Mathematisches Forschungsinstitut Oberwolfach

Report No. 06/2010

DOI: 10.4171/OWR/2010/06

Statistical Issues in Prediction: what can be learned for individualized predictive medicine?

Organised by
Leonhard Held (Zürich)
Robin Henderson (Newcastle upon Tyne)
Ulrich Mansmann (München)

January 24th – January 30th, 2010

Abstract. Error is unavoidable in prediction. And it is quite common, often sizable, and usually consequential. In a clinical context, especially when dealing with a terminal illness, error in prediction of residual life means that patients and families are misinformed about their illness, that they may take foolish actions as a result, and that they may be given inappropriate or needlessly painful treatments or denied appropriate ones. In meteorology, error in prediction of storm paths or extreme events can have devastating consequences. In finance and economics, major policy decisions are taken on the basis of predictions and forecasts. Rational approaches to reduce and assess error in prediction are presented. Ideas are introduced how to relate these statistical strategies with clinical and medical concepts in particular and how to integrate ideas from apparently different areas.

Mathematics Subject Classification (2000): 62N01, 62P10, 92C60.

Introduction by the Organisers

There is a recent resurgence of interest and activity in probability forecasting, which encompasses a wide range of sciences [1]. As far as medicine is concerned, this has been motivated in part by the more routine availability of individual–level genetic information and consequent potential for improved prognosis, diagnosis, and individualized therapy [2]. Further motivation has arisen from dramatic increases in power of the computationally intensive statistical methods needed to determine predictive probability distributions.

The goal of a good probabilistic prediction is to maximize the sharpness of the predictive distributions subject to calibration [3]. Calibration refers to the statistical consistency between the probabilistic forecasts and the observations, and is a joint property of the predictive distributions and the events that materialize. Sharpness refers to the concentration of the predictive distributions, and is a property of the forecasts only: The sharper the distributional forecast, the less the uncertainty, and the sharper, the better, subject to calibration. A number of alternative prediction accuracy measures have been suggested, for example skill scores [4] and the notion of predictiveness [5].

The workshop studied recent developments of tools to quantify the quality of a prediction strategy. These tools have additional impact on guidance in a wealth of applied statistical problems for count data [6], multivariate continuous data [7] and survival data [8]. They range from the evaluation of probabilistic forecasts to model criticism, model comparison and model choice.

Predictive distributions arise naturally in Bayesian modelling approaches. Therefore, the workshop looked at their state–of–the–art and explored interaction between Bayesian ideas and alternative approaches.

The intersection of genomics and medicine has the potential to yield a new set of molecular tools that can be used to individualize and optimize therapy as well as prognosis [9]. Specific prediction problems in individualized medicine are related to individual prognosis. Sharpness of individual prognosis is hampered by the intrinsic large uncertainty of point processes which are so far the methodological backbone of the predictive models. Biomarkers and their relevance for diagnosis, prognosis and clinical patient management pose new challenges on the development of statistical methods for joint models [10, 11, 12]. High–dimensional data from genetic screens are a further aspect to be integrated into the theoretical basis of prediction models for individualized medicine [13, 14].

The workshop also offered contributions in prediction strategies applied in the atmospheric sciences, economics, and finance. Their relevance for the solution of the problems to be handled in individualized predictive medicine was discussed.

The general aspects discussed during the workshop were producing and assessing probabilistic forecasts. Probabilistic forecasts arise natural in a Bayesian framework, taking automatically parameter uncertainty into account. A number of alternative likelihood–based approaches also exist [15]. Technically, probabilistic forecasts are often based on a random sample from the predictive distribution, due to the non–accessibility of the predictive distribution in a closed form. For example, in meteorology so–called ensemble forecasts are often used. These technical problems increase for multivariate forecasts where a closed form of the predictive distribution is rarely available. The selection of useful criteria for the assessment of the quality of probabilistic forecasts is of paramount importance. Proper scoring rules, for example the logarithmic or the Brier score, are accepted tools that address both calibration and sharpness of a prediction. The mathematical theory of proper scoring rules is related to the theory of convex functions, to information measures, and entropy functions [16, 3].

Further complications arise in the selection of the validation set in order to quantify the quality of a predictive model. Cross–validation is at the one end of the spectrum, while truly external validation sets are at the other extreme. As shown by Stone [17], the cross–validated logarithmic score is asymptotically equivalent to Akaike's information criterion (AIC), commonly used for model selection [18]. On the other hand, the Bayesian information criterion (BIC) can be viewed as an approximation to the log marginal likelihood, the sum of the one-step-ahead logarithmic scores [19, 20].

The specific aspects for individualized predictive medicine considered *individual prognosis for event times*, *joint modelling of biomarkers and event time data*, *high–dimensional data for predictive models*, as well as *global assessment of strategies in predictive medicine, clinical studies, and meta–analysis.*

Individual prognosis for event times: Prediction of time–to–event is particularly challenging yet fundamentally important, especially in medicine when there is a need to predict residual lifetime following diagnosis of a potentially terminal disease [8]. Complications include censoring of available data and heteroscedasticity. A variety of measures of predictive accuracy have been suggested e.g. [21, 22, 23] but none has been uniformly accepted. Furthermore, the availability of high dimensional genetic information has brought two unsolved challenges: how best to measure the additional value of genetic information over perhaps simpler to measure characteristics and how to deal with the well-known $p \gg n$ problem for predictive purposes as opposed to estimation and model selection. Dealing with high–dimensional covariate data for non–linear models is beginning to attract attention but numerous problems remain [24]. The usefulness of genetic information for individual level prediction was a core feature of the workshop.

Joint modelling of biomarkers and event time data: Biomarkers — measures of biological health — can be used as measures of disease progression and as prior surrogates for long term events. Examples are CD4 cell counts or other blood markers for HIV/AIDS e.g. [25] or reduction in telomere length as a measure of aging [26]. Methods of the joint analysis of the evolution of longitudinal biomarkers and time–to–event data are being developed [27, 28] but what is not yet clear is how best to exploit biomarker trajectories — not just current values — for predictive purposes. A comprehensive Bayesian approach for individual prediction based on longitudinal biomarker measurements is a promising theoretical approach [10].

High–dimensional data for predictive models: There is a series of studies which demonstrate the clinical value of prognostic models based on data measured by high–throughput biotechnologies. These models are in general derived by applying black–box model free algorithms which do not incorporate subject matter knowledge on the disease of interest [13]. The common feature of these algorithms is a mechanism of regularization which helps to handle the high–dimensionality of the measurements used to derive the prognosis [14]. Research on theoretical grounds of these regularization techniques is of high interest for mathematical statistics with eminent implication for practical applications. Besides using data to predict there is also usually an aims of obtaining information about the underlying data generation mechanism. The first successes in establishing clinically valuable gene signatures are not matched by an elucidation of the disease processes which produce the data. Theoretical guidance and appropriate statistical tools are needed to build the bridge between a gene signature and the functional aspects of the disease under consideration [29].

Global assessment of strategies in predictive medicine, clinical studies, and meta–analysis: The availability of gene signatures which predict specific risks or the response on therapeutic substances is the building stone of individualized medicine. They need a thorough assessment of their clinical relevance. The mathematical and statistical foundations of new design ideas for clinical trials have to be developed and implemented in biometrical practice [2]. Furthermore, there is a lack of mathematical and statistical tools for combining knowledge over a large literature on the relationship between specific biomarkers and response on specific substances. First examples of these meta–analyses are being published and started the discussion on methodological development [30].

References

[1] T. Gneiting, *Editorial: Probabilistic forecasting*, Journal of the Royal Statistical Society: Series A (Statistics in Society) **171** (2008), 319–321.

[2] R. Simon, *Roadmap for developing and validating therapeutically relevant genomic classifiers*, Journal of Clinical Oncology **23** (2005), 7332–7341.

[3] T. Gneiting, A.E. Raftery, *Strictly proper scoring rules, prediction, and estimation*, Journal of the American Statistical Association **102** (2007), 359–378.

