

AXEL MUNK

Einige Bemerkungen zur Mathematischen Statistik

(vorgetragen in der Plenarsitzung am 13. Mai 2011)



Axel Munk, Felix-Bernstein-Professor für Mathematische Statistik an der Georg-August-Universität Göttingen, O. Mitglied der Göttinger Akademie seit 2011

Was ist Mathematische Statistik? In der Tat ist eine Standortbestimmung dieser sich in den letzten Jahrzehnten so rasant entwickelnden Disziplin nicht einfach, und ich werde exemplarisch versuchen, einige Merkmale herauszuarbeiten, die mir – und das ist eine sehr subjektive Betrachtung – als wesentlich erscheinen.

Der Terminus „Statistik“ legt es nahe, anzunehmen, dass die Analyse und Interpretation von Daten (etwa Messungen eines naturwissenschaftlichen Experimentes, einer sozialwissenschaftlichen Erhebung oder Beobachtungen aus einer klinischen Studie) im Vordergrund stehen. Was aber soll die Spezifizierung „mathematisch“ in diesem

Kontext genau bedeuten? Hierzu scheint es mir notwendig, den Begriff der Mathematischen Statistik, wie ich ihn in einem wissenschaftlichen Kontext verstehen möchte, etwas abzugrenzen von dem, was oftmals landläufig unter Statistik oder insbesondere auch unter amtlicher Statistik verstanden wird: das behördliche Zählen, Darstellen und Auswerten von gesellschaftlichen Daten, im englisch-amerikanischen Bereich oftmals als „official statistics“ bezeichnet. Beispielsweise kommentieren die Vereinten Nationen den ersten „World’s Statistics Day“ (man beachte: nicht den ersten „World’s Official Statistics Day“) am 20. 10. 2010 mit den Worten „to raise awareness of the many achievements of official statistics premised on the core value of service, professionalism and integrity“. Dies ist *ein* Aspekt der Statistik, aber diese geht sicherlich weit darüber hinaus.

In der Tat befassen sich viele Disziplinen seit langem mit der Modellierung und Auswertung von Daten, d. h. empirischen Beobachtungen, etwa Teile der Naturwissenschaften, der Computerwissenschaften, der angewandten Mathematik und selbstverständlich genuin all die Disziplinen, in denen derartige Daten anfallen. Die Fülle dieser Gebiete ist mittlerweile so groß, dass es schwieriger scheint, eines zu finden, in dem nicht empirisch gearbeitet wird, als umgekehrt.

Während sich bis zur Mitte des letzten Jahrhunderts die Mathematische Statistik hauptsächlich mit der Entwicklung einer mathematischen Theorie für kleine Datenmengen und recht einfacher Datenstrukturen beschäftigte, ist eine Herausforderung der letzten Jahrzehnte, mathematische Methoden zur Analyse massiver Datensätze und komplexer Datenstrukturen bereitzustellen. Dies hat zu grundlegend neuen statistischen Fragestellungen geführt, und die effiziente Berechenbarkeit statistischer Verfahren hat eine wichtige Stellung eingenommen. Übergänge zu anderen Disziplinen, wie etwa den Computerwissenschaften oder bestimmten Bereichen der angewandten Mathematik, sind dadurch fließender geworden. Sicherlich sind jedoch auch wesentliche Eckpfeiler mathematischer Statistik dieselben geblieben.

Ein wichtiges Merkmal einer „mathematischen statistischen“ Datenanalyse ist der Einsatz von probabilistischen Modellen und Werkzeugen, d. h. solchen, die auf die Verwendung einer mathematischen Theorie der Wahrscheinlichkeit rekurrieren.

Unter Zuhilfenahme vieler Entwicklungen in der Wahrscheinlichkeitstheorie (etwa der stochastischen Analysis oder der Wahrscheinlichkeitstheorie auf Graphen und Netzwerken) hat sich innerhalb der letzten Jahre eine rigorose Statistische Modellbildung zu einem prominenten Thema innerhalb der Statistik entwickelt, nicht zuletzt deshalb, weil immer komplexere Modelle numerisch gerechnet oder mit Computerexperimenten simuliert werden können. Die Kunst dabei ist, ähnlich der Modellbildung, die einer physikalischen Theorie zugrundeliegt, das Modell einerseits fein genug zu gestalten, so dass es alle wesentlichen Detailsigenschaften der Daten erhält, die es beschreiben soll und die etwa für eine Vorhersage oder eine Rekonstruktion benötigt werden, es aber andererseits genügend grob zu gestalten, so dass es ein strukturelles Verständnis des datengenerierenden Prozesses erlaubt, der zu eben diesen Daten geführt hat, oder, mit den Worten A. Einsteins, „so einfach wie möglich, aber auch nicht einfacher“.

