Rational beliefs when the truth is not an option

by Filippo Massari & Jonathan Newton

No statistical model is "true" or "false", "right" or "wrong"; the models just have varying performance, which can be assessed.

Jorma Rissanen - Information and complexity in statistical modeling (2007).

The human mind is programmed for survival, not for truth.

John Gray - Seven types of atheism (2018).

A model

Random unobserved **state** of the world	$\omega \in \Omega$
Decision maker chooses **action** x from finite X	$x \in X$
Gives rise to a **consequence**	$y = f(x, \omega) \in Y$
Playing x gives distribution over consequences	$Q(\cdot \mid x)$
Payoff	$\pi(x, y)$
Strategy σ is a distribution over X	σ
$\theta \in \Theta$ is a parameter vector (beliefs, model etc.)	$\theta \in \Theta$
Each θ associated with distributions over consequences	$Q_\theta(\cdot \mid x)$

Solution concepts for such a setting?

Best response (uncontroversial)

For given beliefs θ, any actions played should be (subjective) best responses to these beliefs.

$$x \in \arg\max_{x \in X} E_{Q_\theta(\cdot|x)}\, \pi(x,\cdot)$$

Let $\boldsymbol{\Lambda}$ be the set of pairs (θ, x) that satisfy this.

If x is part of a pair $(\theta, x) \in \Lambda$, we say that x is **justifiable** and that it is **justified** by model θ.

Solution concepts for such a setting?

Berk-Nash equilibrium (popular)

For a given strategy σ, beliefs should maximize expected log-likelihood.

$$\theta \in \arg\max_{\theta \in \Theta} E_\sigma E_{Q(\cdot|x)} \log Q_\theta(\cdot\,|x)$$

Equivalent to minimizing weighted KL divergence (for given σ).

Justification is that these are the beliefs that a Bayesian learner would learn.

Berk-Nash equilibrium is a multi-selves model

BNE finds Nash equilibrium of a game with two players,

'Payoffs player' who chooses σ given θ, so that for all x in the support of σ,

$$x \in \arg\max_{x \in X} E_{Q_\theta(\cdot|x)} \pi(x,\cdot)$$

'Beliefs player' who chooses θ given σ,

$$\theta \in \arg\max_{\theta \in \Theta} E_\sigma E_{Q(\cdot|x)} \log Q_\theta(\cdot\,|x)$$

Makes one think… **what is the DM's objective ?**

What is best in life?

Conan the barbarian (1982)

Is BNE DM's objective to maximize payoffs?

No, this is the objective of his 'payoffs self'.

Is BNE DM's objective to have accurate beliefs?

No, this is the objective of his 'beliefs self'.

A Berk Nash-Equilibrium DM is instrumentally rational (see Nozick, 1993) for the objective of attaining **consistency** between a payoffs self and beliefs self.

What is instrumentally rational for a DM who cares about **payoffs and accuracy in a first order sense**?

The Pareto frontier

DM cares about payoffs and accuracy.

$$\Pi(\theta, x) = E_{Q(\cdot|x)} \pi(x, \cdot)$$

$$LL(\theta, x) = E_{Q(\cdot|x)} \log Q_\theta(\cdot\,|x)$$

If you prefer KL, can use that instead

Maximize weighted sum such that actions are best responses to beliefs.

$$\Lambda^*_\alpha = \underset{(\theta,x)\in\Lambda}{\arg\max} \quad \alpha\, \Pi(\theta, x) + (1-\alpha)\, LL(\theta, x),$$

for given $0 \le \alpha \le 1$.

This basically sends us to the Pareto frontier in payoff-accuracy space.

Example – Coin Toss.

0.7 0.3

Heads or tails $\mathrm{P}(\omega = H) = 0.7, \mathrm{P}(\omega = T) = 0.3$

Actions (bets) $X = \{H, T\}$

Consequences $y = f(x, \omega) = \omega$

Payoffs $\pi(x, y) = 1 \text{ if } x = y$

$\pi(x, y) = 0 \text{ if } x \neq y$

Models $\Theta = \{\theta^1, \theta^2\}$ s.t. $Q_{\theta^1}(H) = 0.45,\ Q_{\theta^2}(H) = 0.9$

Example – Coin Toss.

Best response gives $\Lambda = \{(\theta^1, T), (\theta^2, H)\}$ and we obtain

$$\Pi(\theta^1, T) = 0.3 < 0.7 = \Pi(\theta^2, H), \qquad LL(\theta^1, T) > LL(\theta^2, H)$$

Giving,

$$\Lambda_\alpha^* = (\theta^1, T) \text{ if } \alpha < \frac{log\frac{1331}{1024}}{4+log\frac{1331}{1024}}, \qquad \Lambda_\alpha^* = (\theta^2, H) \text{ if } \alpha > \frac{log\frac{1331}{1024}}{4+log\frac{1331}{1024}}.$$

With $\Lambda_\alpha^* = \{(\theta^1, T), (\theta^2, H)\}$ on the boundary.

Aside: BNE for this example is (θ^1, T).

About Λ^*

Look at the extremes of the Pareto frontier.

If $\alpha = 0$, consequences independent of actions & this is known, it is like BNE.

Proposition 1 (DM chooses a model for accuracy alone)

Let Q and Q_θ, for all θ, be independent of x (like in the coin toss example). Then Λ_0^* is the set of pure BNE (in the coin toss example, this is (θ^1, T)).

When $\alpha = 1$, it is like NE of a restricted game.

Proposition 2 (DM chooses a model for payoffs alone)

Λ_1^* is the set of pure NE of the game with action set equal to the set of justifiable actions and payoffs given by $\mathrm{E}_{Q(\cdot|x)}\pi(x,\cdot)$ for each such x (in the coin toss example, this is (θ^2, H)).

About Λ^*

Under independence, if the model is well-specified, then these solutions pretty much converge to the same thing.

Proposition 3 (well-specified, independent)

Let Q and Q_θ, for all θ, be independent of x (like in the coin toss example).
Let the problem be well-specified, so that there exists $\theta^\dagger$ such that $Q = Q_{\theta^\dagger}$.

- If x is a BR to $\theta^\dagger$, then $(\theta^\dagger, x) \in \Lambda^*_\alpha$.
- Conversely, if $(\theta^*, x^*) \in \Lambda^*_\alpha$, then x^* is a BR to $\theta^\dagger$. For $\alpha < 1$, $Q_{\theta^*} = Q_{\theta^\dagger}$.

A notable exception to the last statement is $\alpha = 1$. Consider the coin toss example, but with the true model $p(H) = 0.7$ added to the model set. When $\alpha = 1$, the DM is indifferent between the true model and θ^2.

