DISCRETE ACTIONS IN INFORMATION-CONSTRAINED TRACKING PROBLEMS

FILIP MATEJKA AND CHRISTOPHER A. SIMS

ABSTRACT. Optimal actions of an agent facing a Shannon capacity constraint on the translation of an uncertain signal into an action can easily turn out to be discretely distributed, even when the objective function and the initial distribution of uncertainty contain no discrete elements. We show this result analytically in a broad class of cases. It has implications for the interpretation of observed intervals between changes in prices or other economic choice variables in micro-data as indicators of costs of adjustment or of the degree of "stickiness" in responses to aggregate policy changes or business cycle fluctuations.

I. INTRODUCTION

Prices of individual products in most markets do not change continually, but instead stay fixed for spans of time, then jump to new values. We have simple theories that imply such behavior of prices is optimal (menu costs models) or treat such behavior as a constraint (Calvo pricing). As fine-grained microdata on individual product prices has become available, however, we can see that in at least some markets (e.g. grocery stores) prices not only stay fixed for spans of time, when they do change they sometimes move back and forth across a finite array of values. Our simple models explain the fixity for spans of time, but on their face imply that when a

---

price change occurs, the change should be continuously distributed. We don't explain why the price should change, then come back to *exactly the same* price as before the change, for example.

Rational inattention (RI) theory models individuals as having finite information-processing "capacity" in the sense of Shannon (see MacKay (2003) for a textbook treatment). It is a theory about why, when information appears to be available at little or no cost, individuals may not respond to it, or may respond erratically. It has intuitive appeal — most of us, most days, do not look up the term structure of interest rates and make corresponding fine adjustments in our checking account balances, even though, with an internet connection this could be done very easily. RI also has qualitative implications about delay and noise in reactions to information that roughly match empirical relationships among macro variables.[1]

In a simple model he interpreted as a two-period savings problem, Sims (2006) found that numerical solutions for optimal behavior, even when exogenous randomness was continuously distributed, implied discretely distributed behavior. Matějka (2008) explored a model he interpreted as describing a Shannon-capacity constrained monopolistic seller with random costs and showed there again that numerical solution tended to imply discretely distributed behavior. In fact, the time paths of prices emerging from his simple model matched many of the qualitative features of individual product time paths shown in, e.g., Eichenbaum, Jaimovich, and Rebelo (2008).

Matějka (2009) considered the behavior of a monopolistic price setter that is not information-constrained, facing consumers who have finite capacity. The consumers will choose discretely distributed behavior, and this turns out to imply that it is optimal for the seller to set discretely distributed prices.

---

[1]The intuition and this qualitative match are discussed at more length in Sims (2003)

The question of whether "stickiness" reflects something like menu costs, or instead rational inattention, is important for macroeconomic policy modeling. If stickiness reflects RI, its form will change systematically if the stochastic process followed by the economy changes. RI implies that rational expectations models, developed to explain how a change in policy behavior could change the non-policy part of a model, are themselves subject to a similar critique.

Perhaps more important, models that explain stickiness via adjustment costs of one sort or another imply that rapid change in the sticky choice variables is costly or distasteful. RI models, which explain stickiness as reflecting information processing costs, do not imply that rapid change is in itself costly or distasteful. On the other hand, they imply that there is a cost to processing information that existing theories do not take into account. If, for example, an environment of high inflation requires individual consumers and producers to devote more attention to tracking prices, there may be a cost that is not captured in the behavior of direct arguments of production and utility functions. RI therefore does not necessarily generically imply that cyclical fluctuations are more or less important than in adjustment-cost models, but it could imply quite different estimates of welfare costs — and thus different conclusions about optimal policy.

In this paper we consider a restricted class of models that look somewhat like Matějka's information-constrained pricing model and show analytically that in this class of models, for a wide class of situations where all the inputs to the problem — the distribution of exogenous randomness and the objective function — are very smooth, behavior of RI agents is nonetheless discrete. This provides some reassurance that the apparent discreteness in the computational solutions in previous work was not an anomaly, and also may give us some insight into the conditions under

which RI solutions are likely to emerge as discrete. We also display classes of examples in which capacity-constrained agents will choose continuously distributed behavior, to supply qualitative guidance as to when capacity constraints are likely to lead to discrete behavior.

II. THE TRACKING PROBLEM

The problem is

$$
\max_{f,\mu_x} \int U(|x-y|)f(x,y)\,\mu_x(dx)\,\mu_y(dy)
$$

$$
- \alpha^{-1}\left(\int \log(f(x,y))f(x,y)\,\mu_x(dx)\,\mu_y(dy)\right. \tag{1}
$$

$$
\left. + \int \log\left(\int f(x,y')\,\mu_y(dy')\right)f(x,y)\,\mu_x(dx)\,\mu_y(dy)\right)
$$

$$
\text{subject to } \int f(x,y)\,\mu_x(dx) = g(y)\,,\text{ a.s. } \mu_y \tag{2}
$$

$$
f(x,y) \geq 0\,,\text{ all } x,y\,, \tag{3}
$$

where μ_x and μ_y are each a σ-finite Borel measure on $\mathbb{R}$, possibly but not necessarily Lebesgue measure, f is the joint pdf of the choice x and the target y, g is the given pdf for y, before information collection, $U(|y-x|)$ is the objective function being maximized, and α is the inverse of the cost of information (or of the Lagrange multiplier on the information constraint).

