# Extremal Distributions in Information Theory and Hypothesis Testing

Sean Meyn

Department of Electrical and Computer Engineering
University of Illinois
and the Coordinated Science Laboratory

Joint work with Charuhas Pandit, Jianyi Huang, and Venu Veeravalli

NSF support: ECS 02-17836, ITR 00-85929 and CCF 00-49089

# Neyman Pearson Hypothesis Testing

Observations $\boldsymbol{X} = \{X_t : t = 1, 2, \ldots N\}$

$X_t$ takes values in finite alphabet $\mathcal{A}$

$X$ i.i.d. with marginal $\pi_j$ under $H_j$, $j = 0, 1$

Hypothesis test:

$$\phi(x) = 1 \text{ if } x \in A_1 \subset \mathcal{A}^N$$

Error Probabilities

$$\mathsf{P}_{\mathsf{e},0} = \mathsf{P}_0\{\phi(X) = 1\}, \quad \mathsf{P}_{\mathsf{e},1} = \mathsf{P}_1\{\phi(X) = 0\}$$

N-P Criterion: $\inf_{\phi} \mathsf{P}_{\mathsf{e},1}$ subject to $\mathsf{P}_{\mathsf{e},0} \leq \eta$

# Likelihood Ratio Test

Likelihood ratio

$$\ell(a) = \frac{\pi_{1,a}}{\pi_{0,a}}, \quad a \in \mathcal{A}$$

Optimum test (LRT)

$$A_1 = \left\{ x : \frac{1}{N} \sum_{t=1}^{N} \log \ell(X_t) \ \geq \ \lambda \right\}$$

Empirical distribution of $x$

$$\tilde{\mu}_a = \frac{1}{N} \sum_{t=1}^{N} \mathbb{I}_{\{X_t = a\}}$$

Alternative form for LRT:

$$A_1 = \{ x : \langle \tilde{\mu}, \log \ell \rangle \ \geq \ \lambda \}$$

# Robust Hypothesis Testing

$\pi_0$ and $\pi_1$ not known exactly

Uncertainty classes

$\pi_0 \in \mathbb{P}_0 \qquad \pi_1 \in \mathbb{P}_1$

Robust N-P Criterion

$$\inf_{\phi} \sup_{\pi_1 \in \mathbb{P}_1} \mathsf{P}_{e,1} \text{ subject to } \sup_{\pi_0 \in \mathbb{P}_0} \mathsf{P}_{e,0} < \eta$$

Robust test can be found for special uncertainty classes

$\epsilon$-contamination classes, band classes,
T-V neighborhoods, $p$-point classes, etc.

[Huber (65,73)]

Robust test is LRT between *least favorable distributions*

# Moment Classes

Uncertainty classes based on *fixing moments*

$$\mathbb{P}_j = \left\{ \pi : \langle \pi, f_i \rangle = c_i^j, i = 1, ..., n \right\}$$

Natural way to define uncertainty in practical scenarios (e.g. mean and variance constraints)

Markov (1890's)
Whitt (1976)
Smith (1995)
Bertsimas & Sethuraman (2000)
Vandenberghe, Boyd & Comanor (2004)

Applications in detection in sensor networks

*Robust hypothesis testing problem for moment classes does not fall under Huber framework*

# Asymptotic N-P Hypothesis Testing

Sequence of tests

$$\phi_N(x^N) = 1 \ \text{ if } x^N \in A_1^N$$

Number of observations $N \to \infty$

Error exponents:

$$\begin{aligned} \varepsilon_1 \ &:= \ -\liminf_{N\to\infty} \frac{1}{N} \log \mathsf{P}_1\{\phi_N(X^N) = 0\} \\ \varepsilon_0 \ &:= \ -\liminf_{N\to\infty} \frac{1}{N} \log \mathsf{P}_0\{\phi_N(X^N) = 1\} \end{aligned}$$

Asymptotic N-P problem:

$$\sup_{\phi} \varepsilon_1 \qquad \text{subject to} \qquad \varepsilon_0 \ge \eta$$

# Asymptotic N-P Hypothesis Testing

Optimal value of $\varepsilon_1$ :

$$\beta^* = \inf_{\mu \in \mathcal{Q}_\eta(\pi_0)} D(\mu \,\|\, \pi_1)$$

Hoeffding (1965)