Beyond independence – example
(Nyarko,1990; Esponda & Pouzo, 2016)

Monopolist has two possible prices $x \in \{2,10\}$

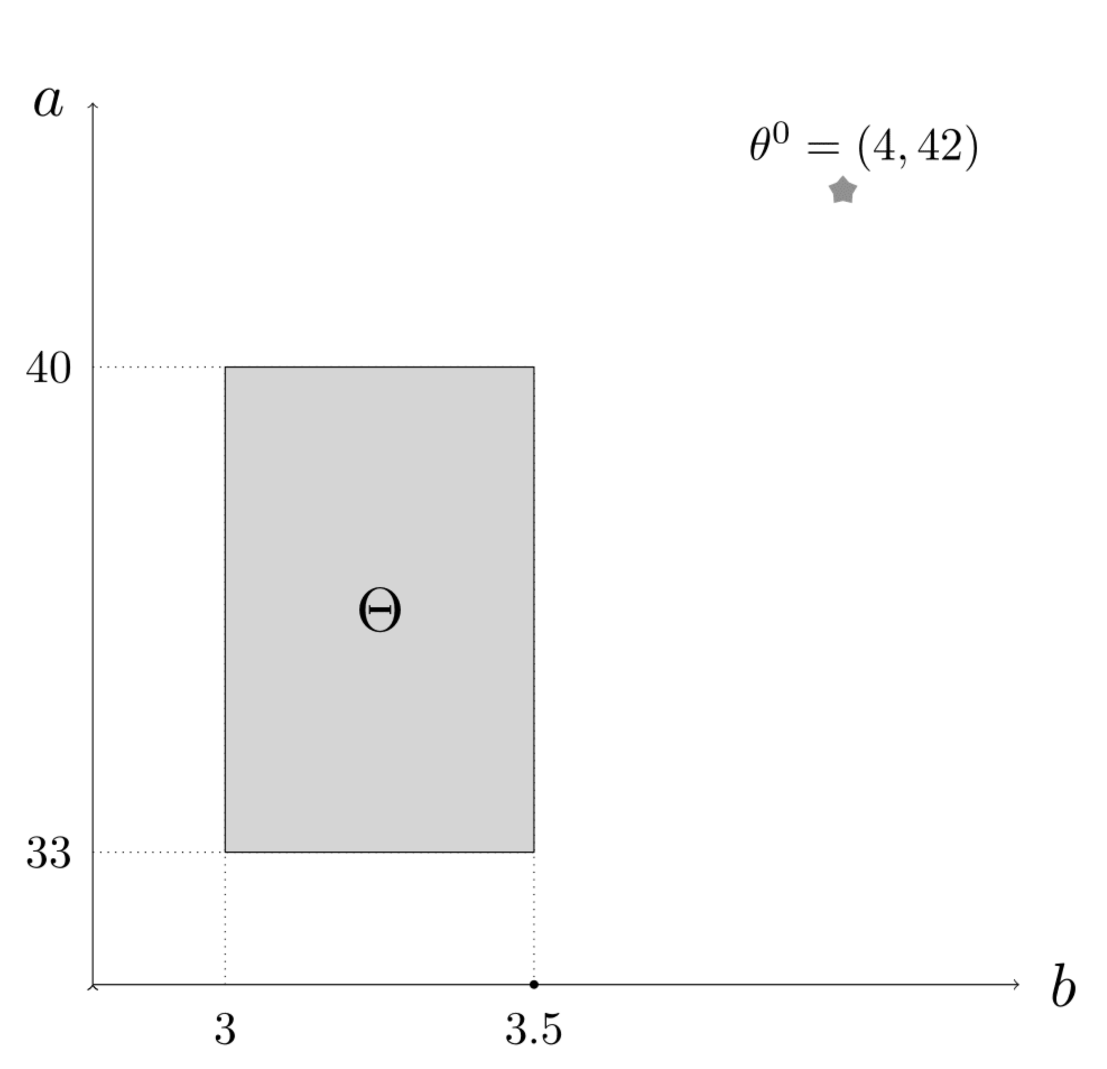
Demand $y = a - bx + \omega$, where $\omega \sim N(0,1)$.

$\theta = (b, a) \in \Theta$ is a parameter vector.

Θ set of parameters that monopolist considers.

Monopolist's payoff $\pi(x, y) = xy$.

True parameters $\theta^0 = (4,42)$.

