14. Strategic Experimentation

ECON 7219 – Games With Incomplete Information

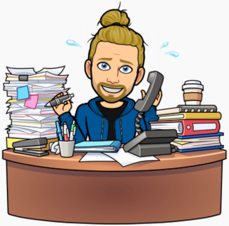
Benjamin Bernard

Getting Lunch on 118 Lane

Choosing a place to get lunch:

- You can either go to one of your staples, where you know the food is good, or you can try new places.
- If you find one dish that is better than what you usually eat, the experimentation was a success.
- A bad dish, however, does not prove that there is no good dish there.

Working on a Proof



Working on a proof:

- You can focus on one of several projects, each having a random reward upon completion and taking a random amount of time to complete.
- A conjecture is either correct (ϑ_C) or it is false (ϑ_F).
- Working on a proof is an experiment that can only reveal state ϑ_C, but it can never be fruitful if the true state is ϑ_F.
- Trying to find a counterexample may reveal ϑ_F if that is the true state.

Adoption of a New Vaccine

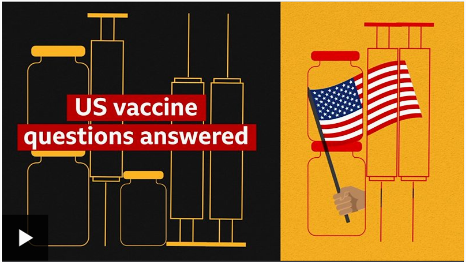
BBC | Sign in | Home | News | Sport | Reel | Worklife | Travel

NEWS

Home | Coronavirus | Video | World | Asia | UK | Business | Tech | Science | Stories | Entertainment & Arts

World | Africa | Australia | Europe | Latin America | Middle East | US & Canada | US Election

Covid: FDA approves Pfizer vaccine for emergency use in US



US Covid vaccine: Three key questions answered

The US Food and Drug Administration has authorised the Pfizer-BioNTech coronavirus vaccine for emergency use.

Which Vaccine Is the Most Effective?

Administering the vaccine:

- Soon vaccines by Pfizer, Moderna, and Oxford University/AstraZeneca are available with prior information about efficacy from clinical trials.
- Each country/state can learn from others who administer the vaccines as we gradually learn more about the true efficacy.
- Will the most effective vaccine be found in equilibrium?

Experimentation vs. Social Learning

Social learning:

- A sequence of short-lived individuals learns about an underlying state by observing the actions of earlier individuals.
- Individuals do not care about the information generated by their action because they each only act once.
- Main question: does learning occur despite heterogeneous preferences?

Strategic experimentation:

- Finitely many players with homogeneous preferences learn about an underlying state by repeatedly carrying out experiments.
- Players observe each others experiments and they care about the information generated because it impacts their continuation value.
- Main question: what are the strategic interactions between the players? Does the equilibrium feature efficient levels of experimentation?

Multi-Armed Bandits

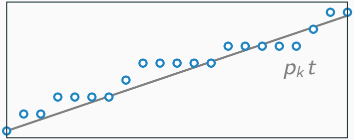
Terminology:

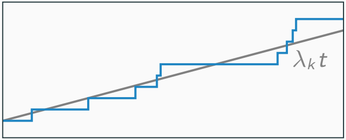
- Experimenters choose to pull one of several one-armed bandits.
- The reward of each arm depends on an unknown state of nature θ.
- Repeatedly pulling the same arm reveals information about its reward.

Example:

- The economic theorist can choose to pull on the "proof arm" or the "counterexample arm," only one of which has a positive payoff.
- Governments choose which vaccine arm to pull.

Types of Information Arrival





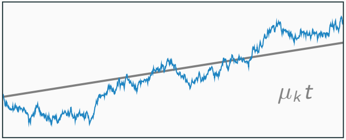
Discrete experimentation:

- Players choose which arm k of the bandit to pull.
- Probability of success p_k depends on the underlying state θ.

Continuous experimentation:

- It is mathematically more tractable to work in continuous time.
- Players decide the share of time to invest into investigating each arm k.
- Frequency λ_k of experimentation successes depends on θ.
- Formally, the arrival times of successes follow a Poisson process.

Estimating the Growth Rate



Which type of corn to cultivate?

- Farmers observe each others choice of crop and yield.
- Strategic experimentation in these contexts are better modeled using continuously distributed signals, whose mean is unknown.
- Formally, the signal is a Brownian motion with unknown drift rate.

Today's goal:

- Analyze the more intuitive Poisson model while introducing the notation of stochastic processes, which the Brownian case requires.

Poisson Processes

Stochastic Process

Definition 14.1

Fix a probability space $(\Omega, \mathcal{F}, P)$ and a measure space $(\mathcal{X}, \Xi)$. An $\mathcal{X}$-valued stochastic process X is a product-measurable map $X : \Omega \times [0, \infty) \to \mathcal{X}$.

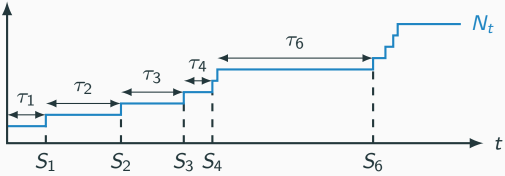
Function of two arguments:

- A stochastic process is a family of random variables $X = (X_t)_{t \geq 0}$.
- For fixed $\omega \in \Omega$, the function $X(\omega) : [0, \infty) \to \mathcal{X}$ is the path of X.
- We could also understand X as a function-valued random variable.

Product measurability:

- Product measurability guarantees that we can quantify $P(X_t \in B)$ and the duration of time that X spent in B for any $B \in \Xi$.

