Multiterminal Estimation - Extensions and a Geometric Interpretation.

Rebecka Jörnsten
Bin Yu
Departments of Statistics
Rutgers University and UC Berkeley

http://www.stat.rutgers.edu/~rebecka

Outline

Multiterminal Data Compression

Deals with the problem of optimal inference based on multi-source lossy data.

Multiterminal framework: Provides a measure of loss of statistical information due to the separate compression of multiple sources.

- Best case scenario
- Baseline for comparison of methods

Multiterminal Estimation Problem: Level of compression/loss of statistical estimation efficiency we want/can accept?

Multiterminal Data Compression

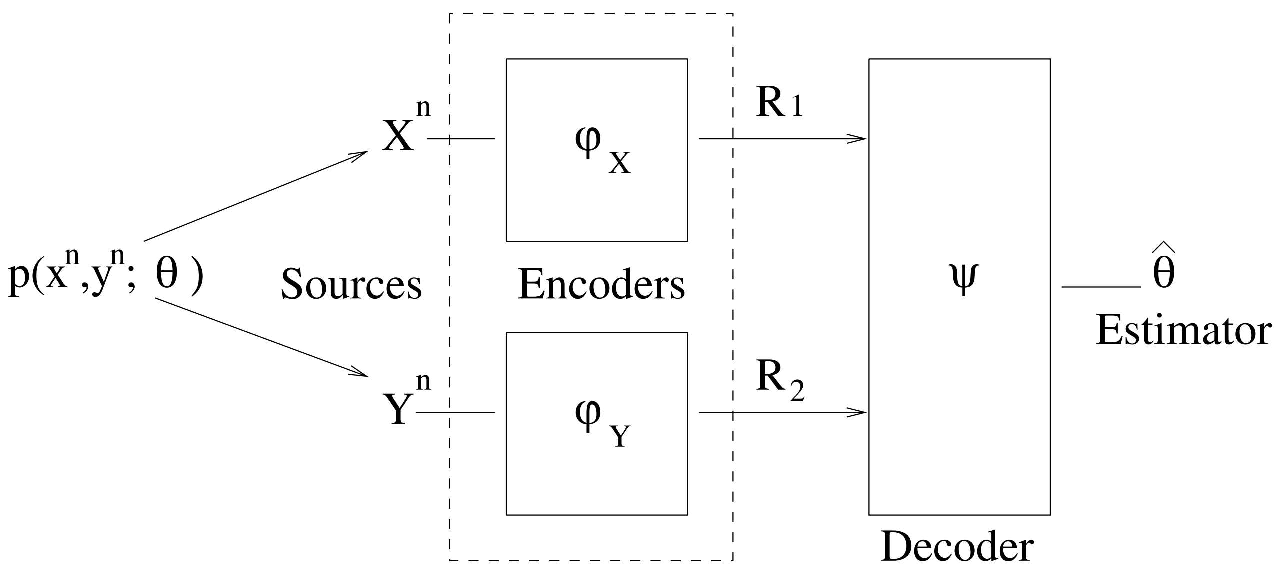
- Remote sensing, Signal Detection, Pattern Recognition.
  Multiterminal hypothesis testing
- CEO problem, Denoising, Signal Retrieval.
  Multiterminal Source Coding

Multiterminal Estimation is by comparision much less studied. Two approaches exist:

-Zhang and Berger (1988)

-Han and Amari (1995,1998).

Problem Setup



The data: $(X^n, Y^n) \sim p_\theta(x^n, y^n)$, where $\mathcal{X} = 1, \dots, \mathcal{K}$ and $\mathcal{Y} = 1, \dots, \mathcal{L}$

Encoders: $\varphi_X : X^n \to \varphi_X(X^n)$, where $\frac{1}{n} \log_2 ||\varphi_X|| \leq R_1 + \delta$

$\varphi_Y : Y^n \to \varphi_Y(Y^n)$, where $\frac{1}{n} \log_2 ||\varphi_Y|| \leq R_2 + \delta$

Decoder: $\psi(\varphi_X(X^n), \varphi_Y(Y^n))$ Estimator: $\hat{\theta}(\psi(\varphi_X(X^n), \varphi_Y(Y^n)))$

Problem: $$\min_{\varphi_X, \varphi_Y, \psi, \hat{\theta}} \lim_{n \to \infty} nV\left(\hat{\theta}(\psi(\varphi_X(X^n), \varphi_Y(Y^n)))\right) = V(\theta | R_1, R_2)$$

Lossless/Lossy Compression and Estimation

Data: Observe sample (x^n, y^n) with marginal and joint types denoted by $\lambda_x, \lambda_y, \lambda_{xy}$.