In der Tat ist eine der wichtigsten Erkenntnisse der mathematischen Statistik, dass diese Phänomene Hand in Hand gehen: Nur die „richtige“ Reduktion der Komplexität von beobachteten Daten durch ein statistisches

Modell erlaubt eine Vorhersage mit statistisch kontrollierbarem Fehler. Ein derartiges statistisches Modell wird anhand von Daten ermittelt und validiert, und damit ist Mathematische Statistik sicherlich auch eine empirische Disziplin oder, genauer, eine Disziplin, die über empirische Untersuchungen mit anderen Disziplinen interagiert.

Eine der treffendsten Beschreibungen wird dem zeitgenössischen Wahrscheinlichkeitstheoretiker und Statistiker P. Diaconis (Stanford) zugeschrieben: „(Mathematische) Statistik ist die Physik der Zahlen.“

In der Tat sind wesentliche Strukturmerkmale beider Disziplinen sehr eng miteinander verwandt. Statistik basiert einerseits auf beobachteten Daten (Experiment) und andererseits auf einer Theorie, diese Daten zu modellieren und Entscheidungen mit kontrollierbarem Fehler – etwa eine Vorhersage aufgrund einer beobachteten Zeitreihe oder eine medizinische Diagnose mittels eines Klassifikationsverfahrens – zu treffen. Dabei spielen die Mathematik als die präziseste Sprache, über die wir verfügen, und Computer als unerlässliches Werkzeug zur empirischen Klärung und Beurteilung der Qualität solcher Entscheidungen eine maßgebliche Rolle.

Bioäquivalenz. Ein Beispiel aus der pharmazeutischen Statistik. Seit etwa Mitte der achtziger Jahre hat sich in der pharmazeutischen Forschung die Herstellung sogenannter Generika etabliert, meist bedingt durch das Auslaufen verschiedener Patente, die auf Wirkstoffe erteilt wurden. Generikahersteller nutzen dies aus, um diese Präparate zu kopieren. Dabei ist die Herstellung des Wirkstoffes oftmals sehr einfach und gut bekannt (etwa Azetylsalizylsäure als Wirkstoff in Kopfschmerzmitteln). Wichtig für die Wirksamkeit des Medikamentes ist jedoch auch, wie der Wirkstoff im Metabolismus absorbiert wird, das heißt, die Art und Weise, wie der Wirkstoff durch ein Präparat verabreicht wird, spielt eine wichtige Rolle. Ein Hustenspray hat sicher eine andere Wirkung als die Einnahme einer Flüssigkeit. Die Generikahersteller sind deshalb in der Pflicht, nachzuweisen, dass das kopierte Präparat im Blut ein ähnliches Absorptionsprofil (Bioverfügbarkeit) wie das Originalpräparat hinterlässt. Dies ist weltweit behördlich geregelt, etwa durch Bestimmungen der FDA (Food and Drug Administration) in den USA oder durch das BfArM (Bundesinstitut für Arzneimittel und Medizinprodukte) in Deutschland. Einen solchen Nachweis nennt man einen Bioäquivalenznachweis, und er erlaubt ein stark vereinfachtes und damit zeit- und kostengünstigeres Prüfverfahren, als es bei einer „normalen“ Neuzulassung notwendig würde. Somit stellt sich die Aufgabe, aus den Daten (etwa den integrierten Zeitkonzentrationskurven des Wirkstoffes im Blut) einer Probandengruppe, die das Originalpräparat (O), und einer, die das Generika (G) erhält, statistisch gesichert (d. h. innerhalb einer

