Strategic Experimentation with Undiscounted Bandits*

Godfrey Keller† Sven Rady‡

April 2004, January 2005

Preliminary and Incomplete

Abstract

We analyse a game of strategic experimentation with two-armed bandits when there is no discounting but there is background information. The payoffs from both the safe arm and the risky arm have quite general specifications, as does the noise in the observations. With the players using stationary Markov strategies, we describe all the Markov perfect equilibria, and show that, under all these specifications, the optimal policy is strikingly similar and depends only on the current expected payoff from the risky arm and the full-information payoff.

Keywords: Strategic Experimentation, Two-Armed Bandit.

JEL Classification Numbers: C73, D83.

*Our thanks for very helpful discussions and suggestions are owed to workshop participants at Nuffield College, Oxford. The first author would like to thank CES, University of Munich, for their hospitality.

†Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK.

‡Department of Economics, University of Munich, Kaulbachstr. 45, D-80539 Munich, Germany.

Introduction

We analyze the undiscounted version of a class of continuous-time two-armed bandit models, in which a number of players act non-cooperatively. (We introduce background information to ensure that the problem is well-posed.) Undiscounted models are easier to solve because the best responses of the players do not depend on continuation values. The reasons for studying these types of model are two-fold: first, in models where both the low-discounting and no-discounting cases have been solved, the low-discounting solution converges to the no-discounting solution as the discount rate tends to zero; consequently, in cases where the discounted model is hard to solve, the modeller might benefit from the clues available from the solution to the undiscounted model. Secondly, we are able to draw a very clear connection from the basic set-up of the model to its solution, highlighting precisely which particular aspects of the set-up are driving which features of the solution.

We focus on three examples. In the first, the noise is Brownian motion with unknown drift, and the agents' prior belief about this drift is a two-point distribution. In the second, the noise comes from a Poisson process with unknown intensity, and the agents' prior belief about this intensity is again a two-point distribution. In the third, the noise again comes from a Poisson process with unknown intensity, but now the agents' prior belief about this intensity is characterized by a Gamma distribution. The equilibria in the general specification exhibit striking similarities, but some differences. As a result, we can say to what extent the way that the noise is modelled matters, and to what extent the distribution of the agents' prior belief.

The rest of the article is organized as follows. The next section sets up the general model and provides details of the three examples. Then we establish the efficient benchmark where players cooperate in order to maximize joint expected payoffs before we look at the strategic problem and provide an inefficiency result. Finally, we characterize all the Markov perfect equilibria of the non-cooperative game and offer some concluding remarks.

1 Undiscounted Bandits

Time $t \in [0, \infty)$ is continuous, and there is no discounting. There are $N \geq 1$ players, each of them endowed with one unit of a perfectly divisible resource per unit of time. Each player faces a two-armed bandit problem where she continually has to decide what

fraction of the available resource to allocate to each arm.

If a player uses the safe arm S over an interval $[t, t+dt)$, the payoff increment is $dy_t = \tilde{s}\,dt + \nu(t)$, where $\nu(t)$ is IID noise with mean 0, and $\mathrm{E}[\tilde{s}] = s$, fixed and known to all players. If a player uses the risky arm R over an interval $[t, t+dt)$, the payoff increment is $dy_t = \tilde{\mu}\,dt + \nu(t)$, where $\nu(t)$ is again IID noise with mean 0, and $\mathrm{E}[\tilde{\mu}] = \mu$, fixed but unknown. So s and μ are the expected flow equivalents of the two arms, and if a player allocates the fraction $k_t \in [0,1]$ of the resource to R over an interval of time $[t, t+dt)$, and consequently the fraction $1-k_t$ to S, then her expected payoff increment conditional on μ is $[(1-k_t)s + k_t\mu]\,dt$.

Regardless of the players' choices, $(k_{1,t}, \ldots, k_{N,t})$, all the players observe a background signal which is a perfect substitute for $k_{0,t}$ units of the resource being allocated to R; assume that $k_{0,t} = k_0 > 0$. This ensures that the players eventually learn the value of μ, even if they all play S all the time.