Counting Process



Counting process:

- A counting process $N = (N_t)_{t \geq 0}$ is an $\mathbb{N}$-valued and non-decreasing stochastic process with $N_0 = 0$ a.s.
- N_t is the number of events that have occurred on or before time t.

Arrival times:

- The k^{th} jump time of N is called the k^{th} arrival time S_k with $S_0 = 0$.
- The time between arrival k and $k-1$ is the k^{th} interarrival time τ_k.

Poisson Process

Definition 14.2

A time-homogeneous Poisson process with intensity λ is a counting process, whose interarrival times are $Exp(\lambda)$-distributed.

Definition 14.3

A time-homogeneous Poisson process with intensity λ is a stochastic process $N = (N_t)_{t \geq 0}$ such that $N_0 = 0$ a.s. and for any $s \leq t$, the increment $N_t - N_s$ is independent of $\sigma(N_s)$ and $Poisson(\lambda(t-s))$-distributed.

- The two definitions are both very useful and they are equivalent.
- Seeing the equivalence is quite instructive for understanding why Poisson processes are so frequently used.

Memorylessness

Lemma 14.4

The exponential distribution is memoryless, i.e., for any Exp(λ)-distributed random variable τ, we have $P(\tau \leq t + s \,|\, \tau \geq s) = P(\tau \leq t)$.

Relevance to Poisson processes:

- Knowing that an event has not arrived yet does not reveal any information about when the next event will arrive.
- This is the crucial ingredient for independence of increments.

Proof: From Bayes' rule we obtain

$$\frac{P(s \leq \tau \leq t + s)}{P(\tau \geq s)} = \frac{1 - \mathrm{e}^{\lambda(s+t)} - (1 - \mathrm{e}^{\lambda s})}{e^{\lambda s}} = 1 - \mathrm{e}^{\lambda t}.$$

Arrival Times

Lemma 14.5

The sum S_n of n independent $Exp(\lambda)$ distributed random variables follows a $\Gamma(n, \lambda)$-distribution with distribution function

$$P(S_n \leq s) = \int_0^s \frac{\lambda \mathrm{e}^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!} \,\mathrm{d}x.$$

Proof:

- The moment-generating function of the $Exp(\lambda)$-distribution is

$$\mathbb{E}[\mathrm{e}^{u\tau}] = \int_0^\infty \mathrm{e}^{ut}\mathrm{e}^{-\lambda t}\,\mathrm{d}t = \int_0^\infty \mathrm{e}^{(u-\lambda)t}\,\mathrm{d}t = \frac{\lambda}{\lambda - u}$$

- The moment-generating function of the $\Gamma(n, \lambda)$-distribution is

$$\mathbb{E}\left[e^{uS_n}\right] = \int_0^\infty \frac{\lambda \mathrm{e}^{(u-\lambda)x}(\lambda x)^{n-1}}{(n-1)!}\,\mathrm{d}x = \left(\frac{\lambda}{\lambda - u}\right)^n \frac{1}{(n-1)!}\int_0^\infty e^{-t}t^{n-1}\,\mathrm{d}t.$$

Definition 14.2 implies Definition 14.3

Poisson distribution:

- Let N be a counting process with $Exp(\lambda)$ interarrival times.
- Using integration by parts, we obtain

$$P(N_t = n) = P(S_n \leq t) - P(S_{n+1} \leq t)$$
$$= \int_0^t \frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}\, dx - \int_0^t \frac{\lambda e^{-\lambda x}(\lambda x)^n}{n!}\, dx = \frac{e^{-\lambda t}(\lambda t)^n}{n!}.$$

Independent increments:

- For any $s > 0$, the process $\widetilde{N}$ defined by $\widetilde{N}_t := (N_{s+t} - N_s)$ is a counting process with $Exp(\lambda)$-distributed interarrival times $\widetilde{\tau}_k$ for $k \geq 2$.
- By the memorylessness property, $\widetilde{\tau}_1$ is also $Exp(\lambda)$-distributed.
- Therefore, $(\widetilde{N}_t)_{t \geq 0}$ is a Poisson process and, in particular,

$$P(N_t - N_s = n \,|\, \sigma(N_s)) = P\left(\widetilde{N}_{t-s} = n\right) = \frac{e^{-\lambda t}(\lambda(t-s))^n}{n!}.$$

Definition 14.3 implies Definition 14.2

Non-decreasing counting process:

- Let N have independent Poisson increments with $N_0 = 0$.
- In particular, N_t is *Poisson*(λt)-distributed and, hence $\mathbb{N}$-valued.
- For any $s \leq t$, independence of increments implies monotonicity via

$$P(N_t - N_s \geq 0) = P(N_{t-s} \geq 0) = 1.$$

Exponential interarrival times:

- Independence of increments imply

$$P(\tau_k > t) = P\big(N_{S_{k-1}+t} - N_{S_{k-1}} = 0 \,\big|\, \sigma(N_{S_{k-1}})\big) = P(N_t = 0) = \mathrm{e}^{-\lambda t}.$$

Why are Poisson Processes so Frequently Used?

Discrete time:

- The public signal is a sequence $(Y_t)_{t \geq 0}$ of random variables such that the distribution of Y_t depends on A_t and θ.
- Moreover, Y_t is conditionally independent of Y_s for $s < t$.

Analogue in continuous time:

- The distribution of $Y_t - Y_s$ depends on $A_x, x \in (s, t]$ and θ
- The increments $Y_t - Y_s$ are conditionally independent of Y_x for $x < s$.
- In a repeated or stationary environment, increments have to be independent and identically distributed, given the same chosen actions.