Lossless: estimator $\tilde{\theta}(x^n, y^n) = \tilde{\theta}(\lambda_x, \lambda_y, \lambda_{xy})$, with types restricted to $k = 1, \cdots, K$ and $l = 1, \cdots, L$ (m-coordinates).

Lossy: Defined by a *stochastic mapping*,

$$\text{test-channels} \stackrel{def}{=} \begin{array}{l} p_\theta(u|x) : \mathcal{X} \to \mathcal{U} \\ p_\theta(v|y) : \mathcal{Y} \to \mathcal{V} \end{array}$$

where $\mathcal{U} = 0, \cdots, N$, $\mathcal{V} = 0, \cdots, M$. $U \to X \to Y \to V$ form a Markov chain.

Test-channel output: the *codewords* U^n, V^n

Compression ctd

- Separate compression: $p_\theta(u,v|x,y) = p_\theta(u|x)p_\theta(v|y)$
- θ is unknown, can't have the test-channels depend on it
- fixed test-channels inefficient approach

Universal Coding scheme, fix the test-channels for each data type:

$$p_{\lambda_x}(u|x), \;\; p_{\lambda_y}(v|y)$$

For large n, we can approximate the test-channels by the *conditional types*: $\tilde{p}_{\lambda_x}(u|x) = \frac{\lambda ux}{\lambda_x}, \tilde{p}_{\lambda_y}(v|y) = \frac{\lambda vy}{\lambda_y}$

Lossy Compression:

$$(X^n, Y^n) \overset{R_1,R_2}{\longrightarrow} (U^n, V^n) \overset{m-coord}{=} (\lambda_u, \lambda_v, \lambda_{uv})$$
$$\overset{0-rate}{\longrightarrow} (\lambda_x, \lambda_y)$$

Estimation based on lossy compression: $\hat{\theta}(\lambda_x, \lambda_y, \lambda_{uv})$

The Problem: Find bounds on asymptotic efficiency of $\hat{\theta}(\lambda_x, \lambda_y, \lambda_{uv})$

Zhang and Berger (1988)

- Encoder ensemble moments for additive estimators
- Restrict $\hat{\theta}(\lambda_x, \lambda_y, \lambda_{uv})$ to a linear functional of $\tilde{\theta}(\lambda_x, \lambda_y, \lambda_{xy})$
- Simplicity: single-letter characterization

Han and Amari (1995, 1998)

- Asymptotic distribution of $(\lambda_x, \lambda_y, \lambda_{uv}) \rightarrow$ MLE, Cramer-Rao bound
- Universal *decoding* argument
- Complexity: uncomputable for sources $>$ binary

