relations between

information

and

estimation

in the presence of

feedback

Tsachy Weissman

talk at workshop on:

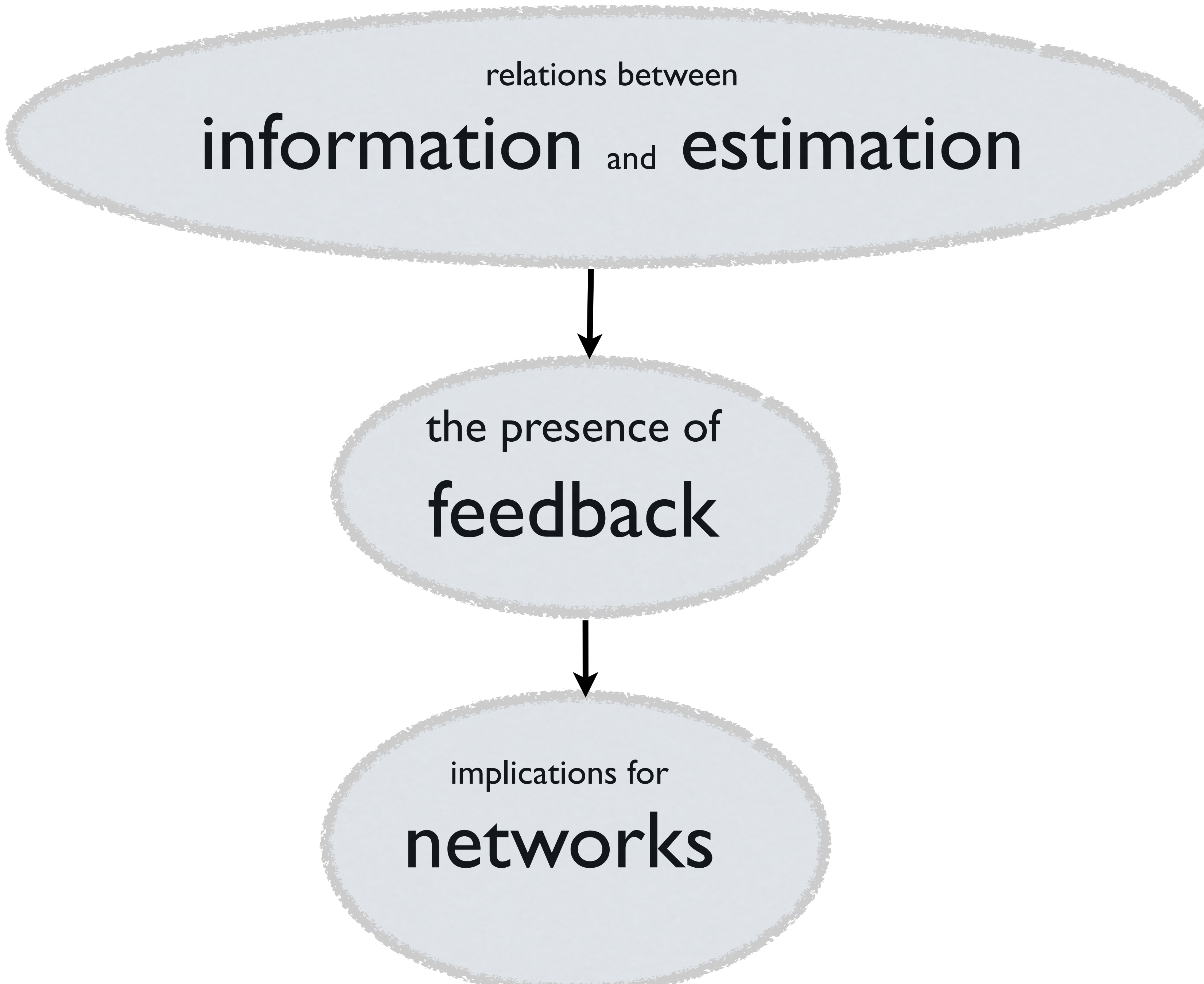
Information and Control in Networks

LCCC - Lund Center for Control of Complex Engineering Systems

information, control, networks

- real-time and limited delay communication
- feedback communications
- “action” in information theory
- relations between information and estimation (w. feedback + networks)

Outline



relations between
information and estimation
the presence of
feedback
implications for
networks

Haves and Have-Nots (in this talk)

we'll have:

- some theorems
- cute (and meaningful) relations
- an algorithmic framework

we won't have:

- account of related literature
- stipulations
- proofs
- algorithms
- data

“de Bruijn’s identity”

[A. J. Stam 1959]:

X independent of $Z \sim \mathcal{N}(0,1)$

$$\frac{d}{dt} h\left(X + \sqrt{t} Z\right) = \frac{1}{2} J(X + \sqrt{t} Z)$$

$G_{uo}S_{hamai}V_{erdu}$ setting

$$Y = \sqrt{\gamma} \cdot X + W$$

W is a standard Gaussian, independent of X

$$I(\gamma) = I(X;Y)$$

$$\mathsf{mmse}(\gamma) = E\left[(X - E[X|Y])^2\right]$$

[Guo, Shamai and Verdú 2005]:

$$\frac{d}{d\gamma} I(\gamma) = \frac{1}{2} \mathsf{mmse}(\gamma)$$

GSV in continuous time

$$dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \leq t \leq T$$

$$I(\gamma) = I(X^T; Y^T)$$

$$\text{mmse}(\gamma) = E\left[\int_0^T (X_t - E[X_t | Y^T])^2 dt\right]$$

[Guo, Shamai and Verdú 2005], [Zakai 2005]:

$$\frac{d}{d\gamma} I(\gamma) = \frac{1}{2}\mathsf{mmse}(\gamma)$$

or in its integral version

$$I(\mathsf{snr}) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma) d\gamma$$

Duncan

$$dY_t = X_t dt + dW_t, \qquad 0 \le t \le T$$

W is standard white Gaussian noise, independent of X

[Duncan 1970]:

$$I(X^T; Y^T) = \frac{1}{2} E\left[\int_0^T (X_t - E[X_t|Y^t])^2 dt\right]$$

SNR in Duncan

$$dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \le t \le T$$

$$I(\gamma) = I(X^T; Y^T)$$

$$\mathsf{cmmse}(\gamma) = E\left[\int_0^T (X_t - E[X_t | Y^t])^2 dt\right]$$

[Duncan 1970]:

$$I(\gamma) = \frac{\gamma}{2} \cdot \mathsf{cmmse}(\gamma)$$

Recap

[Duncan 1970]:

$$I(\gamma) = \frac{\gamma}{2} \cdot \mathsf{cmmse}(\gamma)$$

[Guo, Shamai and Verdú 2005], [Zakai 2005]:

$$I(\mathsf{snr}) = \frac{1}{2} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma) d\gamma$$

$$\Rightarrow ?$$

Relationship between cmmse and mmse

[Guo, Shamai and Verdú 2005]:

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{1}{\mathsf{snr}} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma) d\gamma$$

Mismatch

$$Y = \sqrt{\gamma} \cdot X + W$$

W is a standard Gaussian, independent of X

What if $X \sim P$ but the estimator thinks $X \sim Q$?

$$\mathsf{mse}_{P,Q}(\gamma) = E_P\left[(X - E_Q[X|Y])^2\right]$$

A representation of relative entropy [Verdu 2010]:

$$D(P\|Q) = \int_0^\infty [\mathsf{mse}_{P,Q}(\gamma) - \mathsf{mse}_{P,P}(\gamma)]d\gamma$$

$$D(P_{Y_{\mathsf{snr}}}\|Q_{Y_{\mathsf{snr}}}) = \int_0^{\mathsf{snr}} [\mathsf{mse}_{P,Q}(\gamma) - \mathsf{mse}_{P,P}(\gamma)]d\gamma$$

Causal vs. Non-causal Mismatched Estimation

$$dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \leq t \leq T$$

W is standard white Gaussian noise, independent of X

$$\mathsf{cmse}_{P,Q}(\gamma) = E_P \left[\int_0^T (X_t - E_Q[X_t | Y^t])^2 dt \right]$$

$$\mathsf{mse}_{P,Q}(\gamma) = E_P \left[\int_0^T (X_t - E_Q[X_t | Y^T])^2 dt \right]$$

Causal vs. Non-causal Mismatched Estimation

$$dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \leq t \leq T$$

W is standard white Gaussian noise, independent of X

$$\mathsf{cmse}_{P,Q}(\gamma) = E_P \left[\int_0^T (X_t - E_Q[X_t | Y^t])^2 dt \right]$$

$$\mathsf{mse}_{P,Q}(\gamma) = E_P \left[\int_0^T (X_t - E_Q[X_t | Y^T])^2 dt \right]$$

Relationship between $\mathsf{cmse}_{P,Q}$ and $\mathsf{mse}_{P,Q}$?