tolerierbaren Fehlerschranke) zu schließen (statistisch zu testen), dass diese sich nicht um mehr als eine bestimmte Toleranzgrenze in den Mittelwerten unterscheiden. Dies ist zunächst eine etwas ungewöhnliche Fragestellung, da statistische Testverfahren traditionell für eine Situation konzipiert sind, in der es darum geht, einen Unterschied nachzuweisen (etwa die Verbesserung der Heilungsrate durch eine Behandlung oder die Wirksamkeit einer Intervention, z. B. eines bestimmten Trainings bei Sportlern) und nicht eine Ähnlichkeit. Bei Bioäquivalenzprüfungen ist es jedoch genau die Ähnlichkeit, die man nachweisen muss. Initiiert durch diese Problematik, sind verschiedene solcher statistischen Testverfahren zum Bioäquivalenznachweis vorgeschlagen worden, und einige davon haben sich in der Praxis gut bewährt und werden heute weltweit vielfach eingesetzt (vgl. Senn 1997). Eine „mathematisch statistische“ Herangehensweise an das Problem wäre, zu versuchen, die Frage zu beantworten, ob es ein „optimales“ Testverfahren, d. h. eines mit kleinstmöglichen Irrtumswahrscheinlichkeiten, für diese Fragestellung gibt. Dies zwingt uns zuallererst, ein Modell präzise zu spezifizieren, und das meistverwendete – und einfachste – Modell in diesem Kontext ist, anzunehmen, dass die Mittelwerte der logarithmierten Daten normalverteilt sind. Der Bioäquivalenznachweis ist nun dann erbracht, wenn statistisch gesichert gezeigt werden kann, dass die zugehörige Differenz der Erwartungswerte innerhalb einer tolerierbaren Schranke Δ um Null liegt, gebräuchlich ist etwa $\Delta = \pm \text{Ln}(1.25)$. Nun stellt sich die Frage nach einem optimalen Verfahren – schon in dieser sehr einfachen Situation – als ein überraschend schwieriges Problem heraus. Dieses offenbar sehr hoch gesteckte Ziel kann in einem nächsten Schritt abgeschwächt werden, und man versucht, Verfahren zu finden, die zumindest gewissen Minimalanforderungen genügen. Eine Minimalanforderung, die man bisher als solche für diese Situation angesehen hat, ist die Unverfälschtheit eines Tests. Grob gesprochen, bedeutet dies, dass ein Testverfahren so konzipiert sein sollte, dass die Wahrscheinlichkeit, sich aufgrund der Daten richtigerweise für die Bioäquivalenz zu entscheiden, nicht kleiner sein soll als die Fehlerwahrscheinlichkeit, sich für die Bioäquivalenz der beiden Medikamente zu entscheiden, obwohl diese in Wirklichkeit nicht zutrifft. Dies liest sich kompliziert, lässt sich jedoch mathematisch sehr einfach hinschreiben. Viele prominente Testverfahren, die bei den bis dahin üblichen Fragestellungen zur Prüfung einer Unterschiedlichkeit eingesetzt werden – etwa der heutzutage immer noch sehr populäre Studentische t-Test, der von dem bei der Brauerei Guinness tätigen Chemiker W. S. Gosset 1908 unter dem Pseudonym „Student“ publiziert wurde, oder auch die sogenannten F-Tests aus der Varianzanalyse –, haben diese Eigenschaft der Unverfälschtheit.

Versucht man nun, unverfälschte Tests für das Bioäquivalenzproblem zu entwickeln, so stellt sich heraus, dass dies ein wesentlich komplizierteres Unterfangen ist als das Testen der Unterschiedlichkeit, für das Gossets t -Test in gewisser Hinsicht eine optimale Antwort gibt (was erst etwa 50 Jahre später nach Einführung dieses Tests rigoros bewiesen wurde). Für das Bioäquivalenzproblem kann man eine Idee verwenden, die auf Hodges & Lehmann (1954) zurückgeht, und erhält eine vollständige geometrische Charakterisierung aller unverfälschten Tests. Benötigt werden nur die Mittelwertdifferenz D der Daten aus den beiden Gruppen und die geschätzte gemeinsame Standardabweichung S . Die geometrische Aufgabe lautet dann: Finde eine um die S -Achse symmetrische Menge in der oberen Halbebene, die noch eine weitere Eigenschaft hat. Diese Eigenschaft lässt sich am einfachsten bei zwei Beobachtungen formulieren – eine praktisch irrelevante Situation, aber sie gibt einem die Leitidee für alle anderen Stichprobenumfänge, etwa 24, so wie sie in vielen Bioäquivalenzstudien heutzutage verwendet werden. Diese symmetrische Menge muss aus jedem Halbkreis um die Koordinate $(D, S) = (1, 0)$ ein Stück proportional zur vorgegebenen Fehlerschranke des Tests ausschneiden (vgl. Abbildung 1). Hat man eine solche Menge gefunden, so beschreibt der innere Bereich (in Abbildung 1 orange eingefärbt) genau den Bereich, in dem man sich für die Bioäquivalenz entscheidet. Es ist erstaunlich schwierig, solche Mengen zu konstruieren, eine weitere wird in Abbildung 2 gezeigt. Heutzutage ist immer noch offen, ob diese Mengen – und damit die zugehörigen unverfälschten Tests für das Bioäquivalenzproblem – für alle Stichproben-

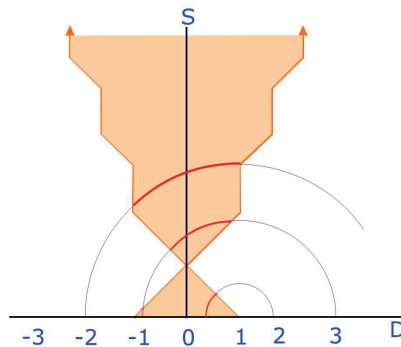


Abbildung 1: Die oben gezeigte Menge (orange) hat die Eigenschaft, dass jeder Halbkreis um den Wert $D = 1$ ein Stück (rot markiert) ausschneidet, welches proportional zum Winkel $\frac{\pi}{4}$, d.h. 45° ist. Die entsprechende Fehlerschranke des Tests ist $\frac{1}{4}$. In diesem Beispiel ist $\Delta = 1$.