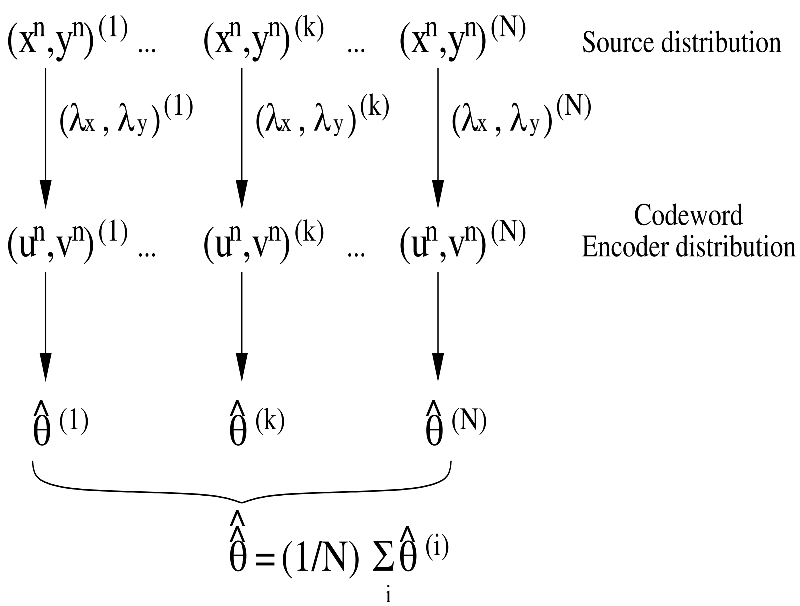
Zhang and Berger

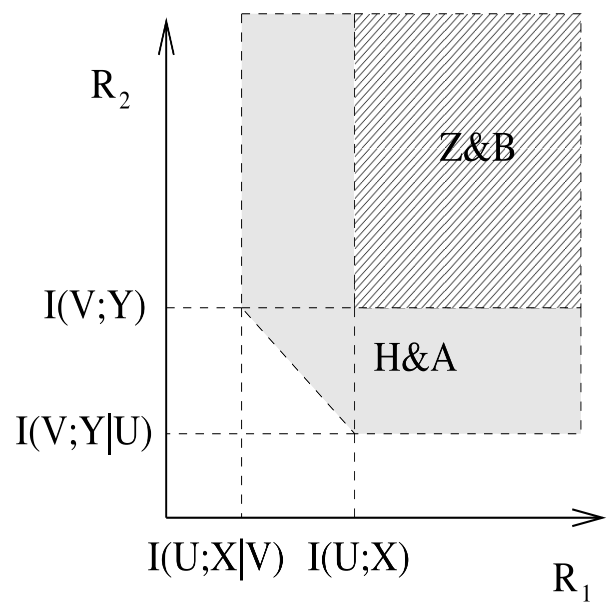
There exists a code (f, g) with rates (R_1, R_2) s.t.

$$nV[\hat{\theta}_{\lambda_x,\lambda_y}(f(X^n), g(Y^n))] \leq V[\hat{\theta}(U,V)] + E[E(\tilde{\theta}(X,Y)|X)]^2 + E[E(\tilde{\theta}(X,Y)|Y)]^2 - E[E(\hat{\theta}(U,V)|UX)]^2 - E[E(\hat{\theta}(U,V)|VY)]^2 + O(n^{-1/2})$$

Rate compatibility conditions: $R_1 \geq I(U;X)$ and $R_2 \geq I(V;Y)$

Key: Compute *ensemble moments* over random codebooks.





Han and Amari 1995,1998

$(\lambda_x, \lambda_y, \lambda_{uv}) = \bar{\Lambda}_f$: HA show $\bar{\Lambda}_f \underset{n\to\infty}{\sim} N(\bar{p}_{\theta,f}, H_f G H_f^T)$

There exists a code (f, g) with rates R_1, R_2 and a decoder h s.t.

$$nV[\hat{\theta}_{\lambda_x,\lambda_y}(h(f(X^n), g(Y^n)))) \leq \left\{\nabla\bar{p}_{\theta,f}(H_f G H_f^T)^{-1}\nabla\bar{p}_{\theta,f}^T\right\}^{-1} + O(n^{-1/2})$$