Relationship between $\mathsf{cmse}_{P,Q}$ and $\mathsf{mse}_{P,Q}$

[Weissman 2010]:

$$\mathsf{cmse}_{P,Q}(\mathsf{snr}) \quad = \quad \frac{1}{\mathsf{snr}} \int_0^{\mathsf{snr}} \mathsf{mse}_{P,Q}(\gamma) d\gamma$$

Relationship between $\mathsf{cmse}_{P,Q}$ and $\mathsf{mse}_{P,Q}$

[Weissman 2010]:

$$\begin{aligned}\mathsf{cmse}_{P,Q}(\mathsf{snr}) &= \frac{1}{\mathsf{snr}}\int_0^{\mathsf{snr}} \mathsf{mse}_{P,Q}(\gamma)d\gamma \\ &= \frac{2}{\mathsf{snr}}\left[I(\mathsf{snr}) + D\left(P_{Y^T}\|Q_{Y^T}\right)\right]\end{aligned}$$

Implications and Applications

- many

Minimax (causal) Estimation

$$\mathsf{minimax}(\mathcal{P}, \mathsf{snr}) \stackrel{\triangle}{=} \min_{\{\hat{X}_t(\cdot)\}_{0\le t\le T}} \max_{P\in\mathcal{P}} \left\{ E_P\left[\int_0^T \ell(X_t, \hat{X}_t(Y^t))dt\right] - \mathsf{cmse}_{P,P}(\mathsf{snr}) \right\}$$

Minimax (causal) Estimation

$$\mathsf{minimax}(\mathcal{P}, \mathsf{snr}) \triangleq \min_{\{\hat{X}_t(\cdot)\}_{0\le t\le T}} \max_{P\in\mathcal{P}} \left\{ E_P\left[\int_0^T \ell(X_t, \hat{X}_t(Y^t))dt\right] - \mathsf{cmse}_{P,P}(\mathsf{snr}) \right\}$$

classical

Minimax (causal) Estimation

$$\mathsf{minimax}(\mathcal{P}, \mathsf{snr}) \triangleq \min_{\{\hat{X}_t(\cdot)\}_{0 \le t \le T}} \max_{P \in \mathcal{P}} \left\{ E_P \left[\int_0^T \ell(X_t, \hat{X}_t(Y^t)) dt \right] - \mathsf{cmse}_{P,P}(\mathsf{snr}) \right\}$$

classical

ours

Minimax (causal) Estimation

$$\mathsf{minimax}(\mathcal{P}, \mathsf{snr}) \stackrel{\triangle}{=} \min_{\{\hat{X}_t(\cdot)\}_{0\le t\le T}} \max_{P\in\mathcal{P}} \left\{ E_P\left[\int_0^T \ell(X_t, \hat{X}_t(Y^t))dt\right] - \mathsf{cmse}_{P,P}(\mathsf{snr}) \right\}$$

classical

ours

Redundancy-Capacity theory

Minimax (causal) Estimation

$$\text{minimax}(\mathcal{P}, \text{snr}) \triangleq \min_{\{\hat{X}_t(\cdot)\}_{0\le t\le T}} \max_{P\in\mathcal{P}} \left\{ E_P\left[\int_0^T \ell(X_t, \hat{X}_t(Y^t))dt\right] - \text{cmse}_{P,P}(\text{snr}) \right\}$$

classical

ours

Redundancy-Capacity theory

Shannon

Minimax (causal) Estimation

$$\text{minimax}(\mathcal{P}, \text{snr}) \triangleq \min_{\{\hat{X}_t(\cdot)\}_{0\le t\le T}} \max_{P\in\mathcal{P}} \left\{ E_P\left[\int_0^T \ell(X_t, \hat{X}_t(Y^t))dt\right] - \text{cmse}_{P,P}(\text{snr}) \right\}$$

$$\text{minimax}(\mathcal{P}, \text{snr}) \overset{\text{classical}}{=} \min_{Q} \max_{P\in\mathcal{P}} \left[\text{cmse}_{P,Q}(\text{snr}) - \text{cmse}_{P,P}(\text{snr})\right]$$

$$\overset{\text{ours}}{=} \frac{2}{\text{snr}} \min_{Q} \max_{P\in\mathcal{P}} D\left(P_{Y_{\text{snr}}^T} \| Q_{Y_{\text{snr}}^T}\right)$$

$$\overset{\text{Redundancy-Capacity theory}}{=} \frac{2}{\text{snr}} \max\left\{I\left(\Theta; Y_{\text{snr}}^T\right) : \Theta \text{ is a } \mathcal{P}\text{-valued RV}\right\}$$

$$\overset{\text{Shannon}}{=} \frac{2}{\text{snr}} C\left(\left\{P_{Y_{\text{snr}}^T}\right\}_{P\in\mathcal{P}}\right)$$

