Approximating a Continuous Utility Function over a Set of Lotteries*

Konrad Grabiszewski†

This Version: January 2012

Abstract

We consider the problem of approximating a continuous utility function, V, defined over the set of lotteries, $P(X)$, where the set of prizes, X, is a finite set or a compact metric space. To do this, we derive a functional form of an approximating function, $\hat{V}$. If h is a real-valued, continuous, and injective function on X, then $\hat{V}$ is a linear combination of the raw moments of h and their powers. Since we consider a global approximation over the whole set of lotteries, our result complements the literature focusing on "local approximation" initiated by [11].

Keywords: Expected Utility Theory; Approximating Utility Function

JEL classification: D80; D81

*I thank Adam Brandenburger, David Pearce, Ben Polak, David Ross and Ennio Stacchetti for their valuable comments. Support from the Asociación Mexicana de Cultura A.C. and NYU's Stern School of Business is gratefully acknowledged.

†Address: Department of Business Administration, ITAM, Av. Camino a Sta. Teresa No. 930, Col. Magdalena Contreras, C.P. 10700 Mexico D.F., Mexico, konrad.grabiszewski@itam.mx, http://allman.rhon.itam.mx/~konrad.grabiszewski/

1 Introduction

Let X be the set of prizes and $P(X)$ the collection of lotteries over X. Assume that the agent's utility function, V, defined over $P(X)$ is continuous. The functional form of V is not known and, as a result, it is not possible to estimate the agent's utility function. In such cases, this paper suggests an approximation of V by $\hat{V}$ in which the functional form of $\hat{V}$ is known and can be estimated. As such, the approximation method employed in this paper is a useful tool for analyzing the agent's behavior.

This paper focuses on a global approximation of V—that is, an approximation over the whole set, $P(X)$. We endow $P(X)$ with the weak-* topology. The objective is to approximate V using a linear combination of functions that we construct. (Later we discuss how to construct such functions.) With a given set of such real-valued functions, $\{f_1, ..., f_n, ...\}$, which are defined on $P(X)$, an approximation of V is $\hat{V} = a_0 + a_1 f_1 + ... + a_n f_n$. Knowing the agent's choices, it is possible to use this functional form to estimate the parameters a_0,...,a_n.

As an example of our approach, consider a two-element set, $X = \{x, y\}$. A generic element of $P(X)$ is a pair, (p, q), where p and q are weights assigned to x and y, respectively. Since $p + q = 1$, $P(X)$ can be identified with the segment, $[0, 1]$. Let V be a continuous utility function over $P(X)$. Since $P(X)$ is equivalent to $[0, 1]$, we define V as a function of one variable, p. The Weierstrass Approximation Theorem tells us that there exists an integer n, such that V is approximated by a polynomial, $\hat{V}(p) = a_0 + a_1 p + ... + a_n p^n$. However, we do not want to express $\hat{V}$ in terms of p since, for X of cardinality larger than 2, probability is not a scalar and, in consequence, $\hat{V}(p)$ could not be a function with values in $\mathbb{R}$. Alternatively we can express $\hat{V}$ in terms of moments of some function h. Let h be a real-valued function on X, such that $h(x) > h(y)$. The expected value of h under p is $E(h, p) = h(x)p + h(y)(1 - p)$. This yields a formula for p, $p = \frac{E(h,p)-h(y)}{h(x)-h(y)}$, which we substitute into $\hat{V}(p)$ in order to obtain $\hat{V}(p) = b_0 + b_1 E(h, p) + ... + b_n E^n(h, p)$. In short, we can conclude that a continuous utility function on $P(X)$ for a two-element X is approximately an n-degree polynomial whose variable is the expected value of h.

This paper analyzes the global approximation of a continuous utility function defined over the set of lotteries. As such, this analysis complements the literature that focuses on local approximation. This literature begins with [11], who shows that a Fréchet differentiable utility function on $P(X)$ behaves locally in accordance with the expected utility hypothesis. Key later developments were contributed by [1], [6], [4], [12], [7], [15], [5], and [3].

