Robust Latent Data Representations

Larry Samuelson[1] Jakub Steiner[2,3]

[1]Yale
[2]Zurich U
[3]Cerge-Ei

Jerusalem, October 2026

The Question

suppose data invalidate your statistical theory

how do you reason about

- the data
- vs data-generating process?

Berk-Nash analyst

- focuses on the data-generating process
- leaves the wedge between prediction and data unexplained

e.g. principal observes outputs of her workforce

- no combination of hidden efforts explains the output frequencies
- beliefs about individual and aggregate effort?

1 Sample Model

2 Axioms

3 MLE-Bayes Procedure

4 Representation

5 Stereotypes Application

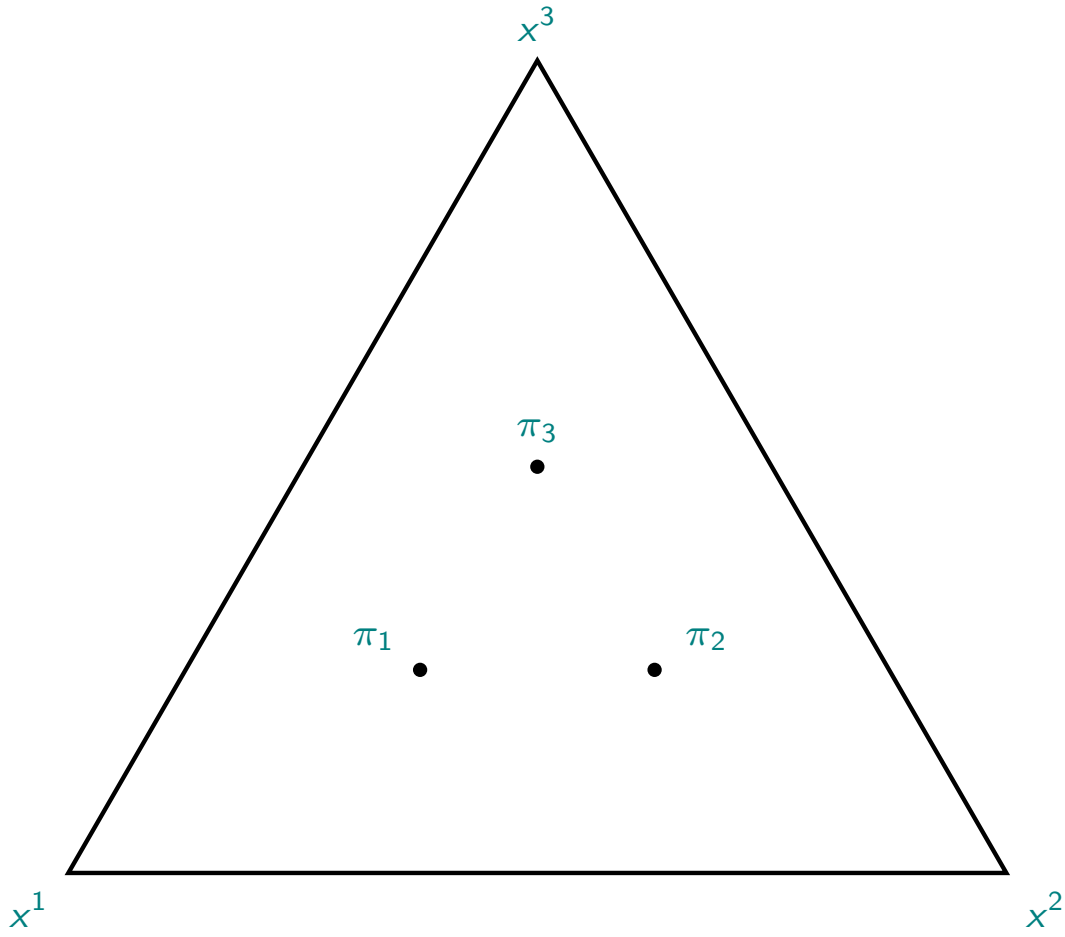
6 Moral-Hazard Application

7 Maximal Type Conservatism

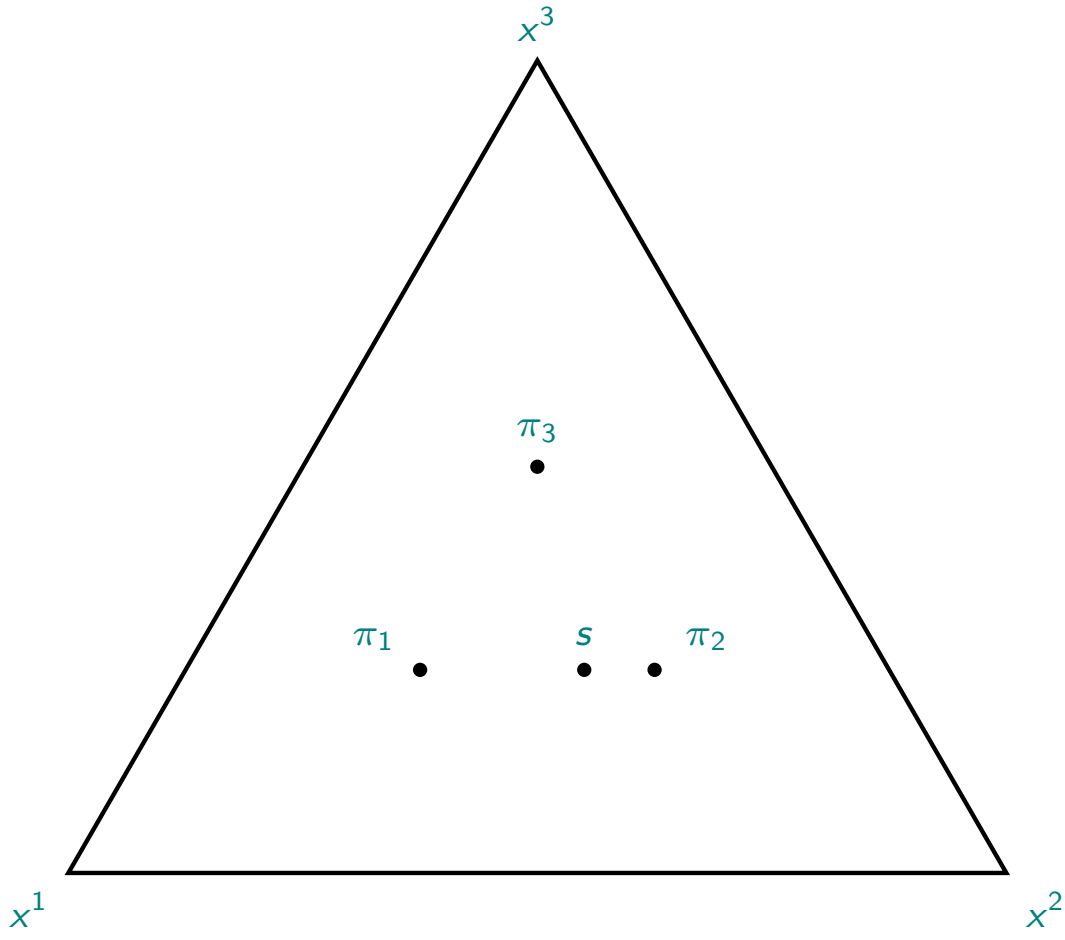
8 About Proof

9 Literature

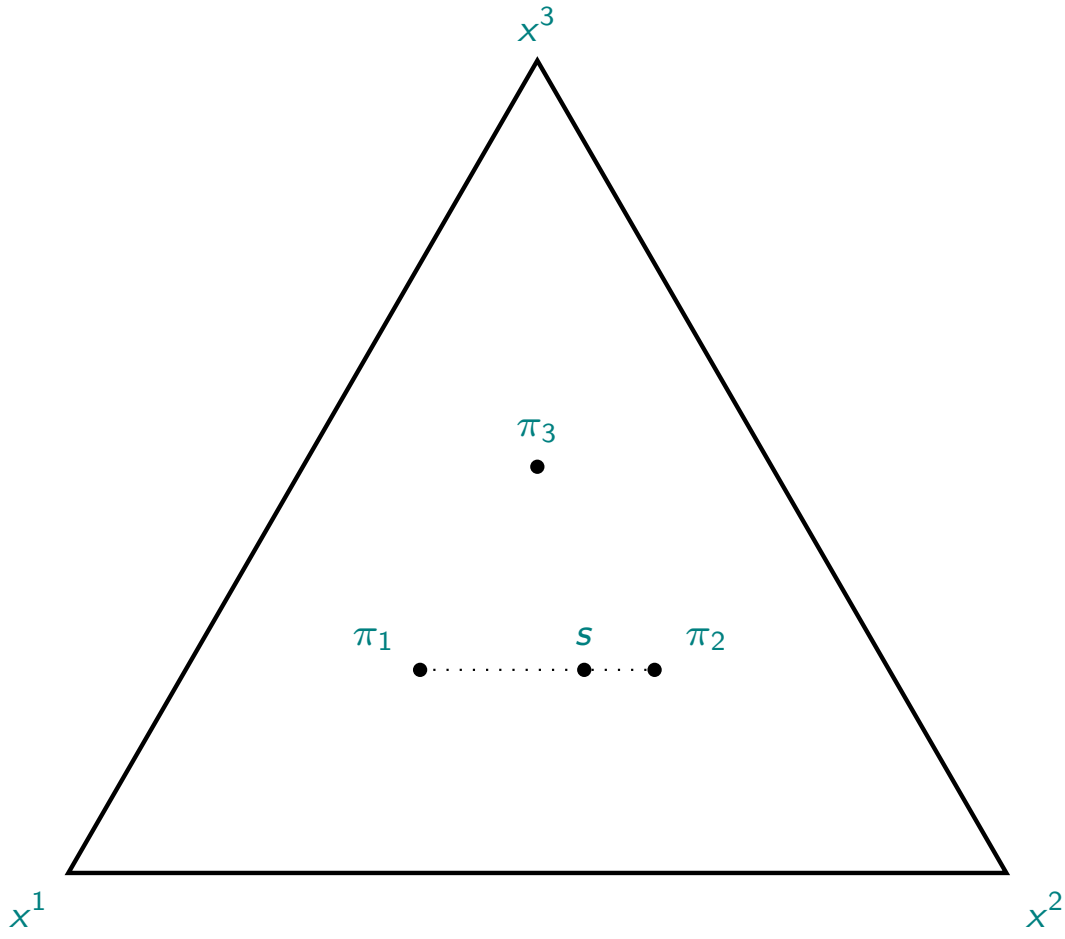
Sample and Primitive Types



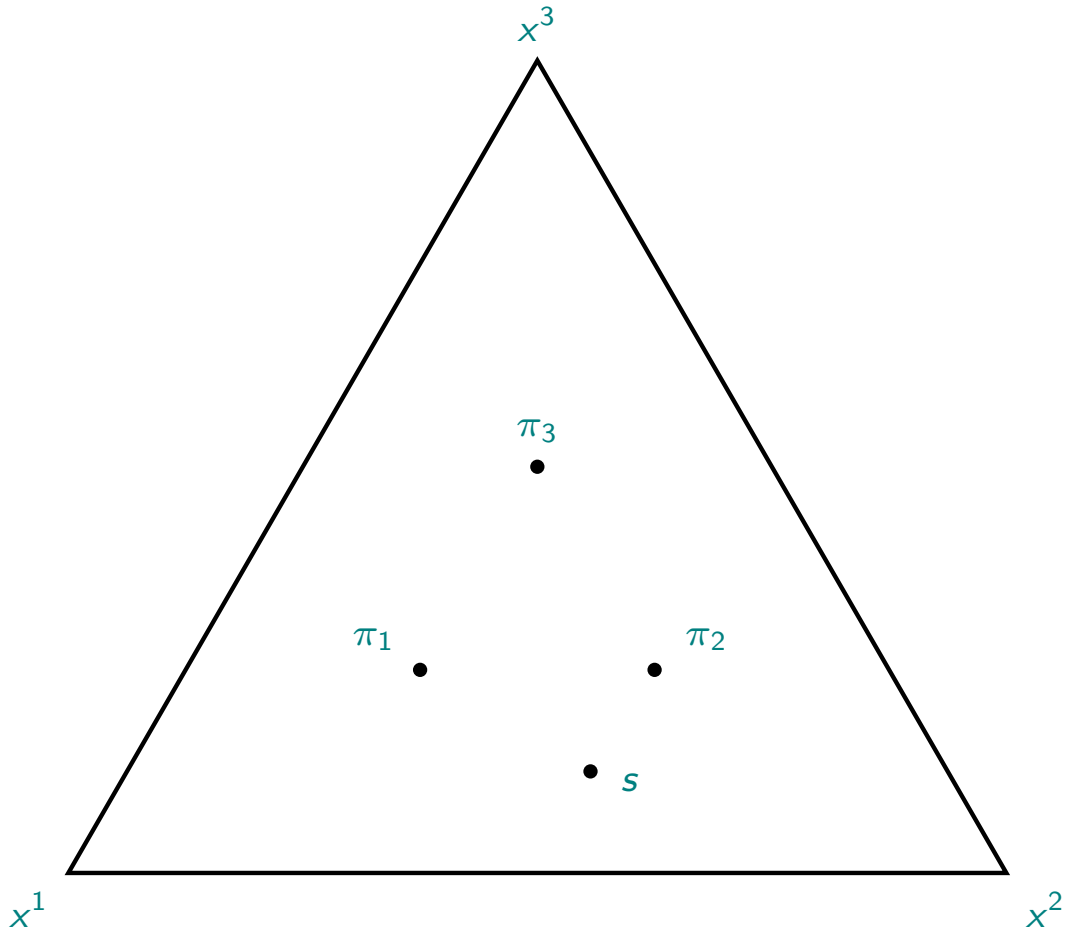
Sample and Primitive Types



Sample and Primitive Types



Sample and Primitive Types



Sample and Primitive Types

observations $x^1, \ldots, x^n$ from finite $\mathcal{X}$

sample $s(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x^i = x}, \qquad x \in \mathcal{X}$

- $s \in \operatorname{int} \Delta(\mathcal{X})$

primitive theory $\mathcal{T} = \{\pi_z\}_{z \in \mathcal{Z}}$, finite $\mathcal{Z}$

