Correlations in Populations: Information-Theoretic Limits

Don H. Johnson

Ilan N. Goodman

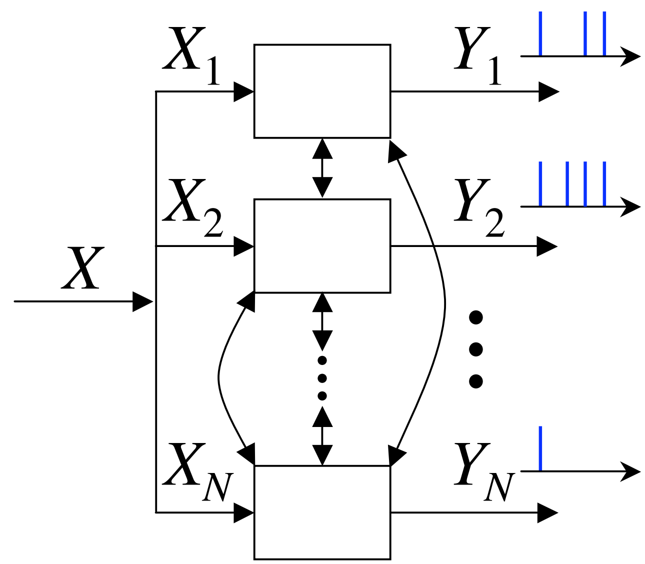
dhj@rice.edu

Department of Electrical & Computer Engineering

Rice University, Houston, Texas

Population coding

- Describe a population as parallel point process channels
- Variations
  - Separate inputs
  - Common input
  - Dependence among channels
- What do information theoretic considerations suggest is best?



Modeling approach

- We would like to use point process models for the outputs
  - Technically *very* difficult to describe connection-induced dependencies
  - Use simpler Bernoulli models, capable of describing complex correlation structures
- Assume homogeneous populations

$$P(X_1, X_2, \ldots, X_N) = P(X_1)P(X_2)\cdots P(X_N)$$

$$P(X_j = 1)$$

A note on modeling

- Correlation, orthogonal model

$$P(X_1, X_2, \ldots, X_N) =$$

$$P(X_1)P(X_2)\cdots P(X_N)\left[1 + \sum_{i>j} \frac{\rho^{(2)} \cdot (X_i - p_i)(X_j - p_j)}{\sqrt{p_i(1-p_i)p_j(1-p_j)}}\right.$$

$$\left. + \sum_{i>j>k} \frac{\rho^{(3)} \cdot (X_i - p)(X_j - p)(X_k - p)}{\sqrt{p_i(1-p_i)p_j(1-p_j)p_k(1-p_k)}}\right]$$