[4] W. Briggs, D. Ruppert, *Assessing the skill of yes/no predictions*, Biometrics **61** (2005), 799–807.

[5] Y. Huang, M.S. Pepe, Z. Feng, *Evaluating the predictiveness of a continuous marker*, Biometrics **63** (2007), 1181–1188.

[6] C. Czado, T. Gneiting, L. Held, *Predictive model assessment for count data*, Biometrics **65** (2009), 1254–1261.

[7] T. Gneiting, L.I. Stanberry, E.P. Grimit, L. Held, N.A. Johnson, *Assessing probabilistic forecasts of multivariate quantities, with applications to ensemble predictions of surface winds (with discussion)*, Test **17** (2008), 211–264.

[8] R. Henderson, N. Keiding, *Individual survival time prediction using statistical models*, Journal of Medical Ethics **31** (2005), 703–706.

[9] A. Dupuy, R. Simon, *Critical Review of Published Microarray Studies for Cancer Outcome and Guidelines on Statistical Analysis and Reporting*, Journal of the National Cancer Institute **99** (2007), 147–157.

[10] J.M. Taylor, D.P. Ankerst, R.R. Andridge, *Validation of biomarker–based risk prediction models*, Clinical Cancer Research **14** (2008), 5977–5983.

[11] C. Proust-Lima et J.M.G. Taylor, *Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach.* Biostatistics, **10** (2009), 535–549.

[12] Y. Zheng, T. Cai, M.S. Pepe, W.C. Levy, *Time–dependent predictive values of prognostic markers with failure time outcome*, Journal of the American Statistical Association **103** (2008), 362–368.

[13] M. Schumacher, H. Binder, T. Gerds, *Assessment of survival prediction models based on microarray data*, Bioinformatics **23** (2007), 1768–1774.

[14] A. Benner, M. Zucknick, T. Hielscher, C. Ittrich, U. Mansmann, *High-dimensional Cox models: the choice of penalty as part of the model building process*, Biometrical Journal **52** (2010), 50–69.

[15] G.A. Young, R.L. Smith, *Essentials of Statistical Inference*, (2005), Cambridge University Press, Cambridge.

[16] L. J. Savage, *Elicitation of personal probabilities and expectations*, Journal of the American Statistical Association **66** (1971), 783–801.

[17] M. Stone, *An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion*, Journal of the Royal Statistical Society, Ser. B **39** (1977), 44–47.

[18] G. Claeskens and N. J. Hjort, *Model Selection and Model Averaging*, (2008), Cambridge University Press, Cambridge.

[19] A. P. Dawid, *Statistical theory: The prequential approach*, Journal of the Royal Statistical Society, Ser. A **147** (1984), 278–292.

[20] R. E. Kass, A. E. Raftery, *Bayes factors*, Journal of the American Statistical Association **90** (1995), 773–795.

[21] P.J. Heagerty, Y. Zheng, *Survival model predictive accuracy and ROC curves*, Biometrics **61** (2005), 92–105.

[22] S. Rosthoj, N. Keiding, *Explained variation and predictive accuracy in general parametric statistical models: the role of model misspecification*, Lifetime Data Analysis **10** (2004), 461–472.

[23] P. Royston, W. Sauerbrei, *A new measure of prognostic separation in survival data*, Statistics in Medicine **23** (2004), 723–748.

[24] S. Nygaard, O. Borgan, O.C. Lingjaerd, H.L. Storvold, *Partial least squares Cox regression for genome–wide data*, Lifetime Data Analysis **14** (2008), 179–195.

[25] X. Song, C.Y. Wang, *Semiparametric approaches for joint modeling of longitudinal and survival data with time–varying coefficients*, Biometrics **64** (2008), 557–568.

[26] T. DeMeyer, E.R. Rietzshel, M.L. De Buyzere, E. Van Criekinge, S. Bekaert, *Studying telomeres in a longitudinal population based study*, Frontiers in Bioscience **13** (2008), 2960–2970.

[27] R.M. Elashoff, G. Li N. Li, *A joint model for the longitudinal and survival data in the presence of competing failure types*, Biometrics **64** (2008), 762–771.

[28] P. Diggle, D. Farewell, R. Henderson, *Analysis of longitudinal data with dropout: objectives, assumptions and a proposal (with discussion)*, Applied Statistics **56** (2007), 499–550.

[29] M. Hummel, K.H. Metzeler, C. Buske, S.K. Bohlander, U. Mansmann, *Association between a prognostic gene signature and functional gene sets*, Bioinformatics and Biology Insights **2** (2008), 329–341.

[30] C.Y. Li, M. Mao, L. Wei, *Genes and (common) pathways underlying drug addiction*, PLoS Computational Biology **4** (2008), e2.

Workshop: Statistical Issues in Prediction: what can be learned for individualized predictive medicine?

Table of Contents

Abstracts

When is a risk prediction model useful in the individual?

Martin Schumacher

In many areas of epidemiology and clinical medicine, risk prediction models are used to individually predict the risk of the occurrence or the course of a particular disease, respectively. Popular examples are the Gail model for the development of breast cancer or the Framingham risk score for dying on cardiovascular disease. Mathematically, a risk prediction model is a rule that yields a predicted probability of the event of interest in a certain time period given the covariate information of an individual available at the time of prediction. It is common practice that from a significant effect of a risk factor and/or a good calibration of the model it is assumed that it also has good predictive ability. This, however, is generally not the case. In order to judge the predictive ability, an unbiased estimation of measures of prediction error is necessary that satisfy several requirements, e.g. consider time in an adequate way and allow a comparison with suitable benchmark values. In some examples from epidemiology and clinical medicine, it is shown that the predictive ability of most current risk prediction models is poor or at least modest and can only marginally be improved by adding genomic markers. Finally, some requirements are derived that make a good prediction model also useful in the individual. For achieving this ultimate goal, further steps towards improvement of current risk prediction models, the evaluation of suitable criteria and identification of appropriate study designs have to be taken.

References

[1] S.G. Baker, *Putting risk prediction in perspective: Relative utility curves*, Journal of the National Cancer Institute **101** (2009), 1538–1542.

[2] H. Binder, M. Schumacher, *Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models*, BMC Bioinformatics **9** (2008), 10–19.

[3] M.H. Gail, *Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model*, Journal of the National Cancer Institute, **101** (2009), 959–963.

[4] T.A. Gerds, T. Cai, M. Schumacher, *The performance of risk prediction models*, Biometrical Journal **50** (2008), 457–479.

[5] J.P.A. Ioannidis, *Personalized genetic prediction: too limited, too expensive, or too soon?*, Annals of Internal Medicine **150** (2009), 139–141.

[6] R.J. Marshall, *Cardiovascular risk can be represented by scaled rectangle diagrams*, Journal of Clinical Epidemiology **62** (2009), 998–1000.

[7] K. Offit, *Breast cancer single–nucleotide polymorphisms: statistical significance and clinical utility*, Journal of the National Cancer Institute **101** (2009), 973–975.

[8] M.S. Pepe, H.E. Janes, *Gauging the performance of SNPs, biomarkers, and clinical factors for predicting risk of breast cancer*, Journal of the National Cancer Institute **100** (2008), 978–979.

[9] E.W. Steyerberg, *Clinical prediction models*, (2009), Springer, New York.

[10] E.W. Steyerberg, A.J. Vickers, N.R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M.J. Pencina, M.W. Kattan, *Assessing the performance of prediction models: A framework for traditional and novel measures*, Epidemiology **21** (2010), 128–138.

[11] J.M.G. Taylor, D.P. Ankerst, R.R. Andridge, *Validation of biomarker–based risk prediction models*, Clinical Cancer Research **14** (2008), 5977–5983.