- π_z are primitive types (full support)
- affine independence

Definition

Sample Model

is the analyst's belief

$$q^s_{XZ} \in \Delta(\mathcal{X} \times \mathcal{Z})$$

about the representative sample element. It is the belief over

$$(x^i, z^i) = (x, z), \qquad \text{for uniformly random } i$$

e.g.: what were the effort and the output of my representative worker?

derived distributions:

- sample types $q^s_{X|Z=z}$; may differ from the primitive types π_z
- sample prior q^s_Z

1 Sample Model

2 Axioms

3 MLE-Bayes Procedure

4 Representation

5 Stereotypes Application

6 Moral-Hazard Application

7 Maximal Type Conservatism

8 About Proof

9 Literature

Realism

A1

$q_X^s = s$.

belief about the output matches the observed output distribution

Diagnostic Conservatism

A2

The analyst preserves the diagnostic likelihood ratio:

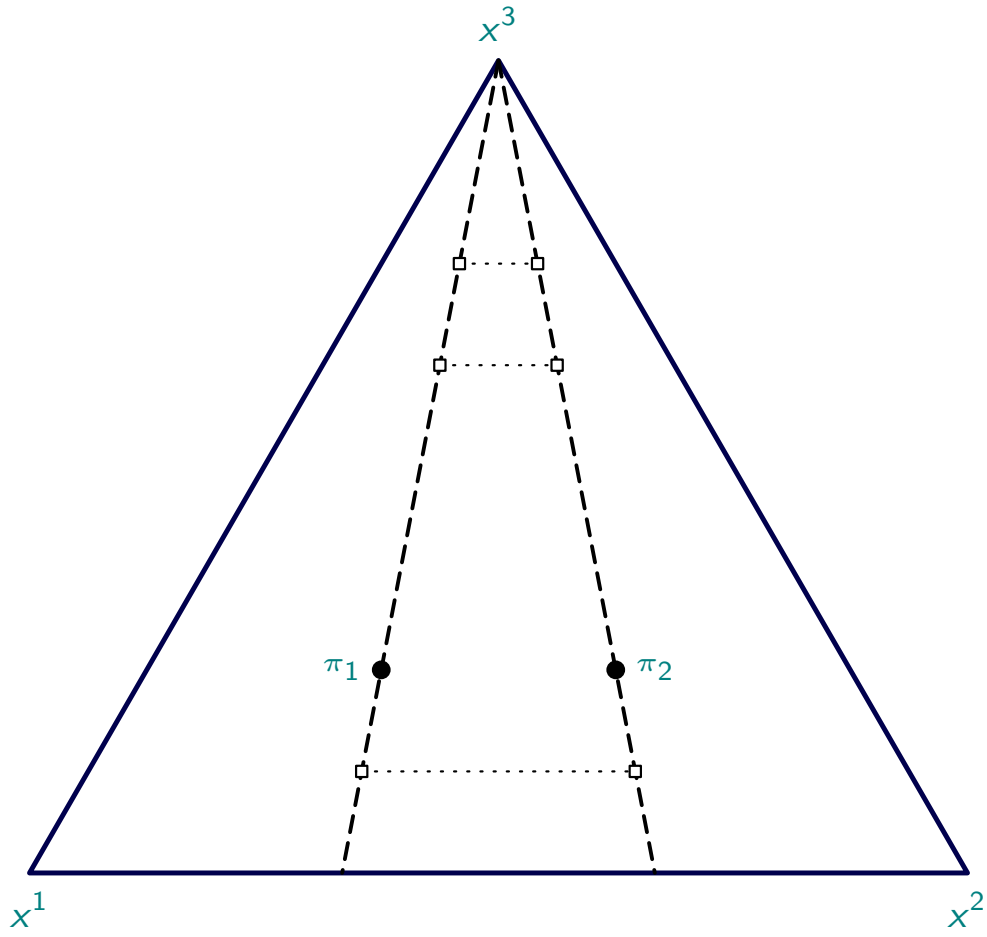
$$\frac{q^s_{X|Z=z}(x)}{q^s_{X|Z=z'}(x)} = \frac{\pi_z(x)}{\pi_{z'}(x)} \qquad \text{for all } z, z' \in \operatorname{supp} q^s_Z \text{ and } x \in \mathcal{X}.$$

e.g., the principal is rigid on how each x is diagnostic of high vs low effort

normative argument:

- realism may force type adjustments $q^s_{X|Z=z}(x) = \pi_z(x) d_z(x)$.
- since z isn't observed, the analyst chooses $d(x)$ independent of z

