# Homework 4 - Suggested Solutions[1]

## Question 1.

In the lectures, I set up the optimization problem for a decision maker who has Shannon Entropy costs of attention in the case there were two states and two acts to choose from. I claimed that, if both acts were chosen, then posterior beliefs would satisfy

$$\frac{\gamma_1^t}{\gamma_1^s} = exp\left(\frac{U_1^a - U_1^b}{\kappa}\right)$$

$$\frac{\gamma_2^t}{\gamma_2^s} = exp\left(\frac{U_2^a - U_2^b}{\kappa}\right)$$

### Part 1

Prove this result.

Optimal behavior solves the following problem

$$\begin{aligned}\{P(t), \gamma_1^t, \gamma_1^s\} &= argmax_{\{P(t),\gamma_1^t,\gamma_1^s\}} P(t)\left[\gamma_1^t U_1^a + (1-\gamma_1^t)U_2^a\right] + (1-P(t))\left[\gamma_1^s U_1^b + (1-\gamma_1^s)U_2^b\right] \\ &- \kappa\left(P(t)\left[\gamma_1^t \ln\gamma_1^t + (1-\gamma_1^t)\ln(1-\gamma_1^t)\right] + (1-P(t))\left[\gamma_1^s \ln\gamma_1^s + (1-\gamma_1^s)\ln(1-\gamma_1^s)\right]\right)\end{aligned}$$

subject to

$$P(t)\gamma_1^t + (1-P(t))\gamma_1^s = \mu_1$$

If we write down the Langragean and we keep it in terms of $\gamma_2^t, \gamma_2^s$ without substituting $\gamma_1^i + \gamma_2^i = 1$ for $i = s, t$; but imposing these constrains explicitly we have that

$$\begin{aligned}\mathcal{L} &= P(t)\left[\gamma_1^t U_1^a + \gamma_2^t U_2^a\right] + (1-P(t))\left[\gamma_1^s U_1^b + \gamma_2^s U_2^b\right] \\ &- \kappa\left(P(t)\left[\gamma_1^t \ln\gamma_1^t + \gamma_2^t \ln\gamma_2^t\right] + (1-P(t))\left[\gamma_1^s \ln\gamma_1^s + \gamma_2^s \ln\gamma_2^s\right]\right) \\ &+ \lambda_1\left(P(t)\gamma_1^t + (1-P(t))\gamma_1^s - \mu_1\right) \\ &+ \lambda_2\left(\gamma_1^t + \gamma_2^t - 1\right) + \lambda_3\left(\gamma_1^s + \gamma_2^s - 1\right)\end{aligned}$$

[1] If you find any typo please email us: Maria_Jose_Boccardi@Brown.edu or Victor_Aguiar@Brown.edu

Where the first order conditions are given by

$$\frac{\partial \mathcal{L}}{\partial \gamma_1^t} = 0 \quad \Leftrightarrow \quad P(t)\left(U_1^a - \kappa\left(\ln \gamma_1^t + 1\right) + \lambda_1\right) + \lambda_2 = 0 \tag{1}$$
$$\frac{\partial \mathcal{L}}{\partial \gamma_1^s} = 0 \quad \Leftrightarrow \quad (1-P(t))\left(U_1^b - \kappa\left(\ln \gamma_1^s + 1\right) + \lambda_1\right) + \lambda_3 = 0 \tag{2}$$
$$\frac{\partial \mathcal{L}}{\partial \gamma_2^t} = 0 \quad \Leftrightarrow \quad P(t)\left(U_2^a - \kappa\left(\ln \gamma_2^t + 1\right)\right) + \lambda_2 = 0 \tag{3}$$
$$\frac{\partial \mathcal{L}}{\partial \gamma_2^s} = 0 \quad \Leftrightarrow \quad (1-P(t))\left(U_2^b - \kappa\left(\ln \gamma_2^s + 1\right)\right) + \lambda_3 = 0 \tag{4}$$
$$\frac{\partial \mathcal{L}}{\partial P(t)} = 0 \tag{5}$$
$$\gamma_1^t + \gamma_2^t = 1 \tag{6}$$
$$\gamma_1^s + \gamma_2^s = 1 \tag{7}$$
$$P(t)\gamma_1^t + (1-P(t))\gamma_1^s = \mu_1 \tag{8}$$