Strong Converse

"strong redundancy-capacity" result of
[Merhav and Feder 1995] applied here implies:

$$\forall \varepsilon > 0 \text{ and } \textit{any} \text{ filter } \{\hat{X}_t(\cdot)\}_{0 \le t \le T},$$

$$E_P\left[\int_0^T \ell(X_t, \hat{X}_t(Y^t))dt\right] - \mathsf{cmse}_{P,P}(\mathsf{snr}) \ge (1-\varepsilon) \cdot \mathsf{minimax}(\mathcal{P}, \mathsf{snr})$$

for all $P \in \mathcal{P}$ with the possible exception of sources in a subset $\mathcal{B} \subset \mathcal{P}$ where

$$w^*(\mathcal{B}) \le e \cdot 2^{-\varepsilon \cdot \mathsf{minimax}(\mathcal{P}, \mathsf{snr})},$$

w^* being the capacity achieving prior

Example

Given:

orthonormal signal set $\{\phi_i(t), 0 \le t \le T\}_{i=1}^{n}$

$$X_t = \sum_{i=1}^{n} B_i \cdot \phi_i(t)$$

$$\mathcal{P} = \left\{ \text{laws } P \text{ on } X^T : E_P\|B\|^2 \le n\beta \text{ and } E_P\|B\|_0 \le n\alpha \right\}$$

$$\max I(X^T; Y^T) = \ ?$$

Example (cont.)

$$Y_i = \int_0^T \phi_i(t) dY_t \quad 1 \leq i \leq n$$

are sufficient statistics for Y^T,

$$\Longrightarrow$$

$$I(X^T; Y^T) = I(B^n; Y^n)$$

$$\Longrightarrow$$

$$\max I(X^T; Y^T) = \max I(B^n; Y^n) = \max\{I(B; Y) : B^2 \leq \beta, P(B = 0) \geq (1 - \alpha)\}$$

latter considered and numerically solved in:

Lei Zhang and Dongning Guo, "Capacity of Gaussian Channels with Duty Cycle and Power Constraints", IEEE Int. Symposium on Information Theory 2011

Example (cont.)

thus the minimax filter here is the Bayes filter assuming:

$$X_t = \sum_{i=1}^{n} B_i^* \cdot \phi_i(t)$$

where B_i^* are iid according to the capacity achieving distribution of [Zhang and Guo, 2011]

cf. [Albert No + T.W., ISIT 2013]...

(well) beyond Gaussian noise

- Poisson channel
- Lévy-type channels:
  - Input-Output relationship expressed via Lévy-type stochastic integral
  - can obtain formulae via Lévy-Khintchine-type decompositions

✓ information

control

networks

The presence of Feedback

The presence of Feedback

- what of what we’ve seen carries over to presence of feedback?

Duncan

$$dY_t = X_t dt + dW_t, \qquad 0 \le t \le T$$

W is standard white Gaussian noise, independent of X

[Duncan 1970]:

$$I(X^T;Y^T) = \frac{1}{2}E\left[\int_0^T (X_t - E[X_t|Y^t])^2 dt\right]$$

Breaks down in presence of feedback!

cont time directed info

[W., Permuter, Kim 2012]

$$I\left(X_0^T \to Y_0^T\right) := \inf_{\mathbf{t}} I_{\mathbf{t}}\left(X_0^T \to Y_0^T\right)$$

where

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

Duncan with feedback

Theorem [W., Permuter, Kim 2012]

Let $\{(X_t, B_t)\}_{t=0}^{T}$ *be adapted to the filtration* $\{\mathcal{F}_t\}_{t=0}^{T}$, *where* X_0^T *is a signal of finite average power* $\int_0^T E[X_t^2]dt < \infty$ *and* B_0^T *is a standard Brownian motion. Let* Y_0^T *be the output of the AWGN channel whose input is* X_0^T *and whose noise is driven by* B_0^T, *i.e.,*

$$dY_t = X_t dt + dB_t.$$

Suppose that the regularity assumptions of Proposition 2 are satisfied for all $0 < t < T$. *Then*

$$\frac{1}{2}\int_0^T E\left[(X_t - E[X_t|Y_0^t])^2\right]dt = I(X_0^T \to Y_0^T)$$

compare with [Kadota, Zakai, Ziv 1971]

GSV in continuous time

$$dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \le t \le T$$

$$\frac{d}{d\gamma} I(\gamma) = \frac{1}{2} \mathsf{mmse}(\gamma)$$

or in its integral version

$$I(\mathsf{snr}) = \frac{1}{2} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma) d\gamma$$

GSV in continuous time

$$dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \le t \le T$$

$$\frac{d}{d\gamma} I(\gamma) = \frac{1}{2} \mathsf{mmse}(\gamma)$$

or in its integral version

$$I(\mathsf{snr}) = \frac{1}{2} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma) d\gamma$$