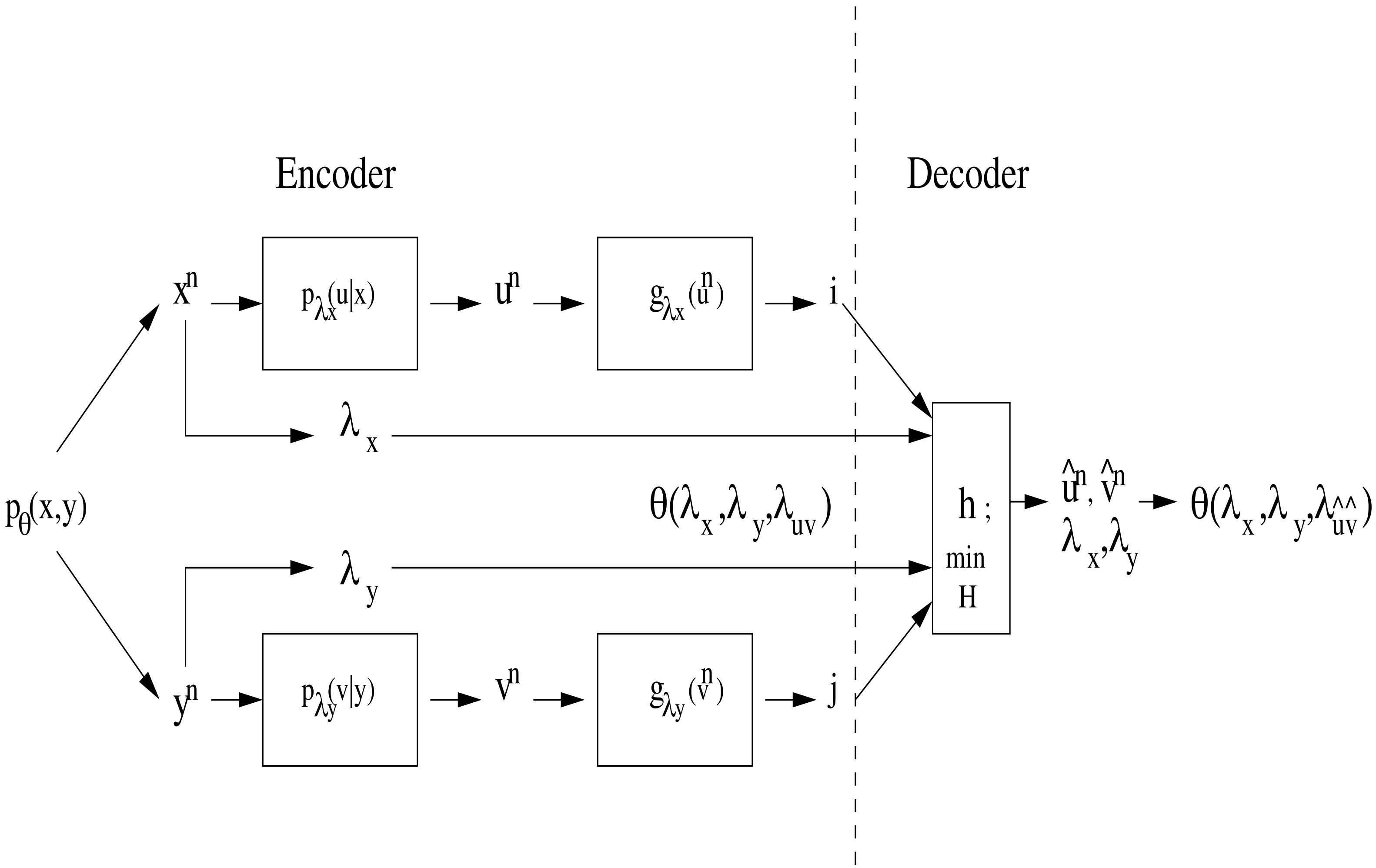
Rate compatibility conditions:

$R_1 \geq I(U;X|Y)$,$R_2 \geq I(V;Y|X)$,$R_1 + R_2 \geq I(UV;XY)$

Codebook constraint on type distribution Λ_f enters via projection operator $H_f : \sum_{uxyv} \lambda_{uxyv} = 1,\ \lambda_{ux} = \tilde{p}(u|x)\lambda_x,\ \lambda_{vy} = \tilde{p}(v|y)\lambda_y$

Computing H_f in terms of the m-coordinates is a non-trivial exercise.