Memoryless property:

- The exponential distribution is the only such continuous distribution.
- Poisson processes are the only counting processes with independent and stationary increments finitely many events on finite time intervals.

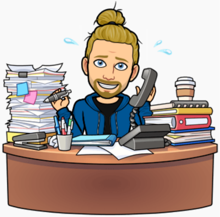
Developing a Vaccine

Developing a vaccine:

- Different approaches to finding a vaccine could be fruitful:
    - Use inactivated coronavirus cells to trigger an immune response.
    - Develop Adenoviruses that help fight the coronavirus.
    - Inject mRNA with genetic instructions to build antibodies.
- The number of vaccines developed using these approaches are Poisson processes, whose intensities depend on a variety of factors.

Source: https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html

Working on a Proof



Proving a conjecture:

- The number of proofs found is a Poisson process, whose intensity depends on the correctness and the difficulty of the conjecture, as well as how much time you spend working on it.

Frequency of events:

- Which vaccine is developed or which proof is found depends on how much time the experimenter invests.
- We need to allow for time-varying intensities of events.

Time-Inhomogeneous Poisson Process

Definition 14.6

A time-inhomogeneous Poisson process with instantaneous intensity $\lambda = (\lambda_t)_{t\geq 0}$ is a process N such that $N_0 = 0$ and for any $s \leq t$, the increment $N_t - N_s$ is independent of $\sigma(N_s)$ and $\mathit{Poisson}\big(\int_s^t \lambda_x \,\mathrm{d}x\big)$-distributed.

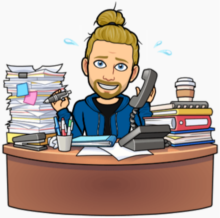
Remark:

- Events occur more frequently on intervals where λ_t is high.
- The distribution of interarrival time τ_k satisfies

$$P(\tau_k \leq t \,|\, \tau_k \geq s) = 1 - \exp\big(- \int_s^t \lambda_x \,\mathrm{d}x\big).$$

- Conditional on $\tau_k \geq s$, the density at time $t \geq s$ is $\lambda_t \exp\big(- \int_s^t \lambda_x \,\mathrm{d}x\big)$.
- In the limit as $s \nearrow t$, we obtain the instantaneous intensity λ_t.

Random Sizes of Rewards



Developing a vaccine:

- A laboratory may trade off the "investment arm" with a constant stream of revenue and the "research arm" with an unknown revenue.
- A rational laboratory will pull the arm with the higher expected profit.

Working on a proof:

- The quality of the publication is random even if you finish the proof.
- A strategic researcher will work on the project with the highest expected gains per invested time.

Compound Poisson Process

Definition 14.7

A compound Poisson process with jump size distribution F and instantaneous intensity λ is the process $Y = (Y_t)_{t \geq 0}$, defined by

$$Y_t = \sum_{k=1}^{N_t} X_k,$$

where N is a Poisson process with instantaneous intensity λ and $(X_k)_{k \geq 0}$ is a sequence of i.i.d. random variables $X_k \sim F$ independent of N.

Two-dimensional uncertainty:

- Compound Poisson processes allow us to model uncertainty about the likelihood of a success as well as the ensuing monetary reward.
- Today we will restrict attention to one-dimensional uncertainty.

Many Labs Are Working on the Vaccine

Information spillover:

- Suppose labs $i = 1, \ldots, n$ are working on mRNA-based vaccines.
- The breakthroughs of any of these labs is a Poisson process with instantaneous intensity $\lambda_t^i = A_t^i \lambda_\theta$.
- The common component λ_θ determines how difficult it is to find an mRNA-based vaccine for any laboratory.
- What is the aggregate information revealed about λ_θ?

Superposition of Compound Poisson Processes

Lemma 14.8

Let Y^1 and Y^2 be independent compound Poisson processes with identical distribution of jumps F and instantaneous intensities $\lambda^1 = (\lambda^1_t)_{t\geq 0}$ and $\lambda^2 = (\lambda^2_t)_{t\geq 0}$. Then $Y^1 + Y^2$ is a compound Poisson processes with instantaneous intensity $\lambda^1 + \lambda^2$ and jumps distributed according to F.

Proof:

- The first arrival time of an event is

$$\begin{aligned} P\big(\min\big(\tau^1_1, \tau^2_1\big) \leq t\big) &= 1 - P\big(\tau^1_1 > t\big)P\big(\tau^2_1 > t\big) \\ &= 1 - \mathrm{e}^{-\int_0^t \lambda^1_s \, \mathrm{d}s}\mathrm{e}^{-\int_0^t \lambda^2_s \, \mathrm{d}s} = 1 - \mathrm{e}^{-\int_0^t \lambda^1_s + \lambda^2_s \, \mathrm{d}s}. \end{aligned}$$

- The other interarrival times follow from independence of increments.

Two-Armed Bandit Problem

Risky and a safe arm

The two arms:

- Players $i = 1, \dots, n$ choose among a risky arm and a safe arm.
- The risky arm can be good $\vartheta = 1$ or bad $\vartheta = 0$ with prior probability μ_0 indicating that the arm is good.

Division of resources:

- At each moment of time $t \in [0, \infty)$, player i chooses which fraction $A_t^i \in [0, 1]$ of his/her resources to invest into the risky arm.
- Player i's safe arm yields a deterministic flow payoff $(1 - A_t^i)g\,\mathrm{d}t$.
- Player i's risky arm yields lump-sum rewards at intensity $A_t^i \lambda_\vartheta\,\mathrm{d}t$, whose sizes are drawn independently from a distribution F on $\mathbb{R} \setminus \{0\}$.
- We assume that $\lambda_0 f < g < \lambda_1 f$, where f is the expectation of F.