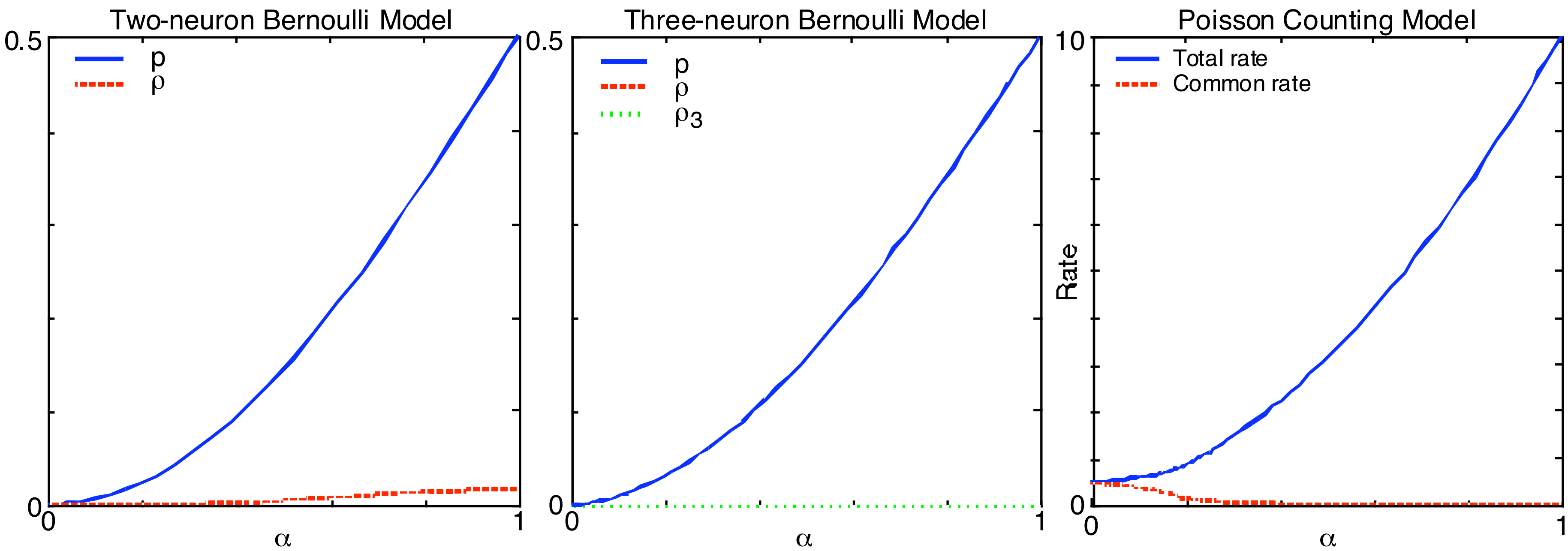
- Exponential model

$$P(X_1, X_2, \ldots, X_N) \propto$$

$$\exp\left\{\sum_i \theta_i X_i + \sum_{i,j} \theta_{ij} X_i X_j + \sum_{i,j,k} \theta_{ijk} X_i X_j X_k + \cdots\right\}$$

Fisher information analysis

- How should the stimulus be encoded in spike rate to achieve *constant* Fisher information?
- Input structure not important



Kullback-Leibler distance and data analysis

$$D_X(\alpha_1 \| \alpha_0) = \sum_x p(x;\alpha_1) \log \frac{p(x;\alpha_1)}{p(x;\alpha_0)}$$

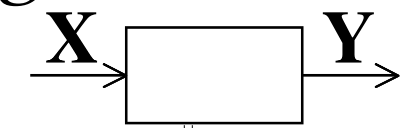
- α_0, α_1 two different stimulus conditions
- $p(x;\alpha)$ - response probabilities
- K-L distance is the "exponential rate" of a Neyman-Pearson classifier's false-alarm probability

$$P_F \sim 2^{-N D_X(\alpha_1 \| \alpha_0)} \text{ for fixed } P_M$$

- Distance resulting from information perturbations is proportional to Fisher information

$$D_X(\alpha_0 + \delta\alpha \| \alpha_0) \propto F(\alpha_0) \cdot (\delta\alpha)^2$$

Data Processing Theorem Redux



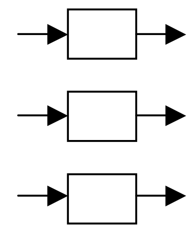
$$\gamma_{\mathbf{X},\mathbf{Y}}(\alpha_0,\alpha_1) = \frac{D_{\mathbf{Y}}(\boldsymbol{\alpha}_1 \| \boldsymbol{\alpha}_0)}{D_{\mathbf{X}}(\boldsymbol{\alpha}_1 \| \boldsymbol{\alpha}_0)}$$

$$0 \le \gamma_{\mathbf{X},\mathbf{Y}}(\alpha_0,\alpha_1) \le 1$$

- "Systems cannot create information"
- Basis for a system theory for information processing and determining which structures are inherently more effective

