Reprinted From

A Modern Course on

STATISTICAL DISTRIBUTIONS IN SCIENTIFIC WORK

Edited by

G. F. PATIL, S. KOTZ, AND J. K. ORD

Published by

D. Reidel Publishing Company
38 Papaperspad, P. O. Box 17, Dordrecht, Holland
306 Dartmouth Street, Boston, Massachusetts 02116
STATISTICAL DISTRIBUTIONS IN SCIENTIFIC WORK

Based on the NATO Advanced Study Institute
A Modern Course on Statistical Distributions in Scientific Work
and The International Conference on Characterizations
of Statistical Distributions With Applications

Held at
The University of Calgary, Calgary, Alberta, Canada
July 29-August 10, 1974

Sponsored by
International Statistical Institute
The Pennsylvania State University
The University of Calgary
Indian Statistical Institute

With the Support of
North Atlantic Treaty Organization
National Research Council of Canada
United States Army Research Office

DIRECTOR
G. P. Patil

SCIENTIFIC DIRECTORS
G. P. Patil, S. Kotz, J. K. Ord

JOINT DIRECTORS
E. G. Evans (Local Chairman), J. K. Wani, P. C. Consul

ADVISORS
T. Cacoullos, C. D. Kemp, L. Olkin
J. B. Douglas, I. Kotz, C. R. Rao
A. Haal, E. Lukacs, L. R. Shenton
W. L. Hurkness, L. J. Martin, D. A. Sprott
N. L. Johnson, W. Molensar, N. Teicher
J. E. Mosimann
STATISTICAL MODELLING: AN ALTERNATIVE VIEW

Keith Ord and G.P. Patil
Warwick University and The Pennsylvania State University

SUMMARY

Modelling is an interactive process involving the use of data to select a distribution as well as to examine its validity. The paper outlines the statistical requirements of such an approach, and indicates the extent to which these are available. Alternative classes of distribution from which a model could be chosen are described, and alternative methods of model selection considered.

KEY WORDS

Families of distributions; model selection; likelihood ratio; moment ratios; probability plotting; probability ratios.

1. INTRODUCTION

The approach to modelling suggested by many introductory statistical texts is usually of the form:

```
scheme A
    \rightarrow formulate the hypothesis
    \------------------------------
    \-----------------------------
    collect the data
    \------------------------------
    \-----------------------------
    test the hypothesis
```

Thus the examples presented are usually of the form "given a particular body of data, test the goodness of fit of the (blank)
distribution. The better texts suggest that this is not the end of the process and that the cycle should be repeated (dotted line in scheme A), albeit with a fresh set of data which is never available for textbook examples and often not available in practical situations either. Even if we accept the full version of scheme A, (at least) two important questions remain unanswered:

(i) how do we choose the right model to start with? (model selection)
(ii) if the selected model fits, why? (discrimination between chance mechanisms).

In an era when large simulation studies are becoming increasingly important, good descriptions of random inputs are crucial, and a poor choice of model may vitiate the experiment. Further, it is not enough to know that a distribution provides a reasonable fit if we are unable to say something about the underlying chance mechanism which gave rise to the data. A flexible family of distributions may satisfy the descriptive requirement, but only a detailed study of the chance mechanisms generating that family can tell us whether a resulting "good fit" is significant in scientific terms. Often our background information will serve to limit the choice of distribution to a reasonable sub-class, as for the failure time problem discussed by Shapiro (1975) in this volume. Therefore, it seems reasonable to go as far as our a priori reasoning will take us and then to use the data to help us to select a model from the remaining sub-class. In summary, we might suggest the alternate procedure

scheme B formulate the initial model (class of distributions) 
  collect data
  select the model from the sub-class (particular distribution)
  test the goodness of fit of that distribution,
  formulate hypothesis (ascertain chance mechanism)

at least as the initial approach to a problem, with subsequent collection of data and remodelling where possible.