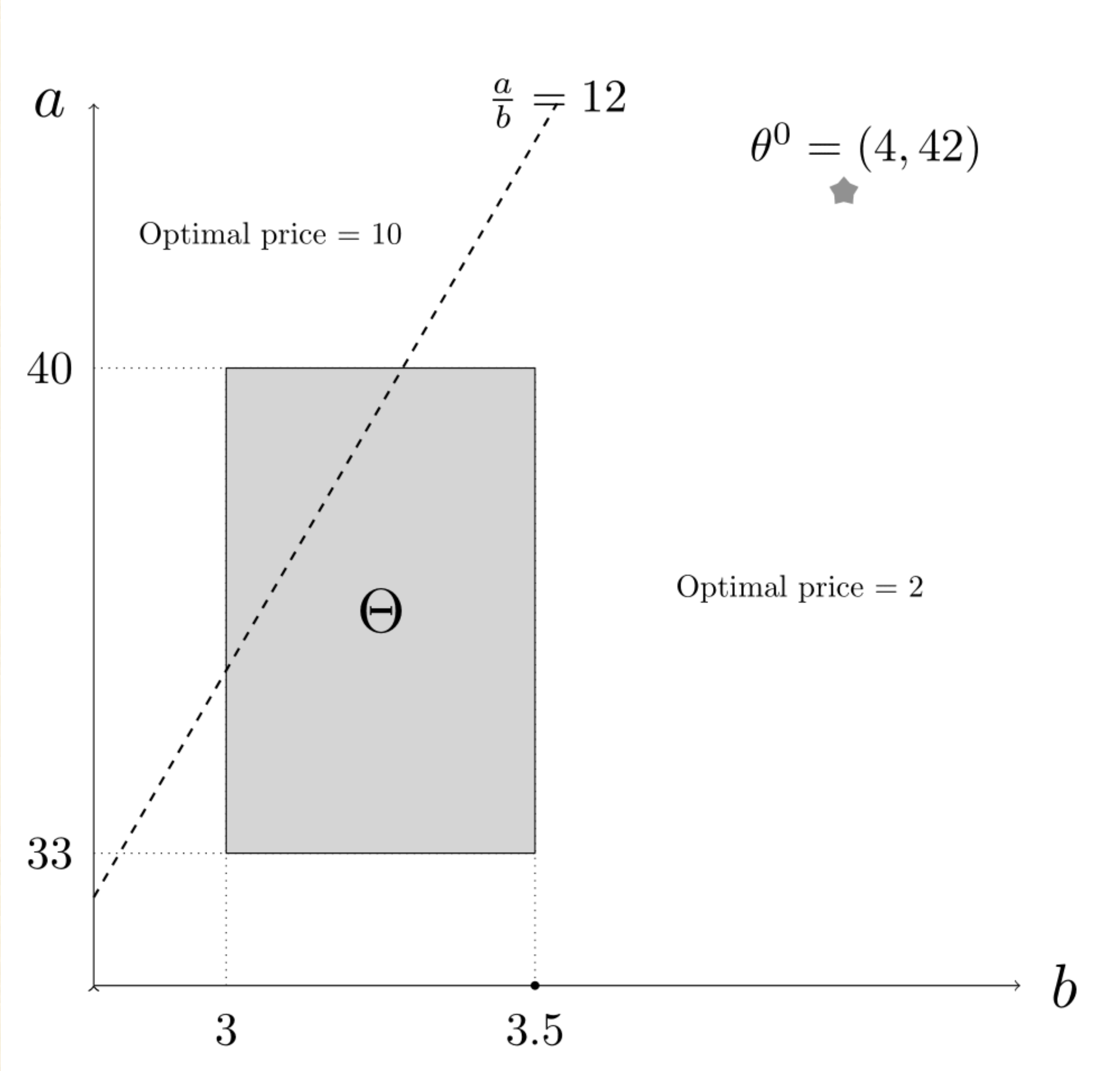


Cycling (Nyarko,1990)

Monopolist has two possible prices $x \in \{2,10\}$

Demand $y = a - bx + \omega$, where $\omega \sim N(0,1)$.

$\frac{a}{b} = 12$ gives threshold parameter values.



Cycling (Nyarko,1990)

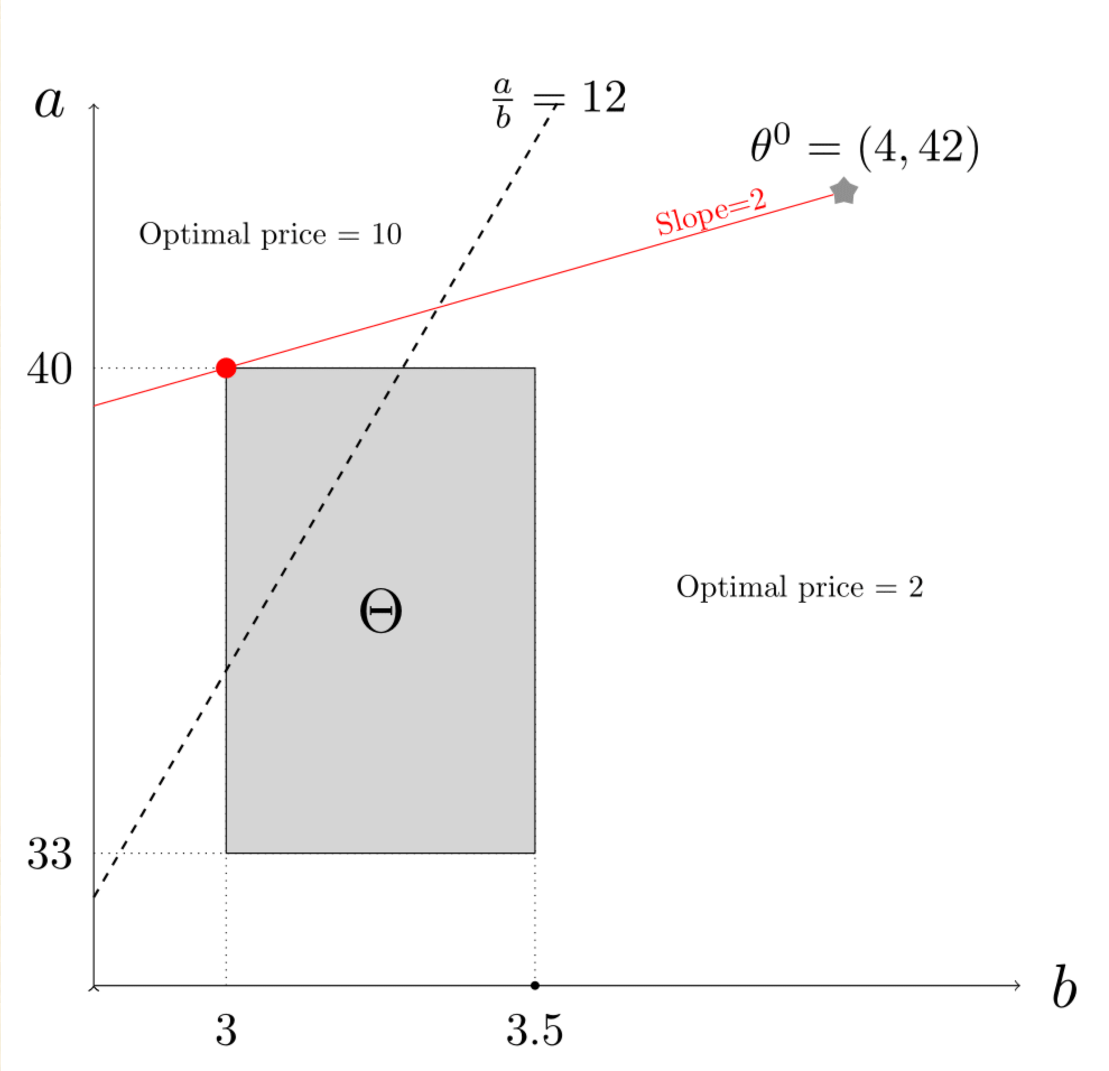
Monopolist has two possible prices $x \in \{2,10\}$

Demand $y = a - bx + \omega$, where $\omega \sim N(0,1)$.

$\frac{a}{b} = 12$ gives threshold parameter values.

Setting $x = 2$ leads to parameters $\theta = (3,40)$ being learned by a standard Bayesian.

However, under these parameters, $x = 10$ maximizes payoff.