From equations (1) and (3) we get

$$\lambda_1 = U_2^a - U_1^a - \kappa\left(\ln \gamma_2^t - \ln \gamma_1^t\right) \tag{9}$$

and from equations (2) and (4) we get

$$\lambda_1 = U_2^b - U_1^b - \kappa\left(\ln \gamma_2^s - \ln \gamma_1^s\right) \tag{10}$$

Combining equations (9) and (10) we get that

$$U_2^a - U_2^b - (U_1^a - U_1^b) = \kappa\left[\ln\left(\frac{\gamma_2^t}{\gamma_2^s}\right) - \ln\left(\frac{\gamma_1^t}{\gamma_1^s}\right)\right] \tag{11}$$

From condition (5) combined with (6) and (7) we get that

$$\gamma_1^t\left(U_1^a - U_2^a - \kappa\ln\left(\frac{\gamma_1^t}{1-\gamma_1^t}\right) + \lambda_1\right) - \gamma_1^s\left(U_1^b - U_2^b - \kappa\ln\left(\frac{\gamma_1^s}{1-\gamma_1^s}\right) + \lambda_1\right) = U_2^b - U_2^a - \kappa\ln\left(\frac{1-\gamma_2^s}{1-\gamma_2^t}\right) \tag{12}$$

which combined with equations (9) and (10) and using again (6) and (7) we get that

$$U_2^a - U_2^b - \kappa\ln\left(\frac{\gamma_2^t}{\gamma_2^s}\right) = 0 \tag{13}$$

and therefore

$$\frac{\gamma_2^t}{\gamma_2^s} = exp\left(\frac{U_2^a - U_2^b}{\kappa}\right)$$

On the other hand combining equations (1) and (2) we have that

$$U_1^a - U_1^b - \kappa\ln\left(\frac{\gamma_1^t}{\gamma_1^s}\right) = \frac{\lambda_3}{1-P(t)} - \frac{\lambda_2}{P(t)} \tag{14}$$

and combining (3) and (4)

$$U_2^a - U_2^b - \kappa \ln\left(\frac{\gamma_2^t}{\gamma_2^s}\right) = \frac{\lambda_3}{1-P(t)} - \frac{\lambda_2}{P(t)} = 0 \tag{15}$$

where the last inequality follows from equation (13). Then, from (14) it is the case that

$$U_1^a - U_1^b - \kappa \ln\left(\frac{\gamma_1^t}{\gamma_1^s}\right) = 0$$

and therefore

$$\frac{\gamma_1^t}{\gamma_1^s} = exp\left(\frac{U_1^a - U_1^b}{\kappa}\right)$$

## Part 2

Use this result (and Bayes rule) to find an expression for $\pi_1(t)$ and $\pi_2(t)$, i.e. the probability of receiving signal $t$ in each state (note you should write a formula that describes this values as a function of the primitives of the model - the utilities of different acts in different states and the costs $\kappa$)

By using Bayes rule we have that

$$\pi_1(t) = \frac{\gamma_1^t P(t)}{\mu_1}$$

$$\pi_2(t) = \frac{\gamma_2^t P(t)}{\mu_2} = \frac{\gamma_2^t P(t)}{1-\mu_1}$$

$$\pi_1(s) = \frac{\gamma_1^s P(s)}{\mu_1} = \frac{\gamma_1^s (1-P(t))}{\mu_1}$$

$$\pi_2(s) = \frac{\gamma_2^s P(s)}{\mu_2} = \frac{\gamma_2^s (1-P(t))}{1-\mu_1}$$

