

# 2009

**DAG voor STATISTIEK en BESLISKUNDE**

maandag 6 april 2009

Vrije Universiteit, Amsterdam

---

hoofdsprekers

**Mark van der Laan  
&  
Paul Embrechts**



# DAG VOOR STATISTIEK EN BESLISKUNDE 2009

## Toegang

De lezingen zijn vrij toegankelijk voor alle belangstellenden, ook van buiten de Vereniging.

## Locatie

De Dag voor Statistiek en Besliskunde is dit jaar te gast bij de Vrije Universiteit te Amsterdam. De Dag voor Statistiek en Besliskunde zal plaatsvinden in het Hoofdgebouw, auditorium en zalen.

Vrije Universiteit Amsterdam, Hoofdgebouw, De Boelelaan 1105, 1081 HV Amsterdam. Het ochtendprogramma vindt plaats in het auditorium, de slotlezing in de aula. De zalen zijn bewegwijzerd. Voor routebeschrijving en plattegrond zie <[www.vu.nl/organisatie/index.cfm](http://www.vu.nl/organisatie/index.cfm)>.

## Lunch en borrel

De lunch kan in de VU locatie worden genuttigd (voor eigen rekening). De slotborrel wordt aangeboden door de Vereniging.

## Taal

De voertaal voor de algemene gedeelten is Nederlands. De meeste lezingen zijn echter in het Engels.

## Organisatiecomité

Aad van der Vaart, Richard Gill, Cees Diks, Gerrit Stermerdink en de Sectiecoördinatoren

## Informatie

Cees Diks, e-mail <[c.g.h.diks@uva.nl](mailto:c.g.h.diks@uva.nl)>, <<http://home.uva.nl/c.g.h.diks>>

## VVS

Vereniging voor Statistiek en Operationele Research, Postbus 244, 6700 AE Wageningen, telefoon 0317 - 419572, fax 0317 - 421364, <admin@vvs-or.nl>. Raadpleeg onze site <www.vvs-or.nl> over hoe u lid kunt worden van de VVS of een abonnement kunt nemen op een van de VVS-periodieken.

## Algemene Leden Vergadering

De Algemene Leden Vergadering vindt plaats voorafgaand aan het lezingenprogramma. De vergaderstukken kunnen vanaf 2 weken voor de vergadering worden gedownload van de website van de VVS-OR, of worden opgevraagd bij de voorzitter van de vereniging.

## Wandelgangen

De Dag voor Statistiek en Besliskunde van de VVS-OR is dé ontmoetingsplaats voor mensen die de statistiek en de operations research een warm hart toedragen danwel in de genoemde vakgebieden werkzaam zijn; leden en niet-leden ontmoeten elkaar in de wandelgangen. Ook kunt u de stands van enkele exposanten bezoeken.



*De Dag voor Statistiek en Besliskunde 2009 wordt mede mogelijk gemaakt door de Vrije Universiteit en Cosinus Computing BV, Drunen.*



## DAG VOOR STATISTIEK EN BESLISKUNDE 2009

maandag 6 april 2009 in Amsterdam

- 09.20 – 10.20 Algemene Leden Vergadering
- 10.00 – 10.30 Pauze met koffie en thee
- 10.30 - 10.35 Opening Dag voor Statistiek en Besliskunde 2009
- 10.35 – 11.30 Lezing door Mark van der Laan (zie pagina 5)
- 11.30 – 12.00 Uitreiking VVS-Scriptieprijs met lezing door prijswinnaar
- 12.00 – 13.30 Lunch
- 13.30 – 15.15 Parallelsessies (zie pagina 8 e.v.)  
*De parallelsessies worden georganiseerd door de Secties van de VVS-OR. Bovendien zijn er een speciale sessies georganiseerd door het Centraal Bureau voor de Statistiek en het Nederlands Forensisch Instituut.*
- 15.15 – 15.30 Pauze met koffie en thee
- 15.30 – 16.30 Lezing door Paul Embrechts (zie pagina 7)
- 16.30 – 17.30 Borrel

# Plenaire lezingen

**MARK VAN DER LAAN**

*University of California, Berkeley*

## **Towards robust black-box algorithms for causal effects that preserve meaningful statistical inference**

Current statistical practice to assess an effect of an intervention or exposure on an outcome of interest often involves either maximum likelihood estimation for a priori specified regression model, or, manual and/or data adaptive interventions to fine tune a choice of model. In both cases, bias in the point estimates and the estimate of the signal to noise ratio are rampant, causing an epidemic of false claims based on data analyses.