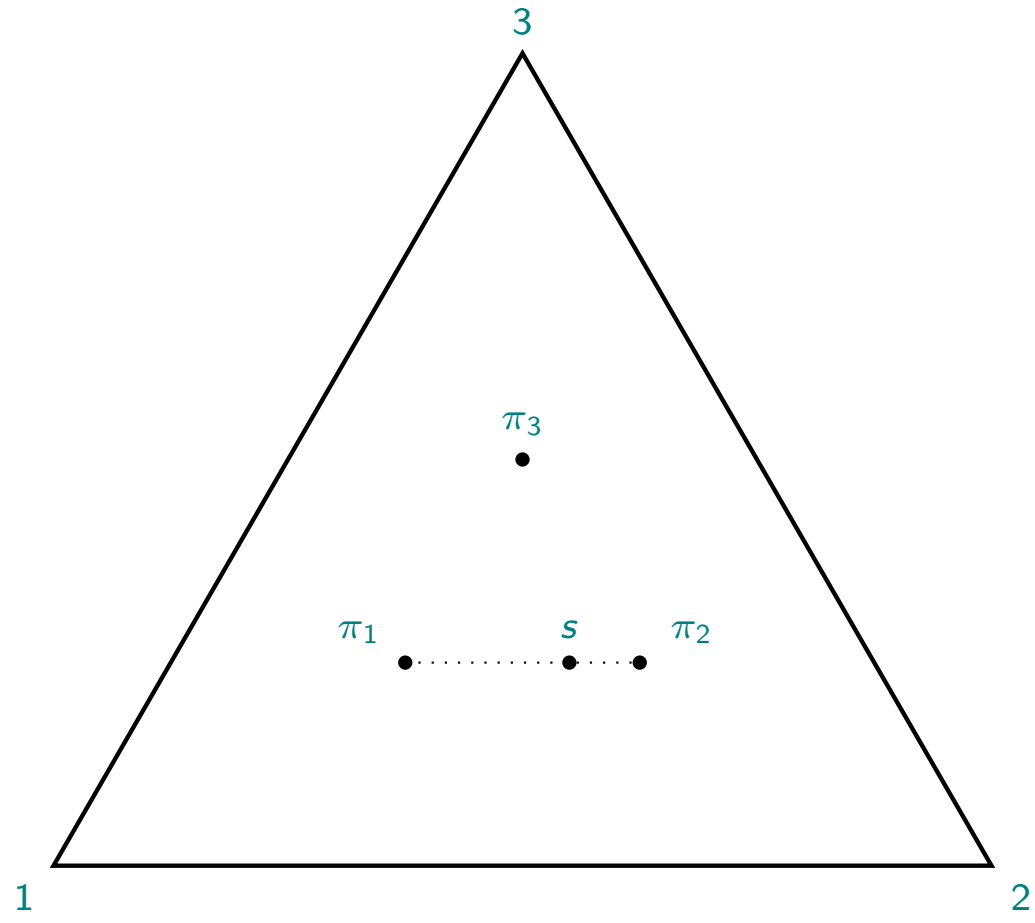
Diagnostic Conservatism



Perfect Fit

A3

If $s \in \mathrm{co}(\mathcal{T})$, then $q^s_{X|Z=z} = \pi_z$ for all $z \in \operatorname{supp} q^s_Z$.



Continuity

A4

The map $s \mapsto q_Z^s$ is continuous on $\operatorname{int} \Delta(\mathcal{X})$.

1 Sample Model

2 Axioms

3 MLE-Bayes Procedure

4 Representation

5 Stereotypes Application

6 Moral-Hazard Application

7 Maximal Type Conservatism

8 About Proof

9 Literature

MLE-Bayes Procedure

stage 1: maximum-likelihood estimation.

generative prior:

$$p_Z^s := \arg\max_{p_Z \in \Delta(\mathcal{Z})} \sum_{x \in \mathcal{X}} s(x) \ln p_X(x) \qquad \text{where } p_X := \sum_{z \in \mathcal{Z}} p_Z(z) \pi_z$$

generative model:

$$p_{XZ}^s(x, z) := p_Z^s(z) \pi_z(x)$$

stage 2: Bayes' updating.

MLE-Bayes sample model:

$$q_{XZ}^{\text{MLE}, s}(x, z) = s(x) p_{Z|X=x}^s(z)$$

The Running Example

Principal observed empirical frequencies s of outputs within her workforce.