The objective function is concave in the measure on xy space defined by f, μ_y and $\mu_x{}^2$, and the constraints are linear, so we can be sure that a solution to the first order conditions (FOC's) is a solution to the problem. However the non-negativity

---

[2]Expected utility is linear in this probablity measure, and mutual information between two random variables is a convex function of their joint distribution, so expected utility minus θ times the mutual information is concave in the measure.

constraints can be binding, so that exploration of which constraints are binding may make solution difficult.

The FOC's of the problem with respect to f imply that at all values of x, y with $f(x, y) > 0$ and $g(y) > 0$

$$U(x, y) = \theta(y) + \alpha^{-1} \log \left(\frac{f(x, y)}{\int f(x, y) \, dy} \right) \tag{4}$$

$$\therefore f(x, y) = p(x) e^{\alpha U} h(y) \tag{5}$$

$$\therefore \quad p * e^{\alpha U} \cdot h = g, \tag{6}$$

where p is the pdf of the action x and h is a function that is non-zero where g is non-zero, zero otherwise. The "$*$" symbol denotes convolution. At points x where $f(x, y) = 0$, the FOC's require that the left hand side of (4) be less than or equal to the right hand side. Note that if $p(x) = \int f(x, y) \, dy > 0$, the right hand side of (4) is minus infinity wherever $f(x, y) = 0$, so with bounded U we can conclude that $f(x, y) = 0$ for a particular x, y only if $f(x, y) = 0$ for all y, i.e. $p(x) = 0$.

At points x with $p(x) = 0$, the right-hand side of (4) is undefined. However we can reparameterize $f(x, y)$ as $p(x) q(y \mid x)$ and take the first order condition with respect to p. Since at points with $p(x) = 0$ the value of $q(\cdot \mid x)$ makes no marginal contribution to the objective function or the constraints, the first order condition with respect to p at points with $p(x) = 0$ becomes

$$\max_q E[U(x, y) - \theta \log(q(y \mid x)) - \lambda(y) \mid x] \leq 0. \tag{7}$$

These FOC's do not take explicit account of the possibility of varying μ_x. Adding or deleting a point x with non-zero μ_x probability is accounted for, since that can be treated as setting $p(x)$ to a zero or non-zero value at a point where $\mu_x(x) > 0$. If μ_x puts discrete probability π on a point x_0, we can, though, derive an additional FOC by considering changing the location of x_0. If we change the location of x_0 to a nearby

$x^* \neq x_0$ that initially had probability zero (though possibly a non-zero density value w.r.t. Lebesgue measure), while keeping the pdf of $y \mid x_0$ and of $y \mid x^*$ the same, we leave the mutual information between x and y the same and continue to satisfy the boundary condition (2), but we change the expected value of U. The derivative of the expected value of U w.r.t. x_0 when x_0 is changed in this way is

$$\mu_x(x_0) \int \frac{\partial U(x_0, y)}{\partial x_0} f(x_0, y) \, dy \,, \tag{8}$$

which then implies that $E[\partial U / \partial x \mid x]$ must be zero at every point x_0 that has positive probability.

III. When does x have a discrete distribution?

Even when y is continuously distributed, it turns out to be optimal to make x discretely distributed, under some fairly general conditions. We now proceed to give some sufficient conditions for discreteness of x.

Convolution tends to smooth a function, making it "more differentiable", so convoluting an analytic function with some weighting function leaves it analytic. To prove this formally, we first cite a known result:

Lemma III.1. *f is analytic on an open subset S of the real line, if and only if on any compact subset of $K \subset S$ there is a $C > 0$ such that for all $x \in K$, $\sum C^j j! f^{(j)}(x) < \infty$.*

Proof. See lemma 1.2.10 in Krantz and Parks (2002). □

Lemma III.2. *If*

   (i) *p is a finite measure on the real line $\mathbb{R}$ and*

   (ii) *f is a bounded real-valued function, analytic on the whole of $\mathbb{R}$,*

then $p * f$ is analytic on $\mathbb{R}$.

Proof. With f analytic on the entire real line, if p has bounded support, we can, for any bounded interval $J \in \mathbb{R}$, find a larger interval $M \supset J$ such that for $x \in J$

$$\frac{d^n}{dx^n}(p * f)(x) = \int_{z \in M} f^{(n)}(x - z)p(dz) \,.$$

Let C_M be the constant C from Lemma III.1 applied to f and the interval M. Then, since each derivative of $p * f$ at points in J is an average of the corresponding derivatives of f at points in M, C_M can play the role of C in lemma III.1 to show that $p * f$ satisfies the condition of the lemma for J. But since we can do this for any bounded interval J, we conclude that $p * f$ is analytic on the whole of $\mathbb{R}$.