Cycling (Nyarko,1990)

Monopolist has two possible prices $x \in \{2,10\}$

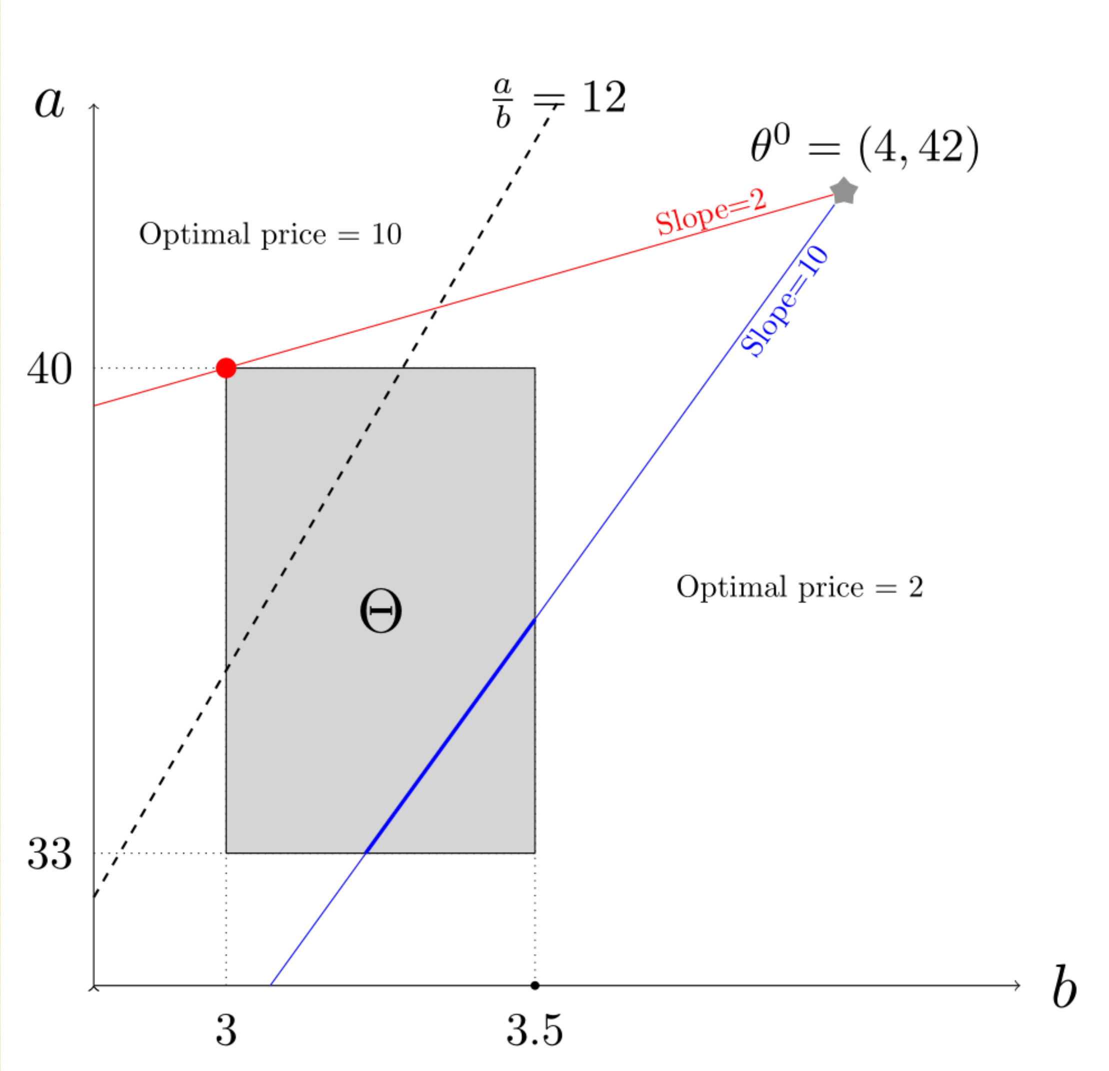
Demand $y = a - bx + \omega$, where $\omega \sim N(0,1)$.

$\frac{a}{b} = 12$ gives threshold parameter values.

Setting $x = 2$ leads to parameters $\theta = (3,40)$ being learned by a standard Bayesian.

However, under these parameters, $x = 10$ maximizes payoff.

Similarly, $x = 10$ leads to a Bayesian learning parameters at which $x = 2$ maximizes payoff.



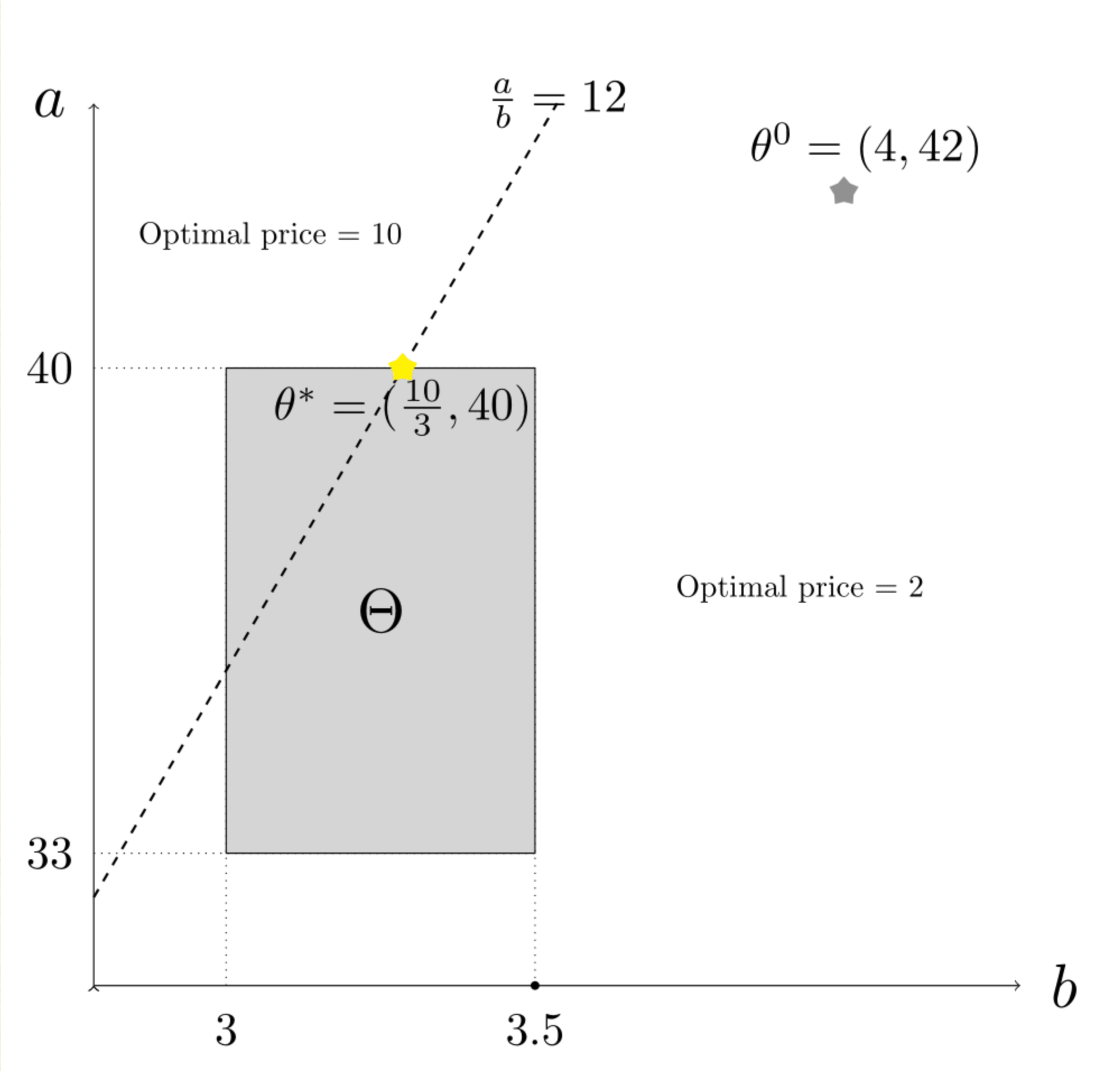
Berk-Nash Equilibrium

(Esponda & Pouzo, 2016)

