Growth and Likelihood

Larry Samuelson, Jakub Steiner

Paris, October 2023

Stochastic Growth

stochastic growth: multiplication of random variables

in economics

- wealth is a product of random returns

in statistics

- likelihood of sample is a product of likelihoods of random data points

growth-rate maximization in both cases

Consistency Principles

Optimal policy seeks consistency with outcomes it generates.

in economics

- meritocracy: consistency between wealth and merit shares

in predictive coding

- consistency between prediction and sensory information

Literature

economics

- redistribution may enhances growth: Stiglitz'69, Aghion et al'95, Barro'00
- empiric of meritocracy: Almås et al'20, Cappelen et al'23, Andre'22

information theory:
Kelly'56, Csiszar&Tusnady'84, Cover&Thomas'06

machine learning and predictive coding:
Dayan et al'95, Friston'05, Kingma&Welling'14, Aridor, da Silveira&Woodford'20,23

On the Notation

probabilities $p(x)$ and likelihoods $p(y \mid x)$ induce

$$\begin{aligned} p(x, y) &= p(x)p(y \mid x) \\ p(y) &= \sum_x p(x, y) \\ p(x \mid y) &= \frac{p(x, y)}{p(y)} \end{aligned}$$

On the Notation

probabilities $q(x)$ and generalized likelihoods $q(y \mid x) \geq 0$ induce

$$
\begin{aligned}
q(x, y) &= q(x)q(y \mid x) \\
q(y) &= \sum_x q(x, y) \\
q(x \mid y) &= \frac{q(x, y)}{q(y)}
\end{aligned}
$$

On the Notation

probabilities $\mathbf{q}(x)$ and generalized likelihoods $q(y \mid x) \geq 0$ induce

$$\begin{aligned} q(x, y) &= \mathbf{q}(x) q(y \mid x) \\ q(y) &= \sum_x q(x, y) \\ \mathbf{q}(x \mid y) &= \frac{q(x, y)}{q(y)} \end{aligned}$$

properly normalized distributions in **bold**

Model

agents $i \in I$, discrete time

stationary allocation $\mathbf{q}(i)$

- each i receives share $\mathbf{q}(i)$ of the aggr. wealth each morning

gross return $r(i, \omega_t) \geq 0$

- iid shocks $\omega_t \sim \mathbf{p}^0(\omega)$
- each t, wealth of each agent i is multiplied by $r(i, \omega_t) \geq 0$

planner maximizes long-run growth rate

$$\max_{\mathbf{q}(i), r(\omega|i)} \mathrm{E}_{\mathbf{p}^0} \ln \left(\sum_i \mathbf{q}(i) r(i, \omega) \right)$$

Model

agents $i \in I$, discrete time

stationary allocation $\mathbf{q}(i)$

- each i receives share $\mathbf{q}(i)$ of the aggr. wealth each morning

gross return $q(\omega_t \mid i) \geq 0$

- iid shocks $\omega_t \sim \mathbf{p}^0(\omega)$
- each t, wealth of each agent i is multiplied by $q(\omega_t \mid i) \geq 0$

planner maximizes long-run growth rate

$$\max_{\mathbf{q}(i), q(\omega|i)} \mathrm{E}_{\mathbf{p}^0} \ln \left(\sum_i \mathbf{q}(i) q(\omega \mid i) \right)$$

Model

agents $i \in I$, discrete time

stationary allocation $\mathbf{q}(i)$

- each i receives share $\mathbf{q}(i)$ of the aggr. wealth each morning

gross return $q(\omega_t \mid i) \geq 0$

- iid shocks $\omega_t \sim \mathbf{p}^0(\omega)$
- each t, wealth of each agent i is multiplied by $q(\omega_t \mid i) \geq 0$

planner maximizes long-run growth rate

$$\max_{q(i,\omega) \in \mathcal{Q}} \mathrm{E}_{\mathbf{p}^0} \ln \left(\sum_i q(i,\omega) \right)$$