Because f is real-analytic on $\mathbb{R}$, it has an analytic extension to an open subset S of the complex plane $\mathbb{C}$ that contains $\mathbb{R}$. Let B be a strict upper bound on $|f|$ over $\mathbb{R}$. The set $S \cap \{z \in \mathbb{C} \mid |f(z)| < B\}$ is an open subset of $\mathbb{C}$ containing $\mathbb{R}$ and on which f is bounded. So we can without loss of generality assume f is bounded by B on S. If p does not have bounded support, consider the sequence of functions $p_n * f$, where p_n is the restriction of p to the interval $(-a_n, a_n)$ and $a_n \to \infty$ as $n \to \infty$. (I.e. for a set $A \in \mathbb{R}$, $p_n(A) = p(A \cap (-a_n, a_n))$.) Because p is a finite measure and f is bounded,

$$|p_n * f - p * f| = |(p_n - p) * f| < \left(p((-\infty, -a_n)) + p((a_n, \infty)) \right) \cdot B \xrightarrow[n\to\infty]{} 0 \,,$$

and it is thus clear that $p_n * f$ converges to $p * f$ uniformly over S. But the uniform limit of analytic functions on an open subset of $\mathbb{C}$ is itself analytic (Ahlfors, 1953, Section 1.4), and therefore $p * f$ is real-analytic on $\mathbb{R}$.   □

Now we can specify a first fairly wide class of cases in which the solution of our tracking problem will necessarily make μ_x discrete.

Proposition 1. *Suppose in the tracking problem* (1)-(3)

   (i) *$e^{U(\cdot)}$ is analytic and positive on the entire real line;*

  (ii) *U is non-increasing in its argument and not constant; and*

(iii) *g has bounded support.*

Then p is concentrated on a finite set of points within the support of g.

Proof. First, p surely has bounded support, contained within the support of g. If p put positive measure on any set outside of the support of g, we could modify the joint distribution by moving the probability on that set onto points y in the support of g that would make $U(|x - y|)$ larger for all x in the support of g. That there is such a point inside the support of g for every y outside the support we know from the monotonicity of U. To be precise, let $S \subset \mathbb{R}$ be the support of g, S_L be the set of points below S and S_H be the set of points above S. If $f(x, y)$ were a solution to the tracking problem with

$$\int_{S_L} f(x, y)dx + \int_{S_H} f(x, y)dx > 0,$$

we could set $f(x, y) = 0$ for all $x \in S_L$ and all $x \in S_H$ and increase the discrete probability on the greatest lower bound and least upper bound of S by

$$\int_{S_L} f(x, y)dx \quad \text{and} \quad \int_{S_H} f(x, y)dx,$$

respectively. This necessarily reduces or leaves unchanged the mutual information between x and y. To see this, let $q(y \mid x)$ be the conditional pdf of $y \mid x$ before we rearrange the probabilities, and note that for each x in the interior of S, the conditional distribution of $x \mid y$ is unchanged. The new conditional pdf at the boundaries is given by

$$q^*(y \mid x = \inf(S)) = \frac{\int_{S_L} p(x)q(y \mid x)dx}{\int_{S_L} p(x)dx}$$

and a similar expression for the upper bound. The mutual information between X and Y can be characterized as the expected reduction in entropy between the conditional and unconditional distributions for Y. Since entropy is concave, the entropy of an average, like q^* above, is always larger than or equal to the average of the

entropies. Thus rearranging the probabilities as we have proposed never increases mutual information.

We know from the FOC's that for points y within the support of g,

$$h(y) = \frac{g(y)}{p * e^{\alpha U}(y)} \, .$$

This will be a μ_y-integrable function because g is a density w.r.t. μ_y and by the assumption that e^U is positive. Then, again from the FOC's, we have, at points where $p(x) > 0$,

$$e^{\alpha U} * h(x) = 1 \, .$$

Since h is integrable and we have assumed $e^{\alpha U}$ analytic on $\mathbb{R}$, by Lemma III.2 $e^{\alpha U} * h$ is analytic on $\mathbb{R}$. Thus if there is any sequence of points $\{x_i\}$ with a limit point x_∞ in $\mathbb{R}$ on which $p(x_i) > 0$, for all i, $e^{\alpha U} * h$ is constant on all of $\mathbb{R}$. But with $e^{\alpha U}$ monotone decreasing and h having bounded support, $h * e^{\alpha U}$ cannot be constant. Near $x = 0$ it is a weighted sum of values of $e^{\alpha U}$ that are all larger than or equal to, and some strictly larger than, the values of $e^{\alpha U}$ whose weighted sum make up $h * e^{\alpha U}(x)$ for $|x|$ large. So we have reached a contradiction under the assumption that there is a sequence of x's with a limit point such that $p(x_i) > 0$, all i. Therefore there must be no such sequence. That implies, since the support of p is bounded by that of g, that p is concentrated on a finite set of points within the support of g. $\qquad\square$

A second broad class of cases for which we can characterize solutions arises when g and e^U are both analytic. We use $\tilde{f}$ to denote the Fourier transform of f.