Recursive Definition?

Available information:

- The players observe the division of resources A^i by each player i and the payouts $Y^i = (Y^i_t)_{t\geq 0}$ from his/her risky arm.

Sequential definition:

- Agent observes realizations $a_0, y_0, \ldots, a_{t-1}, y_{t-1}$ to make decision A_t.
- We could define A_t and Y_t in sequence $A_0, Y_0, A_1, Y_1, A_2, \ldots$
- Will this cause a recursive definition in continuous time?

Better approach:

- We define $Y = (Y_t)_{t\geq 0}$ as a fixed sequence of random variables.
- The choice of strategy then simply changes Y's distribution.

Information and Strategies

Non-recursive definition:

- Fix a probability space $(\Omega, \mathcal{F}, P)$ containing θ and a compound Poisson process $Y = (Y_t)_{t \geq 0}$ with intensity 1 and jumps drawn from F.
- Let $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ denote the filtration generated by Y.
- A strategy $A^i = (A^i_t)_{t \geq 0}$ is a predictable process $A^i : \Omega \times [0, \infty) \to \mathcal{A}^i$.

Analyzing this notion of strategy:

- The strategy specifies for every outcome ω, which action should be taken at any time t.
- Predictability means that A^i_t is measurable with respect to $\mathcal{F}_{t-}$:
  - For any $a \in [0, 1]$, the set $\{A^i_t = a\}$ lies in $\mathcal{F}_{t-}$.
  - Recalling that $\mathcal{F}_{t-}$ is the collection of information sets just before time t, this means that A^i_t maps information sets in $\mathcal{F}_{t-}$ to $\mathcal{A}^i$.

Induced Probability Measure

Induced probability measure:

- Let $N = (N_t)_{t \geq 0}$ denote the Poisson process that drives Y.
- Strategy profile A induces a probability measure P_A, under which N (and hence Y) has instantaneous intensity $\sum_i A_t^i \lambda_\theta \, \mathrm{d}t$.
- Formally, the density process of P_A with respect to P is given by

$$\frac{\mathrm{d}P_{A,t}}{\mathrm{d}P} = \exp\left(\int_0^t 1 - \bar{A}_s \lambda_\theta \, \mathrm{d}s\right) \prod_{0<s\leq t} \big(1 + (\bar{A}_{s-}\lambda_\theta - 1)\Delta N_s\big),$$

where we denote $\bar{A}_s = \sum_i A_s^i$ and $\Delta N_s = N_s - N_{s-}$.

Decomposing the density process:

- The continuous change of the density process is $\exp\big(\int_0^t 1 - \bar{A}_s \lambda_\theta \, \mathrm{d}s\big)$.
- A jump in N multiplies the density process by $\bar{A}_{t-}\lambda_\theta$.

Beliefs

Belief process:

- The agent's beliefs about the state of nature are $\mu_t = P_A(\theta = 1 \,|\, \mathcal{F}_t)$.
- The expected intensity of successes of the risky arm at time t is

$$\begin{aligned}\mathbb{E}_A[\lambda_\theta \,|\, \mathcal{F}_t] &= \lambda_0 P_A(\theta = 0 \,|\, \mathcal{F}_t) + \lambda_1 P_A(\theta = 1 \,|\, \mathcal{F}_t) \\ &= (1 - \mu_t)\lambda_0 + \mu_t \lambda_1 =: \lambda(\mu_t).\end{aligned}$$

Beliefs after a success:

- If a success occurs on the risky arm at time t, the beliefs change to

$$\mu_t = \frac{\lambda_1 \mu_{t-}}{\lambda_1 \mu_{t-} + \lambda_0 (1 - \mu_{t-})} = \frac{\lambda_1 \mu_{t-}}{\lambda(\mu_{t-})}.$$

- It will be convenient to denote abbreviate $j(\mu) := \frac{\lambda_1 \mu}{\lambda(\mu)}$.

Continuous Changes in Beliefs

Impact of actions on beliefs:

- Conditional on no event arriving at time t, the density process is

$$Z_{A,t} := \exp\Big(\int_0^t 1 - \bar{A}_s \lambda_\theta \, \mathrm{d}s\Big).$$

- Therefore, the belief process must satisfy

$$\mu_t = \frac{\mathbb{E}\big[Z_{A,t} 1_{\{\theta=1\}} \,\big|\, \mathcal{F}_t\big]}{\mathbb{E}[Z_{A,t} \,|\, \mathcal{F}_t]} = \frac{\mathrm{e}^{-\int_0^t \bar{A}_s \lambda_1 \, \mathrm{d}s}}{\mathrm{e}^{-\int_0^t \bar{A}_s \lambda(\mu_s) \, \mathrm{d}s}} = \mathrm{e}^{\int_0^t \bar{A}_s (\lambda(\mu_s) - \lambda_1) \, \mathrm{d}s}.$$

Solving for the belief process:

- It follows from the chain rule that

$$\mathrm{d}\mu_t = \mu_t \bar{A}_t (\lambda(\mu_t) - \lambda_1) \, \mathrm{d}t = -\bar{A}_t (\lambda_1 - \lambda_0) \mu_t (1 - \mu_t) \, \mathrm{d}t.$$

- Using separation of variables, we obtain

$$\mu_t = \tfrac{X_t}{1+X_t}, \qquad X_t = \tfrac{\mu_0}{1-\mu_0} \exp\big(-\int_0^t \bar{A}_s (\lambda_1 - \lambda_0) \, \mathrm{d}s\big).$$