For the sake of presentation let $e_1 \equiv exp\left(\frac{U_1^a - U_1^b}{\kappa}\right)$ and $e_2 \equiv exp\left(\frac{U_2^a - U_2^b}{\kappa}\right)$. Then from previous part we know that

$$\frac{\gamma_1^t}{\gamma_1^s} = exp\left(\frac{U_1^a - U_1^b}{\kappa}\right) \Leftrightarrow \gamma_1^t = exp\left(\frac{U_1^a - U_1^b}{\kappa}\right)\gamma_1^s \Leftrightarrow \gamma_1^t = e_1\gamma_1^s \tag{16}$$

$$\frac{\gamma_2^t}{\gamma_2^s} = exp\left(\frac{U_2^a - U_2^b}{\kappa}\right) \Leftrightarrow \gamma_2^t = exp\left(\frac{U_2^a - U_2^b}{\kappa}\right)\gamma_2^s \Leftrightarrow \gamma_2^t = e_2\gamma_2^s \tag{17}$$

$$P(t)\gamma_1^t + (1-P(t))\gamma_1^s = \mu_1 \Leftrightarrow P(t) = \frac{\mu_1 - \gamma_1^s}{\gamma_1^t - \gamma_1^s} \tag{18}$$

$$\gamma_1^t + \gamma_2^t = 1 \Leftrightarrow \gamma_2^t = 1 - \gamma_1^t \tag{19}$$

$$\gamma_1^s + \gamma_2^s = 1 \Leftrightarrow \gamma_2^s = 1 - \gamma_1^s \tag{20}$$

Substituing (16) and (17) into (19) we get that

$$\gamma_2^s = \frac{1 - e_2\gamma_1^s}{e_2}$$

And combining with (20) we get

$$\gamma_1^s = \frac{1-e_2}{e_1-e_2}$$

Substituting back in (20) we get

$$\gamma_2^s = \frac{e_1-1}{e_1-e_2}$$

Substituting these last two expressions into (16) and (17) we get

$$\gamma_1^t = e_1\frac{1-e_2}{e_1-e_2}$$

$$\gamma_2^t = e_2\frac{e_1-1}{e_2-e_2}$$

Plugging back into (18) we get

$$P(t) = \frac{(e_1-e_2)\,\mu_1 - (1-e_2)}{(e_1-1)\,(1-e_2)}$$

Finally we can get the expressions for $\pi_1(t)$ and $\pi_2(t)$ we get

$$\begin{aligned}
\pi_1(t) &= \frac{\gamma_1^t P(t)}{\mu_1} \\
&= \left[e_1\frac{1-e_2}{e_1-e_2}\right]\frac{(e_1-e_2)\,\mu_1-(1-e_2)}{\mu_1\,(e_1-1)\,(1-e_2)} \\
&= \left[\frac{e_1}{e_1-e_2}\right]\frac{(e_1-e_2)\,\mu_1-(1-e_2)}{\mu_1\,(e_1-1)}
\end{aligned}$$

## Part 3

Show that, in the simple case in which $U_1^a = U_2^b = c$, $U_2^a = U_1^b = 0$ and $\mu_1 = 0.5$ the probability of choosing the correct act in each state is given by $\frac{exp\left(\frac{c}{k}\right)}{1+exp\left(\frac{c}{k}\right)}$

We need to find expressions for $\pi_1(t)$ and $\pi_2(s)$.

Instead using the expression derived in the previous part (though it should give you the same results), note that using $U_1^a = U_2^b = c$, $U_2^a = U_1^b = 0$ and $\mu_1 = 0.5$, we get that

$$exp\left(\frac{U_1^a - U_1^b}{\kappa}\right) = exp\left(\frac{c}{k}\right)$$

$$exp\left(\frac{U_2^a - U_2^b}{\kappa}\right) = exp\left(-\frac{c}{k}\right) = \frac{1}{exp\left(\frac{c}{k}\right)}$$

furthermore, and for easiness in the notation let $e \equiv exp\left(\frac{c}{k}\right)$, then from equations (16-20) we get that

$$\gamma_1^t = \frac{e}{1+e} \tag{21}$$

$$\gamma_1^s = \frac{1}{1+e} \tag{22}$$

$$\gamma_2^t = \frac{1}{1+e} \tag{23}$$

$$\gamma_2^s = \frac{e}{1+e} \tag{24}$$

$$P(t) = \frac{1}{2} \tag{25}$$

Then

$$\pi_1(t) = \frac{\gamma_1^t P(t)}{\mu_1} = \frac{\frac{e}{1+e}\frac{1}{2}}{\frac{1}{2}} = \frac{e}{1+e}$$

$$\pi_2(s) = \frac{\gamma_2^s P(s)}{\mu_2} = \frac{\gamma_2^s(1-P(t))}{1-\mu_1} = \frac{\frac{e}{1+e}\frac{1}{2}}{\frac{1}{2}} = \frac{e}{1+e}$$