Breaks down in presence of feedback

GSV in continuous time with DI?

$$I(X^T \to Y^T) \stackrel{?}{=} \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\mathsf{d}\gamma$$

No. In general

$$I(X^T \to Y^T) \neq \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\mathsf{d}\gamma$$

and so

$$\mathsf{cmmse}(\mathsf{snr}) \neq \frac{1}{\mathsf{snr}}\int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)d\gamma$$

I.e., breakdown in presence of feedback

Mismatched setting

a fortiori, in presence of feedback, in general

$$\mathsf{cmse}_{P,Q}(\mathsf{snr}) \quad \neq \quad \frac{1}{\mathsf{snr}} \int_0^{\mathsf{snr}} \mathsf{mse}_{P,Q}(\gamma) d\gamma$$

Mismatched setting

a fortiori, in presence of feedback, in general

$$\mathsf{cmse}_{P,Q}(\mathsf{snr}) \quad \neq \quad \frac{1}{\mathsf{snr}} \int_0^{\mathsf{snr}} \mathsf{mse}_{P,Q}(\gamma) d\gamma$$

end of story?

Mismatched setting (cont.)

$$\text{cmse}_{P,Q} - \text{cmse}_{P,P} = D(P_{Y^T} \| Q_{Y^T})$$

holds with or without FB, appears in TW2010 implicitly and explicitly in workshop book chapter [Asnani, Venkat, W. 2012]

(why?)

implications and apps

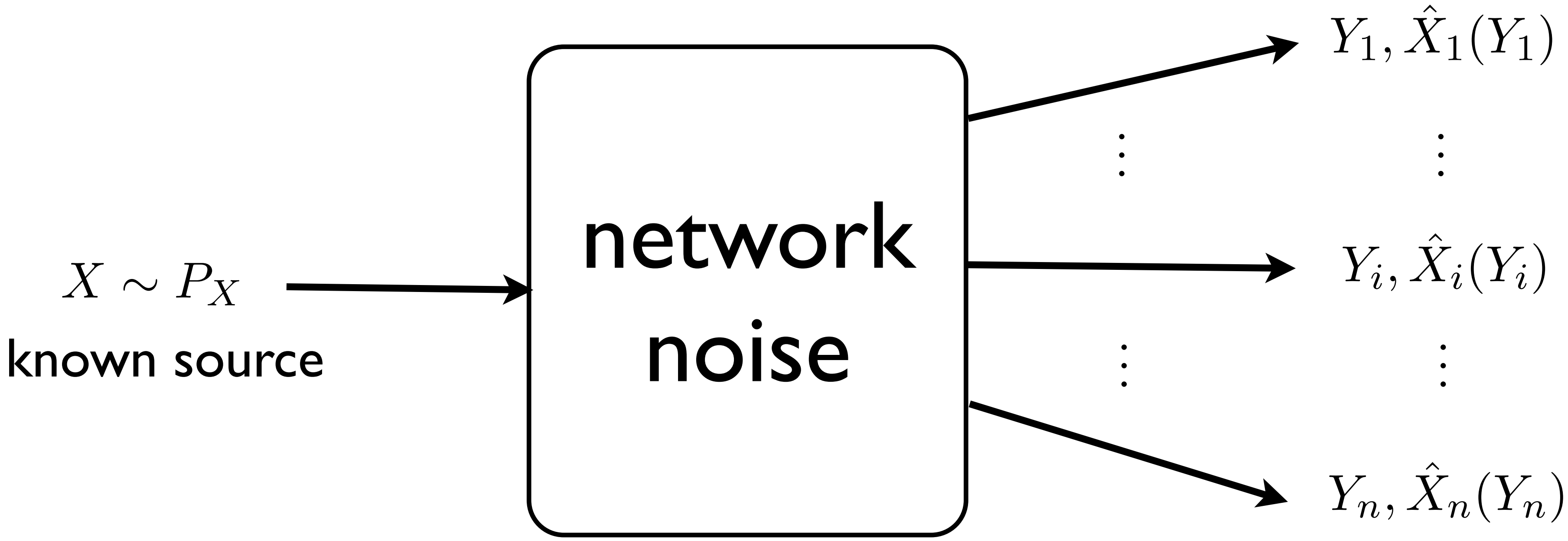
- minimax estimation setting carries over
- directed info maximization instead of mutual info but same idea
- similar extensions to the more general channels