2 Main result

Let Y be a metric space and $C(Y)$ the collection of real-valued, continuous, and bounded functions defined over Y. Set $C(Y)$ is endowed with the topology of uniform convergence. That is, a sequence of functions in

$C(Y)$, $\{V_n\}$, converges to a function in $C(Y)$, V, if and only if $\sup_{y\in Y}|V_n(y)-V(y)|$ goes to zero, as n goes to infinity. We say that $\hat{V}\in C(Y)$ approximates $V\in C(Y)$ with ε degree of accuracy if $\sup_{y\in Y}|\hat{V}(y)-V(y)|<\varepsilon$. The existence of such a $\hat{V}$ is guaranteed by the Stone-Weierstrass Approximation Theorem (see, for instance, Theorem 34 in [14]), which requires that (a) the set Y be a compact space, and (b) there be algebra in $C(Y)$ that contains constant functions and separates points in Y.

In our case, condition (a) is easily satisfied. Note that Y is the set of lotteries on X, $P(X)$. It is a well known fact in the probability theory that if X is a compact metric space and $P(X)$ is endowed with the weak-* topology, then $P(X)$ is also a compact metric space (see, for instance, Theorem 6.4 in [13]). Hence, we require that the underlying set X be a compact metric space.

As for condition (b), suppose that there exists a real-valued, continuous, and injective function, $h:X\to\mathbb{R}$. (Later, we will discuss what restrictions on X the existence of h imposes.) Fix that function, and taking a positive integer k, let $E(h^k,p)$ denote the kth raw moment of h under $p\in P(X)$. In other words, $E(h^k,p)$ is a real-valued function defined on $P(X)$ as $E(h^k,p):=\int_X h^k dp$. Since we can construct h and its raw moments, our objective is to express $\hat{V}$, an approximation of V, as a linear combination of $E(h^k,p)$'s and powers of $E(h^k,p)$'s. Let Ω be a countable set consisting of functions $E(h^k,p)$'s. In the case of finite X with cardinality m, Ω must contain the first $m-1$ raw moments—that is, $\Omega=\{E(h,p),...,E(h^{m-1},p)\}$. However, in the case of infinite X, we need all raw moments—that is, $\Omega=\{E(h,p),...,E(h^{m-1},p),...\}$. We say that ψ is a polynomial generated by Ω if ψ is a function defined on $P(X)$ and if there is a positive integer, r, such that $\{E^{i_1}(h,p),...,E(h^{i_r},p)\}$ is a subset of Ω and ψ has a functional form, $\psi=a_0+\sum_{k=1}^{r}(a_{1,k}f_{i_k}^k+...+a_{r,k}f_{k_r}^k)$, where a_0,$a_{1,1}$,..., $a_{r,r}$ are the parameters belonging to real line. Let $\Psi(\Omega)$ be the set of all polynomials generated by Ω. As the Main Result shows, $\Psi(\Omega)$ is dense in $C(P(X))$.

In the Main Result that follows, we consider two cases: finite X and infinite X. The results and proofs differ between these cases. If X is finite, then we deduce that required Ω is finite, and our proof relies on linear algebra. If X is infinite, then Ω must be infinite, and use probability theory to prove the result.

It is necessary to discuss the limits imposed by the existence of the real-valued, continuous, and injective function, h, defined on X. If X is countable, then, a from topological perspective, X is a subset of a real line, and consequently, there must be an injective function, $h:X\to\mathbb{R}$. However, if X is uncountable, then imposing the existence of h implies specific topological restrictions on X. Note that X and $h(X)$ are homeomorphic (see, for instance, Proposition 4.1 in [10]). Since $h(X)$ is a subset of $\mathbb{R}$, not every X will satisfy our demands. What we need is X with a topological dimension that is the same as, or smaller than, the topological dimension of $\mathbb{R}$. Here it is important to note that topological dimension, also known as covering dimension, is the fundamental concept of dimension theory (see, for instance, Definition 1.6.7 in [8]). We say that a topological space X has topological dimension n if every finite open cover of X has a finite refinement of order that is no more than n. The order of the set of subsets is the largest number n, such that $n+1$ of its elements have non-empty intersection. For example, the empty set has dimension -1, any countable set has dimension 0, and $\mathbb{R}^n$ has dimension n (see, for instance, Theorem 1.8.3 in [8]). If X has dimension n, then every closed subset of X has a dimension of, at most, n (see, for instance, Theorem 3.1.3 in [8]). In our case, this means that $h(X)$ can have dimensions -1, 0, or 1. If two spaces are homeomorphic, then they have the same dimension. In our case, this means that the dimension of X will be, at most, 1. Hence, X must be either countable or similar to $[0, 1]$. If X were, for instance, $[0, 1] \times [0, 1]$, then its dimension would be 2 (see, for instance, Corollary 1.8.4 in [8]). In that case, the approximation method pursued in this paper would not apply.

