

## Zur Analyse von Mobilitätsdaten

Blasius, Jörg

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

### Empfohlene Zitierung / Suggested Citation:

Blasius, J. (1995). Zur Analyse von Mobilitätsdaten. In H. Sahner, & S. Schwendtner (Hrsg.), *27. Kongreß der Deutschen Gesellschaft für Soziologie - Gesellschaften im Umbruch: Sektionen und Arbeitsgruppen* (S. 307-314). Opladen: Westdt. Verl. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-137287>

### Nutzungsbedingungen:

*Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.*

*Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.*

### Terms of use:

*This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.*

*By using this particular document, you accept the above-stated conditions of use.*

## **XII. Sektion Modellbildung und Simulation**

*Leitung: Ulrich Mueller*

### **Modelle der räumlichen und sozialen Mobilität**

#### **1. Zur Analyse von Mobilitätsdaten**

*Jörg Blasius*

##### *Einleitung*

Die Analyse von Mobilitätsdaten hat eine lange Tradition. Zu den wichtigsten Arbeiten gehört die von Blau und Duncan (1967) - die Beschreibung der beruflichen Mobilität in den Vereinigten Staaten mit Hilfe der multidimensionalen Skalierung (MDS). Dieser klassische Ansatz, mit dem (Un)-Ähnlichkeiten im Mobilitätsverhalten zwischen Objekten (z.B. Berufsgruppen) durch Distanzen in einem niederdimensionalen Raum dargestellt werden können, wurde von vielen Autoren übernommen; so auch von Best (1990) für seine Analyse der beruflichen Mobilität von Mitgliedern der Frankfurter Nationalversammlung (FNV). In getrennten Analysen beschreibt Best die inter- und die intragenerationale Mobilität der Mitglieder der FNV; als Ähnlichkeitsmaße für die MDS verwendete er den Dissimilaritätsindex und den Phi-Koeffizienten.

In einer Reanalyse dieser Daten wird die Korrespondenzanalyse verwendet. Der Vorteil dieses Verfahrens liegt zum einen darin, daß die Daten nicht vor der Analyse in Ähnlichkeitsmaße transformiert werden müssen - als Eingabedaten können die Häufigkeiten der Kontingenztafeln verwendet werden. Zum anderen können die inter- und die intragenerationale Mobilität innerhalb einer Analyse beschrieben werden. Sowohl bei der MDS als auch bei der Korrespondenzanalyse wird mit der Berufsstabilität bzw. mit der Berufsvererbung der größte Teil der Variation der Daten erklärt. Berufliche Veränderungen sind im Gegensatz zum Verbleiben im Beruf relativ selten und daher in der Auswertung erst mittels höherdimensionaler Lösungen interpretierbar. In diesem Papier wird ein Vorschlag gemacht, mit dem die berufliche Mobilität der Mitglieder der FNV beschrieben und grafisch dargestellt werden kann, ohne daß die Berufsstabilität bzw. die Berufsvererbung einen Einfluß auf die Ergebnisse hat.

##### *Daten*

Die FNV war das erste demokratische Parlament in Deutschland; es tagte vom Mai 1848 bis zum Mai 1849. Die Wahlen waren im wesentlichen Honorationswahlen, an denen 80% der

männlichen Bevölkerung teilnehmen durften (vgl. ausführlich Langewiesche 1981). Die 809 Parlamentarier der FNV übten insgesamt 927 verschiedene Berufe aus, die von Best (1990) in 10 Berufsgruppen zusammengefaßt wurden. Die Gruppen sind differenziert nach funktionalen Bereichen und innerhalb dieser weiter unterschieden nach funktionalen Differenzierungen. So z.B. der Staatsdienst, der nach Bildung (einschließlich Wissenschaft), Justiz, Verwaltung und Militär unterteilt wird. Die verbleibenden sechs Berufsgruppen sind die Angehörigen des Klerus, die Unternehmer, die Gutsbesitzer, die Advokaten, die freiberuflich Tätigen (bezogen auf die Mitglieder der FNV hauptsächlich Publizisten, Journalisten, Literaten und Künstler) sowie die Residualkategorie "Kleinbürger". Sowohl für die Analyse der intergenerationalen als auch der intragenerationalen Mobilität kann Best mit Hilfe der MDS vier Cluster nachweisen, die den sozialen Grundfunktionen der strukturell funktionalen Systemtheorie von Parsons (1985) entsprechen.