Blahut (1974)

where

$$\mathcal{Q}_\eta(\pi_0) := \{\mu : D(\mu \,\|\, \pi_0) < \eta\}$$

Optimal test sequence

$$\begin{aligned} A_1^N &= \Big\{ x^N \in \mathcal{A}^N : \frac{1}{N}\sum_{t=1}^{N} \log \ell(x_t) \ \geq\ \eta - \beta^* \Big\} \\ &= \big\{ x^N \in \mathcal{A}^N : \langle \tilde{\mu}_N, \log \ell \rangle \ \geq\ \eta - \beta^* \big\} \\ &= \big\{ x^N \in \mathcal{A}^N : \langle \tilde{\mu}_N, \log \ell \rangle \ \geq\ \langle \mu^*, \log \ell \rangle \big\} \end{aligned}$$

# Asymptotic N-P Hypothesis Testing

Optimal value of $\varepsilon_1$ :

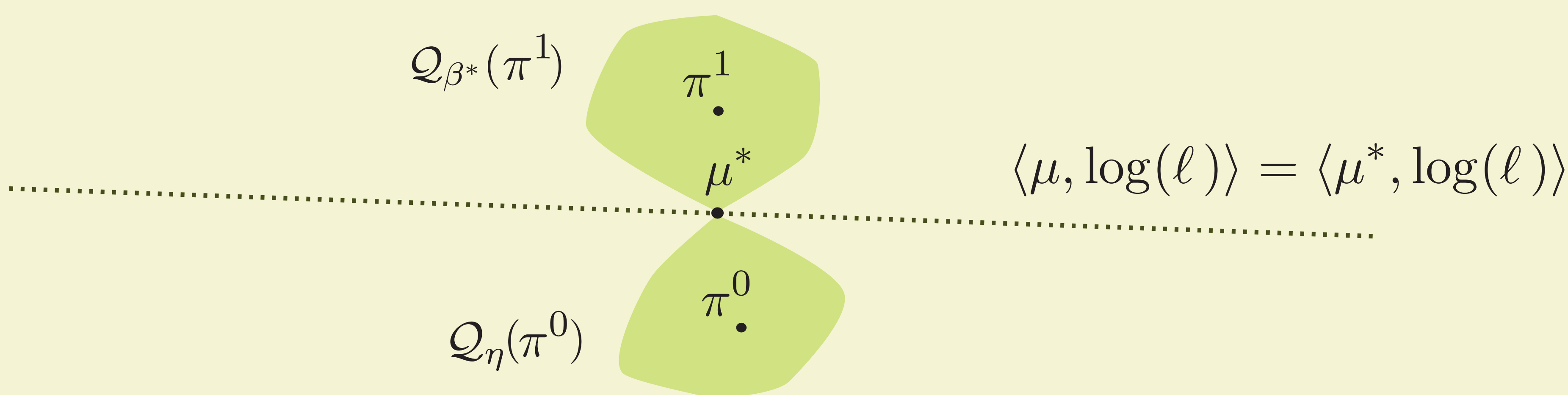
$$\beta^* = \inf_{\mu \in \mathcal{Q}_\eta(\pi_0)} D(\mu \parallel \pi_1)$$

Hoeffding (1965)

Blahut (1974)

where

$$\mathcal{Q}_\eta(\pi_0) := \{\mu : D(\mu \parallel \pi_0) < \eta\}$$



# Asymptotic Robust N-P Hypothesis Testing

## Moment Classes

$\pi_0$ and $\pi_1$ not known exactly: $\quad \pi_0 \in \mathbb{P}_0 \quad \pi_1 \in \mathbb{P}_1$

$$\mathbb{P}_j = \left\{ \pi : \langle \pi, f_i \rangle = c_i^j \,, \; i = 1, ..., n \right\}$$

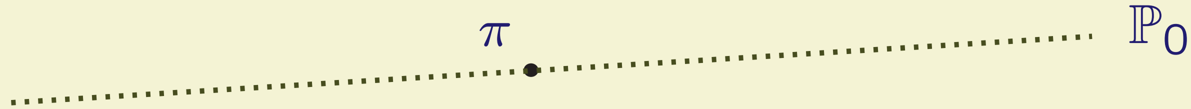
## Asymptotic Robust N-P Criterion

$$\sup_{\phi} \inf_{\pi_1 \in \mathbb{P}_1} \varepsilon_1 \;\; \text{subject to} \;\; \inf_{\pi_0 \in \mathbb{P}_0} \varepsilon_0 \; \geq \; \eta$$

# Divergence Sets

Divergence set containing $\mathbb{P}_0$ :

