# EDITING

## WHEN WRITING FOR THE STATISTICAL SCIENCES

Don Ylvisaker

# EDITING

## EDITING

### EDITING

#### EDITING

# EDIT TO

MAINTAIN INTEREST AND ACCURACY

IMPROVE DIRECTNESS AND READABILITY

# AT EACH ITERATION CONSIDER

RETHINKING ORGANIZATION AND TEXT

READING THE TEXT ALOUD FOR ITS CADENCE

ALTERNATING DRAFTS WITH COAUTHORS

RECRUITING YOUR OWN REFEREE

# PARAGRAPHS

SELF-CONTAINED UNIT CONCERNING AN IDEA

TOO LONG?

TOO SHORT?

NECESSARY?

# SENTENCES

## VOICE – ACTIVE OR PASSIVE

WE CANNOT APPLY THE BOUND DIRECTLY OR

THE BOUND CANNOT BE DIRECTLY APPLIED

## TENSE – PRESENT, PAST, FUTURE, . . .

EXPONENTIAL SMOOTHING SPLINES ARE DEVELOPED . . . OR

EXPONENTIAL SMOOTHING SPLINES WERE DEVELOPED

# SOME SUGGESTIONS

Ten definitions is, by definition, too many. Use "define" sparingly, use "take", "let", or "write"

"Take ø = ß*∂ . . . " No need for a definition.

BE IN POSITION TO STATE YOUR MAIN POINTS CLEARLY AND SUCCINCTLY – CENTRAL ASSUMPTIONS AND NOTATION HAVING GONE BEFORE.

# WORDS AND PHRASES TO AVOID

Avoid "**THAT IS**"

" . . . whenever the right-hand side vanishes as ∂ goes to zero, that is, if there exists a ∂ . . . "

Choose between

". . . whenever the right-hand side vanishes as ∂ goes to zero."

" . . . if there exists a ∂ such that for . . . "

# WORDS AND PHRASES TO AVOID

Avoid "**IN ORDER TO**", "**LET US**", and "**FOLLOWING**"

"~~In order~~ [T]o achieve our aims, ~~let us~~ consider the[se] ~~following~~ assumptions.

(A1) . . .

Avoid "**IT IS WORTH NOTING THAT**"

"~~It is worth noting that~~ [T]he condition (A3) is not restrictive . . ."

# WORDS AND PHRASES TO AVOID

Avoid "**INTERESTINGLY**"

"~~Interestingly,~~ [O]ur result can be used to

Avoid "**NOTE THAT**"

"~~Note that~~ [T]he iteration changes the value of . . . "

# WORDS AND PHRASES TO AVOID

Avoid "**RECALL THAT**"

"Recall that our intention is to . . . "

Avoid "**WILL BE**" and "**BECOMES**"

"If we take a = 1, b ~~becomes~~ (is) a constant."

"If we take a = 1, b ~~will be~~ (=) 0"

# WORDS AND PHRASES TO AVOID

Avoid "**OF COURSE**"

"~~Of course,~~ [W]e only need to show that . . . "

Avoid "**WE CAN SEE THAT** "

"From Figure 1, ~~we can see that~~ y increases as x . . .

~~Theorem 1 describes the form of the solution and states that it exists and is unique.~~

Theorem 1: The solution takes the form ____. It exist and is unique.

We start with some regularity conditions which will be used later to establish our main theorem on the asymptotic distribution of $Y_n$ defined in (1.1).

We need some regularity conditions to establish the asymptotic distribution of $Y_n$.

THE ELLIPTICAL DISTRIBUTION FAMILY CONTAINS MUCH BROADER CLASS OF DISTRIBUTIONS THAN THE NORMAL DISTRIBUTION FAMILY, SUCH AS MIXUTRES OF NORMAL DISTRIBUTIONS, THE MULTI-UNIFORM DISTRIBUTION, THE MULTIVARIATE t-DISTRIBUTION, THE MULTI-UNIFORM DISTRIBUTION ON THE UNIT SPHERE, AND PEARSON TYPE II DISTRIBUTIONS, AMONG OTHERS.

THE ELLIPTICAL DISTRIBUTION FAMILY CONTAINS MIXTURES OF NORMALS, THE MULTI-UNIFORM, THE MULTIVARIATE t, THE MULTI-UNIFORM ON THE UNIT SPHERE, AND THE PEARSON TYPE II DISTRIBUTIONS, AMONG OTHERS.

THE CVX PROGRAM IN MATLAB (HAS) ~~SHOULD BE ABLE TO SOLVE THESE PROBLEMS BUT THERE ARE~~ SOME TECHNICAL ISSUES.

