Constrained Data-Fitters

Larry Samuelson, Jakub Steiner

Austin, October 2024

The Map is Not the Territory

Alfred Korzybski

1 Approximate Updates and Estimation

2 Microfoundations

3 (not so) Rational Expectations

4 Optimal Simplicity

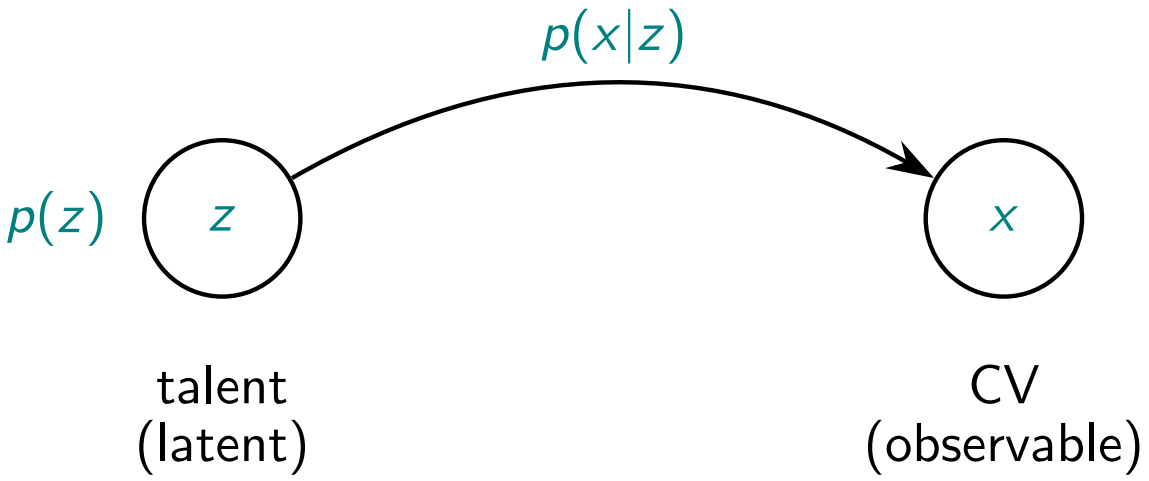
5 Posterior Approach

6 Misspecification and Beyond

7 Tricks from Rational Inattention

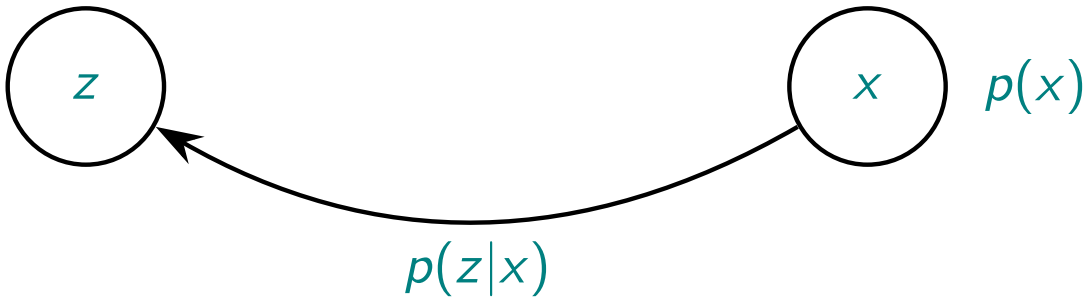
Statistical Model

a "map"

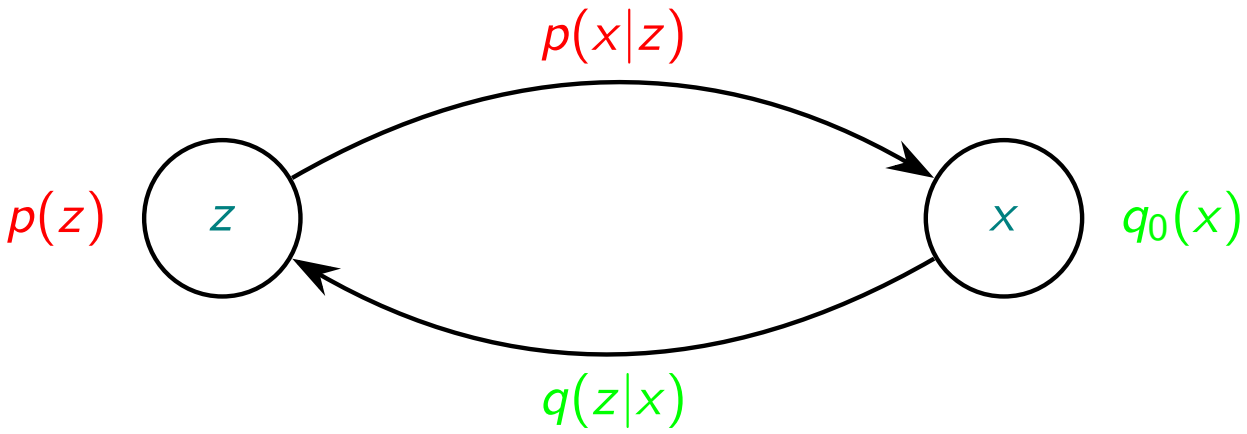


Posterior Approach

same model, different perspective



Two Models



generative model vs recognition model

$p(x, z) = p(z)p(x \mid z)$ $q(x, z) = q_0(x)q(z \mid x)$

Example

generative model:

- $p(z)$ prior distribution of talent in population
- $p(x \mid z)$ stochastic CV of each talent type

recognition model:

- large sample of the job applicants' CVs
- $q_0(x)$ – empirical distribution of the sample
- belief $q(z \mid x)$ about talent z of a candidate with CV x