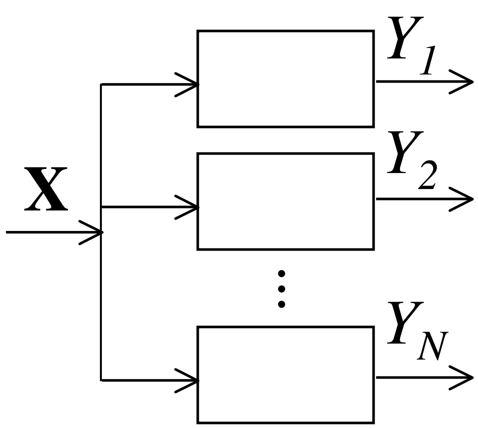
Population encoding properties from a K-L distance perspective

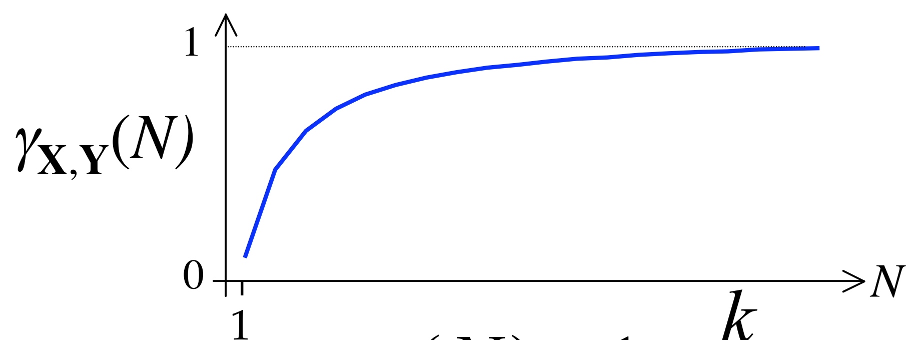
- Individual inputs don't necessarily achieve maximal information transfer



$$\gamma_{\mathbf{X},\mathbf{Y}}(N) \le \max_i \gamma_{X_i,Y_i}$$

- Explicitly indicating that the inputs encode a single quantity reveals that *perfect* fidelity is possible





$$\gamma_{\mathbf{X},\mathbf{Y}}(N) \approx 1 - \frac{k}{N}$$

or

$$\gamma_{\mathbf{X},\mathbf{Y}}(N) \approx 1 - e^{-kN}$$

Another viewpoint: Channel Capacity

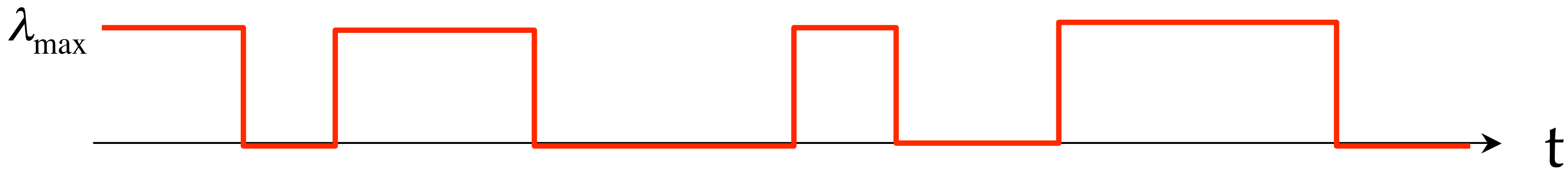
- Capacity for the *stationary* point process channel is known
  - If $0 \le \lambda_t \le \lambda_{\max}$ is the "power" constraint

$$C\ (\text{bits/s}) = \frac{\lambda_{\max}}{e \ln 2} = \frac{\lambda_{\max}}{1.88417}$$

  - If we additionally constrain average rate

$$C\ (\text{bits/s}) = \begin{cases} \dfrac{\lambda_{\max}}{e \ln 2}, & \bar{\lambda} > \lambda_o \\ \dfrac{\bar{\lambda}}{\ln 2} \ln \dfrac{\lambda_{\max}}{\bar{\lambda}}, & \bar{\lambda} < \lambda_o \end{cases}, \quad \lambda_o = \frac{\lambda_{\max}}{e}$$