In this talk we present our efforts to construct machine learning algorithms for estimating a causal effect that take away the need for specifying regression models, while still providing maximum likelihood based estimators and inference. Two fundamental concepts underlying this methodology are the very aggressive use of cross-validation to select optimal combinations of many model fits, and subsequent targeted maximum likelihood estimation to target the fit towards the causal effect of interest.

We illustrate this method in observational studies for assessing the effect of an intervention on adherence to drug regimen in HIV infected patients, and for discovery of mutations in the HIV virus that cause resistance to a particular drug regimen. We also illustrate the performance on FDA approved clinical trials, simulated data imitating postmarket safety analysis, and the analysis of single nucleotide polymorphisms.

MARK VAN DER LAAN is Professor of Biostatistics and Statistics at UC Berkeley. His research interests include statistical methods in genomics (i.e., computational biology), survival analysis, censored data, semiparametric models and causal inference, data adaptive loss based estimation, and

multiple testing. Mark is also the founder of Target Analytics, a company that offers statistical consulting services and translates state-of-the-art causal inference methodology into user-friendly software products. He has been Associate Editor for *Biometrics* (1997-2003) and *Lifetime Data Models* (1996-2000), and Vice-President of the Bay-Area Chapter of the American Statistical Association. He is currently Associate Editor for *Journal of Statistical Planning and Inference*, *Statistical Applications in Genetics and Molecular Biology*, and the *Annals of Statistics*, and is Director of the Biostatistics and Computing core of the Superfund Research Program of a Center on Genomics in Environmental Science in the School of Public Health, headed by Professor Martyn Smith. He is also founding editor of a new electronic International Journal in Biostatistics, which is led by Professor Nick Jewell. Mark received the 2004 Mortimer Spiegelman Award, the Van Dantzig Award on April 11, 2005, was selected to present the 2005 Lefkopoulou distinguished lectureship, Harvard University, Boston, September 15, 2005, and the long-term Jiann-Ping Hsu/Karl E. Peace Endowed Chair in Biostatistics starting 2005, the 2005 COPPS Snedecor Award and the 2005 (COPSS) Presidential Award.

**Paul Embrechts**

*ETH, Zürich*

## Mathematical Finance Before and 'After' the Credit Crisis.

We are currently witnessing a deep economic crisis worldwide which started with the so-called subprime crisis in the USA. Mathematics is (though only marginally) involved through the pricing and hedging of some of the more complicated products underlying the market for credit risk, e.g. CDOs, CDSs, MBSs. Also statistics plays an important role in the calibration of such models to data.

Examples will be given of some of the 'bad models' used, or still in use. In this talk I will also review the current state of the application of statistical methodology to questions within quantitative risk management and draw conclusions on the kind of research I feel of importance for the years to come.

PAUL EMBRECHTS is Professor of Mathematics at the ETH Zürich specialising in actuarial mathematics and quantitative risk management. Previous academic positions include the Universities of Leuven, Limburg and London (Imperial College). Dr. Embrechts has held visiting appointments at the University of Strasbourg, ESSEC Paris, the Scuola Normale in Pisa (Cattedra Galileiana), the London School of Economics (Centennial Professor of Finance) and has an Honorary Doctorate from the University of Waterloo. He is an Elected Fellow of the Institute of Mathematical Statistics, Honorary Fellow of the Institute and the Faculty of Actuaries, Corresponding Member of the Italian Institute of Actuaries and is on the editorial board of numerous scientific journals. He belongs to various national and international research and academic advisory committees.

He co-authored the influential books *Modelling of Extremal Events for Insurance and Finance*, Springer, 1997 and *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton UP, 2005. Dr. Embrechts consults for a number of leading financial institutions and insurance companies, and is a member of the Board of Directors of companies in insurance and finance.

