

Two Uses for Measurable Selections

KCB
2012
v. 2017.09.22::16.58

1 Stochastic Taylor's Theorem

Taylor's Theorem is standard, see, e.g., [4, Theorem 10, p. 179].

1 Taylor's Theorem *Let $h: [a, b] \rightarrow \mathbf{R}$ be continuous and possess a continuous n^{th} -order derivative on (a, b) . Fix $c \in [a, b]$ and let $c + x$ belong to $[a, b]$. Then there is a point ξ satisfying $\xi \in [0, x]$ (where $[0, x]$ is the line segment joining 0 and x , regardless of the sign of x), and*

$$h(c + x) = h(c) + \sum_{k=1}^{n-1} \frac{1}{k!} h^{(k)}(c) x^k + \frac{1}{n!} h^{(n)}(c + \xi) x^n.$$

The next result seems obvious, but I don't know an easy proof. It is implicitly assumed by Pratt [5] in his analysis of risk aversion.

2 Stochastic Taylor's Theorem *Let $h: [a, b] \rightarrow \mathbf{R}$ be continuous and possess a continuous n^{th} -order derivative on (a, b) . Fix $c \in [a, b]$ and let X be a random variable on the probability space (S, \mathcal{S}, P) such that $c + X(s)$ belongs to $[a, b]$ for all s . Then there is a measurable function ξ satisfying $\xi(s) \in [0, X(s)]$ for all s (where $[0, X(s)]$ is the line segment joining 0 and $X(s)$, regardless of the sign of $X(s)$), and*

$$h(c + X(s)) = h(c) + \sum_{k=1}^{n-1} \frac{1}{k!} h^{(k)}(c) X^k(s) + \frac{1}{n!} h^{(n)}(c + \xi(s)) X^n(s).$$

Proof: Taylor's Theorem asserts that there is such a $\xi(s)$ for each s , the trick is to show that there is a measurable version. To this end define the correspondence $\varphi: S \rightarrow \mathbf{R}$ by $\varphi(s) = [0, X(s)]$. It follows from [1, Theorem 14.78, p. 501] that φ is measurable and it clearly has compact

values. Set $g(s) = h(c + X(s)) - h(c) - \sum_{k=1}^{n-1} \frac{1}{k!} h^{(k)}(c) X^k(s)$, $f(s, x) = \frac{1}{n!} h^{(n)}(c + x) X^n(s)$. Then g is measurable and f is a Carathéodory function. (See sections 14.12–14.14 of [1] for the definitions of measurable correspondences and Carathéodory functions.) By Filipov's Implicit Function Theorem [1, Theorem 14.89, p.507] there is a measurable function ξ such that for all s , $\xi(s) \in \varphi(s)$ and $f(s, \xi(s)) = g(s)$, and we are done. ■

2 Spaces of Probability Measures

Let $f: X \rightarrow Y$ be a measurable mapping between measurable spaces (X, Σ_X) and (Y, Σ_Y) . Then a probability P on (X, Σ_X) induces a probability $P \circ f^{-1}$ on (Y, Σ_Y) via

$$P \circ f^{-1}(A) = P(f^{-1}(A)).$$

The probability $P \circ f^{-1}$ is sometimes called the **distribution of f under P** . When X and Y are metrizable spaces and f is continuous, then we have the following result (cf. [1, Lemma 12.12, p. 419]). It has application to the construction of universal type spaces in a framework similar to Brandenburger and Dekel [2].

3 Theorem *Let X and Y be metrizable spaces, and let $f: X \rightarrow Y$ be continuous. Define $\hat{f}: \mathcal{P}(X) \rightarrow \mathcal{P}(Y)$ via $\hat{f}(P) = P \circ f^{-1}$. Then \hat{f} is continuous. If f is injective, then \hat{f} is injective. If f is surjective and Y is Polish, then \hat{f} is surjective. If f is a homeomorphism, then so is \hat{f} .*

Proof: We first note that if P is a regular Borel measure on X , then $P \circ f^{-1}$ is a regular Borel measure on Y . This is just [1, Theorem 11.22].

To see that \hat{f} is continuous, observe that if h is a bounded continuous real function on Y , then $h \circ f$ is a bounded continuous real function on X . Furthermore for any $P \in \mathcal{P}(X)$,

$$\int_X h \circ f(x) dP(x) = \int_Y f(y) d(P \circ f^{-1})(y) = \int_Y f(y) d(\hat{f}(P))(y).$$

(See, e.g., [1, Theorem 10.46, p. 365].) Thus if $P_n \rightarrow P$, it follows that $\hat{f}(P_n) \rightarrow \hat{f}(P)$.