Das Cluster der Unternehmer und Kleinbürger entspricht der durch das wirtschaftliche Subsystem wahrgenommenen Anpassungsfunktion (adaption). Das zweite Cluster umfaßt die im Bildungsbereich Tätigen, es entspricht der durch das kulturelle Subsystem wahrgenommenen Funktion der Normerhaltung und Spannungsbewältigung (pattern maintenance); hierzu gehören die im Wissenschaftsbereich Tätigen, die Freiberuflichen (Journalisten, Publizisten, ...) und die Angehörigen des Klerus, die in der Mitte des 19. Jahrhunderts auch eine wichtige Bildungsfunktion hatten. Zum dritten Cluster werden die Angehörigen aus Verwaltung und Justiz sowie die Advokaten gerechnet, es entspricht der Funktion der sozialen Kontrolle und Aufrechterhaltung von Gesetznormen (integration). Das vierte Cluster, die Gutsbesitzer und Militärs, beinhaltet die Exponenten der traditionellen Führungsgruppen mit der Funktion der Zielerreichung (goal attainment). Da nur wenige Mitglieder der FNV als Beruf Unternehmer bzw. Kleinbürger angaben und da von diesen nur eine sehr geringe Mobilität zu dritten Berufen besteht, verzichten wir in den nachfolgenden Analysen auf diese beiden Kategorien.

### *Methode*

Die Korrespondenzanalyse ist ein exploratives Verfahren zur Darstellung der Zeilen und Spalten beliebiger Kontingenztafeln innerhalb einer Abbildung. Als Eingabeinformationen werden in der Regel die absoluten Werte der Zellen verwendet; bei den Kontingenztafeln selbst kann es sich sowohl um einzelne als auch um zusammengesetzte Tabellen handeln. Durch die Möglichkeit der Verwendung derartiger Tabellen kann z.B. die inter- und die intragenerationale Mobilität simultan beschrieben werden (zur Interpretation von zusammengesetzten Tabellen Blasius 1994). Ähnlich wie bei der Hauptkomponentenanalyse oder der MDS handelt es sich bei der Korrespondenzanalyse um ein Datenreduktionsverfahren, bei dem eine Vielzahl vorhandener Informationen mit möglichst wenigen, orthogonalen Faktoren beschrieben werden soll (vgl. ausführlich Greenacre 1984).

Die Korrespondenzanalyse basiert auf einer Chi-Quadrat-Metrik, d.h. analysiert (und grafisch dargestellt) werden die quadrierten Abweichungen von Erwartungs- und empirischen Werten (dividiert durch die jeweiligen Erwartungswerte). Je größer diese Werte, d.h. je größer die quadrierten Abweichungen von Erwartungs- und empirischen Werten sind, desto mehr tragen die

entsprechenden Komponenten zur Determination des Abbildungsraums bei. Die größten Abweichungen bei Mobilitätstabellen sind in der (den) Hauptdiagonale(n) der Kontingenztabelle(n) zu erwarten; die Berufsstabilität bzw. die Berufsvererbung ist in den meisten Fällen wesentlich höher als die Mobilität zu anderen Berufen.

### Ergebnisse

Zur simultanen Beschreibung der inter- und der intragenerationalen Mobilität wird der Hauptberuf im Jahre 1848/49 als Spaltenvariable verwendet (in den Tabellen und Abbildungen abgekürzt mit [h]; mit den Ausprägungen "Justiz" [j], "Verwaltung" [v], "Bildung" [b], "Militär" [m], "Klerus" [k], "Gutsbesitzer" [g], "Advokat" [a] und "freie Berufe" [f]), als Zeilenvariablen die entsprechenden Berufe zur Zeit des Eintritts in das Berufsleben (Erstberuf, [e]) sowie die entsprechenden Berufe des Vaters [v]. Die Matrix der Eingabedaten hat somit 16 Zeilen und acht Spalten (Tabelle 1).