Evolution of Beliefs

Lemma 14.9

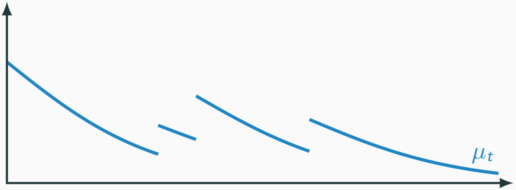
The belief process satisfies the following stochastic differential equation

$$\mathrm{d}\mu_t = -\bar{A}_t(\lambda_1 - \lambda_0)\mu_t(1-\mu_t)\,\mathrm{d}t + \left(\frac{\lambda_1 \mu_{t-}}{\lambda(\mu_{t-})} - \mu_{t-}\right)\mathrm{d}N_t.$$

Interpretation:

- While no success arrives, the beliefs decrease continuously.
- Beliefs decrease faster if:
  - More total resources $\bar{A}_t$ are invested into the risky arm,
  - The difference in intensities $\lambda_1 - \lambda_0$ is larger,
  - μ_t is closer to $\frac{1}{2}$.
- Upon a success, the beliefs jump upwards to $j(\mu_{t-})$.

Visualization of Belief Process



Mathematical tractability:

- Because of independent increments, the beliefs decrease stationarily.
- Despite the conceptual difficulty of the model, continuous time allows us to find a closed-form expression for the beliefs.

Relation to social learning:

- Knowing how public beliefs evolve in this model, can we recover the behavior of short-lived players as in the social learning model?

Myopic Maximizers

Lemma 14.10

A short-lived player chooses the risky arm only if $\mu_t \geq \mu^m := \frac{s-\lambda_0 f}{f(\lambda_1-\lambda_0)}$.

Proof:

- A short-lived player maximizes the expected flow-payoff, given μ_t.
- Solving $f\lambda(\mu_t) > s$ for μ^m yields the above myopic cutoff μ^m.

Consequence:

- If $\mu_0 > \mu^m$, then a short-lived player will choose the risky arm until beliefs reach μ^m and then stops forever.
- The cascade set $[0, \mu^m]$ is reached with positive probability.

Long-Lived Players

Discounted payoffs:

- Suppose the agent discounts future payoffs at discount rate $r > 0$.
- The discounted future payoff at time t is

$$U_t^i(A) := \int_t^\infty r\mathrm{e}^{-r(s-t)}(1-A_s^i)g\,\mathrm{d}s + \int_t^\infty r\mathrm{e}^{-r(s-t)}\,\mathrm{d}Y_s^i.$$

- The discounted expected future payoff (or continuation value) is

$$\begin{aligned} W_t^i(A) &= \int_t^\infty r\mathrm{e}^{-r(s-t)}\mathbb{E}_A\left[(1-A_s^i)g + A_s^i\lambda_\theta f \,\middle|\, \mathcal{F}_t\right]\mathrm{d}s \\ &= \int_t^\infty r\mathrm{e}^{-r(s-t)}\mathbb{E}_A\left[(1-A_s^i)g + A_s^i\lambda(\mu_s) f \,\middle|\, \mathcal{F}_t\right]\mathrm{d}s \end{aligned}$$

where we have used Fubini's theorem, that $\mathrm{d}Y_s^i - A_s^i\lambda_\theta f\,\mathrm{d}s$ is a martingale increment, and the tower property of the conditional expectation.

Markov Strategies

Markov strategies:

- The continuation value of player i depends on A^{-i} only through μ.
- As a first pass at the problem, we look for strategies $A^i : [0,1] \to \mathcal{A}$ that are Markovian in the public beliefs μ_{t-}.
- The corresponding equilibrium notion is a Markov perfect equilibrium.

Approach:

- The efficient level of experimentation is attained as the solution to the cooperative problem, where players maximize the sum of payoffs.
- Find the symmetric Markov perfect equilibrium and see how it differs.

Cooperative Experimentation

Efficient Level of Experimentation

Theorem 14.11 (Keller and Rady, 2010)

Let K_n be the positive solution K to

$$\tfrac{r}{N} + \lambda_0 - K(\lambda_1 - \lambda_0) = \lambda_0 \left(\tfrac{\lambda_0}{\lambda_1}\right)^K.$$

In the n-agent cooperative problem, there exist cut-off belief

$$\mu_n^* = \frac{K_n(s - \lambda_0 f)}{(K_n + 1)(\lambda_1 f - s) + K_n(s - \lambda_0 f)} < \mu^m$$

such that R is uniquely optimal for $\mu_{t-} > \mu_n^*$ and S is uniquely optimal for $\mu_{t-} < \mu_n^*$. Moreover, the continuation value in the optimum is

$$V_n^*(\mu) = \lambda(\mu) f + (s - \lambda(\mu) f) \frac{1 - \mu}{1 - \mu_n^*} \left(\frac{(1 - \mu)\mu_n^*}{(1 - \mu_n^*)\mu} \right)^{K_n}$$

if $\mu > \mu_n^*$ and $V_n^*(\mu) = s$ otherwise.