The value of adding SNP data to a model for breast cancer risk for public health decisions

Mitchell H. Gail

This work compares the National Cancer Institute's 'Breast Cancer Risk Assessment Tool' (BCRAT), a model that predicts breast cancer risk with standard epidemiologic predictors such as family history, with a model that also includes 7 single nucleotide polymorphisms(SNPs) that are associated with breast cancer (BCRAT+7). I described the assumptions for specifying the joint risk from the SNPs and the factors in BCRAT and the corresponding distribution of risks in the population, F. Using this distribution, I showed that adding SNPs improved the discriminatory accuracy, measured as the area under an ROC–type curve, from 0.607 for BCRAT to 0.632 for BCRAT+7. To determine whether this small increase in discriminatory accuracy translates into improvements in public health decisions, I used the distribution F to compute expected losses in three applications: (1) deciding whether to take tamoxifen to prevent breast cancer; (2) deciding whether or not to recommend mammographic screening; and (3) allocating resources for screening mammography under cost constraints. For the first application, BCRAT+7 reduced expected deaths in a population of women aged 50–59 years by 0.1%, compared to BCRAT. For the second application, the improvement in expected losses from using BCRAT+7 was 0.8%, compared to BCRAT. In the third application, I assumed that there was only enough money to give mammograms to half the population and that risk assessment cost 2% as much as a mammogram. The proportions of deaths prevented, compared to giving all women mammograms, were: 0.500 if the mammograms were allocated at random; 0.632 if women are first ranked according to their BCRAT risk and then mammograms are given to those at highest risk, until the money runs out; and 0.667 if this procedure is followed with BCRAT+7 instead. The procedure based on BCRAT+7 is thus 5.5% better than that based on BCRAT, but this calculation ignored the fact that SNP measurements are currently too expensive, compared to mammography, for this risk based strategy. Based on these calculations, I concluded that SNPs do not add enough information for public health decisions to warrant their use outside the research setting.

References

[1] M.H. Gail, *Discriminatory accuracy from single–nucleotide polymorphisms in models to predict breast cancer risk*, Journal of the National Cancer Institute **100** (2008), 1037–1041.

[2] M.H. Gail, *Value of adding single–nucleotide polymorphism genotypes to a breast cancer risk model*, Journal of the National Cancer Institute **101** (2009), 959–963.

Two criteria for evaluating risk prediction models

Ruth M. Pfeiffer

We propose and study two criteria to assess the usefulness of a risk model that predicts risk of disease incidence for screening and prevention, or the usefulness of prognostic models for management following disease diagnosis. The first criterion, the proportion of cases followed, $PCF(q)$, is the proportion of individuals who will develop disease who are included in the proportion q of individuals in the population at highest risk as determined by the model. The second new complementary criterion, is the proportion needed to follow–up, $PNF(p)$, namely the proportion of the general population at highest risk as determined by the model that one needs to follow in order that a proportion p of those destined to become cases will be followed. Letting F denote the distribution of risk in the population, for a well calibrated model the distribution of risk, R, in the cases is given by $G(r) = P(R \leq r|Y = 1) = \frac{1}{\mu}\int_0^r tdF(t)$, where μ denotes the mean risk in the population. Denote the $(1-q)$th population quantile by ξ_{1-q}, then $PCF(q) = 1 - G(\xi_{1-q}) = 1 - G \circ F^{-1}(1-q) = 1 - \frac{1}{\mu}\int_0^{\xi_{1-q}} tdF(t)$. The proportion needed to follow–up is $PNF(p) = 1 - F \circ G^{-1}(1-p)$. We show the relationship of those two criteria to the Lorenz curve and its inverse, and present distribution theory for estimates of PCF and PNF. We develop new methods, based on influence functions, for inference for a single risk model, and also for comparing the PCFs and PNFs of two risk models, both of which were evaluated in the same validation data. We assess our methods using simulated data, and we compute PCF and PNF for data from a validation study for a model that predicts the absolute risk of colorectal cancer.

References

[1] R.M. Pfeiffer, M.H. Gail, *Two criteria for evaluating risk prediction models*, submitted.

[2] M.H. Gail, R.M. Pfeiffer, *On criteria for evaluating models of absolute risks*, Biostatistics **6** (2005), 227–239.

[3] C.M. Goldie, *Convergence theorems for empirical Lorenz curves and their inverses*, Advances in Applied Probability **9** (1977), 765–791.

[4] A.N. Freedman, M.L. Slattery, R. Ballard–Barbash, G. Willis, B.J. Cann, D. Pee, M.H. Gail, R.M. Pfeiffer, *A colorectal cancer risk assessment tool*, Journal of Clinical Oncology **27** (2009), 686–693.

[5] Y. Park, A.N. Freedman, M.H. Gail, D. Pee, A. Hollenbeck, A. Schatzkin, R.M. Pfeiffer, *Validation of a colorectal cancer risk prediction model among whites 50 years old and over*, Journal of Clinical Oncology **27** (2009), 694–698.

Predictive models in prostate cancer

Jeremy M. Taylor

We consider joint longitudinal–survival models for describing the pattern of PSA values and the recurrence of prostate cancer for patients treated with radiation therapy for prostate cancer. A non–linear random effects is used for the longitudinal PSA data, and a time dependent proportional hazards model is used for the recurrence of prostate cancer. The models are fit using Bayesian MCMC methods. The models are used for individual prediction of prostate cancer recurrence for patients who have a series of PSA values and wish to predict future disease progression. The predictions are implemented on a website, https://psacalc.sph.umich.edu/ The models are validated by building them on pooled datasets of over 2000 patients, and tested on two external datasets. The model validates well on one dataset, but not so well on the other. A complication with the validation is that the censoring is dependent.

Development and validation of a dynamic predictive tool derived from the joint modelling of longitudinal marker and time–to–event

Cecile Proust-Lima
(joint work with Jeremy M.G. Taylor)

Joint models for longitudinal and time–to–event data offer a natural framework to describe the risk of a clinical event according to the repeated measures of a biomarker. Based on these models, dynamic prognostic tools can be built that may be updated each time a new biomarker data is collected. In Prostate Cancer study, such tool would be of great interest since the biomarker PSA, which is routinely and repeatedly measured on patients treated by a radiation therapy, has been shown to be highly associated with Prostate Cancer recurrence. However, for now, prognostic models for recurrence of Prostate Cancer only use information collected at diagnosis. Indeed, the development of dynamic prognostic tools has been limited by the numerical complexity induced by the joint models estimation. In this talk, we propose a dynamic prognostic tool derived from a joint latent class model [3] that avoids the numerical complexity due to the shared random–effects in standard joint modelling. We show how to compute such dynamic prognostic tool and provide 95% confidence bands using an approximation of the posterior distribution of the predicted probability of event. The main problem when developing a prognostic tool, either dynamic or static, remains in its validation on external data and its comparison with other prognostic tools. In survival analysis, predictive accuracy measures were proposed to evaluate the predictive performances of prognostic tools [1, 4] and were extended to evaluate dynamic tools that can be updated during the follow–up [2, 5]. In the present work, we use some of these measures to validate the dynamic prognostic tool we proposed and compare its predictive performances with those of proportional hazard models that include either only baseline covariates or baseline covariates and the level of PSA at the point of prediction. The latter prognostic tool is derived from a landmarking analysis that directly fits the predictive model for the individuals still at risk at the landmark point. From maximum likelihood estimates of these models obtained on a large cohort of patients treated after the diagnosis of a localized prostate cancer, we evaluate the predictive ability of the derived prognostic tools on 2 independent cohorts. We show that the dynamic prognostic tool based on the joint model reduces the error of prediction (whatever the specification of its estimator and especially in the way it handles the censored data). This underlines the relevance of the dynamic prognostic tool in this context, and shows especially that the entire trajectory of PSA is of interest when predicting the risk of recurrence of Prostate cancer. The use of the dynamic prognostic tool is also illustrated at the individual level.

References

[1] E. Graf, C. Schmoor, W. Sauerbrei, M. Schumacher, *Assessment and comparison of prognostic classification schemes for survival data*, Statistics in Medicine **18** (1999), 2529-45.

[2] R. Henderson, P. Diggle, A. Dobson, *Identification and efficacy of longitudinal markers for survival*, Biostatistics **3** (2002), 33-50.

[3] H. Lin, B.W. Turnbull, C.E. McCulloch, E.H. Slate, *Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate–specific antigen readings and prostate cancer*, Journal of the American Statistical Association **97** (2002), 53-65.