✓ information

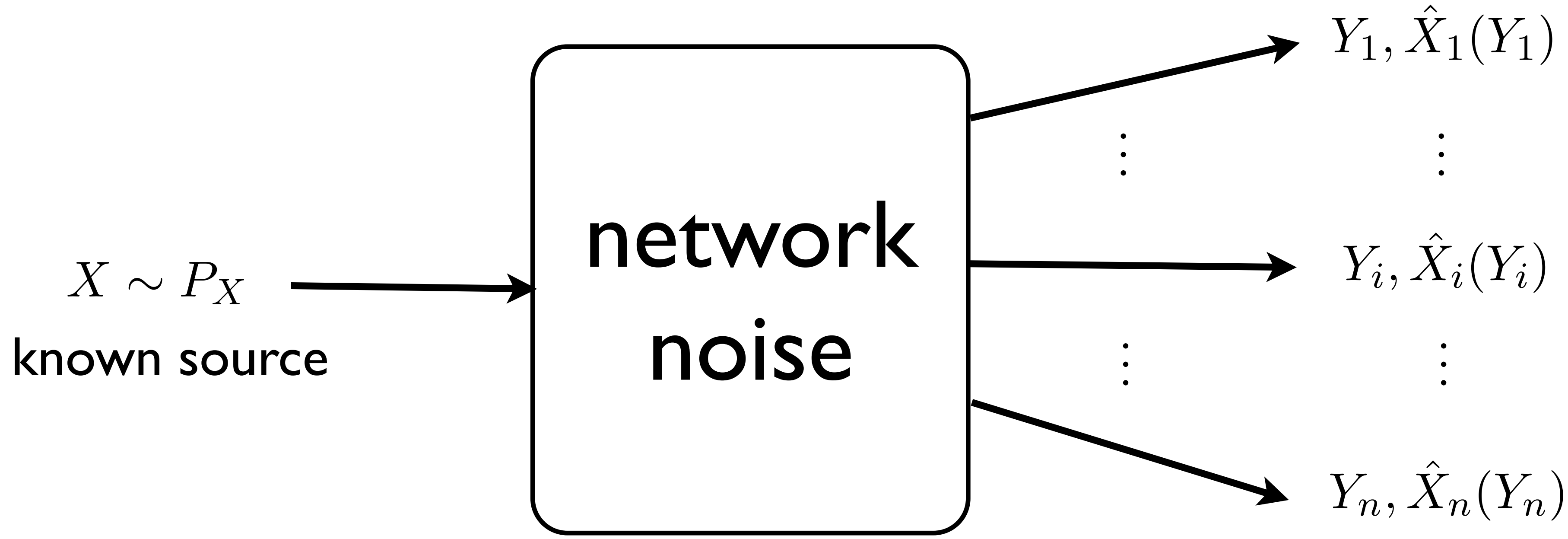
✓ control

networks

Distributed estimation (known source)

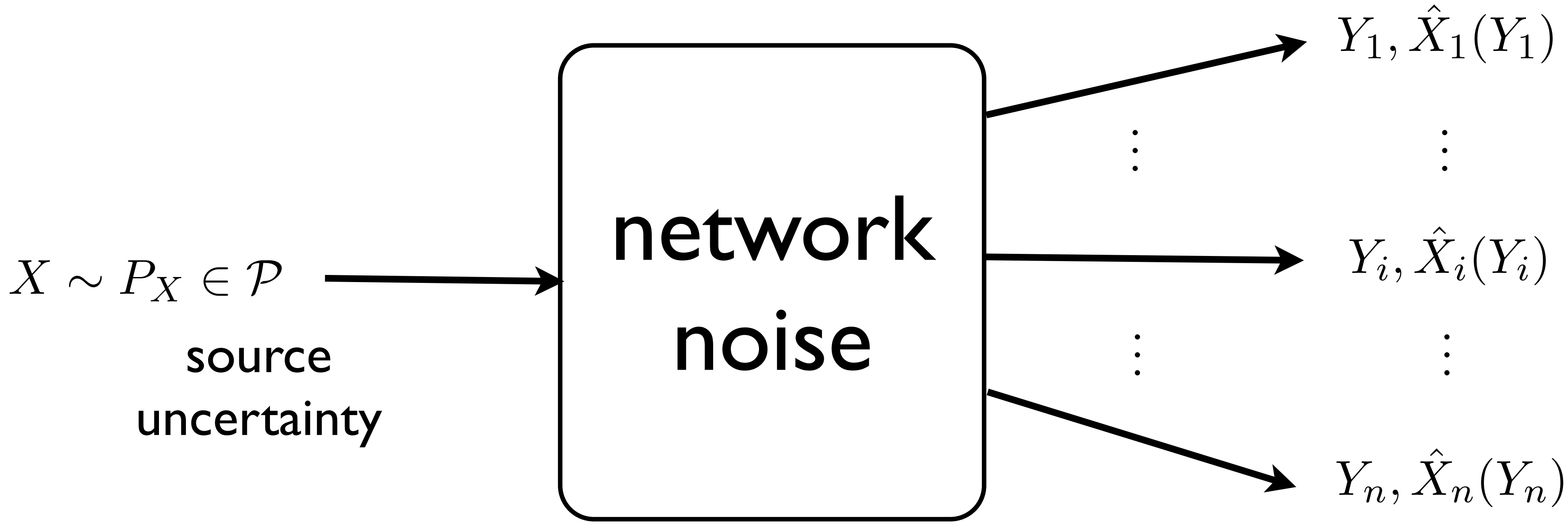


Distributed estimation (known source)

