A GUIDE TO SURVIVAL
BAYESIAN ANALYSIS OF SURVIVAL DATA

DANI GAMERMAN
Cidade Universitária, Rio de Janeiro - Brazil

MIKE WEST
Institute of Statistics & Decision Science, Duke University

ANDY POLE
Institute of Statistics & Decision Sciences, Duke University
& Department of Statistics, University of Warwick, U.K.

DP #90-B02
A GUIDE TO SURVIVAL

BAYESIAN ANALYSIS OF SURVIVAL DATA

DANI GAMERMAN\textsuperscript{1}, MIKE WEST\textsuperscript{1}, and ANDY POLE\textsuperscript{*}

Bayesian Forecasting Group
University of Warwick\textsuperscript{*}

July 1988

\textsuperscript{1}Now at: Departamento de Métodos Estatísticos, Instituto de Matemática, Caixa Postal 68530, 21944 Cidade Universitária, Rio de Janeiro — RJ, Brasil.

\textsuperscript{2}Now at: Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina, 27706, U.S.A.

\textsuperscript{*}Department of Statistics, University of Warwick, Coventry, West Midlands, CV4 7AL, U.K.
CONTENTS

1. INTRODUCTION: WHAT IS SURVIVAL? ........................................... 1

SECTION A

STATISTICAL MODELS, INPUTS AND OUTPUTS

2. SURVIVAL MODELS ............................................................... 2
3. STATISTICAL ANALYSIS ....................................................... 2
4. INITIAL MODEL INPUTS ....................................................... 3
5. GRAPHICS: COMMUNICATION AND PRESENTATION ...................... 4
6. OUTPUT OF NUMERICAL INFORMATION .................................... 4

SECTION B

SURVIVAL FACILITIES

7. SURVIVAL MENU STRUCTURE: AN OVERVIEW ............................ 5
8. DETAILED MENU DESCRIPTION ........................................... 9
9. PLOTTER OUTPUT ............................................................. 14
10. INTERACTIVE GRAPHICS ................................................... 15
11. MISCELLANEOUS ............................................................. 16

SECTION C

ILLUSTRATIONS

12. EXAMPLE OF A REFERENCE ANALYSIS .................................... 17
13. EXAMPLE OF STANDARD ANALYSIS ..................................... 25
14. DATA CLASSIFIED BY FACTORS ........................................... 31
1. INTRODUCTION: WHAT IS SURVIVAL?

SURVIVAL is a micro-computer package for Bayesian analysis of survival data. The package is written in APL, comprising a collection of functions that can be used to perform a variety of activities in data management, modelling and analysis, and to produce many numeric and graphical displays. Survival modelling is based on dynamic regression models for the analysis of survival/failure data and provides a useful framework for data analysis in medicine, reliability, economics and other areas of science. The dynamic models considered have time-varying effects of either fixed or time-dependent regressor variables, including both continuous and categorical variables. SURVIVAL is completely menu driven, providing an easy to use interface for all users, expert or otherwise, and thus no knowledge of APL is required. In addition to a full, hierarchical system of menus, SURVIVAL incorporates standard spreadsheet interfaces for data management.

Illustrations of some of the facilities in SURVIVAL, based on earlier versions of the package, appear in references [1], [2] (Chapter 8), [3] and [4].

Briefly, the functions available in SURVIVAL provide facilities along the following lines:

1. interactive data input and output to files, data description and display in numeric and graphical forms;
2. graphical displays, summaries of data variables and simple, exploratory statistical analyses, including the product limit estimator and the Bayesian nonparametric A(n) predictor;
3. data manipulations such as calculations, transformations, grouping and creation of new variables and selection of subsets of the data;
4. dynamic model definition, specifying components for regression effects of independent variables, time grids and parametric evolution;
5. specification of prior distributions on model components;
6. sequential model estimation;
7. preposterior prediction of survival times for prior assessment;
8. numerical and graphical displays of features of fitted model, including on-line and smoothed estimates of regression effects and individual hazard functions;
9. graphical displays of predictive features from the fitted model for given individuals;
10. numerical summaries, in tabular form, and model information saved on disk file or printed;
11. interactive manipulation of graphical displays for report production using direct screen copy to dot-matrix printer or post-processing, within SURVIVAL, to high-resolution, colour pen-plotter;
12. various, minor facilities for enhancements and tidying-up operations.
2. SURVIVAL MODELS

The dynamic models available in SURVIVAL are described in references [1],[2] and [3]. Reference [2] is an extended version of [1], dealing with a wider class of models for a number of statistical problems related to the regression analysis in SURVIVAL. A brief description of the SURVIVAL models is given below. This provides the basic information required by users and can be complemented, if necessary, by any of the references above.

The models assume (censored or uncensored) survival times are observed along with their censoring indicators and a number of related variables and indexed by \( j = 1, \ldots, n \). Each survival time \( Y_j \) has piecewise-exponential hazard function

\[
\lambda^{(j)}(t) = \lambda^{(j)}_i, \quad t \in I_i = (t_{i-1}, t_i], \quad i = 1, 2, \ldots,
\]

for \( t_0 = \) time origin and ordered times \( t_i \). The related variables relevant to the model are denoted \( z_j = (1, z_{j1}, \ldots, z_{jp}) \) and linked to the hazard via

\[
\log \lambda^{(j)}_i = z_j \beta_i, \quad i = 1, 2, \ldots,
\]

where \( \beta_i \) is a \( p + 1 \) column vector containing the baseline hazard parameter \( \beta_{i0} \) and the regression coefficients \( \beta_{i1}, \ldots, \beta_{ip} \) for interval \( I_i \). If a categorical variable, or factor, with \( q \) levels is included in the regression, \( q - 1 \) regression coefficients are included in the model, one for each level other than the first whose effect is taken as zero. If the values of \( z_j \) are time-dependent, then \( z_j \) is replaced by \( z_{ji} \) in the expression above where \( z_{ji} \) is the vector containing the values of the covariates in interval \( I_i \).

Unlike common standard approaches, the regression coefficients are allowed to change with time providing for non-proportionality of hazards. The nature of this time variation is governed by a stochastic structure relating successive values via

\[
\beta_i = \beta_{i-1} + w_i
\]

where \( w_i \) is a zero mean random term.

3. STATISTICAL ANALYSIS

Analysis is sequential. The observed data is factorised along time intervals \( I_i \) with each factor containing events occurring in the given interval. Eventually, there is a time, say \( t_N \), after which no further information is collected from the data. The analysis can therefore be performed sequentially as in time series. At each time \( t_i \), prior information and past data constitute the current information denoted by \( D_i \). This information is summarised in terms of posterior distributions for the model parameters.
As time evolves to times $t_{i+1}, t_{i+2}, \ldots$, each coefficient may change and new data is processed to revise, or update, these posterior distributions as the information set updates. The sequences of point estimates of these coefficients (posterior means for each $t$) define estimated on-line time trajectories for the parameters; on-line variances provide measures of uncertainty in these estimates.

Having processed all the available data up to the end of the series $i = N$, retrospective assessments may be based on the smoothed estimates of model parameters that are derived from the smoothed or posterior distributions for parameters at each time $t$ given all the information $D_N$. These are obtained via a backward filtering algorithm.

A sequence of point estimates of the parameters based on smoothed distributions can also be formed; again, smoothed variances provide a measure of the uncertainty about these estimates. These constitute smoothed time trajectories for the model parameters.

4. INITIAL MODEL INPUTS

Initial model specification requires the following inputs from the user:

(1) **Data**
   
   A data file is selected on the designated disk drive, and data read in or entered manually from the keyboard. Previously used data is automatically erased, and may also be erased manually, as may details of previous analyses.

(2) **Model components**
   
   The covariates to be included in the model are selected from the variables originally available in the data set and the ones created by the user.