# Programma's van de secties *voor abstracts zie pagina 13 en verder*

## Speciale Sessie georganiseerd door het CBS

- 13.30-14.05 **Bart F.M. Bakker, Centraal Bureau voor de Statistiek / Sector Statistische Analyse Personen**  
Naar een methodologisch raamwerk voor het gebruik van administratieve gegevens voor sociaal-wetenschappelijk onderzoek.
- 14.05-14.40 **Edwin de Jonge, Centraal Bureau voor de Statistiek / Sector Methodologie**  
Data in zicht!
- 14.40-15.15 **Jan Mol, Centraal Bureau voor de Statistiek / Centrum voor Beleidsstatistiek**  
Remote Access: de toegangspoort naar de schatkamer van het CBS

## Speciale Sessie georganiseerd door het Nederlands Forensisch Instituut (NFI)

### theme FORENSIC STATISTICS

Forensic Statistics is an emerging area of applied probability and statistics. It focuses on the probabilistic aspects of forensic science, that is, the use of scientific knowledge in criminal cases. DNA profiles, fingerprints and footprints are well-known examples, but forensic sciences also includes forensic chemistry, physics, pathology, toxicology and digital technology. Two main areas of forensic statistics can be distinguished: the 'normal' statistical work like sampling, design of experiments, and hypotheses testing, and a new area called interpretation of evidence. The aim of the latter is to quantify the value of the evidence used in court. This has raised some very interesting questions, to which this session will devote attention.

- 13.30-14-00 **Klaas Slooten en Marjan Sjerps, Nederlands Forensisch Instituut**  
Forensic DNA statistics
- 14.00-14.25 **Annabel Bolck, met Céline Weyermann, Laurence Dujourdy, Ivo Alberink, Pierre Esseiva en Jorrit van den Berg, Nederlands Forensisch Instituut**  
Likelihood Ratios for Comparing XTC
- 14.25-14.50 **Ivo Alberink, Nederlands Forensisch Instituut**  
Meetonzekerheid en Likelihood Ratios bij lengtemetingen aan personen in digitale beelden
- 14.50-15.15 **Richard Gill, Leiden University**  
Forensic statistics: ready for consumption?

## Sectie Mathematische Statistiek

### theme UNCERTAINTY IN FINANCE -- TAILS WAGGING DOGS

- 13.30-14.05 **Peter Spreij, Universiteit van Amsterdam**  
Negative volatility for a two-dimensional square root stochastic differential equation
- 14.05-14.40 **Michel Vellekoop, Universiteit Twente**  
Two tales of too heavy tails
- 14.40-15.15 **Jasper Anderluh**  
Parisian options and excursions of Brownian motion

## **Sociaal-Wetenschappelijke Sectie**

### **thema (GEZOND) ETEN: DENKEN EN DOEN**

- 13.30-14.05 **Frank Busing, Methoden en Technieken van Psychologisch Onderzoek, Universiteit Leiden**  
Restricted Unfolding: Preference Analysis with Optimal Transformations of Preferences and Attributes
- 14.05-14.40 **Garmt Dijksterhuis, Unilever Food & Health Research Institute**  
Food Consumer Science: Where Do the Data Come From?
- 14.40-15.15 **Frank van Lenthe, Afdeling Maatschappelijke GezondheidsZorg, Erasmus MC, Rotterdam**  
Methodological challenges of our obesogenic environment

## Economische Sectie

### theme EMPIRICAL ANALYSIS OF LABOR MARKET PHENOMENA

- 13.30-14.05 **Bas van der Klaauw, VU**  
Analyzing female labor supply. Evidence from a Dutch tax reform
- 14.05-14.40 **Monique de Haan, UvA**  
The Effect of Parents' Schooling on Child's Schooling: a Nonparametric Bounds Analysis
- 14.40-15.15 **Marloes de Graaf-Zijl, UvA**  
Stepping-stones for the unemployed: the effect of temporary jobs on the duration until (regular) work