Monopolist can adopt parameters θ^* that support mixing between the two prices.

Mixing probabilities must be such that a standard Bayesian monopolist learns θ^*.

Elegant solution, but leads to the curious outcome where the monopolist mixes between actions that give different objective payoffs.

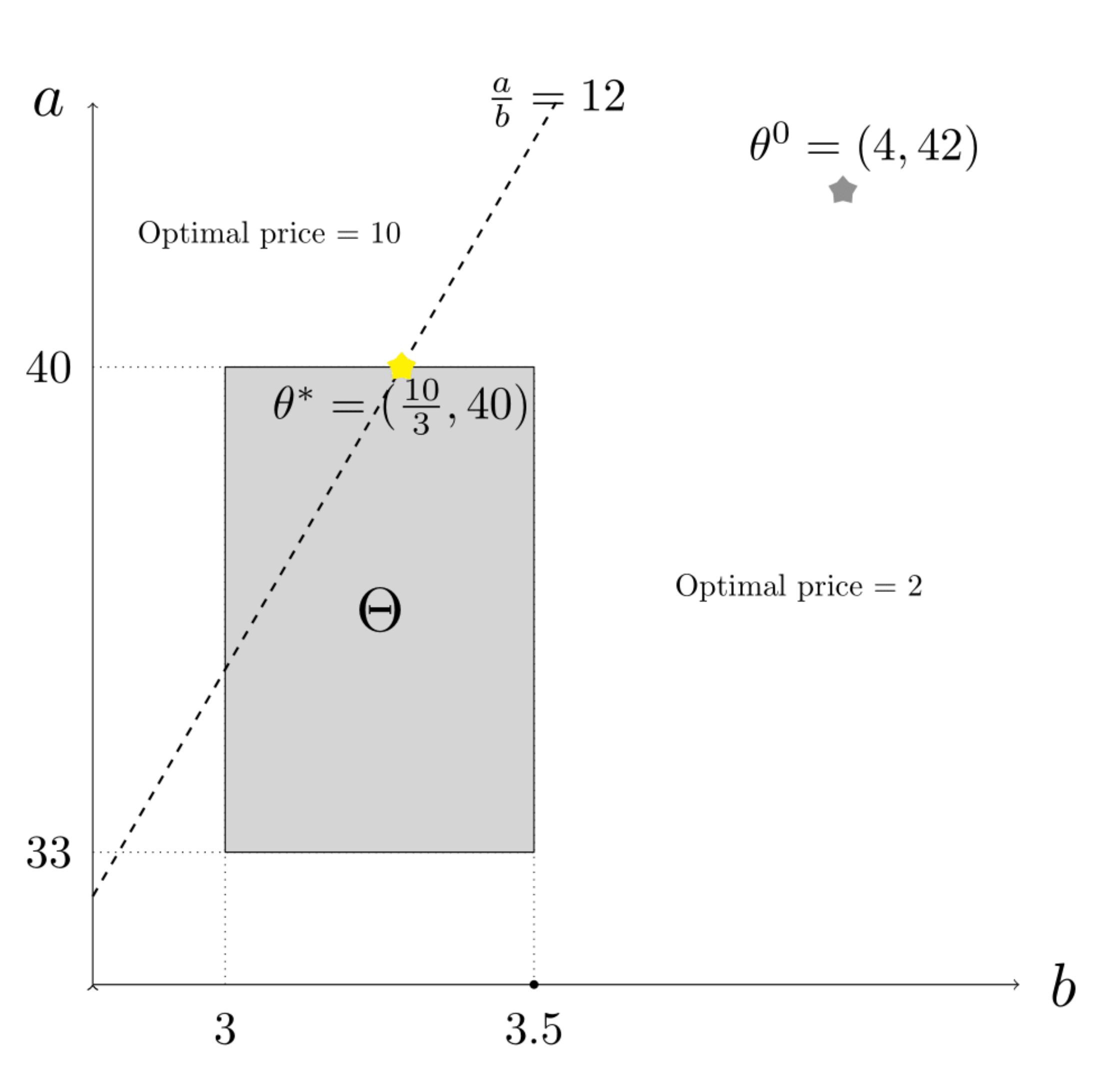


Recall, BNE finds Nash equilibrium of a game between two players

- One player who maximizes payoffs by choosing a strategy given beliefs.
- Another player who maximizes log-likelihood (minimizes K-L) by choosing beliefs given a strategy.

The objective functions of these players are optimized independently. There may exist a profitable 'coalitional' deviation.

In this example, keeping parameters $\boldsymbol{\theta}^*$ while increasing the probability of $\boldsymbol{x = 2}$ increases payoff while decreasing K-L.



Λ_α^* for $0 \le \alpha < 1$.