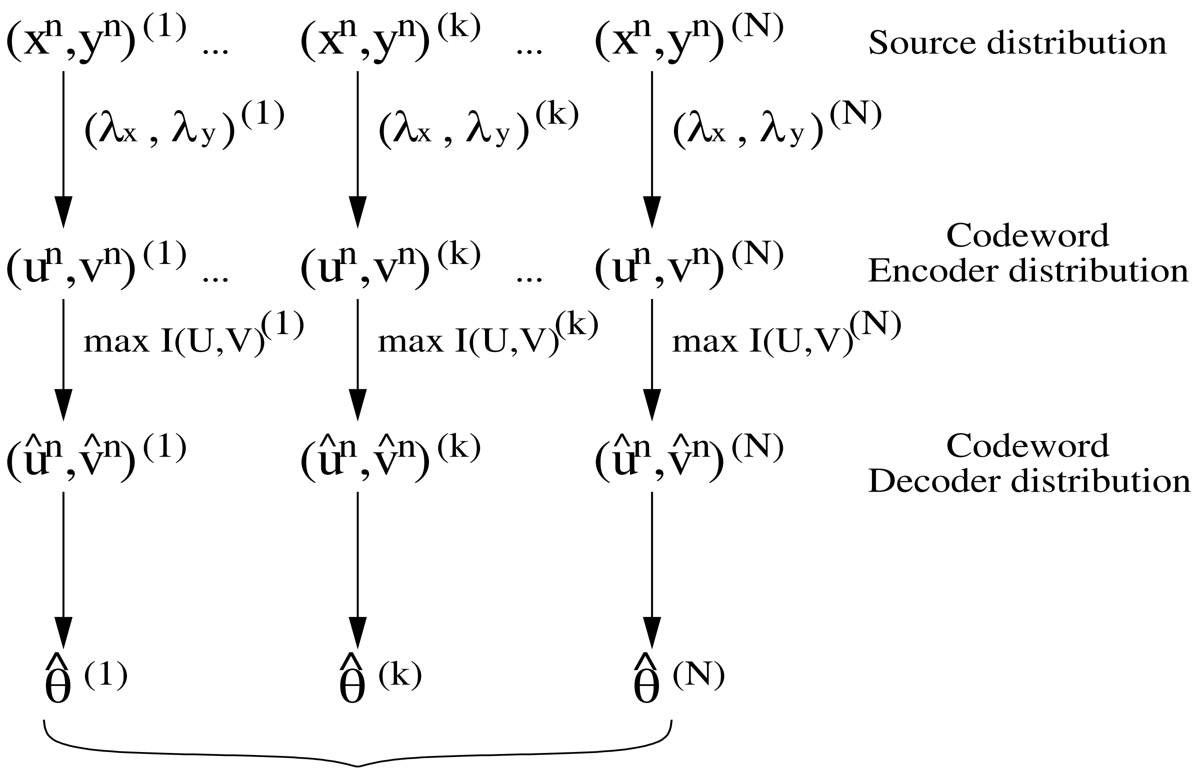
Han and Amari



Extending Zhang and Berger

- Push *Ensemble moment* computation to the decoder side.
- Use test-channels that map from/to exponential families.
- Consider estimates $\hat{\theta}$ that are "nice" functionals of mean-value parameters (additive and asympt. unbiased MLEs)
- A delta-method argument gives an *approximation* of the Cramer-Rao bound of Han and Amari, and
  1. it's *computable* for large source alphabets
  2. it has a *single-letter characterization* (suffices to compute moments under $p_\theta(u, x, y, v)$, cmp. $p_\theta(\Lambda_f)$).

Extending Zhang and Berger



$$\hat{\hat{\theta}} = (1/N) \sum_i \hat{\theta}^{(i)}$$

New upper bound on $V(\theta|R_1, R_2)$

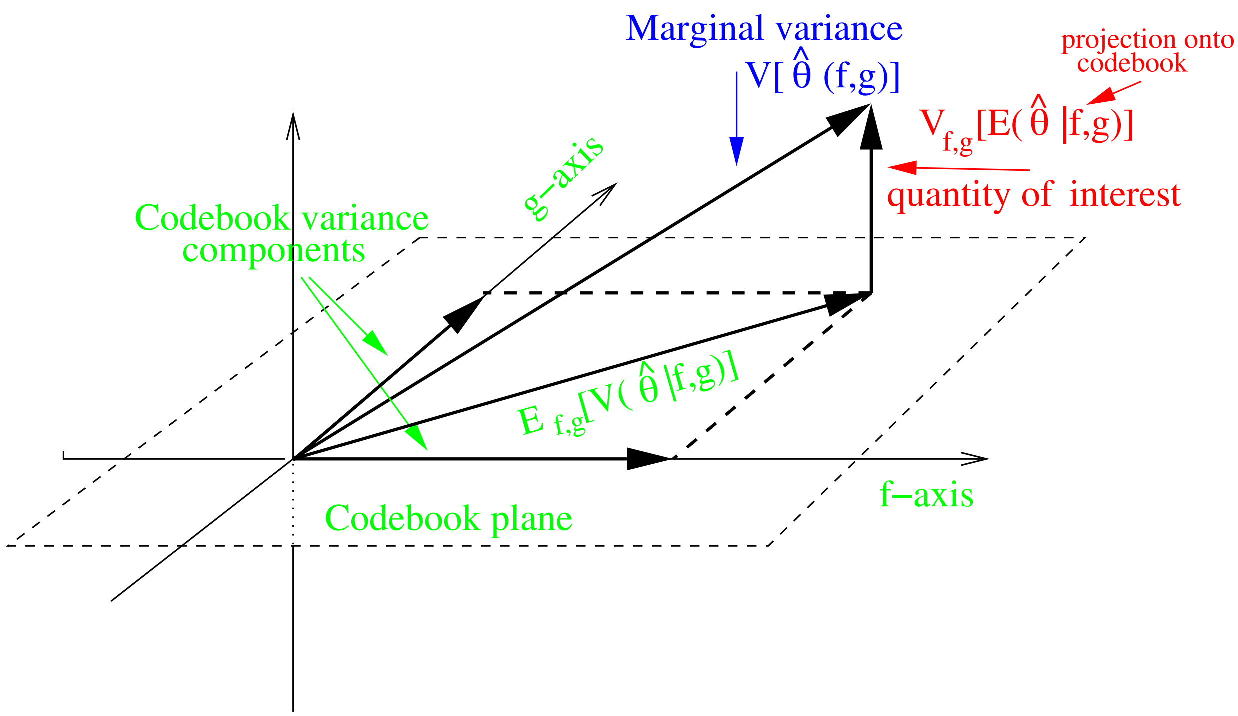
There exists a code (f, g) with rates R_1 and R_2 and a decoder h s.t.