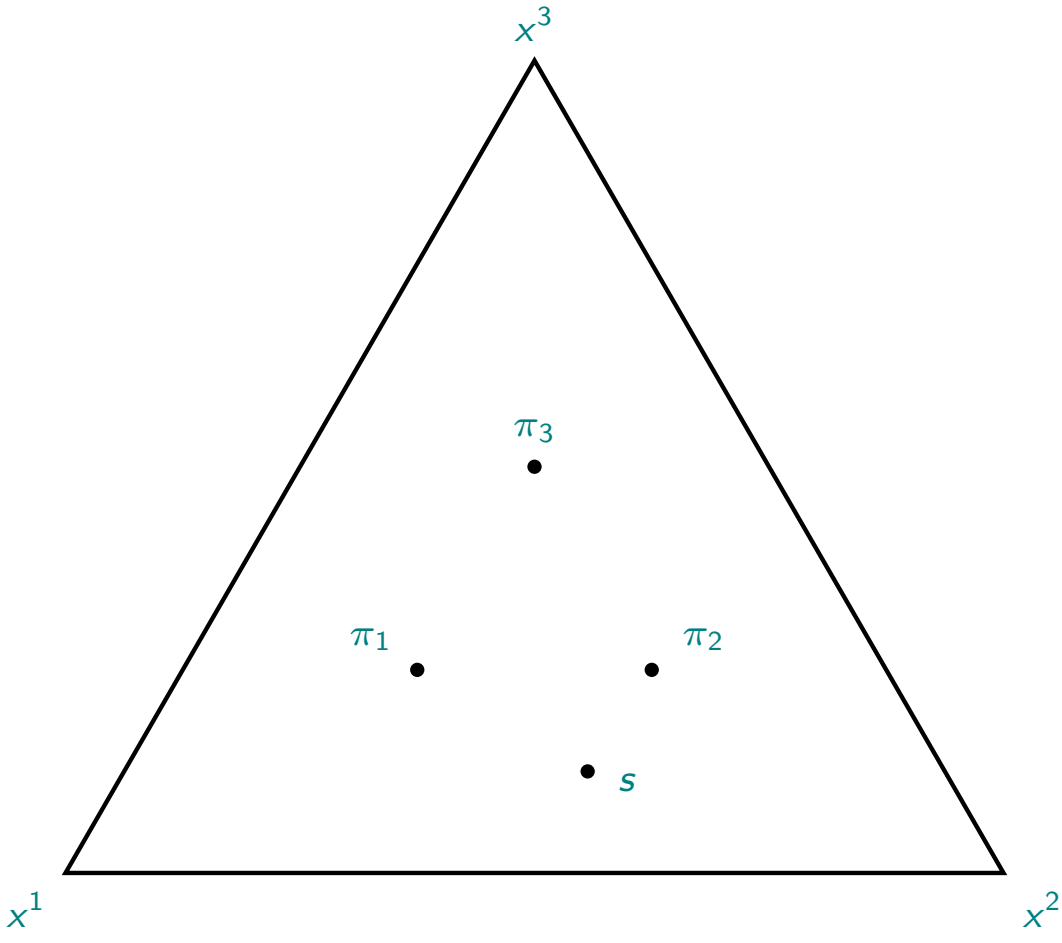
generative model p_{XZ}

- principal's belief about a new worker

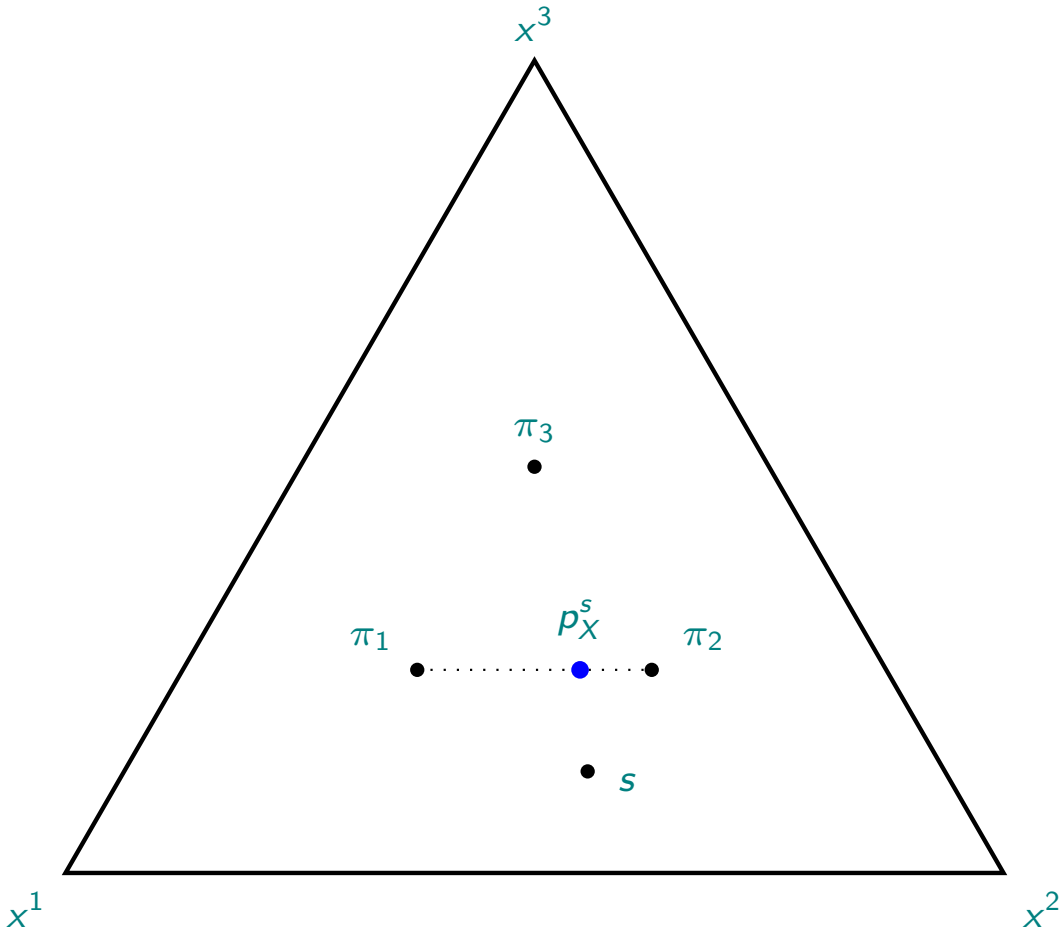
sample model q_{XZ}

- principal's belief about a current worker

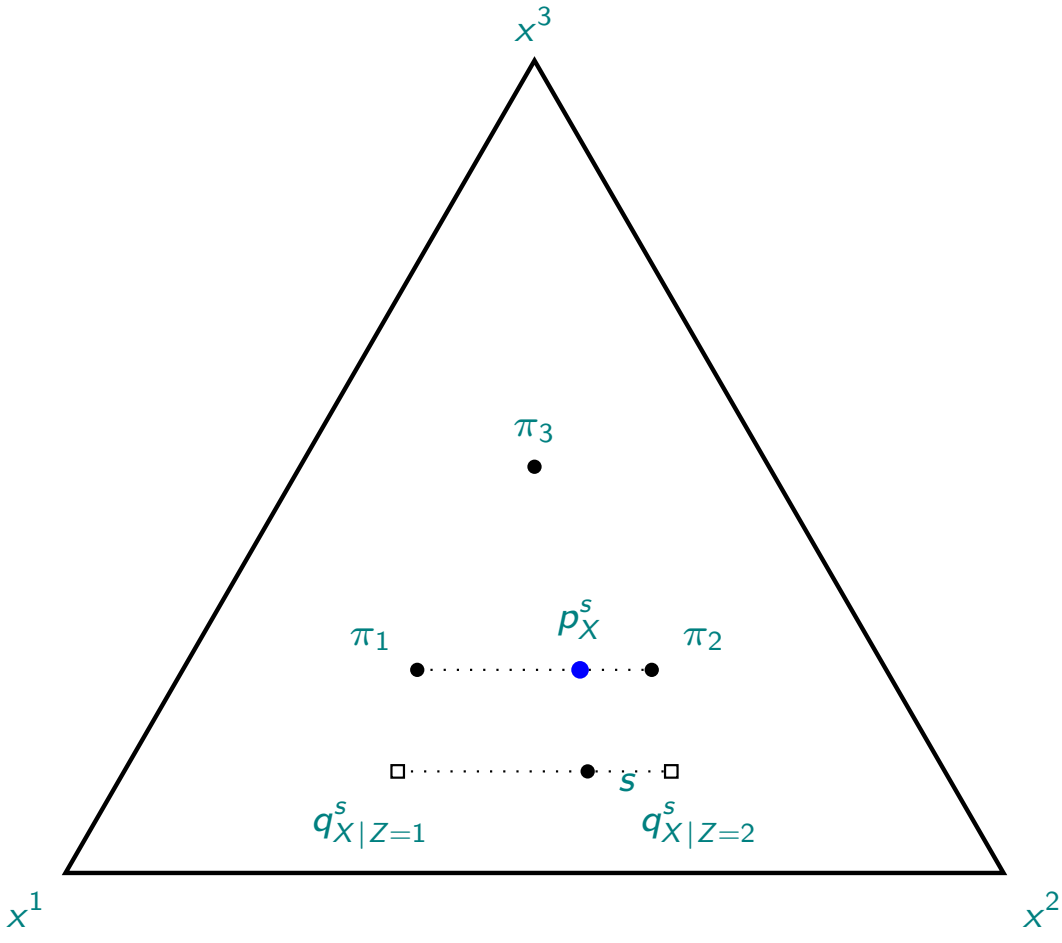
MLE-Bayes Procedure



MLE-Bayes Procedure



MLE-Bayes Procedure



1 Sample Model

2 Axioms

3 MLE-Bayes Procedure

4 Representation

5 Stereotypes Application

6 Moral-Hazard Application

7 Maximal Type Conservatism

8 About Proof

9 Literature

Representation Theorem

Theorem

$s \mapsto q^s_{XZ}$ satisfies A1–A4 if and only if $q^s_{XZ} = q^{\mathrm{MLE},\, s}_{XZ}$.

1 Sample Model

2 Axioms

3 MLE-Bayes Procedure

4 Representation

5 Stereotypes Application

6 Moral-Hazard Application

7 Maximal Type Conservatism

8 About Proof

9 Literature

Stereotypes: The Setting

the analyst observes $x = (x^1, x^2)$

- $x^1 \in \{L, H\}$ is a labor output
- $x^2 \in \{0, 1\}$ is a demographic group

she reasons about latent talent $z \in \{\ell, h\}$

her primitive theory:

	$(L,0)$	$(H,0)$	$(L,1)$	$(H,1)$
π_ℓ	$3/5$	$1/5$	$1/10$	$1/10$
π_h	$3/10$	$1/10$	$1/10$	$1/2$

Stereotype

diagnostic content of the labor output:

$$\lambda(x) := \frac{\pi_h(x)}{\pi_\ell(x)}$$

group 0: $$\lambda(L,0) = \lambda(H,0) = \frac{1}{2}$$

- high output is interpreted as luck, not talent

group 1: $$1 = \lambda(L,1) < \lambda(H,1) = 5$$

- output is strongly diagnostic of talent

What Evidence Corrects

sample:

$$s = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$$

MLE-Bayes sample types:

	$(L,0)$	$(H,0)$	$(L,1)$	$(H,1)$
$q^s_{X\|Z=\ell}$	$1/3$	$1/3$	$1/4$	$1/12$
$q^s_{X\|Z=h}$	$1/6$	$1/6$	$1/4$	$5/12$

predictive meanings of the talent labels get adjusted

What Evidence Does Not Correct

diagnostic likelihood ratios are preserved:

$$\frac{q^s(H,0 \mid h)}{q^s(H,0 \mid \ell)} = \frac{q^s(L,0 \mid h)}{q^s(L,0 \mid \ell)} = \frac{1}{2}$$

$$1 = \frac{q^s(L,1 \mid h)}{q^s(L,1 \mid \ell)} < \frac{q^s(H,1 \mid h)}{q^s(H,1 \mid \ell)} = 5$$

1 Sample Model

2 Axioms

3 MLE-Bayes Procedure

4 Representation

5 Stereotypes Application

6 Moral-Hazard Application

7 Maximal Type Conservatism

8 About Proof

9 Literature

Standard Moral-Hazard Setting

continuum of risk-neutral workers

latent efforts: $\mathcal{Z} = \{\ell, h\}$, $c_\ell = 0 < c_h$

labor outcomes $x^i \in \mathcal{X} = \{L, H, N\}$

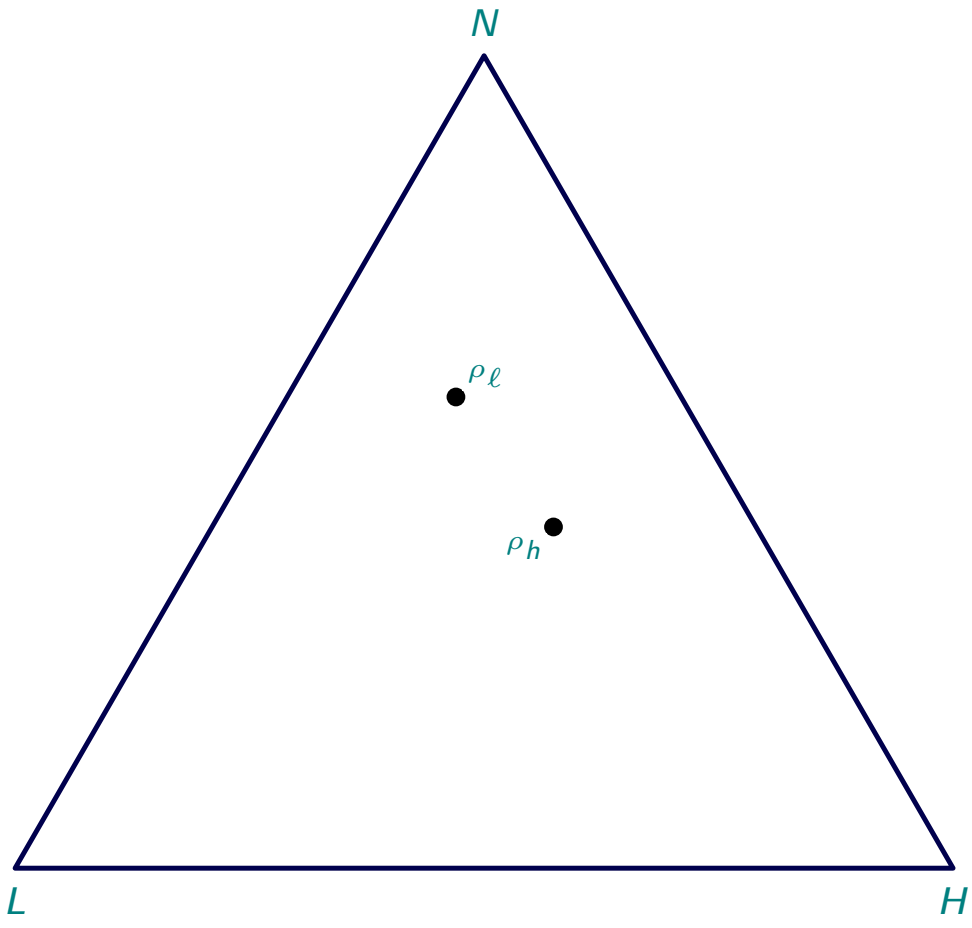
technology $\rho_z \in \Delta(\mathcal{X})$

principal chooses $w(\cdot)$ to maximize $\mathsf{E}[r(x) - w(x)]$ s.t. limited liability