(3) **Time grid**
   
   The times $t_1, \ldots, t_N$ specifying the temporal factorisation are specified. These can be based on equal-length intervals (or a combination of such), on the observed death/failure times or indeed in any arbitrary setting. The specification of the time grid does not depend on the model components selected and can be set prior to the selection of covariates.

(4) **Initial priors**
   
   Initial moments, conditional on the prior information $D_0$ are required for the model parameters. They consist of mean and standard deviation (s.d.) for each of the parameters taken to be initially uncorrelated.

(5) **Stochastic evolution**
   
   The rates at which the model parameters vary stochastically over time are determined by the variance matrix of an additive error. This can be specified directly or by use of discount factors. The variance matrix affects the uncertainty about the model parameters additively whilst discount factors are applied multiplicatively to the parametric variance using the notion of information decay. Thus, a discount factor of 0.9 implies a loss of 10% of the information due to evolution in time. Both approaches adjust automatically for intervals of unequal lengths. Finally note that proportional hazard models are obtained by corresponding specification of 0 entries in the additive setting or of discount factors of 1 for the regression coefficients.
5. GRAPHICS: COMMUNICATION AND PRESENTATION

Much of the output information from SURVIVAL is in graphical form. Preliminary data exploration uses simple plotting facilities for scatter plots and log plots of data variables. Simple preliminary statistical analyses and other functions also use graphical displays.

After fitting and smoothing, several plots of time trajectories of model parameters are available which provide plots over time of estimated regression effects. Also, plots of the estimated hazard function for given values of the regressor variables are available. In prediction both before and after analysing the data, plots of the predictive hazard, density and survival functions are available.

Graphs are undeniably useful for communication and presentation of inferences. When a graph is drawn in SURVIVAL, the user may interact to modify the display for copying. On completion of any graph, a cross-hair cursor appears in the lower left-hand corner of the screen. This may be moved around the screen using the Arrow keys on the numeric keypad and character input from the keyboard allows a variety of additions, deletions and other modifications to be made. The screen may then be saved for later processing, copied directly to a dot-matrix printer, or saved for post-processing to a multi-colour, high resolution pen plotter. Interaction on graphical displays is detailed in the graphics section of the catalogue of SURVIVAL functions below. One important keyboard input is the Q character, to quit graphical mode and return control to the menu system.

6. OUTPUT OF NUMERICAL INFORMATION

Several facilities exist for viewing and storing numeric data and information relating to fitted models. Paginated tabular displays can be produced to provide:

- variables from the data set;
- on-line or smoothed time trajectories of model components;
- summaries of posterior distributions on model parameters.

These tabular displays can also be output to a printer or DOS file respectively, the latter of obvious use for communication with other packages. Generally, SURVIVAL functions exist to read and write model information and summaries to disk files. This allows the user the option of processing new data with a model previously fitted to historical data.
SECTION B

SURVIVAL FACILITIES

7. SURVIVAL MENU STRUCTURE: AN OVERVIEW

SURVIVAL is menu driven. At any stage in a user session, a variety of SURVIVAL facilities are available for selection from one of several, hierarchically linked menus. Some facilities may not be present at any given time if prerequisite functions have not previously been executed. The menu selection procedure takes care of this automatically, allowing only the use of permitted functions at any time.

Menu options are displayed at the top of the screen and can be selected in one of two ways:

- by using the keypad arrows to move the cursor, highlighting the required option, and pressing the Enter key; or
- by typing the initial letter, and others, as necessary, if two or more options have the same initial letter(s).

Note that the leftmost option is highlighted initially and so will be the default selection if the Enter key is pressed directly.

After loading SURVIVAL, the initial, or Root, menu is displayed, and the user must answer several initial questions. The responses determine variables that control the global environment for the session, concerning the following features.

(a) Graphics display

In the current implementation, SURVIVAL may be used with any of the following four common graphics adaptors/monitors.

- IBM Colour Graphics Adaptor — CGA — and fully compatible adaptors
- IBM Enhanced Graphics Adaptor — EGA
- Hercules Monochrome Adaptor
- Olivetti M24 Monochrome Monitor

It is extremely important that the correct graphics configuration is selected since failure to do so will result in unpredictable results during graphical displays, and probably lead to suspension of the session, requiring a re-boot from scratch of the system. If an incorrect selection is made, or suspected, it can be altered later (before using any graphical facilities) by selecting the Gr-config option from the Housekeeping menu described in Section 8.

(b) Plotfiles

This controls the option for saving graphics commands for possible later processing to provide pen plotter output. Answering Yes means that the user may save any graph displayed during the session, simply by asking for the save facility when the graph is initially constructed. Answering No simply means that this facility will be unavailable. As with the graphics configuration, this option may be reset within the Housekeeping menu.
Exhibit 1 provides an overall representation of the linkages between menus in the hierarchical structure of SURVIVAL. Each box represents a given option and the hierarchy is represented by the interconnection of these boxes. At the lowest level, which is entered immediately after the system initialisation procedure described above, is the Root menu. At the beginning of a SURVIVAL session this menu contains just three options: Data, Housekeeping and Exit. The other options, enclosed in brackets in Exhibit 1, only become available after certain prerequisite actions have been performed within SURVIVAL. This *conditional option suppression* mechanism is used throughout SURVIVAL and ensures that only those facilities which are meaningful at any stage of an analysis are presented to the user. For example, Fit (which, as the name suggests, performs the model fit to the data) cannot be accessed until a model has been defined, and model definition, which is controlled by the Model option, itself becomes available only after a dataset has been entered into the SURVIVAL memory. A complete description of these option prerequisites is given in the detailed discussion of the options which follows.
Exhibit 1. An overview of the SURVIVAL menu structure

ROOT

DATA  (MODEL)  (FIT)  (OUTPUT)  HOUSEKEEPING  EXIT

DATA

INPUT/OUTPUT  (CREATE)  (EXPLORE)  (REMOVE)  QUIT

INPUT/OUTPUT

READ  (WRITE)  KEYBOARD  (AMEND)  (SELECT)  QUIT

ASFD  NATIVE  QUIT

ASF

NATIVE

QUIT

EXPLORE

X-Y  LOGPLOT  SUMMARY/HISTOGRAM  PL  EMPERICAL HAZARD  A (n)  QUIT

REVIEW  PLOTTER

OVERALL  (FACTOR)  QUIT

OVERALL  (FACTOR)  QUIT

OVERALL  (FACTOR)  QUIT

MODEL

COVARIATES  GRID  PRIOR  EVOLUTION  QUIT

EQUAL LENGTHS

DEATH TIMES

ARBITRARY

QUIT

ADDITIVE

DISCOUNT

QUIT
Exhibit 1. (continued)

- OUTPUT
  - (TRAJECTORY)
  - (HAZARD)
  - TABLES
  - (LIKELIHOOD)
  - (PREDICT)
  - QUIT

- (PREDICT)
  - (POSTERIOR)
  - PREPOSTERIOR
  - QUIT

- HOUSEKEEPING
  - PRECISION
  - CLEAN
  - GR-CONFIG
  - PLOTFILE
  - QUIT
(ii) [Write]
Available only when there is a dataset in the SURVIVAL memory. This option
allows the current dataset to be written to a file; the same formats as under the
Read option are available.