THE CHAO ESTIMATOR FOR THE NUMBER OF SPECIES ~~IS HIGHLY FAVORED BY SCIENTISTS, AND~~ PLAYS AN IMPORTANT ROLE IN CONSERVATION BIOLOGY. WE SHOW THAT (ITS) ~~THE~~ ASYMPTOTIC VARIANCE ~~OF THE CHAO ESTIMATOR~~ IS NOT ESTIMABLE, BUT . . .

# ABSTRACT 1

In the analysis of semicompeting risks data, the illness-death models provide a general frame-work to characterize the progression of a disease. In this work, a class of semiparametric models is proposed to study the event history under the illness-death model process, in which we consider that some subjects may experience “illness" then “death" but the others may experience “death" directly. The dependence between the successive events and the marginal distributions of event times are of interest. Besides, the subsequent event, death time after illness, is censored by a dependent variable related to the occurrence time of illness, and is observable only if the first event has occurred. To accommodate such dependent censoring effect, a copula model is applied for the successive events and the semiparametric transformation models are used for the marginal distributions. Unbiased estimating equations are developed through the nonparametric maximum likelihood method and their asymptotic properties follow directly from martingale theory. Simulation results indicate that the proposed method perform well under practical sample sizes. An application of clinical study with chronic myeloid leukemia is reported for illustration.

# ABSTRACT 1a

Illness-death models provide a framework for characterizing the progression of a disease In the analysis of semi-competing risks data. We propose a class of semi-parametric models to study event histories under a illness-death model in which some subject experience “illness” and then “death”, while others experience only “death”. The dependence between successive events is of interest, as is the marginal distribution of event times. In particular, death time is censored by a variable related to the occurrence time of illness, and is observed only if the first event has occurred. To accommodate the censoring effect, a copula model is employed for successive events, and semi-parametric models are used for the marginal distributions. Unbiased estimating equations are developed through a nonparametric maximum likelihood method, and their asymptotic properties follow from martingale theory. Simulation results indicate that the proposed method performs well under practical sample sizes, and an application to a chronic myeloid leukemia dataset is reported to illustrate its use.

# ABSTRACT 2

Semi-parametric modeling of interval-valued data is of great practical importance, as exampled by applications in economic and financial data analysis. We propose a flexible semi-parametric modeling of interval-valued data by integrating the partial linear regression model and the Center and Range method, and investigate its estimation procedure. Furthermore, we introduce a new test statistic that allows one to decide between a parametric linear model and a semi-parametric model, and approximate its null asymptotic distribution based on wild Bootstrap method to obtain the critical values. Simulations results are carried out to evaluate the performance of the proposed methodology and the new test. Moreover, four real datasets are analyzed to document its practical applications.

# ABSTRACT 2a

Interval-valued data are increasingly common in such fields as engineering, finance, and medicine. We propose a partial linear regression model for them. A semi-parametric test is introduced in this context, and the wild Bootstrap is applied to approximate its asymptotic distribution. Simulations have been conducted to demonstrate the effectiveness of the test, and its usefulness is suggested by applications to four data sets.

# ABSTRACT 3

The emergence of the big data era and the popularity of social media play a capital role in growing research interest of many sectors, i.e., the pattern recognition and the social networks analyses. In the real world, any social organization may be made of a wide variety of communities including schools, families, firms, etc. Communities are vital elements of every organization. And community detection has drawn much attention in business and social science particularly. Under a Poisson random graph model, the scan statistics had been considered as a useful tool to investigating the statistical significance of both structure and attribute of clusters in networks. However, the Poisson random graph assumption may not hold in all networks. In this paper, we first develop a generalized scan statistics method by considering the individual diversity of each edge. Second, we construct the random connection probability model and the logit model. To demonstrate the effectiveness of the proposed method, a simulation study conducted showed that the generalized method had better detection results, and an empirical study showed an improvement of the proposed method compared with the existing methods.

# ABSTRACT 3a

The growth of social networks, in combination with the increasing sophistication of Big Data tools, has led to a burgeoning interest in an enriched understanding of relationships among people, institutions, and more. A relevant setting for such a study is graph theory, together with its random counterpart. Poisson random graphs have been employed to investigate scan statistics that can measure the significance of clusters of nodes based on structure and attributes. Our interest is in the diversity that might be found in the edges of a graph, and we generalize scan statistics to account for it. We construct a random connection probability model and a logit model and show the effectiveness of scan statistics in cluster detection in these settings. Simulations and empirical studies suggest wide applicability of our methods.