The purpose of such a scheme is to recognize a good fitting distribution and then to search among the possible generating mechanisms for the most appropriate explanation of the process. It may be, of course, that a "good fit" is all that is required (as in some Monte Carlo studies) but usually this final stage will be crucial. To make scheme B operational, we need both a systematic study of chance mechanisms [as in the papers by Patil and Boswell (1975), Patil, Boswell and Friday (1975) in this volume] and the means to discriminate between alternative chance
mechanisms which lead to the same distribution(s). However, the requirements do not stop there, as we also require flexible classes of distributions as preliminary models, and extended goodness-of-fit procedures which allow for model selection within a class prior to examining the individual model.

We recognize that the process described here is a departure from the conventional scheme A and has, in some respects, a weaker philosophical basis. However, we feel that the approach is closer in spirit to the way scientists try to develop their subjects, and we feel that it is, at least, worthy of more detailed consideration.

In section 2 we describe various classes of distributions that could form the basis for such an approach, while in section 3 different approaches to model selection are considered, although this topic is likely to remain an art form for some time to come. Finally, a simple example is described to demonstrate broad features of the approach. Work upon different chance mechanisms leading to certain classes of distributions is continuing, and the paper by Kemp and Kemp (1975) should be consulted in this context. Problems of discrimination between chance mechanisms will not be explicitly discussed here; however, the pioneering work of Bates and Neyman (1952) and Bates (1955) should be noted.

2. DIFFERENT FAMILIES AND THEIR HISTORY

Perhaps the most famous class of distributions is the Pearson family, defined by the differential equation for \( f(x) \), the probability density function,

\[
\frac{d \ln f(x)}{dx} = (a-x)/ \left( b_0 + b_1 x + b_2 x^2 \right)
\]

(1)

where \( a, b_0, b_1 \) and \( b_2 \) are parameters and the range of \( x \) may be finite or infinite depending upon the parameter values. This class is centred on the normal and includes the beta, gamma, F, reciprocal gamma and Student's t along with various special cases and the mystical Type IV distribution. Full listings are available, for example, in Johnson and Kotz (1970, pp.10-13) or in Ord (1972, pp.6-7); these books are subsequently referred to as J2 and O respectively. Pearson's approach was that of curve fitting and motivation for the model selected was pure. In their exchange upon the relative merits of the methods of moments and maximum likelihood, it is noteworthy that both Fisher and Pearson used the third and fourth moments to select the model (O, pp.10-12), thus the goodness of fit of the model fitted by ML could not be justified by traditional means.

The Johnson family (J2, pp.22-27, O, pp.36-43) was based upon
translations, or transformations, of the normal curve, considering the four possibilities

\[ y = x \text{ (normal); } \beta x, \text{ (lognormal); } \tanh^{-1}(2x-1), \]

(type \( S_0 \); \( \sinh^{-1}x \text{ (type } S_1) \)) \hspace{1cm} (2)

where \( y \) is taken to be normally distributed. These four types match the Pearson system in flexibility, as measured by coverage of the \( \beta_1, \beta_2 \) plane (see O, p. 37), where \( \beta_1 = \mu_1^2/\mu_2^2 \), \( \beta_2 = \mu_4/\mu_2^2 \) and \( \mu_4 \) denotes the 4th moment about the mean. Also, they require only standard functions and a normal random number generator for Monte Carlo work.

A third family is that of Burr (JK2, pp. 30-1, 0, pp. 43-4) based upon the distribution function

\[ f(x) = 1 - (1+x^c)^{-k} \quad x \geq 0, \quad c, k > 0 \]

\[ = 0 \quad x < 0. \] \hspace{1cm} (3)

Equation (3) is readily inverted to give the percentile function

\[ x = \left( (1-F)^{-1/k} - 1 \right)^{1/c} \] \hspace{1cm} (4)

so that only a uniform random number generator is required. Unfortunately, the present Burr system does not cover the whole \( \beta_1, \beta_2 \) plane, being particularly deficient for the area occupied by \( J \) and \( U \) shaped beta distributions. However, the original Burr scheme was developed from the differential equation

\[ \frac{dP}{dx} = F(1-F) g(x), \] \hspace{1cm} (5)

centering on the logistic when \( g(x) = x \), so that a more general scheme along these lines remains a possibility.