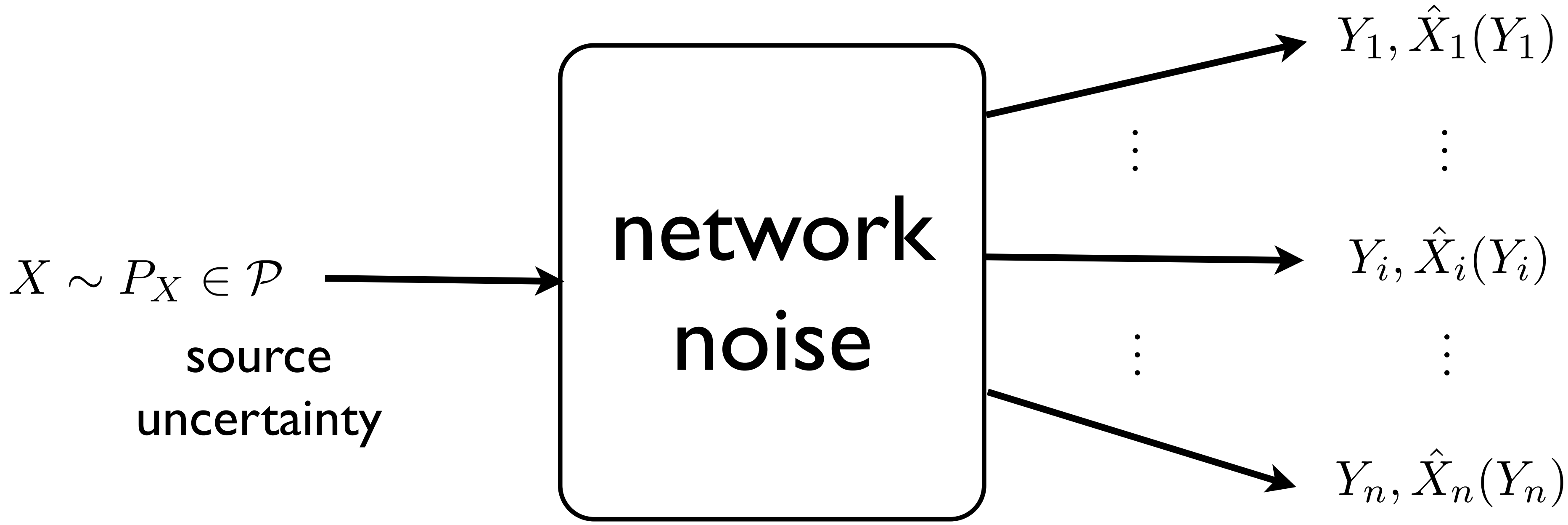


can (and should) be greedy!

Distributed estimation (source uncertainty)