(iii) Keyboard
Takes information about a dataset from the keyboard. Alongside each variable
name, a number has to be entered. This should be 0 for continuous variables and
an integer q for factors with q levels. (Note that the only values allowed for the
entries of a factor with q levels are the integers between 0 and q - 1 denoting the
levels.) Data entry for the variables is controlled in the manner of a spreadsheet.
Data entries are input by typing in the values and pressing Enter; the cursor is
then moved between entries using the cursor arrows on the numeric keypad. Note
that entries as yet unset appear as 0. Missing values are entered as -999. Input is
completed, and the dataset may be stored on disk, after pressing the ESC key and
selecting the file format. The first column is reserved for individual indexation,
the second is for entry of the survival/failure times, the third for the respective
censoring indicators (1 for observed death/failure and 0 for censored times) and
the remaining ones for the value of the covariates, if any. An affirmative answer
to the query about the presence of time-dependent covariates, will subsequently
present another spreadsheet for data entry. The time at which the change in the
values of the covariates takes place and the changed values of the covariates are
input only for the appropriate individuals. Note that all covariate values have
to be input for individuals whose covariates change with time. If more than one
change occurs for any given individual(s), new spreadsheets appear for data entry
after request.

(iv) [Amend]
Provides spreadsheet display of the data allowing the user to amend/edit values at
will. Also allows user to extend the dataset by specifying the number of additional
individuals and then inputting new values onto the spreadsheet display. Used in
conjunction with Select this option provides a comprehensive data editing facility.
On completion, the revised dataset may be written to file as described above.

(v) [Select]
Chooses a subset of the current data for analysis. The result affects all subsequent
SURVIVAL operations including Write. Options are:
- Index: Selects according to a specified range of individuals indices.
- Times: Selects according to a specified range of survival/failure times.
- Deaths: Selects the censored individuals.
- Covariates: Selects according to a specified range of a selected variable.
- Quit: Returns control to Input/Output menu.

The original dataset can only be recovered by re-entry (using Keyboard or Read).

(vi) Quit
Returns control to the Data menu.

(2) [Create]

Allows operations on data series with addition, subtraction, multiplication, division,
logarithmic, exponential and power transformations and grouping operations to pro-
duce new variables for analysis. The user inputs a name for the new variable, and
then a definition using keys +, -, x (or *) and ÷ (or /) and keywords, LOG, EXP,
CENTRAL (for standardisation to zero arithmetic mean), STD (for standardisation
also to unit standard deviation), POWER and GROUP to define new variables. Note
that APL reads from right to left. Suppose, for example, that the dataset contains two variables AGE and TYPE, with \( j \)th elements \( \text{AGE}(j) \) and \( \text{TYPE}(j) \) respectively. A variable named INDEX is created as follows. Firstly,

Variable Name INDEX

Formula \((\log \text{CENTRAL AGE}) - 2 \times (\text{STD TYPE}) \text{ POWER } 0.5\)

results in a series having \( j \)th element

\[ \text{INDEX}(j) = \log |\text{AGE}(j) - m_1| \div 2 \times ((\text{TYPE}(j) - m_2) \div s)^{0.5}, \]

where \( m_1 \) is the arithmetic mean of the elements of AGE and \( m_2 \) and \( s \) are the arithmetic mean and standard deviation of the elements of TYPE. In another example, to categorise AGE into groups 0-10, 11-20, 21-30,..., 81-90, over 90, a variable named AGEGROUP is created as follows:

Variable Name AGEGROUP

Formula \( \text{AGE GROUP} 10 20 30 40 50 60 70 80 90 \)

Since the values of AGEGROUP are integers between 0 and 9, SURVIVAL prompts for confirmation of this variable as a factor. An affirmative answer creates a factor with 10 levels, a negative one leads to creation of a continuous variable assuming values 0, 1, 2, ..., 9. Using Create, current data variables may be overwritten but confirmation of this is required. The new series become an integral part of the current dataset available everywhere.

(3) [Explore]

Available only when there is a data set in the SURVIVAL memory. This option provides access to a number of graphical and statistical data analyses that can be used as a preliminary step in model building. The options are:

- X-Y: Produces a scatter plot of two variables. Uncensored observations are represented by + and censored observations by o.
- Logplot: Produces a scatter plot of two variables, the first which is logged. Uncensored observations are represented by + and censored observations by o.
- Summary/Histogram: Simple summary statistics for any selected variable including TIMES (survival/failure times) and DEATHS (censoring indicators). If TIMES is selected, only uncensored times are considered.
- Empirical Hazard: Displays the empirical hazard functions for the survival times either classified according to a selected factor or pooled together.
- PL: Calculates and displays the Product-Limit estimator (see reference [8]) for the survival times either classified according to a selected factor or pooled together.
- A(n): Calculates and displays the A(n) predictor for the survival function (see reference [9]) for survival times either classified according to a selected factor or pooled together. (Note that the calculations required for the A(n) predictor are very time consuming when censoring is present; indeed, this option should not be selected in datasets of moderate sizes if censoring is present).
- Review: Facility for retrieval of a specified file of plotting commands from disk. These instructions are then processed and the resulting graph displayed on screen. This may be amended, reassembled or plotted as the user requires.
- Plotter: Facility for retrieval of a specified file of plotting commands from disk. These instructions are then processed and the resulting code sent to the pen plotter. There is an automatic check to ensure that the plotter is on line. See also Section 9.
- Quit: Return control to Data menu.

(4) [Remove]

Erase all existing data and model information from memory.
(5) Quit
Returns control to Data menu.

8.3 MODEL MENU

This is available only after a dataset has been entered into the SURVIVAL memory using Read or Keyboard under the Data menu. The option provides access to a set of routines from which to define a model and set global conditions for a subsequent analysis.

(1) Covariates
Lists all the existing variables for selection of the regressors to be included in the model.

(2) Grid
Selects the type of grid used in the analysis subsequent to specification of the time origin, set to 0 by default. Options are:
- Equal lengths: Specifies the length and number of intervals up to the last observed survival times allowing for the groups of intervals with the same length.
- Deaths: Sets the time grid based on [a specified integer multiple of] the successive uncensored survival/failure times.
- Arbitrary: Allows entry of any arbitrary time grid from a spreadsheet.
- Quit: Returns control to Model menu.

(3) [Prior]
This option may be accessed only after the model components have been defined. Sets the mean and standard deviation of all model parameters with an uninformative/reference setting as default. The moments of the baseline hazard and the continuous covariates coefficients are input on the first screen, followed by those for the effects of any factors included in the model on subsequent screens.

(4) [Evolution]
This option may be accessed only after the model components have been defined. Selects the type of evolution adopted in the model. The options are:
- Additive: Specifies the upper diagonal part of the covariance matrix of the additive error.
- Discount: Specifies the discount factors for the model parameters. Successive parameters with the same discount are taken as a block.
- Quit: Returns control to Model menu.

In both specifications, the evolution is based on a specified time length and a static model is set as default.

(5) Quit
Returns control to the Root menu.

8.4 FIT MENU

This option may only be accessed after a model has been defined, which in turn requires a dataset in the SURVIVAL memory. The model is then fitted, displaying the interval currently being processed and then smoothed. On completion of the fitting process a simple measure of predictive fit — the marginal likelihood of the model from the data — is displayed. On quitting this menu, the results of a fit, including estimated time trajectories, hazard functions, predictions and so forth, are available under the options of the Output menu.
8.5 OUTPUT MENU

Provides the results from the fit, the model and the data used.

(1) [Trajectory]
   Available only after a model has been fitted. Provides a plot of the estimated time
   trajectories of a specified model parameter. The options are:
   • Smoothed: Plots the smoothed mean trajectories along with 2 s.d. limits.
   • Smoothed/On-line: Plots in addition, the on-line mean trajectory.
   • Quit: Returns control to the Output menu.

(2) [Hazard]
   Available only after a model has been fitted. Provides plots of the estimated hazard
   function for a number of individuals given their regressor values. After specifying the
   number of cases considered, the respective values of the covariates are input via the
   spreadsheet.