$\mathcal{Q}$: set of feasible policies $q(i,\omega)$

Model

agents $i \in I$, discrete time

stationary allocation $\mathbf{q}(i)$

- each i receives share $\mathbf{q}(i)$ of the aggr. wealth each morning

gross return $q(\omega_t \mid i) \geq 0$

- iid shocks $\omega_t \sim \mathbf{p}^0(\omega)$
- each t, wealth of each agent i is multiplied by $q(\omega_t \mid i) \geq 0$

planner maximizes long-run growth rate

$$\max_{q(i,\omega)\in\mathcal{Q}} \mathrm{E}_{\mathbf{p}^0} \ln q(\omega)$$

$\mathcal{Q}$: set of feasible policies $q(i, \omega)$

Example

$\omega \sim \mathbf{p}^0 \in \Delta(I)$

fixed returns: $q(\omega \mid i) = \mathbb{1}_{\omega=i}$

unconstrained allocation: $\mathbf{q}(i) \in \Delta(I)$

optimal allocation: $\mathbf{q}^*(i) = \mathbf{p}^0(i)$

equivalent to Kelly's betting

Merit Distribution

share of aggr. wealth produced by agent i in a period with shock ω

$$\frac{\mathbf{q}(i)q(\omega \mid i)}{\sum_j \mathbf{q}(j)q(\omega \mid j)} := \mathbf{q}(i \mid \omega)$$

definition

merit distribution: share of aggr. wealth produced by i in a random period

$$\mathbf{m}_{\mathbf{q}}(i) = \mathrm{E}_{\mathbf{p}^0(\omega)}\, \mathbf{q}(i \mid \omega)$$

Naive Meritocracy Principle

Proposition

Growth-maximizing allocation $\mathbf{q}^*$ minimizes KL-divergence from the induced merit:

$$\mathbf{q}^*(i) \in \operatorname*{arg\,min}_{\mathbf{q}(i) \in \mathcal{A}^*} \mathrm{KL}\left(\mathbf{m}_{\mathbf{q}^*}(i) \parallel \mathbf{q}(i)\right);$$

$\mathcal{A}^*$ is the set of feasible allocations given optimized return $q^*(\omega \mid i)$.