More recently, the percentile function has been used to define the useful Tukey lambda family (see Fiszberg, 1975). This family has the advantage of very rapid computer implementation, but it also fails to cover the \( J \) and \( U \) shaped regions of the beta distribution.

All the families mentioned so far refer to continuous variates, and there is generally a paucity of such developments for count data. One example is that of Ord (1967a) which uses the difference equation

\[ f(x) - f(x-1) = (a-x) f(x-1)/(b_0 + b_1x + b_2x(x-1)), \]

where \( a \) and the \( b_i \) are parameters. Equation (6) includes the
standard 'urn schema' sampling distributions. For further
details see Johnson and Kotz (1969, pp.36-39) or 0, chapter 5.
This family relates closely to the class of hypergeometric series
distributions (Kemp and Kemp, 1956, 1975). The $B_1$, $B_2$ plane for
this family does not partition into non-overlapping regions, but
the $S$, $I$ plane (where $S = \mu G/\mu I$, $I = \mu G/\mu I$ and $\mu I$ is the mean)
is more useful, as shown in 0, (p.98).

Several other families have appeared in the literature and
the selection in this paper is a subjective one based upon the
criteria of flexibility and ease of computer application. For
example, the whole class of series expansions has been ignored,
see JK2, pp.16-22 and 0, pp.25-35.

3. METHODS OF MODEL SELECTION

It is apparent from section 2 that one approach to model
selection is to use the $B_1$, $B_2$ or $S$, $I$ charts. Subsequent para-
meter estimation could then be carried out by maximum likelihood
or another efficient method, while recognizing that the usual
critical points for goodness of fit tests would be no longer valid.

An alternative approach is to use probability plots to examine
a series of alternatives, as done in the context of re-
liability by Shapiro (1975). The introduction of formal testing procedures
may be misleading unless the investigator can draw up a preferred
listing, as the lack of symmetry in the testing framework may not
be justified.

A different plotting approach is that considered by Ord
(1676) and Gart (1970). For the discrete Pearson family given
in (6), we find $b_2 = 0$ for most distributions of interest defined
on the non-negative integers, and we may define

$$u(x) = (a+b_1-1)/((1-b_2)x)/((b_1-1) + b_2 x).$$

When $b_2 = 0$, this reduces to $u(x) = c_0 + c_1 x$ and $c_1 > 0$
for the negative binomial, Poisson and binomial distributions
respectively. Other typical shapes are given in 0, (p.104). Model
reselection proceeds by plotting the empirical $u(x)$ values.
Greater stability may be achieved by smoothing, $v(x+1) = (u(x) +
2u(x+1))/2$ for example. Ratios based on $f(x+1) \leq 5$ or thereabouts
tend to be unreliable and subject to large sampling fluctuations.

Hinz and Garland (1957) have used functions of probability
ratios and of factorial cumulant ratios as a basis for initial
selection between contagious distributions. They then used these
functions to develop minimum chi-square estimators which have a
good performance in none parts of the parameter space.
A similar approach may be tried for continuous distributions. We discuss two different approaches relating to grouped and ungrouped observations respectively.

(i) Grouped data. Let \( p_j \) denote the proportion of observations which land in the \( j \)-th interval with limits \([x_{j-1}, x_j] \) and width \( h_j \). A smoothed estimate of the density \( f(x_j) \) at \( x_j \) is \( \frac{1}{2}(p_{j-1} + p_j) \) where \( p_j \) = \( p_j / h_j \). Then, an estimate of \( d \ln f/\text{dx} \) at \( x_j \) is given by

\[
w(x_j) = \frac{4(p_{j-1}^2 - p_{j-1})/(h_{j-1} + h_j)(p_{j-1}^2 + p_j^2)}{w(x_j)} = \frac{4(p_{j-1}^2 - p_{j-1})/(h_{j-1} + h_j)(p_{j-1}^2 + p_j^2)}{w(x_j)}.
\]

while we would substitute \( x_j \) on the right hand side of equation (1) if the Pearson family was being considered.