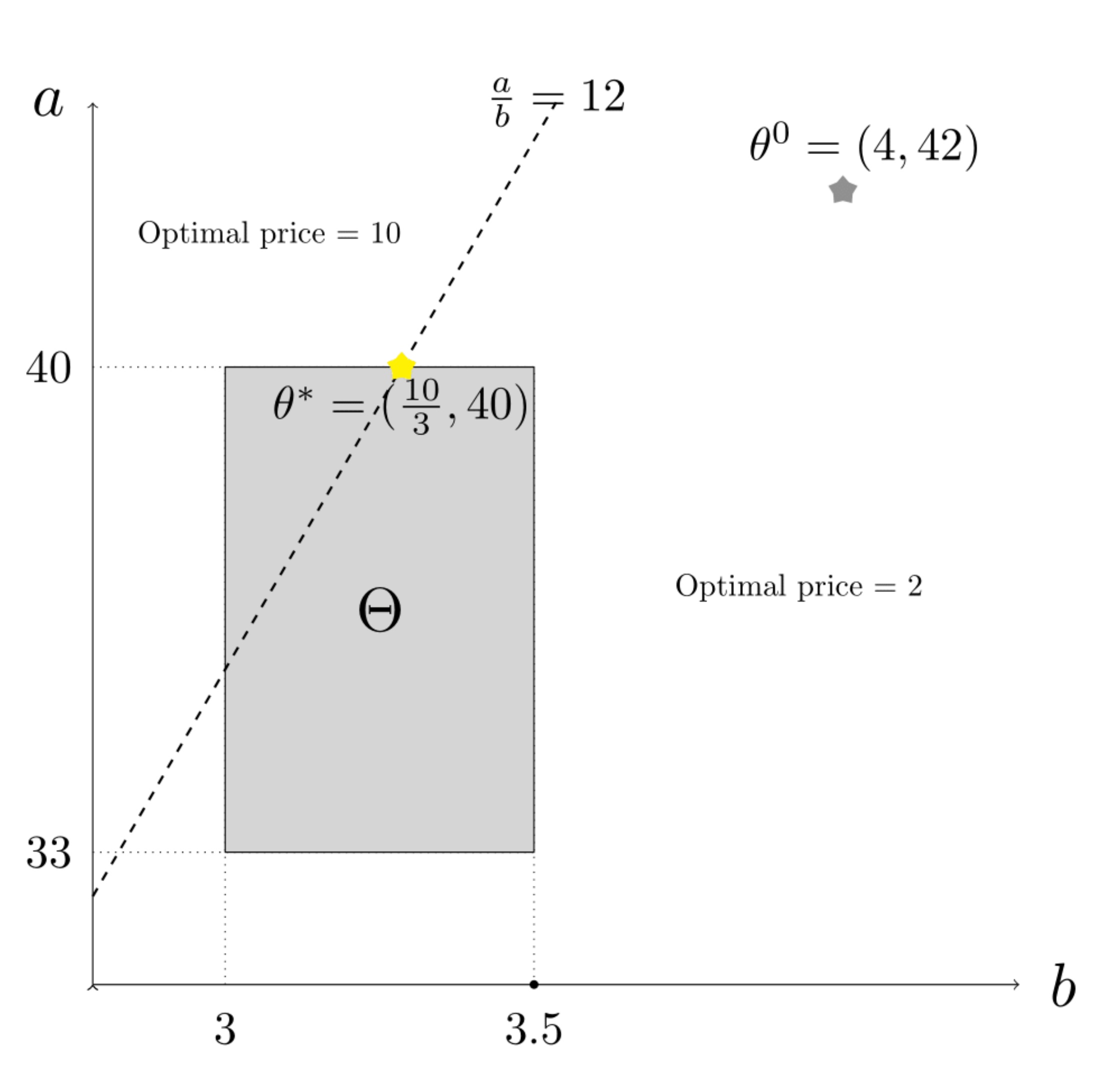
Selects the same parameters as BNE (i.e. θ^*), while $x = 2$ is played with probability one.

This gives strictly higher payoffs & strictly lower K-L divergence than the BNE.

(In general, trade offs exist – e.g. coin toss)

What has been sacrificed?

θ^* is not what a standard Bayesian would learn if we fixed $x = 2$.

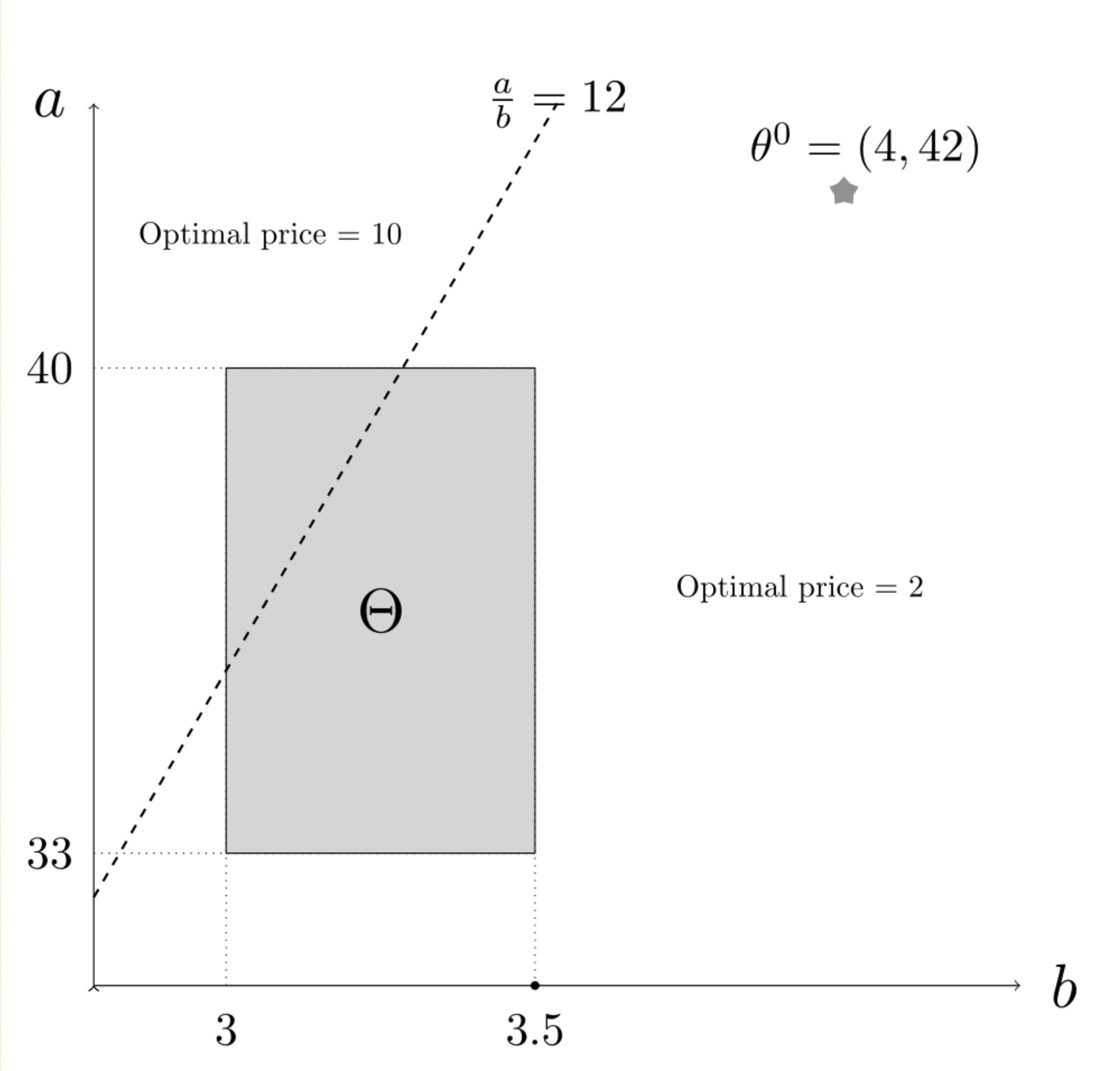


Λ_α^* for $\alpha = 1$.