(3) Tables
   Provides a variety of tabulated displays with options to print and/or save to a file.
   Tables are displayed on a screen one page at a time. Type 'Y', the Enter key or the
   Space bar to display the next page; any other key quits the current table display. Tables
   provided are:
   • [On-line]: Available only after a fit; displays the on-line estimates of the model
     parameters. If the model has more than seven parameters, then a series of tables
     is produced.
   • [Smoothed]: Available only after a fit; displays the smoothed estimates of the
     model parameters. If the model has more than seven parameters, then a series of
     tables is produced.
   • [Posterior]: Available only after a fit; the final posterior estimates — means and
     standard deviations — of the model components are displayed.
   • [Correlation]: Available only after a fit; displays the lower triangular part of the
     final posterior correlation matrix of the model parameters. If the model has more
     than ten parameters, then up to ten may be selected for display.
   • Data: Displays up to seven selected data variables.
   • Print: Turns the printer on if it is currently off and turns it off if it is currently
     on. Succeeding selection of tables may thus be printed as well as displayed on the
     terminal.
   • Store: Saves the most recently displayed table in a file.
   • Quit: Returns control to Output menu.

(4) [Likelihood]
   Available after a model has been fitted; displays the logarithm of the model likelihood
   from the data.

(5) [Predict]
   This option may be accessed after a model has been defined. It produces plots of the
   predictive distribution of a given number of individuals based on the total information
   available. Selection of number of cases and their covariate values are as in the Hazard
   option — see Section 2 above. The options are:
   • [Posterior]: Available only after a model has been fitted; obtains the predictive
     distribution given the total information available after fitting the data.
   • Predict posterior: Obtains predictive distribution given only the prior information.
   • Quit: Returns to Output menu.

After the selection of the type of prediction, the following features of the predictive
distribution are available:
- Survival: Plots the predictive survival function.
- Hazard: Plots the predictive hazard function.
- Density: Plots the predictive density function.
- Quit: Returns control to the Predict menu.

(6) Plotter
   Already described in the Explore menu in Section 8.2.

(7) Quit
   Returns control to Root menu.

9. PLOTTER OUTPUT

To achieve a hardcopy of a graphics screen on a pen plotter two stages are necessary. First, when the screen is drawn the plotting commands must be saved on file. This is only possible if the < Plot file > prompt is not disabled (see Section 7) and if Yes is selected in response to the prompt. This action produces an ASCII file of plotting commands in a proprietary format. Second, this file must be retrieved by the Plotter option in the Data/Explore or Output menus. When Plotter is selected the file containing the plotting commands is first specified and retrieved and a list of plotter options is displayed. These options, described here, may be altered by pressing return to change the values of the option indicated by the cursor, cycling through the possible settings of that option. Movement between options is effected using the cursor ↑ and ↓ keys. The options are:

(1) Colour: On or Off — if off, then all colour changes from the default are ignored.
(2) Line Type: On or Off — if off, then all lines will be plotted as full even though they were originally dotted or dashed on the screen.
(3) Text: On or Off — if off, no text is plotted.
(4) Slant: On or Off — On to produce slanted text output.
(5) Rectangle: On or Off — on to produce a rectangular frame for the plot, off to produce a square frame.
(6) Position: Possibilities are
   - 1 ... using the full A4 page in landscape mode
   - 2 1 or 2 2 ... Uses half the A4 page in portrait mode, upper half for 2 1 and lower half for 2 2,
   - 4 1, 4 2, 4 3 or 4 4 ... Uses one quarter of the A4 page in landscape mode.

(7) Pen Speed: P or T ... P for paper speed, T for Transparencies. The latter speed is slower than that for paper to accomodate the extra drying time of the special transparency ink (and is also useful for drawing on paper when the pens are nearing the end of their useful lives).

(8) Arc Angle: An integer controlling the smoothness of appearance of any circles drawn ...
     smoothness increases as arc angle decreases. An arc angle of 5, the minimum, produces the smoothest circle but circle drawing at this level of smoothness is extremely slow.

After these options have been chosen by pressing ESC, the command file is processed and the plotter activated. A device driver for the Hewlett Packard 7475A pen plotter is incorporated within SURVIVAL, and this will also work with any other plotter driven by the Hewlett Packard Graphics Language (HPGL). No plans exist at present to provide support for alternative plotting languages — users who require this facility should contact Andy Pole.
10. INTERACTIVE GRAPHICS

Once a graph has been completed under SURVIVAL control, interactive graphics mode is automatically invoked. In this mode, the user may interact with the displayed graph to alter the display and add to it, producing final annotated plots for output to printer or plotter. The modes of interaction and facilities provided are as follows.

10.1 CURSOR CONTROL

The cross-hair cursor, appearing initially in the lower left-hand corner of the screen, may be moved around the screen using the cursor keypad arrows ←, →, ↑, and ↓. The extent of movement may be varied using Control and Shift keys simultaneously with the cursor arrows. The number of pixels moved will be 1 if no control key is used, 10 using Control/Cursor arrow, or 50 using Shift/Cursor.

10.2 INTERACTIVE ANNOTATION OPTIONS

The various options detailed below are selected by typing the letter of the option required, followed by whatever additional keyboard input, if any, required to specify the extent and nature of the annotation chosen.

A : Draw an arrow from the previous marked position (marked using option M) to the current cursor position.
C : Draw a line from the previous marked position to the cursor and extend it right across the plotting window.
D : Draw a line from the previous marked position to the cursor.
E : Erase that portion of the display in the rectangular area having diagonally opposite corners given by the previous and current cursor positions (nb. as yet this facility will NOT produce erasures on pen-plotter reproductions).
F : Place cursor in the Y axis at ordinate .1, .2,...,9 defined by input after F. Used in plots of the survival function where the cursor can then be moved horizontally until coinciding with the desired curve. After pressing return, displays the value of the required decile.
H : Help! Exits graphics mode temporarily to display a summary list of these interactive graphics options. Display is restored after striking any key.†
L : Change the linetype for all subsequent annotation. Following L, input a number defining the required linetype as follows: 0 for solid line, 1 for dotted line, and 2,3,..., for various dashed forms ... try some and see. Invalid entries will produce the default full line.
M : Define the current cursor position as the marked position. This option has no apparent effect until one of options A, C, D, E or R is used.
O : Centre cursor at top of screen.
P : Print the graph on dot matrix printer via parallel port connection (a message is displayed if the printer is not on line).
Q : Quit interactive graphics session, losing displayed graph.
R : Return cursor to the previous marked position.
S : Draw a symbol centred at the current cursor position. After typing S, the symbol is chosen by typing an integer, as follows: 0 for *, 1 for x, 2 for o, 3 for Δ, 4 for △, 5 for □, 6 for o, 7 for *, and 8 for +.

†Note on problem with EGA.
T : Write text with upper left-hand corner at the current cursor position. After typing T, text is input, and appears directly on the screen, until Enter is pressed.

W : Where am I? The current position of the cursor, in the coordinates of the graph displayed, is given in the top left-hand corner of the screen. If a graph is being reviewed then these coordinates are in terms of the default scaling 0-639, 0-399.

Z : Zoom subsequent text output to size 1, 2, ..., 9, defined by input after Z. Any input not an integer in this range is taken as 1.

11. MISCELLANEOUS

11.1. VARIABLE SELECTION

There are only two possible modes of selection of options/variables etc. in SURVIVAL menus. These are

- single variable option: highlight variable name and press Enter.
- multiple variable option: highlight each variable in turn, pressing Enter for each. This marks the selected variables with \( \ll \) — this marker is a toggle that may be turned on and off by repeated use of Enter with the highlighted variable. Press ESC to complete selection of indicated variables.

11.2. SURVIVAL AUDIO VISUAL CUES

Audible tones are used by SURVIVAL to indicate various successful operations and most errors. The tones used are

- *blank* tone for invalid data entered during a windowed input session, such as with specification of priors, for example;
- *beep* tone when moving between windows for data entry, at completion of a plot, and on certain other occasions.