NOTE: If you substitute in the result of the previous part just note that $e_1 = e$ and $e_2 = \frac{1}{e}$, then

$$\begin{aligned}
\pi_1(t) &= \left[\frac{e_1}{e_1-e_2}\right]\frac{(e_1-e_2)\,\mu_1-(1-e_2)}{\mu_1\,(e_1-1)} \\
&= \left[\frac{e}{e-\frac{1}{e}}\right]\frac{\left(e-\frac{1}{e}\right)\frac{1}{2}-\left(1-\frac{1}{e}\right)}{\frac{1}{2}e-1} \\
&= \left[\frac{e^2}{e^2-1}\right]\frac{\frac{e^2-1}{e}-2\left(\frac{e-1}{e}\right)}{e-1} \\
&= \left[\frac{e}{e^2-1}\right]\frac{e^2-1-2\,(e-1)}{e-1} \\
&= \left[\frac{e}{e^2-1}\right](e+1-2) \\
&= \left[\frac{e}{e^2-1}\right](e-1) \\
&= \left[\frac{e}{e+1}\right]
\end{aligned}$$

## Part 4

In a recent experiment, I recorded the fraction of correct responses in each state for four different levels of reward. The results of the experiment are given in the following table assume $U(\$x) = x$. Is this data

| Reward (\$) | % Correct |
|---|---|
| 2 | 74.8 |
| 10 | 81.9 |
| 20 | 83.3 |
| 30 | 86.7 |

consistent with your findings from section 3 (i.e.) can the same $\kappa$ explain behavior at the 4 different reward levels?

If we assume the model is true, then the implicit $\kappa$, given the data is given by (let $p$ be the percentage of correct responses ($\pi_1(t)$ and $\pi_2(s)$))

$$p = \frac{exp\left(\frac{c}{\kappa}\right)}{1 + exp\left(\frac{c}{\kappa}\right)} \Leftrightarrow exp\left(\frac{c}{\kappa}\right) = \frac{p}{1-p} \Leftrightarrow \kappa = \frac{c}{\ln\left(\frac{p}{1-p}\right)}$$

Then the implicit $\kappa$ for the different rewards are given by Clearly, the answer is no, the same $\kappa$ cannot

| Reward ($) | % Correct |
|---|---|
| 2 | 1.8383 |
| 10 | 6.6243 |
| 20 | 12.4452 |
| 30 | 16.0026 |

explain this behavior.

### Part 5

If not, are my subjects increasing their attention as rewards increase more quickly or more slowly than the Shannon model predicts?

Given that the implicit $\kappa$ is increasing with respect to the rewards, the subjects seem to be increasing attention in a slower fashion that the model predicts. Alternatively you can use the $\kappa$ implied by the first result and see which one should be the predicted probability of choosing the correct act, then you can also see that the predicted probability is consistently above the observed one.

## Question 2

So far in class we have considered models which have, at their heart, some form of optimizing behavior An alternative is the fast and frugal heuristics approach, described here: http://www.cs.utah.edu/~miriah/uncertainty/Fast-and-frugal.pdf . Read this article and write a 2-3 page essay contrasting these two different approaches Focus particularly on when (if ever) you think either approach would be applicable, the usefulness of the two approaches as modelling strategies (for example in their ability to make out of sample predictions) and how you might go about testing for which of the two schools of models does the best job in a given situation.

**Fast and Frugal vs. optimizing**.

Gigerenzer and Todd (GT) formulate a research program that consists mainly on the study of a different alternative to optimization models of choice in the form of "heuristics" or approximations to real human choice procedures that are limited in time, attention and cognitive capabilities. In this text, I will contrast their research program with the optimization methodology for consumer behavior modelling. It is my belief and luckily of many others that a research program, as the one proposed by GT, has to be judged in the basis of two main criteria: (i) Falsifiability, and prediction power . (ii) Generality and internal consistency [Popper(2005)][Friedman(1953)].