[4] M. Schemper, R. Henderson, *Predictive accuracy and explained variation in Cox regression*, Biometrics **56** (2000), 249-55.

[5] R. Schoop, E. Graf, M. Schumacher, *Quantifying the predictive performance of prognostic models for censored survival data with time–dependent covariates*, Biometrics **64** (2008), 603-10.

Development and evaluation of empirical models for prediction

Tianxi Cai

Clinical trials or studies of biomarkers incremental value focus primarily on estimating average effects. However, a treatment reported to be effective may not be beneficial to all patiens. Markers shown as potentially useful for improving the prediction of clinical outcomes may not be equally useful to the entire population. Traditional approaches to evaluating treatment benefit and added value of new markers fit regression models and assess the significance of the corresponding coefficient. Such methods while useful for hypothesis testing, may have limited ability to quantify the importance of new treatment or markers due to restrictive model assumptions. In this talk I discuss various approaches to quantifying treatment/marker benefit over subpopulations indexed by predictive covariates. We propose robust procedures for evaluating treatment/marker benefit over subpopulations via two–step procedures where we employ statistical models to approximate how such benefits may change over covariate labels and subsequently use a non–parametric procedure to obtain consistent model free approach to estimate the subgroup specific benefit. Simultaneous confidence interval procedures were proposed as tools for identifying subgroups that may or may not benefit from the new treatment/marker.

Quantifying the uncertainty of individual risk predictions

Thomas Gerds

The performance of risk prediction models can be estimated and compared using bootrap–crossvalidation. In praxis, this often yields similar performances of several modelling strategies. However, one of the rival strategies will eventually perform best — on the population average. For patient individual predictions it may be of interest to know if different models have different prediction variability. Confidence levels can be obtained at patient individual characteristics, from the same bootrap–crossvalidation procedure which was used to assess prediction performance. The ideas are illustrated with examples in fertility and stroke prediction.

Disease progression models: assessing their relevance for individualized medicine

Jörg Rahnenführer

Human tumors are often associated with typical genetic events. The identification of characteristic pathogenic routes in such tumors can improve the prediction of survival times and help choosing optimal therapies. We have developed models for estimating pathways of chromosomal alteration from cross–sectional data. Such models can be validated both statistically and biologically. The model further allows the introduction of a genetic progression score (GPS) that quantifies univariately the progression status of a disease. The clinical relevance was shown for various tumor entities. We present a simulation study that examines model stability and the necessary sample size for recovering the true relationship between genetic progression and survival.

Testing the prediction error difference between predictors

Mark Van de Wiel

We develop an inference framework for the difference in errors between two prediction procedures. These two procedures may differ in any aspect and possible utilize different sets of covariates. We apply training and testing on the same data set which is accommodated by sample splitting. For each split, both procedures predict the response for the same samples, which results in peaked residuals to which a signed–rank test is applied. Multiple splits result in multiple p–values. The median p–value and the mean inverse Normal one are proposed as summary test statistics, for which we prove bounds on the type I error rate. Simulation studies confirm the conservative nature of these bounds. Moreover, the multi–split approach has superior power w.r.t. the single split approach. Our inference framework is applied to genomic–survival data sets to study two issues: compare lasso and ridge regression; and decide upon use of both methylation and gene expression markers as the latter only. In the latter case, significance was established, which should support the use of the more expensive prediction method that uses both marker types.

Model dependence of statistical predictions

John Copas

Most statistical methods assume that the date are randomly sampled from some fixed model. In practice, however, there is uncertainty in the model as well as in the data. Before using a model, it is standard practice to check its goodness of fit to the data. We may forget that for any given set of data there will be many other models which also fit the data just as well. Our interest is to see whether such models all give the same inference (model rebustness) or a wide variety of inferences (model dependence).

We develop a simple theory to illustrate these points, based on confidence intervals for a specific parameter of interest. We study the union of confidence intervals from all models which would be accepted by goodness–of–fit test G, and then minimize the length of this union interval by optimizing over G. The resulting interval is similar to the non–parametric confidence interval based on the data alone, but with the variance parameter doubled. This surprisingly simple result raises questions about whether statisticians should adopt a 'worse case' or a 'concensus' view of model uncertainity, and whether models must necessarily invoke assumptions which go beyond the information in the data.

Semiparametric mixed models with Dirichlet process mixture and P–spline priors

Ludwig Fahrmeir

Longitudinal data often require a combination of flexible trends and individual–specific effects. We propose a fully Bayesian MCMC approach, using (Bayesian) P–splines for modelling nonlinear trends, while a Dirichlet process mixture specification allows for an adaptive amount of deviation from normality of random effects. We investigate properties through a simulation study and present an application to childhood obesity.

Nonparametric predicitve inference via Bayesian additive regression trees

Edward I. George

(joint work with Hugh Chipman, Robert McCulloch)

Consider the canonical regression setup where one has data on y, a variable of interest, and $x_1, \ldots, x_p$, p potential predictor variables. For the general purpose of discovering the form of $f(x_1, \ldots, x_p) \equiv E(Y \mid x_1, \ldots, x_p)$ and making predictive inference about a future y, we develop a Bayesian 'sum–of–trees' model where each tree is constrained by a regularization prior to be a weak learner, and fitting and inference are accomplished via an iterative Bayesian backfitting MCMC algorithm that generates samples from a posterior. Effectively, BART is a nonparametric Bayesian regression approach which uses dimensionally adaptive random basis elements. Motivated by ensemble methods in general, and boosting algorithms in particular, BART is defined by a statistical model: a prior and a likelihood. This approach enables full posterior inference including point and interval estimates of the unknown regression function as well as the marginal expects of potential predictors. By keeping track of predictor inclusion frequencies, BART can also be used for model free variable selection. BART's many features are illustrated with a bake–off against competing methods on 42 different data sets, with a simulation experiment and on a drug discovery classification problem.

Modelling interactions with continuous covariates

Willi Sauerbrei
(joint work with Patrick Royston)

In regression models continuous predictors are often either categorized or linearity is assumed. However, both approaches can have major disadvantages and modelling non–linear functions may improve the fit. The multivariable fractional polynomial (MFP) approach determines simultaneously a suitable functional form and deletes uninfluential variables [2, 3]. Extensions of MFP have been developed to investigate for interactions of continuous covariates with treatment (or more generally with a categorical variable, MFPI) and for two continuous covariates (MFPIgen). Both strategies allow to adjust for other covariates when investigating for interactions. After an introduction to MFP the two interaction approaches will be illustrated in two large studies analyzed with the Cox–model and respectively the logistic model.

References

[1] P. Royston, W. Sauerbrei, *A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials*, Statistics in Medicine **23** (2004), 2509–2525.

[2] P. Royston, W. Sauerbrei, *Multivariable Model–Building — A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables* (2008), Wiley.

[3] W. Sauerbrei, P. Royston, H. Binder, *Selection of important variables and determination of functional form for continuous predictors in multivariable model building*, Statistics in Medicine **26** (2007), 5512–5528.

[4] W. Sauerbrei, P. Royston, K. Zapien, *Detecting an interaction between treatment and a continuous covariate: a comparison of two approaches*, Computational Statistics and Data Analysis **51** (2007), 4054–4063.

Evaluating point forecasts

Tilmann Gneiting

Typically, point forecasting methods are compared and assessed by means of an error measure or scoring function, such as the absolute error or squared error. The individual scores are then averaged over forecast cases, to result in a summary measure of the predictive performance. I demonstrate that this common practice can lead to grossly misguided influences, unless the scoring function and the forecasting task are carefully matched.

Effective point forecasting requires that the scoring function be specified ex ante, so that the forecaster can employ the Bayes predictor, or that the forecaster receives a directive in the form of a statistical functional, for which the scoring function is consistent, in the sense that the expected score is minimized by following the directive. Expectations, ratios of expectations, quantiles and expectiles allow for an explicit characterization of the respective consistent scoring functions, which can be understood as a Choquet representation.