(ii) Ungrouped data. A non-parametric estimator of the density function at \( x_j \) such as

\[
p_j(h) = \frac{\text{number of observations in interval } [x_{j-h}, x_j+h]}{2h} = m(x_j, h)/2h, \text{ say,}
\]

may be used. Similarly, we could estimate \( d \ln f/\text{dx} \) at \( x_j \) by

\[
w(x_j) = \frac{m(x_j + \frac{1}{2}h, \frac{1}{2}h) - m(x_j - \frac{1}{2}h, \frac{1}{2}h)}{h p_j(h)}.
\]

The terms \( w(x) \) have the theoretical form \( w(x) = (x-y)/\sigma^2 \) for the normal distribution. For distributions defined on the positive half-line the form \( w(x) = x w(x) \) is more useful. The form of \( w(x) \) for some well-known distributions is as follows (\( c_0, c_1 \) and \( c_2 \) represent positive constants):

\[
\begin{align*}
c_0 - c_1 x & \quad \text{(gamma)} \\
- c_1 x & \quad \text{exponential) } \\
- c_0 + c_1 x & \quad \text{(type V or reciprocal gamma) } \\
(c_0 - 1) + (c_1 - 1) x / (1-x) & \quad \text{(beta of first kind on } [0, 1]) \\
(c_0 - 1) - c_2 x / (c_2 + x) & \quad \text{(beta of second kind or F) } \\
- c_0 + c_2 x & \quad \text{(lognormal) } \\
- c_0 + c_0 c_1 / (c_1 + x^2) & \quad \text{(Student's t)}
\end{align*}
\]

At this level of rigour little will be lost by using the extreme order statistics to define the range. Prior transformations for \( x \) could then be used to compare the Johnson types, using the normal. Again the \( u_j \) might be smoothed by pairwise averaging.
STATISTICAL MODELLING: AN ALTERNATIVE VIEW

For reliability problems, we could plot the observed and theoretical hazard rate functions \( p_j(h)/(1-F_j) \) against \( g(x_j) \) for different models.

3.1. Use of the Likelihood ratio

At a more formal level, we might adopt the approach of Cox (1961, 1962) in comparing separate families of hypotheses, see also Atkinson (1970). Their general approach is to consider the alternate densities \( f(x, \alpha) \) and \( g(x, \beta) \), where \( \alpha \) and \( \beta \) may be vectors, and a composite version such as

\[
h(x) = h(x, \lambda, \alpha, \beta) = \lambda f(x, \alpha) + (1-\lambda) g(x, \beta) \tag{11}
\]

or

\[
h(x) = (f(x, \alpha))^\lambda (g(x, \beta))^{1-\lambda}. \tag{12}
\]

They then derive likelihood ratio (LR) test procedures, but with an asymmetry that either \( f \) or \( g \) must be the null hypothesis. In terms of discrimination between models, a symmetric criterion is more appropriate (as noted by Cox (1962) and by Barnett in the discussion on Atkinson's paper). Suppose that \( f, g \) and \( h \) have \( a, b \) and \( c \) unspecified parameters respectively. Thus, for (11) or (12), \( c = a + b + 1 \), but more general forms for \( h \) might be considered. The LR criterion produces the log-likelihood ratio, \( \Delta(f, h) \), say, and \(-2\Delta(f, h)\) is distributed asymptotically as \( \chi^2 \) with \((c-a)\) degrees of freedom (or \((c-b)\) for \( g \)). For discrimination purposes, one could choose the model with density function \( f_j \), which maximized

\[
h(f_j, h) = \frac{1}{2}(c - a_j)
\]

or equivalently

\[
\log L(f_j) - \frac{1}{2} a_j, \tag{13}
\]

where \( L(f_j) \) is the likelihood function for \( f_j \). The composite model \( h \) does not appear in (13). This approach is essentially that of support tests put forward by Edwards (1971, chapter 9). The procedure is well defined and objective once the class of alternatives is specified, and allows the investigator to test whether a more complex model is worthwhile. An increase in the factor \( \frac{1}{2} \) in (13) would allow a built-in bias against more complex models in much the same way one can establish variable thresholds in stepwise regression analysis. The drawback of such procedures is simply that of the effort involved; all models must be fitted before a choice can be made.