Visual indicators of SURVIVAL activities are also used from time to time, such as

- a message on clear screen if invalid actions are requested by the user;
- flashing messages to indicate calculations or other operations, such as plotting, in progress;
- flag displayed at the top of the screen when the printer is in operation, except when printing a graphics screen.
ILLUSTRATIONS

This Section illustrates in detail how to perform analyses of data sets using SURVIVAL. The datasets used in this Section, named WBC, UNEMSAMP and NASOCANC, are analysed purely for illustrative purposes, demonstrating the use of SURVIVAL facilities, and in no way represent any kind of final analyses accepted as adequate or appropriate summaries of the data series. Even so, the usefulness of the models is clearly apparent from the examples where the results are relatively satisfactory.

These examples cover different analyses of the data. The first uses a reference (vague initial prior, static evolution) analysis of the data requiring minimal inputs from the user. The second and third analyses illustrates many of the modelling inputs and outputs available from SURVIVAL using proper, informative priors as are typically available from the context of the problem and previous data analyses etc., and a dynamic evolution whose use is often vital to the predictive performance of the model. While the second analysis is performed over data with time-dependent covariates the third is performed over data with factor covariates.

12. EXAMPLE OF A REFERENCE ANALYSIS

After loading SURVIVAL and completing the initialisation procedure (described in Section 7 above) the screen display contains the following three lines of information:

SURVIVAL 1.1

SELECT

Root menu

Data Housekeeping Exit

The Data entry is highlighted, and a brief explanation of the highlighted option appears immediately below, in this case

Data handling and analysis routines

The illustrated analysis proceeds as follows.

(1) The option is chosen to input Data into the SURVIVAL memory. For the illustration we shall assume that this data resides in a data file called WBC.ASF (an APL component file) — which of course means that it has been created in an earlier SURVIVAL session. Select Data from the Root menu and the Data menu is displayed:

Input/Output Quit

(Data)

Recall that the other possible options in the Data menu are not available at this stage since there is no dataset in the SURVIVAL memory as yet.

(2) Select Input/Output from the Data menu to obtain the following (subset of) options:

Read Keyboard Quit

(Data Input/Output)

(3) Select Read and the list of supported SURVIVAL file types is displayed:

ASF file Native file Quit

(Data Input/Output)
The data file for this illustration is WBC.ASF — an APL component file — and hence the appropriate selection from the Data/ Input/Output/ Read menu is ASF. Having selected the type of file for SURVIVAL to read a list of disk drives is displayed — this time a vertical list is used rather than the horizontal kind but the method of option selection is the same (as described above). For a machine with a hard disk drive and floppy drives a typical display will be,

A:  
B:  
C: (Select this option for the hard disk)  
D:  
E:  

A list of files with the extension .ASF will now be displayed and to select the file WBC simply use the cursor movement keys to highlight this file name and press return. On completion of this file selection procedure control returns to the Data/ Input/Output menu which displays the additional options Write, Amend and Select.

To obtain a table with the data we move through the menu hierarchy to the Root menu using Quit successively:

- Read Write Keyboard Amend Select Quit (Data Input/Output)  
- Input/Output Create Explore Remove Quit (Data)  

finally producing the display

Data Model Output Housekeeping Exit (Root menu)

Select the Output option which then displays:

Tables Quit (Output)  

Select the Tables option displaying:

Data Print Store Quit (Tables)  

If the Data option is selected, a display like

TIMES  
DEATHS  
AG  
WBC

appears. By pressing Enter for all lines followed by ESC (as described in Section 11.1), all variables are selected producing Table 1. This is a paginated display that can be copied to printer or disk file as explained in Section 8.5.

The wide range of values of WBC suggests the use of a logarithmic transformation. This is done in the Create option of the Data menu. To reach it, we have to move through the hierarchy back to Root menu using Quit successively:

- Data Print Store Quit (Tables)  
- Tables Quit (Output)  

and select Data followed by the Create option in the Data menu.

This prompts for a variable name and a formula to which we input

Variable name LOGWBC  
Formula LOG WBC
followed by the ESC key to return control to Data menu. This creates the variable LOGWBC with the values of the logarithm of WBC.

(10) To display a scatterplot of the survival times against LOGWBC, select Explore which yields

X-Y Logplot Summary/Histogram Empirical-Hazard PL A(n) Review Plotter Quit

(Explore)

and select X-Y. A vertical list of all variables is displayed, and pressing Enter at the TIMES line selects the survival times for the Y axis. Similarly, pressing Enter at the LOGWBC line will subsequently select the values for the X axis. This produces Figure 1 (after responding appropriately to the Plotfile inquiry if this facility was not suppressed during initialisation). Quit from interactive graphics mode (option Q) and return control to the Root menu by quitting the Data menu.

(11) A model can now be set by selecting Model which displays

Covariates Grid Quit

(Model)

Select Covariate which produces the vertical display
AG
WBC
LOGWBC

Pressing Enter at the AG and LOGWBC lines followed by the ESC key selects them as the covariates of the model and returns control to Model menu displaying

Covariates Grid Prior Evolution Quit

(Model)

The new options are now available as a result of the specification of the regressors.

(12) The time grid is set by Grid and, as time origin for this data is 0, press the ESC key to obtain:

Equal-Length Deaths Arbitrary Quit

(Grid)

Select the Deaths option and press the ESC key again to set the time grid to include all uncensored survival times. This also returns control to Model menu.

(13) In order to set the prior moments, select Prior. For a reference analysis, the vague distribution implied by the default is appropriate and all we have to do is to press the ESC key to return control to the Model menu.

(14) Finally, the parametric evolution can be set by the Evolution option which yields

Additive Discount Quit

(Evolution)

Choose Discount, press the ESC key (to set discount factors over periods of 1 week) and press the ESC key to set discount factors of 1 for the model parameters. This sets a static evolution implying parameters constant over time.

(15) The model is now completed, and can be fitted to the data by quitting the Model menu to:

Data Model Fit Output Housekeeping Exit

(Root menu)

The Fit option is now available and can be selected. While the model is being sequentially fitted, the cycle currently being processed is displayed and after fitting and smoothing are completed the value of the model log-likelihood is displayed.

(16) The results can now be analysed with the Output menu which now displays:
Trajectory Hazard Tables Likelihood Predict Review Plotter Quit (Output)

In particular, Figure 2 is produced using the Trajectory option yielding:

Smoothed Smoothed/On-line Quit (Trajectory)

Selection of the Smoothed/On-line option produces the plot of the smoothed and on-line trajectories of a selected parameter. In Figure 2.a, the AG effect is selected and in Figure 2.b, the LOGWBC effect is selected. Observe that a static evolution automatically implies a constant smoothed trajectory for the parameter. After quitting the graphics mode and the Trajectory menu, control returns to the Output menu.

(17) Numerical summaries of the analysis can also be obtained using the Tables menu:

On-line Smoothed Posterior Correlation Data Print Store Quit (Tables)

Selection of the On-line option produces Table 2 and the Posterior option produces Table 3. The Quit option returns control to the Output menu.

(18) The results of the analysis can also be assessed by the predictions obtained in the Predict menu. In this case, two individuals with covariate values (1,10) and (0,10) are considered. After input of the number of cases and the covariate values (in a spreadsheet) successively followed by pressing the ESC key, the following is displayed:

Posterior Preposterior Quit (Predict)

Selection of Posterior leads to the calculation of the predictive distribution for the above individuals given the data. The features available for plots are displayed:

Survival Hazard Density Quit (Posterior)

Figure 3 is the plot of the survival functions of the individuals selected with the Survival option.