Choice of the Two Models

let's proceed backwards

given the generative model, choose the recognition model

- that is most consistent with the generative model
- subject to a cognitive constraint

choose the generative model with the best subjective fit to data

- accounting for own cognitive constraint during updating

Constrained Updating

variational Bayes methods, Jordan et al.'99

updates solve

$$\max_{\left(\tilde{q}(z|x)\right)_x} \quad \mathrm{E}_{\tilde{q}(x,z)} \ln p(\hat{x}, \hat{z}) + \mathrm{H}\left(\tilde{q}(x,z)\right)$$

$$\text{s.t.} \quad \tilde{q}(x,z) \in \mathcal{Q}$$

the maximizer: constrained updates

the value: constrained likelihood

Constrained Updating

variational Bayes methods, Jordan et al.'99

recognition model solves

$$\max_{\tilde{q}(x,z)} \quad \mathrm{E}_{\tilde{q}(x,z)} \ln p(\hat{x}, \hat{z}) + \mathrm{H}\left(\tilde{q}(x,z)\right)$$

$$\text{s.t.} \quad \tilde{q}(x,z) \in \mathcal{Q}$$

$$\tilde{q}(x) = q_0(x)$$

empirical constraint

Constrained Updating

variational Bayes methods, Jordan et al.'99

recognition model solves

$$\max_{\tilde{q}(x,z)} \quad \mathrm{E}_{\tilde{q}(x,z)} \ln p(\hat{x}, \hat{z}) + \mathrm{H}\left(\tilde{q}(x,z)\right)$$

$$\text{s.t.} \quad \tilde{q}(x,z) \in \mathcal{Q}$$

$$\tilde{q}(x) = q_0(x)$$

updating constraint

Constrained Updating

variational Bayes methods, Jordan et al.'99

recognition model solves

$$\max_{\tilde{q}(x,z)} \quad \mathrm{E}_{\tilde{q}(x,z)} \ln p(\hat{x}, \hat{z}) + \mathrm{H}\left(\tilde{q}(x,z)\right)$$

$$\text{s.t.} \quad \tilde{q}(x,z) \in \mathcal{Q}$$

$$\tilde{q}(x) = q_0(x)$$

reconstruction term

Constrained Updating

variational Bayes methods, Jordan et al.'99

recognition model solves

$$\max_{\tilde{q}(x,z)} \quad \mathrm{E}_{\tilde{q}(x,z)} \ln p(\hat{x}, \hat{z}) + \mathrm{H}\left(\tilde{q}(x,z)\right)$$

$$\text{s.t.} \quad \tilde{q}(x,z) \in \mathcal{Q}$$

$$\tilde{q}(x) = q_0(x)$$

regularization term

Constrained Updating

variational Bayes methods, Jordan et al.'99

recognition model solves

$$\min_{\tilde{q}(x,z)} \quad \mathrm{KL}\left(\tilde{q}(x,z) \parallel p(x,z)\right)$$

$$\text{s.t.} \quad \tilde{q}(x,z) \in \mathcal{Q}$$

$$\tilde{q}(x) = q_0(x)$$

Approximate Maximum Likelihood Estimation

variational autoencoder, Kingma&Welling'13

principle

Choose the generative model that maximizes constrained likelihood.

$$
\begin{aligned}
\min_{\tilde{p}(x,z),\tilde{q}(x,z)} \quad & \mathrm{KL}\left(\tilde{q}(x,z) \parallel \tilde{p}(x,z)\right) \\
\text{s.t.} \quad & \tilde{p}(x,z) \in \mathcal{P} \\
& \tilde{q}(x,z) \in \mathcal{Q} \\
& \tilde{q}(x) = q_0(x)
\end{aligned}
$$

information geometry

Some Updating Constraints

no constraint: $\mathcal{Q} = \Delta(X \times Z)$

- Bayesian updates, $q(z \mid x) = p(z \mid x)$
- unconstrained likelihood, $\mathrm{E}_{q_0(x)} \ln p(\hat{x}) + \text{const.}$

Some Updating Constraints

computational constraint: $\mathcal{Q}$ is a family of tractable distributions

Some Updating Constraints

causal constraint; e.g.:

- $z = (z_1, z_2)$
- q must comply with directed acyclical graph $z_1 \leftarrow x \rightarrow z_2$
- $\Leftrightarrow$ factorization constraint $q(x, z_1, z_2) = q(x)q(z_1 \mid x)q(z_2 \mid x)$

1 Approximate Updates and Estimation

2 Microfoundations

3 (not so) Rational Expectations

4 Optimal Simplicity

5 Posterior Approach

6 Misspecification and Beyond

7 Tricks from Rational Inattention

Overview

$x_1, \ldots, x_n$ – CVs of job applicants

fixed generative model $p(x, z)$

joint distribution $q(x, z)$ of CVs and talents among the applicants?

$$\begin{array}{ll} \max\limits_{\tilde{q}(x,z)} & p\text{-likelihood} \\ \text{s.t.} & \text{you comprehend } \tilde{q}(x, z) \\ & \tilde{q}(x) \text{ matches the sample of CVs} \end{array}$$

Estimation

sample $(x_i)_{i=1}^n$ $\dashrightarrow$ extended sample $(x_i, z_i)_{i=1}^n$

p-likelihood of all extended samples with frequencies $q(x, z)$

$$\ell_n(q) := \prod_{x,z} p(x, z)^{q(x,z)n} \times \mathcal{N}_n(q)$$

the estimate:

$$q_n(x, z) \in \arg\max_{\tilde{q} \in \mathcal{Q}_n} \ell_n(q)$$

Permutations
illustration

four extended samples with a same joint distribution

x^4	0	1	0	1
z^4	0	0	1	1
z^4	1	0	0	1
z^4	0	1	1	0
z^4	1	1	0	0

Limit

let $\mathcal{Q}_n$ approximate $\mathcal{Q}$ details

proposition

$$
\begin{aligned}
q_n(x,z) &\rightarrow \text{recognition model } q(x,z) \\
\frac{1}{n}\ln \ell_n(q_n) &\rightarrow \text{constrained likelihood} + \text{const.}
\end{aligned}
$$

because

$$
\begin{aligned}
\frac{1}{n}\ln \prod_{x,z} p(x,z)^{q(x,z)n} &\times \mathcal{N}_n(q) \\
\rightarrow \mathrm{E}_{q(x,z)} \ln p(\hat{x},\hat{z}) &+ \mathrm{H}\left(q(x,z)\right) - \mathrm{H}\left(q_0(x)\right)
\end{aligned}
$$

1 Approximate Updates and Estimation

2 Microfoundations

3 (not so) Rational Expectations

4 Optimal Simplicity

5 Posterior Approach

6 Misspecification and Beyond

7 Tricks from Rational Inattention

Rational Expectations
definition

definition

The agent has rational expectations if she isn't systematically surprised:

$$p(z) = \mathrm{E}_{q_0(x)}\, q(z \mid x) \equiv q(z).$$

in general, our agent will not have RE because she

- doesn't use Bayes' rule
- has a wrong model

Flexibility Assumption

definition

$\mathcal{P}$ has unconstrained margin:

- all $p(z)$ are feasible for any choice of $\big(p(x \mid z)\big)_z$.

agent with unconstrained margin

- can conceive any prior $p(z)$
- has preconceived knowledge about implications of z to x

Rational Expectations

result

proposition

If $\mathcal{P}$ has unconstrained margin, the agent forms rational expectations.

proof: optimize over $\tilde{p}(z)$

$$\mathrm{KL}\left(q(x,z) \parallel \tilde{p}(x,z)\right) = \mathrm{KL}\left(q(z) \parallel \tilde{p}(z)\right) + \sum_z q(z)\,\mathrm{KL}\left(q(x \mid z) \parallel p(x \mid z)\right)$$

standard Bayes' plausibility is forced by the Bayes' rule

- RE can fail in our framework, but hold at the optimum

a popular non-Bayesian intuition in support of RE:

- systematically surprised agent should adjust her prior ✓

1 Approximate Updates and Estimation

2 Microfoundations

3 (not so) Rational Expectations

4 Optimal Simplicity

5 Posterior Approach

6 Misspecification and Beyond

7 Tricks from Rational Inattention

Example

$x = \text{CV}$, $z = (\text{aptitude, grit})$

generative model:

- any $p(\text{aptitude, grit})$
- a constraint imposed on $p(\text{CV} \mid \text{aptitude, grit})$

recognition model:

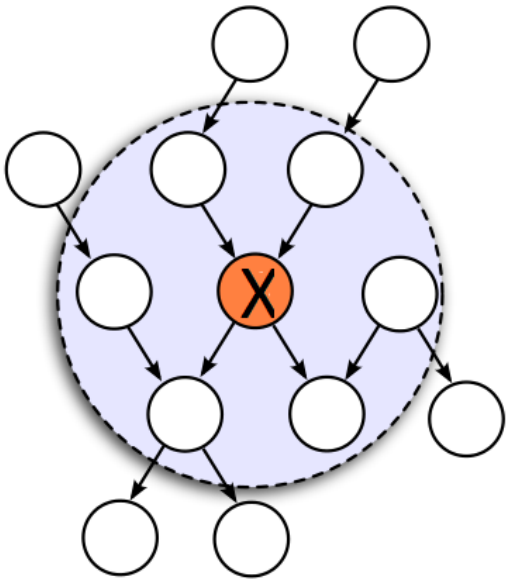
- $\text{CV} \rightarrow \text{aptitude} \rightarrow \text{grit}$

deterministic collapse

Grit is a deterministic function of aptitude under both p and q.

grit has no explanatory power at the recognition stage
$\Rightarrow$ it is not used in the generative stage

Deterministic Collapse



Markov boundary: minimal set that contains all information about X

assumptions

- $\mathcal{P}$ has unconstrained margin
- q complies with a causal network, or is more deterministic

proposition

The agent models latent variables from outside the Markov Boundary as a deterministic function of the latent variables from within the boundary.

1 Approximate Updates and Estimation

2 Microfoundations

3 (not so) Rational Expectations

4 Optimal Simplicity

5 Posterior Approach

6 Misspecification and Beyond

7 Tricks from Rational Inattention

Posterior Approach
information design

posterior representation: $q(z)$, $\left(q(x \mid z)\right)_z$, and $\left(p(x \mid z)\right)_z$

- specifies both models $p(x, z)$ and $q(x, z)$

lemma: posterior-separable objective

If $\mathcal{P}$ has unconstrained margin, then the model-fitting problem becomes

$$\max_{\tilde{q}(z),(\tilde{q}(x|z))_z,(\tilde{p}(x|z))_z} \quad \mathrm{E}_{\tilde{q}(z)}\left[\mathrm{E}_{\tilde{q}(x|\hat{z})} \ln \tilde{p}(\hat{x} \mid \hat{z}) + \mathrm{H}\left(\tilde{q}(x \mid \hat{z})\right)\right]$$

$$\text{s.t.} \quad \text{the constraints}$$

Deterministic Collapse

proof

$\mathcal{P}$ has unconstrained margin $\Rightarrow$ posterior-separable objective

the chain constraint $x \to z_1 \to z_2 \Rightarrow q(x \mid z_1, z_2) = q(x \mid z_1)$

consider a solution p and q

for each z_1, re-optimize $q(z_2 \mid z_1)$, fixing $q(z_1)$, $q(x \mid z)$ and $p(x \mid z)$

objective $\mathrm{E}_{q(z_1)} \left[\mathrm{E}_{q(x, z_2 \mid z_1)} \ln p(x \mid z_1, z_2) + \mathrm{H}\left(q(x \mid z_1)\right) \right]$

deterministically pick the maximizer $z_2^* = d(z_1)$

1 Approximate Updates and Estimation

2 Microfoundations

3 (not so) Rational Expectations

4 Optimal Simplicity

5 Posterior Approach

6 Misspecification and Beyond

7 Tricks from Rational Inattention

Two Frictions

		$\mathcal{P}$	
		Well-specified	**Miss-specified**
$\mathcal{Q}$	**Bayes' Rationality**	Wald'49	Berk'66
	Updating Friction	model-fitting	model-fitting

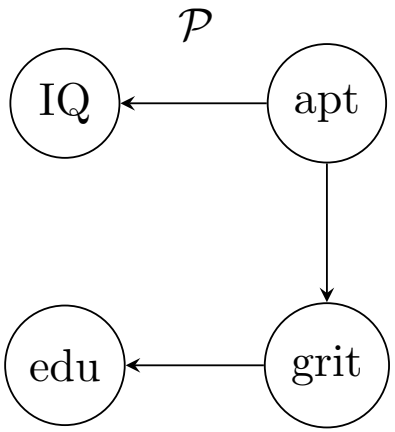
Two Frictions

		$\mathcal{P}$	
		Well-specified	**Miss-specified**
$\mathcal{Q}$	**Bayes' Rationality**	Wald'49	Berk'66
	Updating Friction	model-fitting	model-fitting