Motivating ridge regression

Jelle Goeman

We consider the problem of shrinkage. Ridge regression, a well–known shrinkge method, is often motivated by a bias–variance trade–off argument. By a toy example we show that the same bias–variance trade–off argument can be used to motivate other methods that do not, like ridge regression, shrink towards zero, but towards any other value. To complement this motivation, we propose an alternative argument that bounds the norm of the regression coefficients on the basis of an argument using the marginal distribution of the response Y and the predictor variables $\mathbf{X}$.

Measures of prediction error for survival data with longitudinal covariates

Erika Graf

(joint work with Rotraut Schoop)

Prognostic models play an important role in medical research. They can be useful to establish a link between a patient's characteristics and patient's future survival, and also for example for risk classification or therapy assignment. It is well known that point predictions of survival endpoints are hardly of any 'real use', due to the inherent variability of human survival. However, probabilistic prediction of survival chances may still be valuable, and their prediction error should be assessed. To develop an adequate measure of predictive error in a survival context with longitudinal covariates, these issues should be taken into account. *i)* Misspecification should be picked up, *ii)* the chronological order in which information (covariates and survival) arises should be accounted for and *iii)* scoring should be dealt with appropriately. Estimation of the Brier Score with an inverse probability of censoring weighting has the three desired properties and is therefore recommended as a measure of prediction error. Simulation studies show that the estimator is centered around the true parameter with reasonable variability, depending on the percentage of censored observation.

Score regression: detect miscalibration of normal probability forecasts

Kaspar Rufibach
(joint work with Leonhard Held and Fadoua Balabdaoui)

Typically, calibration of probabilistic forecasts is assessed via looking at probability integral transformation (PIT). We propose a new approach based on scoring rules. Specifically, for a normal predictive distribution one can compute the expectation of the Log– and Continuous ranked probability Score and set up a linear regression model based on the relationship between the expected scores and the predictive standard deviation. We illustrate the new and some further approaches in two case studies and show that score regression on either score substantially outperform existing methods, in particular PIT, in terms of power to detect miscalibration.

Predictive model selection in linear mixed–effects models

Julia Braun
(joint work with Leonhard Held)

Considering predictions for longitudinal data, there are two possible aims: either predictions for future timepoints of an individual that is already included in the data set or the prediction of a whole trajectory of a new individual can be desired. Performing choice of linear mixed–effects models with serial correlation is a challenging task in both of these situations. Apart from the selection of covariates, also the choice of the random effects and the residual correlation structure should be possible. The application of classical model choice criteria such as AIC or BIC is not obvious, and many versions do exist. We propose a predictive cross–validation approach to model choice which makes use of the logarithmic and the continuous ranked probability score [3, 2]. In contrast to full cross–validation, the model has to be fitted only once, which enables fast computations, especially for large data sets [4]. The proposed methodology is applied to search for the best model to predict HIV progression based on CD4+ count data obtained from the Swiss HIV Cohort Study (SHCS).

References

[1] J. Braun, L. Held, *Predictive cross–validation for choice and assessment of linear mixed–effects models with application to HIV progression data*, (2010), Technical Report, Institute for Social and Preventive Medicine, Biostatistics Unit, University of Zurich.

[2] T. Gneiting, F. Balabdaoui, A.E. Raftery, *Probabilistic forecasts, calibration and sharpness*, Journal of the Royal Statistical Society, Ser. B **69** (2007), 243–268.

[3] T. Gneiting, A.E. Raftery, *Strictly proper scoring rules, prediction and estimation*, Journal of the American Statistical Association **102** (2007), 359–378.

[4] E. Marshall, D. Spiegelhalter, *Approximate cross–validatory predictive checks in disease mapping models*, Statistics in Medicine **22** (2003), 1649–1660.

On the prognostic value of survival models with application to gene expression signatures

Thomas Hielscher

(joint work with Manuela Zucknick, W. Werft, Axel Benner)

As part of the validation of any statistical model, it is good statistical practice to quantify the prediction accuracy and amount of prognostic information represented by the model; this includes gene expression signatures derived from high–dimensional microarray data. Several approaches exist for right–censored survival data measuring the gain in prognostic information compared to established clinical parameters or biomarkers in terms of explained variation or explained randomness. They are either model–based or use estimates of prediction accuracy. As these measures differ in their underlying mechanisms, they vary in their interpretation, assumptions and properties, in particular in how they deal with the presence of censoring. It remains unclear, under which conditions and to which extent they are comparable. We present a comparison of several common measures and illustrate their behaviour in high-dimensional situations in simulation examples as well as in an application to a real gene expression microarray data set.

Geostatistical model averaging

Will Kleiber

We introduce new methodology to produce calibrated and sharp predictive distributions for temperature based on an ensemble of forecasts. The method extends the Bayesian model Averaging approach of Raftery et al. (2005) to allow for locally varying statistical parameters. We view the bias correction and predictive variance as spectral Gaussian processes. In an example for 48 hour ahead temperature forecasts, the method produces locally calibrated predictive distributors which are significantly sharper than a global parameter approach.

References

[1] Raftery, A.E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005). *Using Bayesian Model Averaging to Calibrate Forecast Ensembles.* Monthly Weather Review, **133**, 1155–1174.

Using data of clinical trials to explore individualized therapies in patients with colorectal cancer

Rüdiger P. Laubender

Responder analysis is a statistical concept often required by physicians. It is used for modelling individual response to drug treatment in fields like pharmacogenetics, pharmacogenomics, individualized therapies and oncology. However, individual response to treatment cannot be established from data generated by an experiment using a parallel–group design. In such a design we observe for a patient only one of the potential outcomes, either Y_1 (outcome under active treatment) or Y_0 (outcome under control treatment or placebo), but not both simultaneously. Hence, we cannot observe the individual treatment effect for any patient.

This fact is known in the field of causality based on counterfactuals as the 'Fundamental Problem of Causalitiy' [1]. To solve this problem, Holland provides two 'solutions': Two scientific solutions and two statistical solutions. Essentially, the problem is to identify the joint distribution of the potential outcomes Y_1 and Y_0 from two marginal distributions, each for Y_1 and Y_0 obtained from a parallel–group design. One of Holland's statistical solutions is the assumption of constant treatment effect for all patients. This assumption can be relaxed by conditioning on prognostically important covariates and covariate–by–patient interactions in order to approximate individual treatment effects. An example for this is given developed within the framework of linear models [2, 3].

It is planned to use data from three oncologic trials to explore the limitations of identifying individual treatment effects. Further, the methodology for identifying the joint distribution of treatment outcomes from two marginal distributions will be extended to the framework of logistic regression and regressions for survival data.

References

[1] P.W. Holland, *Statistics and causal inference*, Journal of the American Statistical Association **81** (1986), 945–960.

[2] G.L. Gadbury, H.K. Iyer, *Unit–treatment interaction and its practical consequences*, Biometrics **56** (2000), 882–885.

[3] G.L. Gadbury, H.K. Iyer, D.B. Allison, *Evaluating subject–treatment interaction when comparing two treatments*, Journal of Biopharmaceutical Statistics **11** (2001), 313–333.

Predictive assessment of Bayesian hierarchical models

Daniel Sabanes-Bove

(joint work with Leonhard Held, Ludwig Fahrmeir)

Bayesian hierarchical models are increasingly used in many applications. In parallel, the desire to check the predictive capabilities of these models grows. However, classic Bayesian tools for model selection, as the marginal likelihood of the models, are often unavailable analytically, and the models have to be estimated with MCMC methodology. This also renders leave–one–out cross–validation of the models infeasible for realistically sized data sets. In this talk we therefore propose approximate cross–validation sampling schemes based on work by Marshall and Spiegelhalter [2], for two model classes: conjugate change point models are applied to time series, while general linear mixed models are used to analyze longitudinal data. The quality of the models' predictions for the left–out data is assessed with calibration checks and proper scoring rules. In several case studies we experienced that the approximate cross–validation results are typically close to the exact cross–validation results, and are much better suited for predictive model assessment than analogous posterior–predictive results, which can only be used for goodness–of–fit checks. One case–study on the Nile discharge data [1] is presented in the talk, and an application with childhood obesity data [3] where the exact leave–one–out scheme is infeasible demonstrates the practical use of the approximate method.