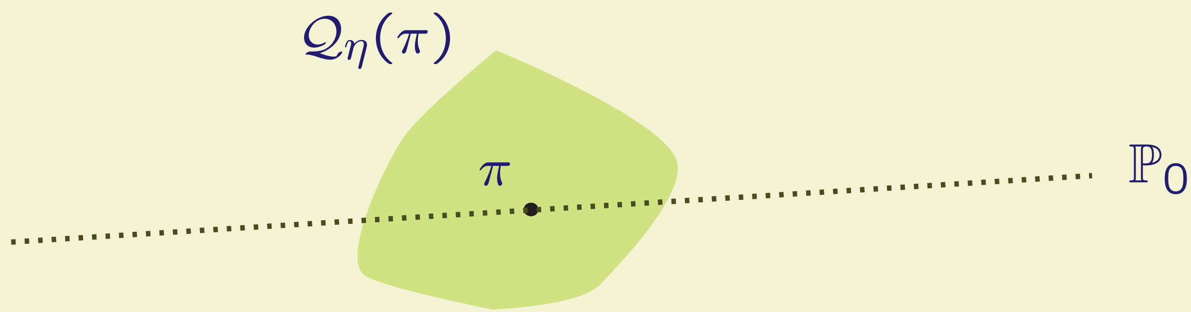
$$\mathcal{Q}_\eta(\mathbb{P}_0) = \bigcup_{\pi \in \mathbb{P}_0} \mathcal{Q}_\eta(\pi) = \bigcup_{\pi \in \mathbb{P}_0} \{\mu : D(\mu \,\|\, \pi) < \eta\}$$



# Divergence Sets

Divergence set containing $\mathbb{P}_0$ :

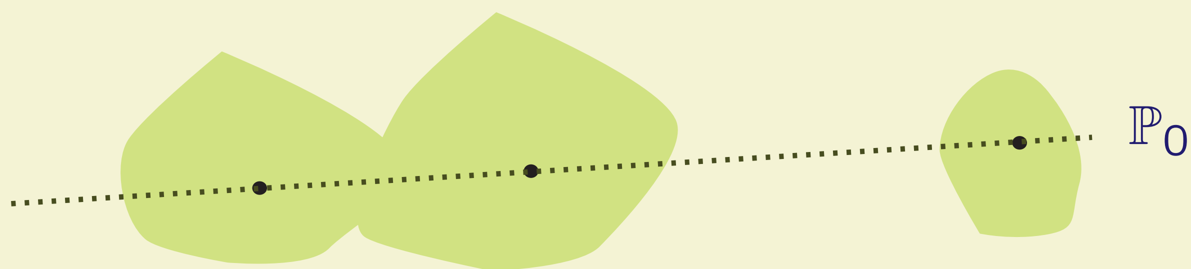
$$\mathcal{Q}_\eta(\mathbb{P}_0) = \bigcup_{\pi \in \mathbb{P}_0} \mathcal{Q}_\eta(\pi) = \bigcup_{\pi \in \mathbb{P}_0} \{\mu : D(\mu \parallel \pi) < \eta\}$$



# Divergence Sets

Divergence set containing $\mathbb{P}_0$ :

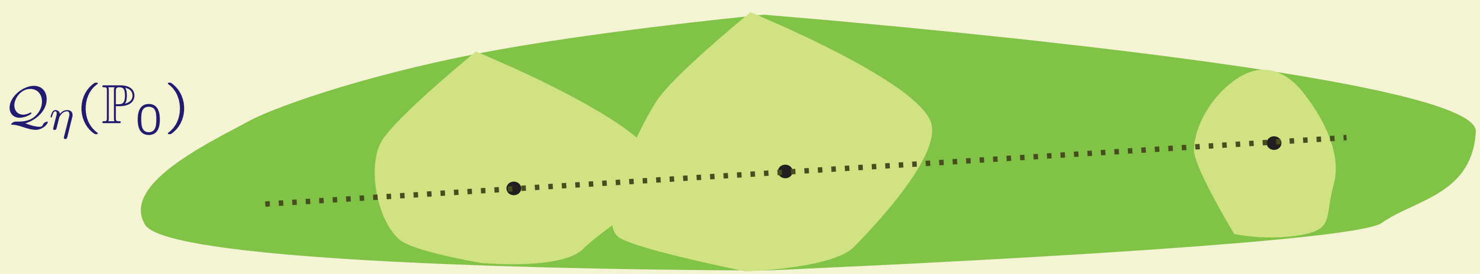
$$\mathcal{Q}_\eta(\mathbb{P}_0) = \bigcup_{\pi \in \mathbb{P}_0} \mathcal{Q}_\eta(\pi) = \bigcup_{\pi \in \mathbb{P}_0} \{\mu : D(\mu \,\|\, \pi) < \eta\}$$