The players start with a common prior belief about μ, and thereafter they all observe each other's actions and outcomes, so they hold common posterior beliefs throughout time. Assume that at time t the players believe that μ has CDF $H(\cdot;\pi_t)$, where π_t is a sufficient statistic for the observations on R and the background signal up to time t, and H represents a conjugate family of distributions. We assume that $s \in \mathrm{supp}\, H(\cdot;\pi_t)$, so each player prefers R, if it is 'good', to S, and prefers S to R, if it is 'bad'. Let $m(\pi_t)$ denote the expected payoff from R, and let $f(\pi_t)$ denote the full-information payoff:

$$\begin{aligned} m(\pi) &= \mathrm{E}_\pi[\mu], \\ f(\pi) &= \mathrm{E}_\pi[\mu \vee s]. \end{aligned}$$

Assume that both $m(\cdot)$ and $f(\cdot)$ are monotonic. (Establish that $\mathrm{E}[f(\pi_t)] = f(\pi_0)$.)

Given a player's actions $\{k_t\}_{t\geq 0}$ such that k_t is measurable with respect to the information available at time t, her objective is to maximize

$$\mathrm{E}\left[\int_0^\infty \left[(1-k_t)s + k_t m(\pi_t) - f(\pi_t)\right]\,dt\right],$$

where the expectation is over the stochastic processes $\{k_t\}$ and $\{\pi_t\}$. This highlights the potential for the sufficient statistic to serve as a state variable. It also shows that a player's payoff depends on others' actions only through their impact on the evolution of the sufficient statistic.

Let $K_t = \sum_{n=1}^N k_{n,t}$. Assume that, when R is used with intensity K over an interval

of length dt, π transits to π', conditional on what is observed. Let $u_n(\pi)$ denote the value function of player n, and assume that the expected continuation value is then given by

$$\mathrm{E}\left[u_n(\pi') \mid K, \pi\right] = u_n(\pi) + (K + k_0)\, C_n(\pi)\, dt$$

for some function $C_n(\cdot)$.

Examples

For more details of Example 1, see Bolton and Harris (1999, 2000); with regard to Example 2(a), see Keller, Rady and Cripps (2005).

Example 1: Brownian noise

With $dZ_n(t) \sim \mathrm{IIN}(0, dt)$,

$$\begin{aligned} S: &\quad s\, dt + \sigma\, dZ_n(t), \\ R: &\quad \mu\, dt + \sigma\, dZ_n(t), \\ \text{b/g}: &\quad \sqrt{k_0}\, \mu\, dt + \sigma\, dZ_0(t). \end{aligned}$$

- $\mu \in \{\mu_0, \mu_1\}$ and $\mu_0 < s < \mu_1$.

  At time t, players believe that $\Pr[\mu = \mu_1] = \pi_t$, so

$$H(\mu; \pi) = \begin{cases} 0 & \text{if } \mu < \mu_0, \\ 1 - \pi & \text{if } \mu_0 \le \mu < \mu_1, \\ 1 & \text{if } \mu_1 \le \mu. \end{cases}$$

and

$$\begin{aligned} m(\pi) &= \pi\mu_1 + (1 - \pi)\mu_0, \\ f(\pi) &= \pi\mu_1 + (1 - \pi)s. \end{aligned}$$

Moreover, it follows from Liptser and Shiryayev (1977, Chapter 9) that

$$\mathrm{E}[d\pi_t] = 0 \quad \text{and} \quad \mathrm{Var}[d\pi_t] = (K_t + k_0)\left(\Delta\mu\, \sigma^{-1}\, \pi_t(1 - \pi_t)\right)^2 dt,$$

where $\Delta\mu = \mu_1 - \mu_0$.

Now, using Itô's lemma,

$$u_n(\pi + d\pi) = u_n(\pi) + u_n'(\pi)\, d\pi + \tfrac{1}{2} u_n''(\pi)\, d\pi^2$$

so

$$\mathrm{E}[u_n(\pi + d\pi)] = u_n(\pi) + (K + k_0)\left[\tfrac{1}{2} u_n''(\pi)\left(\Delta\mu\, \sigma^{-1}\, \pi(1-\pi)\right)^2\right] dt.$$

Example 2: Poisson noise

Lump-sums arrive according to Poisson processes, and these lump-sums have a mean of 1 over occurrences.

$$\begin{aligned} S: &\quad \text{Poisson process has parameter } s \\ R: &\quad \text{Poisson process has parameter } \mu \\ \text{b/g}: &\quad \text{Poisson process has parameter } k_0\mu \end{aligned}$$

If a player used a time-invariant allocation $k_t = k$, then the delay between the arrivals of successive lump-sums on a risky arm would be exponentially distributed with parameter $k\mu$; see Karlin and Taylor (1981, p.146), for instance. We write K_t^+ for $K_t + k_0$.