# ABSTRACTS

**Ivo Alberink**, Nederlands Forensisch Instituut

## **Meetonzekerheid en Likelihood Ratios bij lengtemetingen aan personen in digitale beelden**

In de forensische wereld begint meer en meer de nadruk te liggen op de invoering van deugdelijke validatie-procedures en kwantificering van bewijswaarde. Meer concreet is er behoefte aan kennis over meetonzekerheid, Likelihood Ratios en operator-effecten. Een goede illustratie van deze thema's vormt het uitvoeren van lengtemetingen aan personen in digitale (meest CCTV) beelden. Metingen worden hierbij uitgevoerd door verschillende onderzoekers, die cilinders rond personen plaatsen in een 3D model van de plaats delict. Hierbij spelen alle drie genoemde thema's een rol. De onderzoekers doen metingen aan de persoon in de betwiste beelden alsmede aan figuranten in beelden van reconstructie-opnamen. In de presentatie wordt ingegaan op het verkrijgen van betrouwbaarheidsintervallen voor werkelijke lengtes in geval van kleine steekproeven (weinig figuranten) en aanwezigheid van systematische afwijkingen. Verder wordt ingegaan op het bepalen van numerieke Likelihood Ratios gegeven reconstructie-metingen en populatiedata.

**Jasper Anderluh**

## **Parisian options and excursions of Brownian motion**

In this talk we introduce the Parisian option and explain how valuation of the Parisian option is connected to excursion theory for Brownian motion. The valuation is done in the most basic setting, i.e. the Black-and-Scholes world, in which the stock price is modeled by a geometric Brownian motion. Even for this case, there exist no explicit formulas and in this talk the option pricing is done by Laplace transforms. The talk ends with the calculation of the Laplace transforms that are relevant for the double-sided Parisian option. These Laplace transforms deliver also nice formulas for certain Brownian motion excursion-related probabilities.

**Bart F.M. Bakker**, Centraal Bureau voor de Statistiek / Sector Statistische Analyse Personen

### **Naar een methodologisch raamwerk voor het gebruik van administratieve gegevens voor sociaal-wetenschappelijk onderzoek.**

In steeds meer empirisch sociaal-wetenschappelijk onderzoek wordt gebruik gemaakt van administratieve gegevens. De voordelen daarvan zijn evident: de kosten voor de verzameling van de gegevens is een fractie van de kosten van enquetes, de mensen waarvan de informatie wordt gebruikt hoeven er niet mee lastig gevallen te worden en de informatie beslaat dikwijls de gehele (sub)populatie waardoor onderzoek naar kleine groepen mogelijk is. Bij veel onderzoekers die geen ervaring hebben met het gebruik van administratieve gegevens bestaat het naïeve idee dat onderzoek doen met deze gegevens niet meer is dan koppelen van verschillende administratieve bestanden en analyseren daarvan. Ervaren gebruikers van administratieve gegevens erkennen inmiddels dat ook voor het gebruik hiervan voor sociaal-wetenschappelijk onderzoek een methodologie nodig is.

Tot op heden zijn vooral ad hoc methoden ontwikkeld om een specifiek probleem op te lossen en zijn best practices gepubliceerd. In mijn presentatie zal ik een voorstel doen voor een theoretisch raamwerk voor registermethodologie. Daarin worden de mogelijke foutenbronnen geïnventariseerd, gecategoriseerd en met elkaar in verband gebracht. Aan de hand van enkele voorbeelden wordt aangegeven op welke wijze dit raamwerk in de praktijk kan worden toegepast.

**Annabel Bolck, met Céline Weyermann, Laurence Dujourdy, Ivo Alberink, Pierre Esseiva en Jorrit van den Berg**, Nederlands Forensisch Instituut