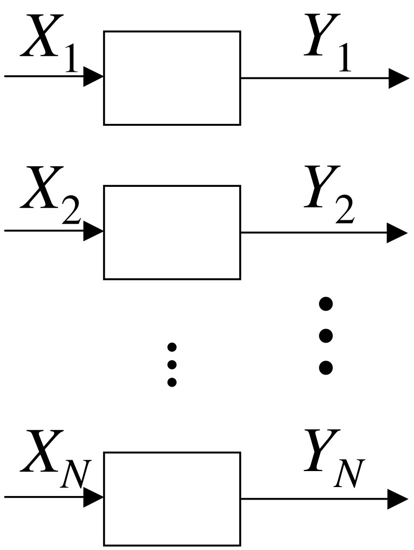
- Capacity achieved by a Poisson process driven by a random telegraph wave

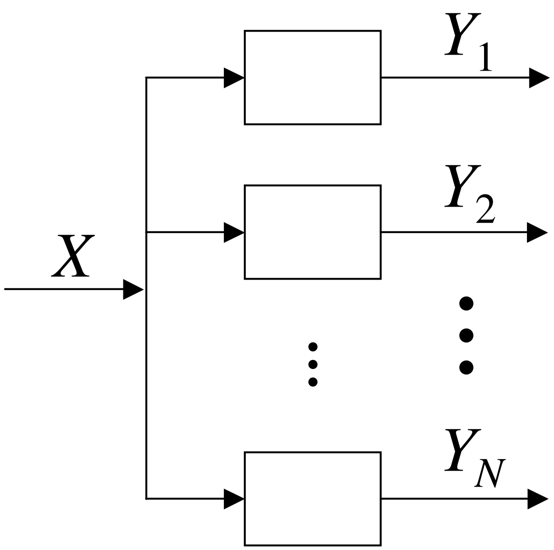


Channel capacity of populations

- Use a Bernoulli model and investigate the small probability limit to determine capacity for parallel *Poisson* channels
- The two input structures have the *same* capacity