References

[1] G.W. Cobb, *The problem of the Nile: conditional solution to a changepoint problem*, Biometrika **65** (1978), 243–251.

[2] E.C. Marshall, D.J. Spiegelhalter, *Approximate cross–validatory predictive checks in disease mapping models*, Statistics in Medicine **22** (2003), 1649–1660.

[3] N. Fenske, L. Fahrmeir, P. Rzehak, M. Hardle, *Detection of risk factors for obesity in early childhood with quantile regression methods for longitudinal data*, (2008), Technical Report 38, Department of Statistics, University of Munich.

Independence screening for high–dimensional prognostic Cox models

Manuela Zucknick
(joint work with Thomas Hielscher, Axel Benner)

In clinical applications of high–throughput technologies for generating genomic data, one aim is the development of prognostic models for patient survival. In this context, two — possibly competing — objectives exist. While a model should be as useful as possible for survival prognosis, it should also be small, i.e. only contain a small set of genomic variables relevant for prognosis to enable biological interpretation.

Sparse penalised likelihood methods are well suited for this task. As examples of this class of models, we chose to study the lasso and SCAD, because of their known asymptotic properties of model consistency and in the case of SCAD — under certain assumptions — also asymptotic unbiasedness. While this so–called oracle property makes SCAD an attractive penalty choice in certain — very sparse — data settings, it does require an initial screening step to reduce the number of potential predictors to a number smaller than the sample size. In the context of linear models, Fan and Lv [1] proposed to use simple marginal statistics for screening, proving the sure screening property in the case of independent data.

We investigated in simulation studies, how well the sure screening property holds, if this unrealistic assumption of independence is violated. We compared several adaptations of Fan and Lv's method to the setting of Cox models, since we are interested in modelling patient survival. Because the ultimate aim is to fit models with good prognostic value, we also studied the effect of the screening methods on the prognostic value of the final models, as measured by the probability to reject the global null hypothesis that all regression coefficients are zero under various alternatives.

References

[1] J. Fan, J. Lv, *Sure independence screening for ultrahigh dimensional feature space*, Journal of the Royal Statistical Society, Series B: Statistical Methodology **70** (2008), 849–911.

Competing risks and time–dependent covariates

Per K. Andersen

(joint work with Giuliana Cortese)

In survival analysis, hazard regression models enable inclusion of time–dependent covariates. However, when including internal/endogeneous covariates, the one–to–one correspondance between hazard function and cumulative survival probability is lost. This problem persists in models for competing risks. In this paper we explore some methods by which cumulative incidences in competing risks models may be estimated in the presence of an internal time–dependent covariate. One method is based on an extension of the competing risks multi–state model while the other builds on van Houwelingen's 'landmark' approach [1]. Data from a bone marrrow transplantation study are used to illustrate the techniques.

References

[1] H.C. van Houwelingen, *Dynamic Prediction by Landmarking in Event History Analysis*, Scandinavian Journal of Statistics **34** (2007), 70–85.

Choice of prognosis estimators based on Kullback–Leibler risk

Daniel Commenges

A general approach to statistical inference can be based on Kullback–Leibler risk. A model (g) is a family of distributions $(g^{\beta})_{\beta\in B}$. The maximum likelihood estimator minimizes an estimator of the Kullback–Leibler risk. The maximum likelihood estimator converges toward the distribution g^{β_0} which has the minimum Kullback–Leibler risk in (g). We can define the expected Kullback–Leibler risk for it: $EKL(g^{\hat{\beta}_n})$. Akraike criterion is an estimator of this risk, up to a constant. A normalized difference of Akaike criterions estimates the difference of EKL between two models and an an asymptotic distribution can be given. In regression problems two cases can be distinguished. In the standard case Akaike criterion can be used because it can be considered as applying to a 'reduced model', which is a joint model in which the marginal distribution of the explanatory variable is known.

In complex prognosis models where a latent process is assumed to link the marker values and the event to be predicted, a corrected version of Akaike criterion must be developed. An adapted cross–validation criterion can also be proposed. Both criterions are asymptotically equivalent.

A Measure of Explained Variation for Survival Models

Janez Stare

Coefficient of determination, usually denoted as R^2, is a popular measure of the overall performance of a linear regression model. It is based on the fact that a measure of variation, sum of squared distances from the mean in this case, can be decomposed into the explained and unexplained part. R^2 is then simply a ratio of the explained part and the total variation, thereby giving us an intuitively attractive measure of the proportion of explained variation. It is then natural that similar measures are sought for other models. Unfortunately, the simple decomposition doesn't work in nonlinear models. A search for some measure of explained variation, or something similar, in survival analysis has been going on for almost 30 years. First intuitive attempts were followed by papers trying to fix the statistically non–desirable properties, dependency on censoring being the most obvious.

In this work we present a new approach, based on ranks, in which our goal is to have a measure which can handle time dependent effects and covariates, while having all the necessary statistical properties. At each event time, we calculate the differences between the predicted rank of a failed subject under the null and the fitted model, and under the null and the perfect model. These differences are then summed up over all event times, and the first sum is the divided by the second, giving our measure. The contributions to the sum are appropriately weighted to make the measure independent of censoring. We can give the measure the explained variation interpretation in the sense that variation is any measure of the degree to which a distribution is not degenerate. We provide the population value and a variance estimator. A side effect of our approach is that the measure contains the popular C index as a special case when there is no time dependency. We are reluctant to call our measure a generalization of the concordance index, as concordance doesn't make much sense in the presence of time dependent covariates and/or effects.

Obtaining probabilistic forecasts from an ensemble of point forecast

Thordis L. Thorarinsdottir

Ensembles of point forecasts appear in many different applications: ensembles of dynamical weather prediction models have been developed, in which multiple estimates of the current state of the atmosphere are used to generate a collection of deterministic forecasts; in economics, large surveys are regularily conducted, where experts or non–experts give their past estimates for future values of a large number of economic variables. However, such ensemble systems are often uncalibrated and biased. We propose a novel way of statistically post–processing ensemble forecasts by using heteroskedastic regression which allows for censoring and/or asymmetry in the resulting predictive distribution. The results show a substantial imporovement in the predictive performance over the unprocessed ensembles.

A new strategy for meta–analysis of continuous covariates in observational studies

Patrick Royston

While meta–analyses of data from randomized controlled trials is well–established and familiar to statisticians and many health professionals, meta–analysis of dose–response relationships in observational studies is much less developed. We present a suggestion for meta–analysing the influence of a continuous covariate in an epidemiological or prognostic setting. We model the effect of the covariate using fractional polynomical functions in each study, and combine estimated functions using pointwise weighted averaging. Confounders, which may differ across studies, are modelled per study and their linear predictors are used to adjust for confounding in each study. Fixed or random–effects models can be used. We apply the methods, as a motivating example, to registry data from the US SEER database for breast cancer patients, showing how the prognostic influence of number of positive lymph nodes and of age can be usefully summarised across registries.

Study designs for evaluating putative predictive markers

Patrick Bossuyt

We argue that prediction models should be evaluated from a conceptualist perspective, not an essentialist one. We discussed study designs and metrics to do so. The key issue is to look at predictive markers and models on treatment selection indicators.

Developing valid prediction models: a proposal for a framework with 7 steps

Ewout W. Steyerberg

Prediction models are increasingly developed in many medical fields, including cancer, cardiovascular disease, and many others. Methodological reviews have consistently shown many shortcomings in currently published models.

A number of requirements need to be fulfilled to develop a valid model. First, predictors need to be available that have a strong relationship with the outcome. Only then subjects with and without the outcome can be discerned. Next, an adequate sample size needs to be available. Third, a sensible modelling strategy needs to be followed that systematically considers a number of steps, such as dealing with missing predictor values, transformation of continuous variables, selection of predictors and model specification, concern for overoptimistic estimation of effects, validation, and presentation of the final model. Ideally, a developed model is subsequently externally validated in another setting to assess its generalisability, and eventually tested for clinical impact. This final application may require simplification of a model to a simple decision rule.