(a) $\mu \in \{\mu_0, \mu_1\}$ and $\mu_0 < s < \mu_1$.

At time t, players believe that $\Pr[\mu = \mu_1] = \pi_t$, so

$$H(\mu;\pi) = \begin{cases} 0 & \text{if } \mu < \mu_0, \\ 1 - \pi & \text{if } \mu_0 \le \mu < \mu_1, \\ 1 & \text{if } \mu_1 \le \mu. \end{cases}$$

and

$$\begin{aligned} m(\pi) &= \pi\mu_1 + (1-\pi)\mu_0, \\ f(\pi) &= \pi\mu_1 + (1-\pi)s. \end{aligned}$$

Now, w.p. $K_t^+ m(\pi_t)\, dt$ a lump-sum arrives on R (or b/g) and

$$\pi_t' = \pi_t + \Delta\pi_t = \mu_1\pi_t / m(\pi_t),$$

and w.p. $1 - K_t^+ m(\pi_t)\,dt$ no lump-sum arrives on R (or b/g) and

$$\pi_t' = \pi_t + d\pi_t = \pi_t - K_t^+\,\Delta\mu\,\pi_t(1-\pi_t)\,dt.$$

So

$$\begin{aligned}
\mathrm{E}[u_n(\pi')] &= K^+ m(\pi) u_n(\mu_1\pi/m(\pi))\,dt \\
&\quad + \left(1 - K^+ m(\pi)\,dt\right)\left(u_n(\pi) - K^+\,\Delta\mu\,\pi(1-\pi)u_n'(\pi)\,dt\right) \\
&= u_n(\pi) + K^+\left[m(\pi)\left(u_n(\mu_1\pi/m(\pi)) - u_n(\pi)\right) - \Delta\mu\,\pi(1-\pi)u_n'(\pi)\right]dt.
\end{aligned}$$

(b) $\mu \geq 0$ has a Gamma distribution and $s > 0$.

At time t, players believe that $\mu \sim \mathrm{Ga}(\alpha_t, \beta_t)$, so $\pi = (\alpha, \beta)$, the pdf of μ is given by

$$h(\mu;\pi) = \frac{\beta^\alpha}{\Gamma(\alpha)}\mu^{\alpha-1}e^{-\mu\beta}$$

with $\alpha > 0, \beta > 0$, and note that $\mathrm{E}_\pi[\mu] = \alpha/\beta$ and $\mathrm{Var}_\pi[\mu] = \alpha/\beta^2$. Then

$$\begin{aligned}
m(\pi) &= \alpha/\beta, \\
f(\pi) &= \int_0^\infty (\mu \vee s)\,h(\mu;\pi)\,d\mu
\end{aligned}$$

(Check monotonicity of $f(\cdot)$. With

$$\frac{\partial f}{\partial \alpha} = \int_0^\infty (\mu \vee s)\ln\left(\frac{\mu}{\alpha/\beta}\right)h(\mu)\,d\mu, \quad \frac{\partial f}{\partial \beta} = \int_0^\infty (\mu \vee s)\left(\frac{\alpha}{\beta} - \mu\right)h(\mu)\,d\mu,$$

expect $\partial f/\partial\alpha > 0$, $\partial f/\partial\beta < 0$.)