Main Result

1. ***Finite case:*** *Let X be a finite set with cardinality larger than 1. Let V be a continuous utility function on $P(X)$. Let h be a real-valued injective function defined on X. Let $\Omega = \{E(h, p), ..., E(h^{m-1}, p)\}$. Then, for each $\varepsilon > 0$, there exists $\psi \in \Psi(\Omega)$, such that $\sup_{p\in P(X)} |\psi(p) - V(p)| < \varepsilon$.*

2. ***Infinite case:*** *Let X be a compact metric space. Let V be a continuous utility function on $P(X)$. Assume that there exists a real-valued, continuous, and injective function h defined on X. Let $\Omega = \{E(h, p), ..., E(h^{m-1}, p), ...\}$. Then, for each $\varepsilon > 0$, there exists $\psi \in \Psi(\Omega)$, such that $\sup_{p\in P(X)} |\psi(p) - V(p)| < \varepsilon$.*

To prove the Main Result, first note that $\Psi(\Omega)$ contains constant functions and is an algebra: If $a, b \in \mathbb{R}$ and $\psi_1, \psi_2 \in \Psi(\Omega)$, then $a\psi_1 + b\psi_2 \in \Psi(\Omega)$ and $\psi_1\psi_2 \in \Psi(\Omega)$. Next we prove that $\Psi(\Omega)$ separates points in $P(X)$. We will consider two cases: finite X (Lemma 2.1) and infinite X (Lemma 2.2).

Lemma 2.1.

Let X be a finite set with cardinality larger than 1. Let h be a real-valued injective function defined on X. Then, for distinct $p, q \in P(X)$, there exists $k \in \{1, ..., m - 1\}$, such that $\sum_{i=1}^m h^k(x_i)p_i \neq \sum_{i=1}^m h^k(x_i)q_i$.

Lemma 2.2.

Let X be a compact metric space. Assume that there exists a real-valued, continuous, and injective h defined on X. Then, for distinct $p, q \in P(X)$, there exists a positive integer k, such that $\int_X h^k dp \neq \int_X h^k dq$.

Proof of Lemma 2.1:

For simplicity, let h_k denote $h(x_k)$. Without losing generality, assume that $h_1 > h_2 > ... > h_m$. Then take two probability measures on X, p and q, such that $p \neq q$. Suppose that, for each $k = 1, ..., m - 1$, it is true that $\sum_{i=1}^m h^k(x_i)p_i = \sum_{i=1}^m h^k(x_i)q_i$. Then the following must be true.

$$\begin{cases} (p_1 - q_1)h_1 + ... + (p_m - q_m)h_m = 0 \\ \vdots \\ (p_1 - q_1)h_1^{m-1} + ... + (p_m - q_m)h_m^{m-1} = 0 \end{cases} \tag{1}$$

Next, we re-write (1) to reflect the fact that $p_m = 1 - p_1 - ... - p_{m-1}$ and $q_m = 1 - q_1 - ... - q_{m-1}$.

$$\begin{cases} (p_1 - q_1)(h_1 - h_m) + ...(p_{m-1} - q_{m-1})(h_{m-1} - h_m) = 0 \\ \vdots \\ (p_1 - q_1)(h_1^{m-1} - h_m^{m-1}) + ...(p_{m-1} - q_{m-1})(h_{m-1}^{m-1} - h_m^{m-1}) = 0 \end{cases} \tag{2}$$

We define two matrices,

$$y_m := \begin{pmatrix} p_1 - q_1 \\ \vdots \\ p_{m-1} - q_{m-1} \end{pmatrix} \quad \text{and} \quad A_m := \begin{pmatrix} h_1 - h_m & \cdots & h_{m-1} - h_m \\ \vdots & & \vdots \\ h_1^{m-1} - h_m^{m-1} & \cdots & h_{m-1}^{m-1} - h_m^{m-1} \end{pmatrix},$$

which allow us express (2) as a matrix system, $A_m y_m = 0$. According to Cramer's Rule (see, for instance, Theorem 4.9 in [9]), if the determinant of A_m is non-zero, then there exists the unique solution of $A_m y_m = 0$. In our case, $y_m = 0$ would be this solution. However, if $y_m = 0$, then $p = q$, which contradicts the initial assumption. Hence, it remains to be proven that A_m is an invertible matrix. The proof, by induction, is presented next.