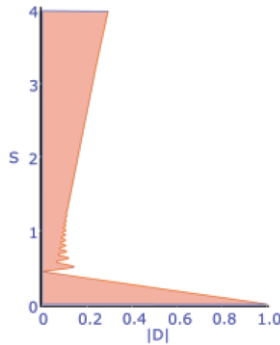


Abbildung 2: Numerisch bestimmter Verwerfungsbereich eines unverfälschten Tests, Fehlerniveau 0.05, 6 Beobachtungen, $\Delta = 1$. Gezeigt ist die rechte Seite; die linke Seite erhält man durch Spiegelung.

umfänge und alle Testniveaus existieren. Viele gelöste Fälle findet man in (Brown et al. 1997).

Die Übersetzung dieses sehr einfach zu formulierenden statistischen Problems in ein rein geometrisches Problem erlaubt einem also einen verblüffenden neuen Zugang zu der Frage nach einem möglichst guten Testverfahren für den Bioäquivalenznachweis. Hat man diesen Zusammenhang verstanden, so kann man damit nun einige weitere Aussagen treffen. Beispielsweise lässt sich folgendes „Äquivalenzparadox“ mittels dieses Argumentes beweisen (vgl. Munk, 2001).

Jeder unverfälschte Test – auch in Situationen, in denen wir noch nicht wissen, ob ein solcher Test existiert oder nicht und wie er gegebenenfalls aussieht – verletzt notwendigerweise eine äußerst plausible „Common Sense“-Eigenschaft: Falls die Streuung der Daten steigt, sollte die Entscheidungskraft des Tests fallen, oder anders: Jeder unverfälschte Test hat notwendigerweise folgende paradoxe Eigenschaft: Es gibt Werte der Mittelwertdifferenz D , für die der Test bei einer beobachteten Streuung für die Bioäquivalenz entscheidet, bei größerer Streuung dagegen und bei noch größerer Streuung gleichfalls dafür.

Man stelle sich folgende Situation vor: Zwei Hersteller führen eine Bioäquivalenzstudie durch, erhalten dieselbe empirische Mittelwertdifferenz für den Unterschied der integrierten Konzentrationsprofile, etwa $D = 0.5$, aber unterschiedliche Versuchsstreuungen. Der erste Hersteller hat eine Standardabweichung von 0.4 gemessen und ist damit nicht in der Lage, Äquivalenz nachzuweisen. Der Zweite hat einen „schlechteren“ Versuch gemacht: die Standardabweichung ist nun $S = 1.5$ (größere Streuung,

weniger informativ), er kann aber nun die Bioäquivalenz zur gleichen Fehlerschranke, die der erste Hersteller verwendet hat, für sich beanspruchen und darf somit letztlich sein Medikament auf dem Markt anbieten. Das Äquivalenzparadox besagt nun, dass eine solche Situation nie ausgeschlossen werden kann. Was folgt daraus? Dies ist offensichtlich ein Beispiel, in dem ein lange etabliertes statistisches Kriterium (Unverfälschtheit) notwendigerweise zu „paradoxen“ Testentscheidungen führt. Gegeben, die mathematische Analyse ist korrekt, so muss also offensichtlich das Kriterium konzeptuelle Schwächen haben (eine ausführliche Diskussion hierzu findet sich in Berger et al. 1996 und, aus bayesianischer Sicht, in Perlman & Wu 1997). Dies trifft auch sicherlich zu. Der Begriff der Unverfälschtheit wird dennoch heutzutage in vielen Statistiklehrbüchern recht unkritisch verwendet. Aber die Situation ist leider noch etwas komplizierter: Man kann unverfälschte Verfahren konstruieren, die also notwendigerweise das oben beschriebene „Common Sense“-Kriterium verletzen (vgl. Brown et al. 1998), die aber auch stets in ihrer Entscheidung einen kleineren mittleren Fehler machen als die üblichen Testverfahren, die heutzutage in der Praxis eingesetzt werden. Warum sollte man darauf verzichten?