I propose a framework including 7 logically distinct steps for prognostic modelling, and illustrate these steps in a case study of patients with an acute myocardial infarction. The framework may not only be useful for model development, but also to critique developed models, to guide reporting of models, and to define issues to be addressed in protocols for prediction models.

References

[1] E.W. Steyerberg, . *Clinical prediction models: a practical approach to development, validation, and updating*, (2009), New York, Springer.

Dynamic prediction of cure

Hans C. van Houwelingen

Cure models for survival data divide the patients in two groups: those who are cured by the treatment and will not die from the disease and those who are not cured and might die from the disease. Since the latter might not die within the follow–up period, it is impossible to estimate the fraction cured unless very strong assumptions are made on the shape of the survival function. The paper proposes an operational definition of cure related to the probability of dying within a fixed window of $w = 5$ or 10 years. Patients are declared to be cured if that probability is small enough. It is shown how the probability of dying within w years can be assessed dynamically during the follow–up using the landmark methodology of Van Houwelingen [1].

References

[1] H.C. van Houwelingen, *Dynamic Prediction by Landmarking in Event History Analysis*, Scandinavian Journal of Statistics **34** (2007), 70–85.

Predicting survival from genomic data

Ornulf Borgan

Six methods for building survival prediction models from clinical covariates and gene expression measurements were discussed. The methods, all based on Cox's regression model, were variable selection, unsupervised and supervised principal components regression and partial least squares regression, ridge regression, and the lasso. For all methods it was described how one may combine classical clinical covariates with genomic data in a clinico–genomic prediction model using both types of covariates, but applying dimension reduction only to the high–dimensional genomic variables. The performance of the methods were compared using three data sets, and the comparison showed that ridge regression had the best performance and univariate selection had the poorest performance among the six methods [3]. These conclusions were not dependent on which one of three commonly used criteria that was applied to assess the prediction performance of the models [2]. Finally it was pointed out that simulation studies are of limited use when comparing the performance of methods for building survival prediction models, and that studies of real data sets are to be preferred [1].

References

[1] O. Borgan, H. Bovelstad, *The role of simulations in studying methods for survival prediction from gene expression data*, (2010), Manuscript in preparation, Department of Mathematics, University of Oslo.

[2] H.M. Bovelstad, O. Borgan, *Assessment of evaluation criteria for survival prediction from genomic data*, (2010), Manuscript, Department of Mathematics, University of Oslo.

[3] H.M. Bovelstad, S. Nygard, O. Borgan, *Survival prediction from clinico–genomic models — a comparative study*, BMC Bioinformatics **10** (2009), article 413.

Genetic interactions

Wolfgang Huber

Individuals within a population vary across many phenotypes. For instance, humans vary with respect to their susceptibility to different diseases, cancers vary with respect to their response to drugs. There are both genetic and environmental contributions to this variation. Typically, the resulting phenotypes are complex, combinatorial functions of the genetic and environmental variables. Broadly speaking, the aim of my research is to dissect these different sources of variation. Specifically, I describe an experiment that systematically explores pairwise interactions of gene perturbations by RNA interference on the growth rate and cell cycle in Drosophila cell lines. These data pose some interesting challenges to statistical analysis and modelling. They provide unprecedented insights on the modular architecture of cellular processes, and allow to place individual genes on a functional map.

Signaling, drugs, and cancer

Rainer Spang

Tumours arise from dysfunctional cellular communication. Normal cells grow when receiving growth signals. Tumour cells grow without these signals. They proliferate although they should not. They receive signals of cell death but do not respond to them properly. They can escape immune responses by sending out signals that modulate the immune system. In summary, a tumour can only survive as a tumour, if it perturbs the molecular communication in cells and between cells.

The molecular survival mechanisms in turn are Achilles heals of tumours, making them targets for cancer treatment. An obstacle is that different tumours activate different survival pathways. While the survival mechanism of somebody's breast cancer might be similar to that of someone else's colon tumour, different tumours of the same entity can use completely different survival strategies. How can we find the molecular weaknesses of an individual tumour?

Genomic high through put data allows us to monitor the molecular makeup of tumours. Survival mechanisms leave their traces in this data forming characteristic patterns. I described some novel statistical approaches to associate patterns in high dimensional genomic data with molecular mechanisms of tumour survival. Moreover, I discussed a network inference method that can be used to model the flow of information in cells bases on the nested structure of intervention effects.

Technological advances in genomics and their impact for personalized medicine

Julien Gagneur

I will review technological developments of the last 5 years in genomics and molecular biology with potential impact on personalized medicine. Four technologies appear to become likely major players. High–throughput sequencing gives access to individual genotypes as well as dynamic molecular states of the cell such as transcription or chromatin modifications. Synthetic biology will enable directed explorations of genetic variations. Microfluidics extends by several orders of magnitudes the amount of samples handled in a single experiment. Finally, the generation of induced pluripotent stem cells will give flexible access to patient–specific cells of any type.

Genetic diversity of pathogen populations within patients

Niko Beerenwinkel

Many human diseases are the result of evolving pathogens, including cancer cells in a tumor and infectious parasites, such as bacteria and viruses. Treatment of these constantly changing ensembles of individuals is complicated by evolutionary escape from the selective pressure of drugs and immune responses. We present statistical and computational tools for estimating the diversity of a pathogen population from next–generation DNA sequencing data. We employ a infinite–dimensional probabilistic clustering method based on the Dirichlet process mixture in order to separate technical sequencing errors from true biological variation and to reconstruct the haplotype structure of the population.

High–dimensional Cox models: the choice of penalty as part of the model building process

Axel Benner
(joint work with Manuela Zucknick, Thomas Hielscher, Carina Ittrich, Ulrich Mansmann)

The Cox proportional hazards regression model is the most popular approach to model covariate information for survival times. In this context, the development of high–dimensional models where the number of covariates is much larger than the number of observations ($p >> n$) is an ongoing challenge. A practicable approach is to use ridge penalized Cox regression in such situations. Beside focussing on finding the best prediction rule, one is often interested in determining a subset of covariates that are the most important ones for prognosis. This could be a gene set in the biostatistical analysis of microarray data. Covariate selection can then, for example, be done by L1–penalized Cox regression using the lasso.

Several approaches beyond the lasso, that incorporate covariate selection, have been developed in recent years. This includes modifications of the lasso as well as non–convex variants like SCAD. The purpose of our paper is to implement them practically into the model building process when analyzing high–dimensional data with the Cox proportional hazards model.

To evaluate penalized regression models beyond the lasso we included SCAD variants and the adaptive lasso. We compare them with 'standard' applications like ridge regression, the lasso, and the elastic net. Predictive accuracy, features of variable selection, and estimation bias will be studied to assess the practical use of these methods.

We observed that the performance of SCAD and adaptive lasso is highly dependent on non–trivial pre–selection procedures. A practical solution to this problem does not yet exist. Since there is high risk of missing relevant covariates when using SCAD or adaptive lasso applied after an inappropriate initial selection step, we recommend to stay with lasso or the elastic net in actual data applications. But with respect to the promising results for truly sparse models we see some advantage of SCAD and adaptive lasso, if better pre–selection procedures would be available. This requires further methodological research.

Populational inference in presence of non–ignorable missing data with individualized modelling prediction of latent variables

Haiqun Lin

We propose a method of using latent variable prediction for population inference of longitudinal data in the presence of non–ignorable missing data. Under the situation of missing completely at random, generalized estimating equation (GEE) provide a valid approach (Liang & Zeger 1986). Under the situation of missing at random, the weighted GEE proposed by Robins and colleagues is a valid approach (Robins, Rotnetzky and Zhu 1993). For non–ignorable missing data, we interpret the latent variable modelling approach as an individualized prediction of probability of not missing into weighted GEE. Our estimation of population parameter is consistent. Our method is evaluated with simulation study and illustrated with a longitudinal clinical trial data with a large fraction of dropouts.