Now, w.p. $K_t^+ m(\pi_t)\,dt$ a lump-sum arrives on R (or b/g) and

$$\pi_t' = (\alpha_t + \Delta\alpha_t, \beta_t + d\beta_t) = (\alpha_t + 1, \beta_t + K_t^+\,dt),$$

and w.p. $1 - K_t^+ m(\pi_t)\,dt$ no lump-sum arrives on R (or b/g) and

$$\pi_t' = (\alpha_t, \beta_t + d\beta_t) = (\alpha_t, \beta_t + K_t^+\,dt).$$

(For the evolution of π_t, see for example DeGroot, 1970, Chapter 9.) So

$$\mathrm{E}[u_n(\pi')] = K^+ m(\pi) u_n(\alpha+1, \beta + K^+\,dt)\,dt + \left(1 - K^+ m(\pi)\,dt\right)u_n(\alpha, \beta + K^+\,dt)\,dt$$

$$
\begin{aligned}
&= K^+ m(\pi) u_n(\alpha+1,\beta)\,dt + \left(1 - K^+ m(\pi)\,dt\right)\left(u_n(\pi) + K^+\,\partial u_n(\pi)/\partial\beta\,dt\right)\\
&= u_n(\pi) + K^+\left[m(\pi)\left(u_n(\alpha+1,\beta) - u_n(\pi)\right) + \partial u_n(\pi)/\partial\beta\right]dt.
\end{aligned}
$$

2 The Cooperative Problem

Suppose that the N players work cooperatively, i.e. want to maximize the *average* expected payoff by jointly choosing the action profiles $\{(k_{1,t},\ldots,k_{N,t})\}_{t\geq 0}$. This is a dynamic programming problem with the current value of π as the state variable.

If current actions are $(k_1,\ldots,k_N)$, the average expected payoff increment is given by $\left[(1-\frac{K}{N})s + \frac{K}{N}m(\pi)\right]\,dt$ with $K = \sum_{n=1}^N k_n$. As the expected continuation value is of the form $u(\pi) + (K+k_0)\,C(\pi)\,dt$, the cooperative's problem reduces to choosing the optimal level of the overall allocation K given the current state π.

By the Principle of Optimality, the value function of the cooperative, expressed as average payoff per agent, satisfies

$$
u(\pi) = \max_{K\in[0,N]}\left\{\left[(1-\tfrac{K}{N})s + \tfrac{K}{N}m(\pi) - f(\pi)\right]\,dt + \mathrm{E}\left[u(\pi')\mid K,\pi\right]\right\}
$$

leading to the Bellman equation

$$
0 = \max_{K\in[0,N]}\left\{s - f(\pi) + \tfrac{K}{N}(m(\pi)-s) + (K+k_0)\,C(\pi)\right\}.
$$

Since the maximand in the Bellman equation is an affine function of K, it is immediate that it is always optimal to choose either $K=0$ (all agents use S exclusively), or $K=N$ (all agents use R exclusively). As $k_0>0$, the Bellman equation can be rearranged as

$$
\begin{aligned}
0 &= \max_{K\in[0,N]}\left\{\frac{s - f(\pi) + \frac{K}{N}(m(\pi)-s)}{K+k_0}\right\} + C(\pi)\\
&= \max_{K\in[0,N]}\left\{\frac{\frac{k_0}{N}(s-m(\pi)) - (f(\pi)-s)}{K+k_0}\right\} - \tfrac{1}{N}(s-m(\pi)) + C(\pi).
\end{aligned}
$$

Proposition 2.1 (Cooperative solution) *In the N-agent cooperative problem, the state space can be divided into two regions such that in one region it is optimal for all to play S exclusively and in the other it is optimal for all to play R exclusively. The boundary between these two regions is given by values π_N^* satisfying*

$$
\tfrac{k_0}{N}(s - m(\pi_N^*)) = f(\pi_N^*) - s. \tag{1}
$$

PROOF: When the numerator in the reworked Bellman equation is positive, it is optimal to minimize the denominator by choosing $K = 0$, and when the numerator is negative, it is optimal to maximize the denominator by choosing $K = N$. Since both $m(\pi)$ and $f(\pi)$ are monotonic in π, values of π_N^* that make the numerator zero form the boundary between the two regions. ■

This solution exhibits all of the familiar properties: the optimal strategy has a threshold where the agents change irrevocably from R to S; there are occasions where the agents make a mistake by changing from R to S although the risky action is actually better; the probability of mistakes decreases as the reward from the safe action decreases, and as the number of agents increases.