3.2. Example
The time elapsed between marriage and the birth of the first child for 251 mothers is recorded in Table 1. The u<sub>q</sub> and v<sub>j</sub> ratios suggest the gamma or exponential model. The exponential model has intuitive appeal as indicative of random arrivals (the data relate to a pre-birth-control era). The expected frequencies in Table 1 show an excess in the first class, suggesting a slower rate of arrivals in the very early months. The observed hazard rates also reflect the higher risk in the second and third years of marriage. The yields figures in the last row for u and v probably reflect premature closure of the upper interval. The real deficiency of the model is that it does not reflect variable entry times into 'at risk' situations, for which a mixture model would be more appropriate. Indeed, ignoring the data for the first year provides a closer fit, but still fails to reflect the second year peak. Despite the simplicity of the suggested model, it was overlooked in the original analysis.

### Table 1: Number of wives in age group 30-34 tabulated according to the number of years between marriage and the birth of the first child (Elderton and Johnson, 1963, p.79)

<table>
<thead>
<tr>
<th>Years between</th>
<th>Number</th>
<th>u(x&lt;sub&gt;j&lt;/sub&gt;)</th>
<th>v(x&lt;sub&gt;j&lt;/sub&gt;)</th>
<th>expected frequencies (exponential)</th>
<th>observed hazard rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>marriage and birth of first child, y</td>
<td>wives</td>
<td>(x&lt;sub&gt;j&lt;/sub&gt; = y&lt;sub&gt;j&lt;/sub&gt;-1/2)</td>
<td>(x&lt;sub&gt;j&lt;/sub&gt; = y&lt;sub&gt;j&lt;/sub&gt;-1/2)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>0-1</td>
<td>60</td>
<td>-</td>
<td>-</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>98</td>
<td>1</td>
<td>1</td>
<td>1.24</td>
<td>1.88</td>
</tr>
<tr>
<td>2-3</td>
<td>55</td>
<td>2</td>
<td>2</td>
<td>-1.33</td>
<td>-0.61</td>
</tr>
<tr>
<td>3-4</td>
<td>21</td>
<td>4</td>
<td>4</td>
<td>-2.72</td>
<td>-2.02</td>
</tr>
<tr>
<td>4-5</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>-1.33</td>
<td>-2.02</td>
</tr>
<tr>
<td>5-6</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>-4.24</td>
<td>-2.78</td>
</tr>
<tr>
<td>6-7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-5.33&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-4.63&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>7-8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6.33&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.05&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Total</td>
<td>251</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* class frequencies too small to place any reliance upon these numbers
(1) first arrivals assumed after 8 months or more of marriage
(2) data for first year omitted

**Acknowledgement**

This work was made possible by a research grant from the NATO Scientific Affairs Division, to whom the authors are indebted.
REFERENCES


Patil, G.P., Boswell, M.T. and Friday, D. (1975). Chance mechanisms in the computer generation of random variables. (This volume)

Ramberg, J. (1975). A probability distribution with application to Monte Carlo simulation studies. (This volume)

Shapiro, S.S. (1975). Application of statistical distributions to engineering problems. (This volume)
A Modern Course on

STATISTICAL DISTRIBUTIONS IN SCIENTIFIC WORK

Contents

VOLUME 1: MODELS AND STRUCTURES

Section

1. INAUGURAL ADDRESS - G. P. Patil

2. POWER SERIES AND RELATED FAMILIES - S. W. Joshi, T. Cacoullos,
   A. W. Kemp and C. B. Kemp, P. C. Consul and L. R. Shenton,
   J. Gurland and R. Tripathi, G. P. Patil and V. Seshadri

3. RECENT TRENDS IN UNIVARIATE MODELS - S. J. Press, J.
   Behboodian, K. V. Mardia, F. McNulty, J. R. Huynen, and
   E. Hansen, T. P. Hettmansperger and M. A. Keenan, A. L. Rukhin

