{n} I(Y^n \to X^n) = \quad H(Y_1|X_0) - H(Y_1|X_1) = h(p * q) - h(q),$$

$$p * q = (1-p)q + (1-q)p$$

Stock Market [Barron/Cover88], [Cover/Thomas06, ch. 13]

- $\mathbf{X_i} = (X_{i,1}, X_{i,2}, ..., X_{i,m})$ - the relative price at time i.

$$X_{i,k} = \frac{\text{stock-}k \text{ at the end of day } i}{\text{stock-}k \text{ at the end of day } i-1}$$

- Y_i causal side information
- $\mathbf{b}(\mathbf{x}^{i-1}, y^i)$ is the portfolio. It is non-negative and $\sum_{k=1}^{m} b_k(\mathbf{x}^{i-1}, y^i) = 1$.
- The goal is to maximize the growth rate, i.e.,

$$\max_{\{\mathbf{b}(\mathbf{x}^{i-1}, y^i)\}_{i=1}^{n}} E\left[\sum_{i=1}^{n} \log(\mathbf{b}^t(\mathbf{X}^{i-1}, Y^i)\mathbf{X}_i)\right]$$

Theorem

The increase in growth rate in $n-$epoch time investments due to side information is bounded by $\Delta W \leq \frac{1}{n} I(Y^n \to \mathbf{X}^n)$.

Lossless Compression

- X_i source
- Y_i causal side information
- encoder and decoder are instantaneous and variable length.
  - Encoder: $M_i = g(X^i, Y^i)$
  - Decoder: $\hat{X}_i = f(M^i, Y^i)$
- the transmission rate

$$R = \frac{1}{n} \sum_{i=1}^{n} \log |M_i|$$

$$\frac{1}{n} H(X^n || Y^n) \leq R \leq \frac{1}{n} H(X^n || Y^n) + 1$$

Theorem

The decrease in the transmission rate due to causal side information is $\frac{1}{n} I(Y^n \to \mathbf{X}^n) + c$, where $|c| \leq 1$.

Summary

$I(Y^n;X^n) \triangleq H(X^n) - H(X^n|Y^n)$ amount of uncertainty about X^n reduced by knowing Y^n
$I(Y^n \to X^n) \triangleq H(X^n) - H(X^n||Y^n)$ amount of uncertainty about X^n reduced by knowing Y^n causally.

- Gambling with causal side information:

$$b(x^n||y^n) = p(x^n||y^n),$$

$$\Delta W = \frac{1}{n} I(Y^n \to X^n).$$

- Portfolio theory: $\Delta W \leq \frac{1}{n} I(Y^n \to \mathbf{X}^n).$
- Instantaneous compression: $\Delta R \leq \frac{1}{n} I(Y^n \to X^n) + c$

Summary

$I(Y^n; X^n) \triangleq H(X^n) - H(X^n|Y^n)$ amount of uncertainty about X^n reduced by knowing Y^n
$I(Y^n \to X^n) \triangleq H(X^n) - H(X^n||Y^n)$ amount of uncertainty about X^n reduced by knowing Y^n causally.

- Gambling with causal side information:

$$b(x^n||y^n) = p(x^n||y^n),$$

$$\Delta W = \frac{1}{n} I(Y^n \to X^n).$$

- Portfolio theory: $\Delta W \leq \frac{1}{n} I(Y^n \to \mathbf{X}^n)$.
- Instantaneous compression: $\Delta R \leq \frac{1}{n} I(Y^n \to X^n) + c$

Thank you for attending the talk!