Take $m = 2$, and note that A_2 is invertible as $A_2 = h_1 - h_2$, due to the injectivity of h, is non-zero. Assume that A_{m-1} is invertible. We need to prove that the determinant of A_m is non-zero.

Using the Newton's binomial formula, $h_i^l - h_j^l = (h_i - h_j)(h_i^{l-1} + h_i^{l-2} h_j + ... + h_i h_j^{l-2} + h_j^{l-1})$, we can rewrite A_m as a product of two matrices, $A_m = B_m C_m$.

$$B_m \quad := \begin{pmatrix} 1 & \cdots & 1 \\ h_1 + h_m & \cdots & h_{m-1} + h_m \\ \vdots & & \vdots \\ h_1^{m-3} + h_1^{m-1} h_m + ... + h_1 h_m^{m-2} + h_m^{m-3} & \cdots & h_{m-1}^{m-3} + h_{m-1}^{m-2} h_m + ... + h_{m-1} h_m^{m-2} + h_m^{m-3} \\ h_1^{m-2} + h_1^{m-3} h_m + ... + h_1 h_m^{m-3} + h_m^{m-2} & \cdots & h_{m-1}^{m-2} + h_{m-1}^{m-3} h_m + ... + h_{m-1} h_m^{m-3} + h_m^{m-2} \end{pmatrix}$$

$$
C_m \quad := \quad \begin{pmatrix} h_1 - h_m & 0 & \cdots & 0 & 0 \\ 0 & h_2 - h_m & \cdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & h_{m-2} - h_{m-1} & 0 \\ 0 & \cdots & \cdots & 0 & h_{m-1} - h_m \end{pmatrix}
$$

The determinant of A_m is the product of the determinants of B_m and C_m. The determinant of C_m is the product of its diagonal elements. Since h is injective, the determinant of C_m is non-zero, though, it remains to be shown that the determinant of B_m is also non-zero. Note that B_m is an $(m-1) \times (m-1)$ matrix, and recall that adding a multiple of a row (or column) to another row (or column) does not change the determinant. Then consider the following sequence of manipulations of B_m.

1. Multiply row $(m-2)$ by h_m and subtract it from row $(m-1)$. In column 1, we obtain $h_1^{m-2} + h_1^{m-3} h_m + ... + h_1 h_m^{m-3} + h_m^{m-2} - (h_1^{m-3} + h_1^{m-1} h_m + ... + h_1 h_m^{m-2} + h_m^{m-3}) h_m = h_1^{m-2} + h_1^{m-3} h_m + ... + h_1 h_m^{m-3} + h_m^{m-2} - h_1^{m-3} h_m - h_1^{m-1} h_m^2 - ... - h_1 h_m^{m-1} - h_m^{m-2} = h_1^{m-2}$. In column l, we have h_l^{m-2}.

2. Multiply row $(m-3)$ by h_m and subtract it from row $(m-2)$. If we follow the same analysis as above, in column l, we have h_l^{m-3}.

3. Continue this procedure until we subtract row 1, multiplied by h_m, from row 2. The result is the following matrix.

$$
\hat{B}_m \quad := \quad \begin{pmatrix} 1 & \cdots & 1 & 1 \\ h_1 & \cdots & h_{m-2} & h_{m-1} \\ \vdots & & \vdots & \\ h_1^{m-2} & \cdots & h_{m-2}^{m-2} & h_{m-1}^{m-2} \end{pmatrix}
$$

4. In matrix $\hat{B}_m$, subtract the last column from each of the other columns. The result is the following

matrix.

$$\tilde{B}_m \quad := \quad \begin{pmatrix} 0 & \cdots & 0 & 1 \\ h_1 - h_{m-1} & \cdots & h_{m-2} - h_{m-1} & h_{m-1} \\ \vdots & & \vdots & \\ h_1^{m-2} - h_{m-1} & \cdots & h_{m-2}^{m-2} - h_{m-1} & h_{m-1}^{m-2} \end{pmatrix}$$

Note that the first row of matrix $\tilde{B}_m$ consists of zeros but in the last column there is a 1. Hence, using the cofactor expansion along the first row, we find that the determinant of $\tilde{B}_m$ is non-zero if the determinant of the following matrix is also non-zero.