### **Likelihood Ratios for Comparing XTC**

In Forensic Science often items are compared to investigate whether they have a common source. XTC tablets are compared by looking at their external characteristics, active compound percentage, organic impurities and elemental composition. In many countries distance measures (or similarity

indices) are calculated based on all these characteristics to conclude that tablets (probably) come from the same production facility or not. This is a typical conclusion. In many forensic fields it is nowadays generally accepted to use a Bayesian approach, in which not a conclusion about the hypothesis is provided by the forensic experts, but a probability statement about the evidence. In that case verbal scales or numerical Likelihood Ratios (LR) can be used as a measure for the strength of the evidence. In this presentation it is shown how we can construct Likelihood Ratios for determining the strength of evidence in comparing MDMA tablets. The LR will in most cases concern continuous characteristics, and because of the large amount of characteristics will eventually be multivariate. The LR can be constructed as a ratio of the probability density of one (reference) batch and the probability density over all batches. Various assumptions can be made concerning the population parameters of the distributions, resulting in (slightly) different LR formulas and differences between the actual calculated LR's. These results are compared and submitted to sensitivity analyses. These results and the assumptions leading to them are discussed

**Frank Busing**, Methoden en Technieken van Psychologisch Onderzoek,  
Universiteit Leiden

### **Restricted Unfolding: Preference Analysis with Optimal Transformations of Preferences and Attributes**

The fundamentals of preference mapping are revisited in the context of a new restricted unfolding method that has potential for wide application to product optimisation for consumers. Since more of an attribute is not necessarily preferred, the unfolding distance model provides estimates of ideal points for consumers and therefore provides a more adequate representation of the preference relationships, compared to conventional internal preference mapping. Compared to other ideal point methods, the new unfolding technique offers advantages in terms of allowing for the ordinal nature of the ratings, rather than implicitly assuming that ratings are linear. The proposed restricted unfolding model incorporates property fitting, both

passive, as a separate, second step, and active, as a restriction on the product locations. This is also available as a restriction on the respondents' locations and as such establishing a link between internal and external preference mapping.

**Garnt Dijksterhuis**, Unilever Food & Health Research Institute

### **Food Consumer Science: Where Do the Data Come From?**

As the field of Food Consumer Science is about understanding and predicting food related behaviour, its research methods should be valid with respect to behaviour. Choosing the best research method requires a clearly formulated research question. Without a clear goal a method cannot be judged to be good or bad. Measuring the behaviour itself is, per definition, the most valid method. Psychological theory amply illustrates that one is mostly not aware of one's own motivations to show a certain behaviour. It is concluded from this that methods that rely on consumers' reflecting on their own behaviour, or on their hypothesised behaviour, can not be very good in predicting actual behaviour. Three criteria to judge methods on are introduced and used to illustrate that a large number of methods that are often applied in Food Consumer Science may be invalid with respect to actual food related behaviour. It may even be impossible to gain insight about future behaviour without ever studying that very behaviour in an experiment.

**Richard Gill**, Leiden University

### **Forensic statistics: ready for consumption?**

Forensic statistics gets its unique nature from the court setting in which a forensic statistician must communicate with a board of judges or with the members of a jury. The classical modes of frequentist statistics (hypothesis tests, confidence intervals and so on) are not understood; a full Bayesian approach is not allowed. What is left is the likelihood ratio. I will describe some examples and pitfalls, show that even in the least controversial situations, many issues still need to be addressed, among others, those connected

to extrapolation, choice of loss functions, role of nuisance parameters and prior distributions, and robustness to model assumptions. I will give some new results on empirical Bayes validation via nonparametric estimation of the likelihood ratio of a model based likelihood ratio (!).

**Edwin de Jonge**, Centraal Bureau voor de Statistiek / Sector Methodologie  
**Data in zicht!**

Datavisualisatie is sinds lange tijd een belangrijk hulpmiddel bij het maken van statistiek. Dit geldt niet alleen tijdens de analyse, het is ook een waardevol middel om de resultaten van de analyse zichtbaar te maken. Tegenwoordig is internet het medium waarop statistiek gepubliceerd wordt. De ontwikkelingen in web technologie hebben voor nieuwe mogelijkheden (of oude in een nieuwe jasje) gezorgd, waarmee statistiek getoond kan worden. Net als andere statistische bureaus verkent het CBS de mogelijkheden en probeert haar statistiek helder weer te geven. Voor een goede visualisatie is datakennis, web-kennis, maar ook methodologische kennis nodig. De presentatie gaat in op de huidige visualisatie initiatieven van het CBS en een aantal van de methodologische aspecten.