This data has also been analysed with different prior and evolution settings in an early version of SURVTVAL reported in [2].
<table>
<thead>
<tr>
<th>OBS No.</th>
<th>TIMES</th>
<th>DEATHS</th>
<th>AG</th>
<th>WBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>100000</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1000000</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>27000</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>10000</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>28000</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>21000</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>17000</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>19000</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>26000</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1000000</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>52000</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1500</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>31000</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>6000</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>9000</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>4000</td>
</tr>
<tr>
<td>17</td>
<td>22</td>
<td>1</td>
<td>1</td>
<td>35000</td>
</tr>
<tr>
<td>18</td>
<td>22</td>
<td>1</td>
<td>0</td>
<td>5300</td>
</tr>
<tr>
<td>19</td>
<td>26</td>
<td>1</td>
<td>1</td>
<td>32000</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>79000</td>
</tr>
<tr>
<td>21</td>
<td>39</td>
<td>1</td>
<td>1</td>
<td>5400</td>
</tr>
<tr>
<td>22</td>
<td>43</td>
<td>1</td>
<td>0</td>
<td>1000000</td>
</tr>
<tr>
<td>23</td>
<td>56</td>
<td>1</td>
<td>1</td>
<td>9400</td>
</tr>
<tr>
<td>24</td>
<td>56</td>
<td>1</td>
<td>0</td>
<td>4400</td>
</tr>
<tr>
<td>25</td>
<td>65</td>
<td>1</td>
<td>1</td>
<td>2300</td>
</tr>
<tr>
<td>26</td>
<td>65</td>
<td>1</td>
<td>1</td>
<td>1000000</td>
</tr>
<tr>
<td>27</td>
<td>65</td>
<td>1</td>
<td>0</td>
<td>3000</td>
</tr>
<tr>
<td>28</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>4300</td>
</tr>
<tr>
<td>29</td>
<td>108</td>
<td>1</td>
<td>1</td>
<td>10500</td>
</tr>
<tr>
<td>30</td>
<td>121</td>
<td>1</td>
<td>1</td>
<td>10000</td>
</tr>
<tr>
<td>31</td>
<td>134</td>
<td>1</td>
<td>1</td>
<td>2600</td>
</tr>
<tr>
<td>32</td>
<td>143</td>
<td>1</td>
<td>1</td>
<td>7000</td>
</tr>
<tr>
<td>33</td>
<td>156</td>
<td>1</td>
<td>1</td>
<td>750</td>
</tr>
</tbody>
</table>
TABLE 3

POSTERIOR SUMMARY

<table>
<thead>
<tr>
<th>Component</th>
<th>Mean</th>
<th>S.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASELINE</td>
<td>-2.574</td>
<td>1.282</td>
</tr>
<tr>
<td>AG</td>
<td>-2.145</td>
<td>0.349</td>
</tr>
<tr>
<td>LOGWBC</td>
<td>-0.024</td>
<td>0.131</td>
</tr>
</tbody>
</table>

Figure 3

KEY

1

2

S(t)

0  0.2  0.4  0.6  0.8  1

0  25  50  75  100  125  150

WEEKS
13. EXAMPLE OF STANDARD ANALYSIS

The data file in this section contains the duration of unemployment spell of up to one year of 200 individuals along with their age and log (income whilst unemployed). The latter variable is quarterly adjusted and therefore is time-dependent, and is also unavailable before 28 days. Analysis of larger datasets containing the data used in this Section can be found in [2] and [3]. The procedure starts as in the previous Section up to item (3). Proceed as follows:

(4) The data file for this Section is UNEMSAMP.DAT — an ordinary DOS file in the SURVIVAL format detailed in the Appendix — hence the appropriate selection now is Native file. This leads to the display

A: (Select this option for the default floppy disk)
B:
C:
D:
E:

(5) A list of files with the extension .DAT is now displayed and UNEMSAMP can be selected as before, returning control to the Data/ Input/Output menu.

(6) If no preliminary investigation or selection of the data is required, we can proceed to the modelling stage by successively quitting:

- Read
- Write
- Keyboard
- Amend
- Select
- Quit
(Data Input/Output)

- Input/Output
- Create
- Explore
- Remove
- Quit
(Data)

to produce

Data Model Output Housekeeping Exit (Root menu)

(7) Select Model and then Covariates in the Model menu to produce

AGE
UNEMPINC

Select them both by pressing Enter at their lines. After pressing the ESC key, control is returned to the Model menu.

(8) Select Grid now and input 28 followed by ESC to set the time origin correctly. This is followed by the option Equal-Lengths from the Grid menu. To set weekly intervals for the first semester and fortnightly for the second semester enter 7 (length of interval) and 22 (number of intervals) followed by the ESC key. As this does not cover the time span of the data, another line is presented for input of 14 (length of interval) and 20 (number of interval) followed by ESC key. Since the grid now covers the time span of the data, control is automatically returned to the Model menu.

(9) Select Prior and input the following moments:

- BASELINE hazard: mean $-1$, standard deviation 1.
- AGE effect: mean $-0.025$, standard deviation 0.015.
- UNEMPINC effect: mean $-0.5$, standard deviation 0.25.

using the $\rightarrow$, $\leftarrow$, $\downarrow$ and $\uparrow$ keys. After completing the input, press the ESC key to return control to the Model menu.

(10) Finally, select Evolution followed by the Additive option. Input now 7 to set the additive variance matrix over a period of a week and press the ESC key. Input values $0.01$, $1E - 5$ and $0.001$ successively followed by Enter at the positions $(1,1)$, $(2,2)$ and $(3,3)$ respectively using the $\rightarrow$, $\leftarrow$, $\downarrow$ and $\uparrow$ keys to move. This sets a diagonal variance matrix with diagonal elements $10^{-2}$, $10^{-4}$ and $10^{-3}$.
(11) The model is completed and its implications can be assessed by observing the prediction it implies for given individuals. To do that, quit the Model menu, to return to the Root menu displaying

Data Model Fit Output Housekeeping Exit (Root menu)

and select the Output option displaying

Tables Review Predict Plotter Quit (Output)

(12) Select the option Predict and consider four individuals with covariates: 1-(20,3); 2-(20,3.912); 3-(40,3); and 4-(40,3.912). The values 3 and 3.912 correspond to weekly unemployment income of roughly £20 and £50. The option Posterior calculates the predictive distribution of the unemployment spells given the covariates values based only on the prior assumptions without use of the data. Once selected, the user is prompted for the number of cases. Inputting 4 followed by the ESC key leads to a spreadsheet display where the above covariates values can be entered. After calculations are performed, the following display appears

Survival Hazard Density Quit (Posterior)

Selection of the Survival option produces Figure 4. After quitting the interactive graphics mode, we can return to the Root menu using Quit successively.

(13) The model can be fitted to the data using the Fit option of the Root menu. A display shows the cycle of estimation giving an idea of the stage the analysis is at.

(14) After fitting is completed, the results of the analysis can be obtained with the Output menu now displaying

Trajectory Hazard Tables Likelihood Review Predict Quit (Output)

(15) The parametric trajectories can be plotted using the Trajectory option yielding

Smoothed Smoothed/On-line Quit (Trajectory)

Select Smoothed and press Enter at the BASELINE line to obtain Figure 5.a. This is a plot of the smoothed mean trajectory of the baseline parameter. After quitting the graphics mode, control returns to the Trajectory menu and the same procedure can be repeated with the other parameters to obtain Figures 5.b and 5.c respectively. Control is returned to the Output menu after quitting the Trajectory menu.

(16) The smoothed estimate of the hazard function can be obtained with the Hazard option. After inputting the number of cases followed by the ESC key and the value of the covariates for each case (as in Predict), the hazard function for each case is plotted. Figure 6 shows the output for the four cases considered in (12) above. After quitting the graphics mode, control is returned to the Output menu.