Proposition 2. *Suppose in the tracking problem*

   (i) $e^{\alpha U(|x|)}$ *is analytic in x on the whole real line $\mathbb{R}$,*

   (ii) $e^{\alpha U(|x|)}$ *is integrable,*

(iii) $e^{\alpha U(|x|)}$ *is monotone increasing for negative arguments and decreasing for positive arguments and not constant, and*

(iv) $g(y)$ *is analytic in y on $\mathbb{R}$.*

then if $\tilde{g}/\widetilde{e^{\alpha}U}$ is the characteristic function of a probability distribution on $\mathbb{R}$, that distribution is the marginal distribution of x in the solution to the tracking problem. Otherwise, the solution makes the marginal distribution of x discrete, putting probability one on a set with no limit points.

Proof. From Lemma III.2 we know that, whatever the solution, it will imply a marginal probability distribution π for x, and $\pi * e^{\alpha U}$ will be analytic. Furthermore, because $e^{\alpha U}$ is analytic and positive on the whole real line, $\pi * e^{\alpha U}$ must also be positive everywhere. Thus $h = g/(\pi * e^{\alpha U})$ being the ratio of two analytic functions in which the denominator is always postive, is itself analytic. But we know that in the solution $h * e^{\alpha U}(x) = 1$ at all points x with non-zero probability density. Thus if π puts positive probability on a set of points containing a limit point, $e^{\alpha U} * h$ is constant. This is possible, because of the fact that $e^{\alpha U}$ decreases as we move away from zero, only if h is constant. In that case $\pi * e^{\alpha} U = g$ and therfore

$$\tilde{\pi} = \tilde{g}/\widetilde{e^{\alpha U}}.$$

If the right-hand side of this expression fails to be the characteristic function of a probablity distribution on $\mathbb{R}$, we have reached a contradiction. In that case, therefore, it must be that π puts probability one on a set that has no limit points. $\square$

Note that the marginal distribution π of x could turn out to be discrete in any case, because it is possible that $\tilde{g}/\widetilde{e^{\alpha U}}$ is the characteristic function of a discrete distribution. Note also that this proposition gives us a simple rule to apply when g and e^{U} are both analytic: form $\tilde{g}/\widetilde{e^{\alpha U}}$, take its inverse Fourier transform, and verify that the result is a probability distribution. If so, one has solved the problem. If not, the

solution will be discrete, but finding its exact form probably requires a numerical approach.

A necessary condition that a function $\tilde{\phi}$ be the characteristic function of a distribution is that it is continuous, bounded, and takes the value 1 at 0. But this leaves open the possibility that its inverse Fourier transform might not be everywhere positive, and there is no check for that simpler than taking the inverse Fourier transform.

When $\tilde{g}/\widetilde{e^{\alpha U}}$ is a characteristic function, the solution implies that we can write

$$y = x + \nu,$$

where ν is independent of x and has a pdf proportional to $e^{\alpha U}$. Note that this does not imply that we can model information flow as an observation of y with i.i.d. error. It is $y \mid x$ that has an i.i.d. error; $x \mid y$ will have a distribution dependent on y except in special cases (e.g. the Gaussian case, with g a normal pdf and g quadratic).

IV. EXAMPLES

Suppose $U = -(y - x)^2$, g is Gaussian with variance σ^2, and the "information per util" parameter is α as in the problem statement (1-3). This will fit the conditions of Proposition 2. Clearly $e^{\alpha U}$ has the shape of a Gaussian pdf in y, centered at x, for each x, and if $p(x)$ is also Gaussian we will have that $p * e^{\alpha U}$ is the pdf of a Gaussian distribution with variance the sum of the variance of the p distribution with $1/(2\alpha)$. However, if $\sigma^2 < 1/(2\alpha)$ this is impossible. The result is that when $\sigma^2 > 1/(2\alpha)$ the solution makes x normally distributed and also makes $y \mid x$ normal. However, if $\sigma^2 \leq 1/(2\alpha)$, the solution is simply $x \equiv E[y]$. That is, if the utility cost of information exceeds $2\sigma^2$, it is optimal to collect no information and set x to a constant, which is of course a discrete distribution, as proposition 2 implies.

This is a special case of a similar result for any infinitely divisible distribution. For example, if g is Cauchy, i.e. with density $1/(\sigma\pi(1 + (y/\sigma)^2)$, if $U(|z|) = -\log(\gamma +$

z^2), and if $\alpha = 1$, then if $\sqrt{\gamma} < \sigma$, there is a solution of the form $y = x + v$ with x and v both Cauchy and v independent of x. Note, though, that unlike the case where U is quadratic, this result holds only with $\alpha = 1$. If $\alpha \neq 1$, $e^{\alpha U}$ has the shape of a t distribution with degrees of freedom $n = 2\alpha - 1$. Obtaining closed form results on whether $\widetilde{\tilde{g}/e^{\alpha U}}$ is the characteristic function of a distribution in this general case seems to be a non-trivial task.