Principle of Optimality

Average payoffs:

- Let $u(A_t, \mu_t) = (1 - \frac{1}{n}\bar{A}_t)g + \frac{1}{n}\bar{A}_t\lambda(\mu_t)$ denote the average flow payoff.
- Let A^* denote the optimal Markovian strategy profile.
- Its continuation value, given that the public beliefs are μ_t, are

$$V(\mu_t) := \frac{1}{n}\sum_{i=1}^{n} W_t^i(A^*) = \int_t^{\infty} r\mathrm{e}^{-r(s-t)}\mathbb{E}_{A^*}[u(A_s^*, \mu_s) \,|\, \mathcal{F}_t]\,\mathrm{d}s.$$

Principle of optimality:

- Because A^* is optimal, for any $h > 0$ we must have

$$V(\mu_t) = \int_t^{t+h} r\mathrm{e}^{-r(s-t)}\mathbb{E}_{A^*}[u(A_s^*, \mu_s) \,|\, \mathcal{F}_t]\,\mathrm{d}s + \mathrm{e}^{-rh}\mathbb{E}_{A^*}[V_n^*(\mu_{t+h}) \,|\, \mathcal{F}_t].$$

- We will solve this equation for $V_n^*(\mu)$.

Change in the Continuation Value

Change in continuation value:

- It follows from Lemma 14.9 that

$$
\begin{aligned}
V(\mu_{t+h}) - V(\mu_t) &= \int_t^{t+h} V'(\mu_s)\,\mathrm{d}\mu_s^c + \sum_{t<s\le t+h} \Delta V(\mu_s) \\
&= -\int_t^{t+h} V'(\mu_s)\bar{A}_s^*(\lambda_1 - \lambda_0)\mu_s(1-\mu_s)\,\mathrm{d}s \\
&\quad + \int_t^{t+h} \big(V(j(\mu_{s-})) - V(\mu_{s-})\big)\,\mathrm{d}N_s.
\end{aligned}
$$

- The $\mathcal{F}_t$-conditional expectation under P_{A^*} of the last term is

$$
\int_t^{t+h} \big(V(j(\mu_s)) - V(\mu_s)\big)\bar{A}_s^*\lambda(\mu_s)\,\mathrm{d}s.
$$

Taking the Limit

Change in continuation value:

- So far we have shown

$$
\begin{aligned}
(1-\mathrm{e}^{-rh})V(\mu_t) &= \int_t^{t+h} r\mathrm{e}^{-r(s-t)}\mathbb{E}_{A^*}[u(A_s^*,\mu_s)\,|\,\mathcal{F}_t]\,\mathrm{d}s \\
&\quad - \mathrm{e}^{-rh}\mathbb{E}_{A^*}\left[\int_t^{\widetilde{t}} V'(\mu_s)\bar{A}_s^*(\lambda_1-\lambda_0)\mu_s(1-\mu_s)\,\mathrm{d}s\,\middle|\,\mathcal{F}_t\right] \\
&\quad + \mathrm{e}^{-rh}\mathbb{E}_{A^*}\left[\int_t^{t+h}\big(V(j(\mu_s))-V(\mu_s)\big)\bar{A}_s^*\lambda(\mu_s)\,\mathrm{d}s\,\middle|\,\mathcal{F}_t\right]
\end{aligned}
$$

- Dividing by h and taking the limit $h\to 0$ yields

$$
\begin{aligned}
rV(\mu_t) &= ru(A_t^*,\mu_t)+\big(V(j(\mu_t))-V(\mu_t)\big)\bar{A}_t^*\lambda(\mu_t) \\
&\quad - V'(\mu_t)\bar{A}_t^*(\lambda_1-\lambda_0)\mu_t(1-\mu_t).
\end{aligned}
$$

Bellman Equation

Bellman equation:

- Finally, dividing by r gives us the Bellman equation

$$V(\mu) = s + \max_{A \in [0,n]} A\Big(b(\mu, V) - \frac{1}{n}c(\mu)\Big), \tag{1}$$

where the opportunity cost of experimentation is $c(\mu) := (s - \lambda(\mu)f)$ and the expected benefit of experimentation is

$$b(\mu, V) := \frac{1}{r}\lambda(\mu)\big(V(j(\mu)) - V(\mu)\big) - \frac{1}{r}V'(\mu)(\lambda_1 - \lambda_0)\mu(1-\mu).$$

Optimal strategy:

- It is uniquely optimal to play $A = n$ if $b(\mu, V) > \frac{1}{n}c(\mu)$ and it is uniquely optimal to play $A = 0$ if $b(\mu, V) < \frac{1}{n}c(\mu)$.
- If $A = 0$ is optimal, then $V(\mu) = s$.

Solving the Bellman Equation

Bellman equation:

- If $A = N$ is optimal, then (1) becomes

$$V'(\mu)(\lambda_1 - \lambda_0)\mu(1-\mu) - \lambda(\mu)\big(V(j(\mu)) - V(\mu)\big) + \frac{r}{N}V(\mu) = \frac{r}{N}\lambda(\mu)f. \quad (2)$$

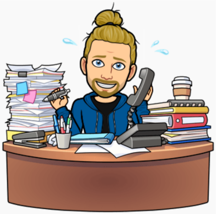
Verifying our candidate solution:

- Let us abbreviate $h(\mu) = (1-\mu)\left(\frac{1-\mu}{\mu}\right)^{K_n}$ so that our candidate is

$$V_n^*(\mu) = \lambda(\mu)f + (s - \lambda(\mu)f)\frac{h(\mu)}{h(\mu_n^*)}.$$

- To show that $V_n^*(\mu)$ is a solution to (2), we show that:
    - $V_p(\mu) = \lambda(\mu)f$ is a particular solution to (2),
    - $V_h(\mu) = (s - \lambda(\mu)f)h(\mu)$ is a solution to the homogeneous equation,
    - $V_n^*(\mu_n^*) = s$ (value matching) and $(V_n^*)'(\mu_n^*)$ (smooth pasting).