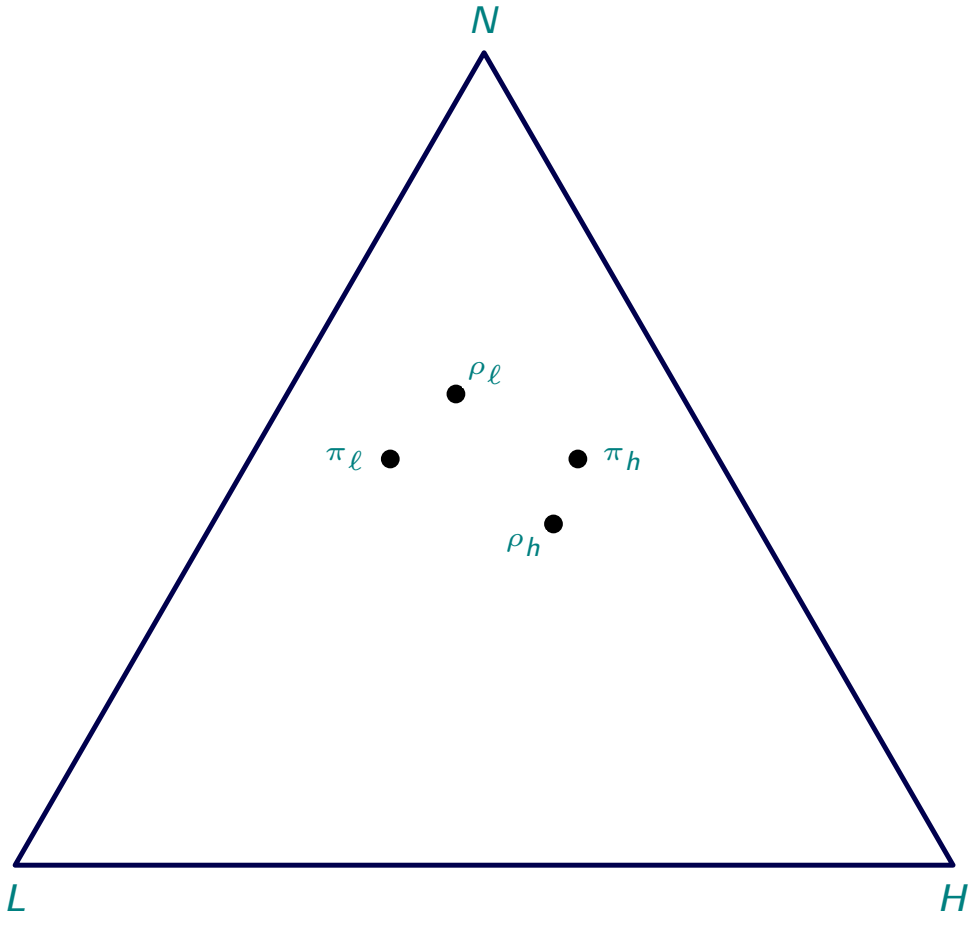
under our parameters, well-specified principal

- induces high effort
- keeps workers indifferent

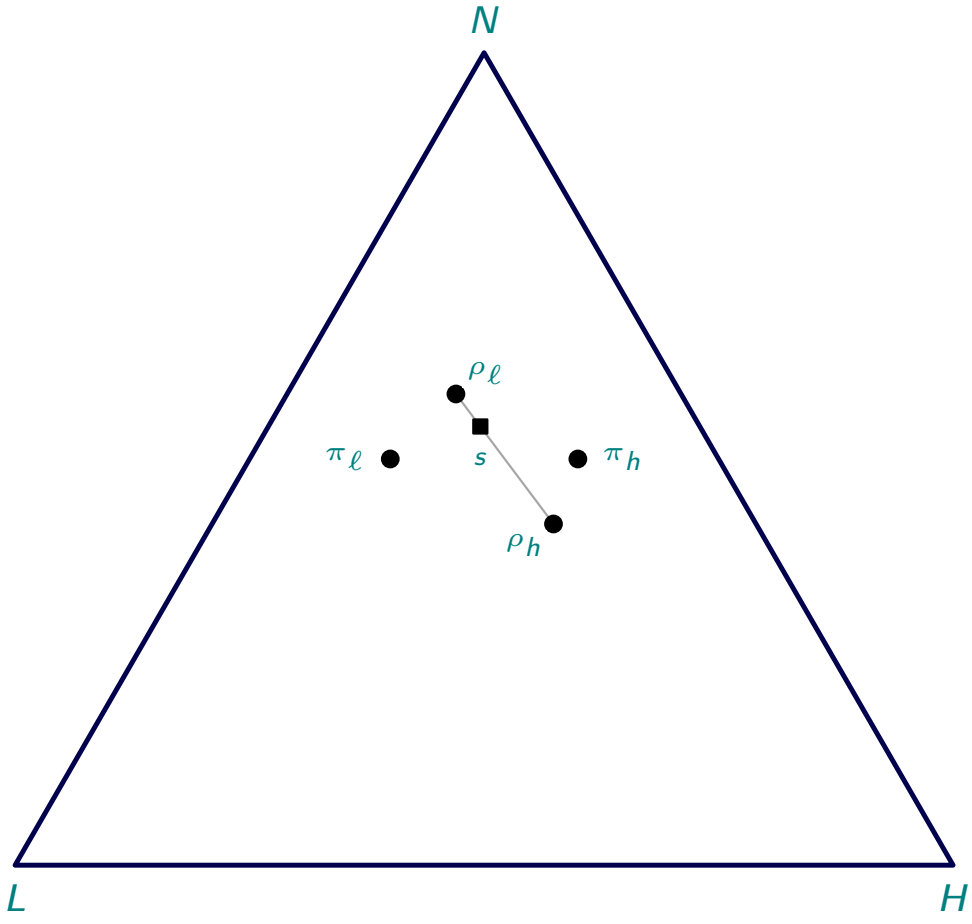
True Types



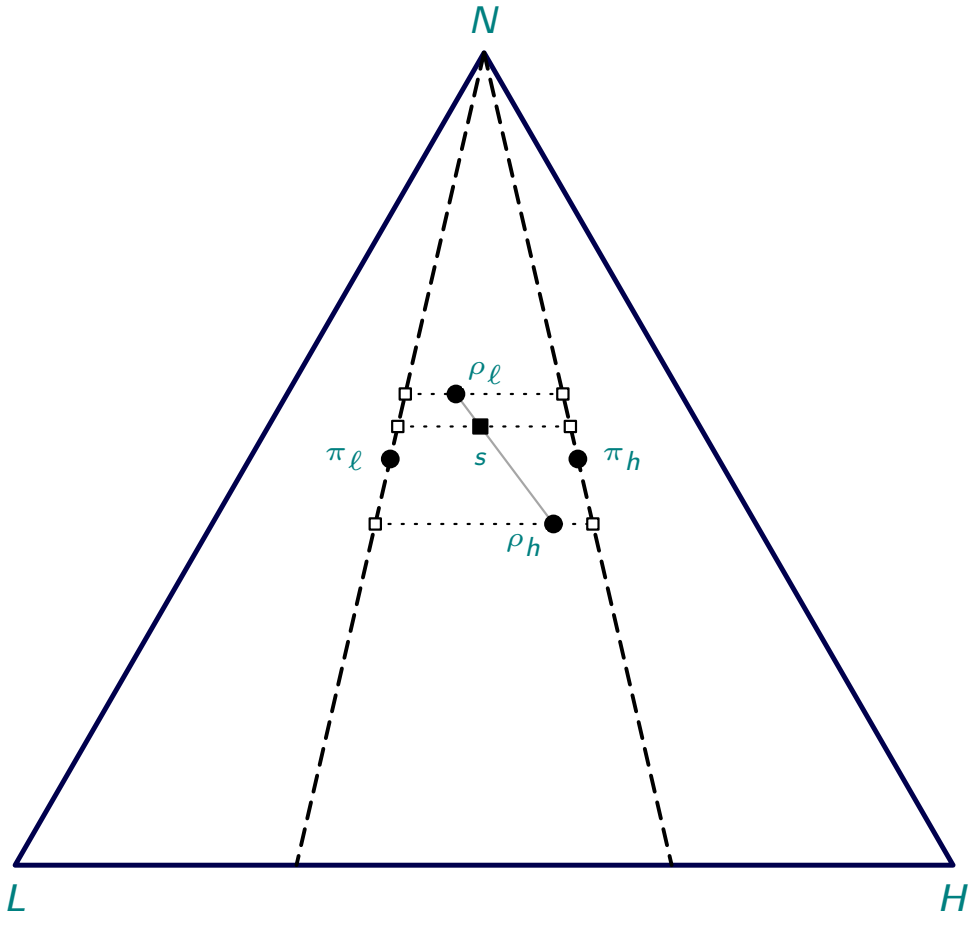
Misspecified Principal



Sample



Sample Model



Sample Equilibrium

a fixed point (e^*, s^*, w^*):