A more interesting case is where g is standard Cauchy and U is quadratic. Then taking the Fourier transform of g we find its characteristic function $\tilde{g} = e^{-|\omega|}$. If there were to be a solution of the form $y = x + v$, we would need to have the characteristic function of p satisfy

$$\tilde{p} = e^{\alpha^{-1}\omega^2 - |\omega|} \ . \tag{9}$$

The right-hand-side of this expression is never bounded and thus can never be the FT of a pdf. Thus no matter what α is, there is no solution of the $y = x + v$ form, and by proposition 2 we know the solution is discrete. But here a degenerate distribution with x concentrated at a single point can never be a solution, as that implies infinite losses (because the Cauchy distribution has infinite variance) and finite losses are possible. Without deriving a solution, we can nonetheless see that it is possible to make losses finite at arbitrarily low information cost. Suppose we divide the real line into intervals with endpoints $na + \frac{a}{2}$ for fixed a and $n = -\infty, \ldots, \infty$. Suppose further that the signal we get tells us simply which of these intervals y is in. We would then optimally set x equal to the conditional mean of y given that y is in the interval. If a is very large, the probability of the interval $(-\frac{a}{2}, \frac{a}{2})$ grows arbitrarily close to one, so that the entropy of the x distribution goes to zero. Also, the conditional entropy of $y \mid x$ grows arbitrarily close to the unconditional entropy of y as a increases, meaning that the mutual information between y and x goes to zero. Yet we have made the conditional second moment $E[(y - x)^2 \mid x]$ finite and bounded for all x. This is not

on optimal solution — optimally the points on which the x distribution concentrates will not be evenly spaced, and certainly the conditional distributions of $y \mid x$ will not have bounded support. But it is clear from proposition 2 that a distribution for x concentrated on a countable set of points, with no limit points, will be the outcome.

Of course we can produce examples with analytic solutions by picking a π and constructing g as $\pi * q$ for some particular q. For example, suppose we pick π_0 to be a discrete distribution with equal weights of $\frac{1}{2}$ on $x = \pm 1$ and g as an equal-weight mixture of two standard normal pdf's with means ± 1. Then $g = \pi_0 * \phi$, where ϕ is the standard normal density. Thus if $U = -x^2$ and $\alpha = \frac{1}{2}$, both g and $e^{\alpha U}$ are analytic and the optimal solution is to make x discretely distributed, with equal weights on ± 1. In other words, $\pi = \pi_0$, so that it is optimal to get a signal that is simply "high" or "low" and take the corresponding action. If $\alpha > \frac{1}{2}$, it is easy to check from the characteristic functions that the optimal π is no longer discrete. It is a mixture of two normals with means ± 1 and variances $1 - 1/(2\alpha)$. If $\alpha < \frac{1}{2}$, $\widetilde{g}/\widetilde{e^{\alpha U}}$ is unbounded, thus not a characteristic function, so again we know that the solution is entirely discrete. In fact, from numerical solutions it seems clear that the solution concentrates the distribution of x on two points $\pm a$, with a approaching one as α approaches $\frac{1}{2}$ from below and approaching zero as some lower bound on α (i.e. upper bound on the cost of information) is approached.

We now proceed to consider some numerical examples.

The Proposition 1 result applies more widely than one might think. Solutions where x has support on a small number of points emerge not when g itself is sharply truncated, but when it is truncated *relative* to the spread of the $e^{\alpha U}$ function. Figure 1 shows the solution for a case where g is a normal pdf truncated at ± 3 standard errors, so that only a tiny fraction of the total probability under the normal density is excluded. The U function is quadratic and has a spread narrower than that of g,
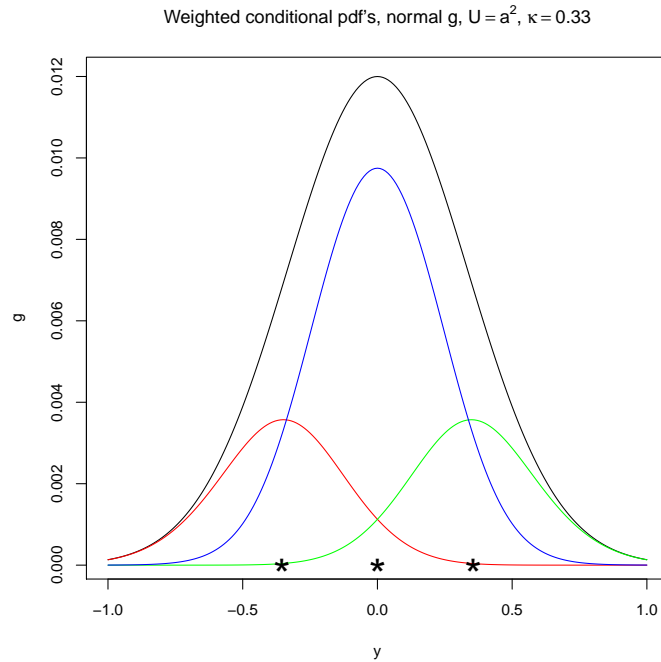
Weighted conditional pdf's, normal g, $U = a^2$, $\kappa = 0.33$

FIGURE 1.

so with unbounded support for g we know from Proposition 2 that a continuously distributed x would emerge. But the actual solution is supported on three points. To interpret the graph, note that the black line is g — a normal density with standard deviation $\frac{1}{3}$ — and the colored lines are the conditional densities for y at the support points of the x distribution, weighted by the probabilities of the support points. The colored densities at each point on the x axis sum to g. The support points themselves are indicated by the black asterisks along the horizontal axis. For this and all the subsequent figures, the displayed solution has been verified to be optimal by searching for a solution with more points of support and finding convergence to the displayed solution. The g distribution is not actually continuous in these exercises, but instead is concentrated on 201 equi-spaced points on (-1,1).