Choosing a Research Project



Project choice:

- You can choose between an easy project, yielding 1 academic prestige unit per semester, and a difficult project that consists of proving some conjecture. Proving the conjecture yields a random reward X with

$$P(X = 10) = 0.1, \qquad P(X = 5) = 0.5, \qquad P(X = 2) = 0.4.$$

- The prior is $\mu_0 = 0.8$, the conjecture takes an expected 1.5 years to prove, and each year of delay reduces the benefit by 20%.
- How long should you try to prove the conjecture?

Non-Cooperative Experimentation

Individual Best Responses

Bellman equation:

- Suppose that players $-i$ play according to $A^{-i}(\mu)$.
- With similar arguments as before, the best response satisfies

$$W_i(\mu) = s + \bar{A}^{-i}(\mu) b(\mu, W_i) + \max_{A^i \in [0,1]} A^i \big(b(\mu, W_i) - c(\mu)\big). \quad (3)$$

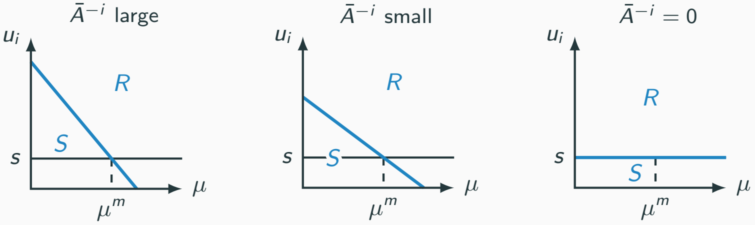
- Note that the cost of experimentation is no longer shared.

Best response:

- The best response correspondence can be written as

$$\mathcal{B}_i(A^{-i}, \mu) = \begin{cases} A^i = 0 & \text{if } W_i(\mu) < s + \bar{A}^{-i}(\mu)c(\mu), \\ A^i \in [0,1] & \text{if } W_i(\mu) = s + \bar{A}^{-i}(\mu)c(\mu), \\ A^i = 1 & \text{if } W_i(\mu) > s + \bar{A}^{-i}(\mu)c(\mu). \end{cases}$$

Best Responses Visualized



Best response:

- In the (μ, W_i)-plane, the indifference points

$$\mathcal{D}(A^{-i}) := \left\{(\mu, W_i) \in [0,1] \times \mathbb{R}^+ \;\middle|\; W_i = s + \bar{A}^{-i} c(\mu)\right\}$$

  form a straight line with slope $\bar{A}^{-i}(\lambda_1 - \lambda_0)f$ going through (μ^m, s).
- Note that the larger A^{-i} is, the larger is the region, where it is optimal to play S. This is the free-rider effect because information is public.

Perfectly Revealing Experiments

Perfectly revealing experiments:

- Suppose that $\lambda_0 = 0$ so that $j(\mu) = 1$ for any μ.
- Then $b(\mu, W_i) = \frac{\lambda_1 \mu}{r}\big(f - W_i(\mu) - (1-\mu)W_i'(\mu)\big)$.

Explicit best response above the diagonal:

- If $W_i(\mu)$ is above the diagonal $\mathcal{D}(A^{-i})$, then $A^i = 1$.
- This implies $W_i(\mu) = \bar{A}b(\mu, W_i) + s - c(\mu)$, which we reorganize to

$$\bar{A}\lambda_1\mu(1-\mu)W_i'(\mu) + (r + \bar{A}\lambda_1\mu)W_i(\mu) = (r + \bar{A}\lambda_1)f\mu.$$

- This ODE has the solution $W_i^+(\mu) := f\mu + C(1-\mu)\left(\frac{1-\mu}{\mu}\right)^{\frac{r}{\lambda\bar{A}}}$.

Explicit Best Responses

Below the diagonal:

- If $W_i(\mu)$ isbelow the diagonal $\mathcal{D}(A^{-i})$, then $A^i = 0$.
- This implies $W_i(\mu) = \bar{A}b(\mu, W_i) + s$, which we reorganize to

$$\bar{A}\lambda\mu(1-\mu)W_i'(\mu) + (r + \bar{A}\lambda_1\mu)W_i(\mu) = rs + \bar{A}\lambda_1 f\mu.$$

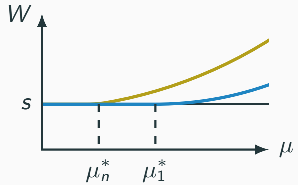
- This ODE has the solution $W_i^-(\mu) := s + \frac{\bar{A}\lambda_1\mu(f-s)}{r+\lambda_1\bar{A}} + C(1-\mu)\left(\frac{1-\mu}{\mu}\right)^{\frac{r}{\lambda\bar{A}}}$.

On the diagonal:

- On the diagonal, player i is indifferent.
- Solving $c(\mu) = b(\mu, W_i)$ for W_i, we obtain

$$W_i^0(\mu) = s + \frac{r+\lambda_1}{\lambda_1}(f-s) + \frac{rs}{\lambda_1}(1-\mu)\ln\left(\frac{1-\mu}{\mu}\right) + C(1-\mu). \quad (4)$$

Bounds for Equilibrium Payoffs



Bounds for equilibrium payoffs:

- Let $W(\mu)$ denote the equilibrium continuation value.
- Since V_n^* is the cooperative solution, we must have $W(\mu) \leq V_n^*(\mu)$.
- With the single-agent strategy, any player can ensure $W(\mu) \geq V_1^*(\mu)$.

Stopping experimentation:

- Let μ_c denote the largest value $\mu \in [\mu_n^*, \mu_1^*]$, where $W(\mu) = s$.
- This point must satisfy $W'(\mu_c) \geq 0$.