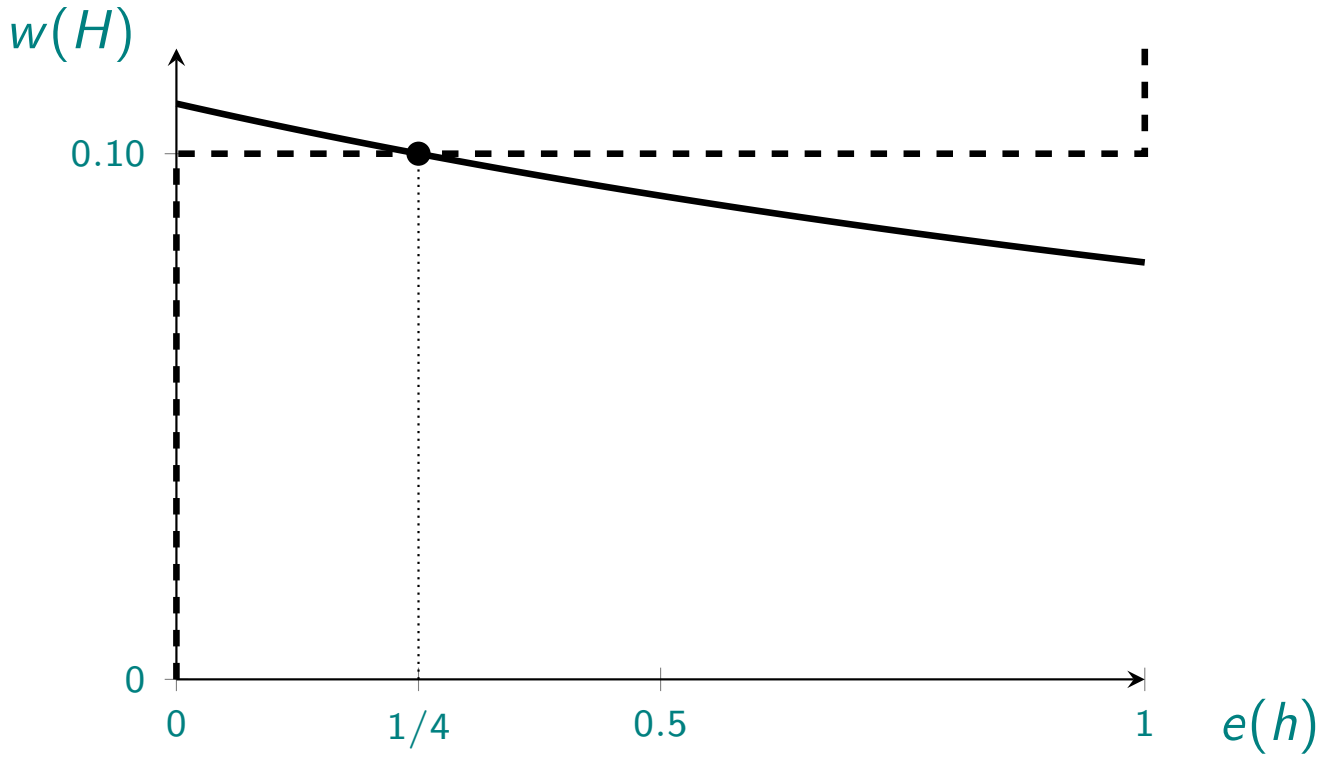
1. given wage profile w^*, workers choose an optimal mixed effort e^*,
2. sample s^* is generated by e^* according to the objective technology,
3. given subjective sample types, w^* is the optimal wage profile.

Equilibrium Characterization

for the considered parametrization:

- unique equilibrium
- the principal chooses the well-specified contract
- interior effort (workers mix)

Sample Equilibrium



Berk-Nash Equilibrium

Berk-Nash principal responds to her generative model

$$p^s_{XZ}(x, z) := p^s_Z(z)\pi_z(x)$$

in this setting, she responds to fixed primitive types π_z

for our parameters, she provides low incentives and workers shirk

she ignores the wedge between the prediction p_X and the sample s

1 Sample Model

2 Axioms

3 MLE-Bayes Procedure

4 Representation

5 Stereotypes Application

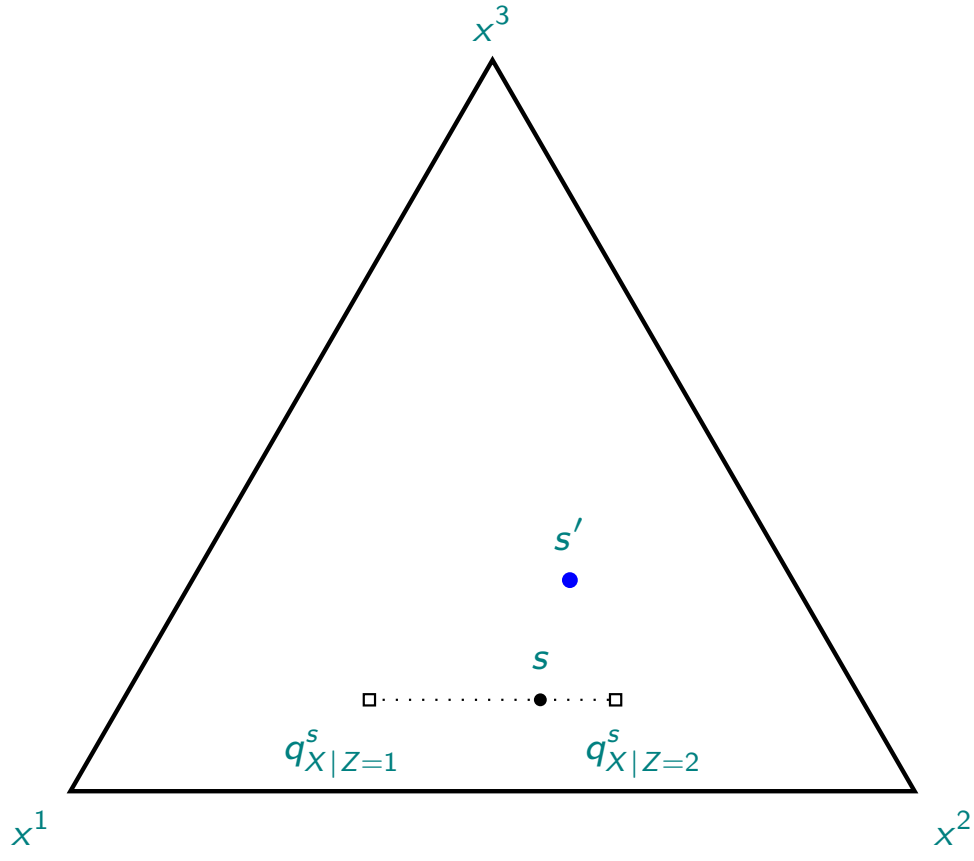
6 Moral-Hazard Application

7 Maximal Type Conservatism

8 About Proof

9 Literature

When Types Must Change



Types Change Only If They Must

Theorem

The analyst retains her type definitions in the sample shift from s to s' if and only if

$$s' \in \operatorname{co}\left(\left\{q^s_{X|Z=z}\right\}_{z \in \operatorname{supp} q^s_Z}\right).$$

1 Sample Model

2 Axioms

3 MLE-Bayes Procedure

4 Representation

5 Stereotypes Application

6 Moral-Hazard Application

7 Maximal Type Conservatism

8 About Proof

9 Literature

Partial Representation

$q_{XZ}^{\mathrm{MLE},\, s,\mathcal{E}}$ – MLE-Bayes sample model restricted to types in $\mathcal{E} \subseteq \mathcal{Z}$

fixed-support representation

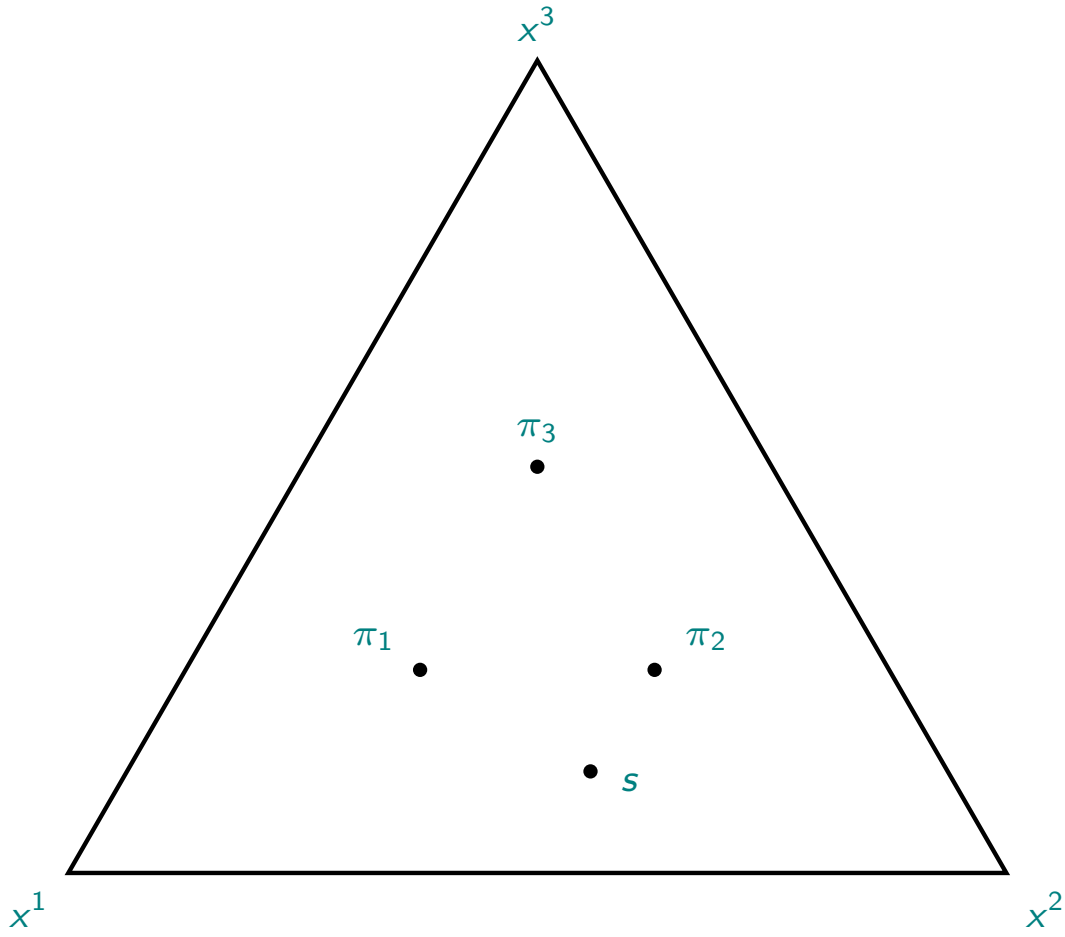
If a sample model q_{XZ}^{s} satisfies A1 and A2 and