(17) Numerical summaries can be obtained with the Tables option. Table 4 is obtained with the Smoothed option and Table 5 with the Correlation option. The Quit option returns control to the Output menu.

(18) Predictions based on the data are obtained with the Posterior option of the Predict menu. Figure 7 is the plot of the predictive density function for the same individuals in (12). It is obtained in a similar way as in (12) but for the selection of the Density instead of the Survival option in the Posterior menu.
### SMOOTHED ESTIMATES SUMMARY

<table>
<thead>
<tr>
<th>TIME (DAYS)</th>
<th>BASELINE HAZARD</th>
<th>AGE EFFECT</th>
<th>UNEMPINC EFFECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>-3.3852</td>
<td>-0.008</td>
<td>-0.400</td>
</tr>
<tr>
<td>42</td>
<td>-3.378</td>
<td>-0.008</td>
<td>-0.392</td>
</tr>
<tr>
<td>49</td>
<td>-3.393</td>
<td>-0.008</td>
<td>-0.389</td>
</tr>
<tr>
<td>56</td>
<td>-3.392</td>
<td>-0.007</td>
<td>-0.378</td>
</tr>
<tr>
<td>63</td>
<td>-3.392</td>
<td>-0.006</td>
<td>-0.366</td>
</tr>
<tr>
<td>70</td>
<td>-3.391</td>
<td>-0.005</td>
<td>-0.350</td>
</tr>
<tr>
<td>77</td>
<td>-3.389</td>
<td>-0.004</td>
<td>-0.330</td>
</tr>
<tr>
<td>84</td>
<td>-3.392</td>
<td>-0.005</td>
<td>-0.322</td>
</tr>
<tr>
<td>91</td>
<td>-3.394</td>
<td>-0.006</td>
<td>-0.309</td>
</tr>
<tr>
<td>98</td>
<td>-3.396</td>
<td>-0.007</td>
<td>-0.304</td>
</tr>
<tr>
<td>105</td>
<td>-4.002</td>
<td>-0.007</td>
<td>-0.291</td>
</tr>
<tr>
<td>112</td>
<td>-4.014</td>
<td>-0.007</td>
<td>-0.278</td>
</tr>
<tr>
<td>119</td>
<td>-4.014</td>
<td>-0.007</td>
<td>-0.261</td>
</tr>
<tr>
<td>126</td>
<td>-3.986</td>
<td>-0.006</td>
<td>-0.236</td>
</tr>
<tr>
<td>133</td>
<td>-3.993</td>
<td>-0.005</td>
<td>-0.223</td>
</tr>
<tr>
<td>140</td>
<td>-3.998</td>
<td>-0.004</td>
<td>-0.209</td>
</tr>
<tr>
<td>147</td>
<td>-4.022</td>
<td>-0.004</td>
<td>-0.201</td>
</tr>
<tr>
<td>154</td>
<td>-4.025</td>
<td>-0.004</td>
<td>-0.189</td>
</tr>
<tr>
<td>161</td>
<td>-4.036</td>
<td>-0.004</td>
<td>-0.183</td>
</tr>
<tr>
<td>168</td>
<td>-4.056</td>
<td>-0.005</td>
<td>-0.182</td>
</tr>
<tr>
<td>175</td>
<td>-4.080</td>
<td>-0.007</td>
<td>-0.181</td>
</tr>
<tr>
<td>182</td>
<td>-4.090</td>
<td>-0.008</td>
<td>-0.176</td>
</tr>
<tr>
<td>196</td>
<td>-4.105</td>
<td>-0.008</td>
<td>-0.164</td>
</tr>
<tr>
<td>210</td>
<td>-4.136</td>
<td>-0.007</td>
<td>-0.157</td>
</tr>
<tr>
<td>224</td>
<td>-4.196</td>
<td>-0.008</td>
<td>-0.164</td>
</tr>
<tr>
<td>238</td>
<td>-4.204</td>
<td>-0.007</td>
<td>-0.156</td>
</tr>
<tr>
<td>252</td>
<td>-4.249</td>
<td>-0.007</td>
<td>-0.162</td>
</tr>
<tr>
<td>266</td>
<td>-4.337</td>
<td>-0.009</td>
<td>-0.183</td>
</tr>
<tr>
<td>280</td>
<td>-4.417</td>
<td>-0.011</td>
<td>-0.205</td>
</tr>
<tr>
<td>294</td>
<td>-4.433</td>
<td>-0.012</td>
<td>-0.206</td>
</tr>
<tr>
<td>308</td>
<td>-4.422</td>
<td>-0.011</td>
<td>-0.200</td>
</tr>
<tr>
<td>322</td>
<td>-4.405</td>
<td>-0.010</td>
<td>-0.194</td>
</tr>
<tr>
<td>336</td>
<td>-4.420</td>
<td>-0.011</td>
<td>-0.198</td>
</tr>
<tr>
<td>350</td>
<td>-4.436</td>
<td>-0.011</td>
<td>-0.204</td>
</tr>
<tr>
<td>364</td>
<td>-4.434</td>
<td>-0.011</td>
<td>-0.202</td>
</tr>
<tr>
<td>365</td>
<td>-4.434</td>
<td>-0.011</td>
<td>-0.202</td>
</tr>
</tbody>
</table>

### TABLE 5

---

**POSTERIOR CORRELATION MATRIX**

---

Components

<table>
<thead>
<tr>
<th></th>
<th>BASELINE</th>
<th>AGE</th>
<th>UNEMPINC</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASELINE</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>-0.266</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>UNEMPINC</td>
<td>-0.550</td>
<td>-0.401</td>
<td>1.000</td>
</tr>
</tbody>
</table>
14. DATA CLASSIFIED BY FACTORS

SURVIVAL handles factor covariates, namely those that classify the survival times into two or more levels, their treatment in data management and model specification being slightly different to that of quantitative variables. To illustrate the handling of factors, consider the dataset in the NASOCANC file. This concerns the survival times of patients suffering from nasopharynx cancer, the data being more fully described and analysed in [6]. NASOCANC contains data on survival times and various covariates for each of 181 patients. A simple, illustrative analysis is performed here on only the male population and takes one of the tumor measurements as a classifying factor. The analysis starts as in previous Sections up to item (3), then proceeds as follows:

(4) The data file for this Section is NASOCANC - an APL component file - and the appropriate selection now is APL file leading to the display

A:(Select this option for the default floppy disk)
B:
C:
D:
E:

(5) A list of files with the extension .ASF is displayed and NASOCANC can be selected as before, returning control to the Data Input/Output menu.

(6) Selection of males only is performed using the Select option:

<table>
<thead>
<tr>
<th>Index</th>
<th>Times</th>
<th>Deaths</th>
<th>Covariate</th>
<th>Quit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Select)</td>
</tr>
</tbody>
</table>

Selection of Covariate will display a list of covariates; in this dataset, these are SEX, AGE, TUMOR1, TUMOR2 and DOSEL. Pressing Enter with SEX highlighted will prompt for upper and lower limits of the selection. Since males have SEX=0 and females SEX=1 any interval including 0 and excluding 1, such as 0 and .5, selects males and removes females from the dataset in memory.

(7) If no further selection is required, we can return to the Data menu by successively quitting at the Select and Data Input/Output menus to produce

Input/Output Create Explore Remove Quit (Data)

(8) Select Create which prompts for a variable name and formula to which we input

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>FTUMOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td>TUMOR2-1</td>
</tr>
</tbody>
</table>

followed by the ESC key. Since values of TUMOR2 are 1, 2, 3 and 4, the values of FTUMOR are 0, 1, 2 and 3. (A formula like TUMOR2 GROUP 1.5 2.5 3.5 would have produced exactly the same values for FTUMOR). SURVIVAL then prompts for confirmation of this created variable as a factor which is done by pressing Enter returning control to the Data menu.