recall: KL-divergence is a pseudo-distance between two distributions

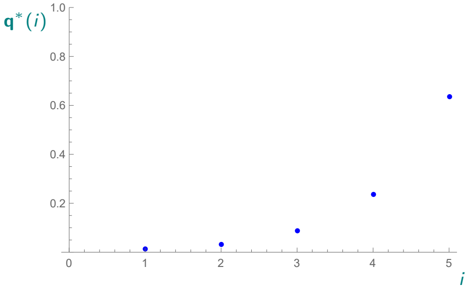
Example

no uncertainty

agent $i = 1, \ldots, 5$ has a return i

an inequality constraint $\mathrm{H}(\mathbf{q}(i)) \geq 1$

growth-optimal allocation



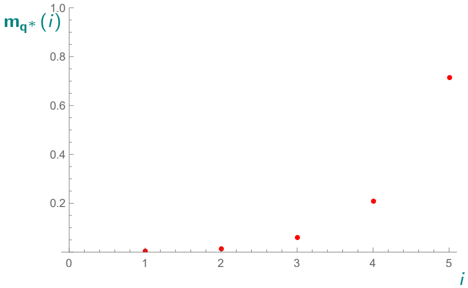
Example

no uncertainty

agent $i = 1, \ldots, 5$ has a return i

an inequality constraint $\mathrm{H}(\mathbf{q}(i)) \geq 1$

induced merit



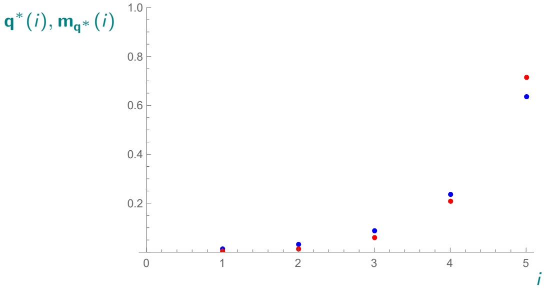
Example

no uncertainty

agent $i = 1, \ldots, 5$ has a return i

an inequality constraint $\mathrm{H}(\mathbf{q}(i)) \geq 1$

naive meritocracy returns the growth-optimal allocation



Naiveté

planner doesn't minimize the wedge between allocation and merit

- the principle ignores endogeneity of merit

Peter Andre: Shallow Meritocracy

- people don't incorporate indirect effects into their merit judgements

sensitive comparative statics

- positive feedback loop

Extension of KL-divergence

$$\mathrm{KL}(\mathbf{p} \parallel \mathbf{q}) := \sum_x \mathbf{p}(x) \ln \frac{\mathbf{p}(x)}{\mathbf{q}(x)}$$

a map $\Delta(X) \times \Delta(X) \to \mathbb{R}_+ \cup \{\infty\}$

the distribution "most consistent" with $\mathbf{q}$ is $\mathbf{q}$

$$\mathbf{q} \in \arg\min_{\mathbf{p}} \mathrm{KL}(\mathbf{p} \parallel \mathbf{q})$$

Extension of KL-divergence

$$\mathrm{KL}(\mathbf{p} \parallel q) := \sum_x \mathbf{p}(x) \ln \frac{\mathbf{p}(x)}{q(x)}$$

a map $\Delta(X) \times \mathbb{R}_+^X \to \mathbb{R} \cup \{\infty\}$

the distribution "most consistent" with q is the normalization of q

$$\frac{q(x)}{\sum_{x'} q(x')} \in \arg\min_{\mathbf{p}} \mathrm{KL}(\mathbf{p} \parallel q)$$

Auxiliary Problem

Theorem

Policy $q^*(i,\omega)$ maximizes the growth rate if and only if solves

$$\min_{\mathbf{p}(i,\omega),q(i,\omega)} \quad \mathrm{KL}\left(\mathbf{p}(i,\omega) \,\|\, q(i,\omega)\right)$$

$$\text{s.t.} \quad \mathbf{p}(\omega) = \mathbf{p}^0(\omega)$$

$$q(i,\omega) \in \mathcal{Q},$$

together with some $\mathbf{p}^*(i,\omega)$.

Additionally, $\mathbf{p}^*(i)$ is the merit distribution induced by $q^*(i,\omega)$.

interpretation of $\mathbf{p}(i,\omega)$?

Dynasties of Dollars

a dynasty originates in \$1 at $t = 1$

it moves from one agent to another

it multiplies by the return of its current owner

define path $\mathbf{p}(i, \omega)$ as

- the empirical frequency of a dynasty being in hands of i and state ω

all paths s.t: $\mathbf{p}(\omega) = \mathbf{p}^0(\omega)$ are followed by some dynasties

Wealth of Paths

wealth of dynasties with path $\mathbf{p}(i, \omega)$ grows at rate,

$$-\operatorname{KL}\left(\mathbf{p}(i, \omega) \parallel q(i, \omega)\right)$$

this summarizes

- frequencies with which the path enjoys returns $q(\omega \mid i)$
- chance of experiencing atypical path $\mathbf{p}(i, \omega)$

the fastest growing path dominates

asymptotically, all wealth is generated by an atypical path of money

Lemma

For any $q : X \to \mathbb{R}_{++}$,

$$\ln \sum_x q(x) = -\min_{\mathbf{p}} \mathrm{KL}(\mathbf{p} \parallel q).$$

1. what are these distributions $\mathbf{p}$?
2. why $\mathbf{p}$ consistent with q?