Das heißt, man erlaubt den Einsatz von statistischen Verfahren, von denen man weiß, dass diese im Mittel stets eine höhere Fehlerquote haben als das oben beschriebene unverfälschte Verfahren. Wenn man nun berücksichtigt, dass jährlich viele hundert solcher Bioäquivalenzstudien durchgeführt werden, so kann man argumentieren, dass sicherlich einige davon durch ein unverfälschtes Verfahren besser ausgewertet worden wären. Dieses Beispiel mag ein wenig die Vorgehensweise Mathematischer Statistik illustrieren: Grundlegend ist eine präzise mathematische Modellierung, um ein klar definiertes Zielkriterium und eine entsprechende mathematische Analyse von statistischen Verfahren zu erhalten. Dies kann sogar die Konstruktion optimaler Verfahren ermöglichen. Dies gelingt nicht immer in vollem Umfang, aber oftmals sind auch schon partielle Erfolge aussagekräftig. Die Existenz oder gar die explizite Konstruktion eines optimalen Bioäquivalenzverfahrens, d. h. eines solchen, welches die Wahrscheinlichkeit, eine Fehlentscheidung zu treffen, unter allen „denkbaren“ Verfahren minimiert, ist übrigens bis heute noch ein offenes Problem.

Hochfrequenzdaten. Als zweites Beispiel möchte ich die Analyse sogenannter Hochfrequenzdaten bei Finanzzeitreihen nennen. Auch hier lassen sich bestimmte Aspekte der Situation erst dann gut verstehen, wenn man eine rigorose mathematische Modellierung und eine auf ihr aufbauende Analyse vornimmt. Zum Hintergrund: Die statistische Modellierung der Preisprozesse von an der Börse gehandelten Wertpapieren – etwa Anleihen

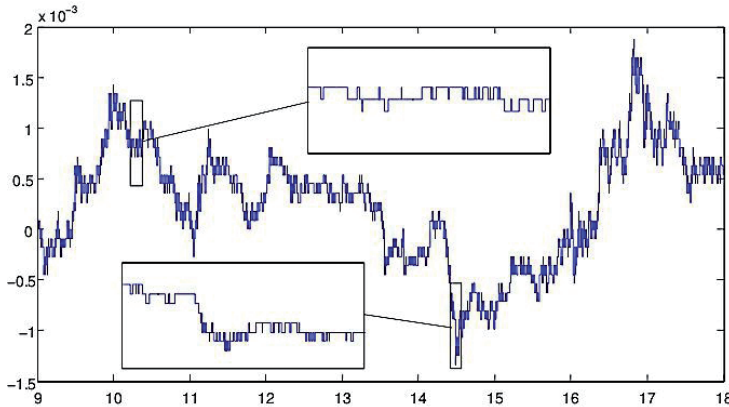


Abbildung 3: Logarithmierte Zuwächse einer Bundesanleihe an einem Handelstag (von 9 bis 18 Uhr). Auf kleinen Zeitskalen ist das hierfür typische Sprungverhalten – gehandelt wird in $\frac{1}{4}$ Cent.

oder Optionen – hat sich in den letzten Jahren rasant entwickelt, und es gibt eine nahezu unüberschaubare Fülle von Modellen. Insbesondere der Bestimmung und Vorhersage der zugehörigen Fluktuationen (Spotvolatilität) des Wertes eines solchen Papiers kommt dabei eine wichtige Funktion zu, da sie direkt in die Ankaufs- und die Verkaufsstrategien des Handels eingeht.

Nun stellt sich empirisch heraus – und dies ist eine universelle Beobachtung bei sehr vielen Finanzprodukten –, dass bei sehr kleinen Zeitskalen (innerhalb derer überhaupt erst seit etwa 10 Jahren gehandelt wird und die als Tick- oder Hochfrequenzdaten bezeichnet werden; vgl. Abbildung 3) das Stabilisierungsverhalten dieser Volatilitäten einem anderen Skalierungsgesetz genügt als bei vergleichsweise großen Zeitskalen. Wenn etwa die aufeinander folgenden stündlichen logarithmierten relativen Zuwächse eines Preisprozesses voneinander abgezogen, quadriert und dann über einen langen Zeitraum gemittelt werden (etwa ein paar Tage), bekommt man ungefähr dieselbe Zahl wie bei viertelstündlichen Messungen. Diese Größe nennt man empirische (integrierte) Volatilität; deren zeitliche Ableitung ist die Spotvolatilität. Bei minütlichen Messungen – die kleinsten Zeitskalen, in denen heutzutage elektronisch gehandelt wird, liegen im Millisekundenbereich – ändert sich das Verhalten der empirischen Volatilität jedoch schon sehr, und diese Zahlen explodieren um so mehr, je kleiner die zeitliche Auflösung wird, für die die Daten zur Berechnung verwendet wird.