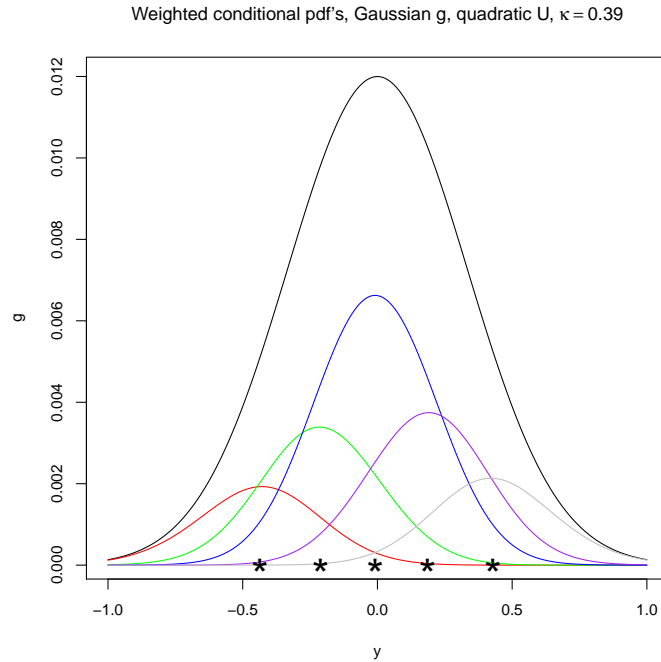Weighted conditional pdf's, Gaussian g, quadratic U, κ = 0.39

FIGURE 2.

Even though the solution has support on only a few points, note that the shape of the conditional distributions for $y \mid x$ are quite similar to normal densities, and in that sense are like what is suggested by the Proposition 2 result when the solution is continuous. We will see this pattern in most of the examples below — solutions are supported on small numbers of points for x as suggested by Proposition 1, while conditional densities for $y \mid x$ tend to resemble $e^{\alpha U(|x-y|)}$, as suggested by Proposition 2.

If we increase α — make information cheaper — the spread of $e^{\alpha U}$ shrinks relative to the (-1,1) interval that supports g, and the solution puts weight on more points, as in Figure 2. The slight asymmetry visible in the plotted solution reflects imprecision in the numerical optimization. The solution can be symmetrized by averaging

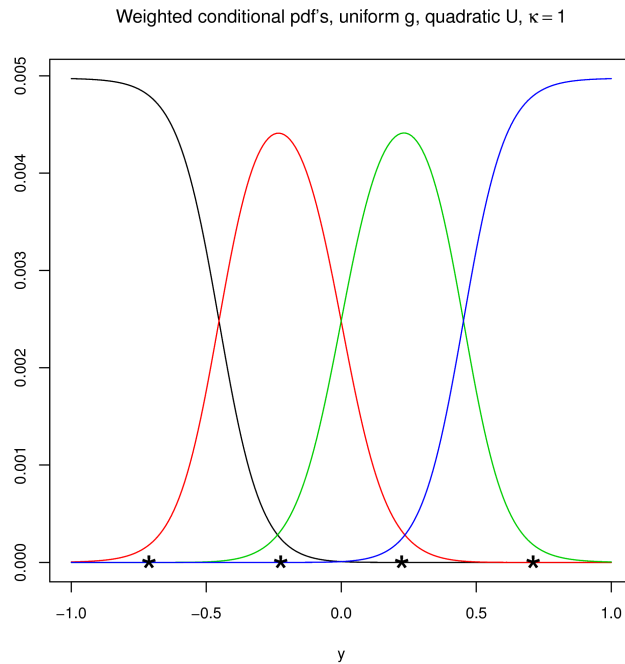Weighted conditional pdf's, uniform g, quadratic U, κ = 1

FIGURE 3.

negative and positive parts of the displayed solution, but this increases the objective function only by 1 digit in the seventh decimal place. The asymmetry therefore reflects the fact that solutions can differ visibly from the computed optimum while delivering objective function values that are extremely close to the optimum. Notice that the values of κ, the mutual information between x and y, shown at the top of Figures 1 and 2 differ by only .06. These κ values are measured in nats, that is, base-e information units. In the more intuitive bits (base-2) unit, the difference would be .09.

In Figures 3 and 4 we see solutions with a uniform g, where the truncation involves a sharply discontinuous dropoff in the density g at the boundaries of its support. Here we see that the κ values are generally higher for a given number of support