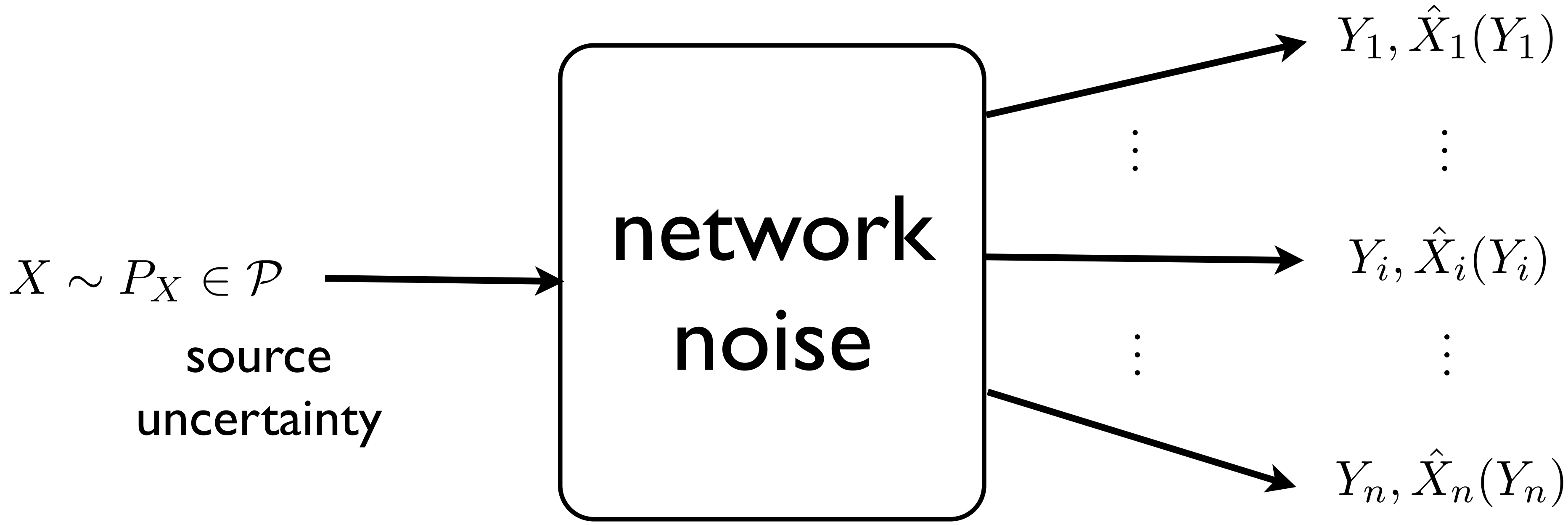


Distributed estimation (source uncertainty)



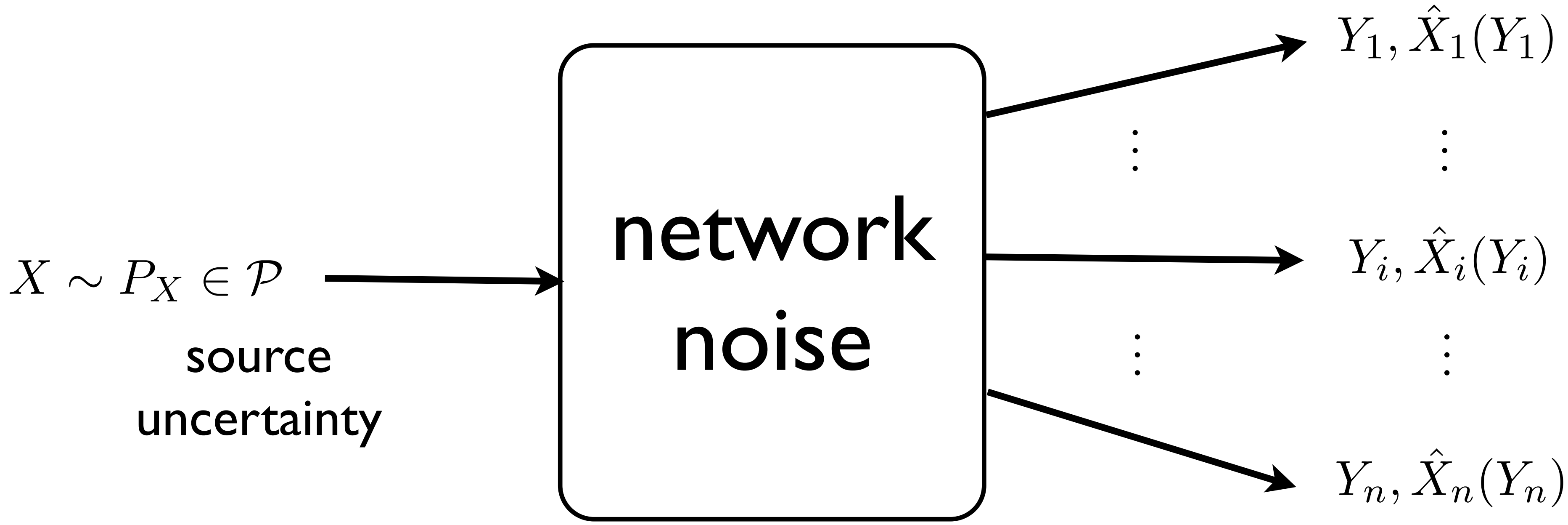
should we be greedy?

Distributed estimation (source uncertainty)



should we be greedy?
no! (in general)

Distributed estimation (source uncertainty)



should we be greedy?

no! (in general)

yes!

(in causal estimation over Gaussian, Poisson, or general Levy-type noise)

minimax estimation for each observation separately would be essentially optimal

✓ information

✓ control

✓ networks

conclusion

- relations between mutual information, relative entropy, and estimation
- findings of pure estimation theoretic significance
- allow the transfer of tools
- much carries over to presence of feedback
- implications for networks