The above proposition determines the *efficient* strategies. More precisely, we can distinguish two aspects of efficiency here. Given an action profile $\{(k_{1,t}, \ldots, k_{N,t})\}_{t\geq 0}$ for the N players, the sum $K_t = \sum_{n=1}^{N} k_{n,t}$ measures how much of the N units of the resource is allocated to risky arms at a given time t – we will call this number the *intensity* of experimentation. On the other hand, the integral $\int_0^T K_t\, dt$ measures how much of the resource is allocated to risky arms overall up to time T – we will call this number the *amount* of experimentation that is performed.

The efficient intensity of experimentation exhibits a bang-bang feature, being N in one region of the state space, and 0 in the other. Thus, the efficient intensity is maximal at early stages, and minimal later on.

As we shall see next, Markov equilibria of the N-player *strategic* problem are never efficient – the intensity of experimentation will always be inefficient because of each player's incentive to free-ride on the efforts of the others.

3 The Strategic Problem

From now on, we assume that there are $N \geq 2$ players acting non-cooperatively. We consider stationary Markovian strategies with the common value of π as the state variable. We describe the best response correspondence, and show that all Markov equilibria are inefficient.

Best responses

Fix a value π. With $k_n \in [0,1]$ indicating player n's action at that value and $K = \sum_{n=1}^{N} k_n$, let $K_{\neg n} = K - k_n$, which summarizes the actions of the other players.

Proceeding in the same way as in the previous section, we find that the rearranged Bellman equation for player n is

$$0 = \max_{k_n \in [0,1]} \left\{ \frac{(K_{\neg n} + k_0)(s - m(\pi)) - (f(\pi) - s)}{K + k_0} \right\} - (s - m(\pi)) + C_n(\pi).$$

Player n's best response, k_n^*, is determined by looking at the numerator of the maximand:

$$k_n^* \begin{cases} = 0 & \text{if } (K_{\neg n} + k_0)(s - m(\pi)) > (f(\pi) - s), \\ \in [0,1] & \text{if } (K_{\neg n} + k_0)(s - m(\pi)) = (f(\pi) - s), \\ = 1 & \text{if } (K_{\neg n} + k_0)(s - m(\pi)) < (f(\pi) - s). \end{cases} \tag{2}$$

Inefficiency of Markov perfect equilibria

If the players use symmetric strategies in equilibrium, then, whenever a positive fraction of the resource is allocated to each arm, that fraction, $k_N^\dagger$, is calculated from the indifference condition in (2) together with $K_{\neg n} = (N-1)k_N^\dagger$, i.e.

$$k_N^\dagger(\pi) = \frac{1}{N-1} \left(\frac{f(\pi) - s}{s - m(\pi)} - k_0 \right),$$

and note that $k_N^\dagger(\pi_1^*) = 0$, i.e. all the players stop using R when a single agent would do so.

If the players use asymmetric strategies in equilibrium, then, whenever all the other players are using S exclusively, so $K_{\neg n} = 0$, player n's decision is the same as in the single-agent case, so she too would switch from R to S when the state is π_1^*. Further, as we will see in the next section, when fewer than $N-1$ players are using S exclusively, in equilibrium the remaining players use symmetric actions near to π_1^* and again stop using R at π_1^*.

Since the region of the state space where an N-agent cooperative plays R increases with N, and any equilibrium of the N-player experimentation game has players using R only where a single agent would, all these equilibria are inefficient. Further, close to the single-agent boundary the equilibrium intensity of experimentation is at most 1, whereas

the intensity of experimentation for the cooperative is N.

4 Equilibria

Sets of mutual best responses are given below and the resulting total experimentation schedule for $N = 3$ is illustrated in Figure 1, which shows the intensity of experimentation