Weighted conditional pdf's, uniform g, quadratic U, κ = 1.2
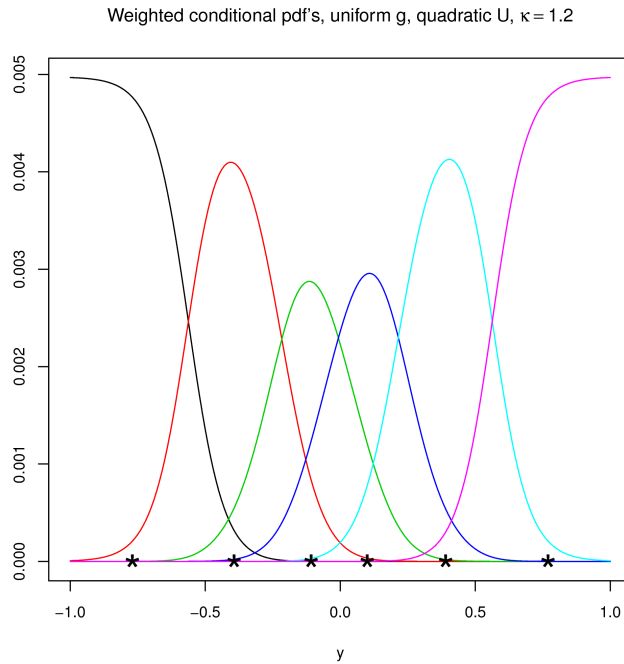
FIGURE 4.

points for x, and the increase in κ in going from 4 to 6 support points is three times as large as that for going from 3 to 5 support points in the normal-g example.

Figures 5 and 6 display solutions with a uniform g and $U(z) = -z^{1.1}$. This makes the peak of the objective function sharp; the first derivative changes very rapidly at the peak. We can see in the figures that this shape of the objective function is reflected in the conditional densities for $y \mid x$ in the solutions, with the conditional pdf's showing kinks.

Figures 7 and 8 show solutions for a truncated Cauchy g with quadratic U. They show that most of the probability distribution for x concentrates on $x = 0$, even as the information constraint is relaxed, and that as the constraint is relaxed, more small-probability x values in the tails of the distribution appear.
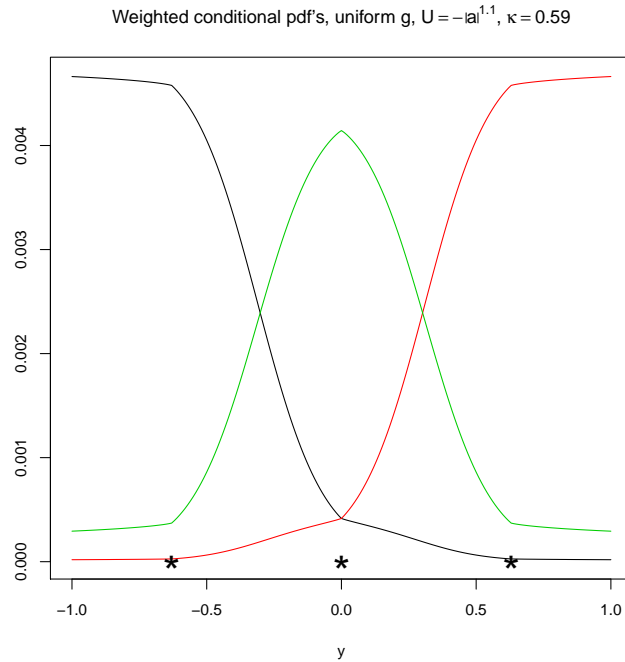
Weighted conditional pdf's, uniform g, $U = -|a|^{1.1}$, $\kappa = 0.59$

FIGURE 5.

V. IMPLICATIONS FOR MODELING ECONOMIC BEHAVIOR

The kind of model explored here, in which a decision-maker reacts to external information subject to a tight information constraint, seems a natural one to apply to the behavior of individuals reacting to fairly frequent and/or numerous economic signals for which the consequences of imprecise responses are modest. This might be true, for example, of price-setters in retail establishments that must set hundreds of prices every day in response to fluctuations in demand and costs for all the items. It might also be true of day-to-day or month-to-month savings and spending decisions of individuals, facing a potentially vast array of information about asset markets.

These models suggest that it is a mistake to identify times at which decision makers' choices change as times at which they fully optimize in reaction to the current state. Rationally inattentive decision makers as modeled here may change their
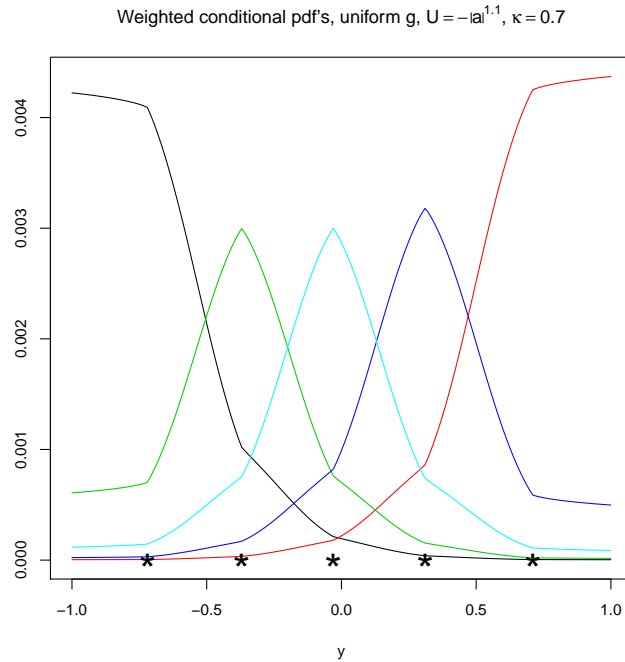
Weighted conditional pdf's, uniform g, $U = -|a|^{1.1}$, $\kappa = 0.7$

FIGURE 6.

choices randomly, even when the state of the world (the draw from the g distribution) is unchanged. As modeled here, they acquire information about the true state every period, but may nonetheless not change their behavior, even though the true state is changing.