Correlation Neglect

$x = (\text{IQ score}, \text{ education})$ and $z = (\text{aptitude}, \text{ grit})$

the true process $q_0(\text{IQ}, \text{edu})$ exhibits correlation



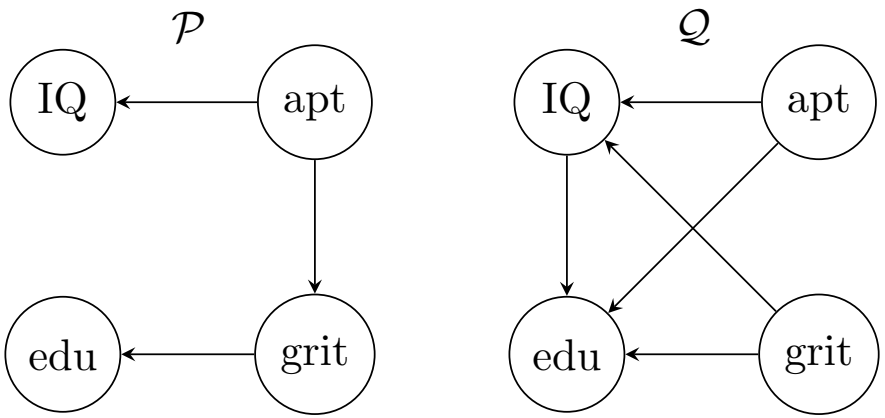
agent is well-specified

- $\Rightarrow$ she learns the true process if updating is unconstrained

Correlation Neglect

$x = (\text{IQ score}, \text{ education})$ and $z = (\text{aptitude}, \text{ grit})$

the true process $q_0(\text{IQ}, \text{edu})$ exhibits correlation



the updating constraint

- $\Rightarrow$ optimal correlation neglect $p(\text{IQ}, \text{edu}) = p(\text{IQ})p(\text{edu})$ proof

The Constraints

generative model:

- aptitude $\rightarrow$ the IQ test result
- grit $\rightarrow$ education
- aptitude and grit may be correlated

recognition model:

- divide the sample of the CVs into arbitrary subsamples
- arbitrarily attach to each subsample a cause $z = (\text{aptitude}, \text{grit})$
- correlation neglect: aptitude and grit independent

correlation neglect the recognition stage
$\Rightarrow$ correlation neglect at the generative stage

1 Approximate Updates and Estimation

2 Microfoundations

3 (not so) Rational Expectations

4 Optimal Simplicity

5 Posterior Approach

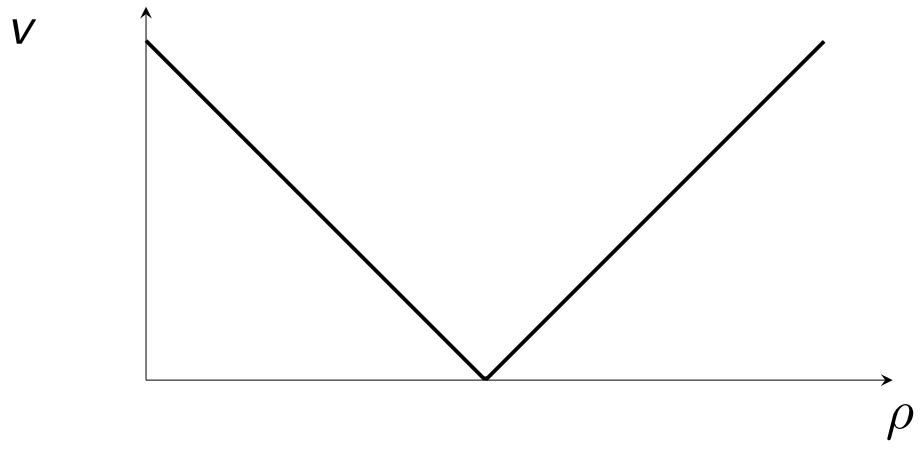
6 Misspecification and Beyond

7 Tricks from Rational Inattention

Rational Inattention

- payoff state $x \sim q_0(x)$
- agent chooses experiment $q(z \mid x)$
- maps the observed signal to action a
- maximizes $\mathrm{E}\left[u(a,x) + \mathrm{H}\left(q(x \mid z)\right)\right]$

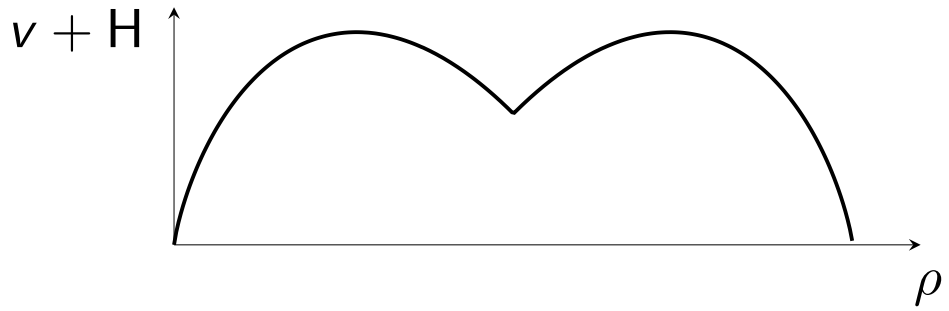
plot $\rho \mapsto \max_a \mathrm{E}_{\rho(x)}\, u(a, \hat{x})$



Rational Inattention

- payoff state $x \sim q_0(x)$
- agent chooses experiment $q(z \mid x)$
- maps the observed signal to action a
- maximizes $\mathrm{E}\left[u(a,x) + \mathrm{H}\left(q(x \mid z)\right)\right]$

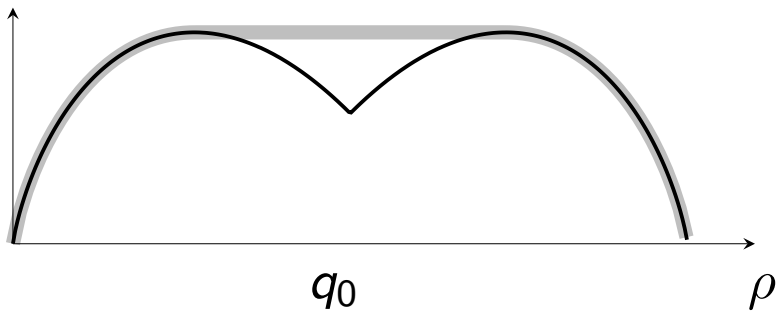
plot $\rho \mapsto \max_a \mathrm{E}_{\rho(x)}\, u(a,\hat{x}) + \mathrm{H}\left(\rho\right)$