Diese empirische Beobachtung widerspricht aber einigen typischen Annahmen, die man über solche Zeitreihen trifft und die wesentlich in die Preistheorie von Finanzprodukten eingehen, insbesondere der Arbitragefreiheit, d. h., dass keine Gewinne aus risikofreien Transaktionen getätigt werden können. In der Tat ist bekannt, dass eine solche Annahme insbesondere auf kleinen Zeitskalen nicht ganz zutrifft. Andere Phänomene wie etwa Rundungsfehler, die sich auf großen Zeitskalen herausmitteln und somit keine Rolle spielen, kommen hier zusätzlich zum Tragen. Nun kann man diese Situation dadurch modellieren, dass weitere kleine zufällige Fehler (Mikrostrukturrauschen) auf kleinen Zeitskalen addiert werden, die sich dann auf großen herausmitteln und keine Rolle mehr spielen (vgl. Zhang et al. 2005). Diese Fehler „verdecken“ gewissermaßen zusätzlich die Spotvolatilität, und völlig andere statistische Skalierungen werden sichtbar. Die klassischen n -Gesetze (n bezeichnet die Anzahl der Daten, aus denen die Volatilität geschätzt wird), werden nun viel „langsamere“ $n^{1/4}$ -Gesetze, und man kann beweisen, dass dies die bestmögliche Schätzungsgeschwindigkeit ist (vgl. Munk & Schmidt-Hieber 2010a, b; Reiss, 2011). Zhang et al. 2005 erhielten ein $n^{1/6}$ -Gesetz für ihr Verfahren, also eine suboptimale Geschwindigkeit. Messungen auf kleiner Zeitskala liefern also mehr Daten, aber die Information aus jeder einzelnen Beobachtung nimmt im Vergleich zu großen Skalen ab, und man schafft es grundsätzlich nie, dieselbe Information aus einer Beobachtung zu extrahieren, wie es bei großen Zeitskalen möglich ist. Dort hat man dann allerdings weniger Daten zur Verfügung. In sehr einfachen Modellen kann man sogar exakte untere Informationsschranken ausrechnen und kann mit recht trickreichen Methoden Schätzer konstruieren, die diese präzise annehmen (vgl. Cai et al. 2010). Dabei stellt sich heraus, dass man Daten auf sehr kleinen Skalen zunächst geeignet mitteln muss, bevor man sie wirklich zur Schätzung verwenden kann, so dass die resultierenden Schätzer nicht mehr „explodieren“. Dies liefert einen Leitfaden für realistischere und kompliziertere Modelle, und eine Kombination mit Techniken aus dem Bereich der Wavelets und der Martingaltheorie erlaubt schließlich eine Schätzung der Spotvolatilität auf sehr kleinen Zeitskalen, welche dann auch die bestmöglichen Schätzungsgeschwindigkeiten annimmt (vgl. Hoffmann et al. 2012, vgl. Abbildung 4).

Das Phänomen, das hier eine tragende Rolle spielt, besteht darin, dass auf kleinen Zeitskalen einerseits die sehr diskrete Struktur der Daten berücksichtigt werden muss und dass andererseits schlecht modellierbare Handelsmechanismen das Bild der Spotvolatilität zusätzlich verrauschen und „unscharf“ werden lassen. Aus einem solchen unscharfen Bild muss

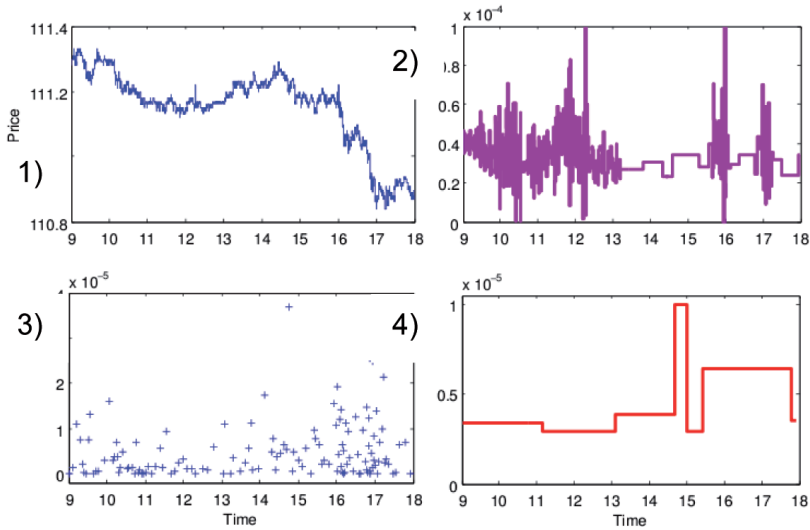


Abbildung 4: Preisprozess einer Bundesanleihe an einem Handelstag. 2) Geschätzte Spotvolatilität aus sekundlichen Daten. 3) Aus 1) geschätzte Pseudodaten, die um das Mikrostrukturrauschen bereinigt sind. 4) Geschätzte Spotvolatilität, um Mikrostrukturrauschen bereinigt. Man sieht deutlich, dass in 2) im Zeitraum zwischen 9 und 13 Uhr die Volatilität völlig falsch geschätzt wird. Dies wird vermutlich durch eine hohe Handelsaktivität verursacht.