High dimensional predictive inference: a decision theoretic perspective

Edward I. George
(joint work with Larry Brown, Feng Liang, Xinyi Xu)

Let X and Y be independent multivariate normal vectors with a common unknown mean. Based on only observing X, we consider the problem of obtaining a predictive density that is close to unknown true density of Y as measured by expected Kullback–Leibler loss. This is the predictive version of the general problem of estimating a multivariate normal mean under quadratic loss, and we will see that a strikingly parallel theory exists for addressing it. To begin with, a natural 'straw man' procedure for this problem is the (formal) Bayes predictive density under the uniform prior which is best invariant and minimax. It turns out that there are wide classes of procedures that dominate this straw man including Bayes predictive densities under superharmonic priors. For the characterization of admissible procedures for this problem, the class of all generalized Bayes rules here is seen to form a complete class, and easily interpretable conditions are seen to be sufficient for the admissibility of a formal Bayes rule. Moving on to the multiple regression setting, our results are seen to extend naturally. Going further, we develop minimax multiple shrinkage predictive estimators for the situation where there is model uncertainty and only an unknown subset of the predictors is thought to be potentially irrelevant.

Reporter: Ulrich Mansmann

Participants

Prof. Dr. Per Kragh Andersen
Department of Biostatistics
University of Copenhagen
Oster Farimagsgade 5
DK-1014 Kobenhavn K

Prof. Dr. Niko Beerenwinkel
ETH Zürich
D-BSSE
Computational Biology Group (CBG)
Mattenstr. 26
CH-4058 Basel

Axel Benner
Deutsches Krebsforschungszentrum (DKFZ)
Abtlg. Biostatistik (C060)
Im Neuenheimer Feld 581
69120 Heidelberg

Prof. Dr. Ornulf Borgan
Department of Mathematics
University of Oslo
P.O.Box 1053
Blindern
N-0316 Oslo

Prof. Dr. Patrick M. Bossuyt
Academic Medical Center
University of Amsterdam
Dept. of Clinical Epidem. & Biostat.
PO Box 22700
NL-1100 DE Amsterdam

Julia Braun
Abteilung Biostatistik, ISPM
Universität Zürich
Hirschengraben 84
CH-8001 Zürich

Prof. Dr. Tianxi Cai
Department of Biostatistics
Harvard School of Public Health
655 Huntington Avenue
Boston MA 02115
USA

Prof. Dr. Saskia le Cessie
Department of Medical Statistics
University of Leiden
Postbus 9600
NL-2300 RC Leiden

Dr. Daniel Commenges
INSERM U 897
Universite de Bordeaux 2
146 Rue Leo Saignat
F-33076 Bordeaux Cedex

Prof. Dr. John Copas
Department of Statistics
University of Warwick
GB-Coventry CV4 7AL

Prof. Dr. Ludwig Fahrmeir
Institut für Statistik
Universität München
Ludwigstr. 33
80539 München

Dr. Julien Gagneur
EMBL Heidelberg
Meyerhofstr. 1
69117 Heidelberg

Dr. Mitchell H. Gail
Division of Cancer Epidemiology and Genetics, National Cancer Institute
Executive Plaza South, Room 8030
6120 Executive Blvd.
Bethesda MD 20892-7244
USA

Prof. Dr. Edward I. George
Department of Statistics
The Wharton School
University of Pennsylvania
3730 Walnut Street
Philadelphia , PA 19104-6340
USA

Prof. Dr. Thomas A. Gerds
Department of Biostatistics
University of Copenhagen
Oster Farimagsgade 5
DK-1014 Kobenhavn K

Prof. Dr. Tilmann Gneiting
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg

Prof. Dr. Jelle Goeman
Leiden University Medical Centre
Medical Statistics
Postbus 9600
NL-2300 RC Leiden

Dr. Erika Graf
Universitätsklinikum Freiburg
Zentrum Klinische Studien
Elsässer Str. 2
79110 Freiburg

Prof. Dr. Leonhard Held
Abteilung Biostatistik, ISPM
Universität Zürich
Hirschengraben 84
CH-8001 Zürich

Prof. Dr. Robin Henderson
School of Mathematics and Statistics
The University of Newcastle
GB-Newcastle upon Tyne NE1 7RU

Thomas Hielscher
Deutsches Krebsforschungszentrum (DKFZ)
Abtlg. Biostatistik (C060)
Im Neuenheimer Feld 581
69120 Heidelberg

Prof. Dr. Hans van Houwelingen
Leiden University Medical Center
PO Box 576
NL-3720 AN Bilthoven

Prof. Dr. Wolfgang Huber
European Bioinformatics Institute
EMBL Outstation - Hinxton
GB-Cambridge CB10 1SD

Will Kleiber
Department of Statistics
University of Washington
Box 35 43 22
Seattle , WA 98195-4322
USA

Rüdiger Laubender
Institut für Medizinische
Informationsverarbeitung, Biometrie
und Epidemiologie
Marchioninistr.15
81377 München

Prof. Dr. Haiqun Lin
Department of Epidemiology and
Public Health
Yale University
60 College Street
New Haven CT 06520-8034
USA

Prof. Dr. Ulrich Mansmann
Institut für Medizinische
Informationsverarbeitung, Biometrie
und Epidemiologie
Marchioninistr.15
81377 München

Dr. Ruth Pfeiffer
Division of Cancer Epidemiology and
Genetics, National Cancer Institute
Executive Plaza South, Room 8030
6120 Executive Blvd.
Bethesda MD 20892-7244
USA

Cecile Proust-Lima
INSERM U 897
Universite de Bordeaux 2
146 Rue Leo Saignat
F-33076 Bordeaux Cedex

Prof. Dr. Jörg Rahnenführer
Fakultät für Statistik
Technische Universität Dortmund
44221 Dortmund

Prof. Patrick Royston
MRC Clinical Trials Unit
222 Euston Road
GB-London NWI 2DA

Dr. Kaspar Rufibach
Abteilung Biostatistik, ISPM
Universität Zürich
Hirschengraben 84
CH-8001 Zürich

Daniel Sabanes Bove
Abteilung Biostatistik, ISPM
Universität Zürich
Hirschengraben 84
CH-8001 Zürich

Prof. Dr. Wilhelm Sauerbrei
Institut für Medizinische Biometrie
und Medizinische Informatik
Klinikum der Universität
Stefan-Meier-Str. 26
79104 Freiburg

Prof. Dr. Michael Schemper
Zentrum für Medizinische Statistik,
Informatik u. Intelligente Systeme
Medizinische Universität Wien
Spitalgasse 23
A-1090 Wien

Prof. Dr. Martin Schumacher
Institut für Medizinische Biometrie
und Medizinische Informatik
Klinikum der Universität
Stefan-Meier-Str. 26
79104 Freiburg

Dr. Lene Theil Skovgaard
Department of Biostatistics
University of Copenhagen
Oster Farimagsgade 5
DK-1014 Kobenhavn K

Prof. Dr. Rainer Spang
Universität Regensburg
Institut für Funktionelle Genomik
Josef-Engert-Str. 9
93053 Regensburg

Prof. Dr. Janez Stare
University of Ljubljana
Institute for Biomedical Informatics
Vrazov trg 2
1104 Ljubljana
SLOVENIA

Prof. Dr. Ewout Steyerberg
Erasmus MC
Department of Public Health
P.O.Box 2040
NL-3000 CA Rotterdam

Prof. Dr. Jeremy M.G. Taylor
Department of Biostatistics
University of Michigan
1420 Washington Heights, M4108
Ann Arbor MI 48109-2029
USA

Dr. Thordis Linda Thorarinsdottir
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg

Dr. Manuela Zucknick
Institut f. Epidemiologie u. Biometrie
Deutsches Krebsforschungszentrum
Im Neuenheimer Feld 280
69120 Heidelberg

Prof. Dr. Mark A. van de Wiel
Dept. of Epidemiology & Biostatistics
VU University Medical Center
PO Box 7057
NL-1007 MB Amsterdam