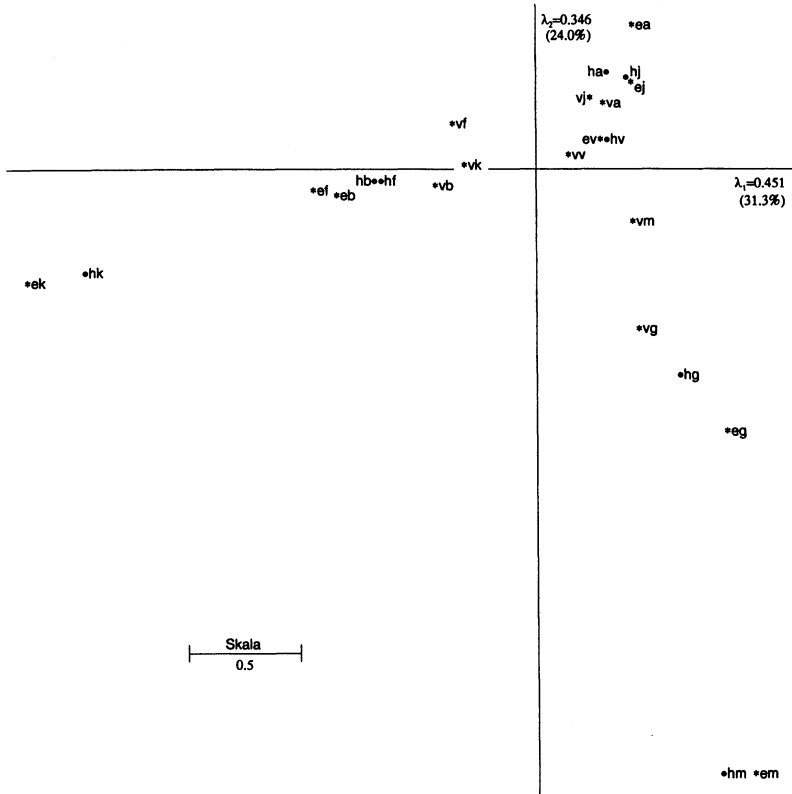
Tabelle 1: Matrix der Eingabedaten

	hj	hv	hb	hm	hk	hg	ha	hf	sum
vj	22	15	12	0	0	4	15	1	69
vv	27	34	20	4	4	8	24	5	126
vb	6	16	9	2	7	0	6	8	54
vm	8	8	1	3	0	1	3	2	26
vk	8	6	13	0	3	4	6	5	45
vg	11	12	6	3	1	35	5	5	78
va	4	10	2	0	0	2	10	2	30
vf	5	6	8	0	2	0	8	8	37
ej	117	83	19	0	1	17	76	6	319
ev	11	37	6	0	0	7	7	6	74
eb	4	3	67	1	4	3	5	13	100
em	0	5	1	16	0	8	1	0	31
ek	0	1	11	0	26	0	1	0	39
eg	0	4	0	0	0	19	1	0	24
ea	8	2	1	0	0	0	22	1	34
ef	0	7	20	0	4	1	0	37	69
sum	231	249	196	29	52	109	190	99	1155

Bereits anhand von Tabelle 1 werden die überdurchschnittlich hohen Werte in den beiden Hauptdiagonalen sichtbar, entsprechend sind in diesen Diagonalfeldern auch die höchsten Chi-Quadrat-Komponenten. Der Anteil der Residuen beträgt allein in diesen 16 Zellen nahezu 80 Prozent, d.h. der weitaus größte Teil der Variation der Daten wird durch die Berufsvererbung und durch die Berufsstabilität erklärt, wobei wiederum letztere den größeren Anteil hat. Dieses

Ergebnis wird auch in der grafischen Darstellung der Ergebnisse der Korrespondenzanalyse sichtbar, bei der das Maximum der Variation mit Hilfe der ersten Dimensionen erklärt wird.