Rational Inattention

- payoff state $x \sim q_0(x)$
- agent chooses experiment $q(z \mid x)$
- maps the observed signal to action a
- maximizes $\mathrm{E}\left[u(a,x) + \mathrm{H}\left(q(x \mid z)\right)\right]$

find the optimal posteriors



Posterior Separable Constraints

$\mathcal{P}$ and $\mathcal{Q}$ are posterior separable if

- p and q are feasible iff $p(x \mid z) \in \bar{\mathcal{P}}$ and $q(x \mid z) \in \bar{\mathcal{Q}}$

posterior separable $\mathcal{P}$:

- agent models the process generating x
- she is endowed with a set of primitive distributions $\bar{\mathcal{P}}$
- can build any mixture distribution from the convex hull of $\bar{\mathcal{P}}$

posterior separable $\mathcal{Q}$:

- agent partitions the observed dataset x^n
- assigns distinct z to each cell
- each cell must have an empirical distribution in $\bar{\mathcal{Q}}$

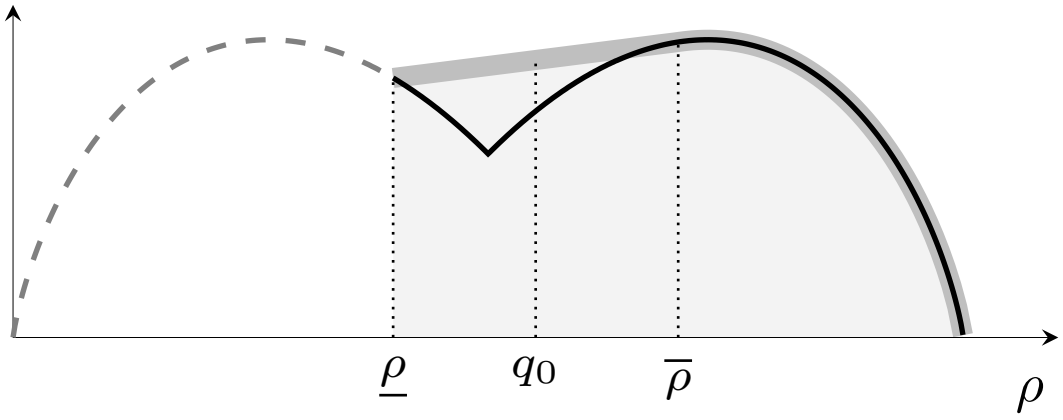
Connection

call the primitive distributions "actions": $\bar{\mathcal{P}} = \{p_a(x)\}_a$

writing $\ln p_a(x) = u(a, x)$, our problem becomes the RI problem:

$$\begin{aligned} \max \quad & \mathrm{E}\left[u\left(\hat{a}, \hat{x}\right) + \mathrm{H}\left(\tilde{q}(x \mid \hat{z})\right)\right] \\ \text{s.t.:} \quad & \mathrm{E}_{\tilde{q}(z)}\, \tilde{q}(x \mid \hat{z}) = q_0(x) \\ & \tilde{q}(x \mid z) \in \bar{\mathcal{Q}} \end{aligned}$$

Concavification of the Augmented Value Function



Base-Rate Neglect

comparative statics w.r.to the true process

local invariance

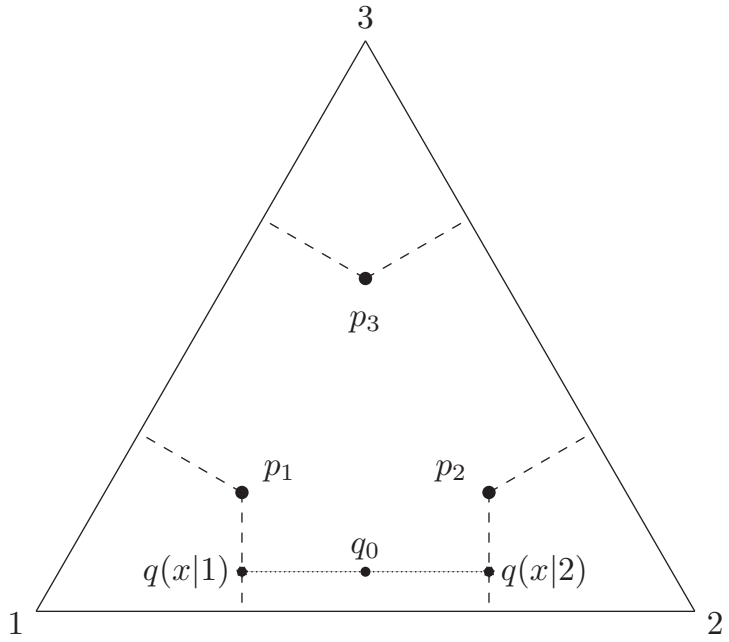
Let true process $q_0^*(x)$ induce posteriors by $p^*(x \mid z)$ and $q^*(x \mid z)$.
For all processes $q_0(x)$ in the convex hull of $\left(q^*(x \mid z)\right)_z$:

$$\begin{aligned} p(x \mid z) &= p^*(x \mid z) \\ q(x \mid z) &= q^*(x \mid z). \end{aligned}$$

Illustration

Matysková&Montes'23

no updating constraint



generative model employs 1, 2, or 3 primitive distributions

- accompanied by a recognition model of the same complexity

Hallucination

optimal recognition model may hallucinate:

- there may exist z and z' such that

$$\begin{aligned} p(x \mid z) &= p(x \mid z') \\ q(x \mid z) &\neq q(x \mid z') \end{aligned}$$

this cannot happen when $\bar{\mathcal{Q}}$ is convex

- akin to the recommendation lemma in RI
- beneficial randomization over $q(x \mid z)$

Literature

non-Bayesian updating: Dominiak, Kovach & Tserenjigmid '21; Jakobsen '21; Ortoleva '12; Zhao '22

machine learning: Caplin, Martin & Marx '23; Zhao, Ke, Wang & Hsieh '20; Aridor, da Silveira & Woodford '24

Bayesian networks: Spiegler '16,'20; Sloman '05; Pearl '88; Ambuehl & Thysen '24; Andre, Haaland, Roth & Wohlfart '23

inconsistent updating: Aina, Amelio & Brütt '23; Bohren & Hauser '23

misspecified learning: Esponda & Pouzo '16; Fudenberg, Lanzani & Strack '21; Frick, Iijima, & Ishii '23

Conclusion

machine learning $\rightarrow$ economics:

- updating and likelihood evaluation are hard
- two distinct statistical models are handy
- tractable constrained updating and model-fitting problems

economics $\rightarrow$ machine learning:

- relaxed constraints can generate solutions with interesting structures
- optimal models are often simple