The GT approach is defined in a very informal way by the authors and in principle is not totally distinguishable from the optimization approach by reason I will explain in the sequel. By means of an example, I will try to underline the main advantages and disadvantages of each approach regarding falsifiability and prediction power. Think about the satisficing model. From a fast and frugal perspective one can imagine a rule of thumb that says that the consumer will stop searching some choice set after $k$ iterations, visiting each object with a random order. She picks some object only if it is good enough in the basis of some criterion. Now think of the satisficing model as an optimal stopping time problem. There is an optimal number of iterations such that $k^* = argmax_{k \in T} V(k, A)$ and $V(k, A)$ is a general function that maps the choice set $A$ and $k$ to some utility value and $T$ is the constraint set that captures the fact that people have limited time. The main insight about this second model is that $k^*$ is the result of optimizing behavior. Now think of a dataset that contains information about the choices of a consumer, you observe the number of iterations $\overline{k}$ and the object the consumer picks $\overline{a}$. The consumer seems to be a satisficer. In the GT model one can fix $k = \overline{k}$ and one can say that $\overline{a}$ fulfills certain criterion by observing which properties has $\overline{a}$ that other objects in $A$ do not. We know that with more data we can recover the value function $V(k, A)$ and the associated utility $u$ and threshold $u^*$ such that $\overline{k} = argmax_{k \in T} \overline{V}(k, A)$ and $u(\overline{a}) > u^*$. In this situation, we can see that the two models "fit" the data perfectly. In fact, we can calibrate our models to data such that no error remains. When the models are equally good at fitting consumer data, we can use another criterion like simplicity. Clearly, the fast and frugal is more simple and requires less data. But the real test for an economic theory is out of the sample prediction. Let's imagine that I change the choice set $B$ such that $B \neq A$. How will the models perform? The fast and frugal model has the same $k = \overline{k}$ irrespective of whether the size or cardinality of $|B| > \overline{k}$, but even with a slight adaptation where you use the proportion $\kappa = \frac{\overline{k}}{|A|}$ such that $\tilde{k} = \kappa |B|$ we can easily see that "out-of-the-sample" predictions become hard due to the simplistic nature of the GT heuristics and its lack of adaptability. However, this itself is not a problem for the GT approach if data shows that the $\kappa$ is constant across choice sets for the consumer, in fact, this is a testable prediction of such a fast and frugal heuristic even if it is non-sensible. The main issue lies with the fuzzy definition of what is "good enough", or what has $\overline{a}$ that other objects do not have. Even if we change the decision procedure to have something like a decision tree it is not difficult to think of situations where our decision three fails to give us any answer when faced with new situations. To fix ideas, think of buying a car and then following decision tree. Buy a car only if the answer to the following three questions is yes: (i) Is the car price less than 100 thousand USD? (ii) Is the car red? (iii) Is the car electric?. This seems reasonable but it raises questions about heterogeneity and about adaptation to situations where the decision tree is not suited for. Think for instance on a situation where you must buy a car but electric cars are discontinued and only solar cars are available, what will this consumer do? The GT heuristic has nothing to say about his situations. Is the optimizing model any better? Well clearly if we identify the function $V$ and $u$ and $u^*$ for the set $A$ our only option is to interpolate and imagine that we can still use our recovered functions to infer from the new situation. This has the requirement that the consumer has stable behavior. This is no small requirement and actually is also required in the GT approach. But after this leap of faith we can actually make predictions that can either succeed or fail, in fact we can predict a new $\hat{k} = argmax_{k \in T} \overline{V}(k, B)$ and $u(\bar{b}) > u^*$. The fact that the optimizing theory can always predict out of the sample is a clear advantage of this approach at least from the point of view of applicability but it must be underlined that there is no a priori reason for this prediction to be accurate. Again this is an empirical matter. The main message of this comparison is that GT heuristics can be as good as optimizing models to explain some dataset and can do this with much more simplicity. In summary, the heuristic approach suffers from a general lack of flexibility to adapt to new situations which limits its applicability and falsifiability due to its fuzziness in how to update the fast and frugal procedures in the light of a new situation (at least as postulated by the authors).