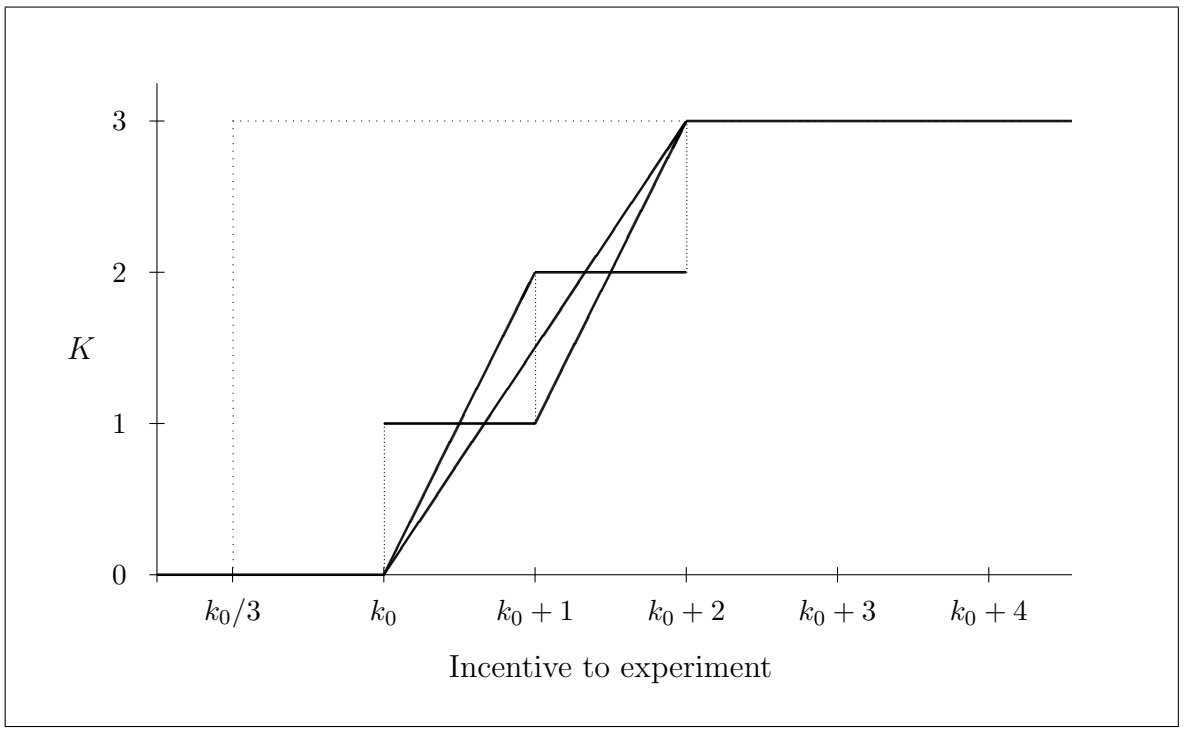

Figure 1: Intensity of experimentation in three-player equilibria

as a function of the incentive to experiment, defined by

$$I(\pi) = \frac{f(\pi) - s}{s - m(\pi)}$$

when $m(\pi) < s$, and ∞ otherwise. Equation (2) can then be rewritten as

$$k_n^* \begin{cases} = 0 & \text{if } I(\pi) < k_0 + K_{\neg n}, \\ \in [0,1] & \text{if } I(\pi) = k_0 + K_{\neg n}, \\ = 1 & \text{if } I(\pi) > k_0 + K_{\neg n}. \end{cases} \tag{3}$$

(See Bolton and Harris, 2000, Sections 6 and 7.)

Let J_0 denote the set of players using S exclusively, and J_1 denote the set of players using R exclusively. Let $j_0 = \#J_0$ and $j_1 = \#J_1$.

Case 1: $j_0 + j_1 = N$.
(a) $j_1 = 0$, $I(\pi) < k_0$;
(b) $0 < j_1 < N$, $k_0 + j_1 - 1 < I(\pi) < k_0 + j_1$;
(c) $j_1 = N$, $k_0 + N - 1 < I(\pi)$.

If $j_1 < N$, then for any player n in J_0 we have $K_{\neg n} = j_1$, so $k_n^* = 0 \Leftrightarrow I(\pi) < k_0 + j_1$.
If $j_1 > 0$, then for any player n in J_1 we have $K_{\neg n} = j_1 - 1$, so $k_n^* = 1 \Leftrightarrow I(\pi) > k_0 + j_1 - 1$.
This case leads to the horizontal sections in the figure.

Case 2: $j_0 + j_1 = N - 1$.
• $0 \leq j_1 < N$, $I(\pi) = k_0 + j_1$, and the player not in J_0 or J_1 chooses *any* $k \in (0, 1)$.