$$\operatorname{supp} q_{Z}^{s} = \mathcal{E},$$

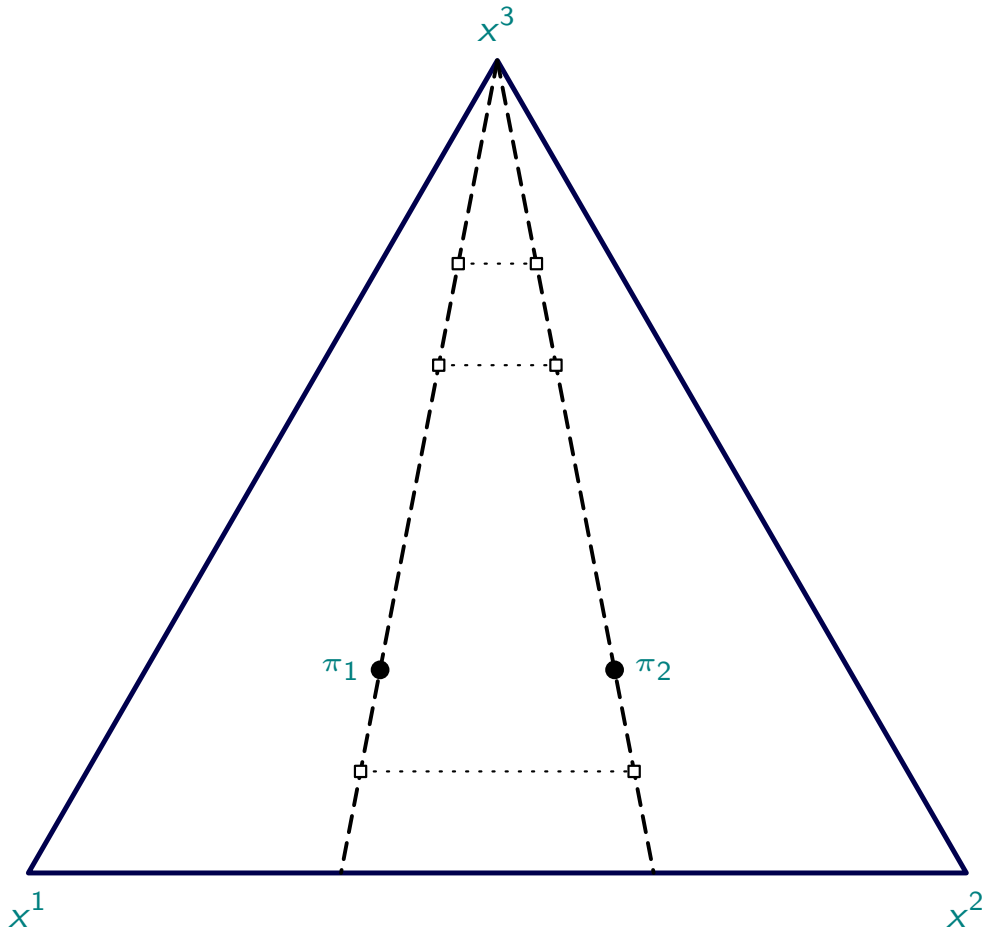
then

$$q_{XZ}^{s} = q_{XZ}^{\mathrm{MLE},\, s,\mathcal{E}}.$$

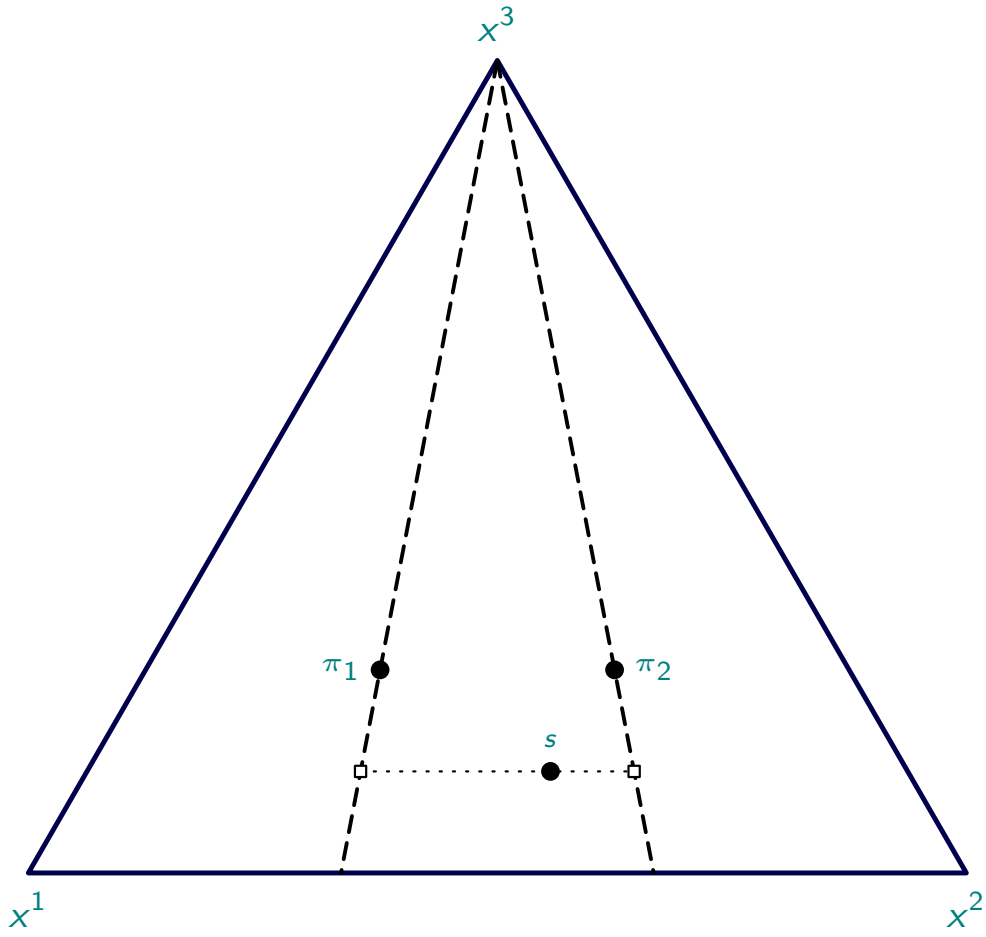
Example



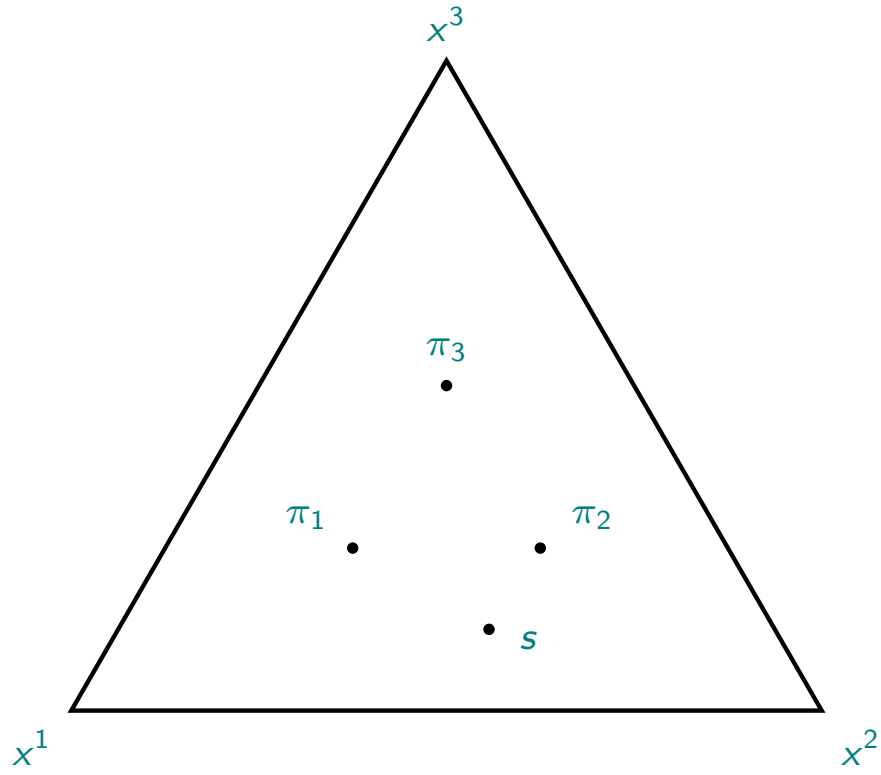
Example



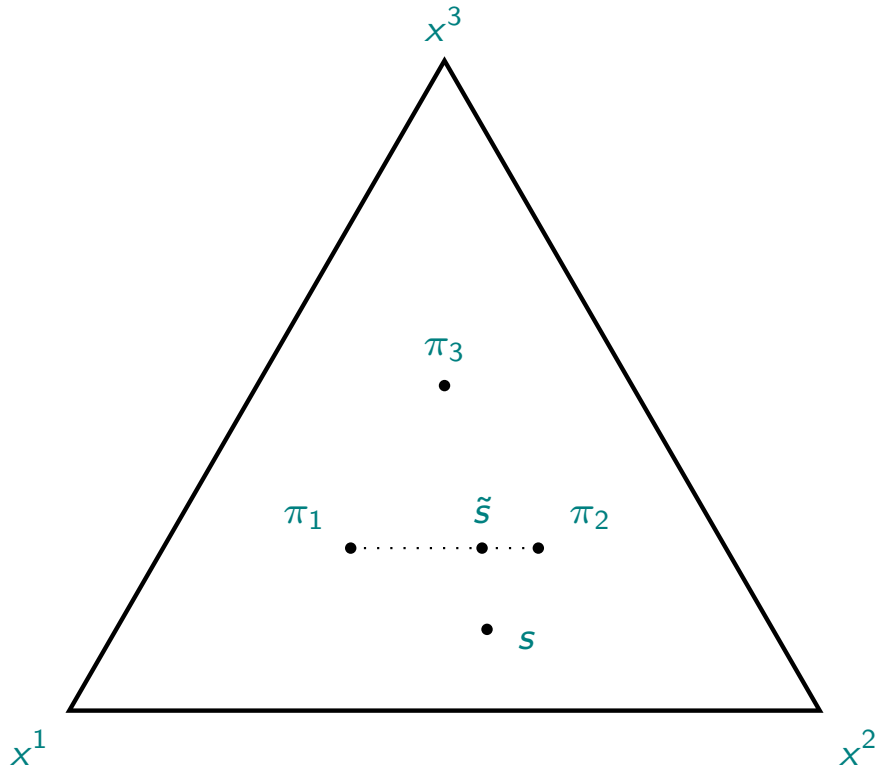
Example



Continuous Expansion

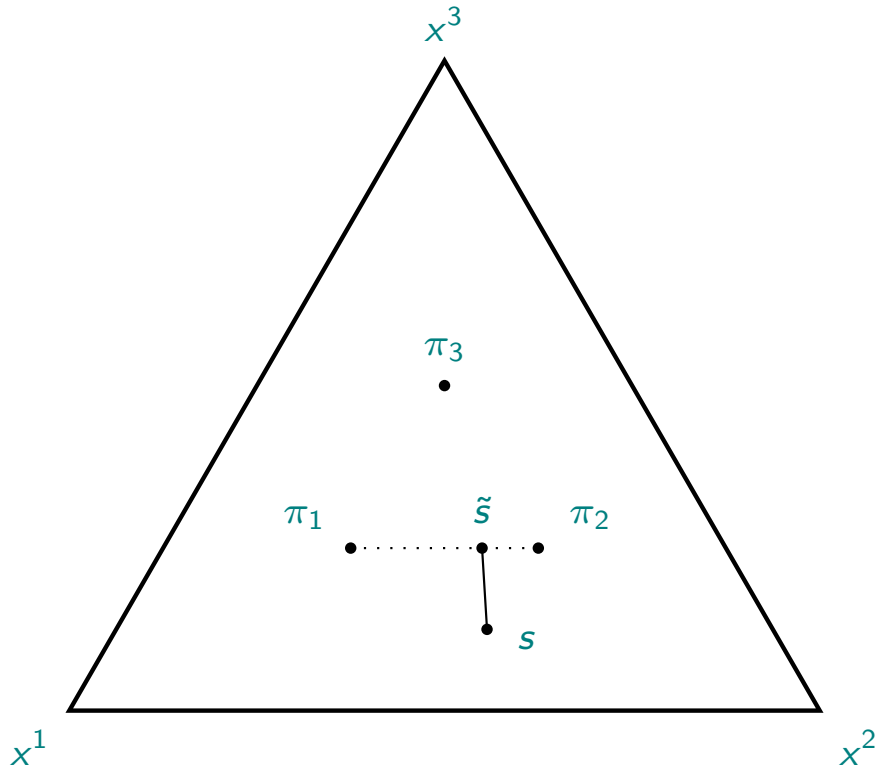


Continuous Expansion



anchor sample $\tilde{s} := p_X^s \in \mathrm{co}(\mathcal{T})$

Continuous Expansion

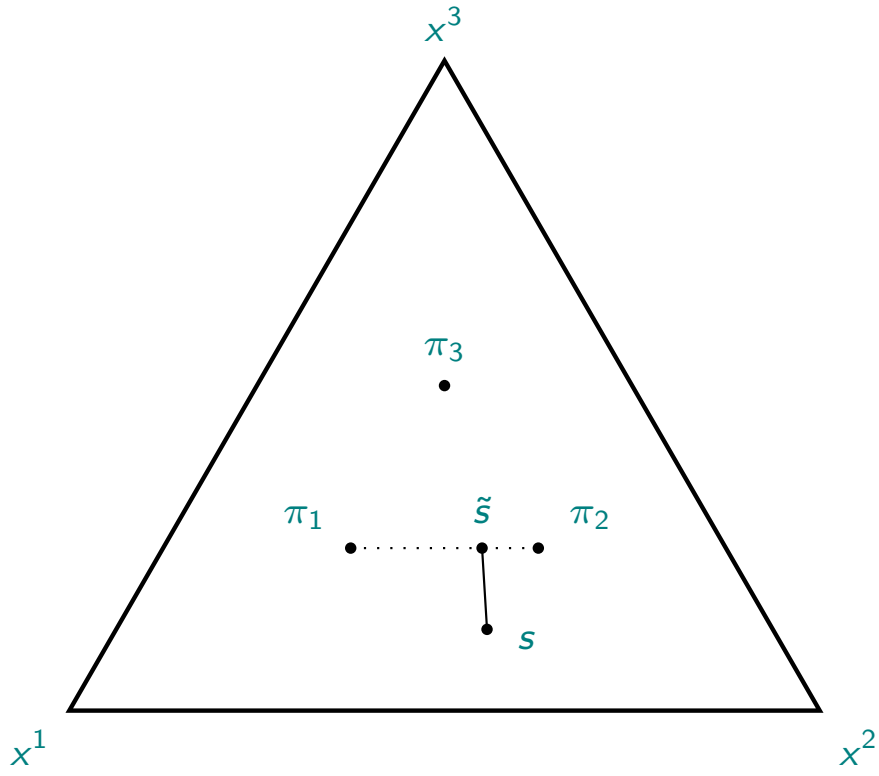


anchor sample $\tilde{s} := p^s_X \in \mathrm{co}(\mathcal{T})$

$\tilde{s} = p^{s'}_X$ for all $s' \in \mathrm{co}\{\tilde{s}, s\}$

constant estimate $q_Z^{\mathrm{MLE},\, s'}$ for all $s' \in \mathrm{co}\{\tilde{s}, s\}$

Continuous Expansion



we prove that $q_Z^{s'} = q_Z^{\mathrm{MLE},\, s'}$ for all $s' \in \mathrm{co}\{\tilde{s}, s\}$

- this holds for $\tilde{s}$
- continuous expansion preserves the set of employed types

1 Sample Model

2 Axioms

3 MLE-Bayes Procedure

4 Representation

5 Stereotypes Application

6 Moral-Hazard Application

7 Maximal Type Conservatism

8 About Proof