Proof

set up a growth process

$$y_t = \left(\sum_x q(x) \right)^t$$

the expression of interest is its growth rate

$$\ln \sum_x q(x) = \frac{1}{t} \ln y_t$$

sum over sequences

$$y_t = \sum_{(x_1,\ldots,x_t)} \prod_{t'} q(x_{t'})$$

the summands depend only the empirical distribution $\mathbf{p}$ of the sequence

Proof

\# of sequences with an empirical distribution $\mathbf{p}$ is $\approx \exp[\mathrm{H}(\mathbf{p})t]$

$$y_t = \sum_{\mathbf{p}} \prod_{x} q(x)^{\mathbf{p}(x)t} \exp[\mathrm{H}(\mathbf{p})t]$$

process y_t is a sum of exponential growths

$$y_t = \sum_{\mathbf{p}} \exp\left[-\mathrm{KL}(\mathbf{p} \parallel q)t\right]$$

the exponential function with the highest exponent dominates

$$y_t \approx \exp\left[-\min_{\mathbf{p}} \mathrm{KL}(\mathbf{p} \parallel q)t\right]$$

Q&A

1. what are these distributions $\mathbf{p}$?
   - each $\mathbf{p}$ corresponds to an empirical distribution of a sequence
   - the original process is a sum of growths across all $\mathbf{p}$
   - the fastest growth dominates
2. why $\mathbf{p}$ consistent with q grows fast?
   - concentrate on x with high $q(x)$
   - but, be random to keep # of sequences high
   - the optimal compromise $\mathbf{p}^*$ matches q

Optimal Path

recall $\mathbf{p}^*(i) = \mathbf{m}_{q^*}(i)$

intuition

- $\mathbf{p}^*(i, \omega)$ is the path that is most consistent with $q^*(i, \omega)$
- for each ω, this requires $\mathbf{p}^*(i \mid \omega) = \mathbf{q}^*(i \mid \omega)$
- the result the follows from definition of merits

$$\mathbf{m}_{q^*}(i) = \sum_{\omega} \mathbf{p}^0(\omega)\mathbf{q}^*(i \mid \omega) = \mathbf{p}^*(i)$$

Proof of Meritocracy Principle

growth-maximizing policy solves

$$\min_{q(i,\omega)} \quad \mathrm{KL}\left(\mathbf{p}^*(i,\omega) \parallel q(i,\omega)\right)$$

$$\text{s.t.} \quad q(i,\omega) \in \mathcal{Q}.$$

Proof of Meritocracy Principle

growth-maximizing policy solves

$$\min_{\mathbf{q}(i)} \left\{ \mathrm{KL}\left(\mathbf{p}^*(i) \parallel \mathbf{q}(i)\right) + \sum_i \mathbf{p}^*(i)\, \mathrm{KL}\left(\mathbf{p}^*(\omega \mid i) \parallel q(\omega \mid i)\right) \right\}$$

$$\text{s.t.} \quad \mathbf{q}(i) \in \mathcal{A}^*.$$

Additional Principle

Reward each agent with high returns in states in which she is productive.

Proposition

Optimal return $q^*(\omega \mid i)$ of each agent i solves

$$\min_{q(\omega|i) \in \mathcal{R}_i^*} \mathrm{KL}\left(\mathbf{m}_{q^*}(\omega \mid i) \parallel q(\omega \mid i)\right),$$

where $\mathbf{m}_{q^*}(\omega \mid i)$ is prob. of state ω conditional on \$ being produced by i.