Mixing

Mixing below μ_1^*:

- Suppose players mix below μ_1^*. Then $W(\mu) = W_i^0(\mu)$.
- Using that $W(\mu_c) = s$ and $W'(\mu_c) \geq 0$ has to solve $c(\mu) = b(\mu, W)$, we deduce that $\mu_c \geq \mu_1^*$ and, hence $\mu_c = \mu_1^*$.

Fixing integration constant:

- Knowing tat $W(\mu) = W_i^0(\mu)$ with $W(\mu_1^*) = s$, we can determine the integration constant C in (4).

Mixing above μ_1^*:

- Using $A^i(\mu) = \frac{1}{n-1} A^{-i}(\mu)$, we deduce that $\lim_{\mu \to \mu^m} A^i(\mu) = \infty$.
- Thus, there exists $\mu_n^\dagger \in (\mu_1^*, \mu^m)$ such that players choose S above $\mu_n^\dagger$.

Perfectly Revealing Experiments

Theorem 14.12 (Keller, Rady, and Cripps, 2005)

Suppose that $\lambda_0 = 0$ and $n \geq 2$. Let $\mu_n^\dagger$ be the solution to

$$\frac{\mu^m}{1-\mu^m} - \frac{\mu_n^\dagger}{1-\mu_n^\dagger} = \frac{r+\lambda_1}{\lambda_1(n-1)}\left(\frac{1}{1-\mu_n^\dagger} - \frac{1}{1-\mu_1^*} - \frac{\mu_1^*}{1-\mu_1^*}\ln\left(\frac{(1-\mu_1^*)\mu_n^\dagger}{(1-\mu_n^\dagger)\mu_1^*}\right)\right).$$

Then $\mu_n^\dagger > \mu_1^*$ and the unique symmetric stationary Markov equilibrium σ and its continuation value $W(\mu)$ are given by:

1. $\sigma^i(\mu) = 1$ *and* $W(\mu) = V_n^*(\mu)$ *if* $\mu_{t-} \geq \mu_n^\dagger$*,*
2. $\sigma^i(\mu) = \frac{1}{n-1}\frac{W(\mu)-s}{c(\mu)}$ *if* $\mu_{t-} \in (\mu_1^*, \mu_n^\dagger)$*, where*

$$W(\mu) = s + \frac{rs}{\lambda_1}\left(\frac{\mu-\mu_1^*}{\mu_1^*} - (1-\mu)\ln\left(\frac{(1-\mu_1^*)\mu}{(1-\mu)\mu_1^*}\right)\right),$$

3. $\sigma^i(\mu) = 0$ *and* $W(\mu) = s$ *if* $\mu_{t-} \leq \mu_1^*$*.*

Interpretation of Theorem 14.12

Equilibrium experimentation:

- Experimentation stops at the single-agent threshold μ_1^*.
- For high beliefs, experimentation is efficient and $W_n(\mu) = V_m^*(\mu)$.
- For intermediate beliefs, players dedicate some but not all of their resources to the risky arm and $\sigma^i(\mu)$ is increasing in μ

Interpretation:

- For beliefs in the range $[\mu_1^*, \mu_n^\dagger]$, players are partially free-riding.
- Experimentation stops at the same threshold μ_1^* as in the single-agent problem: the same total experimentation is performed.
- There is no positive externality from information spillover.

Not Perfectly Revealing Experiments

Theorem 14.13 (Keller and Rady, 2010)

Suppose that experiments are not perfectly revealing, i.e., $\lambda_0 > 0$. Then there is a unique symmetric Markov perfect equilibrium σ with continuation value $W_n(\mu)$, described by $\widetilde{\mu}_n < \mu'_n < \mu^m$ with $\mu^*_n < \widetilde{\mu}_n < \mu^*_1$ and:

1. $\sigma_i(\mu) = 0$ *and* $W_n(\mu) = s$ *if* $\mu \leq \widetilde{\mu}_n$,
2. $\sigma_i(\mu) = \frac{1}{n-1}\frac{W_n(\mu)-s}{c(\mu)}$ *if* $\mu \in (\widetilde{\mu}_n, \mu'_n)$, *where* W_n *solves*

$$W'_n(\mu)(\lambda_1 - \lambda_0)\mu(1-\mu) - \lambda(\mu)\big(W_n(j(\mu)) - W_n(\mu)\big) = r\lambda(\mu)f - rs,$$

3. $\sigma_i(\mu) = 1$ *and* $W_n(\mu) = V^*_n(\mu)$ *if* $\mu \geq \mu'_n$.

Interpretation of Theorem 14.13

Encouragement effect:

- Experimentation proceeds below the single-agent threshold μ_1^*.
- If player i experiments successfully, the posterior increases, which leads to more experimentation by the other players.
- Since future experimentation by others is valuable, player i is willing to experiment below the threshold μ_1^*.
- This is known as the encouragement effect.

Comparison to Theorem 14.12:

- If $\lambda_0 = 0$, a successful experiment unambiguously reveals that the risky arm is good.
- In particular, no future information is to be learned.

Literature

- J. Jacod and A.N. Shiryaev: Limit Theorems for Stochastic Processes, 2nd ed., Springer Berlin, 2003
- E.L. Presman: Poisson Version of the Two-Armed Bandit Problem with Discounting, Theory of Probability and its Applications, **35** (1990), 307–317
- P. Bolton and C. Harris: Strategic Experimentation, Econometrica, **67** (1999), 349–374
- G. Keller, S. Rady, and M. Cripps: Strategic Experimentation with Exponential Bandits, Econometrica, **73** (2005), 39–68
- G. Keller, S. Rady, and M. Cripps: Strategic Experimentation with Poisson Bandits, Theoretical Economics, **5** (2010), 275–311