**Frank van Lenthe**, Afdeling Maatschappelijke GezondheidsZorg, Erasmus MC, Rotterdam

**Methodological challenges of our obesogenic environment**

The prevalence of obesity has increased dramatically in the past decades. The prevention of obesity is among the most important challenges in public health. Effective prevention strategies need to be based on state-of-the-art knowledge of the determinants of obesity. Clearly, obesity is the result of an imbalance between energy intake and energy expenditure, but what are the underlying determinants of this imbalance? There are good reasons to believe that an increasingly unhealthy diet (and a decrease in physical activity) to some extent are rooted in a changed living environment, our so-called obesogenic environment. In this presentation, methodological challenges in

studying the relationship between environmental characteristics and dietary patterns will be discussed.

**Jan Mol**, Centraal Bureau voor de Statistiek / Centrum voor Beleidsstatistiek  
**Remote Access: de toegangspoort naar de schatkamer van het CBS**

Het Centraal Bureau voor de Statistiek beschikt over een zeer grote verzameling databestanden als bron voor haar eigen publicaties. Deze verzameling bestaat niet alleen uit de enquêtes die het CBS zelf uitvoert, maar ook meer en meer uit (integrale) registraties van andere overheidsorganisaties. Om het maatschappelijke nut van deze bestanden zo groot mogelijk te maken, kan het CBS externe onderzoekers ook toegang verlenen tot deze bestanden. Deze toegang is technisch en procedureel zo ingericht dat de flexibiliteit voor de onderzoeker gemaximaliseerd wordt, zonder de vertrouwelijkheid van de gegevensbestanden in het geding te laten komen. Een onderzoeker benadert de databestanden ofwel vanaf beveiligde computers bij het CBS (On Site) ofwel vanaf zijn eigen werkplek (Remote Access). De essentie van deze dienstverlening zal worden belicht samen met enkele voorwaarden voor het gebruik. Er zal worden afgesloten met een korte demonstratie van de Remote Access voorziening.

**Klaas Slooten en Marjan Sjerps**, Nederlands Forensisch Instituut  
**Forensic DNA statistics**

Forensic DNA profiling is used in criminal casework in several ways. When e.g. a blood stain is found at a crime scene, it is possible to make a DNA profile from it and search the national DNA database and other European DNA databases for DNA matches in order to identify a suspect. When no match is found, it is possible to extend the search for relatives of the crime stain donor (familial search). Furthermore, the DNA match between the crime stain and the suspect can be used as evidence in court. DNA profiling is also used in relatedness analyses, for instance in paternity disputes, immigration cases, rape cases, missing person cases, and (disaster) victim identification. These

uses of DNA profiling have raised a number of interesting statistical questions. We will address the hottest debates and recent developments.

**Peter Spreij**, Universiteit van Amsterdam

### **Negative volatility for a two-dimensional square root stochastic differential equation**

In affine term structure models the short interest rate is modeled as an affine transformation of a multi-dimensional square root process. In the one-dimensional case this is known as the Cox-Ingersoll-Ross model. Sufficient conditions to avoid negative volatility factors are the multivariate Feller conditions. We will show their necessity for a 2-dimensional square root SDE in canonical form by presenting a methodology to generate counter examples. This is based on measure transformations and the trivial fact that a random variable assumes negative values if it has negative expectation. We exploit the property that solutions to square root SDEs have expectations which solve a system of linear differential equations.

**Michel Vellekoop**, Universiteit Twente

### **Two tales of too heavy tails**

The desire to incorporate 'heavy tails' in financial asset prices has led to interesting new research topics, but also to some questionable mathematical model assumptions. Our primary example will be a pricing model for derivatives which has recently been proposed. In this model, underlying stock price processes are Markovian, with rates of return that physicists call 'Tsallis-distributed' and which exhibit top-heavy tails. We examine the properties of this process and prove that the closed-form option pricing formulas that have been derived for this model are incorrect. In order to specify a correct model, we prove existence and uniqueness of a strong solution for a certain degenerate stochastic differential equation.