If player n is not in J_0 or J_1 we have $K_{\neg n} = j_1$, so $k_n^* \in (0, 1) \Leftrightarrow I(\pi) = k_0 + j_1$.
For any player n in J_0 we have $K_{\neg n} = j_1 + k$, so $k_n^* = 0 \Leftrightarrow I(\pi) < k_0 + j_1 + k$, i.e. $k > 0$.
For any player n in J_1 we have $K_{\neg n} = j_1 - 1 + k$, so $k_n^* = 1 \Leftrightarrow I(\pi) > k_0 + j_1 - 1 + k$, i.e. $k < 1$.
This case leads to the vertical sections in the figure.

Case 3: $j_0 + j_1 < N - 1$.
• $I(\pi) = k_0 + j_1 + [N - (j_0 + j_1) - 1]k$, $0 < k < 1$, and each player not in J_0 or J_1 chooses the *same* $k \in (0, 1)$ such that the preceding equality is met.

If player n is not in J_0 or J_1 we have $K_{\neg n} = j_1 + [N - (j_0 + j_1) - 1]k$, so $k_n^* \in (0, 1) \Leftrightarrow I(\pi) = k_0 + j_1 + [N - (j_0 + j_1) - 1]k$.
For any player n in J_0 we have $K_{\neg n} = j_1 + [N - (j_0 + j_1)]k$, so $k_n^* = 0 \Leftrightarrow I(\pi) < k_0 + j_1 + [N - (j_0 + j_1)]k$, i.e. $k > 0$.
For any player n in J_1 we have $K_{\neg n} = j_1 - 1 + [N - (j_0 + j_1)]k$, so $k_n^* = 1 \Leftrightarrow I(\pi) > k_0 + j_1 - 1 + [N - (j_0 + j_1)]k$, i.e. $k < 1$.
Futher, $k \in (0, 1) \Leftrightarrow k_0 + j_1 < I(\pi) < k_0 + j_1 + [N - (j_0 + j_1) - 1]$.
This case leads to the sloping sections in the figure.

All the MPE are just combinations of these three cases. In Figure 1 above, the dotted line is the efficient outcome, and we can see the equilibrium experimentation that approaches this the closest as the upper envelope consisting of alternating horizontal and sloping solid lines – this is the equilibrium that maximizes total experimentation at any given belief, and, as such, should also maximize aggregate payoffs.

Concluding Remarks

We have seen that the players' best responses depend only on the current expected payoff from the risky arm and the full-information payoff. In all set-ups where the agents' prior belief about the unknown parameter is a two-point distribution, these two payoffs are identical and thus the equilibrium actions are the same function of the agents' belief – the specification of the noise is irrelevant. Even for very different distributions of the agents' prior belief, the equilibrium actions depend only on the ratio of these two payoff functions adjusted by the safe payoff.

Of course, the evolution of the agents' posterior belief does depend on how the noise is modelled, as do equilibrium payoffs. To calculate the equilibrium payoffs in the three examples, one has to solve a second order ordinary differential equation in example 1 (Brownian noise), a first order ordinary differential-difference equation in example 2(a) (Poisson noise, two-point distribution), and a first order system comprising an ordinary difference equation and a partial differential equation in example 2(b) (Poisson noise, Gamma distribution). With regard to example 1, see Bolton and Harris (2000), and for example 2(a), see Keller, Rady and Cripps (2005); example 2(b) is work in progress.

References

Bolton, P. and C. Harris (1999): "Strategic Experimentation," *Econometrica*, **67**, 349–374.

Bolton, P. and C. Harris (2000): "Strategic Experimentation: the Undiscounted Case," in *Incentives, Organizations and Public Economics – Papers in Honour of Sir James Mirrlees*, ed. by P.J. Hammond and G.D. Myles. Oxford: Oxford University Press, 53–68.

DeGroot, M. (1970): *Optimal Statistical Decisions*. New York: McGraw Hill.

Karlin, S. and H.M. Taylor (1981): *A Second Course in Stochastic Processes,* 2nd edition. New York: Academic Press.

Keller, G., S. Rady and M. Cripps (2005): "Strategic Experimentation with Exponential Bandits," *Econometrica*, **73**, 39–68.

Liptser, R.S. and A.N. Shiryayev (1977): *Statistics of Random Processes I.* New York: Springer-Verlag.