$$
\begin{aligned}
nV[\hat{\theta}(h(f(X^n), g(Y^n)) &\leq V[\hat{\theta}(U,V)] \\
&- E[E(\hat{\theta}(U,V)|UX) - E(\hat{\theta}(U,V)|X)]^2 \\
&- E[E(\hat{\theta}(U,V)|VY) - E(\hat{\theta}(U,V)|Y)]^2 + O(n^{-1/2}) \\
&\underset{exp.\ family}{\overset{orthog.\ codebooks}{=}} V[\hat{\theta}(U,V)] - E_{XY}\left(V_{UV|XY}[\hat{\theta}(U,V)]\right) \\
&\underset{decomp.}{\overset{variance}{=}} V_{XY}\left(E_{UV|XY}[\hat{\theta}(U,V)]\right)
\end{aligned}
$$

Rate compatibility conditions: same as Han and Amari.

Geometric interpretation

$$nV[\hat{\theta}(h(f(X^n), g(Y^n)) \leq V[\hat{\theta}(U,V)] - E_{XY}\left(V_{UV|XY}[\hat{\theta}(U,V)]\right)$$



Or in codebook notation,

$$V[\hat{\theta}(f,g)] - E_{f,g}[V(\hat{\theta}(f,g)|f,g)] = V_{f,g}[E(\hat{\theta}(f,g)|f,g)]$$

Estimation of the crosscorrelation between two separately compressed sources.

Bivariate gaussian model: $p_\theta(x^n, y^n), \theta = (\sigma_x^2, \sigma_y^2, \rho)$

Test-channels: $p(u|x) \sim N(x, \sigma_n^2)$

$$p(v|y) \sim N(x, \sigma_m^2)$$

Bound still applicable for cts sources if we *discretize* (U, X, Y, V).

New bound $\Longrightarrow$ $nV(\hat{\rho}(h(f(X^n), g(Y^n))) \leq$

$$(1-\rho^2)^2 + \left(\frac{1}{2^{2R_1}-1} + \frac{1}{2^{2R_2}-1}\right) + \left(\frac{1}{2^{2R_1}-1}\frac{1}{2^{2R_2}-1}\right)$$
$$-\rho^2\left(\frac{1}{2^{2R_1}-1} + \frac{1}{2^{2R_2}-1} + \frac{1}{2^{2R_2}(2^{2R_1}-1)} + \frac{1}{2^{2R_1}(2^{2R_2}-1)}\right)$$

Compare with Zhang and Berger (stronger rate constraints):

$$nV(\hat{\rho}(f(X^n), g(Y^n)) \leq (1+\rho^2) + (1-\rho^2)\left[\frac{1}{2^{2R_1}-1} + \frac{1}{2^{2R_2}-1}\right]$$
$$+\frac{1}{2^{2R_1}-1}\frac{1}{2^{2R_2}-1}$$

Conclusion

- Simple extension of Zhang and Bergers approach
- Same rate compatibility conditions as Han and Amari
- Computable for most sources
- Has a geometric interpretation

Future Work:

- "Nice" functionals may not correspond to the optimal estimators $\hat{\theta}$
- Best approximating exponential family
- Dependent case