To see that \hat{f} is injective, suppose P and Q belong to $\mathcal{P}(X)$ and $P \neq Q$. Since P and Q are regular, there is some compact subset K of X such that $P(K) \neq Q(K)$. Now $f(K)$ is a compact subset of Y and hence a Borel subset. Since f is injective, $K = f^{-1}(f(K))$. Thus $\hat{f}(P)(f(K)) = P(K) \neq Q(K) = \hat{f}(Q)(f(K))$, so $\hat{f}(P) \neq \hat{f}(Q)$, which shows that \hat{f} is injective.

Now suppose f is surjective and Y is Polish. Let Q be a Borel probability measure on Y , and let Σ_Q denote the collection of Q -measurable sets. (That is, if A belongs to Σ_Q , then there is a Borel subset B of Y such that $A\Delta B$ is included in a Borel set of Q -measure zero. See [1, Definition 8.50, p. 293ff].) Consider the correspondence $\psi : y \mapsto f^{-1}(y)$ from Y into X . Since f is continuous, the graph of ψ is closed in $Y \times X$, so a fortiori the graph belongs to $\mathcal{B}_Y \otimes \mathcal{B}_X \subset \Sigma_Q \otimes \mathcal{B}_X$. Since Y is Polish, by a measurable selection theorem [1, Corollary 14.87, p. 506] there is a selection g from ψ that is $(\Sigma_Q, \mathcal{B}_X)$ -measurable. The selection g acts like an inverse of f : for any y in Y , $f(g(y)) = y$, and for any subset A of X , $f^{-1}(g^{-1}(A)) = A$.

Define the Borel measure P on X by

$$P = Q \circ g^{-1}.$$

I claim that $\widehat{f}(P) = Q$. But first let us verify that P is a Borel measure. For any Borel subset B of X , $g^{-1}(B)$ belongs to Σ_Q , so $Q(g^{-1}(B))$ is defined, and it is easy to see that P is a probability measure. Now $\widehat{f}(P)$ is the Borel probability on Y defined by

$$\widehat{f}(P)(A) = P(f^{-1}(A)) = Q\left(g^{-1}(f^{-1}(A))\right) = Q(A)$$

whenever A is a Borel subset of Y . That is, $\widehat{f}(P) = Q$. Since Q was arbitrary, \widehat{f} is surjective.¹

Finally, assume f is a homeomorphism. Then \widehat{f} is a continuous bijection, and $f^{-1}: Y \rightarrow X$ induces a continuous bijection $\psi: \mathcal{P}(Y) \rightarrow \mathcal{P}(X)$. It suffices to show that $\psi = \widehat{f}^{-1}$. So let $Q = \widehat{f}(P)$, i.e., $Q = P \circ f^{-1}$. Then for any Borel subset A of X ,

$$\psi(Q)(A) = Q \circ f(A) = P \circ f^{-1}(f(A)) = P(A).$$

That is, $\psi = \widehat{f}^{-1}$. ■

References

- [1] C. D. Aliprantis and K. C. Border. 1994. *Infinite dimensional analysis: A hitchhiker's guide*. Number 4 in Studies in Economic Theory. Berlin: Springer-Verlag.

¹I thank Alekos Kechris for pointing out to me that Σ_Q measurability of g is sufficient to obtain this result. He also pointed out that there exists a selection g that is **universally measurable**. That is, g is Σ_Q measurable for every Borel probability measure Q simultaneously. This fact follows from [3, Theorem 18.1, p. 122].

- [2] A. Brandenburger and E. Dekel. 1993. Hierarchies of beliefs and common knowledge. *Journal of Economic Theory* 59(1):189–198.
DOI: [10.1006/jeth.1993.1012](https://doi.org/10.1006/jeth.1993.1012)
- [3] A. S. Kechris. 1995. *Classical descriptive set theory*. Number 156 in Graduate Texts in Mathematics. New York: Springer–Verlag.
- [4] J. E. Marsden. 1974. *Elementary classical analysis*. San Francisco: W. H. Freeman and Company.
- [5] J. W. Pratt. 1964. Risk aversion in the small and in the large. *Econometrica* 32(1/2):122–136. <http://www.jstor.org/stable/1913738>