Approximation

correspondence $\mathcal{Q}(\theta)$, $\theta \in [0, 1]$

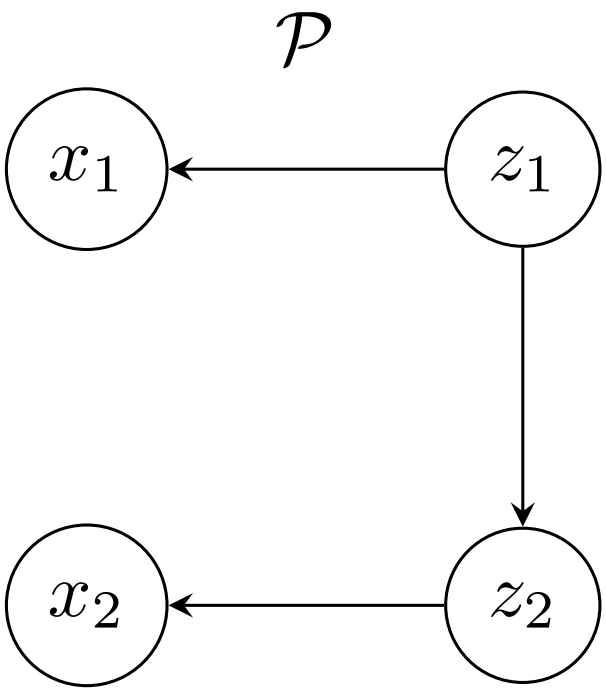
$$\mathcal{Q}(0) = \mathcal{Q} \cap \left\{ \tilde{q}(x, z) : \tilde{q}(x) = q_0(x) \right\}$$

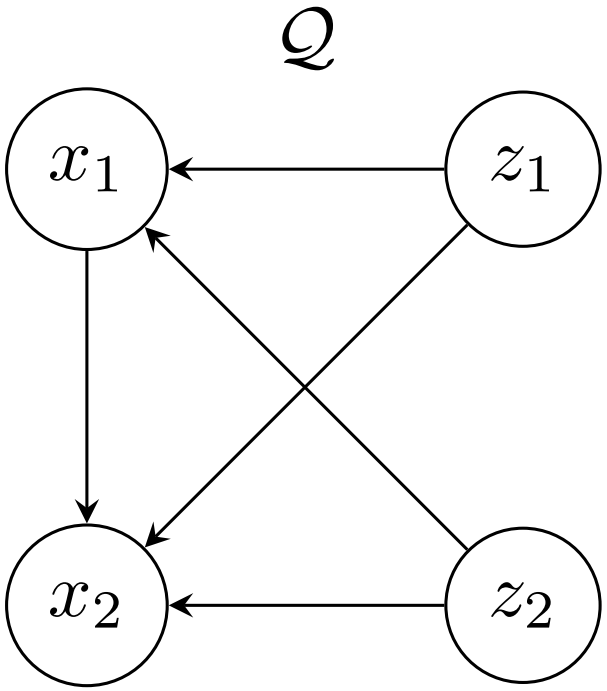
$\mathcal{Q}(\theta) = \mathcal{Q}^{\lfloor \frac{1}{\theta} \rfloor}$ for $\theta > 0$