(9) To display scatterplots of the data, select Explore and then X-Y in the Explore menu. As before, pressing Enter with TIMES highlighted and then subsequently with FTUMOR highlighted produces Figure 8.a. Quit from interactive graphics mode and select X-Y again. Repeating the previous selection procedure with TIMES and AGE produces Figure 8.b.

(10) Quit successively from interactive graphics and the Explore and Data menus to return
control to the Root menu. Select Model and then Covariates which produces the vertical display

SEX
AGE
TUMOR1
TUMOR2
DOSEL
FTUMOR

Pressing Enter at the AGE and FTUMOR lines followed by the ESC key selects them as covariates for the model and returns control to the Model menu.

(11) Select Grid, followed by the ESC key to identify the time origin as zero, and then select the Equal-Lengths option. Enter 12 and 15 followed by ESC to set 15 yearly intervals and return control to the Model menu.

(12) Select Prior and input the following moments:
- BASELINE hazard: mean -4, standard deviation 1.
- AGE effect: mean .5, standard deviation .25.

Pressing ESC allows the completion of the prior specification with the moments associated with the factor FTUMOR:
- LEVEL 1 effect: mean .5, standard deviation .25.
- LEVEL 2 effect: mean 1, standard deviation .25.
- LEVEL 3 effect: mean 1.5, standard deviation .25.

(Note that, as with all factors, LEVEL 0 has zero effect, by convention). Pressing ESC again returns control to Model menu as there are no further factors included in the model.

(13) Select Evolution followed by the Additive option, input 12 and press ESC. Input values .005 along the diagonal to set a diagonal variance matrix over one year with diagonal elements equal to .005.

(14) Again we can follow the same route to the Predict option in the Output menu as in the previous Section to assess the effect of the prior specification in (12). Firstly, consider 4 individuals with covariates: 1-(0,0); 2-(0,1); 3-(0,2); 4-(0,3). Select the Preposterior option, input 4 to set the number of cases and press ESC. A spreadsheet is presented and the above values can be entered. After pressing ESC and the Survival option, Figure 9 is produced. After quitting successively back to the Preposterior option, the effect of AGE can be assessed by repeating the procedure above with 3 individuals with covariates: 1-(1,0); 2-(0,0); 3-(1,0). Selection of the Survival option produces Figure 10. We can now quit successively back to the Root menu.

(15) The model can be fitted and its outputs analysed in the Output menu. The smoothed parametric trajectories can be plotted with the Trajectory option followed by the Smoothed option. Selecting BASELINE produces Figure 11.a and subsequent selection of AGE produces Figure 11.b. Repeating the procedure again with FTUMOR presents the vertical display

1
2
3

Selection of line 1 produces Figure 11.c, subsequent selections of lines 2 and 3 produce Figures 11.d and 11.e.

(16) Predictions based on the data are obtained with the Posterior option of the Predict
menu. Figure 12.a (12.b) is the plot of the predictive survival (hazard) function corresponding to the 4 individuals initially considered in (14) and Figure 13.a (13.b) is the same plot for the 3 individuals considered later in (14).

(17) A comparison with the product-limit estimator is possible by returning to the Explore option in the Data menu. Select the PL option followed by the Factor option then press Enter with FTUMOR highlighted to produce Figure 14.
Appendix: Format of .DAT files

It is recommended that most datasets for use in SURVIVAL be written from SURVIVAL to an APL sequential file, suffixed .ASF, using the Write facility. Subsequent read and write operations will then be performed most efficiently. However, existing datasets in ASCII code on ordinary DOS files may be read into SURVIVAL after reformatting, and data can be written to such files to access the data outside of APL. The format of such files is described here. Existing non-SURVIVAL format files should be copied to a file with suffix .DAT and then edited into this format. SURVIVAL will then read in the data using Read.

The first few lines of a .DAT file describe the dataset, the same description will appear on screen throughout a SURVIVAL session using the dataset. The backslash character \ is used as a delimiter and so should not appear elsewhere unless as a delimiter. Note that any other suitable character will do in place of \. The standard format is as follows:

\ Reference to source of data, a brief descriptor
\ Timing interval, eg. days, weeks, months, ...
\ Number of variables
\ Names of variables (one in each line)
\ Number of observations
\ Data variables, with one row for each observation, columns displaying the observation index, the survival/failure time, censoring indicator and variables in the order listed above
\ Indicator of time-dependence of covariates (1 or 0)
\ Number of changes in the data set implied by time-dependent variables¹
\ Changed values of the variables for the applicable observations, columns displaying the observation index, the time of change and the changed values of the variables in the order listed above¹
\ end

As an illustration, consider the UNEMSAMP dataset of Section 13 as stored on a .DAT file. The full filename is then UNEMSAMP.DAT. The datafile in SURVIVAL format has the appearance given on the next page — note that only the data for the first ten individuals is displayed here for illustration.

¹These entries are only required if the time-dependence indicator is 1.
| 200 INDIVIDUALS, WITH AGES AND UNEMPLOYMENT INCOME DAYS |
| AGE |
| UNEMPINC |
| 200 |
|   1 |   59 |   1 |   27 |  2.839 |
|   2 |   90 |   1 |   55 |  3.412 |
|   3 |  120 |   1 |   34 |  4.174 |
|   4 |  162 |   1 |   44 |  3.932 |
|   5 |  164 |   1 |   33 |  3.838 |
|   6 |   43 |   1 |   64 |  4.092 |
|   7 |  365 |   0 |   31 |  3.870 |
|   8 |   71 |   1 |   57 |  3.573 |
|   9 |  160 |   1 |   62 |  4.469 |
|  10 |  323 |   1 |   34 |  4.090 |
| ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... |
| 264 |
|  3 |   91 |  34 |  4.160 |
|  4 |   91 |  44 |  3.769 |
|  5 |   91 |  33 |  3.862 |
|  7 |   91 |  31 |  3.884 |
|  9 |   91 |  62 |  4.355 |
| 10 |   91 |  34 |  4.088 |
| ... | ... | ... | ... |
| ... | ... | ... | ... |
| ... | ... | ... | ... |

The remainder of the initial and changed data follow in the same format.
Computer requirements

SURVIVAL is written in APL*PLUS/PC© to run on personal micro-computers such as IBM PC, Olivetti M24 (AT&T) and lookalikes. SURVIVAL is loaded as a workspace into APL in computer memory so that users must have the APL*PLUS/PC© language interpreter for their computer. (Note, however, that as SURVIVAL is menu driven, it is not necessary to have a facility for generating APL screen characters. Expert users may, as mentioned in the Introduction, like to interact with SURVIVAL in APL and then will, of course, need character generation). Release 6.3 of APL*PLUS/PC©, the latest available at time of writing, occupies almost 200K of memory, (earlier versions are rather smaller), and SURVIVAL takes up another 170K. A minimum of 520K is recommended for routine use of SURVIVAL to allow for the accommodation of datasets and working variables, and more is obviously needed for larger datasets and models. The package runs most efficiently from a fixed disk although can be used on machines having only single floppy disk drives. In the latter case, the default drive should contain the SURVIVAL disk (after loading APL into memory).

Acknowledgements

SURVIVAL was developed at Warwick University during 1986 - 1988. SURVIVAL owes much of its development to the BATS package (see reference [7]). In particular, the menu system, graphical and data handling functions have been developed from those in BATS. Ewart Shaw (Warwick University) provided original routines on which the extensive graphical facilities are based, and has continued to provide invaluable support on APL generally throughout the project.

The research and development work of Andy Pole was partially supported by the S.E.R.C. and that of Dani Gamerman was supported by CAPES and CNPq-Brazil.

© Copyright of STSC Corp., Berkeley, California, U.S.A.