$$\bar{B}_m \quad := \quad \begin{pmatrix} h_1 - h_{m-1} & \cdots & h_{m-2} - h_{m-1} \\ \vdots & & \vdots \\ h_1^{m-2} - h_{m-1} & \cdots & h_{m-2}^{m-2} - h_{m-1} \end{pmatrix}$$

Note, however, that matrix $\bar{B}_m$ is the same as matrix A_{m-1}. By assumption, the determinant of A_{m-1} is non-zero. Consequently, the determinants of $\bar{B}_m$, $\tilde{B}_m$, $\hat{B}_m$, and B_m are non-zero. This implies that the determinant of A_m is also non-zero. ■

Proof of Lemma 2.2:

Let $\phi : X \to \mathbb{R}$ be defined as a polynomial determined by h, $\phi(x) := a_0 + a_1 h(x) + ... + a_n h^n(x)$. Let $\Phi(X)$ be the collection of all such polynomials. $\Phi(X)$ is algebra, separates points and includes constants. Hence, by the Stone-Weierstrass Approximation Theorem, $\Phi(X)$ is dense in the set of all real-valued continuous functions that are defined over X, $C(X)$. Let $T_p : C(X) \to \mathbb{R}$ be defined as $T_p(f) := \int_X f dp$, noting that T_p is a continuous function. Since we know that a continuous function is determined by its restriction on the dense subset of its domain, each T_p is determined by its restriction on $\Phi(X)$. We know (see, for instance, Theorem 1.2 in [2]) that if $T_p(f) = \int_X f dp = \int_X f dq = T_q(f)\ \forall f \in C(X)$, then $p = q$. However, since $\Phi(X)$ is dense in $C(X)$, it is enough to take only the polynomials from $\Phi(X)$. That is, if $\int_X \phi dp = \int_X \phi dq$ for each $\phi \in \Phi(X)$, then $T_p = T_q$ and, consequently, $p = q$. Take distinct $p, q \in P(X)$. There must be ϕ, such that $\int_X \phi dp \neq \int_X \phi dq$ and, consequently, there must be h^k, such that $\int_X h^k dp \neq \int_X h^k dq$. Otherwise $\int_X \phi dp = \int_X \phi dq$. ■

References

[1] Beth Allen. Smooth preferences and the approximate expected utility hypothesis. *Journal of Economic Theory*, 41:340–355, 1987.

[2] Patrick Billingsley. *Convergence of Probability Measures.* Wiley, 1999.

[3] Kalyan Chatterjee and R. Vijay Krishna. On preferences with infinitely many subjective states. *Economic Theory*, 46:85–98, 2011.

[4] Soon Hong Chew, Larry G. Epstein, and Itzhak Zilcha. A correspondence theorem between expected utility and smooth utility. *Journal of Economic Theory*, 46:186–193, 1988.

[5] Soon Hong Chew and Mao Mei Hui. A Schur concave characterization of risk aversion for non-expected utility preferences. *Journal of Economic Theory*, 67:402–435, 1995.

[6] Soon Hong Chew, Edi Karni, and Zvi Safra. Risk aversion in the theory of expected utility with rank dependent probabilities. *Journal of Economic Theory*, 42:370–381, 1987.

[7] Soon Hong Chew and Naoko Nishimura. Differentiability, comparative statics, and non-expected utility preferences. *Journal of Economic Theory*, 56:294–312, 1992.

[8] Ryszard Engelking. *Theory of Dimensions Finite and Infinite.* Haldermann Verlag, 1995.

[9] Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence. *Linear Algebra.* Prentice Hall, 2003.

[10] A. S. Kechris. *Classical Descriptive Set Theory.* Springer-Verlag, 1995.

[11] Mark J. Machina. "Expected utility" analysis without the independence axiom. *Econometrica*, 50:277–323, 1982.

[12] Mark J. Machina. Comparative statistics and non-expected utility preferences. *Journal of Economic Theory*, 47:393–405, 1989.

[13] K. R. Parthasarathy. *Probability Measures on Metric Spaces.* Academic Press, 1967.

[14] H.L. Royden. *Real Analysis.* Prentice-Hall, 1988.

[15] Tan Wang. l_p-Fréchet differentiable preference and "local utility" analysis. *Journal of Economic Theory*, 61:139–159, 1993.