continuity at $\theta = 0$

back

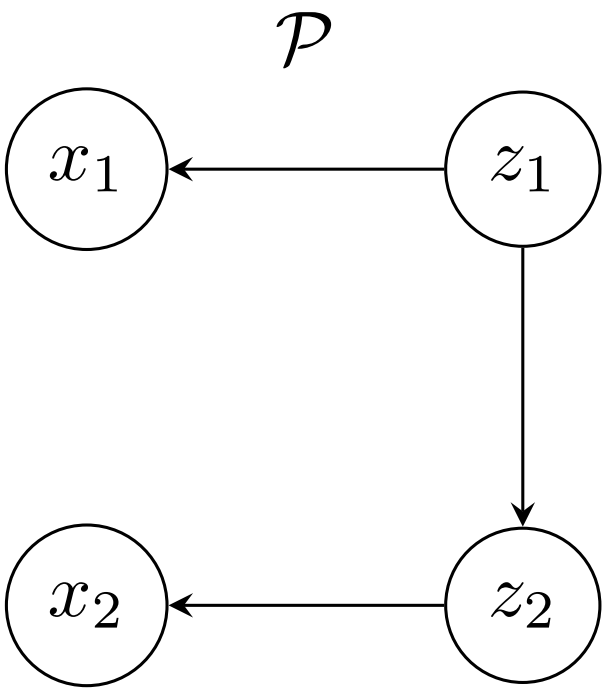
Proof: Correlation Neglect

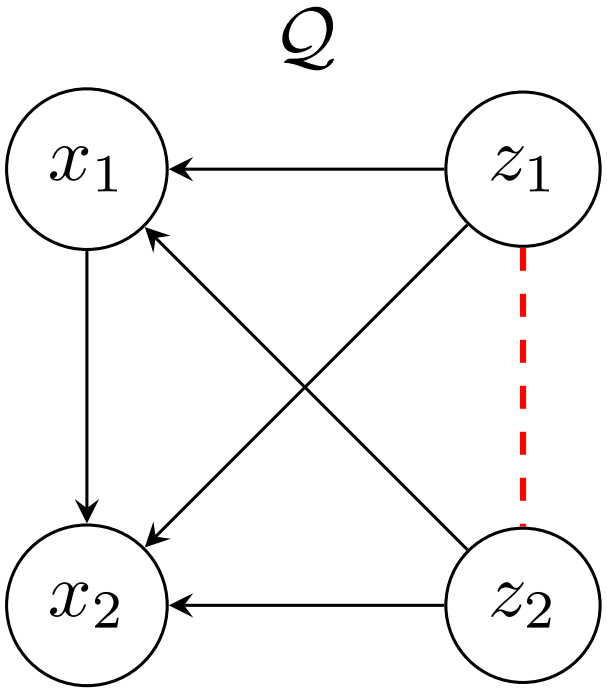




back

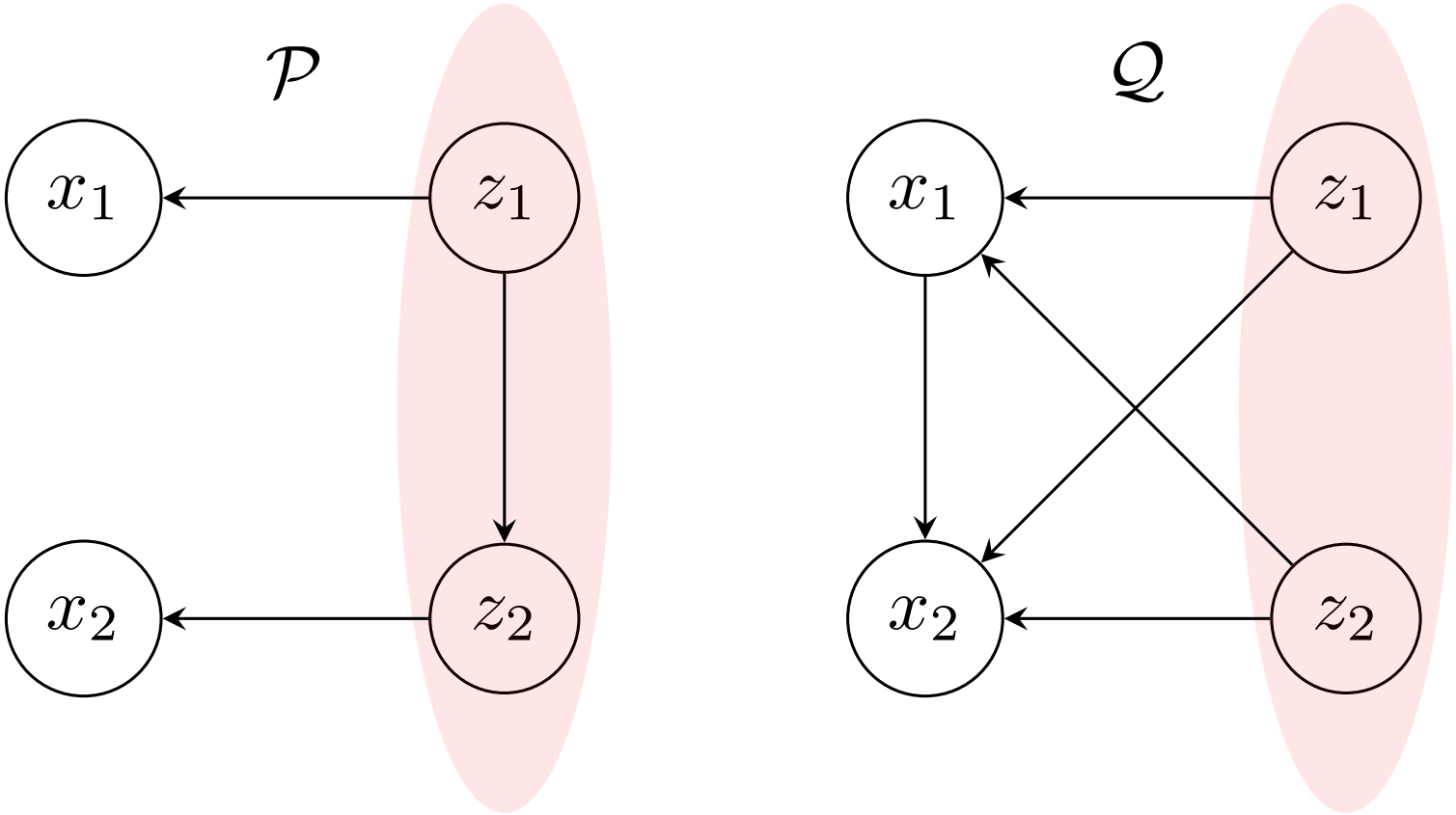
Proof: Correlation Neglect