„invers“ zurückgerechnet werden, und dies lässt sich nur noch mit einer bestimmten Genauigkeit tun.

Statistische Inverse Probleme. Prinzipiell ähnliche Phänomene treten in völlig anderen Bereichen auf, etwa in der Molekularmikroskopie. In der Tat sind dort heutzutage die Mikroskopietechniken und die damit verbundene Auflösung so gut, dass analog zu den Hochfrequenzdaten aus der Finanzstatistik die Diskretheit der Daten (wenig gezählte Photonen in hoher räumlicher und zeitlicher Auflösung), aus denen eine Proteinverteilung zu berechnen ist, explizit berücksichtigt werden muss und unvermeidbare Beugungseffekte zu einem ähnlichen Effekt der Unschärfe führen (vgl. Hell 2009; Hell & Rittweger 2009). In der SMS-Mikroskopie ist die zeitliche Aufnahmedauer, die benötigt wird, um eine räumlich sehr gute Auflösung zu erhalten, sogar so groß, dass die zu bestimmenden Proteinstrukturen sich so stark bewegen, dass dies zu einem zusätzlichen Verwischungseffekt führt, der sich in einer weiteren Unschärfe des Bildes widerspiegelt (vgl. Abbildung 5a,d). Vom statistischen Standpunkt ist dies nicht sehr viel anders als die zusätzliche Unschärfe, die auf sehr kleinen Zeitskalen bei der Schätzung

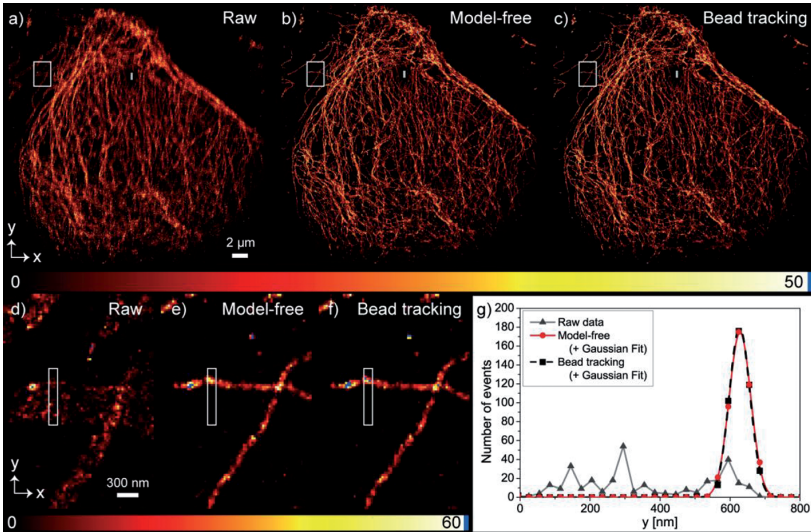


Abbildung 5: SMS-Aufnahme eines β -tubulin Netzwerkes in einer PtK2-Zelle. a) Herkömmliche Rekonstruktion aus den Rohdaten ohne Berücksichtigung der Zeitdynamik. b) Rekonstruktion mittels zeitdynamischer Korrektur, die aus den Daten geschätzt wird. c) Rekonstruktion mittels der Verfolgung eines „Beads“. d)–f) Vergrößerungen der Rechtecke in a)–c). Dabei zeigt d) die unbearbeiteten Daten, e) das Bild nach einer statistischen Zeitdynamik-Korrektur und f) das Bild nach der Korrektur durch „Bead“-Verfolgung. g) Profile in y -Richtung über x integriert innerhalb der markierten Bereiche in d)–f).

der Volatilität entsteht. Diesem Effekt kann man physikalisch dadurch begegnen, dass ein besonders hell leuchtendes Kügelchen (bead) in die Probe implantiert wird und man die Zeitdynamik dieses Kügelchens beobachtet und damit dann das Bild korrigiert (vgl. Abbildung 5c,f). Dies gelingt jedoch genauso gut mit einer rein statistischen Überlegung, mit der man die Bahn der Probe aus allen vorhandenen Daten schätzen kann. Hierzu ist es nicht mehr notwendig, das Kügelchen aufwändig in die Probe einzubauen, und man erhält genauso gute Ergebnisse (vgl. Abbildung 5b,e; vgl. Geisler et al. 2012). Verwandte statistische inverse Probleme entstehen in allen Bereichen der Computertomographie, etwa bei der Magnetresonanztomographie (MRT), wie sie für die diagnostische Bildgebung im Klinikalltag eingesetzt wird. Hier gelang es kürzlich der Arbeitsgruppe um Jens Frahm, durch radiales Abtasten der Signale und die Verwendung von nichtlinearen Regularisierungsverfahren (vgl. Uecker et al. 2008) die für eine gute Bildrekonstruktion notwendige Datenmenge und Rechenzeit sogar soweit zu reduzieren, dass nun dynamische Bildgebung in Echtzeit möglich ist (vgl.