4. MOMENTS-RELATED PROBLEMS - M. S. Ramanujan, C. C. Heyde,
   W. L. Harkness

5. LIMIT DISTRIBUTIONS AND PROCESSES - M. Harris and A. P. Soma,

6. MULTIVARIATE CONCEPTS AND MODELS - S. Kotz, K. Jogdeo,
   J. J. J. Roux, M. Siotani

7. CERTAIN MULTIVARIATE DISTRIBUTIONS - A. Dassauchoy and
   R. Berland, F. C. Durling, R. P. Gupta, C. G. Khatri,
   J. Tiago de Oliveira

8. SAMPLING DISTRIBUTIONS AND TRANSFORMATIONS - M. Shaked,
   K. O. Bowman and W. E. Dusenberry, V. B. Walkar, Y. Prishman,
   J. Ogawa

VOLUME 2: MODEL BUILDING AND MODEL SELECTION

1. MODELLING AND SIMULATION - J. K. Ord and G. P. Patil,
   G. P. Patil and M. Boswell, E. J. Dudewicz, C. F. Patil,
   M. Boswell, and D. Friday, J. S. Ramberg

2. MODEL IDENTIFICATION AND DISCRIMINATION - R. Srinivasan and
   G. E. Antle, M. Georgos, V. Seshadri, and M. Voldovsky,
   S. K. Katti and K. McDonald, J. Gurland and R. C. Bohiya,
   M. L. Tiku, J. J. Cart

3. MODELS IN THE SOCIAL SCIENCES AND MANAGEMENT - J. K. Ord,
   J. K. Ord, R. W. Reesek, C. Chatfield
4. MODELS IN THE PHYSICAL AND BIOMEDICAL SCIENCES - J. E.
Moolman, J. E. Moolman, M. E. Wise, S. Talwalkar, D. M.
Schultz, S. S. Shapiro, A. G. Laurent, B. Elvers
5. MODELS IN THE ENVIRONMENTAL SCIENCES - M. F. Dacey, D. V.
Gokhale, G. Ramachandran, W. G. Warren
6. A MODERN COURSE ON STATISTICAL DISTRIBUTIONS - Participants

VOLUME 3: CHARACTERIZATIONS AND APPLICATIONS

1. LINNICK MEMORIAL INAUGURAL LECTURE - C. R. Rao
2. MATHEMATICAL TOOLS FOR CHARACTERIZATION PROBLEMS - N. Lukacs,
J. Acsél, H. J. Rosseberg
3. CHARACTERIZATIONS USING ORDER STATISTICS - J. Galambos, J.
Govindarajulu, M. Ahsanullah, J. S. Huang and J. S. Hwang,
Z. GovindaRajula, J. S. Huang, and A. K. Mat E. Saleh
4. CHARACTERIZATIONS BY OTHER STATISTICAL PROPERTIES - A. L.
Klebanov, L. Bondesson, C. G. Kharif, B. Gyires
5. CHARACTERIZATIONS ON SPACES AND PROCESSES - I. I. Kelalski,
B. L. S. Prakasa Rao, E. Ahmad, E. Seneta
6. CHARACTERIZATION PROBLEMS FOR DAMAGED OBSERVATIONS - G. P.
Patil and M. V. Ratnaparkhi, R. C. Srivastava and J. Singh,
P. C. Consul, J. K. Ord
7. CHARACTERIZATIONS USING ENTROPY MEASURES AND RELATED PROBLEMS
- D. V. Gokhale, R. Shimizu, A. W. Kemp, J. Acsél
8. CHARACTERIZATIONS FOR DISCRETE DISTRIBUTIONS AND FAMILIES -
A. V. Gomzhe and G. P. Patil, A. N. Nevill and C. D. Kemp,
K. G. Janardan
9. CHARACTERIZATIONS FOR CONTINUOUS DISTRIBUTIONS AND FAMILIES -
K. V. Mardia, M. S. Bingham and K. V. Mardia, A. P. Basu and
H. W. Block, J. K. Wani and G. P. Patil

FOR YOUR LIBRARY COPY

These Proceedings of a Modern Course on Statistical Distribution provide carefully edited materials in three comprehensive volumes on related problems of topics of contemporary interest and concern, such as, Models and Structures, Model Building and Model Selection, and Characterizations and Applications. Every quantitative scientist and statistician should find them very useful and worthwhile for both research and instruction involving any aspects of distributions. Order your copy today.