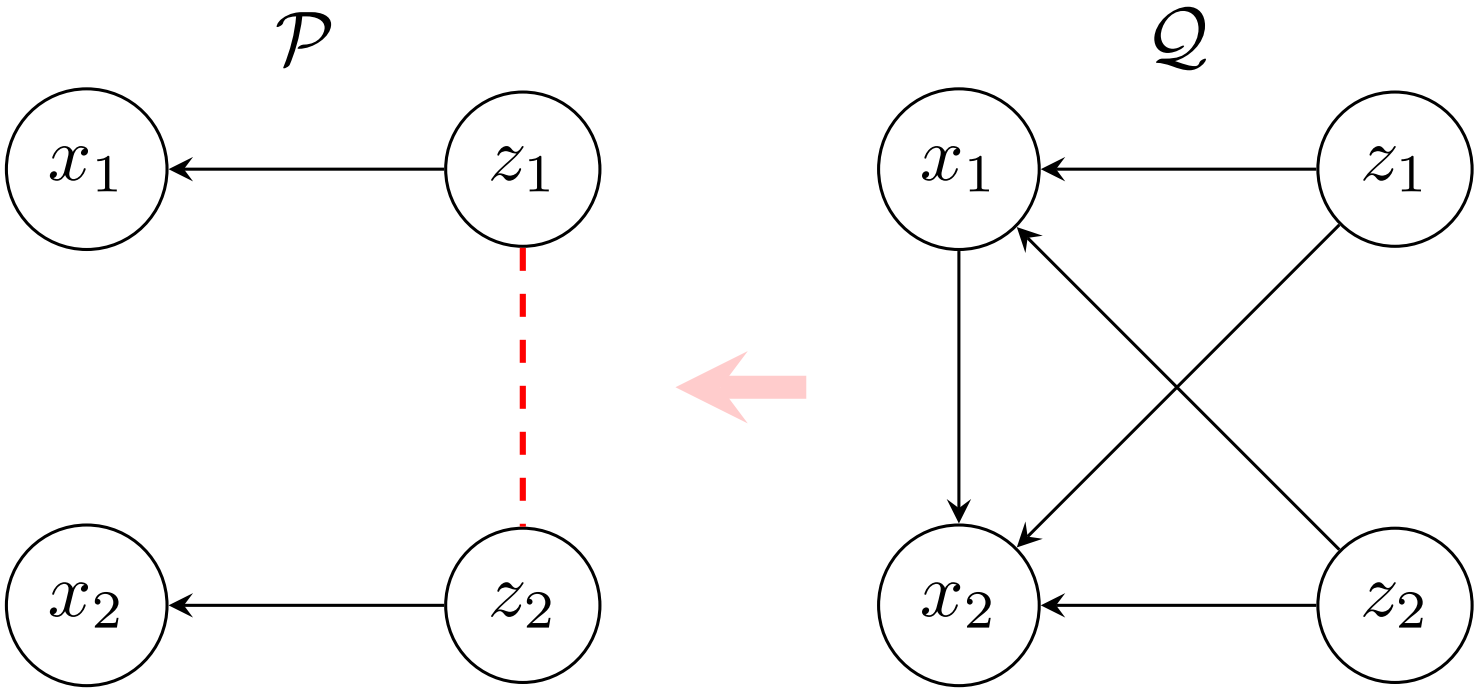
back

Proof: Correlation Neglect



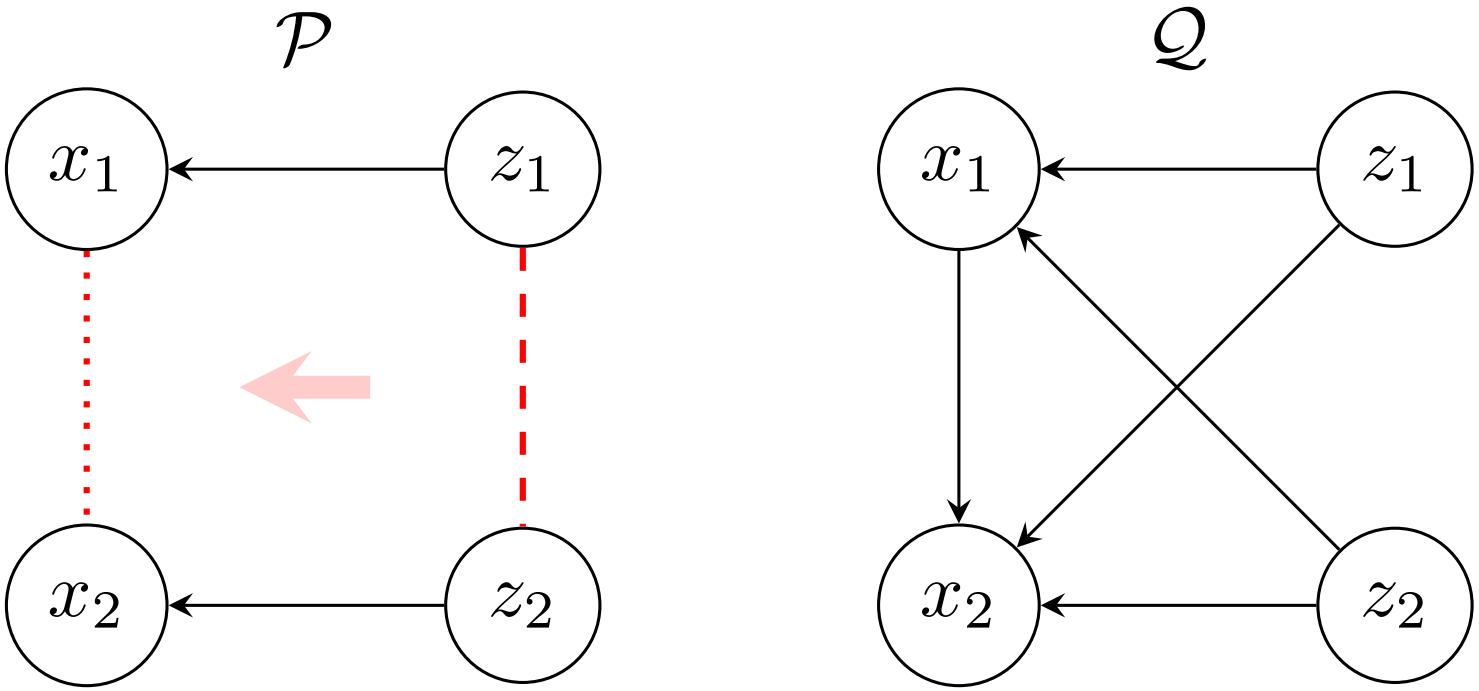
back

Proof: Correlation Neglect



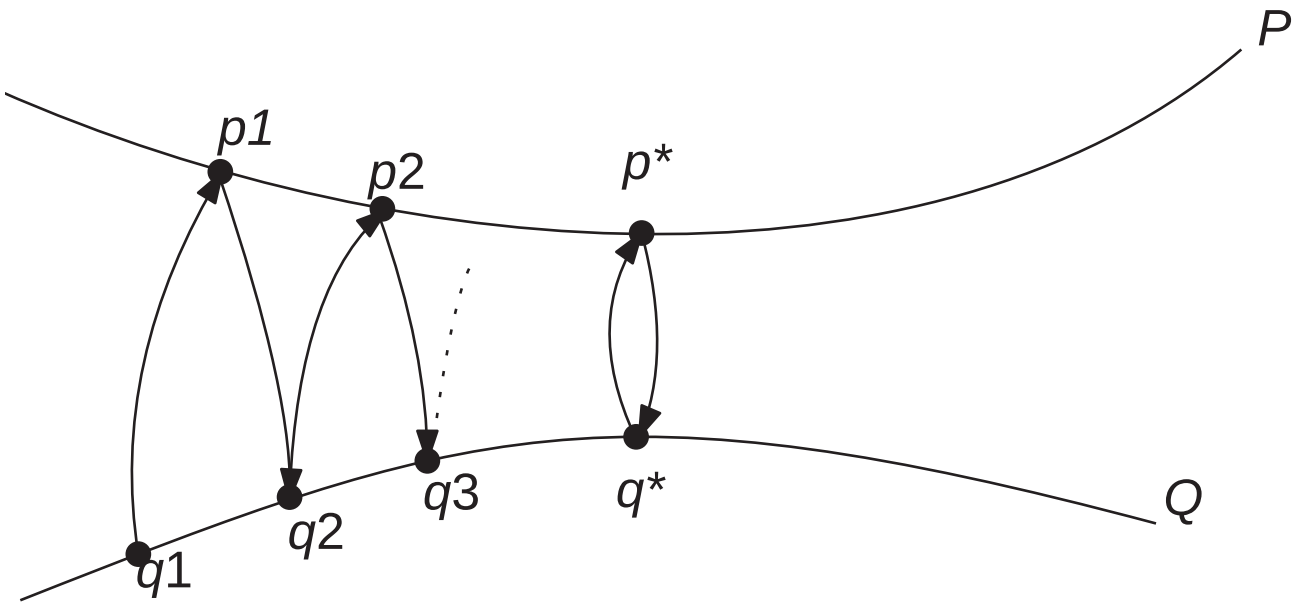
back

Proof: Correlation Neglect



back

Information Geometry



back