Frahm & Uecker 2010). Im Unterschied zum vorherigen Beispiel will man hier die Dynamik gerade sichtbar machen – etwa den Herzschlag – und sie nicht herausrechnen. Die statistischen Modelle hierfür haben aber eine gewisse Verwandtschaft, und man kann sie alle als statistische inverse Probleme beschreiben. Die mathematischen Techniken, mit denen man diese bearbeitet, sind einander deshalb an vielen Stellen ähnlich.

So kann man zusammenfassend sagen, dass Mathematische Statistik Sprache und Techniken bereitstellt, die die Mathematik selbst vielfach bereichern und die es oft erlauben, auf den ersten Blick sehr verschiedene empirische Disziplinen einheitlich zu beschreiben, zu verstehen und Lösungen zu entwickeln. Unter dem Blickwinkel der immer stärker werdenden Differenzierung der empirisch arbeitenden Wissenschaften ist die Perspektive einer Vereinheitlichung sicherlich eine gute Nachricht.

Literatur

- Berger, R. L., HSU, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. With discussion. *Statistical Science* 11, 283–319.
- Brown, L. D., Hwang, J. T. G., Munk, A. (1997). An unbiased test for the bioequivalence problem. *The Annals of Statistics* 25, 2345–67.
- Cai, T., Munk, A., Schmidt-Hieber, J. (2010). Sharp minimax estimation of the variance of Brownian motion corrupted with Gaussian noise, *Statistica Sinica* 20, 1011–1024.
- Fisher Box, J. (1987). „Guinness, Gosset, Fisher, and Small Samples“. *Statistical Science* 2, 45–52.
- Frahm J., Uecker, M. (2010). Echtzeit-MRT: die Zweite. *Jahrbuch der Akademie der Wissenschaften zu Göttingen* 2010, 263–270.
- Geisler, C., Hotz, T., Schönle, A., Hell, S. W., Munk, A., Egner, A. (2012). Drift estimation for single marker switching based imaging schemes. *Optics Express* 20, 7274–7289.
- Hell, S. W. (2009). Far-Field Optical Nanoscopy. In: *Single Molecule Spectroscopy in Chemistry, Physics and Biology*. Springer (Berlin, Germany), 365–398.
- Hell, S. W., Rittweger, E. (2009). Light from the dark. *Nature* 461, 1069–1070.
- Hodges, J. L., Lehmann, E. L. (1954). Testing the approximate validity of statistical hypotheses. *Journal of the Royal Statistical Society, Ser. B* 16, 261–268.
- Hoffmann, M., Munk, A., Schmidt-Hieber, J. (2012). Nonparametric estimation of the volatility under microstructure noise: wavelet adaptation. *Annales de l'Institut Henri Poincaré*, to appear.
- Munk, A. (2001). On a problem in pharmaceutical statistics and the iteration of a peculiar nonlinear operator in the upper complex halfplane. *Nonlinear Analysis* 47, 1513–1523.
- Munk, A., Schmidt-Hieber, J. (2010a). Lower bounds for volatility estimation in microstructure noise models. *Borrowing Strength: Theory Powering Applications – A Festschrift for Lawrence D. Brown*, IMS Collections. 6, 43–55.

- Munk, A., Schmidt-Hieber, J. (2010b). Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise. *Electronic Journal of Statistics* 4, 781–821.
- Perlman, M, Wu, L. (1999). The Emperor's new tests. *Statistical Science* 14, 355–369.
- Reiss, M. (2011) Asymptotic equivalence for inference on the volatility from noisy observations. *Annals of Statistics* 39, 772–802.
- Senn, S. (1997). *Statistical Issues in Drug Development*. Wiley.
- The Student (1908). The probable error of a mean. *Biometrika* 6, 1–25.
- Uecker, M., Hohage, T., Block K. T., Frahm, J. (2008). Image reconstruction by regularized nonlinear inversion – joint estimation of coil sensitivities and image content. *Magnetic Resonance in Medicine* 60, 674–682.
- Zhang, L., Mykland, P.A., Ait-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1396–1411.