Abbildung 1: Grafische Darstellung der Mobilitätstabellen



Bevor wir mit der Beschreibung der Ergebnisse beginnen, sollen ein paar generelle Informationen zur Interpretation von grafischen Darstellungen der Korrespondenzanalyse gegeben werden. Bei Abbildung 1 handelt es sich um eine symmetrische Darstellung - würde mit der hier vorliegenden zweidimensionalen Lösung die gesamte Variation der Daten erklärt, so dürften anhand dieser Grafik sowohl die Abstände zwischen den Zeilen- als auch die Abstände zwischen den Spaltenprofilen als euklidische Distanzen interpretiert werden; in dem vorliegenden Beispiel ist diese Interpretation jedoch der  $n(=7)$ -dimensionalen Lösung vorbehalten. Dementgegen sind die

Distanzen zwischen den Zeilen- und Spaltenprofilen auch im  $n$ -dimensionalen Raum nicht definiert (vgl. ausführlich Greenacre 1984). Soll dennoch auf deren simultane Interpretation nicht verzichtet werden, so können die Vektorendpunkte von Zeilen- und Spaltenmerkmalen auf die Achsen projiziert und anhand ihrer (gemeinsamen) Faktorladungen (der Kosinus der Winkel zum Achsenkreuz) interpretiert werden (vgl. ausführlich Blasius 1994).

Werden die Berufskategorien auf die beiden Achsen projiziert, so können drei Cluster unterschieden werden, die den Grundfunktionen von Parsons entsprechen: im rechten oberen Quadranten sind die neun Ausprägungen von Verwaltung, Advokaten und Justiz (integration), im rechten unteren Cluster sind die "Gutsbesitzer" und "Militärs" (goal attainment) und im linken Abschnitt die Kategorien des Bildungsbereiches (pattern maintenance). Aus diesem Ergebnis kann abgeleitet werden, daß sowohl die inter- als auch die intragenerationale Mobilität innerhalb der Subsysteme größer ist als die zwischen den Subsystemen. Anhand der hier nicht wiedergegebenen numerischen Darstellung kann des weiteren gezeigt werden, daß sowohl die Berufsvererbung als auch die Berufsstabilität größer ist als die Mobilität innerhalb der Subsysteme.

Wie bereits anhand von Tabelle 1 erläutert, werden nahezu 80 Prozent der Gesamtvariation der Daten durch die Berufsvererbung und durch die Berufsstabilität erklärt (siehe die überdurchschnittlich hohen Werte in den beiden Hauptdiagonalen). Die Mobilität zwischen den Berufen (auch innerhalb der Subsysteme) kann daher - wenn überhaupt - nur mit Hilfe höherer Dimensionen und mittels der numerischen Ergebnisse der Korrespondenzanalyse interpretiert werden.