The second criteria that helps to compare two competing theories is that of internal consistency and generality. From the point of view of generality it is interesting to notice that optimizing behavior can contain the GT heuristics approach in principle. To make my point notice that an optimization setting with time constraints, uncertainty and limited attention can generate fast and frugal procedures. The idea of having a fast and frugal action derived from a "run once" optimization has been applied in the work of Gabaix [Gabaix(2012)]. He postulates a two steps optimization procedure when the first one is done once to generate a precompiled attention to price changes that can be used whenever the consumer maximizes its utility given its "perceived" prices. In this case, the first optimization gives the consumer a fast and frugal rule: pay attention to price changes over a threshold $\tau$ that is fixed while maximizing utility in a second step. Think of the hospital decision tree for heart attack patients, it is clear that this "simple" tree is the result of past experience, learning and of course of a will of doctors to reduce time, increase efficiency and the likelihood to identify risk patients. The aforementioned decision tree can yield fast actions but clearly was not designed frugally nor rapidly. In contrast, The fast and frugal heuristics cannot explain optimizing behavior by construction, they reject the idea of optimization. However, it is interesting to notice that fast and frugal heuristic may produce "optimization" like behavior. The fact that zero intelligence systems behave "as if" optimizing is not new. This is the main idea behind the concepts of Evolutionary Nash Equilibrium so pervasive in biology. To understand this, think of the following though experiment. A group of cavemen are trapped in a valley that has fruits never before seen by any of them. There are not animals only trees that yield fruit. How will the cavemen choose? A reasonable simple procedure is to pick at random and test the fruit and eat if I do not get poisoned and if the fruit is tasty. Also let the cavemen communicate the flavor and properties of the "tested" fruits to others. The cavemen will copy those that report that the fruit is tasty and that are not poisoned. The random procedure is similar to "mutation" and the copying is similar to "selection". In fact, it is easy to see that this cavemen community will eventually converge and consume the most tasty fruit that does not harm them. If an observer reaches the island after they have converged they won't be able to tell the difference between optimizers and simple agents with "try and copy" simple rule. Note that variation or mutation together with selection can find the optimal solution to many problems even if they do not maximize willingly. The consumer act "as if" they are maximizing. In summary, in terms of generality the optimization approach seems to beat the GT heuristics since one can always think of a maximization procedure that generates behavior that can be seen as fast and frugal. Finally, the internal consistency becomes the remaining criterion and this seems to be a weakness of the fast and frugal heuristics, in fact, the heuristics name reveals that its a collection of ideas that seem to work but may not be true. Its sole power is to explain or fit data and hopefully to predict behavior but they are not a formal system. The optimizing methodology in contrast has matured to become deeply connected with an axiomatic system, where we can derive all our predictions from a finite set of "axioms" using a language and certain operations. Clearly, not all optimization models are axiomatized but most of them are or can be. This is interesting from the formal point of view since making more clear cut and precise predictions that facilitates falsifiability of the theory and progress towards explaining possible paradoxes of behavior. Moreover, we can assess the empirical content of theories and to judge them in empirical terms more easily. The formalization of a theory is important but is something that the GT heuristics lack right now but can be solved in the future.

In conclusion, the simplicity of GT heuristics are its main attractive and its main distinction with respect to the optimization approach. However, in its current form it lacks flexibility, adaptability and due to its fuzzy definition it lacks clear cut predictions that make difficult to test its performance in out-of-the-sample prediction. Finally, the lack of a formal structure and generality can diminished its interest. Having said that, there is a potential avenue of future research thinking of two-systems approaches that mixed the best of both worlds. Using pre-compiled "run once" optimization procedures that generate fast and frugal procedures and combine this with a second step optimization of a simplified problem.

## References

[Friedman(1953)] FRIEDMAN, M. (1953). The methodology of positive economics. *Essays in positive economics*, **3** (8).

[Gabaix(2012)] GABAIX, X. (2012). A sparsity-based model of bounded rationality, applied to basic consumer and equilibrium theory. *manuscript, New York University*.

[Popper(2005)] POPPER, K. (2005). *The logic of scientific discovery*. Routledge.