Learning: Growth Perspective
Berk'66, White'82

sample $(\omega_1, \ldots, \omega_t)$ from $\mathbf{p}(\omega)$

likelihood of sample under a hypothesis $\mathbf{q}(\omega)$ grows at rate

$$-\,\mathrm{KL}(\mathbf{p} \parallel \mathbf{q}) - \mathrm{H}(\mathbf{p})$$

$\Rightarrow$ a statistician converges to hypothesis $\mathbf{q}^* \in \arg\min_{\mathbf{q}} \mathrm{KL}(\mathbf{p} \parallel \mathbf{q})$

Predictive Coding

variational autoencoder

a system

- samples signals ω from $\mathbf{p}^0(\omega)$
- seeks to form belief about a cause i of the signal ω
- entertains a set $\mathcal{Q}$ of distributions $\mathbf{q}(i, \omega)$

chooses the best fit

$$\mathbf{q}^*(i, \omega) \in \underset{\mathbf{q}(i,\omega)\in\mathcal{Q}}{\arg\min} \, \mathrm{KL}\left(\mathbf{p}^0(\omega) \parallel \mathbf{q}(\omega)\right)$$

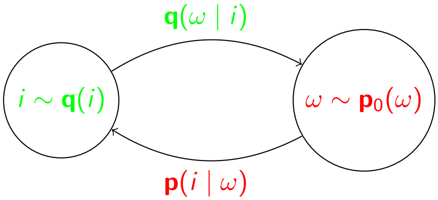
finally, forms Bayesian belief $\mathbf{q}^*(i \mid \omega) = \frac{\mathbf{q}^*(i,\omega)}{\mathbf{q}^*(\omega)}$

Generative and Recognition Models

generative model $\mathbf{q}(i,\omega) \in \mathcal{Q}$: system's internal model of the world

recognition model $\mathbf{p}(i,\omega)$: system's interpretation of the data

- arbitrary belief $\mathbf{p}(i \mid \omega)$ upon observing ω
- data ω are sampled from $\mathbf{p}_0(\omega)$, thus $\mathbf{p}(\omega) \equiv \mathbf{p}_0(\omega)$
- $\mathbf{p}(i,\omega) = \mathbf{p}_0(\omega)\mathbf{p}(i \mid \omega)$



generative and recognition models differ

- but a good pair is as consistent as possible

Variational Characterization

Corollary

The best fit solves

$$\min_{\mathbf{p}(i,\omega),\mathbf{q}(i,\omega)} \quad \mathrm{KL}\left(\mathbf{p}(i,\omega) \,\|\, \mathbf{q}(i,\omega)\right)$$

$$\text{s.t.} \quad \mathbf{p}(\omega) = \mathbf{p}^0(\omega)$$

$$\mathbf{q}(i,\omega) \in \mathcal{Q}.$$

Additionally $\mathbf{p}^*(i \mid \omega)$ is the Bayesian posterior $\mathbf{q}^*(i \mid \omega)$.

proven by a variational argument in machine learning

we provide a growth-based proof

The Connection

optimization of growth rates of multiplicative random processes

- aggregate wealth is a product of random returns

$$\prod_t \left(\sum_i \mathbf{q}(i) r(i, \omega) \right)$$

- likelihood of a sample is a product of likelihoods of data points

$$\prod_t \left(\sum_i \mathbf{q}(i) \mathbf{q}(\omega_t \mid i) \right)$$

economic growth

- policy $q(i, \omega)$
- path $\mathbf{p}(i, \omega)$

predictive sampling

- generative model $\mathbf{q}(i, \omega)$
- recognition model $\mathbf{p}(i, \omega)$

The Connection

optimization of growth rates of multiplicative random processes

- aggregate wealth is a product of random returns

$$\prod_t \left(\sum_i \mathbf{q}(i) q(\omega \mid i) \right)$$

- likelihood of a sample is a product of likelihoods of data points

$$\prod_t \left(\sum_i \mathbf{q}(i) \mathbf{q}(\omega_t \mid i) \right)$$

economic growth

- policy $q(i, \omega)$
- path $\mathbf{p}(i, \omega)$

predictive sampling

- generative model $\mathbf{q}(i, \omega)$
- recognition model $\mathbf{p}(i, \omega)$