# Divergence Sets

Divergence set containing $\mathbb{P}_0$ :

$$\mathcal{Q}_\eta(\mathbb{P}_0) = \bigcup_{\pi \in \mathbb{P}_0} \mathcal{Q}_\eta(\pi) = \bigcup_{\pi \in \mathbb{P}_0} \{\mu : D(\mu \parallel \pi) < \eta\}$$



# Solution to the Robust Neyman-Pearson Problem

Optimal value of $\varepsilon_1$ : There exist $\pi_0^* \in \mathbb{P}_0, \pi_1^* \in \mathbb{P}_1$, and $\mu^*$ solving,

$$\beta^* = \inf_{\pi^1 \in \mathbb{P}_1} \inf_{\mu \in \mathcal{Q}_\eta(\mathbb{P}_0)} D(\mu \,\|\, \pi^1)$$

Optimal robust test: LRT between $\pi_1^*$ and $\pi_0^*$ :

$$\ell(a) = \frac{\pi_{1,a}^*}{\pi_{0,a}^*} = \frac{\gamma^\top f(a)}{\lambda^\top f(a)}$$

Optimal robust test sequence

$$\begin{aligned} A_1^N &= \left\{ x^N \in \mathcal{A}^N : \langle \tilde{\mu}_N, \log \ell \rangle \ \geq \ \eta - \beta^* \right\} \\ &= \left\{ x^N \in \mathcal{A}^N : \langle \tilde{\mu}_N, \log \ell \rangle \ \geq \ \langle \mu^*, \log \ell \rangle \right\} \end{aligned}$$

# Solution to the Robust Neyman-Pearson Problem

Optimal value of $\varepsilon_1$: There exist $\pi_0^* \in \mathbb{P}_0, \pi_1^* \in \mathbb{P}_1$, and $\mu^*$ solving,

$$\beta^* = \inf_{\pi^1 \in \mathbb{P}_1} \inf_{\mu \in \mathcal{Q}_\eta(\mathbb{P}_0)} D(\mu \parallel \pi^1)$$

# Solution to the Robust Neyman-Pearson Problem

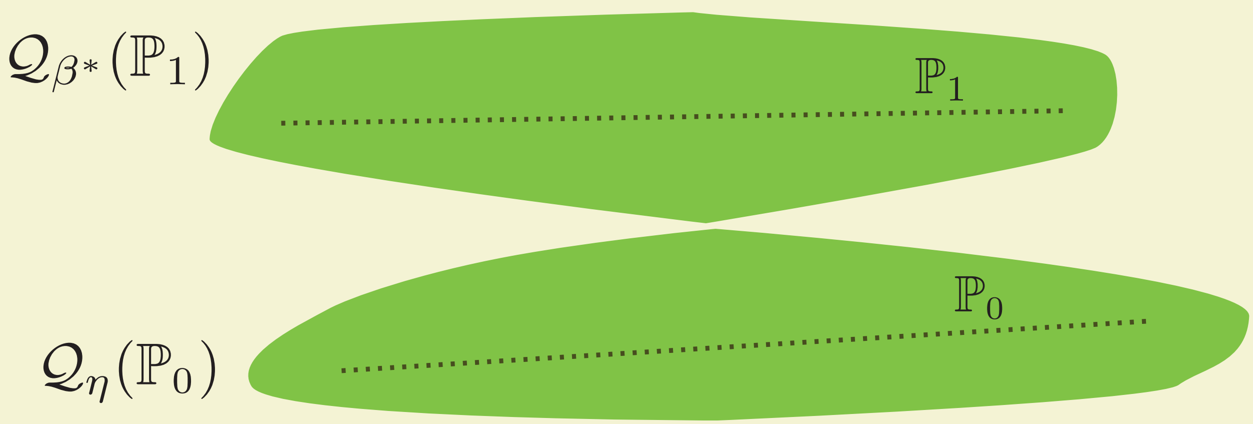
Optimal value of $\varepsilon_1$: There exist $\pi_0^* \in \mathbb{P}_0, \pi_1^* \in \mathbb{P}_1$, and $\mu^*$ solving,

$$\beta^* = \inf_{\pi^1 \in \mathbb{P}_1} \inf_{\mu \in \mathcal{Q}_\eta(\mathbb{P}_0)} D(\mu \parallel \pi^1)$$