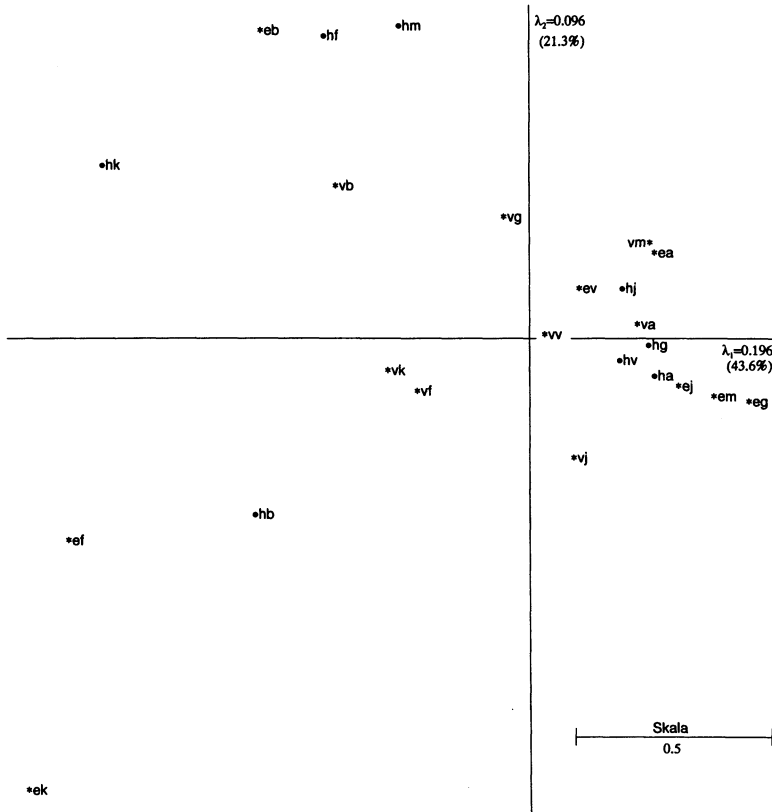
Sollen Aussagen über die Mobilität der Mitglieder der FNV gemacht werden, ohne daß die Berufsvererbung oder die Berufsstabilität einen Einfluß auf die Ergebnisse haben, so müssen die entsprechenden Effekte herausgerechnet werden - z.B. mittels einer Modifikation der Ausgangsdaten. Wird die Matrix der Eingabedaten (Tabelle 1) derartig verändert, daß die Chi-Quadrat-Komponenten der beiden Hauptdiagonalen gleich Null sind, so haben die entsprechenden Zellen keinen Einfluß mehr auf die Determination der geometrischen Orientierung der Achsen - der Einfluß von "Berufsstabilität" und "Berufsvererbung" auf die grafische Darstellung wäre damit neutralisiert. Die Chi-Quadrat-Komponenten der Hauptdiagonalen sind wiederum gleich Null, wenn die Erwartungswerte der entsprechenden Zellen gleich den empirischen Werten sind. Dieses Ziel kann iterativ erreicht werden: im ersten Schritte werden die Erwartungswerte der Ausgangsdaten (von Tabelle 1) berechnet, im zweiten werden die empirischen Werte der Hauptdiagonalen - und nur diese - durch diese ersetzt. Anschließend werden die Erwartungswerte der modifizierten Matrix der Ausgangsdaten berechnet, mit den im zweiten Iterationsschritt berechneten Erwartungswerten der Hauptdiagonalen werden die im ersten Schritt berechneten Erwartungswerte ersetzt. Der Abbruch der Prozedur erfolgt, wenn die Erwartungswerte in den Hauptdiagonalen nach dem  $n$ -ten Schritt (nahezu) gleich denen nach dem  $(n-1)$ -ten Schritt sind.

Tabelle 2: Modifizierte Matrix der Eingabedaten

	hj	hv	hb	hm	hk	hg	ha	hf	sum
vj	10.0	15	12	0	0	4	15	1	57.0
vv	27	34.9	20	4	4	8	24	5	126.9
vb	6	16	8.8	2	7	0	6	8	53.8
vm	8	8	1	0.3	0	1	3	2	23.3
vk	8	6	13	0	1.3	4	6	5	43.3
vg	11	12	6	3	1	3.3	5	5	46.3
va	4	10	2	0	0	2	5.0	2	25.0
vf	5	6	8	0	2	0	8	2.2	31.2
ej	43.1	83	19	0	1	17	76	6	245.1
ev	11	14.0	6	0	0	7	7	6	51.0
eb	4	3	6.5	1	4	3	5	13	39.5
em	0	5	1	0.2	0	8	1	0	15.2
ek	0	1	11	0	0.4	0	1	0	13.4
eg	0	4	0	0	0	0.4	1	0	5.4
ea	8	2	1	0	0	0	3.0	1	15.0
ef	0	7	20	0	4	1	0	2.5	34.5
sum	145.1	226.9	135.3	10.5	25.7	59.7	166.0	59.7	828.9

Durch das Einsetzen der Erwartungswerte in die beiden Hauptdiagonalen der Ausgangsdaten sinkt der Wert für das Gesamt-N von 1155 auf 828,9. Entsprechend der Modifikation der Daten sind die Chi-Quadrat-Komponenten der Hauptdiagonalen Null, die Gesamtvariation der Daten ist deutlich niedriger als in der Ausgangstabelle. Die Matrix mit den modifizierten Hauptdiagonalelementen wird als Eingabeinformation der zweiten Korrespondenzanalyse verwendet (Abbildung 2).