Approximate Bayes-Consistency

analogue of the naive meritocracy principle

misspecification $\Rightarrow$ empirical mean of posteriors $\neq$ prior

$$\mathrm{E}_{\mathbf{p}^0(\omega)}\, \mathbf{p}^*(i \mid \omega) \equiv \mathbf{p}^*(i) \neq \mathbf{q}^*(i)$$

Corollary

Optimal generative prior $\mathbf{q}^*(i)$ maximizes consistency with the average recognition posterior:

$$\mathbf{q}^*(i) \in \operatorname*{arg\,min}_{\mathbf{q}(i) \in \mathcal{A}^*} \mathrm{KL}\left(\mathbf{p}^*(i) \parallel \mathbf{q}(i)\right).$$

Rational Inattention

information-acquisition model of Matějka&McKay'15

trade-off between benefit and cost of information

$$\max_{\mathbf{p}(i,\omega)} \mathsf{E}_{\mathbf{p}}\, u(i,\omega) - \mathsf{I}_{\mathbf{p}(i,\omega)}$$

$$\text{s.t.: } \mathbf{p}(\omega) = \mathbf{p}^0(\omega)$$

Robson, Samuelson & Steiner '23:

- growth optimization is equivalent to an RI problem

Details

RSS is a special case of this paper

- exogenous returns $q(\omega \mid i) \equiv e^{u(i,\omega)}$
- no constraints on allocation: $\mathbf{q}(i) \in \Delta(\mathcal{I})$

Proposition (Robson, Samuelson, Steiner '23)

The optimal money path $\mathbf{p}^*(i, \omega)$ solves the RI problem $(\mathbf{p}^0, u)$.

some dollar dynasties enjoy a joint distribution $\mathbf{p}^*(i, \omega)$

- they achieve extraordinary growth rate $\mathrm{E}_{\mathbf{p}^*}\, u(i, \omega)$
- but such lucky dynasties are rare
- tradeoff as between benefit and cost of information

Multiplicity

growth problem with fixed returns and unrestricted allocation

necessary condition on growth-optimal allocation:

- allocation $\mathbf{q}^*(i)$ and merit $\mathbf{m}_{q^*}(i)$ constitute a fixed point.

multiplicity:

- take arbitrary subset of agents $J \subsetneq I$
- solve the growth problem restricted to agents $I \setminus J$
- extend to I by giving agents in J nothing
- we got a fixed point with agents in I!

Transfers of Results

number of agents with positive wealth is at most $|\Omega|$

follows from RI insights by Caplin and Dean

- \# of used actions in RI is at most $|\Omega|$
- implied by concavification and Carathéodory theorem

sufficient and necessary conditions for optimal allocation:

$$\begin{aligned} \sum_{\omega} \frac{\mathbf{p}^0(\omega)}{q^*(\omega)} q(\omega \mid i) &= 1 \text{ } \mathit{if} \text{ } \mathbf{q}^*(i) > 0, \\ \sum_{\omega} \frac{\mathbf{p}^0(\omega)}{q^*(\omega)} q(\omega \mid i) &\leq 1 \text{ } \mathit{if} \text{ } \mathbf{q}^*(i) = 0. \end{aligned}$$

- included agents have equal marginal contribution to growth
- excluded agents have lower marginal contributions

Learning the Growth-Maximizing Policy

extension of Blahut-Arimoto algorithm

start with an arbitrary interior allocation $\mathbf{q}(i)$

compute induced merit $\mathbf{m}_q(i)$

update to the "fairest" allocation $\mathbf{q}'(i) \in \mathcal{A}$ given $\mathbf{m}_q(i)$

iterate

this converges to the optimal policy (under minor conditions)

- Csiszár & Tusnády '84

Summary

we established equivalence between

- economic growth and
- predictive coding

unified consistency principles that apply to both

$\Rightarrow$ a fairness principle in the economic context

growth-based approach as an alternative to the variational approach