$$C^{(N)} = NC^{(1)} = N \cdot \frac{\lambda_{\max}}{e \ln 2}$$



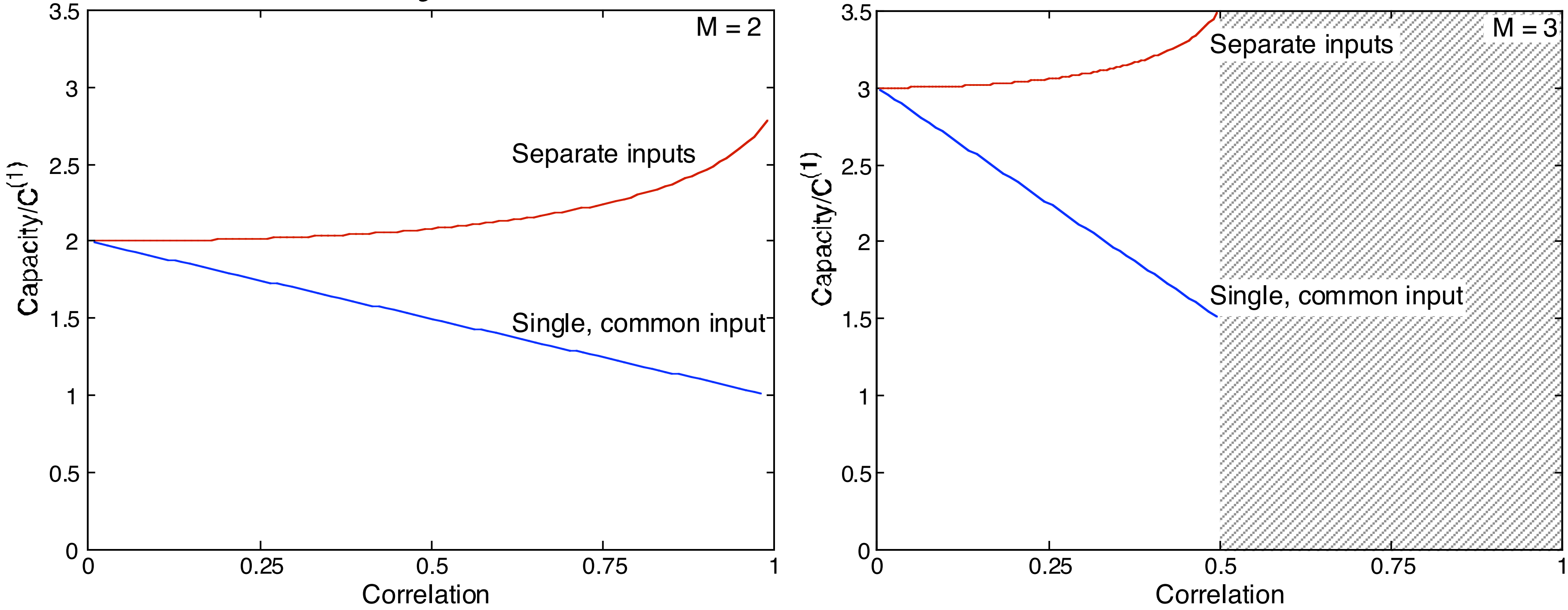


Imposing connection dependence changes the story

- Using Bernoulli models, connection dependence can be added
- Caveat: modeling Poisson processes
- Interesting restrictions arise
  - Capacity depends *only* on pairwise correlations (dependencies)
  - Only *positive* pairwise correlations possible
  - Restricted range of correlation values
    - For homogenous populations: $0 \leq \rho \leq \dfrac{1}{N-1}$
    - For inhomogenous populations: $0 \leq \rho \leq \rho_{\mathrm{max}}$

Capacity results

- Capacity achieved with a homogeneous population
- Correlation affects the two input structures differently



- Qualitatively similar to Gaussian channel results

However…

- As population size increases, introducing connection-induced dependence reduces capacity

⇨ *Capacity unaffected by input- or connection-induced dependence*

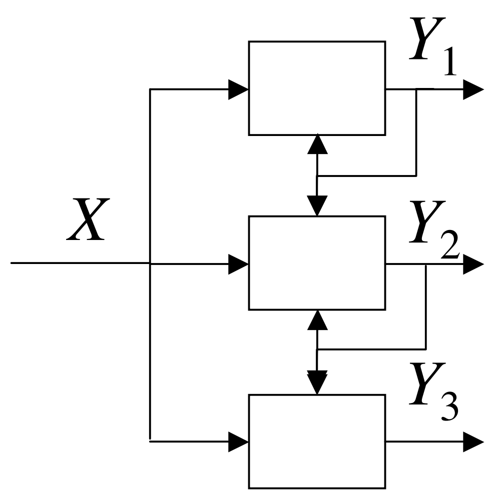
- Fits with previous results derived using Fisher information

Poisson vs. Non-Poisson Models

- Results derived using a Poisson assumption
- How about non-Poisson models?
- Probably impossible to extend Bernoulli approach to interesting non-Poisson cases, but…
- Kabanov showed that the single-channel Poisson capacity bounded the capacity of all other point process models
- Does this bound apply to multi-channel processes as well?

Connection-induced dependence

- Bernoulli model vague about how correlations are induced
- If internal feedback is used…
  - Feedback can increase capacity
  - M. Lexa has shown that internal feedback can increase the performance of distributed classifiers



Conclusions

- From two theoretical viewpoints, connection-induced dependence not required to increase capacity
- Specific forms of dependence *may* increase a population's processing power
- Capacity afforded by non-Poisson models *probably* bounded by Poisson result, but not in detail