9 Literature

Literature

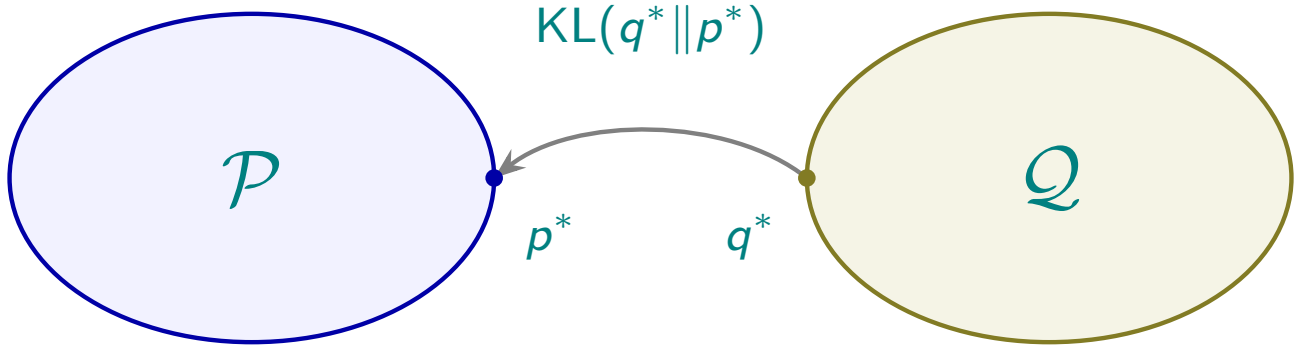
reconciling misspecification: Ortoleva'12, Cho&Kasa'15, Ba'26, Gagnon-Bartsch et al.'18, Karni and Vierø'13, Lanzani'25

variational Bayes: Dempster et al.'77, Jordan et al.'99, Blei et al.'17, Kingma&Welling'13, Aridor et al.'20,'25

mixture estimates: Lazarsfeld&Henry'68, Kiefer&Wolfowitz'56, Lindsay'83, Heckman&Singer'84, Bonhomme et al.'22

formal analogy to rational inattention: Matějka&McKay'15, Caplin&Dean'15

Information Geometry



MLE-Bayes As KL Minimization

generative models: preserve the primitive types

$$\mathcal{P} = \{p_{XZ} : p_{X|Z=z} = \pi_z \text{ for all } z\}.$$

sample models: satisfy realism

$$\mathcal{Q}(s) = \{q_{XZ} : q_X = s\}.$$

latent representation minimizes the KL wedge

$$(p^s_{XZ}, q^s_{XZ}) \in \underset{\substack{p_{XZ}\in\mathcal{P}\\ q_{XZ}\in\mathcal{Q}(s)}}{\arg\min}\ \mathrm{KL}(q_{XZ}\|p_{XZ}).$$

entropic Rational Inattention problem has the same structure

Summary

reasoning about observations that refute one's own statistical model

- adjustments of the predictive content
- rigidity in the diagnostic content