Selects any parameters that support $x = 2$.

Essentially, standard payoff optimization would occur given the (possibly restricted) set of actions that are justifiable given Θ (Proposition 2).

Can think of Θ as a set of 'stories' that justify actions. Arguably, should select which of these stories lead to higher payoffs.



Potential misunderstanding (contradicted by monopoly example)

The paper points out that BNE may lead the DM to adopt a model that induces a worse action than another model within the support of the DM's prior. But this is a well-known fact, and ***it is roughly because the two objectives of true payoff optimization and statistical fit do not always line up.***

BNE need not be efficient as NE of a 2-player game can be Pareto inefficient (cf. Prisoner's dilemma).

Bold statement is true by definition for Λ^*_α as it traverses the Pareto frontier from $\alpha = 0$ to $\alpha = 1$.

The framing of BNE as the outcome of a dual-self exercise is odd*…and does not reflect the vast majority of the literature on BNE, which presents it as the outcome of a learning in games exercise.*

Learning to maximize the goal function – Generalized Bayes’

- Consider a DM who is part of a (possibly infinite) population.
- In every period $t \geq 1$, for every $(\theta, x) \in \Lambda$, there are some members of the population who follow the model θ and play x.
- Every member of the population experiences the same sequence of states $\{\omega_t\}_t$.
- Assume a uniform bound on the absolute value of log-likelihoods.
- DM learns from realized payoffs and log-likelihoods of players in the population.

Generalized likelihood of $(\theta, x) \in \Lambda$ after t periods is

$$gQ_{(\theta,x)}(\omega_1, \dots, \omega_t) = \prod_{\tau=1}^{t} \exp\left(\alpha\, \pi(x, f(x, \omega_\tau)) + (1-\alpha) \log Q_\theta(f(x, \omega_\tau) | x)\right)$$

Proposition 5 (Learnable by generalized Bayesian updating)

Bayesian updating with these likelihoods leads to probability concentrating on models that are arbitrarily close to maximizing the goal function.

Learning to maximize the goal function – Reinforcement learning

$$gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t) = \prod_{\tau=1}^{t} \exp\left(\alpha\, \pi\left(x, f(x,\omega_\tau)\right) + (1-\alpha) \log Q_\theta\left(f(x,\omega_\tau)|x\right)\right)$$

Set $\alpha = 1$ and take log to get

$$\log gQ_{(\theta,x)}(\omega_1,\ldots,\omega_t) = \sum_{\tau=1}^{t} \pi\left(x, f(x,\omega_\tau)\right)$$

Similar to reinforcement algorithm of Erev & Roth (1998).

Consider finite Λ, $\pi > 0$, relax ass^{n} of updating all $gQ_{(\theta,x)}$ every period & instead update them all at once, then in subsequent periods update each w.p. proportional to $\log gQ_{(\theta,x)}$, then Th^{m} 2 of Beggs(2005) implies posterior concentrates on $\Lambda^*_{\alpha=1}$.

Note, this argument also applies for $\alpha < 1$, providing that $\alpha\, \pi + (1-\alpha) \log \mathrm{Q}_\theta$ is suitably normalized to be strictly positive at all outcomes.

Learning to maximize the goal function – Imitation

Consider a population of unit mass.

Let the share of the population following $\lambda \in \Lambda$ at time t be given by $\varsigma_t(\lambda)$.

Assume (for simplicity) that ς_0 has finite support.

In period t, some share of the population is randomly matched to one another.

If pair following λ^1 and λ^2 are matched and λ^1 Pareto dominates λ^2 in terms of $\Pi(\cdot,\cdot)$ and $LL(\cdot,\cdot)$, then the player following λ^2 switches to λ^1.

Let Λ^{dom} be set of λ Pareto dominated by some λ' with $\varsigma_0(\lambda') > 0$.

> *…a mode of behavior is irrational for a given decision maker, if, when the decision maker behaves in this mode and is then exposed to the analysis of her behavior, she feels embarrassed.*
>
> Itzhak Gilboa – Theory of decision under uncertainty (2009, p.139).

Learning to maximize the goal function – Imitation

Consider a population of unit mass.

Let the share of the population following $\lambda \in \Lambda$ at time t be given by $\varsigma_t(\lambda)$.

Assume (for simplicity) that ς_0 has finite support.

In period t, some share of the population is randomly matched to one another.

If pair following $\lambda^1$and λ^2 are matched and λ^1 Pareto dominates λ^2 in terms of $\Pi(\cdot,\cdot)$ and $LL(\cdot,\cdot)$, then the player following λ^2 switches to λ^1.

Let Λ^{dom} be set of λ Pareto dominated by some λ' with $\varsigma_0(\lambda') > 0$.

Proposition X (Learnable by imitation)

As $t \to \infty$, $\varsigma_t(\Lambda^{dom}) \to 0$.

Learning to maximize the goal function – Evolution

Let realized payoffs be understood as evolutionary fitness (assuming $\pi > 0$).

From period $t-1$ to t, the share of the population playing λ changes proportionally to the mean realized payoff of agents that play λ in period $t-1$, with a normalization so that the total mass of the population remains one.

If agents have independent realizations of the state, then the mean realized payoff of those following a model equals the model's expected payoff.

It follows that models with higher expected payoffs lead to higher growth.

Proposition 6 (Learnable by evolution - independent states)

If agents have independent realizations of the state, then as $t \to \infty$, ς_t comes to place all weight on the (θ, x) in the support of ς_0 that maximize $E_{Q(\cdot|x)}\pi(x,\cdot)$.

That is, evolution selects a model that would be chosen by our goal function with $\alpha = 1$.

Learning to maximize the goal function – Evolution

If agents have the same realization of the state, then maximizing expected fitness for a single period is no longer the same as maximizing long run fitness.

Want to maximize geometric mean rather than arithmetic mean, therefore take logs.

Proposition 7 (Learnable by evolution - shared states)

If agents have the same realization of the state, then as $t \to \infty$, ς_t comes a.s. to place all weight on the (θ, x) in the support of ς_0 that maximize $E_{Q(\cdot|x)} \log \pi(x,\cdot)$.

That is, if we transform payoffs by $\log(\cdot)$, then evolution selects a model that would be chosen by our goal function with $\alpha = 1$.