Abbildung 2: Korrespondenzanalyse auf der Basis der modifizierten Datenmatrix



Werden die beiden Korrespondenzanalysen verglichen, so ist in der zweiten Lösung keine eindeutige Clusterstruktur entsprechend der Parsonischen Grundfunktionen erkennbar. Mit dem "Herausnehmen" der Effekte in den Hauptdiagonalen und der damit deutlich geringeren Variation der Daten (vgl. auch die niedrigeren Eigenwerte der zweiten Lösung) sollte der jeweilige Hauptberuf im Jahre 1848/49 [h] nicht mehr mit dem korrespondierenden Erstberuf [e] bzw. dem korrespondierenden Beruf des Vaters [v] in einem Cluster liegen, dennoch bleibt die überdurchschnittlich hohe Mobilität zwischen den Berufen innerhalb der Parsonischen Grundfunktionen sichtbar. Wird exemplarisch der linke untere Quadrant betrachtet, so sind hier der Hauptberuf



"Bildung", die Erstberufe "freie Berufe" und "Klerus" und bei den Berufen des Vaters ebenfalls "freie Berufe" und "Klerus" lokalisiert. D.h., bleibt die Berufsstabilität und die Berufsvererbung außerhalb der Betrachtung, so gaben diejenigen, die 1848/49 hauptberuflich in der Bildung tätig waren, als Erstberuf sowie als Beruf des Vaters überdurchschnittlich oft "Klerus" und "freie Berufe" an. Ähnliches gilt für den Hauptberuf "Justiz" (rechter oberer Quadrant); die Angehörigen dieses Berufes kamen, abgesehen von "Justiz", überdurchschnittlich oft aus der "Verwaltung" bzw. sie waren als "Advokaten" tätig.

### Literatur

- Best, Heinrich (1990), Die Männer von Bildung und Besitz. Düsseldorf.
- Blasius, Jörg (1994), Correspondence Analysis in Social Science Research, in: Michael Greenacre/Jörg Blasius (eds.), Correspondence Analysis in the Social Sciences. Recent Developments and Applications, S. 23-52. London.
- Blau, Peter und Otis Dudley Duncan (1967), The American Occupational Structure, New York.
- Greenacre, Michael (1984), Theory and Applications of Correspondence Analysis, London.
- Langewiesche, Dieter (1981), Die deutsche Revolution von 1848/49 und die vorrevolutionäre Gesellschaft: Forschungsstand und Forschungsperspektiven, in: Archiv für Sozialgeschichte 21: 458-498.
- Parsons, Talcott (1985), Das System moderner Gesellschaften. Weinheim.

Dr. Jörg Blasius, Universität zu Köln, Zentralarchiv für Empirische Sozialforschung, Bachemer Str. 40, D-50931 Köln

## 2. Stochastic Dynamic Programming in Life Course Analysis: Basic Concepts and A simple model

*Ulrich Mueller*

For causal analysis in demography, individual longitudinal data are indispensable. Life table and transition rate methods, however, allow the analysis of single events only, they do not capture the character of the human life course as a adaptive sequence of transitions. Using concepts from evolutionary life history research, a new approach to analyzing whole life courses is presented: from measuring trade-offs between life course traits identifying optimal life courses with dynamic stochastic programming, and modeling the effect of covariates as determining deviations from the optimal sequence.

Stochastic Dynamic Programming is the tool of choice for the problem of optimizing the overall outcome of a sequence of decisions when the optimal choice of later steps depends on earlier steps (Mangel and Clark 1988; Puterman 1994).

Fundamental to Stochastic Dynamic Programming is the principle of optimality:

*"From any point on an optimal trajectory, the remaining trajectory is optimal for the corresponding problem initiated at that point"* (Luenberger 1979, 419).

The trick in this seemingly trivial statement is that, if the trajectory is optimal, the principle works backwards as forwards. Assume the trajectory is a sequence of discrete moves, with discrete outcomes. We consider a possible outcome of the process at terminal time T. Once we