Rationally inattentive decision makers faced with a fat-tailed distribution for the state (as in our Cauchy-g examples) may go for long periods without making any change in their behavior, then make a large and temporary change. The long periods of unchanged behavior are not an indication that changing behavior is "costly", and the large, rare changes are not an indication that changing behavior is not costly. There is no cost of change at all in these models. Apparent inertia in the face of a fat-tailed distribution simply reflects the fact that with information flows valuable,
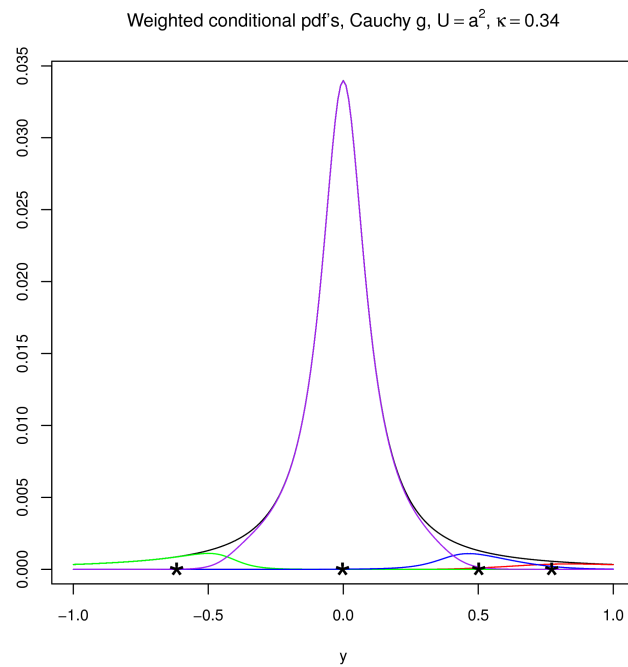
FIGURE 7.

it makes sense to concentrate attention on the rare extreme draws of y, not reacting at all to the usual small fluctuations about the central value.

REFERENCES

AHLFORS, L. V. (1953): *Complex Analysis*. McGraw-Hill, New York, Toronto, London.

EICHENBAUM, M., N. JAIMOVICH, AND S. REBELO (2008): "Reference Prices and Nominal Rigidities," Discussion paper, Northwestern University and Stanford University, NBER Working paper 13829.

KRANTZ, S., AND H. R. PARKS (2002): *A Primer of Real Analytic Functions*. Birkhäuser, second edn.

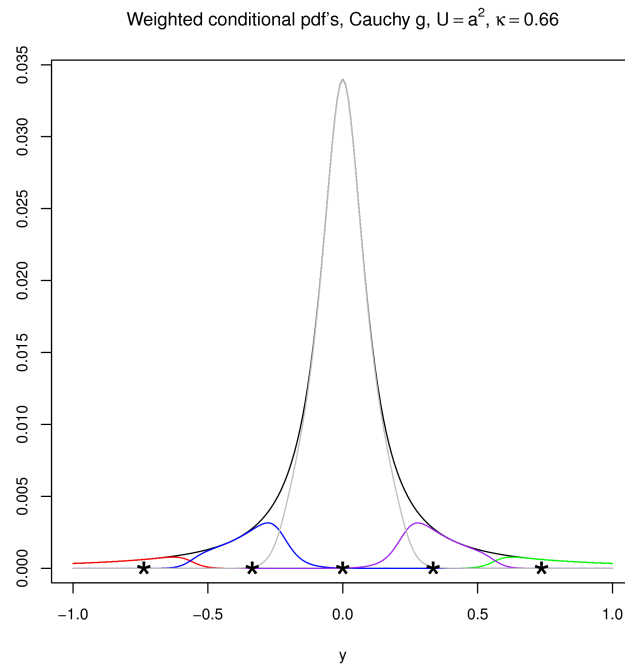MACKAY, D. J. C. (2003): *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Weighted conditional pdf's, Cauchy g, $U = a^2$, $\kappa = 0.66$

FIGURE 8.

MATĚJKA, F. (2008): "Rationally Inattentive Seller: Sales and Discrete Pricing," Discussion paper, PACM, Princeton University, `http://www.pacm.princeton.edu/publications/Matejka_F_2008-wp.pdf`.

——— (2009): "Rigid Pricing and Rationally Inattentive Consumer," Discussion paper, Princeton University.

SIMS, C. A. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3), 665–690.

——— (2006): "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review*, 96(2), 158–163, Proceedings issue.

CERGE-EI, CHARLES UNIVERSITY PRAGUE AND ACADEMY OF SCIENCES OF THE CZECH REPUBLIC, POLITICKYCH VEZNU 7, PRAGUE 11121, CZECH REPUBLIC.

E-mail address: filip.matejka@azpiro.cz

DEPARTMENT OF ECONOMICS, FISHER HALL, PRINCETON UNIVERSITY, PRINCETON, NJ 08544

E-mail address: sims@princeton.edu