Further note, with shared states, mixing across models may be beneficial. This is analogous to the literature on risk preferences and selection for genotypes that generate heterogeneous risk preferences in phenotype (Heller & Nehama, 2022).

Summary

To explore evolutionary robustness and plausibility of learning rules under misspecification.

There is other work considering evolutionary robustness under misspecification (Fudenberg & Lanzani; He & Libgober).

This other work takes Bayesian learning (K-L minimization) as given and explores robustness in other dimensions.

Thank you for listening!

Miscellaneous slides from previous presentations

Probability concentrates on the model with the largest expected log likelihood (Berk, 1966).

Blue model: $0.7\log(0.45) + 0.3\log(0.55) \approx -0.74$
Red model: $0.7\log(0.9) + 0.3\log(0.1) \approx -0.76$

Note, this model, the blue model, minimizes Kullback-Leibler divergence to the true model.

$$\mathrm{KL}(\text{blue}) = 0.7\frac{\log(0.7)}{\log(0.45)} + 0.3\frac{\log(0.3)}{\log(0.55)} \approx 0.92$$

$$\mathrm{KL}(\text{red}) = 0.7\frac{\log(0.7)}{\log(0.9)} + 0.3\frac{\log(0.3)}{\log(0.1)} \approx 2.53$$

Back

Comparing expected log generalized likelihoods:

Blue: $0.7\log\frac{e^{0}}{e^{0}+e^{1}} + 0.3\log\frac{e^{1}}{e^{0}+e^{1}} = 0.3 + \mathit{model\ indep\ terms}$

Red: $0.7\log\frac{e^{1}}{e^{0}+e^{1}} + 0.3\log\frac{e^{0}}{e^{0}+e^{1}} = 0.7 + \mathit{model\ indep\ terms}$

Note that the RHS equals expected payoff under the model plus some model independent terms.

Therefore, updating probabilities in the same way as a Bayesian…
…but using generalized likelihoods instead of standard likelihoods…
…concentrates probability on models that **maximize expected payoff**.

Back

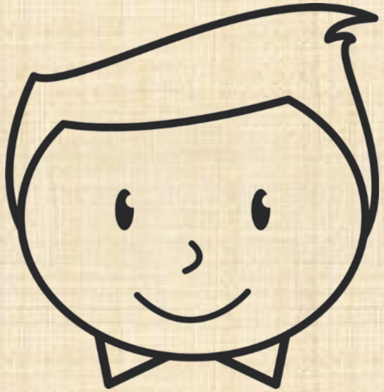
Consider a biased coin that has a true probability of heads of 0.7

0.7

0.3

However Bob only considers the following two possibilities, neither of which corresponds to the truth.

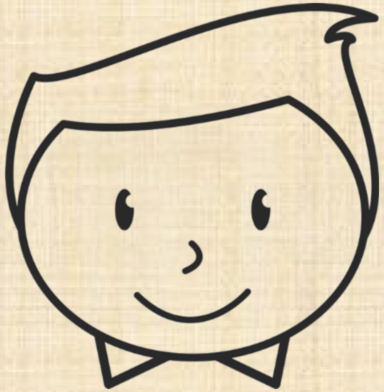
His model is **misspecified**.



Bob

He observes a sequence of tosses of the coin and can bet **heads** or **tails** each time, earning a dollar every time he is correct.

Recall that the true probability of heads is 0.7

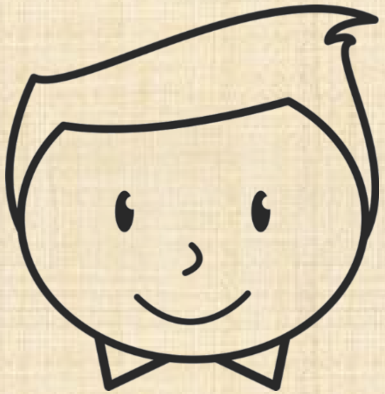


Bob

Bob is a Bayesian learner.

If he places positive initial probability on the **blue** and the **red** models, then over time he places almost all weight on the **blue** model.

Blue model minimizes Kullback-Leibler divergence from the truth.



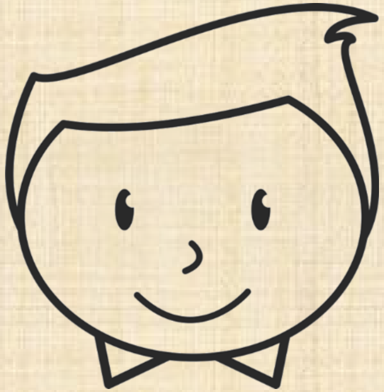
Bob

Detail

Hence he will bet **tails** every period.

As the true probability of **heads** is 0.7, he will earn an average payoff over time of 0.3 dollars per period.

If he had instead learned the **red** model, he would have bet **heads** every period and earned an average payoff of 0.7 dollars!



Bob

Alice is not a Bayesian.

Rather than update her prior probabilities over **red** and **blue** using standard likelihoods, she uses **generalized likelihoods** that depend on payoffs (Grunwald, 1998).

Let's start with the **blue** model.

Recall that a believer in this model will bet on **tails**.

If she followed the **blue** model and **heads** arose, she would obtain a payoff of 0.

If she followed the **blue** model and **tails** arose, she would obtain a payoff of 1.

Take a sequence of observations, say

heads, **tails**, **tails**…

The likelihood of this sequence under the **blue** model is

$$(0.45)(0.55)(0.55)\ \ldots$$

Take a sequence of observations, say
heads, tails, tails…

The **generalized likelihood** under the blue model is

$$\frac{e^{0}}{e^{0}+e^{1}}\cdot\frac{e^{1}}{e^{0}+e^{1}}\cdot\frac{e^{1}}{e^{0}+e^{1}}\cdots$$

Exponents in numerator are the payoffs from the sequence **heads, tails, tails**… when the blue model is followed.

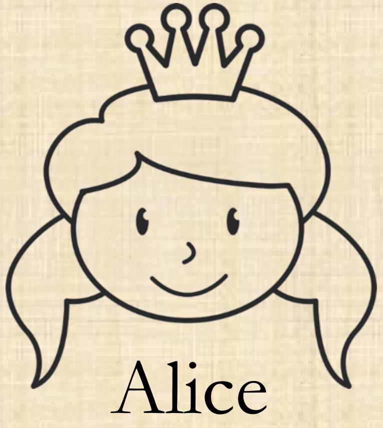
The denominator normalizes.

Do the same for the **red** model.

Now applying standard Bayesian learning using these generalized likelihoods…

…over time she comes to place all weight on the **red** model. Hence she bets **heads** and earns an average payoff of 0.7 dollars.

The model she learns is in some sense **wrong**, but in a pragmatic sense it works out just as well as if she had learnt the correct probability of **heads**.



Detail