# Solution to the Robust Neyman-Pearson Problem

Optimal value of $\varepsilon_1$: There exist $\pi_0^* \in \mathbb{P}_0, \pi_1^* \in \mathbb{P}_1$, and $\mu^*$ solving,

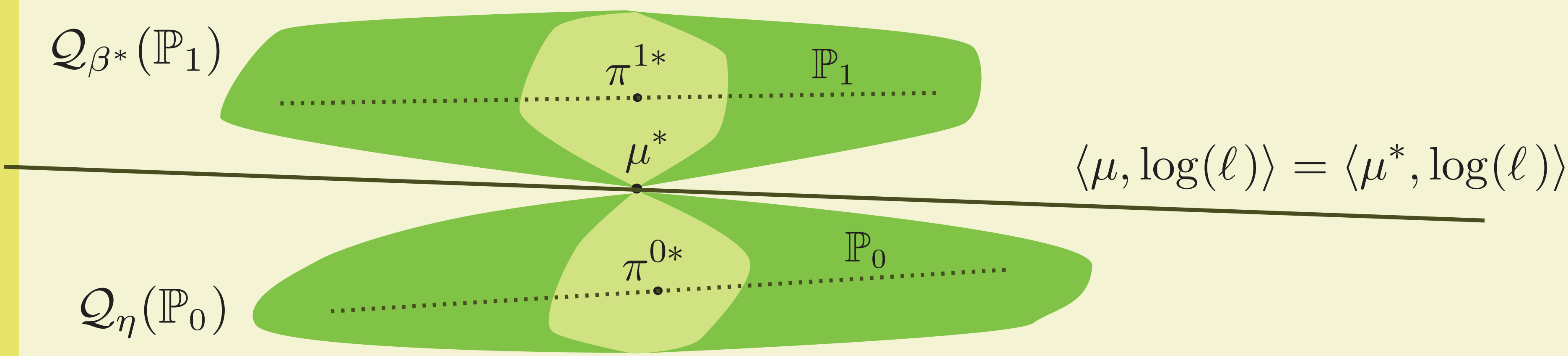
$$\beta^* = \inf_{\pi^1 \in \mathbb{P}_1} \inf_{\mu \in \mathcal{Q}_\eta(\mathbb{P}_0)} D(\mu \,\|\, \pi^1)$$



# Solution to the Robust Neyman-Pearson Problem

Optimal value of $\varepsilon_1$: There exist $\pi_0^* \in \mathbb{P}_0, \pi_1^* \in \mathbb{P}_1$, and $\mu^*$ solving,

$$\beta^* = \inf_{\pi^1 \in \mathbb{P}_1} \inf_{\mu \in \mathcal{Q}_\eta(\mathbb{P}_0)} D(\mu \parallel \pi^1)$$



# Solution Structure

$$\ell(a) = \frac{\pi^*_{1,a}}{\pi^*_{0,a}} = \frac{\gamma^\top f(a)}{\lambda^\top f(a)}$$

Robust test is based on a log-linear combination of data

Optimization is convex: Optimal $\pi^*_0$, $\pi^*_1$, $\mu^*$ and $\lambda$ and $\gamma$ easily computed numerically

Worst case distributions typically discrete: Cutting plane algorithm effective

# Conclusions & Extensions

| | |
|---|---|
| Immediate extensions: | $M$-*ary* hypothesis testing |
| | Non-finite alphabets |
| | Asymmetric moment classes |
| Current research: | Non-i.i.d. observations |
| | Optimizing choice of constraint functions |
| | Applications |

# References


C. Pandit, S.P. Meyn, V.V. Veeravalli, *Robust Neyman-Pearson Hypothesis Testing Based on Moment Classes*, 2004

C. Pandit and S.P. Meyn, *Robust Measurement-Based Admission Control Using Markov's Theory of Canonical Distributions*, 2003

C. Pandit, and S.P. Meyn, *Extremal Distributions and Worst-Case Large-Deviation Bounds*, 2004.

J. Huang and S.P. Meyn, *Characterization and Computation of Optimal Distributions for Channel Coding*.
To appear, IEEE Trans. Info. Theory.


*Preprints available at http://black.csl.uiuc.edu/~meyn*